DE GRUYTER

Mariusz Urbański, Mario Roy, Sara Munday

NON-INVERTIBLE DYNAMICAL SYSTEMS

VOLUME 1: ERGODIC THEORY - FINITE AND INFINITE, THERMODYNAMIC FORMALISM, SYMBOLIC DYNAMICS AND DISTANCE EXPANDING MAPS

EXPOSITIONS IN MATHEMATICS 69/1

Mariusz Urbański, Mario Roy, Sara Munday

Non-Invertible Dynamical Systems

De Gruyter Expositions in Mathematics

Volume 69/1

Mariusz Urbański, Mario Roy, Sara Munday

Non-Invertible Dynamical Systems

Volume 1: Ergodic Theory – Finite and Infinite, Thermodynamic Formalism, Symbolic Dynamics and Distance Expanding Maps

DE GRUYTER

Mathematics Subject Classification 2010
37A05, 37A25, 37A30, 37A35, 37A40, 37B10, 37B25, 37B40, 37B65, 37C05, 37C20, 37C40, 37D20, 37D35, 37E05, 37E10

Authors

Prof. Dr. Mariusz Urbański
University of North Texas
Department of Mathematics
1155 Union Circle #311430
Denton, TX 76203-5017
USA
mariusz@urbanskimath.com

Dr. Sara Munday
John Cabot University
Via della Lungara 233
00165 Rome
Italy
smunday@johncabot.edu

Prof. Dr. Mario Roy
York University
Glendon College
2275 Bayview Avenue
Toronto, M4N 3M6
Canada
mroy@glendon.yorku.ca

ISBN 978-3-11-070264-4
e-ISBN (PDF) 978-3-11-070268-2
e-ISBN (EPUB) 978-3-11-070275-0
ISSN 0938-6572

Library of Congress Control Number: 2021940887

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Typesetting: VTeX UAB, Lithuania
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Mariusz Urbański dedicates this book to his wife, Irena.
À mes parents Thérèse et Jean-Guy, à ma famille et à mes amis, sans qui ce livre n'aurait pu voir la vie... du fond du coeur, merci! Mario

Preface

Dynamical systems and ergodic theory is a rapidly evolving field of mathematics with a large variety of subfields, which use advanced methods from virtually all areas of mathematics. These subfields comprise but are by no means limited to: abstract ergodic theory, topological dynamical systems, symbolic dynamical systems, smooth dynamical systems, holomorphic/complex dynamical systems, conformal dynamical systems, one-dimensional dynamical systems, hyperbolic dynamical systems, expanding dynamical systems, thermodynamic formalism, geodesic flows, Hamiltonian systems, KAM theory, billiards, algebraic dynamical systems, iterated function systems, group actions, and random dynamical systems.

All of these branches of dynamical systems are mutually intertwined in many involved ways. Each of these branches nonetheless also has its own unique methods and techniques, in particular embracing methods which arise from the fields of mathematics the branch is closely related to. For example, complex dynamics borrows advanced methods from complex analysis, both of one and several variables; geodesic flows utilize methods from differential geometry; and abstract ergodic theory and thermodynamic formalism rely heavily on measure theory and functional analysis.

Indeed, it is truly fascinating how large the field of dynamical systems is and how many branches of mathematics it overlaps with. In this book, we focus on some selected subfields of dynamical systems, primarily noninvertible ones.

In the first volume, we give introductory accounts of topological dynamical systems acting on compact metrizable spaces, of finite-state symbolic dynamical systems, and of abstract ergodic theory of measure-theoretic dynamical systems acting on probability measure spaces, the latter including the metric entropy theory of Kolmogorov and Sinai. More advanced topics include infinite ergodic theory, general thermodynamic formalism, and topological entropy and pressure. This volume also includes a treatment of several classes of dynamical systems, which are interesting on their own and will be studied at greater length in the second volume: we provide a fairly detailed account of distance expanding maps and discuss Shub expanding endomorphisms, expansive maps, and homeomorphisms and diffeomorphisms of the circle.

The second volume is somewhat more advanced and specialized. It opens with a systematic account of thermodynamic formalism of Hölder continuous potentials for open transitive distance expanding systems. One chapter comprises no dynamics but rather is a concise account of fractal geometry, treated from the point of view of dynamical systems. Both of these accounts are later used to study conformal expanding repellers. Another topic exposed at length is that of thermodynamic formalism of countable-state subshifts of finite type. Relying on this latter, the theory of conformal graph directed Markov systems, with their special subclass of conformal iterated function systems, is described. Here, in a similar way to the treatment of conformal expanding repellers, the main focus is on Bowen's formula for the Hausdorff dimension

https://doi.org/10.1515/9783110702682-201

of the limit set and multifractal analysis. A rather short examination of Lasota–Yorke maps of an interval is also included in this second volume.

The third volume is entirely devoted to the study of the dynamics, ergodic theory, thermodynamic formalism, and fractal geometry of rational functions of the Riemann sphere. We present a fairly complete account of classical as well as more advanced topological theory of Fatou and Julia sets. Nevertheless, primary emphasis is placed on measurable dynamics generated by rational functions and fractal geometry of their Julia sets. These include the thermodynamic formalism of Hölder continuous potentials with pressure gaps, the theory of Sullivan's conformal measures, invariant measures and their dimensions, entropy, and Lyapunov exponents. We further examine in detail the classes of expanding, subexpanding, and parabolic rational functions. We also provide, with proofs, several of the fundamental tools from complex analysis that are used in complex dynamics. These comprise Montel's Theorem, Koebe's Distortion Theorems and Riemann–Hurwitz formulas, with their ramifications.

In virtually each chapter of this book, we describe a large number of concrete selected examples illustrating the theory and serving as examples in other chapters. Also, each chapter of the book is supplied with a number of exercises. These vary in difficulty, from very easy ones asking to verify fairly straightforward logical steps to more advanced ones enhancing largely the theory developed in the chapter.

This book originated from the graduate lectures Mariusz Urbański delivered at the University of North Texas in the years 2005–2010 and that Sara Munday took notes of. With the involvement of Mario Roy, the book evolved and grew over many years. The last 2 years (2020 and 2021) of its writing were most dramatic and challenging because of the COVID-19 pandemic. Our book borrows widely from many sources including the books [41, 47, 57]. We nevertheless tried to keep it as self-contained as possible, avoiding to refer the reader too often to specific results from special papers or books. Toward this end, an appendix comprising classical results, mostly from measure theory, functional analysis and complex analysis, is included. The book covers quite a many topics treated with various degrees of completeness, none of which are fully exhausted because of their sheer largeness and their continuous dynamical growth.

List of Figures

Introduction to Volume 1

In the first volume of this book, we give introductory accounts of topological dynamical systems acting on compact metrizable spaces, of finite-alphabet symbolic systems, and of ergodic theory of measure-theoretic dynamical systems acting on probability spaces, the latter including the metric entropy theory of Kolmogorov and Sinai. More advanced topics include infinite ergodic theory, general thermodynamic formalism, and topological entropy and pressure. This volume also includes a treatment of several classes of dynamical systems, which are interesting on their own and will be studied at greater length in the second volume: we provide a fairly detailed account of distance expanding maps and discuss Shub expanding endomorphisms, positively expansive maps, and homeomorphisms and diffeomorphisms of the circle.

We now describe the content of each chapter of this first volume in more detail, including their mutual dependence and interrelations.

Chapter 1 – Dynamical systems

In the first few sections of Chapter 1, we introduce the basic concepts in the theory of topological dynamical systems: orbits, periodic points, preperiodic points, ω-limit sets, factors, and subsystems. In particular, we introduce the concept of topological conjugacy and identify the number of periodic points of any given period as a simple (topological conjugacy) invariant. We further examine the following invariants: minimality, transitivity, topological mixing, strong transitivity, and topological exactness. Finally, we provide the first two classes of examples, namely rotations on compact topological groups and some continuous maps on compact intervals.

Chapter 2 – Homeomorphisms of the circle

In Chapter 2, we temporarily step away from the general theory of dynamical systems to consider more specific examples. We investigate homeomorphisms of the unit circle and examine the notions of lift and rotation number for homeomorphisms. Then we study in more detail the subclass of diffeomorphisms of the unit circle. The main result of this chapter is Denjoy's theorem, which states that if a C^2 diffeomorphism has an irrational rotation number, then this diffeomorphism is a minimal system which is topologically conjugate to an irrational rotation.

Chapter 3 – Symbolic dynamics

In Chapter 3, we discuss symbolic dynamical systems. We treat them as objects in their own right, but later (in Chapter 4, among others) we apply the ideas developed here to more general systems. We restrict ourselves to the case of finitely many letters, as symbolic systems born out of finite alphabets give rise to systems acting on compact metrizable spaces. Nevertheless, note that in Chapter 17 of the second volume,

https://doi.org/10.1515/9783110702682-202

we will consider countable-alphabet symbolic dynamics. In Section 3.1, we introduce full shifts. In Section 3.2, we study subshifts of finite type and in particular the characterizations of topological transitivity and exactness in terms of the underlying matrix associated with such systems. Finally, in Section 3.3 we examine general subshifts of finite type.

Chapter 4 – Distance expanding maps

In Chapter 4, we define and give some examples of distance expanding maps. In Section 4.2, we study the properties of their local inverse branches. This is a way of dealing with the noninvertibility of these maps. In Section 4.3, we examine the all important concepts of pseudo-orbit and shadowing. In Section 4.4, we introduce the powerful concept of Markov partitions and establish their existence for open, distance expanding systems. We then show in Section 4.5 how to use Markov partitions to represent symbolically the dynamics of open, distance expanding systems. This is a beautiful application of the symbolic dynamics studied in Chapter 3. The final theorem of the chapter describes the properties of the coding map between the underlying compact metric space (the phase space) and some subshift of finite type (a symbolic space).

Chapter 5 – Expansive maps

In Chapter 5, we introduce the concept of expansiveness. Amidst the large variety of dynamical behaviors, which can be thought of as expansionary in some sense, expansiveness has turned out to be a rather weak but useful notion. Indeed, all distance expanding maps are expansive and so, more particularly, all subshifts over a finite alphabet are expansive. But expansiveness is not so far from expandingness, as we demonstrate in this chapter that every expansive system is in fact expanding with respect to some metric compatible with the topology. This means that many of the results proved in Chapter 4, such as the existence of Markov partitions and of a nice symbolic representation, the density of periodic points, the closing lemma, and the shadowing property, hold for all positively expansive maps. Nevertheless, expansiveness is weaker than expandingness, and we provide at the end of the chapter a class of expansive maps that are not distance expanding. Expansive maps are important for other reasons as well. One of them is that expansiveness is a topological conjugacy invariant. More crucially, the measure-theoretic entropy function is upper semicontinuous within that class of maps. In particular, all expansive maps admit a measure of maximal entropy and, more generally, equilibrium states under all continuous potentials (see Chapter 12).

Chapter 6 – Shub expanding endomorphisms

In Section 6.2, we give a systematic account of Shub's expanding endomorphisms. These maps constitute a large, beautiful subclass of distance expanding maps and are far-reaching generalizations of the expanding endomorphisms of the circle, which will be first introduced in Section 6.1. After a digression into albegraic topology, we establish in Section 6.4 that Shub expanding endomorphisms are structurally stable, form an open set in an appropriate topology of smooth maps, are topologically exact, have at least one fixed point as well as a dense set of periodic points, and their universal covering space is diffeomorphic to $\mathbb{R}^n$.

Chapter 7 – Topological entropy

In Chapter 7, we study the central notion of topological entropy, one of the most useful and widely-applicable topological invariant thus far discovered. It was introduced to dynamical systems by Adler, Konheim, and McAndrew in 1965. Their definition was motivated by Kolmogorov and Sinai's definition of metric/measure-theoretic entropy introduced less than a decade earlier. The topological entropy of a dynamical system, which we introduce in Section 7.2, is a nonnegative extended real number that measures the complexity of the system. Topological entropy is a topological conjugacy invariant but by no means a complete invariant. In Section 7.3, we treat at length Bowen's characterization of topological entropy in terms of separated and spanning sets. In Chapter 11, we will introduce and deal with topological pressure, which is a substantial generalization of topological entropy. Our approach to topological pressure will stem from and extend that for topological entropy. In this sense, this chapter can be viewed as a preparation to Chapter 11.

Chapter 8 – Ergodic theory

In Chapter 8, we move away from the study of purely topological dynamical systems to consider instead dynamical systems that come equipped with a measure. That is, instead of self-maps acting on compact metrizable spaces, we now ask that the self-maps act upon measure spaces. We introduce in Section 8.1 the basic object of study in ergodic theory, namely, invariant measures. We also prove Poincaré's recurrence theorem. Section 8.2 presents the notion of ergodicity and comprises a demonstration of Birkhoff's ergodic theorem. This theorem is one of the most fundamental results in ergodic theory. It is extremely useful in numerous applications. The class of ergodic measures for a given transformation is then studied in more detail. The penultimate Section 8.3 contains an introduction to various measure-theoretic mixing properties that a system may satisfy, and shows that ergodicity is a very weak form of mixing. In the final Section 8.4, Rokhlin's natural extension of any given dynamical system is described and the mixing properties of this extension are investigated.

Chapter 9 –Measure-theoretic entropy

In Chapter 9, we study the measure-theoretic entropy of a (probability) measure-preserving dynamical system, also known as metric entropy or Kolmogorov–Sinai metric entropy. It was introduced by A. Kolmogorov and Ya. Sinai in the late 1950s. Since then, its account has been presented in virtually every textbook on ergodic theory. Its introduction to dynamical systems was motivated by Ludwig Boltzmann's concept of entropy in statistical mechanics and Claude Shannon's work on information theory. We first study measurable partitions in Section 9.2. Then we examine the concepts of information and conditional information in Section 9.3. In Section 9.4, we finally define the metric entropy of a measure-preserving dynamical system. And in Section 9.5, we formulate and prove the full version of Shannon–McMillan–Breiman's characterization of metric entropy. Finally, in Section 9.6 we shed further light on the nature of entropy, by proving the Brin–Katok local entropy formula. Like the Shannon–McMillan–Breiman theorem, the Brin–Katok local entropy formula is very useful in applications.

Chapter 10 – Infinite invariant measures

In Chapter 10, we deal with measurable transformations preserving measures that are no longer assumed to be finite. The outlook is then substantially different than in the case of finite measures. In Section 10.1, we investigate in detail the notions of quasi-invariant measures, ergodicity, and conservativity. We also prove Halmos' recurrence theorem, which is a generalization of Poincaré's recurrence theorem for quasi-invariant measures that are not necessarily finite. In Section 10.2, we discuss first return times, first return maps, and induced systems. We further establish relations between invariant measures for the original transformation and the induced one. In Section 10.3, we study implications of Birkhoff's ergodic theorem for finite and infinite measure spaces. Among others, we demonstrate Hopf's ergodic theorem, which applies to measure-preserving transformations of σ-finite spaces. Finally, in Section 10.4, we seek a condition under which, given a quasi-invariant probability measure, one can construct a σ-finite invariant measure which is absolutely continuous with respect to the original measure. To this end, we introduce a class of transformations, called Martens maps, that have this feature and even more. In fact, these maps have the property that any quasi-invariant probability measure admits an equivalent σ-finite invariant one. Applications of these concepts and results can be found in Chapters 13–14 of the second volume and Chapters 29–32 of the third volume.

Chapter 11 – Topological pressure

Chapter 12 – The variational principle and equilibrium states

In the last two chapters of this first volume, we introduce and extensively deal with the fundamental concepts and results of thermodynamic formalism, including topological pressure, the variational principle, and equilibrium states. This topic has a continuation throughout the whole second volume, first and perhaps most notably, in the first chapter of that volume, which is devoted to the thermodynamic formalism of distance expanding maps and Hölder continuous potentials. It will be enriched by the seminal concepts of Gibbs states and transfer (Perron–Frobenius, Ruelle, Araki) operators.

Thermodynamic formalism originated in the late 1960s with the works of David Ruelle. The motivation for Ruelle came from statistical mechanics, particularly glass lattices. The foundations, classical concepts and theorems of thermodynamic formalism were developed throughout the 1970s by Ruelle, Rufus Bowen, Peter Walters, and Yakov Sinai.

In Chapter 11, we define and investigate the properties of topological pressure. Like topological entropy, this is a topological concept and a topological conjugacy invariant. We further give Bowen's characterization of pressure in terms of separated and spanning sets.

In Chapter 12, we relate topological pressure with metric entropy by proving the variational principle, the very cornerstone of thermodynamic formalism. This principle naturally leads to the concepts of equilibrium states and measures of maximal entropy. Among others, we show that under a continuous potential every expansive dynamical system admits an equilibrium state.

Contents

Volume 2

Volume 3

1 Dynamical systems

In the first few sections of this chapter, we introduce the basic concepts in the theory of topological dynamical systems: orbits, periodic points, preperiodic points, ω-limit sets, factors, and subsystems. In particular, we introduce in Section 1.2 the concept of topological conjugacy and identify the number of periodic points of any given period as a simple (topological conjugacy) invariant. In Section 1.5, we examine the following invariants: minimality, transitivity, topological mixing, strong transitivity, and topological exactness. Finally, in Section 1.6 we provide the first two classes of examples, namely rotations on compact topological groups and some continuous maps of compact intervals.

1.1 Basic definitions

Throughout this book, a (discrete) *topological dynamical system* is a continuous map $T : X \to X$ of a nonempty compact metrizable space X. When emphasis on a metric is desirable, we write (X, d). The study of a dynamical system consists of determining its long-term behaviors, also referred to as asymptotic behaviors. That is, if we denote by T^n the nth *iterate* of T, which is defined to be

$$T^n := \underbrace{T \circ \cdots \circ T}_{n \text{ times}},$$

in order to study a dynamical system T, we investigate the sequence of iterates $(T^n)_{n=0}^{\infty}$. The long-term behavior of a point $x \in X$ can be determined by looking at this sequence of iterates evaluated at the point x.

Definition 1.1.1. Let $x \in X$. The *forward orbit* of x under T is the set

$$\mathcal{O}_+(x) := \{T^n(x) : n \geq 0\}.$$

Moreover, the *backward orbit* of x is the set

$$\mathcal{O}_-(x) := \{T^{-n}(x) : n \geq 0\} = \{T^n(x) : n \leq 0\},$$

while the *full orbit* of x is the set

$$\mathcal{O}(x) := \{T^n(x) : n \in \mathbb{Z}\} = \mathcal{O}_-(x) \cup \mathcal{O}_+(x).$$

The simplest (forward) orbits that may be observed in a dynamical system are those that consist of only finitely many points. Among these are the orbits that are cyclic.

https://doi.org/10.1515/9783110702682-001

Definition 1.1.2. A point $x \in X$ is said to be *periodic* for a system T if

$$T^n(x) = x$$

for some $n \in \mathbb{N}$. Then n is called a *period* of x. The smallest period of a periodic point x is called the *prime period* of x. The set of all periodic points of period n for T shall be denoted by $\mathrm{Per}_n(T)$. In particular, if

$$T(x) = x$$

then x is called a *fixed point* for T. The set of all fixed points will be denoted by $\mathrm{Fix}(T)$. Hence, $\mathrm{Fix}(T) = \mathrm{Per}_1(T)$. Finally, we let $\mathrm{Per}(T) = \bigcup_{n=1}^{\infty} \mathrm{Per}_n(T)$ denote the set of all periodic points for T.

Example 1.1.3.

(a) Define the map $T : [0,1] \to [0,1]$ by setting

$$T(x) := \begin{cases} 2x & \text{if } x \in [0,1/2) \\ 2-2x & \text{if } x \in [1/2,1]. \end{cases}$$

This map is known in the literature as the *tent map*. Its graph, which makes clear the reasoning behind the name, is shown in Figure 1.1. The tent map has for fixed points $\mathrm{Fix}(T) = \{0, 2/3\}$ and has 2^n periodic points of period n for each $n \in \mathbb{N}$. These are given by

$$\mathrm{Per}_n(T) = \left\{0, \frac{k}{2^n-1}, \frac{k}{2^n+1}, \frac{2^n}{2^n+1} : k \in \{2,4,6,\dots,2^n-2\}\right\}.$$

These periodic points are the points of intersection of the graph of T^n with the diagonal line $y = x$.

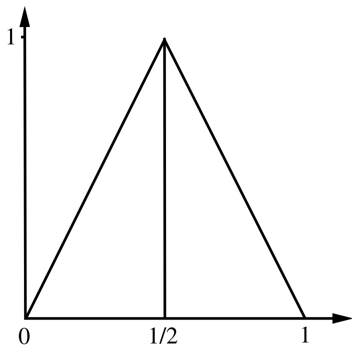

Figure 1.1: The tent map $T : [0,1] \to [0,1]$.

(b) Let $\mathbb{S}^1$ denote the unit circle, where $\mathbb{S}^1 := \mathbb{R}/\mathbb{Z}$, or, equivalently, $\mathbb{S}^1 := [0,1] \pmod 1$. Fix $m \in \mathbb{N}$ and define the map $T_m : \mathbb{S}^1 \to \mathbb{S}^1$ by setting $T_m(x) := mx \pmod 1$. One example of such a map is shown in Figure 1.2. The map T_m is simply a piecewise linear map that sends each interval $[i/m,(i+1)/m]$, for $0 \le i \le m-1$, onto $\mathbb{S}^1$. It can be expressed by the formula

$$T_m(x) = mx - i, \quad \forall x \in \left[\frac{i}{m}, \frac{i+1}{m}\right], \forall 0 \le i \le m-1.$$

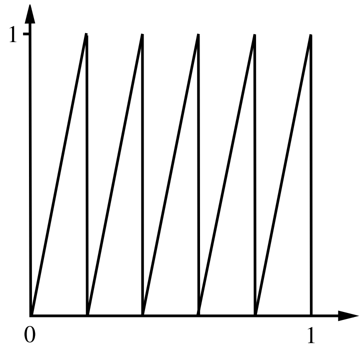

Figure 1.2: The map $T_m : [0,1] \to [0,1]$, where $m = 5$.

The map T_m has $m-1$ fixed points. They are the points of intersection of the graph of T_m with the diagonal line $y = x$. More precisely,

$$\mathrm{Fix}(T_m) = \left\{\frac{i}{m-1} : 0 \le i < m-1\right\}.$$

Similarly, it can be shown that the nth iterate T_m^n has $m^n - 1$ fixed points (see Exercise 1.7.1). We will return to this example later in the book, specifically in Chapters 4 and 9.

We now observe a general fact about convergent sequences of iterates of a point.

Lemma 1.1.4. *Let $T : X \to X$ be a dynamical system. Suppose that there exists $x \in X$ such that*

$$\lim_{n\to\infty} T^n(x) = y.$$

Then y is a fixed point for T.

Proof. Using the continuity of T, we obtain that

$$T(y) = T\Big(\lim_{n\to\infty} T^n(x)\Big) = \lim_{n\to\infty} T^{n+1}(x) = y. \qquad \square$$

Note that this fact applies only when the entire sequence of iterates converges. It does not generally hold for convergent subsequences.

Definition 1.1.5. A point $x \in X$ is said to be *preperiodic* for a system T if one of its (forward) iterates is a periodic point. That is, if there exists $k \in \mathbb{N}$ such that $T^k(x)$ is a periodic point. In other words, this means that there exists $n \in \mathbb{N}$ such that $T^{k+n}(x) = T^k(x)$.

The forward orbit $\mathcal{O}_+(x)$ is finite if and only if x is periodic or preperiodic. Equivalently, the sequence of forward iterates $(T^n(x))_{n=0}^{\infty}$ consist of mutually distinct points if and only if x is neither periodic nor preperiodic. Indeed, $\mathcal{O}_+(x)$ is infinite if and only if the sequence $(T^n(x))_{n=0}^{\infty}$ consist of mutually distinct points.

1.2 Topological conjugacy and structural stability

Suppose that we have two topological dynamical systems, $T : X \to X$ and $S : Y \to Y$. In this section, we describe a particular condition under which these two systems should be considered dynamically equivalent, that is, as dynamically "the same" in some sense. More precisely, we will establish when the orbits of two systems behave in the same way. Establishing an equivalence relation between dynamical systems can be extremely helpful, since it gives us the opportunity to apply our knowledge of systems we understand well to systems we have less information about.

Definition 1.2.1. Two dynamical systems $T : X \to X$ and $S : Y \to Y$ are said to be *topologically conjugate* if there exists a homeomorphism $h : X \to Y$, called a *conjugacy map*, such that

$$h \circ T = S \circ h.$$

In other words, T and S are topologically conjugate if there exists a homeomorphism h such that the following diagram commutes:

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[d, "h"'] & X \arrow[d, "h"] \\ Y \arrow[r, "S"'] & Y \end{tikzcd}$$

Remark 1.2.2.

(a) Topological conjugacy defines an equivalence relation on the space of all dynamical systems (see Exercise 1.7.6).

(b) If two dynamical systems T and S are topologically conjugate via a conjugacy map h, then all of their corresponding iterates are topologically conjugate by means of h. That is,

$$h \circ T^n = S^n \circ h, \quad \forall n \in \mathbb{N}.$$

Therefore, there exists a one-to-one correspondence between the orbits of T and those of S. This is why two topologically conjugate systems are considered dynamically equivalent.

Example 1.2.3. Recall the definition of the tent map from Example 1.1.3. We shall now give an example of another system that is topologically conjugate to the tent map. Define the map $F : [0,1] \to [0,1]$ by setting

$$F(x) := \begin{cases} \frac{x}{1-x} & \text{if } x \in [0, \frac{1}{2}] \\ \frac{1-x}{x} & \text{if } x \in [\frac{1}{2}, 1]. \end{cases}$$

The map F is called the *Farey map* and its graph is shown in Figure 1.3.

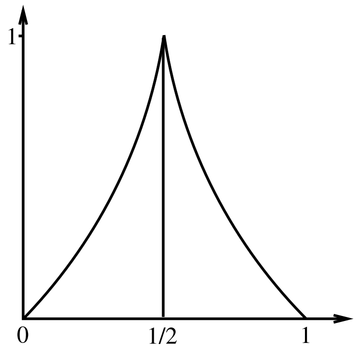

Figure 1.3: The Farey map $F : [0,1] \to [0,1]$.

The Farey map may be familiar to any reader who has studied the continued fraction expansion of real numbers, as it is related to the *Gauss map*, also known as the continued fraction map. Let us just briefly recall that a *continued fraction* is an expression of the form

$$\frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots}}},$$

where $a_i \in \mathbb{N}$ for all $i \in \mathbb{N}$. We write $[a_1, a_2, \ldots]$ for the above expression. It turns out that every continued fraction represents an irrational number in $[0,1]$ and, conversely, every irrational number in $[0,1]$ can be written as a continued fraction. This relation is a bijection.

For $x \in (1/2, 1] \setminus \mathbb{Q}$, the continued fraction representation of x is given by $[1, a_2(x), a_3(x), \ldots]$, where $a_i(x) \in \mathbb{N}$ for all $i \geq 2$. In this case, we deduce that

$$F(x) = \frac{1}{x} - 1 = 1 + [a_2(x), a_3(x), \ldots] - 1 = [a_2(x), a_3(x), \ldots].$$

For $x \in [0, 1/2] \setminus \mathbb{Q}$, the first entry of the continued fraction representation of x is strictly greater than 1 and so it follows that

$$F(x) = \frac{x}{1-x} = \frac{1}{\frac{1}{x} - 1} = [a_1(x) - 1, a_2(x), a_3(x), \ldots].$$

It is known that *Minkowski's question-mark function* is a conjugacy map between the tent map T and the Farey map F. Minkowski's question-mark function is the map $Q : [0,1] \to [0,1]$ defined by

$$Q(x) := -2\sum_{k=1}^{\infty}(-1)^k 2^{-\sum_{i=1}^{k} a_i(x)},$$

whenever $x \in [0,1]\setminus\mathbb{Q}$, where $a_i(x)$ is the ith entry of the continued fraction expansion of x. The map Q is an increasing bijection and is continuous on $[0,1]\setminus\mathbb{Q}$. Recall that a map $f : (Y, d_Y) \to (Z, d_Z)$ between two metric spaces is said to be *Hölder continuous with exponent* α if there exists a constant $C \geq 0$ such that

$$d_Z(f(x), f(y)) \leq C\big(d_Y(x,y)\big)^{\alpha}, \quad \forall x, y \in Y.$$

It was shown by Salem in [62] that the map Q is Hölder continuous with exponent $\log 2/(2\log\gamma)$, where $\gamma := (1+\sqrt{5})/2$ is the golden mean. Furthermore, since $[0,1]\setminus\mathbb{Q}$ is dense in $[0,1]$, the map Q can be uniquely extended to an increasing homeomorphism of $[0,1]$ (this follows from a topological result whose proof is left to Exercise 1.7.5).

Historically, this map was designed by the German mathematician, Hermann Minkowski (1864–1909), to map the rational numbers in $[0,1]$ to the set of dyadic rational numbers $\bigcup_{n=1}^{\infty}\{i/2^n : i = 0, 1, \ldots, 2^n\}$ and the quadratic surds onto the non-dyadic rationals in an order preserving way. The graph of Q is shown in Figure 1.4. For further information on Minkowski's question-mark function, the reader is referred to [48] and [36].

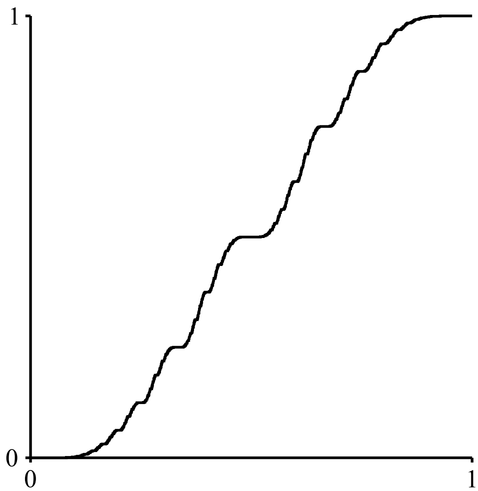

Figure 1.4: Minkowski's question-mark function $Q : [0,1] \to [0,1]$.

Let us now demonstrate that Q really does conjugate the tent and Farey systems. For this, suppose first that $x \in [0,1/2] \setminus \mathbb{Q}$. Then $Q(x) \in [0,1/2]$ and

$$\begin{aligned} T(Q(x)) &= 2\left(-2\sum_{k=1}^{\infty}(-1)^k 2^{-\sum_{i=1}^k a_i(x)}\right) \\ &= -2\left(\sum_{k=1}^{\infty}(-1)^k 2^{-(a_1(x)-1)-\sum_{i=2}^k a_i(x)}\right) \\ &= Q([a_1(x)-1, a_2(x), a_3(x), \ldots]) = Q(F(x)). \end{aligned}$$

Now, suppose that $x \in (1/2,1] \setminus \mathbb{Q}$, that is, $x = [1, a_2(x), a_3(x), \ldots]$. Then $Q(x) \in (1/2,1]$ and

$$\begin{aligned} T(Q(x)) &= 2-2\left(2\cdot 2^{-1} - 2\sum_{k=2}^{\infty}(-1)^k 2^{-1-\sum_{i=2}^k a_i(x)}\right) \\ &= -2\left(\sum_{k=2}^{\infty}(-1)^{k-1} 2^{-\sum_{i=2}^k a_i(x)}\right) \\ &= Q([a_2(x), a_3(x), \ldots]) = Q(F(x)). \end{aligned}$$

Thus, $T(Q(x)) = Q(F(x))$ for all $x \in [0,1] \setminus \mathbb{Q}$. Since this latter set is dense in $[0,1]$, the continuity of T, F, and Q guarantees that $T(Q(x)) = Q(F(x))$ for all $x \in [0,1]$.

Directly from the notion of topological conjugacy, we can derive the following notion of an invariant for a dynamical system.

Definition 1.2.4. A *(topological conjugacy) invariant* is a property of dynamical systems that is preserved under a topological conjugacy map.

Remark 1.2.5.

(a) By definition, topologically conjugate dynamical systems share the same set of topological conjugacy invariants. Thus, if a property is a topological conjugacy invariant and if a given dynamical system has this property while another one does not, then we can immediately deduce that these two dynamical systems are not topologically conjugate.

(b) Among the collection of invariants, there are those which are called *complete invariants*. An invariant is *complete* if two systems that share this invariant are automatically topologically conjugate. Note that this is not true of all invariants, as we will very shortly see. In fact, there are no known complete invariants that exist for arbitrary dynamical systems. Later in this chapter, we shall give examples of topological conjugacy invariants that are complete for a subfamily of dynamical systems.

(c) If $T : X \to X$ is topologically conjugate to $S : Y \to Y$ via a conjugacy map $h : X \to Y$ and if $x \in \mathrm{Per}_n(T)$, then by Remark 1.2.2(b) we deduce that

$$S^n(h(x)) = h(T^n(x)) = h(x),$$

that is, $h(x) \in \mathrm{Per}_n(S)$. Thus, h induces a one-to-one correspondence between periodic points. This correspondence preserves the prime period of a periodic point. Therefore, the number of periodic points of any given period is a topological conjugacy invariant. However, the number of periodic points of any given period is not a complete invariant. Below we give an example of two dynamical systems that have the same number of fixed points despite not being topologically conjugate. Another example will be given in Chapter 3.

(d) The cardinality of X is also an invariant, but again it is not a complete invariant, as we show in the examples below.

Example 1.2.6.

(a) Recall the maps $T_m : \mathbb{S}^1 \to \mathbb{S}^1$ defined in Example 1.1.3. It turns out that the dynamical systems $(\mathbb{S}^1, T_n)$ and $(\mathbb{S}^1, T_m)$ are not topologically conjugate whenever $n \neq m$. Indeed, by Remark 1.2.5(c), we know that if they were topologically conjugate they would have the same number of fixed points. However, T_n has n fixed points, whereas T_m has m.

(b) Let $f : [0,1] \to [0,1]$ be defined by

$$f(x) = \sqrt{x}$$

and let $g : [0,1] \to [0,1]$ be defined by

$$g(x) = 3x(1-x).$$

Then $\mathrm{Fix}(f) = \{0,1\}$ and $\mathrm{Fix}(g) = \{0, 2/3\}$. However, these two systems are not topologically conjugate. This can be seen by supposing $h : [0,1] \to [0,1]$ to be a conjugacy map between f and g. Then we would have to have either $h(0) = 0$ and $h(1) = 2/3$, or $h(0) = 2/3$ and $h(1) = 0$. In either case, it is impossible to construct a homeomorphism of the unit interval into itself satisfying these properties, as such a homeomorphism has to be either strictly increasing or strictly decreasing.

Let us now define the related concept of *structural stability*. Let (X,d) be a compact metric space and let $C(X,X)$ be the space of all continuous maps from X to X. Define the metric d_∞ on $C(X,X)$ by setting

$$d_\infty(T,S) := \sup_{x \in X} d\big(T(x), S(x)\big).$$

The topology on $C(X,X)$ induced by the metric d_∞ is called the *topology of uniform convergence* on X. This terminology is appropriate since $\lim_{n\to\infty} d_\infty(T_n, T) = 0$ if and only if the sequence $(T_n)_{n=1}^\infty$ converges to $T \in C(X,X)$ uniformly. It is not hard to see (and we leave it as an exercise for the reader) that the metric space $(C(X,X), d_\infty)$ is complete and separable.

Let $\mathcal{C}$ be an arbitrary subset of $C(X,X)$. Let τ be a topology on $\mathcal{C}$ which is finer than or coincides with the topology of uniform convergence inherited from $C(X,X)$. We say that an element T of $\mathcal{C}$ is *structurally stable* relative to $\mathcal{C}$ if there exists a neighborhood U of T in the topology τ on $\mathcal{C}$ such that for every $S \in U$ there is a homeomorphism $h = h(S) \in C(X,X)$ for which

$$h \circ T = S \circ h.$$

In other words, T is structurally stable relative to $\mathcal{C}$ if it is topologically conjugate to all systems S in one of its neighborhoods U in the topology τ on $\mathcal{C}$. The system T is *strongly structurally stable* (relative to $\mathcal{C}$) if for every $\varepsilon > 0$ there exists a neighborhood U_ε of T in the topology τ on $\mathcal{C}$ such that for every $S \in U_\varepsilon$ there is a homeomorphism $h \in B_{d_\infty}(\mathrm{Id}_X, \varepsilon)$ for which $T \circ h = h \circ S$. Here, the notation $B_{d_\infty}(g, \varepsilon)$ denotes the ε-ball around the map g:

$$B_{d_\infty}(g,\varepsilon) = \{f \in C(X,X) : d_\infty(f,g) < \varepsilon\}.$$

Later we will provide classes of structurally stable dynamical systems, most notably Shub's expanding endomorphisms (see Chapter 6).

1.3 Factors

A weaker relationship than that of topological conjugacy between two dynamical systems is that of a factor.

Definition 1.3.1. Let $T : X \to X$ and $S : Y \to Y$ be two dynamical systems. If there exists a continuous surjection $h : X \to Y$ such that $h \circ T = S \circ h$, then S is called a *factor* of T. The map h is hereafter called a *factor map*.

In general, the existence of a factor map between two systems is not sufficient to make them topologically conjugate. Nonetheless, if S is a factor of T, then every orbit of T is projected to an orbit of S. As every factor map is by definition surjective, this means that all of the orbits of S have an analogue in T. However, as a factor map needs not be injective, more than one orbit of T may be projected to the same orbit of S. In other words, some orbits of S may have more than one analogue in T. Therefore, the dynamical system T can usually be thought of as more "complicated" than the factor S. In particular, periodic points of period n for T are projected to periodic points for S whose periods are factors of n.

Example 1.3.2. Let $T : X \to X$ be a dynamical system and let $S : Y \to Y$ be given by $Y := \{y\}$ and S equal to the identity map. Then the map $h : X \to Y$ defined by $h(x) := y$ for all $x \in X$ is a factor map. This is, of course, a trivial example. In Chapter 3, we will encounter a class of nontrivial examples.

1.4 Subsystems

Our next aim is to introduce the concept of a subsystem of a dynamical system. In order to do this, we first define the notion of invariance for sets.

Definition 1.4.1. Let $T : X \to X$ be a dynamical system. A subset F of X is said to be
(a) *forward T-invariant* if $T^{-1}(F) \supseteq F$.
(b) *backward T-invariant* if $T^{-1}(F) \subseteq F$.
(c) *completely T-invariant* if $T^{-1}(F) = F$.

If the identity of the map T is clear, then we will sometimes omit it. We also often refer to forward invariant sets simply as "invariant." Note that the condition of being forward invariant is equivalent to $T(F) \subseteq F$.

Remark 1.4.2.
(a) A set is completely invariant if and only if it is both forward and backward invariant.
(b) The closure of an invariant set is invariant.
(c) A set F is invariant if and only if it is equal to the union of the forward orbits of all of its points, that is, $F = \bigcup_{x\in F} \mathcal{O}_+(x)$.
(d) A closed set F is invariant if and only if it is equal to the union of the closure of the forward orbit of all of its points, that is, $F = \bigcup_{x\in F} \overline{\mathcal{O}_+(x)}$. By (c), this means that

$$\overline{\bigcup_{x\in F} \mathcal{O}_+(x)} = F = \bigcup_{x\in F} \overline{\mathcal{O}_+(x)}.$$

We are now in a position to define the concept of subsystem.

Definition 1.4.3. Let $T : X \to X$ be a dynamical system. If $F \subseteq X$ is a closed T-invariant set, then the dynamical system induced by the restriction of T to F, that is, $T|_F : F \to F$ is called a *subsystem* of $T : X \to X$.

Note that as X is a compact metrizable space and, therefore, a compact Hausdorff space, the word "closed" can be replaced by "compact" in the above definition.

Remark 1.4.4. If a dynamical system $S : Y \to Y$ is a factor of a dynamical system $T : X \to X$ via a factor map $h : X \to Y$ and if $Z \subseteq Y$ is forward (resp., backward/completely) S-invariant, then $h^{-1}(Z)$ is forward (resp., backward/completely) T-invariant. Indeed, if $Z \subseteq Y$ is forward S-invariant, that is, $S^{-1}(Z) \supseteq Z$, then

$$T^{-1}(h^{-1}(Z)) = (h \circ T)^{-1}(Z) = (S \circ h)^{-1}(Z) = h^{-1}(S^{-1}(Z)) \supseteq h^{-1}(Z),$$

that is, $h^{-1}(Z)$ is forward T-invariant.

In particular, if $S|_Z$ is a subsystem of $S : Y \to Y$ then $T|_{h^{-1}(Z)}$ is a subsystem of $T : X \to X$. This uses the fact that a compact subset of a Hausdorff space is closed, that the preimage of a closed set under a continuous map is closed, and that a closed subset of a compact space is compact.

If the two systems are topologically conjugate, then any conjugacy map h induces a one-to-one correspondence between the T-invariant sets and the S-invariant sets. In particular, h induces a one-to-one correspondence between the subsystems of T and those of S.

Observe that every orbit is T-invariant, since for every $x \in X$ we have

$$T(\mathcal{O}_+(x)) = \{T(T^n(x)) : n \geq 0\} = \{T^{n+1}(x) : n \geq 0\} \subseteq \mathcal{O}_+(x).$$

By Remark 1.4.2(b), we deduce that the closure of every orbit is T-invariant and, therefore, the restriction of a system to the closure of any of its orbits constitutes a subsystem of that system.

Above and beyond the orbits of a system, the limit points, sometimes called accumulation points, of these orbits are also of interest.

Definition 1.4.5. Let $x \in X$. The set of limit points of the sequence of forward iterates $(T^n(x))_{n=0}^{\infty}$ of x is called the *ω-limit set* of x. It is denoted by $\omega(x)$.

In other words, $y \in \omega(x)$ if and only if there exists a strictly increasing sequence $(n_j)_{j=1}^{\infty}$ of nonnegative integers such that $\lim_{j\to\infty} T^{n_j}(x) = y$.

Remark 1.4.6.
(a) In general, the ω-limit set of a point x is not the set of limit points of the forward orbit $\mathcal{O}_+(x)$ of x. See Exercises 1.7.12, 1.7.13, and 1.7.14.
(b) By the very definition of an ω-limit set, it is easy to see that $\omega(x) \subseteq \overline{\mathcal{O}_+(x)}$. In fact, $\mathcal{O}_+(x) \cup \omega(x) = \overline{\mathcal{O}_+(x)}$. See Exercises 1.7.13 and 1.7.15.

Proposition 1.4.7. *Every ω-limit set is nonempty, closed, and T-invariant. Furthermore, for every $x \in X$ we have that $T(\omega(x)) = \omega(x)$.*

Proof. Let $x \in X$. Since X is compact, the set $\omega(x)$ is nonempty. Moreover, the set $\omega(x)$ is closed as the limit points of any sequence form a closed set (we leave the proof of this fact to Exercise 1.7.16). It only remains to show that $T(\omega(x)) = \omega(x)$. Let $y \in \omega(x)$. Then there exists a strictly increasing sequence $(n_j)_{j=1}^{\infty}$ of nonnegative integers such that $\lim_{j\to\infty} T^{n_j}(x) = y$. The continuity of T then ensures that

$$T(y) = T\Big(\lim_{j\to\infty} T^{n_j}(x)\Big) = \lim_{j\to\infty} T(T^{n_j}(x)) = \lim_{j\to\infty} T^{n_j+1}(x).$$

This shows that $T(y) \in \omega(x)$ and, in turn, proves that $T(\omega(x)) \subseteq \omega(x)$. To establish the reverse inclusion, again fix $y \in \omega(x)$. Then there exists a strictly increasing sequence $(n_j)_{j=1}^{\infty}$ of positive integers such that $\lim_{j\to\infty} T^{n_j}(x) = y$. Consider the sequence $(T^{n_j-1}(x))_{j=1}^{\infty}$. Since X is compact, this sequence admits a convergent subsequence

$(T^{n_{j_k}-1}(x))_{k=1}^{\infty}$, where $(n_{j_k})_{k=1}^{\infty}$ is some subsequence of $(n_j)_{j=1}^{\infty}$. Let $z := \lim_{k\to\infty} T^{n_{j_k}-1}(x)$. Then $z \in \omega(x)$ and

$$T(z) = T\Big(\lim_{k\to\infty} T^{n_{j_k}-1}(x)\Big) = \lim_{k\to\infty} T^{n_{j_k}}(x) = \lim_{j\to\infty} T^{n_j}(x) = y.$$

Consequently, $y \in T(\omega(x))$. This proves that $\omega(x) \subseteq T(\omega(x))$. □

The above proposition shows in particular that the restriction of a dynamical system to any of its ω-limit sets is a subsystem of that system.

If T is a homeomorphism, then we can define the counterpart of an ω-limit set by looking at the backward iterates of a point. For $x \in X$, we define the *α-limit set* of x as the set of accumulation points of the sequence of backward iterates $(T^{-n}(x))_{n=0}^{\infty}$ of x. It is denoted by $\alpha(x)$. In this case, we have that $y \in \alpha(x)$ if and only if there exists a strictly increasing sequence $(n_j)_{j=1}^{\infty}$ of nonnegative integers such that $\lim_{j\to\infty} T^{-n_j}(x) = y$. By the definition of the α-limit set, we have that $\alpha(x)$ is contained in the closure of the backward orbit $\mathcal{O}_-(x)$. The α-limit sets satisfy the same properties under T^{-1} as those of the ω-limit sets under T.

Definition 1.4.8. Let $T : X \to X$ be a topological dynamical system. A point x is said to be *wandering* for T if there exists an open neighborhood U of x such that the preimages of U are mutually disjoint, that is,

$$T^{-m}(U) \cap T^{-n}(U) = \emptyset, \quad \forall m \neq n \geq 0.$$

Accordingly, a point x is called *nonwandering* for T if each of its open neighborhoods U revisits itself under iteration by T, that is, for each neighborhood U of x there is $n \in \mathbb{N}$ such that $T^{-n}(U) \cap U \neq \emptyset$. The *nonwandering set* for T, which consists of all nonwandering points, is denoted by $\Omega(T)$.

Theorem 1.4.9. *The nonwandering set* $\Omega(T)$ *of a system* $T : X \to X$ *enjoys the following properties:*
(a) $\Omega(T)$ *is closed.*
(b) $\emptyset \neq \bigcup_{x\in X} \omega(x) \subseteq \Omega(T)$.
(c) $\mathrm{Per}(T) \subseteq \Omega(T)$.
(d) $\Omega(T)$ *is forward* T*-invariant.*
(e) *If* T *is a homeomorphism, then* $\Omega(T) = \Omega(T^{-1})$ *and is completely* T*-invariant.*

Proof.
(a) The nonwandering set $\Omega(T)$ is closed since its complement, the set of wandering points $X \setminus \Omega(T)$, is open. Indeed, if a point x is wandering, then there exists an open neighborhood U of x such that the preimages of U are mutually disjoint. Therefore, all points of U are wandering as well. So $X \setminus \Omega(T)$ is open.

(b) Let $x \in X$ and $y \in \omega(x)$. Then there is a strictly increasing sequence $(n_k)_{k=1}^{\infty}$ of nonnegative integers such that $\lim_{k\to\infty} T^{n_k}(x) = y$. Thus, given any open neighborhood U of y, there are numbers $n_k < n_l$ such that $T^{n_k}(x) \in U$ and $T^{n_l}(x) \in U$. Then, letting $n = n_l - n_k$ and $z = T^{n_k}(x)$, we have $z \in U$ and $T^n(z) \in U$, that is, $T^{-n}(U) \cap U \neq \emptyset$. As this is true for every open neighborhood U of y, we deduce that $y \in \Omega(T)$. Hence, $\bigcup_{x\in X} \omega(x) \subseteq \Omega(T)$. In particular, $\Omega(T) \neq \emptyset$ since $\omega(x) \neq \emptyset$ for every x.

(c) Since $\omega(x) = \mathcal{O}_+(x) \ni x$ for every periodic point x, all the periodic points of T belong to $\Omega(T)$. More simply, every periodic point is nonwandering as it eventually returns to itself under iteration.

(d) Let $x \in \Omega(T)$ and U an open neighborhood of $T(x)$. Then $T^{-1}(U)$ is an open neighborhood of x. As $x \in \Omega(T)$, there exists $n \in \mathbb{N}$ such that $T^{-n}(T^{-1}(U)) \cap T^{-1}(U) \neq \emptyset$. That is, $T^{-1}(T^{-n}(U) \cap U) \neq \emptyset$, which implies that $T^{-n}(U) \cap U \neq \emptyset$. Since this is true for every open neighborhood U of $T(x)$, we conclude that $T(x) \in \Omega(T)$, and hence $T(\Omega(T)) \subseteq \Omega(T)$.

(e) Suppose T is a homeomorphism. It is easy to show that $\Omega(T^{-1}) = \Omega(T)$. By (d), we then have $T(\Omega(T)) \subseteq \Omega(T)$ and $T^{-1}(\Omega(T)) \subseteq \Omega(T)$. This implies $T^{-1}(\Omega(T)) = \Omega(T)$.

Parts (a) and (d) tell us that the restriction of a system to its nonwandering set forms a subsystem of that system. Part (b) reveals that this subsystem comprises all ω-limit subsystems. □

Finally, we introduce the notion of invariance for a function.

Definition 1.4.10. A continuous function $g : X \to \mathbb{R}$ is said to be T-*invariant* if $g \circ T = g$.

Remark 1.4.11. A function g is T-invariant if and only if $g \circ T^n = g$ for every $n \in \mathbb{N}$. In other words, g is T-invariant if and only if g is constant along each orbit of T, and thus if and only if g is constant on the closure of each orbit of the system T.

1.5 Mixing and irreducibility

In this section, we investigate various forms of topological mixing and irreducibility that can be observed in some dynamical systems. For a dynamical system, topological mixing can intuitively be conceived as witnessing some parts of the underlying space becoming mixed under iteration with other parts of the space. Irreducibility of a dynamical system means that the system does not admit any “nontrivial” subsystem, which takes a different meaning depending on the stronger or weaker form of irreducibility. In any case, the absence of nontrivial subsystems forces irreducible systems to exhibit some form of mixing.

1.5.1 Minimality

We will now define one way in which a dynamical system $T : X \to X$ can be said to be irreducible. As mentioned above, by irreducibility we mean that T admits no nontrivial subsystem, for some sense of nontriviality. One natural form of irreducibility would be that the only subsystems of T are the empty system and T itself. Another way of saying this is that the only closed T-invariant subsets of X are the empty set and the whole of X. The concept we need for this is minimality.

Definition 1.5.1. Let $T : X \to X$ be a dynamical system. A set $F \subseteq X$ is said to be a *minimal set* for T if the following three conditions are satisfied:
(a) The set F is nonempty and closed.
(b) The set F is T-invariant.
(c) If $G \subseteq F$ is nonempty, closed and T-invariant, then $G = F$.

A minimal set F induces the *minimal subsystem* $T|_F : F \to F$. We now address the question of the existence of minimal sets.

Theorem 1.5.2. *Every dynamical system admits a minimal set (that induces a minimal subsystem).*

Proof. Let $\mathcal{F}$ be the family of all nonempty, closed, T-invariant subsets of X. This family is nonempty, as it at least contains X. It is also partially ordered under the relation of backward set inclusion $\supseteq$. We shall use the Kuratowski–Zorn lemma (often referred to simply as Zorn's lemma) to establish the existence of a minimal set. Accordingly, let $\{F_\lambda\}_{\lambda\in\Lambda}$ be a chain in $\mathcal{F}$, that is, a totally ordered subset of $\mathcal{F}$, and let $F = \bigcap_{\lambda\in\Lambda} F_\lambda$. Then F is nonempty and closed (cf. Exercise 1.7.17). Moreover,

$$T(F) = T\Big(\bigcap_{\lambda\in\Lambda} F_\lambda\Big) \subseteq \bigcap_{\lambda\in\Lambda} T(F_\lambda) \subseteq \bigcap_{\lambda\in\Lambda} F_\lambda = F.$$

Thus, F is T-invariant and constitutes the maximal element of the chain $\{F_\lambda\}_{\lambda\in\Lambda}$. Hence, every chain in $\mathcal{F}$ has a maximal element, and by Zorn's lemma, we infer that the family $\mathcal{F}$ has a maximal element under the relation of backward set inclusion $\supseteq$, that is, the family $\mathcal{F}$ has a minimal element under the relation of set inclusion $\subseteq$. This element is a minimal set for T. □

We can now define the concept of minimality for dynamical systems.

Definition 1.5.3. A dynamical system $T : X \to X$ is said to be *minimal* if X is a minimal set for T (and thus is the only minimal set for T).

Minimality is a strong form of irreducibility since minimal systems admit no nonempty subsystems other than themselves.

Let us now give a characterization of minimal sets. In particular, the next result shows that the strong form of irreducibility that we call minimality is also a strong form of mixing, since minimal systems are characterized by having only dense orbits.

Theorem 1.5.4. *Let F be a nonempty closed T-invariant subset of X. Then the following three statements are equivalent:*

(a) *F is minimal.*
(b) *$\omega(x) = F$ for every $x \in F$.*
(c) *$\overline{\mathcal{O}_+(x)} = F$ for every $x \in F$.*

Proof. We shall prove this theorem by establishing the sequence of implications (a)⇒(b)⇒(c)⇒(a).

To begin, suppose that F is a minimal set for T, and let $x \in F$. Then, since F is T-invariant and closed, we obtain that $\omega(x) \subseteq \overline{\mathcal{O}_+(x)} \subseteq F$. Moreover, in light of Proposition 1.4.7, $\omega(x)$ is nonempty, closed and T-invariant. So, by the definition of a minimal set, we must have that $\omega(x) = F$. This proves that (a) implies (b).

Toward the proof of the second implication, recall that as F is T-invariant and closed, we have that $\omega(x) \subseteq \overline{\mathcal{O}_+(x)} \subseteq F$. Thus, if $\omega(x) = F$ then $\overline{\mathcal{O}_+(x)} = F$. This proves that (b) implies (c).

Finally, assume that $\overline{\mathcal{O}_+(x)} = F$ for every $x \in F$. Let $E \subseteq F$ be a nonempty closed T-invariant set. It suffices to show that $E = F$. To that end, let $x \in E$. As E is T-invariant and closed, we have $\overline{\mathcal{O}_+(x)} \subseteq E$. Moreover, $x \in E \subseteq F$ implies that $\overline{\mathcal{O}_+(x)} = F$. Therefore, $F = \overline{\mathcal{O}_+(x)} \subseteq E \subseteq F$, and hence $E = F$. Thus, F is minimal. This proves the remaining implication, namely, that (c) implies (a). □

Remark 1.5.5.

(a) Theorem 1.5.4(c) characterizes a minimal set F by the requirement that the orbit of each point of F stays in F and is dense in F. Another way to think of this is that irreducibility in the sense that a system admits no nonempty subsystem other than itself, is equivalent to mixing in the sense that every orbit is dense. In particular, a consequence of this property is that a minimal system must be surjective. It also implies that an infinite minimal system does not admit any periodic point, as $\overline{\mathcal{O}_+(x)} = \mathcal{O}_+(x)$ for any periodic point x.
(b) Minimality is a topological conjugacy invariant. However, it is not a complete invariant (see Exercise 1.7.19).

1.5.2 Transitivity and topological mixing

In this section, we introduce a weaker form of mixing called transitivity. We have shown in the previous section that minimal systems have only dense orbits; transitive systems are only required to exhibit one dense orbit. Nonetheless, as we shall soon see, the existence of one dense orbit forces the existence of a dense G_δ-set of points with dense orbits.

Definition 1.5.6. Let $T : X \to X$ be a dynamical system.
(a) A point $x \in X$ is said to be *transitive* for T if $\omega(x) = X$.
(b) The system T is called *transitive* if it admits at least one transitive point.
(c) A point $x \in X$ is said to be *weakly transitive* for T if $\overline{\mathcal{O}_+(x)} = X$.
(d) The system T is said to be *weakly transitive* if it admits at least one weakly transitive point.

Remark 1.5.7.
(a) In light of Theorem 1.5.4(b), every minimal system is transitive. There are, of course, transitive systems which are not minimal. For instance, we shall see in Chapter 3 that full shifts are transitive, but not minimal since they admit periodic points.
(b) Transitivity is a topological conjugacy invariant. However, it is not a complete invariant. Indeed, as minimality is not a complete invariant and transitivity is weaker than minimality, transitivity cannot be a complete invariant.
(c) A transitive system is surjective since, given any transitive point x, we have that $T(X) = T(\omega(x)) = \omega(x) = X$.
(d) As $\overline{\mathcal{O}_+(x)} = \mathcal{O}_+(x) \cup \omega(x)$, every transitive system is weakly transitive. Note that there are weakly transitive systems which are not transitive, as Example 1.5.8 below demonstrates.
(e) If $T : X \to X$ is weakly transitive, then every continuous T-invariant function is constant. To see this, let $g : X \to \mathbb{R}$ be a T-invariant continuous function. Also let $x \in X$ be such that the forward orbit of x is dense in X. Then by Remark 1.4.11, we have that $g|_{\mathcal{O}_+(x)} = g(x)$. This means that the continuous function g is constant on a dense set of points, so it must be constant everywhere.

Example 1.5.8. Let $X = \{0\} \cup \{1/n : n \in \mathbb{N}\} \subseteq \mathbb{R}$, and let $T : X \to X$ be defined by $T(0) = 0$ and $T(1/n) = 1/(n+1)$. Then T is continuous. Moreover, as it is not surjective (its range does not include 1), T cannot be transitive. Alternatively, we might argue that $\omega(x) = \{0\}$ for every $x \in X$. However, observe that $\mathcal{O}_+(1) = \{1, 1/2, 1/3, \ldots\} = \{1/n : n \in \mathbb{N}\}$. So, $\overline{\mathcal{O}_+(1)} = X$ and, therefore, T is weakly transitive.

In fact, it turns out that surjectivity is the only difference between weakly transitive and transitive systems, as we now show.

Theorem 1.5.9. *A dynamical system is transitive if and only if it is weakly transitive and surjective.*

Proof. We have already observed in Remark 1.5.7 that transitive systems are weakly transitive and surjective. Suppose now that a system $T : X \to X$ is weakly transitive and surjective. Let $x \in X$ be such that $\overline{\mathcal{O}_+(x)} = X$. Since $X = \overline{\mathcal{O}_+(x)} = \mathcal{O}_+(x) \cup \omega(x)$, we deduce that $X \setminus \mathcal{O}_+(x) \subseteq \omega(x)$.

On one hand, if $T^{-1}(x) \cap \mathcal{O}_+(x) = \emptyset$ then by the surjectivity of T we have that $\emptyset \neq T^{-1}(x) \subseteq X \setminus \mathcal{O}_+(x) \subseteq \omega(x)$. As $\omega(x)$ is T-invariant, we obtain that $\{x\} = T(T^{-1}(x)) \subseteq T(\omega(x)) = \omega(x)$. Using the T-invariance of $\omega(x)$ once again, we deduce that $\mathcal{O}_+(x) \subseteq \omega(x)$. Therefore $X = \overline{\mathcal{O}_+(x)} \subseteq \omega(x) \subseteq X$, that is, $\omega(x) = X$.

On the other hand, if $T^{-1}(x) \cap \mathcal{O}_+(x) \neq \emptyset$, then x is a periodic point, and hence $\omega(x) = \mathcal{O}_+(x)$. It follows that $X = \overline{\mathcal{O}_+(x)} = \mathcal{O}_+(x) \cup \omega(x) = \mathcal{O}_+(x) = \omega(x)$. (That is, every point in X is a periodic point in the orbit of x and X is finite.)

In either case, we have demonstrated that x is a transitive point. □

Let us now introduce the concept of topological mixing. We shall very shortly see the connection between this notion and transitivity.

Definition 1.5.10. A dynamical system $T : X \to X$ is said to be *topologically mixing* if any of the following equivalent statements hold:

(a) For all nonempty open subsets U and V of X, there exists $n \in \mathbb{N}$ such that $T^n(U) \cap V \neq \emptyset$.
(b) For all nonempty open subsets U and V of X, there exists $n \in \mathbb{N}$ such that $U \cap T^{-n}(V) \neq \emptyset$.
(c) For all nonempty open subsets U and V of X and for all $N \in \mathbb{N}$, there exists $n \geq N$ such that $T^n(U) \cap V \neq \emptyset$.
(d) For all nonempty open subsets U and V of X and for all $N \in \mathbb{N}$, there exists $n \geq N$ such that $U \cap T^{-n}(V) \neq \emptyset$.
(e) For all nonempty open subsets U and V of X, there exist infinitely many $n \in \mathbb{N}$ such that $T^n(U) \cap V \neq \emptyset$.
(f) For all nonempty open subsets U and V of X, there exist infinitely many $n \in \mathbb{N}$ such that $U \cap T^{-n}(V) \neq \emptyset$.
(g) $\overline{\bigcup_{n \in \mathbb{N}} T^n(U)} = X$ for every nonempty open subset U of X.
(h) $\overline{\bigcup_{n \in \mathbb{N}} T^{-n}(U)} = X$ for every nonempty open subset U of X.

We leave it to the reader to provide a proof that (a) and (b) are equivalent, (c) and (d) are equivalent and (e) and (f) are equivalent. It is also straightforward to show the chain of equivalences (b)⇔(d) ⇔(f) by using the fact that $W := T^{-(N-1)}(V)$ is an open set. Statement (g) is just a rewriting of (a) while (h) is a reformulation of (b).

Furthermore, note that each of these statements implies that T is surjective. Indeed, since $T(X) \subseteq X$, we have by induction that $T^{n+1}(X) \subseteq T^n(X)$ for all $n \in \mathbb{N}$. So the sequence of compact sets $(T^n(X))_{n=0}^{\infty}$ is descending. By (g), $T(X) = \overline{T(X)} = \overline{\bigcup_{n \in \mathbb{N}} T^n(X)} = X$.

The next theorem gives the promised connection between transitivity and topological mixing, in addition to another characterization of transitive systems in terms of nowhere-dense sets. For more information on nowhere-dense sets, the reader is referred to [77] and [54]. The most relevant fact for us is that a closed set is nowhere-dense if and only if it has empty interior.

Theorem 1.5.11. *If $T : X \to X$ is a surjective dynamical system, then the following statements are equivalent:*
(a) *T is transitive.*
(b) *Whenever F is a closed T-invariant subset of X, either $F = X$ or F is nowhere-dense. In other words, T admits no subsystem with nonempty interior other than itself.*
(c) *T is topologically mixing.*
(d) *$\{x \in X : \overline{\mathcal{O}_+(x)} = X\}$ is a dense G_δ-subset of X.*
(e) *$\{x \in X : \overline{\mathcal{O}_+(x)} = X\}$ contains a dense G_δ-subset of X.*

Proof. We shall prove this theorem by establishing the implications (a)⇒(b)⇒(c)⇒(d)⇒(e)⇒(a).

To prove the first implication, let x be a transitive point and F a nonempty closed T-invariant set. Since x is transitive, we have that $\omega(x) = X$. Suppose that F has nonempty interior. Then there is a nonempty open set $U \subseteq F$. Therefore, there exists $p \geq 0$ with $T^p(x) \in U \subseteq F$. As F is T-invariant, all higher iterates of x lie in F as well. Since F is closed, this implies that $\omega(x) \subseteq F$. Therefore $X = \omega(x) \subseteq F \subseteq X$, that is, $F = X$. Thus, either F has empty interior or $F = X$. That is, either F is nowhere-dense or $F = X$. This proves that (a) implies (b).

For the second implication, suppose that (b) holds and that U and V are nonempty open subsets of X. By the surjectivity of T, the union $\bigcup_{n=1}^{\infty} T^{-n}(V)$ is a nonempty open subset of X. Therefore, the closed set $F := X \setminus \bigcup_{n=1}^{\infty} T^{-n}(V) \neq X$ satisfies

$$\begin{aligned} T^{-1}(F) &= X \setminus T^{-1}\left(\bigcup_{n=1}^{\infty} T^{-n}(V)\right) \\ &= X \setminus \bigcup_{n=2}^{\infty} T^{-n}(V) \supseteq X \setminus \bigcup_{n=1}^{\infty} T^{-n}(V) = F. \end{aligned}$$

This means that $T(F) \subseteq F$, that is, the set F is T-invariant. Thus, either $F = X$ or F is nowhere-dense. As $F \neq X$, the set F is nowhere-dense and its complement $\bigcup_{n=1}^{\infty} T^{-n}(V)$ is dense in X. Consequently, there exists some $n \in \mathbb{N}$ such that $U \cap T^{-n}(V) \neq \emptyset$. So T is topologically mixing. This establishes that (b) implies (c).

Now, suppose that T is topologically mixing and let $\{U_n : n \in \mathbb{N}\}$ be a countable base for the topology of X. Such a base exists since X is a compact metrizable space. Then

$$\{x \in X : \overline{\mathcal{O}_+(x)} = X\} = \bigcap_{n=1}^{\infty} \bigcup_{m=0}^{\infty} T^{-m}(U_n).$$

Since T is topologically mixing, for each $n \in \mathbb{N}$ the open set $\bigcup_{m=0}^{\infty} T^{-m}(U_n)$ intersects every nonempty open set, that is, this set is dense in X. By the Baire category theorem, it follows that $\bigcap_{n=1}^{\infty} \bigcup_{m=0}^{\infty} T^{-m}(U_n)$ is a dense G_δ-set. This proves that (c) implies (d).

The implication (d)⇒(e) is obvious.

Finally, suppose that $\{x \in X : \overline{\mathcal{O}_+(x)} = X\}$ contains a dense G_δ-set. This implies immediately that T is weakly transitive. As T is surjective, we deduce from Theorem 1.5.9 that T is in fact transitive. This demonstrates that (e) implies (a). □

In particular, Theorem 1.5.11 shows that transitivity corresponds to a weaker form of irreducibility than minimality. Indeed, as opposed to minimal systems which admit only the empty set and the whole of X as subsystems, nonminimal transitive systems admit nontrivial subsystems. Each of these subsystems is nevertheless nowhere-dense.

Rotations of the unit circle

We shall now discuss rotations of the unit circle $\mathbb{S}^1$. The unit circle may be defined in many different homeomorphic ways. It may be embedded in the complex plane by defining $\mathbb{S}^1 := \{z \in \mathbb{C} : |z| = 1\}$. It may also be defined to be the set of all angles $\theta \in [0, 2\pi]$ (mod 2π). Alternatively, as we have already seen in Example 1.1.3(b), it may be defined to be the quotient space $\mathbb{S}^1 := \mathbb{R}/\mathbb{Z}$ or as $\mathbb{S}^1 := [0,1]$ (mod 1). We will use the form most appropriate to each specific situation.

Let $\mathbb{S}^1 = [0, 2\pi]$ (mod 2π). Let $\alpha \in \mathbb{R}$ and define the map $T_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ by

$$T_\alpha(x) = x + 2\pi\alpha \pmod{2\pi}.$$

Thus, T_α is the rotation of the unit circle by the angle $2\pi\alpha$. The dynamics of T_α are radically different depending on whether the number α is rational or irrational. We prove the following classical result.

Theorem 1.5.12. *Let $T_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ be defined as above. The following are equivalent:*
(a) $\alpha \in \mathbb{R}\backslash\mathbb{Q}$.
(b) T_α *is minimal.*
(c) T_α *is transitive.*

Moreover, when $\alpha \in \mathbb{Q}$ every point in $\mathbb{S}^1$ is a periodic point with the same prime period.

Proof. We shall prove that (b)⇒(c)⇒(a)⇒(b). Remark 1.5.7 already pointed out that (b)⇒(c).

[(c)⇒(a)] If $\alpha \in \mathbb{Q}$, say $\alpha = p/q$ for some $p, q \in \mathbb{Z}$ with p and q relatively prime and $q > 0$, then $T_\alpha^n(x) = x + 2\pi pn/q$ (mod 2π) for all $x \in \mathbb{S}^1$. In particular, $T_\alpha^q(x) = x + 2\pi p$ (mod 2π) $= x$ (mod 2π) for all $x \in \mathbb{S}^1$. Hence, T_α^q is the identity map, that is, every point in $\mathbb{S}^1$ is a periodic point of (prime) period q. In particular, T_α is not transitive. Therefore, if T_α is transitive then $\alpha \notin \mathbb{Q}$.

[(a)⇒(b)] Suppose now that $\alpha \notin \mathbb{Q}$ and that T_α is not minimal. Let F be a minimal set for T_α. Such a set exists by Theorem 1.5.2 and $F = \omega(x)$ for each $x \in F$ by Theorem 1.5.4. From this and Proposition 1.4.7, we deduce that $T_\alpha(F) = F$. Since T_α

is a bijection (in fact, a homeomorphism), we obtain that $T_\alpha^{-1}(F) = F = T_\alpha(F)$. Consequently, $T_\alpha^{-1}(\mathbb{S}^1 \setminus F) = \mathbb{S}^1 \setminus F = T_\alpha(\mathbb{S}^1 \setminus F)$. As by definition $F \neq \mathbb{S}^1$ and F is closed, the set $\mathbb{S}^1 \setminus F$ is nonempty and open. So it can be written as a countable union of open intervals

$$\mathbb{S}^1 \setminus F = \bigcup_{k=1}^{\infty}(a_k, b_k),$$

where the (a_k, b_k)'s are the connected components of $\mathbb{S}^1 \setminus F$. For each k, we have that $T_\alpha((a_k, b_k)) \subseteq \mathbb{S}^1 \setminus F$. Since (a_k, b_k) is connected and T_α is continuous, the image $T_\alpha((a_k, b_k))$ is also connected. This implies that none of the endpoints a_j, b_j lies in $T_\alpha((a_k, b_k))$. Therefore, there exists a unique $\ell = \ell(k) \in \mathbb{N}$ such that $T_\alpha((a_k, b_k)) \subseteq (a_\ell, b_\ell)$. Since $T_\alpha(a_k)$ and $T_\alpha(b_k)$ are in F, the continuity of T_α implies that $T_\alpha((a_k, b_k)) = (a_\ell, b_\ell)$. By induction on n, there exist unique a_{k_n} and b_{k_n} such that $T_\alpha^n((a_k, b_k)) = (a_{k_n}, b_{k_n})$. We claim that the family

$$\{T_\alpha^n((a_k, b_k)) : n \geq 0\} = \{(a_{k_n}, b_{k_n}) : n \geq 0\}$$

consists of mutually disjoint open arcs. If not, there would exist some $0 \leq p < q$ such that

$$T_\alpha^p((a_k, b_k)) \cap T_\alpha^q((a_k, b_k)) = (a_{k_p}, b_{k_p}) \cap (a_{k_q}, b_{k_q}) \neq \emptyset.$$

As $\{(a_j, b_j) : j \in \mathbb{N}\}$ is a pairwise disjoint family of open arcs, we deduce that

$$T_\alpha^p((a_k, b_k)) = T_\alpha^q((a_k, b_k)).$$

Consequently,

$$(a_k, b_k) = T_\alpha^{-p} \circ T_\alpha^q((a_k, b_k)) = T_\alpha^{q-p}((a_k, b_k)).$$

Writing $r := q - p \in \mathbb{N}$, this means that

$$T_\alpha^r((a_k, b_k)) = (a_k, b_k).$$

Thus, either $T_\alpha^r(a_k) = a_k$ or $T_\alpha^r(a_k) = b_k$. In either case, we would have that $T_\alpha^{2r}(a_k) = a_k$ and so a_k would be a periodic point of period $2r$ for T_α. That is, $T_\alpha^{2r}(a_k) = a_k + 4\pi r\alpha \pmod{2\pi} = a_k \pmod{2\pi}$. This means that $4\pi r\alpha = 0 \pmod{2\pi}$ or, in other words, α would be a rational number, which would contradict our original assumption.

Hence, we have shown that $\{T_\alpha^n((a_k, b_k))\}_{n=0}^{\infty}$ forms a family of mutually disjoint arcs. Moreover, $\mathrm{Leb}(T_\alpha^n((a_k, b_k))) = \mathrm{Leb}((a_k, b_k)) > 0$ for all $n \geq 0$, where Leb denotes the Lebesgue measure on $\mathbb{S}^1$. So the circle, which has finite Lebesgue measure, apparently contains an infinite family of disjoint arcs of equal positive Lebesgue measure. This is obviously impossible. This contradiction leads us to conclude that T_α is minimal whenever $\alpha \notin \mathbb{Q}$. □

Remark 1.5.13. An immediate consequence of the above theorem is that the orbit $\mathcal{O}_+(x)$ of every $x \in \mathbb{S}^1$ is dense in $\mathbb{S}^1$ when α is irrational. This is sometimes called Jacobi's theorem.

Finally, among the many other forms of transitivity, let us mention two. First, strongly transitive systems are those for which all points have a dense backward orbit.

Definition 1.5.14. A dynamical system $T : X \to X$ is *strongly transitive* if any of the following equivalent statements holds:
(a) $\overline{\bigcup_{n\in\mathbb{N}} T^{-n}(x)} = X$ for every $x \in X$.
(b) $\bigcup_{n\in\mathbb{N}} T^n(U) = X$ for every nonempty open subset U of X.

The proof of the equivalency is left to the reader. It is obvious that a strongly transitive system is topologically mixing, and thus transitive when the underlying space X is metrizable. But the converse does not hold in general. In Lemma 4.2.10, we will find conditions under which a transitive system is strongly transitive.

Definition 1.5.15. A dynamical system $T : X \to X$ is *very strongly transitive* if for every nonempty open subset U of X, there is $N = N(U) \in \mathbb{N}$ such that $\bigcup_{n=1}^{N} T^n(U) = X$.

It is clear that a very strongly transitive system is strongly transitive, as the terminology suggests. The converse is not true in general but, given the compactness of X, every open strongly transitive system is very strongly transitive.

1.5.3 Topological exactness

Definition 1.5.16. A dynamical system $T : X \to X$ is called *topologically exact* if for each nonempty open set $U \subseteq X$, there exists some $n \in \mathbb{N}$ such that $T^n(U) = X$.

Note that topologically exact systems are very strongly transitive. However, they may not be minimal. Full shifts, which we shall study in Chapter 3, are topologically exact systems, which are not minimal. On the other hand, there are very strongly transitive systems, which are not topologically exact (see Exercise 1.7.25).

Remark 1.5.17. Topological exactness is a topological conjugacy invariant. However, it is not a complete invariant. We shall see in Chapter 3 examples of topologically exact systems that are not topologically conjugate. Among others, two full shifts are topologically conjugate if and only if they are built upon alphabets with the same number of letters (see Theorem 3.1.14).

1.6 Examples

1.6.1 Rotations of compact topological groups

For our first example, we consider *topological groups*. For a detailed introduction to these objects, the reader is referred to [69] and, for a dynamical viewpoint, to [21].

A topological group is simply a group G together with a topology on G that satisfies the following two properties:
(a) The product map $\pi : G \times G \to G$ defined by setting

$$\pi(g,h) := gh$$

is continuous when $G \times G$ is endowed with the product topology.
(b) The inverse map $i : G \to G$ defined by setting

$$i(g) := g^{-1}$$

is continuous.

Given $a \in G$, we define the map $L_a : G \to G$ by

$$L_a(g) := ag.$$

So L_a acts on the group G by left multiplication by a. The map L_a is often referred to as the left rotation of G by a. The map L_a is continuous since $L_a(g) = \pi(a,g)$. Moreover, observe that $L_a^n = L_{a^n}$ for every $n \in \mathbb{Z}$. In particular, $L_a^{-1} = L_{a^{-1}}$. The rotation L_a is thus a homeomorphism of G. In a similar way, we define the right rotation of G by a to be the continuous map $R_a : G \to G$, where

$$R_a(g) := ga.$$

For rotations of topological groups, transitivity and minimality are one and the same property, as the following theorem shows.

Theorem 1.6.1. *Let $L_a : G \to G$ be the left rotation of a topological group G by $a \in G$. Then L_a is minimal if and only if L_a is transitive. Similarly, the right rotation R_a is minimal if and only if it is transitive.*

Proof. If L_a is minimal, then L_a is transitive by Remark 1.5.7(a). For the converse, let x be a transitive point and let $y \in G$ be arbitrary. According to Theorem 1.5.4, it suffices to show that $\omega(y) = G$. Let $z \in \omega(x)$. Then there exists a strictly increasing sequence $(n_k)_{k=1}^{\infty}$ of nonnegative integers such that $\lim_{k\to\infty} L_a^{n_k}(x) = z$. Observe that

$$\begin{aligned}\lim_{k\to\infty} L_a^{n_k}(y) &= \lim_{k\to\infty} L_a^{n_k}(xx^{-1}y) = \lim_{k\to\infty} a^{n_k}xx^{-1}y = \lim_{k\to\infty} (a^{n_k}x)(x^{-1}y)\\ &= \lim_{k\to\infty} (L_a^{n_k}(x))(x^{-1}y) = \lim_{k\to\infty} R_{x^{-1}y}(L_a^{n_k}(x))\\ &= R_{x^{-1}y}\Big(\lim_{k\to\infty} L_a^{n_k}(x)\Big) = R_{x^{-1}y}(z).\end{aligned}$$

So $R_{x^{-1}y}(z) \in \omega(y)$. Since $z \in \omega(x)$ was chosen arbitrarily, we deduce that $R_{x^{-1}y}(\omega(x)) \subseteq \omega(y)$. As $\omega(x) = G$, we conclude that

$$G = R_{x^{-1}y}(G) = R_{x^{-1}y}\big(\omega(x)\big) \subseteq \omega(y) \subseteq G.$$

Hence, $\omega(y) = G$ for any arbitrary $y \in G$ and so G, and thus L_a, is minimal. The proof that R_α is minimal proceeds analogously and is thus left to the reader. $\square$

We now give a characterization of all minimal rotations of a topological group.

Proposition 1.6.2. *The rotation $L_a : G \to G$ is minimal if and only if*

$$G = \overline{\mathcal{O}_+(e)} = \overline{\{a^n : n \geq 0\}},$$

where e denotes the identity element of G.

Proof. First, observe that $\overline{\mathcal{O}_+(e)} = \overline{\{a^n : n \geq 0\}}$ since $L_a^n(e) = a^n e = a^n$ for each $n \geq 0$. Now suppose that L_a is minimal. By Theorem 1.5.4, we have that $\overline{\mathcal{O}_+(g)} = G$ for every $g \in G$. In particular, $\overline{\mathcal{O}_+(e)} = G$.

For the converse, suppose that $G = \overline{\mathcal{O}_+(e)}$. According to Theorem 1.6.1, it is sufficient to prove that L_a is transitive. By Theorem 1.5.9, since L_a is surjective it is sufficient to show that L_a is weakly transitive. This is certainly the case since $\overline{\mathcal{O}_+(e)} = G$. $\square$

This characterization shows that minimal rotations can only occur in abelian groups. Indeed, if L_a is minimal and $x, y \in G$, then there exist strictly increasing sequences $(n_j)_{j=1}^{\infty}$ and $(m_k)_{k=1}^{\infty}$ of nonnegative integers such that $x = \lim_{j\to\infty} a^{n_j}$ and $y = \lim_{k\to\infty} a^{m_k}$. Using the left and right continuity of the product map, we obtain

$$\begin{aligned}
xy &= \Big(\lim_{j\to\infty} a^{n_j}\Big)y = \lim_{j\to\infty}(a^{n_j}y) = \lim_{j\to\infty}\Big(a^{n_j}\lim_{k\to\infty} a^{m_k}\Big) \\
&= \lim_{j\to\infty}\Big(\lim_{k\to\infty}(a^{n_j}a^{m_k})\Big) = \lim_{j\to\infty}\Big(\lim_{k\to\infty}(a^{m_k}a^{n_j})\Big) \\
&= \lim_{j\to\infty}\Big(\Big(\lim_{k\to\infty} a^{m_k}\Big)a^{n_j}\Big) = \lim_{j\to\infty}(ya^{n_j}) \\
&= y\lim_{j\to\infty} a^{n_j} = yx.
\end{aligned}$$

Rotations of the n-dimensional torus $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$

We shall now study rotations (also sometimes called translations) of the n-dimensional torus

$$\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1,$$

that is, the n-times direct product of $\mathbb{S}^1 := [0,1] \pmod 1$. Let $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n) \in \mathbb{T}^n$ and let $L_\gamma : \mathbb{T}^n \to \mathbb{T}^n$ be the rotation of $\mathbb{T}^n$ by γ, which is defined to be

$$L_\gamma(x_1, x_2, \ldots, x_n) := (x_1 + \gamma_1, x_2 + \gamma_2, \ldots, x_n + \gamma_n) \pmod 1.$$

The proof of the following theorem uses Fourier coefficients and the Hilbert space $L^2(\lambda_n)$ of complex-valued functions whose squared modulus is integrable with respect to the Lebesgue measure λ_n on the n-dimensional torus. A good reference for those unfamiliar with Fourier coefficients or the Hilbert space $L^2(\lambda_n)$ is Rudin [58]. Those unfamiliar with measure theory may wish to consult Appendix A first.

Recall that the numbers $1, \gamma_1, \gamma_2, \dots, \gamma_n$ are said to be *linearly independent over the field of rational numbers* $\mathbb{Q}$ if the equation

$$\alpha_0 + \alpha_1\gamma_1 + \alpha_2\gamma_2 + \cdots + \alpha_n\gamma_n = 0,$$

where the α_j are rational numbers, has for unique solution $\alpha_0 = \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$. Equivalently, the numbers $1, \gamma_1, \gamma_2, \dots, \gamma_n$ are linearly independent over $\mathbb{Q}$ if

$$k_1\gamma_1 + k_2\gamma_2 + \cdots + k_n\gamma_n \in \mathbb{Z},$$

where each $k_j \in \mathbb{Z}$, only when $k_1 = k_2 = \cdots = k_n = 0$.

We now prove the following classical result, which is a significant generalization of Theorem 1.5.12 with a consequently more intricate proof.

Theorem 1.6.3. *Let $L_\gamma : \mathbb{T}^n \to \mathbb{T}^n$ be a translation of the torus. The following statements are equivalent:*
(a) *The numbers $1, \gamma_1, \gamma_2, \dots, \gamma_n$ are linearly independent over $\mathbb{Q}$.*
(b) *L_γ is minimal.*
(c) *L_γ is transitive.*

Proof. According to Theorem 1.6.1, a rotation of a topological group is minimal if and only if it is transitive. Therefore, (b)⇔(c). We shall now prove that (a)⇔(c).

Suppose first that L_γ is transitive. Assume by way of contradiction that $\sum_{j=1}^n k_j\gamma_j \in \mathbb{Z}$, where each $k_j \in \mathbb{Z}$ and at least one of these numbers, say k_i, differs from zero. Let $\varphi : \mathbb{T}^n \to \mathbb{R}$ be the function defined by

$$\varphi(x) = \varphi(x_1, x_2, \dots, x_n) = \sin\left(2\pi \sum_{j=1}^n k_j x_j\right).$$

Since x is in reality the equivalence class $[x] = \{x + \ell : \ell \in \mathbb{Z}^n\}$, we must check that φ is well-defined, that is, is constant on the entire equivalence class. This straightforward calculation is left to the reader as an exercise (see Exercise 1.7.30). As φ is a composition of continuous maps, it is continuous. Moreover, since $\varphi(0, 0, \dots, 0) = 0$ but $\varphi(0, \dots, 0, 1/(4k_i), 0, \dots, 0) = \sin(2\pi k_i \frac{1}{4k_i}) = \sin(\frac{\pi}{2}) = 1$, the function φ is not constant. But given that

$$\varphi(L_\gamma(x)) = \varphi(x_1 + \gamma_1, x_2 + \gamma_2, \dots, x_n + \gamma_n) = \sin\left(2\pi \sum_{j=1}^n k_j(x_j + \gamma_j)\right),$$

and as we have assumed that $\sum_{j=1}^n k_j\gamma_j \in \mathbb{Z}$, we deduce that

$$\varphi(L_\gamma(x)) = \sin\left(2\pi \sum_{j=1}^n k_j x_j + 2\pi \sum_{j=1}^n k_j\gamma_j\right) = \sin\left(2\pi \sum_{j=1}^n k_j x_j\right) = \varphi(x).$$

Hence, φ is L_γ-invariant. To summarize, φ is a nonconstant function, which is invariant under a transitive system. According to Remark 1.5.7(d+e) this is impossible, and thus all k_j must equal 0. Hence, $1, \gamma_1, \gamma_2, \ldots, \gamma_n$ are linearly independent over $\mathbb{Q}$.

For the converse implication, suppose that $1, \gamma_1, \gamma_2, \ldots, \gamma_n$ are linearly independent over $\mathbb{Q}$ and, again for a contradiction, suppose that L_γ is not transitive. By Theorem 1.5.11, we have that L_γ is not topologically mixing. So there exist nonempty open sets U and V contained in $\mathbb{T}^n$ such that

$$\bigcup_{n=1}^{\infty} L_\gamma^n(U) \cap V = \emptyset.$$

This means that $W := \bigcup_{n=1}^{\infty} L_\gamma^n(U) \subseteq X \setminus V$. Then $\overline{W} \subseteq X \setminus V$ since $X \setminus V$ is a closed set. Moreover,

$$L_\gamma(W) = L_\gamma\left(\bigcup_{n=1}^{\infty} L_\gamma^n(U)\right) = \bigcup_{n=1}^{\infty} L_\gamma^{n+1}(U) \subseteq W,$$

and we thus obtain that $\overline{L_\gamma(W)} \subseteq \overline{W}$. Also, the continuity of L_γ ensures that $L_\gamma(\overline{W}) \subseteq \overline{L_\gamma(W)}$. Therefore $L_\gamma(\overline{W}) \subseteq \overline{W}$. We aim to show that this is in fact an equality. To that end, let λ_n denote the Lebesgue measure on $\mathbb{T}^n$, and note that λ_n is translation invariant, which means that

$$\lambda_n(E + v) = \lambda_n(E), \quad \forall E \subseteq \mathbb{T}^n,\ \forall v \in \mathbb{T}^n.$$

So

$$\lambda_n(L_\gamma(\overline{W})) = \lambda_n(\overline{W}). \tag{1.1}$$

If it turned out that $L_\gamma(\overline{W}) \subsetneq \overline{W}$, then there would exist $x \in \overline{W} \setminus L_\gamma(\overline{W})$ and $\varepsilon > 0$ such that $\emptyset \neq B(x, \varepsilon) \cap W \subseteq \overline{W} \setminus L_\gamma(\overline{W})$. But $B(x, \varepsilon) \cap W$ is a nonempty open set, and hence has positive Lebesgue measure. Thus, we would have $\lambda_n(\overline{W} \setminus L_\gamma(\overline{W})) > 0$, which would contradict (1.1). Hence, $L_\gamma(\overline{W}) = \overline{W}$ and, since L_γ is invertible, $L_\gamma^{-1}(\overline{W}) = \overline{W}$.

Now, denote by $\mathbb{1}_A$ the characteristic function of a subset A of $\mathbb{T}^n$, that is,

$$\mathbb{1}_A := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

For any map T and any set A, we have that

$$\mathbb{1}_A \circ T = \mathbb{1}_{T^{-1}(A)}.$$

Since the set $\overline{W}$ is completely L_γ-invariant, we deduce that

$$\mathbb{1}_{\overline{W}} \circ L_\gamma = \mathbb{1}_{L_\gamma^{-1}(\overline{W})} = \mathbb{1}_{\overline{W}}.$$

That is, the function $\mathbb{1}_{\overline{W}}$ is L_γ-invariant.

For every $k \in \mathbb{R}^n$, let $\psi_k : \mathbb{T}^n \to \mathbb{C}$ be defined by

$$\psi_k(x) = e^{2\pi i \langle k,x \rangle} = \cos(2\pi \langle k,x \rangle) + i \sin(2\pi \langle k,x \rangle),$$

where $\langle k,x \rangle = \sum_{j=1}^n k_j x_j$ is the scalar product of the vectors k and x. Then the family $\{\psi_k\}_{k\in\mathbb{Z}^n}$ is an orthonormal basis for the Hilbert space $L^2(\lambda_n)$. Since $\mathbb{1}_{\overline{W}} \in L^2(\lambda_n)$, we can write

$$\mathbb{1}_{\overline{W}}(x) = \sum_{k\in\mathbb{Z}^n} a_k \psi_k(x) = \sum_{k\in\mathbb{Z}^n} a_k e^{2\pi i \langle k,x \rangle} \quad \text{for } \lambda_n\text{-a. e. } x \in \mathbb{T}^n, \tag{1.2}$$

where

$$a_k := \int \mathbb{1}_{\overline{W}}(y) \overline{\psi_k(y)}\, d\lambda_n(y)$$

are the Fourier coefficients of $\mathbb{1}_{\overline{W}}$. Then, for λ_n-a. e. $x \in \mathbb{T}^n$, we have

$$\begin{aligned} \mathbb{1}_{\overline{W}}(x) = \mathbb{1}_{\overline{W}}(L_\gamma(x)) &= \sum_{k\in\mathbb{Z}^n} a_k e^{2\pi i \langle k, x+\gamma \rangle} \\ &= \sum_{k\in\mathbb{Z}^n} a_k e^{2\pi i \langle k,\gamma \rangle} \psi_k(x). \end{aligned} \tag{1.3}$$

Since $\{\psi_k\}_{k\in\mathbb{Z}^n}$ is an orthonormal basis, we deduce from (1.2) and (1.3) that

$$a_k = a_k e^{2\pi i \langle k,\gamma \rangle}$$

for every $k \in \mathbb{Z}^n$. For each such k, this implies that

$$\text{either } a_k = 0 \text{ or } e^{2\pi i \langle k,\gamma \rangle} = 1.$$

In the latter case,

$$\langle k, \gamma \rangle = \sum_{j=1}^n k_j \gamma_j \in \mathbb{Z}.$$

Since the γ_i were assumed to be linearly independent over $\mathbb{Q}$, this implies that $k = (0,0,\ldots,0)$. So, for all $k \neq (0,0,\ldots,0)$, we must be in the former case, that is, we must have that $a_k = 0$. Hence,

$$\mathbb{1}_{\overline{W}}(x) = a_{(0,\ldots,0)} = \lambda_n(\overline{W}) \quad \text{for } \lambda_n\text{-a. e. } x \in \mathbb{T}^n.$$

As W is a nonempty open set, it follows that $\lambda_n(W) > 0$, and thus $\lambda_n(\overline{W}) > 0$. So $\mathbb{1}_{\overline{W}}(x) > 0$ for λ_n-a. e. $x \in \mathbb{T}^n$. However, recall that $\overline{W} \cap V = \emptyset$, and thus $\mathbb{1}_{\overline{W}}(V) = 0$. As V is a nonempty open set, $\lambda_n(V) > 0$ and, therefore, $\mathbb{1}_{\overline{W}}(x) = 0$ on a set of positive measure. We have reached a contradiction. This means that L_γ is transitive. □

1.6.2 Maps of the interval

Although there is no known topological conjugacy invariant which is a complete invariant for the family of all dynamical systems, some conjugacy invariants turn out to be complete invariants for subfamilies of systems. By a complete invariant for a subfamily, we mean that if two systems from the subfamily share this invariant, then they are automatically topologically conjugate.

For instance, we have seen that the number of periodic points of any given period is a topological conjugacy invariant, though not a complete invariant. However, the number of periodic points of a given period sometimes turns out to be a complete invariant if we restrict our attention to an appropriate subfamily of dynamical systems.

In this section, we show that the number of fixed points is a complete invariant for the family of all self-homeomorphisms of compact intervals which fix the endpoints of their domain and whose sets of fixed points are finite. These self-homeomorphisms can be characterized as the strictly increasing continuous self-maps of compact intervals that fix the endpoints of their domain and have only finitely many fixed points.

The proof of the complete invariance of the number of fixed points will be given in several stages.

Theorem 1.6.4. *Any two strictly increasing continuous maps of the unit interval that fix the endpoints of the interval and that have no other fixed points are topologically conjugate.*

Proof. Let $I := [0,1]$, and let $T_1, T_2 : I \to I$ be two strictly increasing continuous maps such that $\mathrm{Fix}(T_1) = \mathrm{Fix}(T_2) = \{0,1\}$. By the intermediate value theorem, each T_i is a surjection. As T_i is a strictly increasing map, it is also injective. Thus, T_i is a bijection and T_i^{-1} exists. Furthermore, T_i^{-1} is a strictly increasing continuous bijection. It follows immediately that T_i^n is a strictly increasing homeomorphism of I for all $n \in \mathbb{Z}$.

Let $x \in (0,1)$. Since $T_1(x) \neq x$, either $T_1(x) < x$ or $T_1(x) > x$. Similarly, either $T_2(x) < x$ or $T_2(x) > x$. Let us consider the case in which $T_1(x) > x$ and $T_2(x) > x$. The proofs of the other three cases are similar and are left to the reader. Then $T_1(y) > y$ for every $y \in (0,1)$; otherwise, the continuous map T_1 would have a fixed point between 0 and 1 by the intermediate value theorem (because of the change of sign in the values of the continuous map $T_1 - \mathrm{Id}_I$). Similarly, $T_2(y) > y$ for each $y \in (0,1)$. As T_i^n is strictly increasing for all $n \in \mathbb{Z}$, it follows easily that for all $m < n$ and all $y \in (0,1)$, we have

$$T_i^m(y) < T_i^n(y)$$

for each $i = 1,2$.

Now, fix $0 < a < 1$. Note once again that $T_i(a) > a$ and let $\triangle_i = [a, T_i(a)]$. We now state and prove three claims that will allow us to complete the proof of the theorem.

Claim 1. *For all $m, n \in \mathbb{Z}$ such that $m < n$, we have*

$$T_i^m(\mathrm{Int}(\triangle_i)) \cap T_i^n(\mathrm{Int}(\triangle_i)) = \emptyset.$$

Proof. Let $j \in \mathbb{Z}$. Since T_i^j is strictly increasing and continuous, we have that

$$T_i^j(\mathrm{Int}(\triangle_i)) = (T_i^j(a), T_i^{j+1}(a)).$$

As $m+1 \le n$, it follows that $T_i^{m+1}(a) \le T_i^n(a)$ and, therefore,

$$T_i^m(\mathrm{Int}(\triangle_i)) \cap T_i^n(\mathrm{Int}(\triangle_i)) = (T_i^m(a), T_i^{m+1}(a)) \cap (T_i^n(a), T_i^{n+1}(a)) = \emptyset. \qquad \square$$

Claim 2. *For all* $x \in (0,1)$, *we have that*

$$\lim_{n\to\infty} T_i^{-n}(x) = 0 \quad \textit{and} \quad \lim_{n\to\infty} T_i^n(x) = 1.$$

Proof. Let $x \in (0,1)$. We shall establish the second limit; the first limit can be ascertained analogously. First, note that the limit must exist because the sequence $(T_i^n(x))_{n=0}^{\infty}$ is (strictly) increasing and bounded above by 1. So, let $y = \lim_{n\to\infty} T_i^n(x)$. According to Lemma 1.1.4, y is a fixed point of T_i. Moreover, y is clearly not equal to 0. Hence, $y = 1$. $\square$

Claim 3.

$$\bigcup_{n=-\infty}^{\infty} T_i^n(\triangle_i) = (0,1).$$

Proof. It is clear that $\bigcup_{n=-\infty}^{\infty} T_i^n(\triangle_i) \subseteq (0,1)$. To prove the opposite inclusion, let $x \in (0,1)$. If $x \in \triangle_i$, we are done. So, let $x \notin \triangle_i$. Then either $x < a$ or $x > T_i(a)$. In the former case, there exists by Claim 2 a largest $n \ge 0$ such that $T_i^n(x) < a$. Therefore, $T_i^{n+1}(x) \ge a$. Moreover, since T_i is strictly increasing, $T_i^n(x) < a$ implies $T_i^{n+1}(x) < T_i(a)$. Hence, $T_i^{n+1}(x) \in \triangle_i$, that is, $x \in T_i^{-(n+1)}(\triangle_i)$. In the latter case, by Claim 2 there exists a largest $n \ge 0$ such that $T_i^{-n}(x) > T_i(a)$. Then $T_i^{-(n+1)}(x) \le T_i(a)$. Moreover, since T_i^{-1} is strictly increasing, $T_i^{-(n+1)}(x) > T_i^{-1}(T_i(a)) = a$. Hence, $T_i^{-(n+1)}(x) \in \triangle_i$, that is, $x \in T_i^{(n+1)}(\triangle_i)$. In all cases, $x \in \bigcup_{n=-\infty}^{\infty} T_i^n(\triangle_i)$.

In order to make the idea behind the sequence of intervals $(T_i^n(\triangle_i))_{n\in\mathbb{Z}}$ clearer, see Figure 1.5.

We will now define a conjugacy map h between T_1 and T_2. First of all, suppose that $H : \triangle_1 \to \triangle_2$ is any homeomorphism satisfying

$$H(a) = a \quad \text{and} \quad H(T_1(a)) = T_2(a).$$

Let $x \in (0,1)$. By Claim 3, there exists $n(x) \in \mathbb{Z}$ such that $x \in T_1^{-n(x)}(\triangle_1)$. We shall shortly observe that $n(x)$ is uniquely defined for all $x \notin \{T_1^n(a) : n \in \mathbb{Z}\}$. When $x \in \{T_1^n(a) : n \in \mathbb{Z}\}$, then $n(x)$ takes the value of any one of two consecutive integers; so it is not uniquely defined and this will require prudence when defining h. Define the conjugacy map h by setting

$$h(x) = \begin{cases} T_2^{-n(x)} \circ H \circ T_1^{n(x)}(x) & \text{if } x \in (0,1) \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x = 1. \end{cases}$$

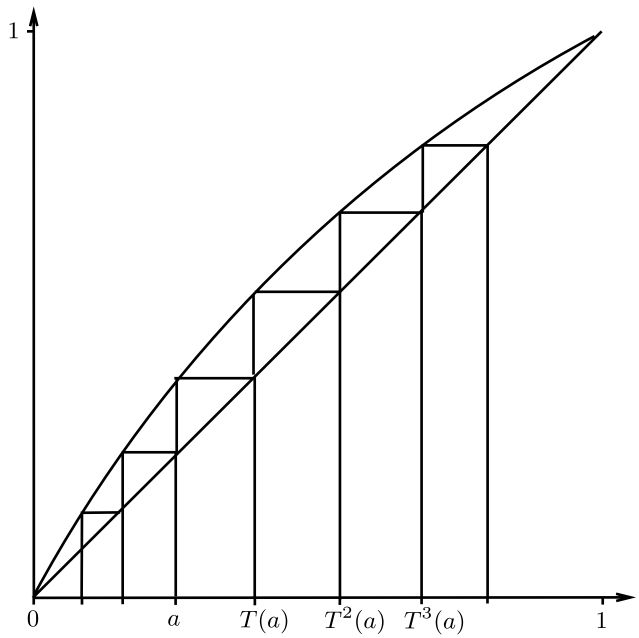

Figure 1.5: The map T is an orientation-preserving homeomorphism of the unit interval that fixes only the endpoints. The vertical dotted lines indicate the intervals $(T^n(\triangle))_{n\in\mathbb{Z}}$, where $\triangle := [a, T(a)]$.

We first check that this map is well-defined. Suppose that $T_1^k(x), T_1^\ell(x) \in \triangle_1$ for some $k < \ell$. Claim 1 above implies that $T_1^k(x), T_1^\ell(x) \in \partial\triangle_1$ and $k+1 = \ell$. It follows that $T_1^k(x) = a$ and $T_1^\ell(x) = T_1^{k+1}(x) = T_1(a)$. Therefore,

$$\begin{aligned}
T_2^{-\ell} \circ H \circ T_1^\ell(x) &= T_2^{-(k+1)}\big(H\big(T_1(a)\big)\big) \\
&= T_2^{-(k+1)}\big(T_2(a)\big) \\
&= T_2^{-k}(a) \\
&= T_2^{-k}\big(H(a)\big) \\
&= T_2^{-k}\big(H\big(T_1^k(x)\big)\big) \\
&= T_2^{-k} \circ H \circ T_1^k(x).
\end{aligned}$$

Thus, the map h is well-defined.

We must also show that $T_2 \circ h = h \circ T_1$. Toward this end, first observe that we have $T_1^{n(x)-1}(T_1(x)) = T_1^{n(x)}(x) \in \triangle_1$, so we can choose $n(T_1(x))$ to be $n(x)-1$, and then we obtain that

$$\begin{aligned}
h \circ T_1(x) &= T_2^{-n(T_1(x))} \circ H \circ T_1^{n(T_1(x))}\big(T_1(x)\big) \\
&= T_2^{-(n(x)-1)} \circ H \circ T_1^{n(x)-1}\big(T_1(x)\big)
\end{aligned}$$

$$= T_2 \circ T_2^{-n(x)} \circ H \circ T_1^{n(x)}(T_1^{-1}(T_1(x)))$$
$$= T_2 \circ h(x).$$

To complete the proof, it remains to show that h is a bijection and that it is continuous (recall that a continuous bijection between compact metrizable spaces is a homeomorphism). That h is continuous follows from the fact that

$$h|_{T_1^{-n}(\triangle_1)} = T_2^{-n} \circ H \circ T_1^n|_{T_1^{-n}(\triangle_1)}$$

for every $n \in \mathbb{Z}$. Indeed, as T_1^n, H, and T_2^{-n} are continuous, the restriction of h to each $T_1^{-n}(\triangle_1)$ is continuous. Using left and right continuity at the endpoints of the intervals $T_1^{-n}(\triangle_1)$, we conclude from Claim 3 that h is continuous on $(0,1)$. The continuity of h at 0 follows from the fact that $\lim_{x\to 0} n(x) = \infty$, that $h(x) \in T_2^{-n(x)}(\triangle_2) = [T_2^{-n(x)}(a), T_2^{-(n(x)-1)}(a)]$, and, by Claim 2, that $\lim_{n\to\infty} T_2^{-n}(a) = 0$. A similar argument establishes the continuity of h at 1.

For the injectivity of h, we shall show that h is strictly increasing. If $0 < x < y < 1$, then $n(x)$ and $n(y)$ can be chosen so that $n(x) \geq n(y)$. If $n(x) = n(y) =: n$, then $h(x) < h(y)$ since the restriction of h to $T_1^{-n}(\triangle_1)$ is strictly increasing (because T_1^n, H and T_2^{-n} are all strictly increasing). If $n(x) > n(y)$, then $T_2^{-n(x)}(\triangle_2)$ lies to the left of $T_2^{-n(y)}(\triangle_2)$ and, as $h(x) \in T_2^{-n(x)}(\triangle_2)$ while $h(y) \in T_2^{-n(y)}(\triangle_2)$, we deduce that $h(x) < h(y)$.

Finally, since h is continuous, $h(0) = 0$ and $h(1) = 1$, the map h is surjective by the intermediate value theorem. □

Corollary 1.6.5. *Any two strictly increasing continuous maps of compact intervals which have the same finite number of fixed points, including both endpoints of their respective domains, are topologically conjugate.*

Proof. We first prove that if $f : I \to I$ is a strictly increasing continuous map which only fixes the points 0 and 1 and if $g : [a,b] \to [a,b]$ is a strictly increasing continuous map which only fixes a and b, then f and g are topologically conjugate. Let $k : [a,b] \to I$ be defined by $k(x) = \frac{x-a}{b-a}$. Then k is a strictly increasing homeomorphism. Consequently, $k \circ g \circ k^{-1}$ is a strictly increasing continuous map of I which only fixes 0 and 1, just like f. By Theorem 1.6.4, there is a conjugacy map $h : I \to I$ between f and $k \circ g \circ k^{-1}$ (i. e., h is a homeomorphism such that $h \circ f = k \circ g \circ k^{-1} \circ h$). It follows that $k^{-1} \circ h : I \to [a,b]$ is a conjugacy map between f and g.

Recall that topological conjugacy is an equivalence relation. It follows from the argument above and the transitivity of topological conjugacy that any two strictly increasing continuous maps of compact intervals which only fix the endpoints of their respective domains, are topologically conjugate.

Now, let $T : I \to I$ and $S : [a,b] \to [a,b]$ be strictly increasing continuous maps with the same finite number of fixed points, which include the endpoints of their respective domains. Denote the sets of fixed points of T and S by $\{0, x_{T_1}, \ldots, x_{T_{k-1}}, 1\}$ and

$\{a, x_{S_1}, \ldots, x_{S_{k-1}}, b\}$, respectively. These fixed points induce two finite sequences of maps

$$\begin{array}{rccclrccc} T_0: & [0, x_{T_1}] & \to & [0, x_{T_1}] & \quad & S_0: & [a, x_{S_1}] & \to & [a, x_{S_1}] \\ T_1: & [x_{T_1}, x_{T_2}] & \to & [x_{T_1}, x_{T_2}] & & S_1: & [x_{S_1}, x_{S_2}] & \to & [x_{S_1}, x_{S_2}] \\ & & \vdots & & & & & \vdots & \\ T_{k-1}: & [x_{T_{k-1}}, 1] & \to & [x_{T_{k-1}}, 1] & & S_{k-1}: & [x_{S_{k-1}}, b] & \to & [x_{S_{k-1}}, b] \end{array}$$

maps which are the restrictions of T and S, and hence are strictly increasing continuous maps of compact intervals which only fix the endpoints of their domains. For each $0 \le i < k$, let h_i be a conjugacy map between T_i and S_i (such a map exists according to the discussion in the previous paragraph). Then set $h : I \to [a, b]$ to be the map defined by $h(x) = h_i(x)$ when $x \in [x_{T_i}, x_{T_{i+1}}]$. This map h is clearly a bijection and is continuous (the continuity of h at the points $\{x_{T_1}, \ldots, x_{T_{k-1}}\}$ can be established by means of left and right continuity). Finally, $h \circ T = S \circ h$ since $h_i \circ T_i = S_i \circ h_i$ for all $0 \le i < k$.

Finally, it follows from this and the transitivity of topological conjugacy that any two strictly increasing continuous maps of compact intervals with the same finite number of fixed points, among whose are the endpoints of their respective domains, are topologically conjugate. □

However, note that any two strictly increasing continuous maps of compact intervals which have the same finite number of fixed points, one of which fixes the endpoints of its domain while the other does not, are not topologically conjugate (see Exercise 1.7.34). In particular, this establishes that the number of fixed points is not a complete invariant for the subfamily of all strictly increasing continuous maps of compact intervals which have the same given finite number of fixed points.

1.7 Exercises

Exercise 1.7.1. In this exercise, we revisit Example 1.1.3(b). Recall that, given $m \in \mathbb{N}$, we defined the map $T_m : \mathbb{S}^1 \to \mathbb{S}^1$ by setting $T_m(x) := mx \pmod 1$. The map T_m is simply a piecewise linear map that sends each interval $[i/m, (i+1)/m]$, for $0 \le i < m$, onto $\mathbb{S}^1$. It can be expressed by the formula

$$T_m(x) = mx - i \quad \text{for all } \frac{i}{m} \le x \le \frac{i+1}{m}.$$

Prove that for every $n \in \mathbb{N}$ the iterates of T can be expressed as

$$T_m^n(x) = m^n x - \sum_{k=1}^{n} m^{n-k} i_{n-k+1}$$

if

$$\frac{1}{m^n} \sum_{k=1}^{n} m^{n-k} i_{n-k+1} \le x \le \frac{1}{m^n}\left(\sum_{k=1}^{n} m^{n-k} i_{n-k+1} + 1\right),$$

where $0 \le i_1, i_2, \ldots, i_n < m$. Deduce that T_m has m^n periodic points of period n.

Exercise 1.7.2. Show that a point $x \in X$ is preperiodic for a system $T : X \to X$ if and only if its forward orbit $\mathcal{O}_+(x)$ is finite.

Exercise 1.7.3. Let $\mathrm{Per}(T)$ be the set of periodic points for a system $T : X \to X$. Prove that

$$\mathrm{PrePer}(T) = \bigcup_{x\in\mathrm{Per}(T)} \mathcal{O}_-(x),$$

where $\mathrm{PrePer}(T)$ denotes the set of all preperiodic points for the system T.

Exercise 1.7.4. Identify all the preperiodic points for the dynamical systems introduced in Example 1.1.3.

Exercise 1.7.5. Prove that if both X and Y are dense subsets of $\mathbb{R}$ and $g : X \to Y$ is an increasing bijection, then g extends uniquely to an increasing homeomorphism $\tilde{g} : \mathbb{R} \to \mathbb{R}$.

Exercise 1.7.6. Prove that topological conjugacy defines an equivalence relation on the space of dynamical systems.

Exercise 1.7.7. Show that if two dynamical systems $T : X \to X$ and $S : Y \to Y$ are topologically conjugate via a conjugacy map $h : X \to Y$, then their corresponding iterates are topologically conjugate by means of the same conjugacy map h.

Exercise 1.7.8. Prove that for every $n \in \mathbb{N}$ there exists a one-to-one correspondence between the periodic points of period n of two topologically conjugate dynamical systems. Show that this implies for every $n \in \mathbb{N}$ the existence of a one-to-one correspondence between the periodic points of prime period n.

Exercise 1.7.9. Prove that if two dynamical systems $T : X \to X$ and $S : Y \to Y$ are topologically conjugate via a conjugacy map $h : X \to Y$, then h induces a one-to-one correspondence between preperiodic points. Deduce that the number of preperiodic points is a topological conjugacy invariant. By means of an example, show also that the number of preperiodic points is not a complete invariant.

Exercise 1.7.10. Suppose that a dynamical system $S : Y \to Y$ is a factor of a system $T : X \to X$. Show that every orbit of T is projected onto an orbit of S. Show also that for all $n \in \mathbb{N}$ every periodic point of period n for T is mapped to a periodic point for S whose period is a factor of n.

Exercise 1.7.11. Prove that the closure of every invariant set is invariant.

Exercise 1.7.12. Show that if $x \in X$ is a periodic point for a system $T : X \to X$, then $\omega(x) = \mathcal{O}_+(x) = \overline{\mathcal{O}_+(x)}$. Observe also that the set of limit points of $\mathcal{O}_+(x)$ is empty. Deduce that $\omega(x)$ does not coincide with the set of limit points of $\mathcal{O}_+(x)$.

Exercise 1.7.13. Prove that if $x \in X$ is a preperiodic point for a system $T : X \to X$, then $\omega(x) \neq \mathcal{O}_+(x) = \overline{\mathcal{O}_+(x)}$. So $\omega(x) \neq \overline{\mathcal{O}_+(x)}$. Moreover, as in Exercise 1.7.12, prove that $\omega(x)$ does not coincide with the set of limit points of the forward orbit $\mathcal{O}_+(x)$.

Exercise 1.7.14. Let $T : X \to X$ be a dynamical system. Prove that $\omega(x)$ is the set of limit points of $\mathcal{O}_+(x)$ if and only if x is not a periodic or preperiodic point.

Exercise 1.7.15. Let $T : X \to X$ be a dynamical system. Show that $\mathcal{O}_+(x) \cup \omega(x) = \overline{\mathcal{O}_+(x)}$ for any $x \in X$.

Exercise 1.7.16. Show that the set of limit points of any set is closed.

Hint: Prove that any accumulation point of accumulation points of a set S is an accumulation point of S.

Exercise 1.7.17. Prove that any intersection of a descending sequence of nonempty compact sets in a Hausdorff topological space is a nonempty compact set.

Exercise 1.7.18. Show that every minimal system is surjective.

Exercise 1.7.19. Prove that minimality is not a complete invariant.

Hint: Construct two finite minimal dynamical systems with different cardinalities.

Exercise 1.7.20. Prove that minimality is not a complete invariant for infinite systems.

Exercise 1.7.21. Let $+_2 : \{0,1\} \to \{0,1\}$ denote addition modulo 2 and endow the set $\{0,1\}$ with the discrete topology. Prove that the dynamical system $T : \mathbb{S}^1 \times \{0,1\} \to \mathbb{S}^1 \times \{0,1\}$ given by the formula

$$T(x,y) := \big(x + \sqrt{2}\,(\mathrm{mod}\ 1), y +_2 1\big)$$

is minimal.

Exercise 1.7.22. Prove or disprove (by providing a counterexample) that if $T : X \to X$ is minimal then $T^2 : X \to X$ is also minimal.

Exercise 1.7.23. Prove that the statements in Definition 1.5.10 are equivalent.

Exercise 1.7.24. Prove that topological transitivity, strong transitivity, very strong transitivity, and topological exactness are topological conjugacy invariants.

Exercise 1.7.25. Construct a very strongly transitive, open system which is not topologically exact. (Recall that a map is said to be open if the image of any open set under that map is open.)

Exercise 1.7.26. Build a strongly transitive system which is not very strongly transitive.

Exercise 1.7.27. Find a transitive system which is not strongly transitive.

Exercise 1.7.28. Let $T : X \to X$ be a dynamical system. Suppose that for every nonempty open subset U of X there exists $n \geq 0$ such that $T^n(U)$ is a dense subset of X. Prove that T is topologically exact.

Exercise 1.7.29. A continuous map $T : X \to X$ is said to be *locally eventually onto* provided that for every nonempty open subset U of X there exists $n \geq 0$ such that

$$\bigcup_{j=0}^{n} T^j(U) = X.$$

Each topologically exact map is locally eventually onto. Provide an example of a locally eventually onto map that is not topologically exact.

Exercise 1.7.30. Let $(k_1, k_2, \ldots, k_n) \in \mathbb{Z}^n$. Let $\varphi : \mathbb{T}^n \to \mathbb{R}$ be the function defined by

$$\varphi(x) = \varphi(x_1, x_2, \ldots, x_n) = \sin\left(2\pi \sum_{j=1}^{n} k_j x_j\right).$$

Show that φ is well-defined.

Exercise 1.7.31. Let $T : X \to X$ be a dynamical system. Prove that a subset Y of X is T-invariant if and only if $\mathbb{1}_Y$ is T-invariant, where $\mathbb{1}_Y$ denotes the characteristic function of the set Y, that is,

$$\mathbb{1}_Y(x) := \begin{cases} 1 & \text{if } x \in Y \\ 0 & \text{if } x \notin Y. \end{cases}$$

Exercise 1.7.32. Establish graphically that Claims 1, 2, and 3 in the proof of Theorem 1.6.4 hold.

Exercise 1.7.33. Prove Theorem 1.6.4 when $T_1(x) > x$ for all $x \in (0,1)$ and $T_2(x) < x$ for all $x \in (0,1)$.

Hint: Prove that Claim 1 still holds. Prove that Claim 2 holds for T_1. However, show that $\lim_{n\to\infty} T_2^n(x) = 0$ and $\lim_{n\to\infty} T_2^{-n}(x) = 1$. Then prove that Claim 3 holds. Finally, show that

$$h(x) = \begin{cases} T_2^{-n(x)} \circ H \circ T_1^{n(x)}(x) & \text{if } x \in (0,1) \\ 1 & \text{if } x = 0 \\ 0 & \text{if } x = 1. \end{cases}$$

is a conjugacy map between T_1 and T_2.

Exercise 1.7.34. Prove that any two strictly increasing continuous maps of compact intervals which have the same finite number of fixed points, one of which fixes the endpoints of its domain while the other does not, are not topologically conjugate.

Hint: Let $T : [a,b] \to [a,b]$ and $S : [c,d] \to [c,d]$ be strictly increasing continuous maps with T fixing both a and b. Suppose that $h : [a,b] \to [c,d]$ is a conjugacy map between T and S. Using the Intermediate Value Theorem (IVT), show that $h(a)$ is an extreme (in other words, the leftmost or rightmost) fixed point of S, while $h(b)$ is the other extreme fixed point. Using the IVT once more, show that $h([a,b]) = [h(a), h(b)]$. Deduce that S fixes both c and d.

2 Homeomorphisms of the circle

In this chapter, we temporarily step away from the general theory of dynamical systems to consider more specific examples. In the preparatory Section 2.1, we first study lifts of maps of the unit circle. Using lifts, we investigate homeomorphisms of the unit circle in Section 2.2. These homeomorphisms constitute the primary class of systems of interest in this chapter. After showing that rotations are homeomorphisms, we introduce Poincaré's notion of *rotation number* for homeomorphisms of the circle. Roughly speaking, this number is the average rotation that a homeomorphism induces on the points of the circle over the long term. In Section 2.3, we examine in more detail diffeomorphisms of the circle. The main result of this chapter is Denjoy's theorem (Theorem 2.3.4), which states that if a C^2 diffeomorphism has an irrational rotation number, then the diffeomorphism constitutes a minimal system which is topologically conjugate to an irrational rotation. Strictly speaking, it suffices that the modulus of the diffeomorphism's derivative be a function of bounded variation. Denjoy's theorem is a generalization of Theorem 1.5.12.

The concept of rotation number generalizes to all continuous degree-one self-maps of the circle. It is then called *rotation interval*. A systematic account of the theory of such maps and, in particular, an extended treatment of the rotation interval, can be found in [3].

2.1 Lifts of circle maps

In this section, we discuss general properties that are shared by all circle maps, be they one-to-one or not. As already mentioned in Section 1.5, the unit circle $\mathbb{S}^1$ can be defined in many homeomorphic ways. Here, we will regard $\mathbb{S}^1$ as the quotient space $\mathbb{R}/\mathbb{Z}$, that is, as the space of all equivalence classes

$$[x] = \{x + n : n \in \mathbb{Z}\},$$

where $x \in \mathbb{R}$, with metric

$$\begin{aligned}\rho([x],[y]) &= \inf\{\big|(x+n)-(y+m)\big| : n,m \in \mathbb{Z}\} = \inf\{|x-y+k| : k \in \mathbb{Z}\}\\ &= \min\{|x-y-1|, |x-y|, |x-y+1|\}.\end{aligned}$$

To study the dynamics of a map on the circle, it is helpful to *lift* that map from $\mathbb{S}^1 \cong \mathbb{R}/\mathbb{Z}$ to $\mathbb{R}$. This can be done via the continuous surjection

$$\begin{array}{rcccl}\pi & : & \mathbb{R} & \longrightarrow & \mathbb{S}^1\\ & & x & \longmapsto & \pi(x) = [x].\end{array}$$

Definition 2.1.1. Let $T : \mathbb{S}^1 \to \mathbb{S}^1$ be a continuous map of the circle. A continuous map $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ is called a *lift* of T to $\mathbb{R}$ if $\pi \circ \widetilde{T} = T \circ \pi$, that is, if the following diagram

https://doi.org/10.1515/9783110702682-002

commutes:

$$\begin{CD}
\mathbb{R} @>{\widetilde{T}}>> \mathbb{R}\\
@V{\pi}VV @VV{\pi}V\\
\mathbb{S}^1 @>>{T}> \mathbb{S}^1
\end{CD}$$

In other words, $T : \mathbb{S}^1 \to \mathbb{S}^1$ is a factor of $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ via the factor map π.

Note that π is countably infinite to one, and hence cannot be a conjugacy map. In fact, $\pi : \mathbb{R} \to \mathbb{S}^1$ is a *covering map*, that is, every $[x] \in \mathbb{S}^1$ has an open neighborhood $U_{[x]}$, which is *evenly covered* by π. Being evenly covered by π means that the preimage $\pi^{-1}(U_{[x]})$ is a union of disjoint open subsets of $\mathbb{R}$, called *sheets* of $\pi^{-1}(U_{[x]})$, each of which is mapped homeomorphically by π onto $U_{[x]}$. Observe further that π is a local isometry when $\mathbb{S}^1 \cong \mathbb{R}/\mathbb{Z}$ is equipped with the metric ρ. More precisely, it is an isometry on any interval of length at most 1/2.

Lemma 2.1.2. *Every continuous map $T : \mathbb{S}^1 \to \mathbb{S}^1$ admits a lift $\widetilde{T} : \mathbb{R} \to \mathbb{R}$.*

Proof. Fix $s_0 \in \mathbb{R}$ and $t_0 \in \pi^{-1}(T(\pi(s_0)))$. Define $\widetilde{T}(s_0) := t_0$. Then $\pi(\widetilde{T}(s_0)) = \pi(t_0) = T(\pi(s_0))$. In other words, $\widetilde{T}(s_0)$ is a lift of s_0. This is the starting point of our lift. By considering successive neighborhoods, we will gradually lift the points of $\mathbb{S}^1$ to $\mathbb{R}$. For each $t \in \mathbb{R}$, let $U_{\pi(t)} \subseteq \mathbb{S}^1$ be the largest open neighborhood centered on $\pi(t)$ which is evenly covered by π, that is,

$$U_{\pi(t)} = \{\pi(r) : t - 1/2 < r < t + 1/2\}.$$

We define the sought-after lift by successive steps, as follows. Let V_0 be the unique sheet of $\pi^{-1}(U_{\pi(\widetilde{T}(s_0))})$, which contains $\widetilde{T}(s_0)$, that is, let

$$V_0 := \{r \in \mathbb{R} : \widetilde{T}(s_0) - 1/2 < r < \widetilde{T}(s_0) + 1/2\}.$$

Since $T(\pi(s_0)) = \pi(\widetilde{T}(s_0)) \in U_{\pi(\widetilde{T}(s_0))}$, since T is continuous and since $U_{\pi(\widetilde{T}(s_0))}$ is open, there exists $s_1' > s_0$ such that $T(\pi(s)) \in U_{\pi(\widetilde{T}(s_0))}$ for all $s_0 \le s < s_1'$. Denote by s_1 the supremum of all such s_1'. For each $s_0 \le s < s_1$, define $\widetilde{T}(s)$ to be the unique point of V_0 such that $\pi(\widetilde{T}(s)) = T(\pi(s))$.

If $s_1 = \infty$, then the lift is defined for all $s \ge s_0$. If $s_1 < \infty$, define

$$\widetilde{T}(s_1) := \lim_{s \nearrow s_1} \widetilde{T}(s).$$

Then

$$\widetilde{T}(s_1) \in \{\widetilde{T}(s_0) \pm 1/2\}$$

and

$$\pi\big(\widetilde{T}(s_1)\big) = T\big(\pi(s_1)\big).$$

Just like we did from s_0, the map $\widetilde{T}$ can then be extended beyond s_1 as follows. Let V_1 be the unique sheet of $\pi^{-1}(U_{\pi(\widetilde{T}(s_1))})$ which contains $\widetilde{T}(s_1)$. Since $T(\pi(s_1)) = \pi(\widetilde{T}(s_1)) \in U_{\pi(\widetilde{T}(s_1))}$, since T is continuous and since $U_{\pi(\widetilde{T}(s_1))}$ is open, there exists $s_2' > s_1$ such that $T(\pi(s)) \in U_{\pi(\widetilde{T}(s_1))}$ for all $s_1 \leq s < s_2'$. Denote by s_2 the supremum of all such s_2'. For each $s_1 \leq s < s_2$, let $\widetilde{T}(s)$ be the unique point of V_1 such that $\pi(\widetilde{T}(s)) = T(\pi(s))$. If $s_2 < \infty$, define

$$\widetilde{T}(s_2) := \lim_{s \nearrow s_2} \widetilde{T}(s).$$

Then

$$\widetilde{T}(s_2) \in \{\widetilde{T}(s_1) \pm 1/2\} \subseteq \{\widetilde{T}(s_0) \pm k/2 : k = 0, 1, 2\}$$

and

$$\pi\big(\widetilde{T}(s_2)\big) = T\big(\pi(s_2)\big).$$

Continuing in this way, either the procedure ends with some $s_n = \infty$ or it does not, in which case a strictly increasing sequence $(s_n)_{n=1}^{\infty}$ is constructed recursively. We claim that $\lim_{n\to\infty} s_n = \infty$. Otherwise, let $s^* = \lim_{n\to\infty} s_n < \infty$. The continuity of T and π then ensures that $T(\pi(s^*)) = \lim_{n\to\infty} T(\pi(s_n)) = \lim_{n\to\infty} \pi(\widetilde{T}(s_n))$. Note also that $\pi(\widetilde{T}(s_n))$ coincides with $\pi(\widetilde{T}(s_0))$ or $\pi(\widetilde{T}(s_0) + 1/2)$ since $\widetilde{T}(s_n) \in \{\widetilde{T}(s_0) \pm k/2 : k \in \mathbb{Z}\}$. Therefore, $T(\pi(s^*))$ coincides with $\pi(\widetilde{T}(s_0))$ or $\pi(\widetilde{T}(s_0) + 1/2)$, and thus the sequence $\pi(\widetilde{T}(s_n))$, $n \geq 0$, is eventually constant. But this is impossible since $\pi(\widetilde{T}(s_{n+1})) \notin U_{\pi(\widetilde{T}(s_n))}$ for all $n \geq 0$ by definition. This contradiction means that the lift can be extended indefinitely to the right of s_0, and a similar argument shows that it can be indefinitely extended to the left as well. □

Reading the proof of the above lemma, the reader may have acquired the intuition that, given a starting point, the lift of a map is unique. Moreover, given that starting points can only differ by an integer, so should entire lifts. This intuition proves to be correct.

Lemma 2.1.3. *Let $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ be a lift of a continuous map $T : \mathbb{S}^1 \to \mathbb{S}^1$. Then $\widehat{T} : \mathbb{R} \to \mathbb{R}$ is a lift of T if and only if $\widehat{T} = \widetilde{T} + k$ for some $k \in \mathbb{Z}$. In particular, given $s \in \mathbb{R}$ and $t \in \pi^{-1}(T(\pi(s)))$, there is a unique lift $\widehat{T}$ so that $\widehat{T}(s) = t$.*

Proof. Suppose first that $\widehat{T} = \widetilde{T} + k$ for some $k \in \mathbb{Z}$. It follows that $\widehat{T}$ is continuous since $\widetilde{T}$ is continuous. Moreover, for every $x \in \mathbb{R}$ we have

$$\pi \circ \widehat{T}(x) = \pi\big(\widetilde{T}(x) + k\big) = \pi\big(\widetilde{T}(x)\big) = T \circ \pi(x).$$

Thus, $\widehat{T}$ is a lift of T. This proves one implication.

For the converse implication, suppose that $\widehat{T}$, like $\widetilde{T}$, is a lift of T. For every $x \in \mathbb{R}$, we then have

$$\pi \circ \widehat{T}(x) = T \circ \pi(x) = \pi \circ \widetilde{T}(x).$$

Therefore, $\widehat{T}(x) - \widetilde{T}(x) \in \mathbb{Z}$ for every $x \in \mathbb{R}$. Define the function $k : \mathbb{R} \to \mathbb{Z}$ by $k(x) = \widehat{T}(x) - \widetilde{T}(x)$. Since both $\widetilde{T}$ and $\widehat{T}$ are continuous on $\mathbb{R}$, so is the function k. Then $k(\mathbb{R})$, as the image of a connected set under a continuous function, is a connected set. But since $\mathbb{Z}$ is totally disconnected, the set $k(\mathbb{R})$ must be a singleton. In other words, the function k must be constant. Hence, $\widehat{T} = \widetilde{T} + k$ for some constant $k \in \mathbb{Z}$. □

Thus, once a lift is found, all the other lifts can be obtained by translating vertically the graph of the original lift by all the integers. In the following lemma, we shall describe an important property that all lifts have in common.

Lemma 2.1.4. *If $T : \mathbb{S}^1 \to \mathbb{S}^1$ is continuous, then the number $\widetilde{T}(x+1) - \widetilde{T}(x)$ is an integer independent of the point $x \in \mathbb{R}$ and of the choice of lift $\widetilde{T}$.*

Proof. For every $x \in \mathbb{R}$,

$$\pi(\widetilde{T}(x+1)) = T(\pi(x+1)) = T(\pi(x)) = \pi(\widetilde{T}(x)).$$

Thus $\widetilde{T}(x+1) - \widetilde{T}(x)$ is an integer. Since $\mathbb{R} \ni x \mapsto \widetilde{T}(x+1) - \widetilde{T}(x) \in \mathbb{Z}$ is a continuous function, it follows, as in the proof of Lemma 2.1.3, that it is constant. This implies the independence from the point $x \in \mathbb{R}$.

If $\widehat{T} : \mathbb{R} \to \mathbb{R}$ is another lift of T, then $\widehat{T} = \widetilde{T} + k$ for some $k \in \mathbb{Z}$ according to Lemma 2.1.3. Therefore,

$$\begin{aligned}(\widehat{T}(x+1) - \widehat{T}(x)) - (\widetilde{T}(x+1) - \widetilde{T}(x))&\\ &= (\widehat{T}(x+1) - \widetilde{T}(x+1)) - (\widehat{T}(x) - \widetilde{T}(x))\\ &= k - k = 0.\end{aligned}$$

This establishes the independence from the choice of lift. □

It then makes sense to introduce the following notion.

Definition 2.1.5. If $T : \mathbb{S}^1 \to \mathbb{S}^1$ is continuous, then the integer number $\widetilde{T}(x+1) - \widetilde{T}(x)$, which is independent of the point $x \in \mathbb{R}$ and of the choice of lift $\widetilde{T}$, is called the *degree* of the map T and is denoted by $\deg(T)$.

We can now reformulate Lemma 2.1.4 as follows.

Lemma 2.1.6. *If $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ is a lift of the continuous map $T : \mathbb{S}^1 \to \mathbb{S}^1$, then*

$$\widetilde{T}(x+1) = \widetilde{T}(x) + \deg(T), \quad \forall x \in \mathbb{R}.$$

By way of an induction argument, this result yields the following corollary.

Corollary 2.1.7. *If $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ is a lift of the continuous map $T : \mathbb{S}^1 \to \mathbb{S}^1$, then*

$$\widetilde{T}(x+k) = \widetilde{T}(x) + k\deg(T), \quad \forall x \in \mathbb{R},\ \forall k \in \mathbb{Z}.$$

Proof. Let $d = \deg(T)$. By Lemma 2.1.6, the statement holds for all $x \in \mathbb{R}$ when $k = 1$. Suppose that the statement holds for all $x \in \mathbb{R}$ for some $k \in \mathbb{N}$. Then

$$\widetilde{T}(x+k+1) = \widetilde{T}\big((x+k)+1\big) = \widetilde{T}(x+k) + d = \big(\widetilde{T}(x) + kd\big) + d = \widetilde{T}(x) + (k+1)d$$

for all $x \in \mathbb{R}$. Thus, the statement holds for $k+1$ whenever it holds for $k \in \mathbb{N}$. By induction, the statement holds for all $x \in \mathbb{R}$ and all $k \in \mathbb{N}$. When $k \le 0$, we have that $-k \ge 0$ and, since the statement holds for $-k$, we obtain

$$\widetilde{T}(x+k) = \widetilde{T}(x+k) + (-k)d + kd = \widetilde{T}\big(x+k+(-k)\big) + kd = \widetilde{T}(x) + kd$$

for all $x \in \mathbb{R}$. □

The degree has the following property relative to the composition of maps.

Lemma 2.1.8. *If $S, T : \mathbb{S}^1 \to \mathbb{S}^1$ are continuous maps of the unit circle, then*

$$\deg(S \circ T) = \deg(S) \cdot \deg(T).$$

Proof. Let $\widetilde{S}$, $\widetilde{T}$ be lifts of S and T, respectively. Then $\widetilde{S} \circ \widetilde{T}$ is a lift of $S \circ T$ since

$$\pi \circ (\widetilde{S} \circ \widetilde{T}) = (\pi \circ \widetilde{S}) \circ \widetilde{T} = (S \circ \pi) \circ \widetilde{T} = S \circ (\pi \circ \widetilde{T}) = S \circ (T \circ \pi) = (S \circ T) \circ \pi.$$

Let $x \in \mathbb{R}$. Using Corollary 2.1.7 once for T and once for S, we obtain that

$$\begin{aligned}
\deg(S \circ T) &= \widetilde{S} \circ \widetilde{T}(x+1) - \widetilde{S} \circ \widetilde{T}(x) \\
&= \widetilde{S}\big(\widetilde{T}(x+1)\big) - \widetilde{S}\big(\widetilde{T}(x)\big) \\
&= \widetilde{S}\big(\widetilde{T}(x) + \deg(T)\big) - \widetilde{S}\big(\widetilde{T}(x)\big) \\
&= \widetilde{S}\big(\widetilde{T}(x)\big) + \deg(T) \cdot \deg(S) - \widetilde{S}\big(\widetilde{T}(x)\big) \\
&= \deg(S) \cdot \deg(T).
\end{aligned}$$ □

This has the following consequence for iterates of maps.

Corollary 2.1.9. *If $T : \mathbb{S}^1 \to \mathbb{S}^1$ is a continuous map of the unit circle, then*

$$\deg(T^n) = \big(\deg(T)\big)^n, \quad \forall n \in \mathbb{N}.$$

As a direct repercussion of Corollaries 2.1.7 and 2.1.9, we have the following fact.

Corollary 2.1.10. *If $T : \mathbb{S}^1 \to \mathbb{S}^1$ is continuous and $\widetilde{T}$ is a lift of T, then*

$$\widetilde{T}^n(x+k) = \widetilde{T}^n(x) + k\big(\deg(T)\big)^n$$

for all $x \in \mathbb{R}$, all $k \in \mathbb{Z}$ and all $n \in \mathbb{N}$.

We can further describe the difference between the values of iterates of various lifts.

Corollary 2.1.11. *Let $T : \mathbb{S}^1 \to \mathbb{S}^1$ be a continuous map and $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ be a lift of T. If $\widehat{T}$ is another lift of T so that $\widehat{T} = \widetilde{T} + k$ for some $k \in \mathbb{Z}$, then*

$$\widehat{T}^n = \widetilde{T}^n + k \sum_{j=0}^{n-1} (\deg(T))^j$$

for all $n \in \mathbb{N}$.

Proof. By hypothesis, the statement holds when $n = 1$. Suppose now that it holds for some $n \in \mathbb{N}$. Let $x \in \mathbb{R}$. Then

$$\begin{aligned}
\widehat{T}^{n+1}(x) &= \widehat{T}^n(\widehat{T}(x)) \\
&= \widetilde{T}^n(\widehat{T}(x)) + k \sum_{j=0}^{n-1} (\deg(T))^j \\
&= \widetilde{T}^n(\widetilde{T}(x) + k) + k \sum_{j=0}^{n-1} (\deg(T))^j \\
&= \widetilde{T}^n(\widetilde{T}(x)) + k(\deg(T))^n + k \sum_{j=0}^{n-1} (\deg(T))^j \\
&= \widetilde{T}^{n+1}(x) + k \sum_{j=0}^{n} (\deg(T))^j.
\end{aligned}$$

The result follows by induction. □

We now observe that the degree, as a map, is locally constant.

Lemma 2.1.12. *If $C(\mathbb{S}^1, \mathbb{S}^1)$ is endowed with the topology of uniform convergence, then the degree map $C(\mathbb{S}^1, \mathbb{S}^1) \ni T \mapsto \deg(T) \in \mathbb{Z}$ is locally constant and hence continuous.*

Proof. We shall regard $\mathbb{S}^1$ as $(\mathbb{R}/\mathbb{Z}, \rho)$, with the metric ρ as defined at the beginning of this section. Let $T, S \in C(\mathbb{S}^1, \mathbb{S}^1)$ be such that $\rho_\infty(T, S) < 1/4$, where we recall that

$$\rho_\infty(T, S) = \sup\{\rho(T(x), S(x)) : x \in \mathbb{S}^1\}.$$

Let $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ be a lift of T. Since $\rho_\infty(T, S) < 1/4$, the restriction of π to the real interval $[\widetilde{T}(0) - \rho_\infty(T, S), \widetilde{T}(0) + \rho_\infty(T, S)]$ is an isometry. Thus, the connected set $\pi([\widetilde{T}(0) - \rho_\infty(T, S), \widetilde{T}(0) + \rho_\infty(T, S)])$, which is "centered" on $\pi(\widetilde{T}(0)) = T(\pi(0)) \in \mathbb{S}^1$, contains the point $S(\pi(0))$, since $\rho(T(\pi(0)), S(\pi(0))) \le \rho_\infty(T, S)$. Therefore, there exists some

$$t \in [\widetilde{T}(0) - \rho_\infty(T, S), \widetilde{T}(0) + \rho_\infty(T, S)]$$

such that $\pi(t) = S(\pi(0))$. There then exists a unique lift $\widetilde{S} : \mathbb{R} \to \mathbb{R}$ of S such that $\widetilde{S}(0) = t$. In particular,

$$\left|\widetilde{T}(0) - \widetilde{S}(0)\right| \leq \rho_\infty(T, S).$$

We shall prove that

$$\left|\widetilde{T}(x) - \widetilde{S}(x)\right| \leq \rho_\infty(T, S), \quad \forall x \in \mathbb{R}. \tag{2.1}$$

Indeed, suppose by way of contradiction that there exists $y \in \mathbb{R}$ such that $|\widetilde{T}(y) - \widetilde{S}(y)| > \rho_\infty(T, S)$. Since the function $x \mapsto |\widetilde{T}(x) - \widetilde{S}(x)|$ is continuous and $|\widetilde{T}(0) - \widetilde{S}(0)| \leq \rho_\infty(T, S) < 1/4$, there must exist $w \in \mathbb{R}$ such that

$$\rho_\infty(T, S) < \left|\widetilde{T}(w) - \widetilde{S}(w)\right| < 1/4.$$

But since the restriction of π to $(\widetilde{T}(w) - 1/4, \widetilde{T}(w) + 1/4)$ is an isometry and $\widetilde{S}(w)$ belongs to that real interval, it follows that

$$\rho_\infty(T, S) \geq \rho\big(T(\pi(w)), S(\pi(w))\big) = \rho\big(\pi(\widetilde{T}(w)), \pi(\widetilde{S}(w))\big) = \left|\widetilde{T}(w) - \widetilde{S}(w)\right| > \rho_\infty(T, S),$$

which is impossible. This proves (2.1). Using that formula, we deduce that

$$\begin{aligned}
\left|\deg(T) - \deg(S)\right| &= \left|\big(\widetilde{T}(x+1) - \widetilde{T}(x)\big) - \big(\widetilde{S}(x+1) - \widetilde{S}(x)\big)\right| \\
&= \left|\big(\widetilde{T}(x+1) - \widetilde{S}(x+1)\big) - \big(\widetilde{T}(x) - \widetilde{S}(x)\big)\right| \\
&\leq \left|\widetilde{T}(x+1) - \widetilde{S}(x+1)\right| + \left|\widetilde{T}(x) - \widetilde{S}(x)\right| \\
&\leq 2\rho_\infty(T, S) < 1.
\end{aligned}$$

As $\deg(S)$ and $\deg(T)$ are integers, we conclude that $\deg(S) = \deg(T)$. □

Remark 2.1.13. Another way of stating (2.1) is to say that if $T, S \in C(\mathbb{S}^1, \mathbb{S}^1)$ are circle maps such that $\rho_\infty(T, S) < 1/4$, then T and S have lifts $\widetilde{T}$ and $\widetilde{S}$ such that $\widetilde{\rho}_\infty(\widetilde{T}, \widetilde{S}) \leq \rho_\infty(T, S)$, where

$$\widetilde{\rho}_\infty(\widetilde{T}, \widetilde{S}) := \sup\left\{\left|\widetilde{T}(x) - \widetilde{S}(x)\right| : x \in [0, 1]\right\} = \sup\left\{\left|\widetilde{T}(x) - \widetilde{S}(x)\right| : x \in \mathbb{R}\right\}.$$

The following lemma gives us more information about lifts and their fixed points.

Lemma 2.1.14. *Every continuous map $T : \mathbb{S}^1 \to \mathbb{S}^1$ with $\deg(T) \neq 1$ has a lift $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ with a fixed point in $[-1/2, 1/2]$. Moreover,*

$$\operatorname{dist}(0, \operatorname{Fix}(\widetilde{T})) \to 0 \quad \textit{whenever } T \to E_k \textit{ uniformly},$$

where $E_k([x]) := [kx]$ and where $k = \deg(T)$.

Proof. Let $k = \deg(T)$ and $\widehat{T} : \mathbb{R} \to \mathbb{R}$ be a lift of T. Define a map $D : \mathbb{R} \to \mathbb{R}$ by setting

$$D(x) := \widehat{T}(x) - x.$$

By definition of k, we have that $\widehat{T}(1/2) = \widehat{T}(-1/2 + 1) = \widehat{T}(-1/2) + k$, and hence

$$D(1/2) = \widehat{T}(1/2) - 1/2 = \widehat{T}(-1/2) + k - 1/2 = D(-1/2) + k - 1.$$

Since $k \neq 1$, the interval $D([-1/2, 1/2])$ has length at least 1, and thus contains an integer, say m. In other words, there exists $x_0 \in [-1/2, 1/2]$ such that $D(x_0) = m$, that is, such that $\widehat{T}(x_0) - m = x_0$. Letting $\widetilde{T} = \widehat{T} - m$, which is also a lift of T by Lemma 2.1.3, we obtain that $\widetilde{T}(x_0) = x_0$ and the first assertion in our lemma is proved.

For the second part, fix $\delta > 0$. One immediately verifies that $\widetilde{E}_k(x) = kx$ is a lift of E_k to $\mathbb{R}$ and

$$\widetilde{E}_k(-\delta) - (-\delta) = -k\delta + \delta = -(k-1)\delta \quad \text{whereas} \quad \widetilde{E}_k(\delta) - \delta = (k-1)\delta.$$

Since $k \neq 1$, the numbers $\widetilde{E}_k(-\delta) - (-\delta)$ and $\widetilde{E}_k(\delta) - \delta$ have opposite signs. Therefore, in view of Remark 2.1.13, if T is sufficiently close to E_k, then there exists a lift $\widetilde{T}$ of T such that $\widetilde{T}(-\delta) - (-\delta)$ and $\widetilde{T}(\delta) - \delta$ have the same signs as $\widetilde{E}_k(-\delta) - (-\delta)$ and $\widetilde{E}_k(\delta) - \delta$, respectively. Consequently, there exists $s \in (-\delta, \delta)$ so that $\widetilde{T}(s) - s = 0$. □

Remark 2.1.15. Lemma 2.1.14 does not generally hold for circle maps of degree 1. Indeed, it clearly does not hold for any irrational rotation.

2.2 Orientation-preserving homeomorphisms of the circle

Every homeomorphism of the unit circle is either *orientation preserving* or *orientation reversing*, which means that either the homeomorphism preserves the order of points on the circle or it reverses their order. In this section, we shall study orientation-preserving homeomorphisms. The differences that occur when considering orientation-reversing homeomorphisms are covered in the exercises at the end of the chapter.

By convention, arcs will be traversed in the counterclockwise direction along the unit circle. That is, the closed arc $[a, b]$ is the arc that consists of all points that are met when moving in the counterclockwise direction from point a to point b, including these two points. The open arc (a, b) simply excludes the endpoints from the closed arc $[a, b]$. Hence, $[\pi(0), \pi(1/2)]$ is the upper-half circle while $[\pi(1/2), \pi(0)]$ corresponds to the lower-half circle. The left half circle is represented by $[\pi(1/4), \pi(3/4)]$ whereas the right-half circle is the arc $[\pi(3/4), \pi(1/4)]$.

Definition 2.2.1. A homeomorphism $f : \mathbb{S}^1 \to \mathbb{S}^1$ is said to be *orientation-preserving* if $f(c) \in (f(a), f(b))$ whenever $c \in (a, b)$.

We will show that this is equivalent to the fact that any lift $\widetilde{f}$ of f is an increasing homeomorphism of $\mathbb{R}$. Recall that lifts exist by Lemma 2.1.2, and are unique up to addition by an integer according to Lemma 2.1.3.

Lemma 2.2.2. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be a homeomorphism of the unit circle. Let $\widetilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of f. Then $\widetilde{f}$ is a homeomorphism of $\mathbb{R}$.*

Proof. Let $\widetilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of f. By definition, $\widetilde{f}$ is surjective and continuous. It remains to show that it is injective. Suppose that this is not the case, that is, there exist $x, y \in \mathbb{R}$ such that $\widetilde{f}(x) = \widetilde{f}(y)$. We claim that there then exist $\widetilde{x}, \widetilde{y} \in \mathbb{R}$ such that $|\widetilde{x}-\widetilde{y}| < 1$ and $\widetilde{f}(\widetilde{x}) = \widetilde{f}(\widetilde{y})$. Indeed, if $|x - y| < 1$ then simply let $\widetilde{x} = x$ and $\widetilde{y} = y$. Otherwise, that is, if $|x - y| \geq 1$ then there exists a unique $k \in \mathbb{N}$ such that $k \leq |x - y| < k + 1$. Without loss of generality, we may assume that $x > y$. Then $y + k \leq x < y + k + 1$. Moreover,

$$\widetilde{f}(y + k) = \widetilde{f}(y) + k\deg(f) = \widetilde{f}(x) + k\deg(f)$$

while

$$\widetilde{f}(y + k + 1) = \widetilde{f}(y) + (k + 1)\deg(f) = \widetilde{f}(x) + k\deg(f) + \deg(f).$$

If $\deg(f) \neq 0$, applying the intermediate value theorem on the intervals $[y + k, x]$ and $[x, y + k + 1]$ gives that there exist $x_1 \in (y + k, x)$ and $x_2 \in (x, y + k + 1)$ such that

$$\widetilde{f}(x_1) = \widetilde{f}(x_2) = \widetilde{f}(x) + k\deg(f)/2.$$

In this case, let $\widetilde{x} = x_1$ and $\widetilde{y} = x_2$.

If $\deg(f) = 0$, then

$$\widetilde{f}(y + k) = \widetilde{f}(y + k + 1) = \widetilde{f}(x).$$

If there exists $y + k < z < y + k + 1$ such that $\widetilde{f}(z) \neq \widetilde{f}(x)$, then applying the intermediate value theorem on the intervals $[y + k, z]$ and $[z, y + k + 1]$ yields points $z_1 \in (y + k, z)$ and $z_2 \in (z, y + k + 1)$ such that

$$\widetilde{f}(z_1) = \widetilde{f}(z_2) = \big(\widetilde{f}(x) + \widetilde{f}(z)\big)/2.$$

In this case, let $\widetilde{x} = z_1$ and $\widetilde{y} = z_2$. Otherwise, $\widetilde{f}$ is equal to $\widetilde{f}(x)$ on the entire interval $[y + k, y + k + 1]$ and we let $\widetilde{x} = x$ and $\widetilde{y} \in (y + k, y + k + 1) \setminus \{x\}$.

In all cases, $|\widetilde{x} - \widetilde{y}| < 1$ and $\widetilde{f}(\widetilde{x}) = \widetilde{f}(\widetilde{y})$. It then follows that

$$f \circ \pi(\widetilde{x}) = \pi \circ \widetilde{f}(\widetilde{x}) = \pi \circ \widetilde{f}(\widetilde{y}) = f \circ \pi(\widetilde{y}).$$

Since f is injective, this means that $\pi(\widetilde{x}) = \pi(\widetilde{y})$. But $\pi(\widetilde{x}) \neq \pi(\widetilde{y})$ since $|\widetilde{x} - \widetilde{y}| < 1$. This contradiction shows that $\widetilde{f}$ is injective. In summary, $\widetilde{f}$ is a continuous bijection of $\mathbb{R}$. Then $\widetilde{f}$ is either strictly increasing or strictly decreasing. In particular, it is a homeomorphism and $\deg(f) \neq 0$. □

Corollary 2.2.3. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be a homeomorphism of the unit circle. Let $\tilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of f. If f is orientation preserving, then $\tilde{f}$ is an increasing homeomorphism of $\mathbb{R}$.*

Proof. By Lemma 2.2.2, we know that $\tilde{f}$ is a homeomorphism of $\mathbb{R}$. Therefore, $\tilde{f}$ is either strictly increasing or strictly decreasing. Suppose for a contradiction that $\tilde{f}$ is strictly decreasing. Changing lift if necessary, we may assume that $\tilde{f}(0) \in (0, 1]$. Since $\tilde{f}$ is continuous and strictly decreasing, there exists $0 < \delta < 1$ such that $0 < \tilde{f}(\delta) < \tilde{f}(0)$. Consider the arc $(a, b) := (\pi(0), \pi(\delta)) = \pi((0, \delta)) \subseteq \mathbb{S}^1$. Then

$$(f(a), f(b)) = (f(\pi(0)), f(\pi(\delta))) = (\pi(\tilde{f}(0)), \pi(\tilde{f}(\delta))).$$

Let $c \in (a, b)$. Then there exists $0 < \tilde{c} < \delta$ such that $\pi(\tilde{c}) = c$. Therefore, $0 < \tilde{f}(\delta) < \tilde{f}(\tilde{c}) < \tilde{f}(0) \leq 1$, and hence

$$f(c) = f(\pi(\tilde{c})) = \pi(\tilde{f}(\tilde{c})) \in \pi((\tilde{f}(\delta), \tilde{f}(0))) = (\pi(\tilde{f}(\delta)), \pi(\tilde{f}(0))).$$

Consequently, $f(c) \notin (f(a), f(b))$. This contradicts the hypothesis that f is orientation preserving. We thus conclude that $\tilde{f}$ must be strictly increasing. □

Let us now show that the degree of every orientation-preserving homeomorphism of the circle is equal to 1, where we recall that the degree of a continuous map $T : \mathbb{S}^1 \to \mathbb{S}^1$ is defined to be the integer $\tilde{T}(x+1) - \tilde{T}(x)$, which is independent of the choice of the point $x \in \mathbb{R}$ and of the choice of lift $\tilde{T} : \mathbb{R} \to \mathbb{R}$ according to Lemma 2.1.4.

Lemma 2.2.4. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be an orientation-preserving homeomorphism. Then* $\deg(f) = 1$.

Proof. Let $\tilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of f. Since $\tilde{f}$ is strictly increasing and $\tilde{f}(1) = \tilde{f}(0) + \deg(f)$, it is clear that $\deg(f) \geq 1$. Suppose that $\deg(f) \geq 2$. Then, as $\tilde{f}$ is continuous and $\tilde{f}(1) = \tilde{f}(0) + \deg(f)$, the intermediate value theorem guarantees the existence of some real number $0 < y < 1$ such that $\tilde{f}(y) = \tilde{f}(0) + 1$. But then

$$f \circ \pi(y) = \pi \circ \tilde{f}(y) = \pi(\tilde{f}(0) + 1) = \pi \circ \tilde{f}(0) = f \circ \pi(0).$$

Since f is injective (after all, it is a homeomorphism), this means that $\pi(y) = \pi(0)$. However, $\pi(y) \neq \pi(0)$ since $0 < y < 1$. This contradiction shows that the assumption $\deg(f) \geq 2$ cannot hold. Thus $\deg(f) = 1$. □

We leave it to the reader to prove that any homeomorphism F of $\mathbb{R}$ with the property that $F(x+1) = F(x) + 1$ for all $x \in \mathbb{R}$ generates an orientation-preserving homeomorphism f of S^1 (see Exercise 2.4.3).

Observe that the inverse $f^{-1} : \mathbb{S}^1 \to \mathbb{S}^1$ of an orientation-preserving homeomorphism $f : \mathbb{S}^1 \to \mathbb{S}^1$ is also an orientation-preserving homeomorphism. Therefore, $\deg(f^{-1}) = \deg(f) = 1$. Moreover, note that if $\tilde{f} : \mathbb{R} \to \mathbb{R}$ is a lift of $f : \mathbb{S}^1 \to \mathbb{S}^1$, then $\tilde{f}^{-1} : \mathbb{R} \to \mathbb{R}$ is a lift of $f^{-1} : \mathbb{S}^1 \to \mathbb{S}^1$. We deduce the following for orientation-preserving homeomorphisms.

Corollary 2.2.5. *Let $\tilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of an orientation-preserving homeomorphism $f : \mathbb{S}^1 \to \mathbb{S}^1$. For all $x \in \mathbb{R}$, all $k \in \mathbb{Z}$ and all $n \in \mathbb{Z}$, we have that*

$$\tilde{f}^n(x+k) = \tilde{f}^n(x) + k.$$

Proof. The result is trivial when $n = 0$. In light of Lemma 2.2.4, the result follows directly from Corollary 2.1.10 for all $n \in \mathbb{N}$. Using f^{-1} instead of f in Lemma 2.2.4 and Corollary 2.1.10, the result follows for $n \leq -1$. □

By induction one can also show the following (cf. Exercise 2.4.1).

Corollary 2.2.6. *Let $\tilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of an orientation-preserving homeomorphism $f : \mathbb{S}^1 \to \mathbb{S}^1$. Given any $k \in \mathbb{N}$, we have that*

$$|x-y| \leq k \Longrightarrow \left|\tilde{f}^n(x) - \tilde{f}^n(y)\right| \leq k, \quad \forall n \in \mathbb{Z}.$$

Moreover, if the left inequality is strict ($<$ instead of $\leq$), then so is the right one.

Corollary 2.2.7. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be an orientation-preserving homeomorphism and $\tilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of f. If $\tilde{g}$ is another lift of f so that $\tilde{g} = \tilde{f} + k$ for some $k \in \mathbb{Z}$, then*

$$\tilde{g}^n = \tilde{f}^n + nk$$

for all $n \in \mathbb{Z}$.

Proof. Apply Corollary 2.1.11 with f and f^{-1} in lieu of T. Recall that $\deg(f) = \deg(f^{-1}) = 1$ since both f and f^{-1} are orientation-preserving homeomorphisms. □

Lemma 2.2.4 implies immediately that $\tilde{f} - \mathrm{Id}_{\mathbb{R}}$ is a periodic function with period 1, where $\mathrm{Id}_{\mathbb{R}} : \mathbb{R} \to \mathbb{R}$ is the identity map (cf. Exercise 2.4.2).

Corollary 2.2.8. *Let $\tilde{f}$ be a lift of f. Then $\tilde{f} - \mathrm{Id}_{\mathbb{R}}$ is a periodic function with period 1. More generally, an increasing homeomorphism $\tilde{g} : \mathbb{R} \to \mathbb{R}$ is a lift of an orientation-preserving homeomorphism of the circle if and only if $\tilde{g} - \mathrm{Id}_{\mathbb{R}}$ is a periodic function with period 1.*

Let us now give the simplest example of an orientation-preserving homeomorphism of the unit circle: a rotation.

Example 2.2.9. Let $\alpha \in \mathbb{R}$. If $f([x]) := [x + \alpha]$ is the rotation of the unit circle by the angle α, then $\tilde{f}(x) = x + \alpha$ is a lift of f and for all $x \in \mathbb{R}$ we have that

$$\lim_{n\to\infty} \frac{\tilde{f}^n(x)}{n} = \lim_{n\to\infty} \frac{x + n\alpha}{n} = \alpha.$$

2.2.1 Rotation numbers

In this section, we introduce a number that allows us to think of the dynamics of a given homeomorphism of the unit circle mimicking, in some sense, the dynamics of a rotation of the circle. Accordingly, this number will be called the *rotation number* of the said homeomorphism.

The first result generalizes to orientation-preserving homeomorphisms the observation about the ratio $\tilde{f}^n(x)/n$ made for rotations of the circle in Example 2.2.9.

Proposition 2.2.10. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be an orientation-preserving homeomorphism of the unit circle and $\tilde{f} : \mathbb{R} \to \mathbb{R}$ a lift of f. Then the following statements hold:*
(a) *The number*

$$\rho(\tilde{f}) := \lim_{n\to\infty} \frac{\tilde{f}^n(x)}{n}$$

exists for all $x \in \mathbb{R}$ and is independent of x.
(b) *If $\tilde{g} = \tilde{f} + k$ for some $k \in \mathbb{Z}$, then $\rho(\tilde{g}) = \rho(\tilde{f}) + k$. That is, the choice of lift only changes ρ by an integer.*
(c) *For every $m \in \mathbb{N}$, we have $\rho(\tilde{f}^m) = m \cdot \rho(\tilde{f})$.*
(d) *The number $\rho(\tilde{f})$ is an integer if and only if f has a fixed point.*
(e) *The number $\rho(\tilde{f})$ is rational if and only if f has a periodic point.*
(f) *Let $x \in \mathbb{R}$. If $q \geq 1$ and r are integers such that $\tilde{f}^q(x) \leq x + r$, then $q\,\rho(\tilde{f}) \leq r$.*
(g) *Let $x \in \mathbb{R}$. If $q \geq 1$ and r are integers such that $\tilde{f}^q(x) \geq x + r$, then $q\,\rho(\tilde{f}) \geq r$.*

Proof. (a) We prove this proposition in two steps. We first assume the existence of $\rho(\tilde{f})$ and prove its independence of the point x chosen. We then prove the existence of $\rho(\tilde{f})$.

Step 1: If $\rho(\tilde{f})$ exists for some $x \in \mathbb{R}$, then it exists for all $y \in \mathbb{R}$ and is the same for all y.

Suppose that $\rho(\tilde{f})$ exists for some $x \in \mathbb{R}$. Choose any $y \in \mathbb{R}$. Then there exists $k \in \mathbb{N}$ such that $|y - x| \leq k$. By Corollary 2.2.6, we have that for every $n \in \mathbb{N}$,

$$\left|\tilde{f}^n(x) - \tilde{f}^n(y)\right| \leq k.$$

Therefore, for all $n \in \mathbb{N}$ we obtain that

$$\frac{\tilde{f}^n(x)}{n} - \frac{k}{n} \leq \frac{\tilde{f}^n(y)}{n} \leq \frac{\tilde{f}^n(x)}{n} + \frac{k}{n}.$$

Passing to the limit as n tends to infinity, we conclude that

$$\lim_{n\to\infty} \frac{\tilde{f}^n(x)}{n} \leq \liminf_{n\to\infty} \frac{\tilde{f}^n(y)}{n} \leq \limsup_{n\to\infty} \frac{\tilde{f}^n(y)}{n} \leq \lim_{n\to\infty} \frac{\tilde{f}^n(x)}{n}.$$

Consequently,

$$\lim_{n\to\infty} \frac{\tilde{f}^n(y)}{n} = \lim_{n\to\infty} \frac{\tilde{f}^n(x)}{n}.$$

Since $y \in \mathbb{R}$ is arbitrary, the proof of Step 1 is complete.

Step 2. The rotation number $\rho(\tilde{f})$ always exists.

According to Step 1, if the limit $\rho(\tilde{f})$ exists for any particular x, it exists for all x. So, without loss of generality, set $x = 0$. Fix momentarily $m \in \mathbb{N}$. Then there exists a unique $k \in \mathbb{Z}$ such that

$$k \le \tilde{f}^m(0) < k + 1.$$

Using Corollary 2.2.5 and the fact that $\tilde{f}^m$ is strictly increasing, we deduce by induction that for any $n \in \mathbb{N}$,

$$nk \le \tilde{f}^{nm}(0) < n(k+1).$$

It follows that

$$\frac{k}{m} \le \frac{\tilde{f}^m(0)}{m} < \frac{k+1}{m} \quad \text{and} \quad \frac{k}{m} \le \frac{\tilde{f}^{nm}(0)}{nm} < \frac{k+1}{m}.$$

Consequently,

$$\left| \frac{\tilde{f}^m(0)}{m} - \frac{\tilde{f}^{nm}(0)}{nm} \right| < \frac{k+1}{m} - \frac{k}{m} = \frac{1}{m}.$$

Interchanging the roles of m and n yields the inequality

$$\left| \frac{\tilde{f}^n(0)}{n} - \frac{\tilde{f}^{nm}(0)}{nm} \right| < \frac{1}{n}.$$

By the triangle inequality, we get that

$$\left| \frac{\tilde{f}^n(0)}{n} - \frac{\tilde{f}^m(0)}{m} \right| < \frac{1}{n} + \frac{1}{m}.$$

This shows that the sequence $(\tilde{f}^n(0)/n)_{n=1}^{\infty}$ is a Cauchy sequence in $\mathbb{R}$ and is therefore convergent. Hence, $\rho(\tilde{f})$ exists.

(b) Suppose that $\tilde{g} = \tilde{f} + k$. By Corollary 2.2.7, we know that $\tilde{g}^n = \tilde{f}^n + nk$ for all $n \in \mathbb{N}$. It follows that

$$\rho(\tilde{g}) = \lim_{n\to\infty} \frac{\tilde{g}^n(x)}{n} = \lim_{n\to\infty} \frac{\tilde{f}^n(x)}{n} + k = \rho(\tilde{f}) + k.$$

(c) Given that $\tilde{f}^m$ is a lift of f^m for every $m \in \mathbb{Z}$, the number $\rho(\tilde{f}^m)$ is well-defined for all m due to statement (a). Using any $x \in \mathbb{R}$, we obtain that

$$\rho(\tilde{f}^m) = \lim_{n\to\infty} \frac{\tilde{f}^{mn}(x)}{n} = \lim_{n\to\infty} \frac{\tilde{f}^{mn}(x)}{mn} \cdot m = \rho(\tilde{f}) \cdot m.$$

(d) Assume first that f has a fixed point. This means that there exists $z \in \mathbb{S}^1$ such that $f(z) = z$. Let $x \in \mathbb{R}$ be such that $\pi(x) = z$. Then

$$\pi(x) = z = f(z) = f(\pi(x)) = \pi(\tilde{f}(x)).$$

Therefore, $\tilde{f}(x) - x = k$ for some $k \in \mathbb{Z}$. Invoking Corollary 2.2.5 once again, we know that $\tilde{f}^n(x) = x + nk$ for each $n \in \mathbb{N}$. Therefore,

$$\rho(\tilde{f}) = \lim_{n\to\infty} \frac{\tilde{f}^n(x)}{n} = \lim_{n\to\infty} \frac{x + nk}{n} = k \in \mathbb{Z}.$$

This proves one implication. To derive the converse, assume that $\rho(\tilde{f}) \in \mathbb{Z}$. We aim to show that f has a fixed point. Replacing $\tilde{f}$ by $\tilde{f} - \rho(\tilde{f})$, which, by Lemma 2.1.3, is also a lift of f, we may assume that $\rho(\tilde{f}) = 0$. Assume by way of contradiction that $\tilde{f}$ has no fixed point. By the intermediate value theorem, this means that either $\tilde{f}(x) > x$ for all $x \in \mathbb{R}$ or $\tilde{f}(x) < x$ for all $x \in \mathbb{R}$. Suppose that $\tilde{f}(x) > x$ for all $x \in \mathbb{R}$ (a similar argument holds in the other case). This implies in particular that $\tilde{f}(0) > 0$, and thus the sequence $(\tilde{f}^n(0))_{n=1}^{\infty}$ is increasing. We further claim that $\tilde{f}^n(0) < 1$ for all $n \in \mathbb{N}$. Indeed, if this were not the case, then we would have $\tilde{f}^N(0) \geq 1$ for some $N \in \mathbb{N}$. We would deduce by induction that $\tilde{f}^{nN}(0) \geq n$ for all $n \in \mathbb{N}$. Hence, we would conclude that

$$\rho(\tilde{f}) = \lim_{n\to\infty} \frac{\tilde{f}^{nN}(0)}{nN} \geq \frac{1}{N} > 0.$$

But this would contradict the fact that $\rho(\tilde{f}) = 0$. Thus, $0 < \tilde{f}^n(0) < 1$ for all $n \in \mathbb{N}$. Summarizing, $(\tilde{f}^n(0))_{n=1}^{\infty}$ is a bounded monotonic sequence and as such is convergent. Let $L := \lim_{n\to\infty} \tilde{f}^n(0)$. Because of the continuity of $\tilde{f}$, Lemma 1.1.4 implies that L is a fixed point of $\tilde{f}$. This contradicts our assumption that $\tilde{f}$ has no fixed point. Thus, $\tilde{f}$ has a fixed point and f, as a factor of $\tilde{f}$, has a fixed point too.

(e) Let $\tilde{f}$ be a lift of f. Note that f has a periodic point if and only if there exists $m \in \mathbb{N}$ for which f^m has a fixed point. By statements (c) and (d), this is equivalent to stating that f has a periodic point if and only if $\rho(\tilde{f})$ is rational.

(f) Suppose that $q \geq 1$ and r are integers such that $\tilde{f}^q(x) \leq x + r$. Using Corollary 2.2.5 and the fact that $\tilde{f}^q$ is increasing, we deduce by induction that $\tilde{f}^{nq}(x) \leq x + nr$ for each $n \in \mathbb{N}$. Then

$$\rho(\tilde{f}) = \lim_{n\to\infty} \frac{\tilde{f}^{qn}(x)}{qn} \leq \lim_{n\to\infty} \frac{x + nr}{qn} = \frac{r}{q}.$$

(g) The proof proceeds analogously to (f) and is left as an exercise for the reader (see Exercise 2.4.4). □

Proposition 2.2.10, in conjunction with Example 2.2.9, suggests the following definition and terminology.

Definition 2.2.11. The *rotation number* $\rho(f)$ of an orientation-preserving homeomorphism $f : \mathbb{S}^1 \to \mathbb{S}^1$ of the unit circle is defined to be $\rho(\widetilde{f}) \pmod 1$, where $\widetilde{f} : \mathbb{R} \to \mathbb{R}$ is any lift of f.

As mentioned at the very beginning of the chapter, the rotation number should be thought of as the average rotation that the homeomorphism induces on the points of $\mathbb{S}^1$ over the long term. Statements (a) and (b) of Proposition 2.2.10 ensure that the rotation number exists and is well-defined. Statements (c), (d), and (e) translate into the assertions below. Note, though, that statements (f) and (g) have no counterparts for $\rho(f)$.

Proposition 2.2.12. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be an orientation-preserving homeomorphism of the unit circle. Then the following statements hold:*
(c′) *For every $m \in \mathbb{N}$, we have that $\rho(f^m) = m \cdot \rho(f) \pmod 1$.*
(d′) *The rotation number $\rho(f)$ is equal to zero if and only if f has a fixed point.*
(e′) *The rotation number $\rho(f)$ is rational if and only if f has a periodic point.*

Given that every orientation-preserving homeomorphism of the circle has an associated rotation number, it is natural to ask whether the rotation number is a topological conjugacy invariant. This is, in fact, very nearly the case, as we now show.

Theorem 2.2.13. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ and $g : \mathbb{S}^1 \to \mathbb{S}^1$ be topologically conjugate orientation-preserving homeomorphisms of the unit circle. If the conjugacy map preserves orientation, then $\rho(f) = \rho(g)$. If the conjugacy map reverses orientation, then $\rho(f) + \rho(g) = 0 \pmod 1$.*

Proof. Let $f, g : \mathbb{S}^1 \to \mathbb{S}^1$ be topologically conjugate orientation-preserving homeomorphisms. Let $h : \mathbb{S}^1 \to \mathbb{S}^1$ be a conjugacy map between them, so that $h \circ f = g \circ h$, and let $\widetilde{h}$ be a lift of h. Then $\widetilde{h}(x + n) = \widetilde{h}(x) + n$ for all $x \in \mathbb{R}$ and $n \in \mathbb{Z}$ if h is orientation preserving. If h is orientation reversing, then $\widetilde{h}(x+n) = \widetilde{h}(x) - n$ for all $x \in \mathbb{R}$ and $n \in \mathbb{Z}$. The proof of this last statement is deferred to Exercise 2.4.5(c). Denoting the integer and fractional parts of x by $\lfloor x \rfloor$ and $\langle x \rangle$, respectively, and observing that $\widetilde{h}(\langle x \rangle)$ lies between $\widetilde{h}(0)$ and $\widetilde{h}(1)$, it follows that for all $x \in \mathbb{R}$,

$$\begin{aligned}
\lim_{x\to\infty} \frac{\widetilde{h}(x)}{x} &= \lim_{x\to\infty} \frac{\widetilde{h}(\lfloor x \rfloor + \langle x \rangle)}{\lfloor x \rfloor + \langle x \rangle} \\
&= \lim_{x\to\infty} \frac{\widetilde{h}(\langle x \rangle) \pm \lfloor x \rfloor}{\lfloor x \rfloor} \cdot \frac{\lfloor x \rfloor}{\lfloor x \rfloor + \langle x \rangle} \\
&= \lim_{x\to\infty} \frac{\widetilde{h}(\langle x \rangle) \pm \lfloor x \rfloor}{\lfloor x \rfloor} \cdot \lim_{x\to\infty} \frac{\lfloor x \rfloor}{\lfloor x \rfloor + \langle x \rangle} \\
&= \pm 1,
\end{aligned}$$

depending on whether $\widetilde{h}$ is orientation preserving (+1) or orientation reversing (−1). The same relation holds for $\widetilde{h}^{-1}$ since it is a lift of h^{-1}. If $\rho(f) = \rho(g) = 0$, then we are done. So, suppose that at least one of $\rho(f)$ and $\rho(g)$ is positive. Without loss of generality, suppose that $\rho(g) > 0$. Let $\widetilde{g}$ be the lift of g such that $\rho(\widetilde{g}) = \rho(g)$. Then the map $\widetilde{f} := \widetilde{h}^{-1} \circ \widetilde{g} \circ \widetilde{h}$ is a lift of f. Indeed, $\widetilde{f}$ is an increasing homeomorphism of $\mathbb{R}$ since $\widetilde{g}$ is an increasing homeomorphism, while $\widetilde{h}$ is either an increasing homeomorphism or a decreasing homeomorphism, depending on the nature of h. Moreover, for any $x \in \mathbb{R}$ notice that

$$\begin{aligned}\pi \circ \widetilde{f}(x) &= \pi \circ \widetilde{h}^{-1} \circ \widetilde{g} \circ \widetilde{h}(x) = h^{-1} \circ \pi \circ \widetilde{g} \circ \widetilde{h}(x)\\ &= h^{-1} \circ g \circ \pi \circ \widetilde{h}(x) = h^{-1} \circ g \circ h \circ \pi(x)\\ &= f \circ \pi(x).\end{aligned}$$

Furthermore,

$$\begin{aligned}\frac{\rho(\widetilde{f})}{\rho(\widetilde{g})} &= \lim_{n\to\infty} \frac{\widetilde{f}^n(x)/n}{\widetilde{g}^n(\widetilde{h}(x))/n} = \lim_{n\to\infty} \frac{\widetilde{f}^n(x)}{\widetilde{g}^n(\widetilde{h}(x))}\\ &= \lim_{n\to\infty} \frac{\widetilde{h}^{-1}(\widetilde{g}^n \circ \widetilde{h}(x))}{\widetilde{g}^n \circ \widetilde{h}(x)} = \pm 1.\end{aligned}$$

Therefore, $\rho(\widetilde{f}) = \pm\rho(\widetilde{g}) = \pm\rho(g)$.

When h, and thus h^{-1}, $\widetilde{h}$ and $\widetilde{h}^{-1}$, is orientation preserving, we have that $\rho(\widetilde{f}) = \rho(g)$ and, hence, $\rho(f) = \rho(\widetilde{f}) \pmod 1 = \rho(g)$.

On the other hand, when h, and thus h^{-1}, $\widetilde{h}$ and $\widetilde{h}^{-1}$, is orientation reversing, we have that $\rho(f) = \rho(\widetilde{f}) \pmod 1 = -\rho(g) \pmod 1$. It hence follows that $\rho(f) + \rho(g) = 0 \pmod 1$. □

The following lemma provides a partial converse to Theorem 2.2.13. It states that the rotation number is essentially a complete invariant for rotations of the circle.

Lemma 2.2.14. *Two rotations of the circle are topologically conjugate if and only if their rotation numbers are equal or sum to zero, modulo 1.*

Proof. By Theorem 2.2.13, two topologically conjugate rotations of the unit circle have rotation numbers that are equal or whose sum is 0 (mod 1). For the converse implication, suppose that $f([x]) = [x + \alpha]$ and $g([x]) = [x + \beta]$ for some $0 < \alpha, \beta < 1$. If $\alpha = \beta$, then f is trivially topologically conjugate to g. If $\alpha + \beta = 0 \pmod 1$, then the map $h([x]) = [-x]$ is a suitable conjugacy map. Indeed,

$$\begin{aligned}h \circ f([x]) &= h([x + \alpha]) = [-x - \alpha]\\ &= [-x + \beta] = g([-x]) = g \circ h([x]).\end{aligned}$$ □

2.3 Minimality for homeomorphisms and diffeomorphisms of the circle

Our first goal in this section is to give a classification of minimal orientation-preserving homeomorphisms of the circle. We will need the following lemma.

Lemma 2.3.1. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be an orientation-preserving homeomorphism and let $\widetilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of f. Let A and B be the sets*

$$A := \{\widetilde{f}^n(0) + m : n, m \in \mathbb{Z}\} \subseteq \mathbb{R}$$
$$B := \{n\rho(\widetilde{f}) + m : n, m \in \mathbb{Z}\} \subseteq \mathbb{R}.$$

If $\rho(f)$ is irrational, then the map $H : A \to B$ defined by

$$H(\widetilde{f}^n(0) + m) = n\rho(\widetilde{f}) + m$$

is well-defined, bijective and increasing.

Proof. The map H is the composition $H_B \circ H_A$ of the two maps

$$H_A : A \to \mathbb{Z}^2, \quad \text{where} \quad H_A(\widetilde{f}^n(0) + m) = (n, m)$$

and

$$H_B : \mathbb{Z}^2 \to B, \quad \text{where} \quad H_B(n, m) = n\rho(\widetilde{f}) + m.$$

Thus, in order to show that H is a well-defined bijection, it suffices to show that H_A and H_B are well-defined bijections. It is clear that the map H_B is well-defined and surjective. To show that it is injective, suppose that $H_B(n_1, m_1) = H_B(n_2, m_2)$, that is, $n_1\rho(\widetilde{f}) + m_1 = n_2\rho(\widetilde{f}) + m_2$. If it were the case that $n_1 \neq n_2$, then we would have

$$\rho(\widetilde{f}) = (m_2 - m_1)/(n_1 - n_2) \in \mathbb{Q},$$

which would contradict the hypothesis that $\rho(f)$ is an irrational number. Thus, $n_1 = n_2$, which implies immediately that $m_1 = m_2$. Hence, H_B is injective. Let us now consider H_A. To prove that H_A is well-defined, assume that $\widetilde{f}^{n_1}(0) + m_1 = \widetilde{f}^{n_2}(0) + m_2$. If $n_1 \neq n_2$, then

$$f^{n_1}(\pi(0)) = \pi(\widetilde{f}^{n_1}(0)) = \pi(\widetilde{f}^{n_2}(0) + m_2 - m_1) = \pi(\widetilde{f}^{n_2}(0)) = f^{n_2}(\pi(0)).$$

Applying f^{-n_2} to both sides yields $f^{n_1-n_2}(\pi(0)) = \pi(0)$, that is, $\pi(0)$ is a periodic point of f. But, according to Proposition 2.2.12, the rotation number of f would then be a rational number. Once again, this would contradict the hypothesis that $\rho(f)$ is irrational. So $n_1 = n_2$, which implies immediately that $m_1 = m_2$. Thus, H_A is well-defined. It is easy to see that H_A is bijective.

To show that the map H is increasing, suppose that $\tilde{f}^n(0) + m < \tilde{f}^k(0) + l$. If $k = n$ then $m < l$ and obviously $n\rho(\tilde{f}) + m < k\rho(\tilde{f}) + l$. If $k < n$ then applying $\tilde{f}^{-k}$ to each side of $\tilde{f}^n(0) + m < \tilde{f}^k(0) + l$ and using Corollary 2.2.5 along with the fact that $\tilde{f}^{-k}$ is increasing, we deduce that $\tilde{f}^{n-k}(0) < l - m$. From part (f) of Proposition 2.2.10, it follows that $(n-k)\rho(\tilde{f}) \leq l-m$. Since $\rho(\tilde{f})$ is irrational, this last inequality must be strict: $(n-k)\rho(\tilde{f}) < l-m$. In other words, $n\rho(\tilde{f})+m < k\rho(\tilde{f})+l$. Similarly, if $k > n$ then applying $\tilde{f}^{-n}$ to each side of $\tilde{f}^n(0)+m < \tilde{f}^k(0)+l$ yields $m-l < \tilde{f}^{k-n}(0)$. From Proposition 2.2.10(g) and the fact that $\rho(\tilde{f})$ is irrational, we conclude that $m-l < (k-n)\rho(\tilde{f})$. In other words, $n\rho(\tilde{f})+m < k\rho(\tilde{f})+l$. In each case, we have shown that $n\rho(\tilde{f})+m < k\rho(\tilde{f})+l$, and hence H is a well-defined, increasing bijection. □

Our next result is the main one of this section. It states that every minimal orientation-preserving homeomorphism of the circle is topologically conjugate to a minimal rotation.

Theorem 2.3.2. *If $f : \mathbb{S}^1 \to \mathbb{S}^1$ is a minimal orientation-preserving homeomorphism of the circle, then f is topologically conjugate to the rotation $R_{\rho(f)} : \mathbb{S}^1 \to \mathbb{S}^1$ of the unit circle by the angle corresponding to the rotation number of f.*

Proof. If f is minimal, then by Remark 1.5.5 it admits no periodic points. In view of Proposition 2.2.12, this implies that $\rho(f)$ is an irrational number. By Lemma 2.3.1, the map H introduced in that lemma is then a well-defined increasing bijection from A to B.

We aim to extend H to a homeomorphism of $\mathbb{R}$ using Lemma 2.3.1 and the fact that an increasing bijection between dense subsets of $\mathbb{R}$ can be uniquely extended to an increasing homeomorphism of $\mathbb{R}$ (see Exercise 1.7.5). Toward that end, we shall now prove that A is dense in $\mathbb{R}$. To begin, choose an arbitrary $x \in \mathbb{R}$. Since f is minimal, we know that $\pi(x)$ belongs to $\omega(\pi(0)) = \mathbb{S}^1$ by Theorem 1.5.4. Therefore, there exists a strictly increasing sequence $(n_k)_{k=1}^{\infty}$ of nonnegative integers such that

$$\pi(x) = \lim_{k\to\infty} f^{n_k}(\pi(0)) = \lim_{k\to\infty} \pi(\tilde{f}^{n_k}(0)).$$

Since $\pi(y) = \pi(z)$ means that $y - z \in \mathbb{Z}$, we have that

$$\lim_{k\to\infty} \operatorname{dist}(\tilde{f}^{n_k}(0) - x, \mathbb{Z}) = 0.$$

Thus, for each $\varepsilon > 0$ there exists $k \in \mathbb{N}$ and $l \in \mathbb{Z}$ such that

$$|\tilde{f}^{n_k}(0) - x - l| < \varepsilon.$$

Therefore, $x \in \overline{A}$. As x was chosen arbitrarily in $\mathbb{R}$, we conclude that $\overline{A} = \mathbb{R}$.

Furthermore, B is also dense because, in light of Theorem 1.5.12, the rotation $R_{\rho(f)} : \mathbb{S}^1 \to \mathbb{S}^1$ is minimal when $\rho(f)$ is an irrational number. Indeed, observe that the set B consists of the orbit of 0 under $R_{\rho(f)}$ translated everywhere by adding each integer.

Since the rotation $R_{\rho(f)}$ is minimal, the orbit of 0 under $R_{\rho(f)}$ is dense in $\mathbb{S}^1$ or, equivalently, in $[0,1]$. As it comprises all integer translations of this orbit, the set B is dense in $\mathbb{R}$.

Because A and B are dense in $\mathbb{R}$, we infer that H extends uniquely to an increasing homeomorphism $\overline{H}:\mathbb{R}\to\mathbb{R}$.

Now, let $x=\tilde{f}^n(0)+m\in A$. Note that

$$H(x+1)=H\big(\tilde{f}^n(0)+m+1\big)=n\rho(\tilde{f})+m+1=H\big(\tilde{f}^n(0)+m\big)+1=H(x)+1.$$

Thus, $H(x+1)=H(x)+1$ for all $x\in A$. By continuity, the extension $\overline{H}$ must satisfy $\overline{H}(x+1)=\overline{H}(x)+1$ for all $x\in\mathbb{R}$. Then $h(\pi(x)):=\pi\circ\overline{H}(x)$ is a well-defined orientation-preserving homeomorphism of the circle. Moreover, for every $x=\tilde{f}^n(0)+m\in A$ we have that

$$\begin{aligned}H\circ\tilde{f}(x)&=H\circ\tilde{f}\big(\tilde{f}^n(0)+m\big)=H\big(\tilde{f}^{n+1}(0)+m\big)\\&=(n+1)\rho(\tilde{f})+m=\big(n\rho(\tilde{f})+m\big)+\rho(\tilde{f})\\&=H\big(\tilde{f}^n(0)+m\big)+\rho(\tilde{f})=T_{\rho(\tilde{f})}\circ H\big(\tilde{f}^n(0)+m\big)\\&=T_{\rho(\tilde{f})}\circ H(x),\end{aligned}$$

where the map $T_{\rho(\tilde{f})}:\mathbb{R}\to\mathbb{R}$ is the translation by $\rho(\tilde{f})$ on $\mathbb{R}$. This shows that $H\circ\tilde{f}(x)=T_{\rho(\tilde{f})}\circ H(x)$ for all $x\in A$. By continuity, $\overline{H}\circ\tilde{f}(x)=T_{\rho(\tilde{f})}\circ\overline{H}(x)$ for all $x\in\mathbb{R}$. Observe also that the real translation $T_{\rho(\tilde{f})}$ is a lift of the circle rotation $R_{\rho(f)}$. It then follows that

$$\begin{aligned}h\circ f\big(\pi(x)\big)&=h\circ\pi\circ\tilde{f}(x)=\pi\circ\overline{H}\circ\tilde{f}(x)\\&=\pi\circ T_{\rho(\tilde{f})}\circ\overline{H}(x)=R_{\rho(f)}\circ\pi\circ\overline{H}(x)\\&=R_{\rho(f)}\circ h\big(\pi(x)\big).\end{aligned}$$

So h is a conjugacy map between f and $R_{\rho(f)}$. □

2.3.1 Denjoy's theorem

The next result is the first for which we need the map f to be a diffeomorphism, rather than merely a homeomorphism. Recall that a diffeomorphism f is a homeomorphism with the property that both f' and $(f^{-1})'$ exist. We will also need the following definition.

Definition 2.3.3. The *(total) variation* $\operatorname{var}(\varphi)$ of a function $\varphi:\mathbb{S}^1\to\mathbb{R}$ is defined to be

$$\operatorname{var}(\varphi):=\sup\left\{\sum_{i=0}^{n-1}\big|\varphi(x_i)-\varphi(x_{i+1})\big| : x_0,x_1,\dots,x_n=x_0 \text{ partition } \mathbb{S}^1, n\in\mathbb{N}\right\},$$

where the supremum is taken over all finite partitions of the circle. If $\operatorname{var}(\varphi)$ is finite, then φ is said to be of *bounded variation*.

The main result of this section is named after the French mathematician, Arnaud Denjoy (1884–1974). Denjoy made outstanding contributions to many areas of mathematics, in particular to the theory of functions of a real variable.

Theorem 2.3.4 (Denjoy's theorem). *Suppose that $f : \mathbb{S}^1 \to \mathbb{S}^1$ is an orientation-preserving C^1 diffeomorphism with derivative f' of bounded variation. If $\rho(f)$ is irrational, then $f : \mathbb{S}^1 \to \mathbb{S}^1$ is minimal.*

Before beginning the proof of Denjoy's theorem, we first establish three lemmas, which will be useful in the proof. In the remainder of this section, we adopt the usual convention that arcs of the unit circle shall be traversed in the counterclockwise direction. For instance, (x, y) is the open arc of the circle generated when moving from x to y along the circle in the counterclockwise direction. Note also that since f is orientation preserving, it holds that $f((x, y)) = (f(x), f(y))$.

Lemma 2.3.5. *Assume that $x_0 \in \mathbb{S}^1$ is such that for some $n \in \mathbb{N}$,*

$$(x_0, f^n(x_0)) \cap \{f^j(x_0) : |j| \le n\} = \emptyset. \tag{2.2}$$

Then, for all $0 \le k \le n$,

$$(f^{k-n}(x_0), f^k(x_0)) \cap \{f^j(x_0) : |j| \le n\} = \emptyset.$$

Proof. Assume that x_0 is as stated above and, by way of contradiction, suppose that there exist $0 \le k \le n$ and $|j| \le n$ such that

$$f^j(x_0) \in (f^{k-n}(x_0), f^k(x_0)). \tag{2.3}$$

Fix the largest k with this property. Of course, $j \ne k$. We shall examine two potential cases.

Case 1: $j \le 0$.

If it turned out that $j \le 0$, then (2.3) and the fact that f preserves orientation would result in

$$f^{j+1}(x_0) \in f((f^{k-n}(x_0), f^k(x_0))) = (f^{k+1-n}(x_0), f^{k+1}(x_0)).$$

Since $-n \le j + 1 \le n$, it follows that (2.3) would also be satisfied with $k + 1$ in place of k. Given that $k \le n$ was chosen to be the maximal number satisfying this property, the only way that this could be true is if $k + 1 > n$, that is, if $k = n$. Hence, (2.3) would reduce to $f^j(x_0) \in (x_0, f^n(x_0))$, which would contradict our original hypothesis (2.2). So this case never takes place.

Case 2: $j > 0$. This case is divided into two subcases, which are illustrated in Figure 2.1.

Note that for any given $x \in \mathbb{S}^1$, we have

$$f^r(x) \ne f^s(x), \quad \forall r, s \in \mathbb{Z}, \ r \ne s. \tag{2.4}$$

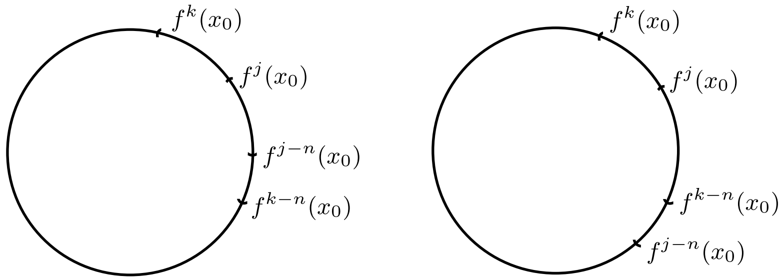

Figure 2.1: On the left, Subcase 2.1: $(f^{j-n}(x_0), f^j(x_0)) \subseteq (f^{k-n}(x_0), f^k(x_0))$. On the right, Subcase 2.2: $f^{k-n}(x_0) \in (f^{j-n}(x_0), f^j(x_0))$.

Otherwise, $f^{r-s}(x) = x$ and f would have a periodic point, that is, $\rho(f)$ would be rational according to Proposition 2.2.12. This would contradict our hypothesis that $\rho(f)$ is irrational.

Subcase 2.1: $f^{j-n}(x_0) \in (f^{k-n}(x_0), f^j(x_0))$.

This means that $(f^{j-n}(x_0), f^j(x_0)) \subseteq (f^{k-n}(x_0), f^j(x_0))$. In light of assumption (2.3), which we recall states that $f^j(x_0) \in (f^{k-n}(x_0), f^k(x_0))$, we actually have that

$$(f^{j-n}(x_0), f^j(x_0)) \subseteq (f^{k-n}(x_0), f^k(x_0)).$$

Then the continuity and orientation-preserving properties of f yield that

$$f^{j-k}([f^{k-n}(x_0), f^k(x_0)]) = [f^{j-n}(x_0), f^j(x_0)] \subseteq [f^{k-n}(x_0), f^k(x_0)].$$

So f^{j-k} maps a closed arc to a closed arc within itself. The intermediate value theorem then asserts that f^{j-k} has a fixed point (recall that $j \neq k$). Hence, f has a periodic point. According to Proposition 2.2.12, this means that $\rho(f)$ is a rational number. This contradicts our hypothesis that $\rho(f)$ is irrational, and thus this subcase cannot occur.

Subcase 2.2: $f^{j-n}(x_0) \notin (f^{k-n}(x_0), f^j(x_0))$.

In other words, $f^{j-n}(x_0) \in [f^j(x_0), f^{k-n}(x_0)]$. Since $f^r(x_0) \neq f^s(x_0)$ for all $r \neq s$, this actually means that $f^{j-n}(x_0) \in (f^j(x_0), f^{k-n}(x_0))$. Equivalently, this means that

$$f^{k-n}(x_0) \in (f^{j-n}(x_0), f^j(x_0)).$$

Then, as $-n \leq k-n \leq n$, we have a relation akin to (2.3) but with $k-n$ in place of j and j in place of k. But since k is maximal with this property, we deduce that $j \leq k$. In fact, as $f^k(x_0) \notin (f^{k-n}(x_0), f^k(x_0))$, we know that $j < k$. Then, by (2.3), we obtain that

$$f^{n+j-k}(x_0) = f^{n-k}(f^j(x_0)) \in f^{n-k}((f^{k-n}(x_0), f^k(x_0))) = (x_0, f^n(x_0)).$$

Since $-n \leq n+j-k \leq n$, this contradicts our original hypothesis (2.2). This shows that this subcase does not happen either.

To summarize, neither Case 1 nor Case 2 can occur. So, whatever $0 \le k \le n$ might be, there is no $-n \le j \le n$ satisfying (2.3). This contradiction completes the proof. □

A rather straightforward consequence of Lemma 2.3.5 is the following.

Lemma 2.3.6. *If $x_0 \in \mathbb{S}^1$ is such that*

$$(x_0, f^n(x_0)) \cap \{f^j(x_0) : |j| \le n\} = \emptyset$$

for some $n \in \mathbb{N}$, then the arcs $\{(f^{k-n}(x_0), f^k(x_0)) : 0 \le k \le n\}$ are mutually disjoint.

Proof. If two such arcs intersected, then an endpoint of one of those arcs would belong to the other arc, which would contradict the previous lemma. □

Finally, we derive an estimate on the derivatives of the iterates of f.

Lemma 2.3.7. *There is a universal constant $v > 0$ with the property that for all $x_0 \in \mathbb{S}^1$ and $n \in \mathbb{N}$ such that*

$$(x_0, f^n(x_0)) \cap \{f^j(x_0) : |j| \le n\} = \emptyset,$$

we have

$$(f^n)'(x_0)(f^{-n})'(x_0) \ge e^{-v}.$$

Proof. Let $x_0 \in \mathbb{S}^1$ and $n \in \mathbb{N}$ be as stated above. Since f preserves orientation, both $(f^n)'(x)$ and $(f^{-n})'(x)$ are strictly positive for all $x \in \mathbb{S}^1$ and all $n \ge 0$. Let

$$a := \inf\{f'(x) \mid x \in \mathbb{S}^1\}.$$

As f is a C^1 function, its derivative f' is continuous, and hence its infimum on the compact set $\mathbb{S}^1$ is achieved. So $a > 0$.

By a simple application of the chain rule, we obtain that

$$\log(f^n)'(x_0) = \log\left(\prod_{k=0}^{n-1} f'(f^k(x_0))\right) = \sum_{k=0}^{n-1} \log f'(f^k(x_0))$$

and

$$\begin{aligned}\log(f^{-n})'(x_0) &= \log\big((f^n)'(f^{-n}(x_0))\big)^{-1} \\ &= -\log(f^n)'(f^{-n}(x_0)) \\ &= -\log\left(\prod_{k=0}^{n-1} f'(f^k(f^{-n}(x_0)))\right) \\ &= -\sum_{k=0}^{n-1} \log f'(f^{k-n}(x_0)).\end{aligned}$$

These two equalities yield the estimate

$$\begin{aligned}\log\Big((f^n)'(x_0)(f^{-n})'(x_0)\Big) &= \log(f^n)'(x_0) + \log(f^{-n})'(x_0)\\ &= \sum_{k=0}^{n-1}\Big(\log f'(f^k(x_0)) - \log f'(f^{k-n}(x_0))\Big)\\ &\geq -\sum_{k=0}^{n-1}\Big|\log f'(f^k(x_0)) - \log f'(f^{k-n}(x_0))\Big|.\end{aligned}$$

But, by the mean value theorem,

$$\Big|\log f'(f^k(x_0)) - \log f'(f^{k-n}(x_0))\Big| = \frac{1}{c_k}\Big|f'(f^k(x_0)) - f'(f^{k-n}(x_0))\Big|,$$

where c_k is between $f'(f^k(x_0))$ and $f'(f^{k-n}(x_0))$. In particular, we have that $c_k \geq a$. Moreover, according to Lemma 2.3.6, the family of arcs $\{(f^{k-n}(x_0), f^k(x_0)) : 0 \leq k \leq n\}$ are mutually disjoint. Consequently, the points $\{f^j(x_0) : |j| \leq n\}$ can be arranged in such a way that they form an ordered partition of $\mathbb{S}^1$ in which $f^k(x_0)$ immediately follows $f^{k-n}(x_0)$ for each $0 \leq k \leq n$. Hence,

$$\sum_{k=0}^{n-1}\Big|f'(f^k(x_0)) - f'(f^{k-n}(x_0))\Big| \leq \operatorname{var}(f').$$

Therefore,

$$\begin{aligned}\sum_{k=0}^{n-1}\Big|\log f'(f^k(x_0)) - \log f'(f^{k-n}(x_0))\Big| &= \sum_{k=0}^{n-1}\frac{1}{c_k}\Big|f'(f^k(x_0)) - f'(f^{k-n}(x_0))\Big|\\ &\leq \frac{1}{a}\sum_{k=0}^{n-1}\Big|f'(f^k(x_0)) - f'(f^{k-n}(x_0))\Big|\\ &\leq \frac{1}{a}\operatorname{var}(f') =: v < \infty.\end{aligned}$$

Then we obtain that

$$\log\Big((f^n)'(x_0)(f^{-n})'(x_0)\Big) \geq -\sum_{k=0}^{n-1}\Big|\log f'(f^k(x_0)) - \log f'(f^{k-n}(x_0))\Big| \geq -v.$$

Hence, $(f^n)'(x_0)(f^{-n})'(x_0) \geq e^{-v}$. Note that v depends only on f. □

We are now in a position to prove Denjoy's theorem.

Proof of Denjoy's theorem. Suppose, by way of contradiction, that the map $f : \mathbb{S}^1 \to \mathbb{S}^1$ is not minimal. According to Theorem 1.5.2, every dynamical system has a minimal set. Call this set $\triangle \subseteq \mathbb{S}^1$. Since we are assuming that f is not minimal, it follows that $\triangle$ is a proper subset of $\mathbb{S}^1$. Recall that $\triangle$ is a closed f-invariant set. Since f is a homeomorphism, not only is $f(\triangle) \subseteq \triangle$ but, in fact, $f(\triangle) = \triangle = f^{-1}(\triangle)$. That is, $\triangle$ is a completely invariant set. Therefore,

$$f(\mathbb{S}^1 \setminus \triangle) = \mathbb{S}^1 \setminus \triangle = f^{-1}(\mathbb{S}^1 \setminus \triangle).$$

Because $\mathbb{S}^1 \setminus \triangle$ is an open subset of the circle, we can write $\mathbb{S}^1 \setminus \triangle = \bigcup_{j=0}^{\infty} I_j$, where the I_j's form a countable union of maximal disjoint open arcs. This implies in particular that $\partial I_j \subseteq \triangle$ for each $j \geq 0$. We then have

$$\sum_{j=0}^{\infty} |I_j| = \mathrm{Leb}(\mathbb{S}^1 \setminus \triangle) \leq 1,$$

where Leb denotes the Lebesgue measure. Hence, $\lim_{j\to\infty} |I_j| = 0$.

We now establish two claims about the I_j's.

Claim 1. *For every $n \in \mathbb{Z}$, there is a unique $j_n \geq 0$ such that $f^n(I_0) \cap I_{j_n} \neq \emptyset$. In fact, $f^n(I_0) = I_{j_n}$.*

By definition, the I_j's are the connected components of $\mathbb{S}^1 \setminus \triangle$. Let n be an integer. Since $\mathbb{S}^1 \setminus \triangle$ is completely f-invariant and f is a homeomorphism, the set $f^n(I_0)$ is a connected component of $\mathbb{S}^1 \setminus \triangle$. Denote by I_{j_n} this unique component. This proves Claim 1.

Claim 2. *If $m \neq k$, then $I_{j_m} \cap I_{j_k} = \emptyset$. Moreover, $\lim_{|n|\to\infty} |I_{j_n}| = 0$.*

Suppose for a contradiction that $I_{j_m} \cap I_{j_k} \neq \emptyset$ for some $k < m$. Since these arcs are connected components of $\mathbb{S}^1 \setminus \triangle$, we have that $I_{j_m} = I_{j_k}$. Then

$$f^{m-k}(I_{j_k}) = f^{m-k}(f^k(I_0)) = f^m(I_0) = I_{j_m} = I_{j_k}.$$

Therefore, $f^{m-k}(\overline{I_{j_k}}) = \overline{I_{j_k}}$. In other words, f^{m-k} maps a closed arc within itself. Thus, f^{m-k} has a fixed point, and hence f has a periodic point. According to Proposition 2.2.12, this means that $\rho(f)$ is a rational number, which contradicts our original assumption that $\rho(f)$ is irrational. Consequently, the family $(I_{j_n})_{n=0}^{\infty}$ consists of mutually disjoint arcs, that is, all the j_n's differ and so there are infinitely many arcs I_{j_n}. It therefore follows that

$$\lim_{|n|\to\infty} |I_{j_n}| = 0.$$

This completes the proof of Claim 2.

Now, for each $n \geq 0$ consider the sets

$$\mathcal{J}_+^n := \{x \in I_0 : (f^n)'(x) \geq e^{-v/2}\} \quad \text{and} \quad \mathcal{J}_-^n := \{x \in I_0 : (f^{-n})'(x) \geq e^{-v/2}\},$$

where $v > 0$ is the universal constant arising from Lemma 2.3.7. We aim to show that there exists a strictly increasing sequence $(n_q)_{q=0}^{\infty}$ of nonnegative integers such that

$$I_0 = \mathcal{J}_+^{n_q} \cup \mathcal{J}_-^{n_q}, \quad \forall q \geq 0.$$

According to Lemma 2.3.7, it suffices to show that

$$\big(x_0, f^{n_q}(x_0)\big) \cap \{f^j(x_0) : |j| \leq n_q\} = \emptyset, \quad \forall x_0 \in I_0, \forall q \geq 0.$$

To prove this, write $I_0 := (a, b)$. We first construct a strictly increasing sequence of nonnegative integers $(p_q)_{q=0}^{\infty}$ and a sequence of integers $(m_{p_q})_{q=0}^{\infty}$, where $|m_{p_q}| \leq p_q$ and $\lim_{q\to\infty} |m_{p_q}| = \infty$, so that

$$\big(b, f^{m_{p_q}}(b)\big) \cap \{f^j(b) : |j| \leq p_q\} = \emptyset, \quad \forall q \geq 0. \tag{2.5}$$

Since $b \in \triangle$ and $\triangle$ is minimal, we know that $\omega(b) = \triangle$ by Theorem 1.5.4. Therefore, there is a strictly increasing sequence $(p_q)_{q=0}^{\infty}$ of nonnegative integers such that $\lim_{q\to\infty} f^{p_q}(b) = b$.

Note that

$$f^j(b) \neq f^k(b), \quad \forall j \neq k. \tag{2.6}$$

Otherwise, $f^{j-k}(b) = b$ for some $j > k$ and f would have a periodic point, that is, $\rho(f)$ would be rational according to Proposition 2.2.12. This would contradict the fact that $\rho(f)$ is irrational. In particular, this means that $f^{p_q}(b) \neq b$ for every q. By passing to a subsequence if necessary, we may thus assume that the points of the sequence $(f^{p_q}(b))_{q=0}^{\infty}$ successively edge closer to b.

For every q, denote by $f^{m_{p_q}}(b)$ the point among the iterates

$$f^{-p_q}(b), f^{-p_q+1}(b), \ldots, f^{-1}(b), f(b), \ldots, f^{p_q-1}(b), f^{p_q}(b)$$

(excluding b) which is closest to the point b. Clearly, $f^{m_{p_q}}(b) \to b$ since $f^{p_q}(b) \to b$ and the points of the sequence $(f^{m_{p_q}}(b))_{q=0}^{\infty}$ successively edge closer to b since the $f^{p_q}(b)$'s do. Thus, the sequence $(m_{p_q})_{q=0}^{\infty}$ accumulates to ∞ or $-\infty$, and hence admits a subsequence converging to ∞ or $-\infty$. By replacing f with f^{-1} if necessary, we may assume without loss of generality that the subsequence in question is positive and converges to ∞. Let us denote that subsequence by the same notation $(m_{p_q})_{q=0}^{\infty}$. Then, by construction of the $f^{m_{p_q}}(b)$'s, relation (2.5) holds. In other words,

$$\{f^j(b) : |j| \leq p_q\} \subseteq [f^{m_{p_q}}(b), b], \quad \forall q \geq 0. \tag{2.7}$$

In fact, by (2.6), we have

$$\{f^j(b) : |j| \leq p_q\} \subseteq \big(f^{m_{p_q}}(b), b\big), \quad \forall q \geq 0. \tag{2.8}$$

Moreover,

$$f^j(b) \in (b,a], \quad \forall j \neq 0. \tag{2.9}$$

Otherwise, $f^j(b) \in (a,b)$ would imply that $f^j((a,b)) \cap (a,b) \neq \emptyset$, which would force j to be equal to zero by Claims 1 and 2.

In particular, this implies that $f^{m_{p_q}}(b) \in (b,a]$. Hence,

$$\left(f^{m_{p_q}}(b), b\right) = (f^{m_{p_q}}(b), a] \cup (a,b), \quad \forall q \geq 0. \tag{2.10}$$

We deduce from (2.8), (2.9) and (2.10) that

$$\{f^j(b) : |j| \leq p_q\} \subseteq \left(f^{m_{p_q}}(b), b\right) \cap (b,a] = (f^{m_{p_q}}(b), a], \quad \forall q \geq 0. \tag{2.11}$$

Finally, by dropping the first few terms of the subsequence $(m_{p_q})_{q=0}^{\infty}$ if necessary, we can also guarantee that all $f^{m_{p_q}}(b)$'s are closer to b than the point a is.

Now, let $x_0 \in (a,b)$. First, note that $f^j(x_0) \notin (a,b)$ for all $j \neq 0$; otherwise, we would have that $f^j((a,b)) \cap (a,b) \neq \emptyset$, and Claims 1 and 2 would force j to be equal to zero. In particular, this implies that

$$(x_0, b) \cap \{f^j(x_0) : |j| \leq m_{p_q}\} \subseteq (x_0, b) \cap \{f^j(x_0) : j \in \mathbb{Z}\} = \emptyset. \tag{2.12}$$

Observe also that $f^j(x_0) \neq b$ for all j since $x_0 \in \mathbb{S}^1 \setminus \triangle$, the set $\mathbb{S}^1 \setminus \triangle$ is completely f-invariant and $b \in \triangle$. In particular, we obtain

$$\{b\} \cap \{f^j(x_0) : |j| \leq m_{p_q}\} \subseteq \{b\} \cap \{f^j(x_0) : j \in \mathbb{Z}\} = \emptyset. \tag{2.13}$$

Finally, we show that $f^{m_{p_q}}(x_0)$ is the point which is closest to b among the points $f^{-p_q}(x_0), f^{-p_q+1}(x_0), \ldots, f^{-1}(x_0), x_0, f(x_0), \ldots, f^{p_q-1}(x_0), f^{p_q}(x_0)$, except possibly x_0 itself. Let $q \geq 0$. Suppose that there exists j with $-p_q \leq j \leq p_q$, $j \neq m_{p_q}$, such that $f^j(x_0)$ is closer to b than $f^{m_{p_q}}(x_0)$. We already know that $f^j(x_0) \neq b$ and $f^j(x_0) \neq a$ by the complete f-invariance of $\mathbb{S}^1 \setminus \triangle$. On one hand, if $f^j(x_0) \in (a,b)$ then $f^j((a,b)) \cap (a,b) \neq \emptyset$. Claims 1 and 2 impose that j equal zero. Then $f^j(x_0) = x_0$. On the other hand, it might turn out that $f^j(x_0) \in (b,a)$. In this case, observe that

$$\begin{aligned} f^{m_{p_q}}(x_0) \in f^{m_{p_q}}\left((a,b)\right) = f^{m_{p_q}}(I_0) = I_{j_{m_{p_q}}} &= \left(f^{m_{p_q}}(a), f^{m_{p_q}}(b)\right) \\ &= (b,a) \cap \left(f^{m_{p_q}}(a), f^{m_{p_q}}(b)\right) \\ &\subseteq \left(b, f^{m_{p_q}}(b)\right). \end{aligned}$$

Since $f^{m_{p_q}}(b)$ is closer to b than the point a is and since $f^j(x_0)$ is assumed to be closer to b than $f^{m_{p_q}}(x_0)$, we would then have $f^j(x_0) \in (b, f^{m_{p_q}}(x_0))$. As $f^j(b) \in (f^{m_{p_q}}(b), a]$ according to (2.11), we would then deduce that $[f^{m_{p_q}}(x_0), f^{m_{p_q}}(b)] \subseteq [f^j(x_0), f^j(b)]$. As f is orientation preserving, this would imply that

$$f^{m_{p_q}-j}\left([x_0, b]\right) = f^{-j}\left([f^{m_{p_q}}(x_0), f^{m_{p_q}}(b)]\right) \subseteq f^{-j}\left([f^j(x_0), f^j(b)]\right) = [x_0, b].$$

Then f would have a periodic point, which is not the case. To summarize, $f^{m_{p_q}}(x_0)$ is indeed the point among $f^{-p_q}(x_0),\dots,f^{p_q}(x_0)$, which is closest to the point b, except possibly x_0 itself. As noted previously, $f^{m_{p_q}}(x_0) \in (b,a)$ for every q. This implies that

$$\big(b, f^{m_{p_q}}(x_0)\big) \cap \{f^j(x_0) : |j| \le m_{p_q}\} = \emptyset. \tag{2.14}$$

From (2.12), (2.13), and (2.14), we conclude that

$$\big(x_0, f^{m_{p_q}}(x_0)\big) \cap \{f^j(x_0) : |j| \le m_{p_q}\} = \emptyset.$$

Note that the sequence $(m_{p_q})_{q=0}^{\infty}$ is independent of $x_0 \in (a,b) = I_0$. Setting $n_q := m_{p_q}$, an application of Lemma 2.3.7 with $n = n_q$ allows us to deduce that $(f^{n_q})'(x_0)\,(f^{-n_q})'(x_0) \ge e^{-v}$ for all $q \ge 0$ and all $x_0 \in I_0$. This implies that $I_0 = \mathcal{J}_+^{n_q} \cup \mathcal{J}_-^{n_q}$, and hence that $\max\{\lambda(\mathcal{J}_+^{n_q}), \lambda(\mathcal{J}_-^{n_q})\} \ge |I_0|/2$ for all $q \ge 0$, where $\lambda = \mathrm{Leb}$ denotes the Lebesgue measure on $\mathbb{S}^1$. If $\lambda(\mathcal{J}_+^{n_q}) \ge |I_0|/2$, then

$$\begin{aligned} |I_{j_{n_q}}| = \big|f^{n_q}(I_0)\big| &= \int_{I_0} (f^{n_q})'(x)\,dx \\ &\ge \int_{\mathcal{J}_+^{n_q}} (f^{n_q})'(x)\,d\lambda(x) \ge \int_{\mathcal{J}_+^{n_q}} e^{-v/2}\,d\lambda(x) \\ &= e^{-v/2}\lambda(\mathcal{J}_+^{n_q}) \ge \frac{e^{-v/2}}{2}|I_0|. \end{aligned}$$

A similar argument yields the same conclusion if $\lambda(\mathcal{J}_-^{n_q}) \ge |I_0|/2$. Thus, $|I_{j_{n_q}}| \ge e^{-v/2}|I_0|/2$ for each q. This contradicts the fact that $\lim_{n\to\infty} |I_{j_n}| = 0$. Therefore, the minimal set $\triangle$ must be $\mathbb{S}^1$, which means that $f : \mathbb{S}^1 \to \mathbb{S}^1$ is minimal. □

As an immediate consequence of Denjoy's theorem, we obtain the following corollary.

Corollary 2.3.8. *Suppose that $f : \mathbb{S}^1 \to \mathbb{S}^1$ is an orientation-preserving C^1 diffeomorphism with derivative f' of bounded variation. If $\rho(f)$ is irrational, then f is topologically conjugate to the rotation around the circle by the angle $\rho(f)$.*

Proof. This follows directly from Theorems 2.3.2 and 2.3.4. □

Remark 2.3.9. Notice that if $\varphi : \mathbb{S}^1 \to \mathbb{R}$ is Lipschitz continuous, then φ is of bounded variation. Indeed, for any finite partition $x_0, x_1, \dots, x_n = x_0$ of the unit circle we have

$$\sum_{i=0}^{n-1} \big|\varphi(x_{i+1}) - \varphi(x_i)\big| \le \sum_{i=0}^{n-1} L|x_{i+1} - x_i| = L,$$

where L is any Lipschitz constant for φ.

As every C^1 function on $\mathbb{S}^1$ is Lipschitz continuous, every C^2 function f has a derivative f' which is C^1, and hence of bounded variation. Thus the previous corollary yields the following further result.

Corollary 2.3.10. *Suppose that $f : \mathbb{S}^1 \to \mathbb{S}^1$ is an orientation-preserving C^2 diffeomorphism. If $\rho(f)$ is irrational, then f is topologically conjugate to the rotation around the circle by the angle $\rho(f)$.*

2.3.2 Denjoy's counterexample

In light of Corollary 2.3.8, it is natural to ask whether all orientation-preserving homeomorphisms with irrational rotation numbers are topologically conjugate to a rotation around the circle by an irrational angle. This is not the case. In fact, there are even orientation-preserving C^1 diffeomorphisms with irrational rotation numbers which are not topologically conjugate to an irrational rotation.

We will now construct an orientation-preserving homeomorphism with irrational rotation number which is not topologically conjugate to an irrational rotation of the circle. This construction is also due to Denjoy. The idea is the following. We know that minimality is a topological invariant. Given that an irrational rotation of the circle is minimal, it suffices to construct an orientation-preserving homeomorphism with irrational rotation number which is not minimal. By Theorem 1.5.4, this reduces to devising an orientation-preserving homeomorphism with irrational rotation number which admits a nondense orbit. We will build such a map by performing a "surgery" on an irrational rotation of the circle. Let $R_\rho : \mathbb{S}^1 \to \mathbb{S}^1$ be an irrational rotation of the circle. Choose arbitrarily $\theta \in \mathbb{S}^1$. Cut the unit circle at the point θ, open it up and insert into the gap an arc I_0. Similarly, cut the circle at the point $R_\rho(\theta)$, open it up and insert into the gap an arc I_1. Perform a similar procedure at every point of the orbit of θ. That is, for every $n \in \mathbb{Z}$ cut the circle at the point $R_\rho^n(\theta)$, open it up and insert into the gap an arc I_n. Make sure to choose the arcs $\{I_n\}$ small enough that $\sum_{n=-\infty}^{\infty} |I_n| < \infty$. These surgeries result in a larger "circle" (more precisely, a simple closed curve). It only remains to extend the original irrational rotation to the larger curve by defining the extension on the union of the I_n's. Since the extension is required to preserve orientation, choose any orientation-preserving homeomorphism h_n mapping I_n to I_{n+1}. The extension is then an orientation-preserving homeomorphism of the larger circle. Note that the extension does not admit any periodic point, since the original map did not and the points in the inserted arcs visit all those arcs in succession without ever coming back to the same arc. Proposition 2.2.12 therefore allows us to infer that the extension must have an irrational rotation number. Moreover, the fact that the interior points of the inserted arcs never come back to their original arc under iteration shows that all these points have nondense orbits.

In fact, the above argument can be modified in such a way that the extension is a C^1 diffeomorphism whose derivative is not of bounded variation (see pp. 111–112 of [19]).

2.4 Exercises

Exercise 2.4.1. In this exercise, you shall prove Corollary 2.2.6. We suggest that you proceed as follows. Let $\tilde{f}$ be a lift of f.

(a) Using Corollary 2.2.5 and the fact that $\tilde{f}$ is increasing, show that if $|x-y| \leq k$ for some $k \in \mathbb{Z}_+$, then $|\tilde{f}(x) - \tilde{f}(y)| \leq k$.

(b) Deduce that if $|x-y| \leq k$ for some $k \in \mathbb{Z}_+$, then $|\tilde{f}^n(x) - \tilde{f}^n(y)| \leq k$ for any $n \in \mathbb{Z}$.

(c) Prove that we can replace $\leq$ by $<$ above.

Exercise 2.4.2. Prove Corollary 2.2.8.

Exercise 2.4.3. Prove that any homeomorphism $F : \mathbb{R} \to \mathbb{R}$ with the property that $F(x+1) = F(x) + 1$ for all $x \in \mathbb{R}$ generates an orientation-preserving homeomorphism $f : S^1 \to \mathbb{S}^1$.

Exercise 2.4.4. Taking inspiration from the proof of statement (f) in Proposition 2.2.10, prove statement (g) of the same proposition.

Exercise 2.4.5. This exercise is concerned with orientation-reversing homeomorphisms of the circle. You will be asked to prove several properties of lifts of such maps. In the end, you will discover that the concept of “rotation” number is useless for orientation-reversing homeomorphisms.

Let $\tilde{f}$ be a lift of an orientation-reversing homeomorphism f. Prove the following statements:

(a) Show that any lift $\tilde{f}$ is a decreasing homeomorphism of $\mathbb{R}$ (cf. Lemma 2.2.3).

(b) Prove that $\deg(f) = -1$ (cf. Lemma 2.2.4).

(c) Show that $\tilde{f}^n(x+k) = \tilde{f}^n(x) + (-1)^n k$ for all $x \in \mathbb{R}$, all $k \in \mathbb{Z}$, and all $n \in \mathbb{Z}$ (cf. Corollaries 2.1.10 and 2.2.5).

(d) Prove that if $|x-y| < k$ for some $k \in \mathbb{Z}_+$, then $|\tilde{f}^n(x) - \tilde{f}^n(y)| < k$ for any $n \in \mathbb{N}$. (This shows that Corollary 2.2.6 still holds.)

(e) If $\tilde{g}$ is another lift of f so that $\tilde{g} = \tilde{f} + k$ for some $k \in \mathbb{Z}$, then $\tilde{g}^n = \tilde{f}^n + k\sin^2(n\pi/2)$ for all $n \in \mathbb{Z}$ (cf. Corollary 2.2.7).

(f) Show that $\tilde{f} + \mathrm{Id}_{\mathbb{R}}$ is a periodic function with period 1. More generally, a decreasing homeomorphism $\tilde{g} : \mathbb{R} \to \mathbb{R}$ is a lift of an orientation-reversing homeomorphism of the circle if and only if $\tilde{g} + \mathrm{Id}_{\mathbb{R}}$ is a periodic function with period 1 (cf. Corollary 2.2.8).

(g) Show that the reflection of the unit circle in the x-axis is an orientation-reversing homeomorphism of $\mathbb{S}^1$. Then find its lifts.

(h) Prove that the number $\rho(\widetilde{f}) := \lim_{n\to\infty} \frac{\widetilde{f}^n(x)}{n}$ exists for all $x \in \mathbb{R}$ and is independent of x (cf. Proposition 2.2.10).
(i) Show that $\rho(\widetilde{f}) = 0$. This demonstrates that the concept of "rotation" number is useless for orientation-reversing homeomorphisms.

Exercise 2.4.6. Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be an orientation-preserving homeomorphism of the circle. Let $\varepsilon > 0$. Show that there exists $\delta > 0$ such that if $g : \mathbb{S}^1 \to \mathbb{S}^1$ is an orientation-preserving homeomorphism which is C^0-δ close to f, then

$$\left|\rho(g) - \rho(f)\right| < \varepsilon.$$

Hint: Reread the proof of statement (a) *in Proposition* 2.2.10.

Exercise 2.4.7. Suppose that $f, g : \mathbb{S}^1 \to \mathbb{S}^1$ are two orientation-preserving homeomorphisms of the circle. Show that if f and g commute, then $\rho(g \circ f) = \rho(g) + \rho(f)$.

Exercise 2.4.8. Suppose that $f : \mathbb{S}^1 \to \mathbb{S}^1$ is an orientation-preserving homeomorphism of the circle such that $\rho(f) \in \mathbb{R} \setminus \mathbb{Q}$. Show that $\{\rho(f^n) : n \in \mathbb{N}\}$ is dense in $\mathbb{S}^1$.

3 Symbolic dynamics

Symbolic dynamics is an extremely powerful tool for the analysis of general dynamical systems. The very rough idea is to break up a space into finitely or countably many parts, assign a symbol to each part and track the orbits of points by assigning sequences of symbols to them, representing the orbits visiting successive parts of the space. In the process, we glean information about the system by analyzing these symbolic orbits.

In this chapter, we exclusively deal with topological aspects of symbolic dynamics. Symbolic dynamics is however equally important, perhaps even more important, in the context of measure-preserving dynamical systems and ergodic theory, particularly thermodynamic formalism. We will see why in Chapters 4, 8 (especially Subsections 8.1.1 and 8.2.3) as well as in Chapters 13 and 17 onward in the second volume.

The first successful use of topological aspects of symbolic dynamics can be credited to Hadamard [26], who applied them to geodesic flows. However, it took another 40 years before the topic received its first systematic account and its name, in the foundational paper by Morse and Hedlund [51]. This paper is the first to treat symbolic dynamical systems as objects of study in and of themselves.

Since then, symbolic dynamics has found ever wider applications within dynamical systems as a whole, while still remaining an active area of research. For a deeper introduction to combinatorial and topological aspects of symbolic dynamics over finite alphabets, we refer the reader to Lind and Marcus [43]. There is also a nice chapter on symbolic dynamics over finite alphabets in the fairly recent book by Brin and Stuck [13]. For a treatment of topological symbolic dynamics over countable alphabets, we refer the reader to Kitchens [39].

In this chapter, we discuss symbolic dynamical systems as objects in their own right, but later on (in Chapter 4, among others mentioned above), we will apply the ideas developed here to more general systems. We restrict ourselves to finitely many letters, as symbolic systems over finite alphabets act on compact metrizable spaces. However, in Chapter 17, we will turn our attention to countable-alphabet symbolic dynamics.

In Section 3.1, we discuss full shifts. In Section 3.2, we talk about subshifts of finite type and in particular the characterizations of topological transitivity and exactness for such systems. Finally, in Section 3.3 we examine general subshifts of finite type.

3.1 Full shifts

Let us begin by introducing the simplest class of symbolic dynamical systems, namely the full shifts.

https://doi.org/10.1515/9783110702682-003

Definition 3.1.1. Let E be a set such that $2 \leq \#E < \infty$, where $\#E$ denotes the cardinality of E. This set will henceforth be referred to as an *alphabet*. The elements of E will be called *letters* or *symbols*.

(a) For each $n \in \mathbb{N}$, we shall denote by E^n the set of all *words* (also called *blocks*) comprising n letters from the alphabet E. For convenience, we also denote the *empty word*, that is, the word having no letters, by ϵ.
For instance, if $E = \{0,1\}$ then $E^1 = E = \{0,1\}$, $E^2 = \{00,01,10,11\}$ and
$$E^3 = \{000,100,010,001,110,101,011,111\}.$$

(b) We will denote by $E^* := \bigcup_{n=1}^{\infty} E^n$ the set of all finite nonempty words over the alphabet E.

(c) The set $E^\infty := E^{\mathbb{N}}$ of all one-sided infinite words over the alphabet E, that is, the set of all sequences or functions from $\mathbb{N}$ to E, is called the *one-sided full E-shift*, or, if no confusion about the underlying alphabet may arise, simply the full shift. In other words,
$$E^\infty := \{\omega = (\omega_j)_{j=1}^{\infty} : \omega_j \in E, \ \forall j \in \mathbb{N}\}.$$
When $E = \{0,1,\dots,n-1\}$ for some $n \in \mathbb{N}$, the full E-shift is usually referred to as the *full n-shift*.

(d) The *length* of a word $\omega \in E^* \cup E^\infty$ is defined in the natural way to be the number of letters that it consists of and is denoted by $|\omega|$. That is, $|\omega|$ is the unique $n \in \mathbb{N}\cup\{\infty\}$ such that $\omega \in E^n$. By convention, $|\epsilon| = 0$.
A word of length n is sometimes called an n-word or n-block. In our notation, the set of all n-words is simply E^n.

Note that for each $n \in \mathbb{N}$ the set E^n is finite, and hence the set E^* of finite words is countable. We can enumerate all the finite words by starting with the 1-words, followed by the 2-words, the 3-words, and so on. As $\#E \geq 2$, a classical Cantor diagonalization argument establishes that the full E-shift is uncountable. More precisely, the full E-shift has the cardinality $\mathfrak{c}$ of the continuum, that is, it is equinumerous to $\mathbb{R}$ and $[0,1]$.

One of the most common examples of a full shift is the full 10-shift, which can serve to encode the decimal expansions of the real numbers between 0 and 1. For instance, the word $0^\infty = 0000\dots$ corresponds to the number 0, while the word $1^\infty = 1111\dots$ represents the number $0.1111\dots = 1/9$. Furthermore, observe that the words $50^\infty = 5000\dots$ and $49^\infty = 4999\dots$ both encode $1/2$ since $0.5000\dots = 0.4999\dots = 1/2$. More generally, a word $\omega \in \{0,1,\dots,9\}^\infty$ encodes the number
$$\sum_{j=1}^{\infty} \omega_j \left(\frac{1}{10}\right)^j.$$
As noted above, this coding is not one-to-one.

Another common example of a full shift is the full 2-shift. In computer science, the symbol 0 might correspond to having a device (e. g., a switch) turned off, while the symbol 1 would then correspond to the device being on. In physics, the Ising model describes particle spins that have only two possible states, up and down. So we can describe the successive states of a device or particle, at regular observation times, using a sequence of 0s and 1s. Words in the full 2-shift are also called binary sequences, as they correspond to the binary expansions of the numbers between 0 and 1. Indeed, a word $\omega \in \{0,1\}^\infty$ may be viewed as representing the real number

$$\sum_{j=1}^{\infty} \omega_j \left(\frac{1}{2}\right)^j.$$

For instance, the word 0^∞ corresponds to the number 0, as in the full 10-shift. However, the word 1^∞, which encoded the number 1/9 as a word in the full 10-shift, represents the number 1 in the full 2-shift; the number 1/9 is instead encoded by the word $(000111)^\infty = 000111000111\ldots$ in the full 2-shift. As for the number 1/2, it is encoded by the words $10^\infty = 1000\ldots$ and $01^\infty = 0111\ldots$. Hence, this binary coding is not one-to-one either.

The idea of approximating a real number by a rational number by cutting it after a certain number of decimals generalizes to the concept of an *initial block*. Initial blocks play an important role in symbolic dynamics.

Definition 3.1.2. If $\omega \in E^* \cup E^\infty$ and $n \in \mathbb{N}$ does not exceed the length of ω, we define the *initial block* $\omega|_n$ to be the initial n-word of ω, that is, the subword $\omega_1\omega_2\ldots\omega_n$.

In a similar vein, words which begin with the same strings of letters are intuitively close to one another and it is therefore useful to identify the initial subword that they share. To describe this, we introduce the wedge of two words.

Definition 3.1.3. Given two words $\omega, \tau \in E^* \cup E^\infty$, we define their *wedge* $\omega \wedge \tau \in \{\epsilon\} \cup E^* \cup E^\infty$ to be their longest common initial block.

The wedge of two words is better understood via examples. If $E = \{1,2,3\}$ and we have two words $\omega = 12321\ldots$ and $\tau = 12331\ldots$, then $\omega \wedge \tau = 123$. On the other hand, if $\gamma = 22331\ldots$ then $\omega \wedge \gamma = \epsilon$. Of course, if two (finite or infinite) words ω and τ are equal, then $\omega \wedge \tau = \omega = \tau$.

So far, we have talked about the set E^∞ and have given a natural sense to the closeness of any two of its words in terms of their common initial block. We will now endow E^∞ with a natural topology. First, the finite alphabet E is endowed with the discrete topology, that is, the topology in which every subset of E is both open and closed. Then, observing that $E^\infty = \prod_{n=1}^{\infty} E_n$, where $E_n = E$ for each $n \in \mathbb{N}$, that is, E^∞ is the product of countably many copies of E, we equip E^∞ with Tychonoff's product topology generated by the discrete topology on each copy of E. More precisely, that topology is determined by a countable base of open sets called *(initial) cylinder sets*.

Definition 3.1.4. Given a finite word $\omega \in E^*$, the *(initial) cylinder set* $[\omega]$ generated by ω is the set of all infinite words with initial block ω, that is,

$$[\omega] = \{\tau \in E^\infty : \tau|_{|\omega|} = \omega\} = \{\tau \in E^\infty : \tau_j = \omega_j,\ \forall 1 \le j \le |\omega|\}.$$

We take this opportunity to introduce more general cylinder sets.

Definition 3.1.5. Given a finite word $\omega \in E^*$ and $m, n \in \mathbb{N}$ such that $n - m + 1 = |\omega|$, the (m, n)*-cylinder set* $[\omega]_m^n$ generated by ω is the set of all infinite words whose subblock from coordinates m to n coincides with ω, that is,

$$[\omega]_m^n = \{\tau \in E^\infty : \tau_j = \omega_{j-m+1} \text{ for all } m \le j \le n\}.$$

In particular, note that $[\omega]_1^{|\omega|} = [\omega]$.

Definition 3.1.6. When equipped with Tychonoff's product topology, that is, the topology generated by the (initial) cylinder sets, the set E^∞ is called the *full E-shift space* or, more simply, full shift space.

We now describe the most fundamental topological properties of full shift spaces. First, note that there are countably many (initial) cylinder sets since there are countably many finite words in E^*. Since these sets form a base for the topology, the space E^∞ is second countable.

Furthermore, since the alphabet E is finite, when endowed with the discrete topology it becomes a compact metrizable space (see Exercise 3.4.2). According to Tychonoff's theorem (see 17.8 in [77]), it then follows that E^∞, as a product of countably many copies of E, is also a compact metrizable space (see Exercise 3.4.2).

Moreover, the full shift space E^∞ is perfect, that is, it contains no isolated point. Indeed, notice that every point $\omega \in E^\infty$ is such that

$$\{\omega\} = \bigcap_{n=1}^{\infty} [\omega|_n].$$

Alternatively, note that every point $\omega \in E^\infty$ is the limit of the sequence of "periodic" points $((\omega|_n)^\infty)_{n=1}^\infty = ((\omega_1 \ldots \omega_n)^\infty)_{n=1}^\infty$.

Finally, since the complement of a cylinder set is a union of cylinder sets, the cylinder sets are both open and closed subsets of the full shift space. Therefore, E^∞ is totally disconnected (see Exercise 3.4.3). Summarizing these properties, we have obtained the following lemma.

Lemma 3.1.7. *The full shift space E^∞ is a totally disconnected, perfect, compact, metrizable space.*

To put this lemma in context, we recall the following definition.

Definition 3.1.8. A *Cantor space* (also frequently called a Cantor set) is a totally disconnected, perfect, compact, second-countable, Hausdorff topological space.

In light of this definition, we can restate Lemma 3.1.7 as follows.

Lemma 3.1.9. *The full shift space E^∞ is a Cantor space.*

Cantor spaces can be characterized as follows:
(a) They are totally disconnected, perfect, compact, metrizable topological spaces.
(b) They are homeomorphic to the middle-third Cantor set.
(c) They are homeomorphic to E^∞ for some finite set E having at least two elements.
(d) They are homeomorphic to E^∞ for every finite set E having at least two elements.

We now introduce a family of metrics on E^∞, each of which reflects the idea that two words are close if they share a long initial block. The longer their common initial subword, the closer two words are.

Definition 3.1.10. For each $s \in (0,1)$, let $d_s : E^\infty \times E^\infty \to [0,1]$ be defined by

$$d_s(\omega,\tau) = s^{|\omega\wedge\tau|}.$$

Remark 3.1.11. If $\omega,\tau \in E^\infty$ have no common initial block, then $\omega \wedge \tau = \epsilon$. Thus, $|\omega\wedge\tau| = 0$ and $d_s(\omega,\tau) = 1$. On the other hand, if $\omega = \tau$ then $|\omega\wedge\tau| = \infty$ and we adopt the convention that $s^\infty := 0$.

Proposition 3.1.12. *For every $s \in (0,1)$, the map $d_s : E^\infty \times E^\infty \to [0,1]$ defined above is an ultrametric, and thus a metric.*

Proof. First, note that $d_s(\omega,\omega) = s^\infty := 0$. Moreover, $d_s(\omega,\tau) = 0$ implies that $|\omega\wedge\tau| = \infty$, that is, $\omega = \tau$. Second, d_s is symmetric, as

$$d_s(\omega,\tau) = s^{|\omega\wedge\tau|} = s^{|\tau\wedge\omega|} = d_s(\tau,\omega).$$

It only remains to show that $d_s(\omega,\tau) \le \max\{d_s(\omega,\rho), d_s(\rho,\tau)\}$ for all $\omega,\rho,\tau \in E^\infty$. Fix $\omega,\rho,\tau \in E^\infty$. Observe that ω and ρ share the same initial block of length $|\omega\wedge\rho|$, while ρ and τ share the same initial block of length $|\rho\wedge\tau|$. This implies that ω, ρ, and τ all share the same initial block of length equal to $\min\{|\omega\wedge\rho|, |\rho\wedge\tau|\}$. Since $0 < s < 1$, we then have that

$$d_s(\omega,\tau) \le s^{\min\{|\omega\wedge\rho|,|\rho\wedge\tau|\}} = \max\{s^{|\omega\wedge\rho|}, s^{|\rho\wedge\tau|}\} = \max\{d_s(\omega,\rho), d_s(\rho,\tau)\}.$$

This shows that d_s is an ultrametric. In particular, it is a metric as the triangle inequality

$$d_s(\omega,\tau) \le d_s(\omega,\rho) + d_s(\rho,\tau)$$

is obviously satisfied. ☐

We have now defined an uncountable family of metrics on E^∞, one for each $s \in (0,1)$. These metrics induce Tychonoff's topology on E^∞ (see Exercise 3.4.4). This implies that these metrics are topologically equivalent. In fact, they are Hölder equivalent but not Lipschitz equivalent (see Exercises 3.4.5 and 3.4.6).

Let us now describe what it means for a sequence to converge to a limit in the full shift space. Let $s \in (0,1)$. Let $(\omega^{(k)})_{k=1}^\infty$ be a sequence in E^∞. Observe that

$$\lim_{k\to\infty} d_s(\omega^{(k)}, \omega) = \lim_{k\to\infty} s^{|\omega^{(k)}\wedge\omega|} = 0 \Leftrightarrow \lim_{k\to\infty} |\omega^{(k)} \wedge \omega| = \infty.$$

In other words, a sequence $(\omega^{(k)})_{k=1}^\infty$ converges to the infinite word ω if and only if for any $L \in \mathbb{N}$, the words in the sequence eventually all have $\omega|_L$ as initial L-block.

Now that we have explored the space E^∞, we would like to introduce some dynamics on it. To this end, we define the shift map, whose action consists in removing the first letter of each word and shifting all the remaining letters one space/coordinate to the left.

Definition 3.1.13. The *full left-shift map* $\sigma : E^\infty \to E^\infty$ is defined by $\sigma(\omega) = \sigma((\omega_j)_{j=1}^\infty) := (\omega_{j+1})_{j=1}^\infty$, that is,

$$\sigma(\omega_1\omega_2\omega_3\omega_4\ldots) := \omega_2\omega_3\omega_4\ldots$$

We will also often refer to this map simply as the shift map.

The shift map is $\#E$-to-one on E^∞. In other words, each word has $\#E$ preimages under the shift map. Indeed, given any letter $e \in E$ and any infinite word $\omega \in E^\infty$, the concatenation $e\omega = e\omega_1\omega_2\omega_3\ldots$ of e with ω is a preimage of ω under the shift map since $\sigma(e\omega) = \omega$.

The shift map is obviously continuous, since two words that are close share a long initial block and thus their images under the shift map, which result from dropping their first letters, will also share a long initial block. More precisely, for any $\omega, \tau \in E^\infty$ with $d_s(\omega, \tau) < 1$, that is, with $|\omega \wedge \tau| \geq 1$, we have that

$$d_s(\sigma(\omega), \sigma(\tau)) = s^{|\sigma(\omega)\wedge\sigma(\tau)|} = s^{|\omega\wedge\tau|-1} = s^{-1}s^{|\omega\wedge\tau|} = s^{-1}d_s(\omega, \tau).$$

So the shift map is Lipschitz continuous with Lipschitz constant s^{-1}. In particular, the shift map defines a dynamical system on E^∞. It is then natural to ask the following question: Given two finite sets E and F, under which conditions are the shift maps $\sigma_E : E^\infty \to E^\infty$ and $\sigma_F : F^\infty \to F^\infty$ topologically conjugate? Notice that the only fixed points of σ_E are the "constant" words e^∞, for each $e \in E$. Hence, the number of fixed points of σ_E is equal to $\#E$. Recall that the cardinality of the set of fixed points $\mathrm{Fix}(T)$ is a topological invariant. So, if $\#E \neq \#F$ then σ_E is not topologically conjugate to σ_F. In fact, as we will see in the following theorem, σ_E is topologically conjugate to σ_F precisely when $\#E = \#F$.

Theorem 3.1.14. *$\sigma_E : E^\infty \to E^\infty$ and $\sigma_F : F^\infty \to F^\infty$ are topologically conjugate if and only if $\#E = \#F$.*

Proof. If σ_E and σ_F are topologically conjugate, then it is clear from the discussion above that $\#E = \#F$. For the converse, assume that $\#E = \#F$. Therefore, there must exist some bijection $H : E \to F$. Now define the mapping $h : E^\infty \to F^\infty$ by concatenation, that is, by setting

$$h(\omega_1\omega_2\omega_3\ldots) := H(\omega_1)H(\omega_2)H(\omega_3)\ldots.$$

Then h is a homeomorphism: that h is a bijection follows from H being a bijection, while the continuity of both h and h^{-1} follows directly from the fact that h is an isometry, as $|h(\omega) \wedge h(\tau)| = |\omega \wedge \tau|$. It remains to show that the following diagram commutes:

$$\begin{array}{ccc} E^\infty & \xrightarrow{\sigma_E} & E^\infty \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle h} \\ F^\infty & \xrightarrow[\sigma_F]{} & F^\infty \end{array}$$

Indeed,

$$\begin{aligned} h \circ \sigma_E(\omega_1\omega_2\omega_3\ldots) &= h(\omega_2\omega_3\omega_4\ldots) \\ &= H(\omega_2)H(\omega_3)H(\omega_4)\ldots \\ &= \sigma_F\big(H(\omega_1)H(\omega_2)H(\omega_3)\ldots\big) \\ &= \sigma_F \circ h(\omega_1\omega_2\omega_3\ldots). \end{aligned}$$

□

3.2 Subshifts of finite type

We now turn our attention to subsystems of full shift spaces. By definition, the subsystems of the full shift space E^∞ are all shift-invariant, compact subsets of E^∞. Recall that a set $F \subseteq E^\infty$ is shift-invariant (i. e., σ-invariant) if $\sigma(F) \subseteq F$.

The notion of forbidden word arises naturally in the study of subsets and subsystems of full shift spaces. Let $\mathcal{F} \subseteq E^*$ be a set of finite words, called *forbidden words* in the sequel. We shall denote by $E^\infty_{\mathcal{F}}$ the set of all those infinite words in E^∞ that do not contain any forbidden word as a subword. In other words,

$$E^\infty_{\mathcal{F}} := \{\omega \in E^\infty : \omega_m\omega_{m+1}\ldots\omega_n \notin \mathcal{F},\ \forall m, n \in \mathbb{N},\ m \le n\}.$$

A subshift is a subset of a full shift that can be described by a set of forbidden words.

Definition 3.2.1. A subset F of the full shift E^∞ is called a *subshift* if there is a set $\mathcal{F} \subseteq E^*$ of forbidden words such that $F = E^\infty_{\mathcal{F}}$. It is sometimes said that $F = E^\infty_{\mathcal{F}}$ is the subshift generated by $\mathcal{F}$.

We briefly examine the relation between sets of forbidden words and the subshifts they generate.

Lemma 3.2.2. *Let $\mathcal{F} \subseteq E^*$ and $\mathcal{G} \subseteq E^*$.*
(a) *If $\mathcal{F} \subseteq \mathcal{G}$, then $E^\infty_\mathcal{F} \supseteq E^\infty_\mathcal{G}$.*
(b) *If $\mathcal{F} \subseteq \mathcal{G}$ and every word in $\mathcal{G}$ admits a subword which is in $\mathcal{F}$, then $E^\infty_\mathcal{F} = E^\infty_\mathcal{G}$.*

Proof.
(a) Suppose that $\mathcal{F} \subseteq \mathcal{G}$. If $\tau \notin \mathcal{G}$ then $\tau \notin \mathcal{F}$, and hence

$$\begin{aligned} E^\infty_\mathcal{G} &= \{\omega \in E^\infty : \omega_m\omega_{m+1}\dots\omega_n \notin \mathcal{G},\ \forall m,n \in \mathbb{N},\ m \le n\} \\ &\subseteq \{\omega \in E^\infty : \omega_m\omega_{m+1}\dots\omega_n \notin \mathcal{F},\ \forall m,n \in \mathbb{N},\ m \le n\} \\ &= E^\infty_\mathcal{F}. \end{aligned}$$

(b) Suppose that $\mathcal{F} \subseteq \mathcal{G}$ and that every word in $\mathcal{G}$ admits a subword which is in $\mathcal{F}$. By (a), we already know that $E^\infty_\mathcal{G} \subseteq E^\infty_\mathcal{F}$. Therefore, it only remains to establish that $E^\infty_\mathcal{G} \supseteq E^\infty_\mathcal{F}$. Let $\omega \in E^\infty \setminus E^\infty_\mathcal{G}$. Then there exist $m,n \in \mathbb{N}$, $m \le n$, such that $\omega_m\omega_{m+1}\dots\omega_n \in \mathcal{G}$. Since every word in $\mathcal{G}$ contains a subword which is in $\mathcal{F}$, there exist $k,l \in \mathbb{N}$ such that $m \le k \le l \le n$ and $\omega_k\omega_{k+1}\dots\omega_l \in \mathcal{F}$. Thus, $\omega \in E^\infty \setminus E^\infty_\mathcal{F}$. This means that $E^\infty \setminus E^\infty_\mathcal{G} \subseteq E^\infty \setminus E^\infty_\mathcal{F}$, and hence $E^\infty_\mathcal{G} \supseteq E^\infty_\mathcal{F}$. □

Taken together, the next two theorems demonstrate that the terms *subsystem of a full shift* and *subshift* can be used interchangeably.

Theorem 3.2.3. *Every subshift of E^∞ is a subsystem of the full shift E^∞.*

Proof. Let F be a subshift of E^∞. This means that there exists $\mathcal{F} \subseteq E^*$ such that $F = E^\infty_\mathcal{F}$. From its definition, it is clear that $E^\infty_\mathcal{F}$ is shift invariant. Moreover, observe that

$$\begin{aligned} E^\infty_\mathcal{F} &= \bigcap_{m=1}^{\infty}\bigcap_{n\ge m}\{\omega \in E^\infty : \omega_m\dots\omega_n \notin \mathcal{F}\} \\ &= \bigcap_{m=1}^{\infty}\bigcap_{n\ge m}\bigcap_{\tau\in E^{n-m+1}\cap\mathcal{F}}\{\omega \in E^\infty : \omega_m\dots\omega_n \ne \tau\} \\ &= \bigcap_{m=1}^{\infty}\bigcap_{n\ge m}\bigcap_{\tau\in E^{n-m+1}\cap\mathcal{F}} E^\infty \setminus [\tau]_m^n. \end{aligned}$$

Since every cylinder set is open in E^∞, the sets $E^\infty \setminus [\tau]_m^n$ are compact. Therefore, $E^\infty_\mathcal{F}$ is an intersection of compact sets and is thereby compact. In summary, $F = E^\infty_\mathcal{F}$ is a compact shift-invariant subset of E^∞. That is, it is a subsystem of the full shift E^∞. □

Theorem 3.2.4. *Let F be a subsystem of the full shift space E^∞. Let*

$$\mathcal{F}_F := \{\tau \in E^* : [\tau] \subseteq E^\infty \setminus F\}.$$

Then

$$F = E^{\infty}_{\mathcal{F}_F}.$$

In other words, the subsystem F of the full shift space E^{∞} coincides with the subshift $E^{\infty}_{\mathcal{F}_F}$ generated by the set of finite words $\mathcal{F}_F$.

Proof. By hypothesis, the set $F \subseteq E^{\infty}$ is σ-invariant and compact. In particular, F is closed. Therefore, $E^{\infty} \setminus F$ is open.

Let $\rho \in E^{\infty} \setminus F$. Since $E^{\infty} \setminus F$ is open, this is equivalent to the existence of $n \in \mathbb{N}$ such that $[\rho|_n] \subseteq E^{\infty} \setminus F$. In turn, this means that $\rho|_n \in \mathcal{F}_F$ and hence $\rho \notin E^{\infty}_{\mathcal{F}_F}$.

Conversely, assume that $\rho \notin E^{\infty}_{\mathcal{F}_F}$. There exist $m, n \in \mathbb{N}$, $m \leq n$, such that $\rho_m\rho_{m+1}\rho_n \in \mathcal{F}_F$. In other terms, $\sigma^{m-1}(\rho)_1\sigma^{m-1}(\rho)_2 \ldots \sigma^{m-1}(\rho)_{n-m+1} \in \mathcal{F}_F$. That is, $\sigma^{m-1}(\rho)|_{n-m+1} \in \mathcal{F}_F$. This is equivalent to $[\sigma^{m-1}(\rho)|_{n-m+1}] \subseteq E^{\infty} \setminus F$. In particular, $\sigma^{m-1}(\rho) \in E^{\infty} \setminus F$. Since F is σ-invariant, this implies that $\rho \in E^{\infty} \setminus F$. □

We shall now study a special class of subshifts. They are called *subshifts of finite type*.

Definition 3.2.5. A subshift F of the full shift E^{∞} is said to be of *finite type* if there is a *finite* set $\mathcal{F} \subseteq E^*$ of forbidden words such that $F = E^{\infty}_{\mathcal{F}}$.

In this case, it easily follows from Lemma 3.2.2(b) that the finite set $\mathcal{F}$ can be chosen so that $\mathcal{F} \subseteq E^q$ for some $q \in \mathbb{N}$. The set $\mathcal{F}$ then induces a function $A : E^q \to \{0,1\}$ whose value is 0 on $\mathcal{F}$ (i. e., for all forbidden words of length q) and 1 on $E^q \setminus \mathcal{F}$ (i. e., for all other words of length q). We will revisit this general framework in the next section, where we will prove that all cases can be reduced to the case $q = 2$. For this reason, we will concentrate on this latter case in this section. Here, rather than using formally a function $A : E^2 \to \{0,1\}$, subshifts of finite type are best understood by means of an incidence/transition matrix. An incidence/transition matrix is simply a square matrix consisting entirely of zeros and ones. To do this, we will work with the full shift on the alphabet $\{1, 2, \ldots, \#E\}$ rather than the full E-shift itself. The incidence/transition matrix determines which letter/number(s) may follow a given letter/number.

Definition 3.2.6. Let A be an incidence matrix of size $\#E \times \#E$. The set of all infinite *A-admissible words* is the *subshift of finite type*

$$E^{\infty}_A := \{\omega \in E^{\infty} : A_{\omega_n\omega_{n+1}} = 1, \ \forall n \in \mathbb{N}\}.$$

E^{∞}_A is a subshift of finite type since $E^{\infty}_A = E^{\infty}_{\mathcal{F}}$, where the set of forbidden words $\mathcal{F}$ is the finite set of two-letter words

$$\mathcal{F} = \{ij \in E^2 : A_{ij} = 0\}.$$

A (finite) word $\omega_1\omega_2 \ldots \omega_n$ is said to be *A-admissible* (or, if there can be no confusion about the matrix A, more simply, *admissible*) if

$$A_{\omega_k\omega_{k+1}} = 1, \quad \forall 1 \leq k < n.$$

The set of all A-admissible n-words will naturally be denoted by E_A^n, while the set of all A-admissible finite words will naturally be denoted by E_A^*. An A-admissible path of length n from $i \in E$ to $j \in E$ is any A-admissible word ω of length n with $\omega_1 = i$ and $\omega_n = j$. Thus, the entry A_{ij} of the matrix A indicates the number of admissible words (or paths) of length 2 from i to j (which is necessarily either 0 or 1). By multiplying the matrix A with itself, we see that $(A^2)_{ij} = \sum_{k=1}^{\#E} A_{ik}A_{kj}$ specifies the number of admissible words of length 3 from i to j, since $A_{ik}A_{kj} = 1$ if and only if ikj is admissible. Similarly, $(A^n)_{ij}$ is the number of admissible words of length $n+1$ from i to j, and $(A^n)_{ij} > 0$ if and only if there is at least one such path.

Note that if a row of A does not contain any 1, then no infinite word can contain the letter corresponding to that row. This letter can then be thrown out of the alphabet because it is inessential. We will henceforth assume that this does not happen, that is, we will assume that all the letters are essential by imposing the condition that every row of A contains at least one 1. This is a standing assumption throughout this book.

Notice that if all the entries of the incidence matrix A are 1s, then $\mathcal{F} = \emptyset$ and $E_A^\infty = E^\infty$. However, if A has at least one 0 entry then E_A^∞ is a proper subshift of E^∞. In particular, if A is the identity matrix then $E_A^\infty = \{e^\infty : e \in E\}$, that is, E_A^∞ is the set of all constant words, which are the fixed points of σ in E^∞.

Alternatively, E_A^∞ can be represented by a directed graph. Imagine that each element e of E is a vertex of a directed graph. Then the directed graph has an edge directed from vertex e to vertex f if and only if $A_{ef} = 1$. The set of infinite A-admissible words E_A^∞ then corresponds to the set of all possible infinite walks along the directed graph. This is sometimes called a *vertex shift*.

Example 3.2.7. Let $E = \{1, 2, 3\}$ and let

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

What is E_A^∞? We strongly advise the reader to draw the corresponding directed graph. According to the incidence matrix, the letter 3 can only be followed by itself since $A_{33} = 1$ while $A_{31} = A_{32} = 0$. This means that vertex 3 of the directed graph has only one outgoing edge, and it is a self-loop. Thus, if $\omega \in E_A^\infty$ starts with a 3, then $\omega = 3^\infty$. According to the incidence matrix again, the letter 1 can only be followed by itself or by 3 since $A_{11} = A_{13} = 1$ whereas $A_{12} = 0$. This means that vertex 1 has two outgoing edges, one being a self-loop while the other terminates at vertex 3. Thus, if ω starts with a 1, then this can be followed by either a 3, in which case it is followed by infinitely many 3s, or by a 1, in which case we face the same choice again. Therefore, the admissible words starting with a 1 are 1^∞ and $1^n 3^\infty$, $n \in \mathbb{N}$. Finally, if ω starts with a 2 then, as $A_{21} = A_{22} = A_{23} = 1$, a 2 can be followed by any other letter. This means that vertex 2 has three outgoing edges, one terminating at each vertex. Hence the admissible words starting with a 2 are 2^∞, 2^m3^∞, 2^m1^∞ and $2^m1^n3^\infty$, where $m, n \in \mathbb{N}$. In summary,

$$E_A^\infty = \{1^\infty, 2^\infty, 3^\infty\} \cup \{2^m1^\infty : m \in \mathbb{N}\} \cup \{2^m1^n3^\infty : m, n \in \mathbb{Z}_+\}.$$

We now study the topological properties of E_A^∞. Since E_A^∞ is a subshift, Theorem 3.2.3 yields that this set is compact and σ-invariant. Nevertheless, we provide below a direct proof of this important fact.

Theorem 3.2.8. *E_A^∞ is a compact σ-invariant set.*

Proof. Let $\omega \in E_A^\infty$. Then $A_{\omega_n\omega_{n+1}} = 1$ for every $n \in \mathbb{N}$. In particular, this implies that $A_{\sigma(\omega)_n\sigma(\omega)_{n+1}} = A_{\omega_{n+1}\omega_{n+2}} = 1$ for all $n \in \mathbb{N}$. Thus, $\sigma(\omega) \in E_A^\infty$ and we therefore have that E_A^∞ is σ-invariant.

In order to show that E_A^∞ is compact, recall that a closed subset of a compact space is compact. As E^∞ is a compact space when endowed with the product topology, it is sufficient to prove that E_A^∞ is closed. Let $(\omega^{(k)})_{k=1}^\infty$ be a sequence in E_A^∞ and suppose that $\lim_{k\to\infty} \omega^{(k)} = \omega$. We must show that $\omega \in E_A^\infty$, or, in other words, we need to show that $A_{\omega_n\omega_{n+1}} = 1$ for all $n \in \mathbb{N}$. To that end, fix $n \in \mathbb{N}$. For each $k \in \mathbb{N}$, we have $A_{\omega_n^{(k)}\omega_{n+1}^{(k)}} = 1$ since $\omega^{(k)} \in E_A^\infty$. Moreover, $\lim_{k\to\infty} |\omega^{(k)} \wedge \omega| = \infty$ since $\omega^{(k)} \to \omega$. So, for sufficiently large k, we have $|\omega^{(k)} \wedge \omega| \geq n+1$. In particular, $\omega_n^{(k)} = \omega_n$ and $\omega_{n+1}^{(k)} = \omega_{n+1}$ for all k large enough. Hence, we deduce that $A_{\omega_n\omega_{n+1}} = A_{\omega_n^{(k)}\omega_{n+1}^{(k)}} = 1$ for all k large enough. Since n was chosen arbitrarily, we conclude that $\omega \in E_A^\infty$.

We now provide an alternative proof of the compactness of E_A^∞. Observe that

$$\begin{aligned} E_A^\infty &= \{\omega \in E^\infty : A_{\omega_n\omega_{n+1}} = 1, \forall n \in \mathbb{N}\} \\ &= \bigcap_{n=1}^\infty \{\omega \in E^\infty : A_{\omega_n\omega_{n+1}} = 1\} \\ &= \bigcap_{n=1}^\infty \Big[\bigcup_{\omega \in E_A^2} [\omega]_n^{n+1} \Big]. \end{aligned}$$

Recall that all cylinders are closed subsets of the compact space E^∞. Therefore, they are all compact. Since the set E_A^2 is finite, the union of cylinders $\bigcup_{\omega\in E_A^2} [\omega]_n^{n+1}$ is compact for all $n \in \mathbb{N}$. As an intersection of these latter sets, E_A^∞ is compact. □

The σ-invariance of E_A^∞ ensures that the map $\sigma : E_A^\infty \to E_A^\infty$ is well-defined. This restriction of the shift map is obviously continuous (in fact, Lipschitz continuous). The compactness of E_A^∞ makes the couple (E_A^∞, σ) a well-defined topological dynamical system. It is a subsystem of the full shift (E^∞, σ).

We shall now determine the condition under which $\sigma : E_A^\infty \to E_A^\infty$ is surjective.

Lemma 3.2.9. *The (restriction of) the shift map $\sigma : E_A^\infty \to E_A^\infty$ is surjective if and only if for every $j \in E$ there exists $i \in E$ such that $A_{ij} = 1$, that is, if and only if every column of A contains at least one* 1.

Proof. First, observe that if for every $j \in E$ there exists some $i \in E$ such that $A_{ij} = 1$, then $\sigma : E_A^\infty \to E_A^\infty$ is surjective. Indeed, given any $\omega \in E_A^\infty$, there exists $e \in E$ such that $A_{e\omega_1} = 1$. Then $e\omega \in E_A^\infty$ and $\sigma(e\omega) = \omega$.

To establish the converse, suppose A has a column consisting solely of 0s, that is, suppose there exists $j \in E$ such that $A_{ij} = 0$ for all $i \in E$. By our standing assumption, every row of A contains at least one 1. So there must be a word of the form $j\omega$ in E_A^∞. However, $j\omega \notin \sigma(E_A^\infty)$ since there is no word of the form $ij\omega$ in E_A^∞. □

After determining when the map $\sigma : E_A^\infty \to E_A^\infty$ is surjective, it is natural to next consider the condition under which the shift map is injective on a subshift of finite type.

Lemma 3.2.10. *The shift map $\sigma : E_A^\infty \to E_A^\infty$ is injective if and only if for every $j \in E$ there exists at most one $i \in E$ such that $A_{ij} = 1$, that is, if and only if A contains at most one* 1 *in each of its columns.*

The proof of this lemma is left to the reader as an exercise (see Exercise 3.4.11).

Corollary 3.2.11. *The shift map $\sigma : E_A^\infty \to E_A^\infty$ is bijective if and only if every column of A contains exactly one* 1. *Given our standing assumption that all letters are essential, we thus have that $\sigma : E_A^\infty \to E_A^\infty$ is bijective if and only if every row and every column of A contains exactly one* 1.

A matrix which has precisely one 1 in each of its rows and each of its columns is called a *permutation matrix*. Such a matrix has the property that A^n is the identity matrix for some $n \in \mathbb{N}$. This means that $\sigma : E_A^\infty \to E_A^\infty$ is bijective if and only if E_A^∞ consists solely of finitely many periodic points.

In addition to being continuous, the shift map is an open map. Recall that a map is said to be open if it sends open sets onto open sets. Moreover, note that as the cylinder sets form a base for the topology on E^∞, their restriction to E_A^∞, which we also call cylinders and which will be denoted by the same notation, constitute a base for the topology on E_A^∞. From this point on, a cylinder $[\omega]_m^n$ will be understood to be a cylinder in E_A^∞. That is to say,

$$[\omega]_m^n := \{\tau \in E_A^\infty : \tau_k = \omega_{k-m+1},\ \forall m \le k \le n\}.$$

Theorem 3.2.12. *The shift map $\sigma : E_A^\infty \to E_A^\infty$ is an open map.*

Proof. As the cylinder sets of length at least two form a base for the topology on E_A^∞, it suffices to prove that the image of an arbitrary cylinder of length at least two is a cylinder. Say $\omega = \omega_1\omega_2\ldots\omega_n$, where $n \ge 2$. Then $\sigma([\omega]) = [\omega_2\ldots\omega_n]$. Indeed, let $\tau \in [\omega]$. Then $\tau = \omega_1\omega_2\ldots\omega_n\tau_{n+1}\tau_{n+2}\ldots$, and thus $\sigma(\tau) = \omega_2\ldots\omega_n\tau_{n+1}\tau_{n+2}\ldots$. Hence $\sigma([\omega]) \subseteq [\omega_2\ldots\omega_n]$. Conversely, let $\gamma \in [\omega_2\ldots\omega_n]$. Since $A_{\omega_1\omega_2} = 1$, we have that $\omega_1\gamma = \omega_1\omega_2\ldots\omega_n\gamma_{n+1}\gamma_{n+2}\cdots \in E_A^\infty$. In fact, observe that $\omega_1\gamma = \omega\gamma_{n+1}\gamma_{n+2}\cdots \in [\omega]$.

Moreover, $\sigma(\omega_1\gamma) = \gamma$. Hence, $\sigma([\omega]) \supseteq [\omega_2 \dots \omega_n]$. We have thus established that

$$\sigma\big([\omega_1\omega_2 \dots \omega_n]\big) = [\omega_2 \dots \omega_n].$$

Hence, the image of a cylinder of length $n \geq 2$ is the cylinder of length $n-1$ obtained by dropping the first symbol. □

3.2.1 Topological transitivity

We now describe the condition on the matrix A under which the subshift of finite type $\sigma : E_A^\infty \to E_A^\infty$ is topologically transitive. Recall that a dynamical system $T : X \to X$ is defined to be transitive if it admits at least one point x with the property that $\omega(x) = X$ (see Definition 1.5.6). We proved in Theorem 1.5.11 that this is equivalent to the system T being topologically mixing, that is, the orbit under T of every nonempty open set encounters every nonempty open set in X (cf. Definition 1.5.10).

Definition 3.2.13. An incidence matrix A is called *irreducible* if for each ordered pair $i, j \in E$ there exists $p := p(i, j) \in \mathbb{N}$ such that $(A^p)_{ij} > 0$.

Observe that an irreducible matrix cannot contain any row or column consisting solely of 0s. In light of Lemma 3.2.9, the irreducibility of a matrix A compels the surjectivity of the shift map $\sigma : E_A^\infty \to E_A^\infty$. We shall now prove that irreducibility of A is equivalent to the transitivity of σ.

Theorem 3.2.14. *The shift map $\sigma : E_A^\infty \to E_A^\infty$ is transitive if and only if the matrix A is irreducible.*

Proof. First, suppose that σ is transitive. By Remark 1.5.7(c), it is surjective, and thus topologically mixing according to Theorem 1.5.11. Fix $i, j \in E$. Then there exists $p := p(i, j) \in \mathbb{N}$ such that $\sigma^p([i]) \cap [j] \neq \emptyset$. So pick $\omega \in [i]$ so that $\sigma^p(\omega) \in [j]$. This means that $\omega_1\omega_2 \dots \omega_{p+1}$ is an admissible word of length $p+1$ from $i = \omega_1$ to $j = \omega_{p+1}$. Thus $(A^p)_{ij} > 0$. Since $i, j \in E$ were chosen arbitrarily, we conclude that A is irreducible.

To prove the converse, suppose that A is irreducible. As the cylinder sets form a base for the topology, it is sufficient to restrict our attention to them. Let $\omega = \omega_1\omega_2 \dots \omega_k \in E_A^*$ and $\tau = \tau_1\tau_2 \dots \tau_l \in E_A^*$. We need to show that there exists some $n \in \mathbb{N}$ such that $\sigma^n([\omega]) \cap [\tau] \neq \emptyset$. Consider the pair of letters ω_k and τ_1. Since A is irreducible, there exists some $p := p(\omega_k, \tau_1) \in \mathbb{N}$ such that $(A^p)_{\omega_k\tau_1} > 0$. This means that there exists a finite word γ of length $p-1$ such that $\omega_k\gamma\tau_1$ is an admissible word of length $p+1$ from ω_k to τ_1. Consequently, $\omega\gamma\tau \in E_A^*$. Concatenate to this word a suffix $\overline{\tau}_{l+1}$ with the property that $A_{\tau_l\overline{\tau}_{l+1}} = 1$. This is possible thanks to our standing assumption that every row of A contains a 1. Continue concatenating in this way to build an infinite admissible word $\overline{\omega} = \omega\gamma\tau\overline{\tau}_{l+1}\overline{\tau}_{l+2} \dots \in E_A^\infty$. Then $\overline{\omega} \in [\omega]$ and

$\sigma^{k+p-1}(\overline{\omega}) = \tau\overline{\tau}_{l+1}\overline{\tau}_{l+2}\ldots \in [\tau]$. Hence $\sigma^{k+p-1}([\omega]) \cap [\tau] \neq \emptyset$. Since $\omega, \tau \in E_A^*$ were arbitrary, we deduce that σ is topologically mixing. The surjectivity of σ is also guaranteed by the irreducibility of A. Therefore, σ is transitive according to Theorem 1.5.11. □

Note that the above theorem does not hold if our standing assumption that every row of A contains at least one 1 is dropped (see Exercise 3.4.12). Also, as mentioned before, it follows immediately from both the transitivity of σ and from the irreducibility of A that every column of A contains at least one 1, or equivalently, that σ is surjective (see Lemma 3.2.9).

3.2.2 Topological exactness

We now describe the condition on the matrix A under which the subshift of finite type $\sigma : E_A^\infty \to E_A^\infty$ is topologically exact. Recall that a dynamical system $T : X \to X$ is said to be topologically exact if every nonempty open set in X is eventually blown up onto the entire space X under iteration that is, for every nonempty open set $U \subseteq X$ there is $N \in \mathbb{N}$ such that $T^N(U) = X$ (cf. Definition 1.5.16). Note that this condition can only be fulfilled if the map T is surjective.

Definition 3.2.15. An incidence matrix A is called *primitive* if there exists some $p \in \mathbb{N}$ such that A^p has only positive entries, which is usually written as $A^p > 0$.

Note that every primitive matrix is irreducible, but that there exist irreducible matrices which are not primitive (see Exercise 3.4.13).

Theorem 3.2.16. *The shift map $\sigma : E_A^\infty \to E_A^\infty$ is topologically exact if and only if the matrix A is primitive.*

Proof. Suppose first that σ is topologically exact. Then for each $e \in E$ there exists some $p_e \in \mathbb{N}$ such that $\sigma^{p_e}([e]) = E_A^\infty$, since $[e]$ is an open set. Define $p := \max\{p_e : e \in E\}$ and note that p is finite. We claim that $A^p > 0$. To see this, fix an ordered pair $i, j \in E$. Since $\sigma^p([i]) = E_A^\infty$, we have $\sigma^p([i]) \cap [j] \neq \emptyset$. So if $\omega \in [i]$ and $\sigma^p(\omega) \in [j]$, then $\omega = i\omega_2\omega_3\ldots\omega_p j\omega_{p+2}\omega_{p+3}\ldots$. In other words, $i\omega_2\omega_3\ldots\omega_p j$ is an admissible word of length $p+1$ from i to j, and hence $(A^p)_{ij} > 0$. Since this is true for all $i, j \in E$, we conclude that $A^p > 0$.

To prove the converse, suppose that A is primitive, that is, suppose there exists $p \in \mathbb{N}$ such that $A^p > 0$. Then for each ordered pair $i, j \in E$ there is at least one admissible word of length $p+1$ from i to j. Since the cylinder sets $\{[\omega]\}_{\omega \in E_A^*}$ form a base for the topology on E_A^∞, it is sufficient to prove topological exactness for cylinder sets. So let $\omega = \omega_1\ldots\omega_n \in E_A^*$ and pick an arbitrary $\tau \in E_A^\infty$. There exists a finite word γ of length $p-1$ such that $\omega_n\gamma\tau_1$ is an admissible word of length $p+1$. Let $\overline{\omega} = \omega\gamma\tau \in E_A^\infty$. Then $\overline{\omega} \in [\omega]$ and $\sigma^{n+p-1}(\overline{\omega}) = \tau$. Therefore, as τ was arbitrarily chosen in E_A^∞, we deduce that $\sigma^{n+p-1}([\omega]) = E_A^\infty$ and σ is topologically exact. □

The reader may be wondering under which condition on A the shift $\sigma : E_A^\infty \to E_A^\infty$ is minimal. The answer to this question can be found in Exercise 3.4.15.

3.2.3 Asymptotic behavior of periodic points

We now prove that for any matrix A, the maximal growth rate of the number of periodic points in E_A^∞ coincides with the logarithm of the spectral radius of A. Recall that the spectral radius of A is defined to be the largest eigenvalue of A (in absolute value). The *spectral radius* $r(A)$ can also be defined by

$$r(A) := \lim_{n\to\infty} \|A^n\|^{1/n}, \tag{3.1}$$

for any matrix norm $\|\cdot\|$ (for a proof of this fact, see Proposition 3.8 in Conway [15]). In what follows, it is convenient to choose the norm to be the sum of the absolute value of the entries of the matrix, that is, for a $k \times k$ matrix B, the norm is

$$\|B\| := \sum_{i,j=1}^{k} |B_{ij}|.$$

Before continuing with the growth rate of the number of periodic points, we first give a lemma which will turn out to be useful over and over again. Although it is a purely analytic result, we include its proof here for completeness. This result will be crucial not only here but also in Chapters 7 and 11. Recall that a sequence $(a_n)_{n=1}^\infty$ of real numbers is said to be *subadditive* if

$$a_{m+n} \le a_m + a_n, \quad \forall m, n \in \mathbb{N}.$$

Lemma 3.2.17. *If $(a_n)_{n=1}^\infty$ is a subadditive sequence of real numbers, then the sequence $(a_n/n)_{n=1}^\infty$ converges and*

$$\lim_{n\to\infty} \frac{1}{n} a_n = \inf_{n\in\mathbb{N}} \frac{1}{n} a_n.$$

If, moreover, $(a_n)_{n=1}^\infty$ is bounded from below, then $\inf_{n\in\mathbb{N}} \frac{1}{n} a_n \ge 0$.

Proof. Fix $m \in \mathbb{N}$. By the division algorithm, every $n \in \mathbb{N}$ can be uniquely written in the form $n = km + r$, where $0 \le r < m$. The subadditivity of the sequence implies that

$$\frac{a_n}{n} = \frac{a_{km+r}}{km+r} \le \frac{a_{km} + a_r}{km+r} \le \frac{ka_m + a_r}{km} = \frac{a_m}{m} + \frac{a_r}{km}.$$

Notice that for all $n \in \mathbb{N}$,

$$-\infty < \min_{0\le s<m} a_s \le a_r \le \max_{0\le s<m} a_s < \infty.$$

Therefore, as n tends to infinity, k also tends to infinity and thereby a_r/k approaches zero by the sandwich theorem. Hence,

$$\limsup_{n\to\infty} \frac{a_n}{n} \le \frac{a_m}{m}.$$

Since $m \in \mathbb{N}$ was chosen arbitrarily, taking the infimum over m yields that

$$\limsup_{n\to\infty} \frac{a_n}{n} \le \inf_{m\in\mathbb{N}} \frac{a_m}{m}.$$

Thus,

$$\limsup_{n\to\infty} \frac{a_n}{n} \le \inf_{m\in\mathbb{N}} \frac{a_m}{m} \le \liminf_{n\to\infty} \frac{a_n}{n} \le \limsup_{n\to\infty} \frac{a_n}{n}.$$

Consequently,

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{m\in\mathbb{N}} \frac{a_m}{m}.$$

This proves the first assertion. The second one is obvious. □

Another purely analytic result that will be needed later is the following.

Lemma 3.2.18. *Let $(b_n)_{n=1}^{\infty}$ be a sequence of positive real numbers. Then*

$$\limsup_{n\to\infty} \frac{1}{n} \log b_n = \inf\left\{p \in \mathbb{R} : \sum_{n=1}^{\infty} b_n e^{-pn} < \infty\right\}.$$

Proof. First, assume that $-\infty < \limsup_{n\to\infty} \frac{1}{n} \log b_n$. Let $P \in \mathbb{R}$ be such that $P < \limsup_{n\to\infty} \frac{1}{n} \log b_n$. Then there exists a strictly increasing sequence $(n_m)_{m=1}^{\infty}$ of positive integers such that $P \le \frac{1}{n_m} \log b_{n_m}$ for all $m \in \mathbb{N}$. That is, $b_{n_m} \ge e^{Pn_m}$ for all $m \in \mathbb{N}$. Consequently,

$$\sum_{n=1}^{\infty} b_n e^{-Pn} \ge \sum_{m=1}^{\infty} b_{n_m} e^{-Pn_m} \ge \sum_{m=1}^{\infty} 1 = \infty.$$

This implies that $P \le \inf\{p \in \mathbb{R} : \sum_{n=1}^{\infty} b_n e^{-pn} < \infty\}$. Since this is true for every $P < \limsup_{n\to\infty} \frac{1}{n} \log b_n$, we deduce that

$$\limsup_{n\to\infty} \frac{1}{n} \log b_n \le \inf\left\{p \in \mathbb{R} : \sum_{n=1}^{\infty} b_n e^{-pn} < \infty\right\}. \tag{3.2}$$

Obviously, this latter inequality holds as well when $-\infty = \limsup_{n\to\infty} \frac{1}{n} \log b_n$.

To prove the opposite inequality, suppose that $\limsup_{n\to\infty} \frac{1}{n} \log b_n < \infty$. Let $P \in \mathbb{R}$ be such that $\limsup_{n\to\infty} \frac{1}{n} \log b_n < P$. Let $Q \in \mathbb{R}$ be such that

$$\limsup_{n\to\infty} \frac{1}{n} \log b_n < Q < P.$$

Then there exists $N \in \mathbb{N}$ such that $\frac{1}{n}\log b_n \le Q$ for all $n \ge N$. That is, $b_n \le e^{Qn}$ for all $n \ge N$. It follows that

$$\begin{aligned}\sum_{n=1}^{\infty} b_n e^{-Pn} &= \sum_{n=1}^{N-1} b_n e^{-Pn} + \sum_{n=N}^{\infty} b_n e^{-Pn}\\ &\le \sum_{n=1}^{N-1} b_n e^{-Pn} + \sum_{n=1}^{\infty} e^{(Q-P)n}\\ &= \sum_{n=1}^{N-1} b_n e^{-Pn} + \sum_{n=1}^{\infty} (e^{Q-P})^n\\ &< \infty\end{aligned}$$

since the geometric series on the right has for ratio $0 < r := e^{Q-P} < 1$. This implies that $\inf\{p \in \mathbb{R} : \sum_{n=1}^{\infty} b_n e^{-pn} < \infty\} \le P$. Since this is true for every $P \in \mathbb{R}$ such that $\limsup_{n\to\infty} \frac{1}{n}\log b_n < P$, we deduce that

$$\inf\left\{p \in \mathbb{R} : \sum_{n=1}^{\infty} b_n e^{-pn} < \infty\right\} \le \limsup_{n\to\infty} \frac{1}{n}\log b_n. \tag{3.3}$$

Obviously, this latter inequality holds as well when $\limsup_{n\to\infty} \frac{1}{n}\log b_n = \infty$.

The result follows from (3.2) and (3.3). □

Let us now come back to the question of the number of periodic points that subshifts of finite type have. The following result holds for all subshifts of finite type. It relates the number of periodic points to the number of finite words, which in turn is related to the underlying matrix. Recall that a sequence $(a_n)_{n=1}^{\infty}$ of real numbers is said to be *submultiplicative* if

$$a_{m+n} \le a_m a_n, \quad \forall m, n \in \mathbb{N}.$$

Note that a sequence $(a_n)_{n=1}^{\infty}$ of positive real numbers is submultiplicative if and only if the sequence $(\log a_n)_{n=1}^{\infty}$ is subadditive.

Lemma 3.2.19. *Let A be an incidence matrix that generates a subshift of finite type $\sigma : E_A^{\infty} \to E_A^{\infty}$. The sequences $(\#E_A^n)_{n=1}^{\infty}$ and $(\|A^n\|)_{n=1}^{\infty}$ are nondecreasing and submultiplicative. Moreover,*

$$\|A^{n-1}\| = \#E_A^n \ge \#\mathrm{Per}_n(\sigma), \quad \forall n \in \mathbb{N}.$$

Proof. Since the matrix A has a 1 in every row by our standing assumption, every admissible word of length n can be extended to an admissible word of length $n+1$. Therefore,

$$\#E_A^n \le \#E_A^{n+1}, \quad \forall n \in \mathbb{N},$$

that is, the sequence $(\#E_A^n)_{n=1}^{\infty}$ is nondecreasing.

Regarding the submultiplicativity of that sequence, notice that each admissible word of length $m+n$ results from an admissible concatenation of an admissible word of length m with an admissible word of length n. This implies that

$$\#E_A^{m+n} \le \#E_A^m \cdot \#E_A^n, \quad \forall m, n \in \mathbb{N}.$$

(However, notice that the concatenation of an admissible word ω of length m with an admissible word τ of length n is an admissible word of length $m+n$ if and only if $\omega_m = \tau_1$. Thus, the equality does not hold in general.)

The nondecreasing and submultiplicative behaviors of the sequence $(\#E_A^n)_{n=1}^{\infty}$ are shared by the sequence $(\|A^n\|)_{n=1}^{\infty}$. Indeed, we have earlier observed that $(A^n)_{ij}$ is the number of admissible words of length $n+1$ starting with the letter i and ending with the letter j. Thus, $\sum_{i,j=1}^{\#E}(A^n)_{ij}$ is the number of words in E_A^{n+1}. This means that

$$\|A^n\| = \sum_{i,j=1}^{\#E} (A^n)_{ij} = \#E_A^{n+1}, \quad \forall n \in \mathbb{N}.$$

It is then obvious that

$$\|A^n\| = \#E_A^{n+1} \le \#E_A^{n+2} = \|A^{n+1}\|, \quad \forall n \in \mathbb{N},$$

that is, the sequence $(\|A^n\|)_{n=1}^{\infty}$ is nondecreasing. Submultiplicativity of that sequence follows from the fact that for all $m, n \in \mathbb{N}$,

$$\|A^{m+n}\| = \#E_A^{m+n+1} \le \#E_A^{m+n+2} \le \#E_A^{m+1} \cdot \#E_A^{n+1} = \|A^m\| \cdot \|A^n\|.$$

Moreover, note that every periodic point of period n is the infinitely repeated concatenation of its initial block of length n. This means that

$$\#\mathrm{Per}_n(\sigma) \le \#E_A^n = \|A^{n-1}\|, \quad \forall n \in \mathbb{N}. \qquad \square$$

Remark 3.2.20. In general, the sequence $(\#\mathrm{Per}_n(\sigma))_{n=1}^{\infty}$ is neither nondecreasing nor submultiplicative. See Exercise 3.4.20.

We can now obtain some information on the growth rate of the number of periodic points.

Theorem 3.2.21. *Let A be an incidence matrix that generates a subshift of finite type $\sigma : E_A^{\infty} \to E_A^{\infty}$. Then*

$$\begin{aligned}\log r(A) = \lim_{n\to\infty} \frac{1}{n} \log \#E_A^n &= \inf\Big\{p > 0 : \sum_{\omega \in E_A^*} e^{-p|\omega|} < \infty\Big\} \\ &\ge \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma) = \inf\Big\{p > 0 : \sum_{n=1}^{\infty} \#\mathrm{Per}_n(\sigma) \cdot e^{-pn} < \infty\Big\},\end{aligned}$$

where $r(A)$ is the spectral radius of A.

Proof. It follows from Lemma 3.2.19 that both of the sequences $(\log \#E_A^n)_{n=1}^\infty$ and $(\log \|A^n\|)_{n=1}^\infty$ are subadditive. By Lemma 3.2.17, we have that both limits $\lim_{n\to\infty} \frac{1}{n} \log \|A^n\|$ and $\lim_{n\to\infty} \frac{1}{n} \log \#E_A^n$ exist. Using Lemma 3.2.19, we deduce that

$$\begin{aligned}
\log r(A) &= \log \lim_{n\to\infty} \|A^n\|^{1/n} \\
&= \lim_{n\to\infty} \log \|A^n\|^{1/n} \\
&= \lim_{n\to\infty} \frac{1}{n} \log \|A^n\| \\
&= \lim_{n\to\infty} \frac{1}{n} \log \#E_A^{n+1} \\
&= \lim_{n\to\infty} \frac{n+1}{n} \cdot \frac{1}{n+1} \log \#E_A^{n+1} \\
&= \lim_{n\to\infty} \frac{n+1}{n} \cdot \lim_{n\to\infty} \frac{1}{n+1} \log \#E_A^{n+1} \\
&= \lim_{n\to\infty} \frac{1}{n} \log \#E_A^n \\
&\geq \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma).
\end{aligned}$$

From Lemma 3.2.18, it follows that

$$\limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma) = \inf\left\{p \in \mathbb{R} : \sum_{n=1}^\infty \#\mathrm{Per}_n(\sigma) \cdot e^{-pn} < \infty\right\}$$

and

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log \#E_A^n &= \inf\left\{p \in \mathbb{R} : \sum_{n=1}^\infty \#E_A^n e^{-pn} < \infty\right\} \\
&= \inf\left\{p \in \mathbb{R} : \sum_{n=1}^\infty \sum_{\omega\in E_A^n} e^{-p|\omega|} < \infty\right\} \\
&= \inf\left\{p \in \mathbb{R} : \sum_{\omega\in E_A^*} e^{-p|\omega|} < \infty\right\}.
\end{aligned}$$

Finally, note that the infima can be restricted to the positive real numbers p since the set E_A^* is infinite and the sets $\mathrm{Per}_n(\sigma)$ are nonempty for infinitely many n. □

The inequality in the statement of the previous theorem turns out to be an equality. We will demonstrate this in two steps. It is first simpler to prove it when the matrix A is irreducible.

Theorem 3.2.22. *Let A be an irreducible matrix. Then*

$$\begin{aligned}
\log r(A) &= \lim_{n\to\infty} \frac{1}{n} \log \#E_A^n = \inf\left\{p > 0 : \sum_{\omega\in E_A^*} e^{-p|\omega|} < \infty\right\} \\
&= \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma) = \inf\left\{p > 0 : \sum_{n=1}^\infty \#\mathrm{Per}_n(\sigma) \cdot e^{-pn} < \infty\right\}.
\end{aligned}$$

Proof. Since A is irreducible, there exists $p \geq 2$ such that for every $1 \leq i, j \leq \#E$ there is an admissible word of length at least 2 and at most p that begins with i and ends with j. Therefore, for any admissible word ω of length n there is an admissible word $\widetilde{\omega}$ of length at least 2 and at most p beginning with $\widetilde{\omega}_1 = \omega_n$ and ending with $\widetilde{\omega}_{|\widetilde{\omega}|} = \omega_1$. Consequently, the word $(\omega_1\omega_2 \ldots \omega_n\widetilde{\omega}_2\widetilde{\omega}_3 \ldots \widetilde{\omega}_{|\widetilde{\omega}|-1})^\infty$ is an admissible periodic point of period $n + |\widetilde{\omega}| - 2$, with $n \leq n + |\widetilde{\omega}| - 2 \leq n + p - 2$. This shows that every $\omega \in E_A^n$ generates at least one periodic point whose period is between n and $n + p - 2$, with different words $\omega \in E_A^n$ producing different periodic points $(\omega\widetilde{\omega}_2\widetilde{\omega}_3 \ldots \widetilde{\omega}_{|\widetilde{\omega}|-1})^\infty$. Thus, if for every $n \in \mathbb{N}$ we choose $m(n)$ to be such that $n \leq m(n) \leq n + p - 2$ and

$$\#\mathrm{Per}_{m(n)}(\sigma) = \max_{n \leq m \leq n+p-2} \#\mathrm{Per}_m(\sigma),$$

we obtain that

$$\begin{aligned}\|A^{n-1}\| = \#E_A^n &\leq \#\mathrm{Per}_n(\sigma) + \#\mathrm{Per}_{n+1}(\sigma) + \cdots + \#\mathrm{Per}_{n+p-2}(\sigma)\\ &\leq (p-1)\#\mathrm{Per}_{m(n)}(\sigma).\end{aligned}$$

Using this estimate, we can make the following calculation:

$$\begin{aligned}\log r(A) &= \lim_{n\to\infty} \frac{1}{n-1} \log\|A^{n-1}\| \leq \limsup_{n\to\infty} \frac{1}{n-1} \log\big[(p-1) \cdot \#\mathrm{Per}_{m(n)}(\sigma)\big]\\ &= \limsup_{n\to\infty} \frac{1}{n-1}\big[\log(p-1) + \log \#\mathrm{Per}_{m(n)}(\sigma)\big]\\ &= \lim_{n\to\infty} \frac{1}{n-1} \log(p-1) + \limsup_{n\to\infty} \frac{1}{n-1} \log \#\mathrm{Per}_{m(n)}(\sigma)\\ &= 0 + \limsup_{n\to\infty}\left[\frac{m(n)}{n-1} \cdot \frac{1}{m(n)} \log \#\mathrm{Per}_{m(n)}(\sigma)\right]\\ &= \lim_{n\to\infty} \frac{m(n)}{n-1} \cdot \limsup_{n\to\infty} \frac{1}{m(n)} \log \#\mathrm{Per}_{m(n)}(\sigma)\\ &\leq 1 \cdot \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma)\\ &= \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma).\end{aligned}$$

Combining this inequality with the opposite one proved in Theorem 3.2.21 completes the proof. □

In order to prove that the equality holds in full generality, we need to decompose reducible systems into maximal (in the sense of inclusion) irreducible subsystems. For this, we separate the letters of E into equivalence classes by means of a relation called *communication*.

Definition 3.2.23.

(a) A letter $i \in E$ *leads to* a letter $j \in E$ if there exists $p = p(i,j) \in \mathbb{N}$ such that $(A^p)_{ij} = 1$.

(b) A letter i is said to *communicate* with a letter j if i leads to j and j leads to i.

(c) A letter which communicates with itself or any other letter is called *communicating*.
(d) Otherwise, the letter is said to be *noncommunicating*.

The relation of communication defines an equivalence relation on the set of communicating letters. The corresponding equivalence classes are called *communication classes*.

Theorem 3.2.24. *Let A be an incidence matrix that generates a subshift of finite type $\sigma : E_A^\infty \to E_A^\infty$. Then*

$$\begin{aligned}\log r(A) &= \lim_{n\to\infty} \frac{1}{n} \log \# E_A^n = \inf\Big\{p > 0 : \sum_{\omega\in E_A^*} e^{-p|\omega|} < \infty\Big\}\\ &= \max_{\substack{C\\ \textit{comm. class}}} \lim_{n\to\infty} \frac{1}{n} \log \# C_A^n\\ &= \max_{\substack{C\\ \textit{comm. class}}} \limsup_{n\to\infty} \frac{1}{n} \log \# \mathrm{Per}_n(\sigma_C)\\ &= \limsup_{n\to\infty} \frac{1}{n} \log \# \mathrm{Per}_n(\sigma) = \inf\Big\{p > 0 : \sum_{n=1}^{\infty} \# \mathrm{Per}_n(\sigma)\cdot e^{-pn} < \infty\Big\}\\ &= \max_{\substack{C\\ \textit{comm. class}}} \log r(A|_C).\end{aligned}$$

Proof. Suppose that E admits k communication classes $C_1, C_2, \ldots, C_k$. Clearly, $(C_l)_A^\infty \subseteq E_A^\infty$ for each $1 \le l \le k$. Moreover, $(C_l)_A^\infty \cap (C_m)_A^\infty = \emptyset$ for all $l \ne m$ since $C_l \cap C_m = \emptyset$ for all $l \ne m$. Note further that the submatrix $A|_{C_l} : C_l \times C_l \to \{0,1\}$ is irreducible for each $1 \le l \le k$ by the very definition of communication classes. Therefore, Theorem 3.2.22 asserts that

$$\begin{aligned}\log r(A|_{C_l}) &= \lim_{n\to\infty} \frac{1}{n} \log \#(C_l)_A^n = \inf\Big\{p > 0 : \sum_{\omega\in (C_l)_A^*} e^{-p|\omega|} < \infty\Big\}\\ &= \limsup_{n\to\infty} \frac{1}{n} \log \# \mathrm{Per}_n(\sigma_{C_l}) \qquad (3.4)\end{aligned}$$

for every $1 \le l \le k$.

Let $I = E \setminus \bigcup_{l=1}^k C_l$. This set consists of all noncommunicating letters. If $\omega \in E_A^*$, then each noncommunicating letter can appear at most once in ω. Moreover, if ω contains at least one letter from the class C_l then ω can be uniquely written as $\beta_l\alpha_l\gamma_l$, where $\alpha_l \in (C_l)_A^*$ and $\beta_l, \gamma_l \in (E \setminus C_l)_A^*$, that is, α_l comprises only letters from the class C_l while none of the letters of β_l and γ_l are from C_l. Note that β_l and/or γ_l may be the empty word, while α_l is the longest subword of ω that has letters from C_l only. For each $\omega \in E_A^*$ and $1 \le l \le k$, let $\alpha_l(\omega)$ be the longest subword of ω in $(C_l)_A^*$. Note that $\alpha_l(\omega)$ may be the empty word for some l's. Then ω can be uniquely written as a concatenation of the subwords $\alpha_l(\omega)$, $1 \le l \le k$, and no more than $k+1$ subwords of noncommunicating letters, each of which consists of at most $\#I$ letters. Therefore, the map

$$\begin{array}{cccccccccccc} \alpha & : & E_A^* & \longrightarrow & (C_1)_A^* & \times & (C_2)_A^* & \times & \cdots & \times & (C_k)_A^* \\ & & \omega & \longmapsto & \big(\alpha_1(\omega) & , & \alpha_2(\omega) & , & \ldots & , & \alpha_k(\omega)\big) \end{array}$$

is such that each element of $(C_1)_A^* \times \cdots \times (C_k)_A^*$ has at most $(\#I \cdot \#I! + 1)^{k+1}$ preimages. Indeed, suppose that $\alpha(\tau) = \alpha(\omega)$ for some $\tau, \omega \in E_A^*$. Then $\alpha_l(\tau) = \alpha_l(\omega) =: \alpha_l$ for all $1 \le l \le k$. If $\alpha_l \neq \epsilon \neq \alpha_m$ for some $l \neq m$ and if τ contained the subword $\alpha_l \beta \alpha_m$ whereas ω contained the subword $\alpha_m \gamma \alpha_l$, then the word $\alpha_l \beta \alpha_m \gamma \alpha_l$ would be in E_A^*. This would imply that the classes C_l and C_m communicate, which would contradict their very definition. This reveals that the α_l, $1 \le l \le k$, must appear in the same order in both τ and ω. That is, τ and ω can only differ in the subwords of noncommunicating letters they contain. Now, there are at most $k+1$ subwords of noncommunicating letters in any word. And each of these subwords contains at most $\#I$ letters. Let $1 \le L \le \#I$. The number of words of length L with distinct letters drawn from I is at most $\#I \cdot (\#I - 1) \cdots (\#I - L + 1) \le \#I!$. Thus the number of nonempty words of length at most $\#I$ with distinct letters drawn from I is at most $\#I \cdot \#I!$. Add 1 for the empty word. This is an upper estimate of the number of possibilities for each instance of a subword of noncommunicating letters. Since there are at most $k+1$ such instances, a crude upper bound on the number of preimages for any point is $B := (\#I \cdot \#I! + 1)^{k+1}$.

For all $p > 0$, it then follows that

$$\begin{aligned} \sum_{\omega \in E_A^*} e^{-p|\omega|} &\le \sum_{\omega \in E_A^*} e^{-p \sum_{l=1}^{k} |\alpha_l(\omega)|} \\ &= \sum_{\omega \in E_A^*} \prod_{l=1}^{k} e^{-p|\alpha_l(\omega)|} \\ &\le B \prod_{l=1}^{k} \sum_{\omega_l \in (C_l)_A^*} e^{-p|\omega_l|}. \end{aligned} \tag{3.5}$$

Let

$$P > \max_{1 \le l \le k} \lim_{n \to \infty} \frac{1}{n} \log \#(C_l)_A^n.$$

We infer from (3.4) that

$$\sum_{\omega_l \in (C_l)_A^*} e^{-P|\omega_l|} < \infty, \quad \forall 1 \le l \le k$$

and thus by (3.5) we deduce that

$$\sum_{\omega \in E_A^*} e^{-P|\omega|} < \infty.$$

According to Theorem 3.2.21, this implies that

$$P > \lim_{n\to\infty} \frac{1}{n} \log \#E_A^n.$$

Since this is true for every $P > \max_{1\le l\le k} \lim_{n\to\infty} \frac{1}{n} \log \#(C_l)_A^n$, we obtain that

$$\lim_{n\to\infty} \frac{1}{n} \log \#E_A^n \le \max_{1\le l\le k} \lim_{n\to\infty} \frac{1}{n} \log \#(C_l)_A^n.$$

The opposite inequality is obvious. Hence,

$$\lim_{n\to\infty} \frac{1}{n} \log \#E_A^n = \max_{1\le l\le k} \lim_{n\to\infty} \frac{1}{n} \log \#(C_l)_A^n. \tag{3.6}$$

On the other hand, note that $\mathrm{Per}(\sigma_{C_l}) \cap \mathrm{Per}(\sigma_{C_m}) = \emptyset$ for all $l \ne m$ since $C_l \cap C_m = \emptyset$ for all $l \ne m$. Moreover, since a periodic point can comprise neither noncommunicating letters nor letters from two distinct communicating classes, we have that

$$\mathrm{Per}(\sigma_E) = \bigcup_{l=1}^{k} \mathrm{Per}(\sigma_{C_l}).$$

Therefore,

$$\max_{1\le l\le k} \#\mathrm{Per}_n(\sigma_{C_l}) \le \#\mathrm{Per}_n(\sigma_E) = \sum_{l=1}^{k} \#\mathrm{Per}_n(\sigma_{C_l}) \le k \max_{1\le l\le k} \#\mathrm{Per}_n(\sigma_{C_l}).$$

It follows immediately that

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log \max_{1\le l\le k} \#\mathrm{Per}_n(\sigma_{C_l}) &\le \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma_E)\\ &\le \limsup_{n\to\infty} \frac{1}{n} \log\Big[k \max_{1\le l\le k} \#\mathrm{Per}_n(\sigma_{C_l})\Big].\end{aligned}$$

Using Exercise 3.4.14, it follows that

$$\begin{aligned}\max_{1\le l\le k} \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma_{C_l}) &\le \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma_E)\\ &\le \lim_{n\to\infty} \frac{\log k}{n} + \max_{1\le l\le k} \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma_{C_l}).\end{aligned}$$

Hence,

$$\limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma_E) = \max_{1\le l\le k} \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma_{C_l}). \tag{3.7}$$

Using Theorem 3.2.21 and relations (3.4), (3.6), and (3.7), the result follows. □

The use of the lim sup in Theorem 3.2.22 is indispensable. Indeed, there are transitive subshifts of finite type for which the limit does not exist (see Exercise 3.4.20). However, the limit does exist for all topologically exact subshifts of finite type.

Theorem 3.2.25. *Let A be a primitive matrix. Then*

$$\begin{aligned}\log r(A) &= \lim_{n\to\infty}\frac{1}{n}\log\#E_A^n = \inf\Big\{p>0:\sum_{\omega\in E_A^*}e^{-p|\omega|}<\infty\Big\}\\&= \lim_{n\to\infty}\frac{1}{n}\log\#\mathrm{Per}_n(\sigma) = \inf\Big\{p>0:\sum_{n=1}^{\infty}\#\mathrm{Per}_n(\sigma)\cdot e^{-pn}<\infty\Big\}.\end{aligned}$$

Proof. In Lemma 3.2.19, we observed that $\#\mathrm{Per}_n(\sigma) \le \#E_A^n = \|A^{n-1}\|$ for all $n\in\mathbb{N}$. Now we use the primitivity of the matrix to establish a similar inequality in the other direction. Since A is primitive, there exists $P\in\mathbb{N}$ such that $A^P>0$. Consequently, for every $1\le i,j\le \#E$ there is an admissible word of length $P+1$ that begins with i and ends with j. Therefore, for any $\omega\in E_A^n$ there is a word $\widetilde{\omega}\in E_A^{P+1}$ which begins with $\widetilde{\omega}_1=\omega_n$ and ends with $\widetilde{\omega}_{P+1}=\omega_1$. Then the word $(\omega_1\omega_2\ldots\omega_n\widetilde{\omega}_2\widetilde{\omega}_3\ldots\widetilde{\omega}_P)^\infty$ is an admissible periodic point of period $n+P-1$. This shows that every $\omega\in E_A^n$ generates at least one periodic point of period $n+P-1$, with different words $\omega\in E_A^n$ producing different periodic points $(\omega\widetilde{\omega}_2\widetilde{\omega}_3\ldots\widetilde{\omega}_P)^\infty\in\mathrm{Per}_{n+P-1}(\sigma)$. Hence, $\#E_A^n\le\#\mathrm{Per}_{n+P-1}(\sigma)$. Using this and Lemma 3.2.19, we get

$$\#E_A^n\le\#\mathrm{Per}_{n+P-1}(\sigma)\le\#E_A^{n+P-1}\le\#E_A^n\cdot\#E_A^{P-1},\quad\forall n\in\mathbb{N}.$$

Hence,

$$\frac{1}{n}\log\#E_A^n\le\frac{1}{n}\log\#\mathrm{Per}_{n+P-1}(\sigma)\le\frac{1}{n}\log\#E_A^n+\frac{1}{n}\log\#E_A^{P-1},\quad\forall n\in\mathbb{N}.$$

It follows from the squeeze theorem that

$$\lim_{n\to\infty}\frac{1}{n}\log\#\mathrm{Per}_{n+P-1}(\sigma)=\lim_{n\to\infty}\frac{1}{n}\log\#E_A^n.$$

Therefore,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\log\#\mathrm{Per}_n(\sigma) &= \lim_{n\to\infty}\frac{1}{n+P-1}\log\#\mathrm{Per}_{n+P-1}(\sigma)\\&= \lim_{n\to\infty}\Big[\frac{n}{n+P-1}\cdot\frac{1}{n}\log\#\mathrm{Per}_{n+P-1}(\sigma)\Big]\\&= \lim_{n\to\infty}\frac{n}{n+P-1}\cdot\lim_{n\to\infty}\frac{1}{n}\log\#\mathrm{Per}_{n+P-1}(\sigma)\\&= \lim_{n\to\infty}\frac{1}{n}\log\#E_A^n.\end{aligned}$$

The rest follows from Theorem 3.2.22. □

3.3 General subshifts of finite type

Recall that a subshift of finite type is a subshift that can be described by a finite set $\mathcal{F}$ of forbidden words. In this case, the set $\mathcal{F}$ can be chosen so that $\mathcal{F} \subseteq E^q$ for some $q \in \mathbb{N}$. The set of forbidden words then induces a function A from E^q to $\{0,1\}$, where the function A takes the value 0 on the set $\mathcal{F}$ of forbidden words and takes the value 1 on the set $E^q \setminus \mathcal{F}$ of all admissible words. For the sake of simplicity, we concentrated on the case $q = 2$ in the previous section. We then pointed out that we would prove that all cases can be reduced to that case. We now do so.

Fix an integer $q \geq 2$ and a function $A : E^q \to \{0,1\}$. Let

$$E_A^\infty := \{\omega \in E^\infty : A(\omega_n, \omega_{n+1}, \ldots, \omega_{n+q-1}) = 1, \forall n \in \mathbb{N}\}.$$

E_A^∞ is a subshift of finite type since it consists of all those infinite words that do not contain any word from the finite set of forbidden words

$$\mathcal{F} = \{\omega \in E^q : A(\omega_1, \omega_2, \ldots, \omega_q) = 0\}.$$

Theorem 3.3.1. *The shift map $\sigma : E_A^\infty \to E_A^\infty$ is topologically conjugate to a shift map $\widetilde{\sigma} : \widetilde{E}_{\widetilde{A}}^\infty \to \widetilde{E}_{\widetilde{A}}^\infty$, where $\#\widetilde{E} = (\#E)^q$ and $\widetilde{A}$ is an incidence matrix on $\widetilde{E}$.*

Proof. Set $\widetilde{E} := E^q$ as an alphabet. In other words, a letter in the alphabet $\widetilde{E}$ is a word of length q over the alphabet E. Define the incidence matrix $\widetilde{A} : \widetilde{E} \times \widetilde{E} \to \{0,1\}$ by

$$\widetilde{A}_{\tau\rho} = \begin{cases} 1 & \text{if } A(\tau) = 1 = A(\rho) \text{ and } \tau_2\tau_3 \ldots \tau_q = \rho_1\rho_2 \ldots \rho_{q-1} \\ 0 & \text{otherwise} \end{cases}$$

for every $\tau, \rho \in \widetilde{E}$. Let

$$\widetilde{E}_{\widetilde{A}}^\infty := \{\widetilde{\omega} \in \widetilde{E}^\infty : \widetilde{A}_{\widetilde{\omega}_n \widetilde{\omega}_{n+1}} = 1,\ \forall n \in \mathbb{N}\}$$

be the subshift of finite type generated by the matrix $\widetilde{A}$. Define the map $H : E_A^\infty \to \widetilde{E}$ by $H(\omega) = \omega|_q$. That is, the map H associates to every A-admissible infinite word its initial subword of length q. Let $h : E_A^\infty \to \widetilde{E}_{\widetilde{A}}^\infty$ be defined by the concatenation

$$h(\omega) = H(\omega) H(\sigma(\omega)) H(\sigma^2(\omega)) \ldots .$$

In other words, for any $\omega \in E_A^\infty$, we have that

$$h(\omega) = (\omega_1 \ldots \omega_q)(\omega_2 \ldots \omega_{q+1})(\omega_3 \ldots \omega_{q+2}) \ldots \in \widetilde{E}_{\widetilde{A}}^\infty.$$

We claim that h is a homeomorphism. Indeed, h is injective since if ω and τ are two distinct elements of E_A^∞, then $\omega_k \neq \tau_k$ for some $k \in \mathbb{N}$. It immediately follows that $\sigma^{k-1}(\omega)_1 = \omega_k \neq \tau_k = \sigma^{k-1}(\tau)_1$, and hence, by definition, $H(\sigma^{k-1}(\omega)) \neq H(\sigma^{k-1}(\tau))$. Thus, $h(\omega)_k \neq h(\tau)_k$ and so $h(\omega) \neq h(\tau)$. To show that the map h is surjective, let

$\widetilde{\tau} \in \widetilde{E}_{\widetilde{A}}^{\infty}$ be arbitrary. Recall that $\widetilde{\tau}_k \in \widetilde{E} = E^q$ for each $k \in \mathbb{N}$, that is, $\widetilde{\tau}_k$ is a word of length q from the alphabet E. Also bear in mind that by the definition of $\widetilde{E}_{\widetilde{A}}^{\infty}$, the word consisting of the last $q-1$ letters of $\widetilde{\tau}_k$ is equal to the initial $(q-1)$-word of $\widetilde{\tau}_{k+1}$. Construct the infinite word τ by concatenating the first letters of each word $\widetilde{\tau}_k$ in turn, that is,

$$\tau = (\widetilde{\tau}_1)_1(\widetilde{\tau}_2)_1(\widetilde{\tau}_3)_1 \dots .$$

Then, for every $n \in \mathbb{N}$, we have that

$$\begin{aligned} A(\tau_n, \tau_{n+1}, \dots, \tau_{n+q-1}) &= A\big((\widetilde{\tau}_n)_1, (\widetilde{\tau}_{n+1})_1, \dots, (\widetilde{\tau}_{n+q-1})_1\big) \\ &= A\big((\widetilde{\tau}_n)_1, (\widetilde{\tau}_n)_2, \dots, (\widetilde{\tau}_n)_q\big) \\ &= A(\widetilde{\tau}_n) \\ &= 1. \end{aligned}$$

Therefore, $\tau \in E_A^{\infty}$. Furthermore,

$$\begin{aligned} h(\tau) &= H(\tau)H\big(\sigma(\tau)\big)H\big(\sigma^2(\tau)\big)\dots \\ &= \big((\widetilde{\tau}_1)_1(\widetilde{\tau}_2)_1\dots(\widetilde{\tau}_q)_1\big)\big((\widetilde{\tau}_2)_1(\widetilde{\tau}_3)_1\dots(\widetilde{\tau}_{q+1})_1\big)\big((\widetilde{\tau}_3)_1(\widetilde{\tau}_4)_1\dots(\widetilde{\tau}_{q+2})_1\big)\dots \\ &= \big((\widetilde{\tau}_1)_1(\widetilde{\tau}_1)_2\dots(\widetilde{\tau}_1)_q\big)\big((\widetilde{\tau}_2)_1(\widetilde{\tau}_2)_2\dots(\widetilde{\tau}_2)_q\big)\big((\widetilde{\tau}_3)_1(\widetilde{\tau}_3)_2\dots(\widetilde{\tau}_3)_q\big)\dots \\ &= (\widetilde{\tau}_1)(\widetilde{\tau}_2)(\widetilde{\tau}_3)\dots \\ &= \widetilde{\tau}. \end{aligned}$$

Since $\widetilde{\tau}$ is arbitrary, this demonstrates that h is surjective.

Moreover, h is continuous. To see this, let $\omega, \tau \in E_A^{\infty}$. Denote the length of their wedge by $Q = |\omega \wedge \tau|$. If $Q \geq q$, then $|h(\omega) \wedge h(\tau)| = Q - q + 1$, for if ω and τ share the same first Q letters, then $h(\omega)$ and $h(\tau)$ share the same first $Q - q + 1$ letters. The fact that h is a homeomorphism follows from the fact that it is a continuous bijection between two Hausdorff compact topological spaces. It only remains to show that the following diagram commutes:

$$\begin{tikzcd} E_A^{\infty} \arrow[r, "\sigma"] \arrow[d, "h"'] & E_A^{\infty} \arrow[d, "h"] \\ \widetilde{E}_{\widetilde{A}}^{\infty} \arrow[r, "\widetilde{\sigma}"'] & \widetilde{E}_{\widetilde{A}}^{\infty} \end{tikzcd}$$

Indeed, for every $\omega \in E_A^{\infty}$ we have

$$\begin{aligned} h \circ \sigma(\omega) &= H\big(\sigma(\omega)\big)H\big(\sigma(\sigma(\omega))\big)H\big(\sigma^2(\sigma(\omega))\big)\dots \\ &= H\big(\sigma(\omega)\big)H\big(\sigma^2(\omega)\big)H\big(\sigma^3(\omega)\big)\dots \\ &= \widetilde{\sigma}\big(H(\omega)H\big(\sigma(\omega)\big)H\big(\sigma^2(\omega)\big)\dots\big) \\ &= \widetilde{\sigma} \circ h(\omega). \end{aligned}$$

This completes the proof. $\square$

3.4 Exercises

Exercise 3.4.1. Fix $n \in \mathbb{N}$. Show that the full n-shift can be used to encode all the numbers between 0 and 1. That is, show that to every number in $[0, 1]$ can be associated an infinite word in $\{0, \ldots, n-1\}^\infty$.

Exercise 3.4.2. Prove that the discrete topology on a set X, that is, the topology in which every subset of X is both open and closed, is metrizable by means of the distance function $d : X \times X \to \{0, 1\}$ defined by

$$d(x_1, x_2) = \begin{cases} 1 & \text{if } x_1 \neq x_2 \\ 0 & \text{if } x_1 = x_2. \end{cases}$$

Then show that the product $\prod_{n=1}^{\infty} X$ is metrizable.

Exercise 3.4.3. Show that the family $\{[\omega] : \omega \in E^*\}$ of all initial cylinders forms a base of open sets for Tychonoff's product topology on E^∞. Deduce that all cylinders $\{[\omega]_m^n : \omega \in E^*, m, n \in \mathbb{N}\}$ are both open and closed sets. Deduce further that the space E^∞ is totally disconnected.

Exercise 3.4.4. Prove that the metrics d_s, $s \in (0, 1)$, introduced in Definition 3.1.10 induce Tychonoff's product topology on E^∞.

Exercise 3.4.5. Prove that the metrics d_s, for each $s \in (0, 1)$, are Hölder equivalent. That is, show that for any pair $s, s' \in (0, 1)$ there is an exponent $\alpha \geq 0$ and a constant $C \geq 1$ such that

$$C^{-1}\big(d_{s'}(\omega, \tau)\big)^{\alpha} \leq d_s(\omega, \tau) \leq C\big(d_{s'}(\omega, \tau)\big)^{\alpha}, \quad \forall \omega, \tau \in E^\infty.$$

Exercise 3.4.6. Show that the metrics d_s, $s \in (0, 1)$, are not Lipschitz equivalent. That is, prove that for any pair $s, s' \in (0, 1)$ there is no constant $C \geq 1$ such that

$$C^{-1} d_{s'}(\omega, \tau) \leq d_s(\omega, \tau) \leq C d_{s'}(\omega, \tau), \quad \forall \omega, \tau \in E^\infty.$$

Exercise 3.4.7. Prove directly that the space E^∞ is separable. That is, find a countable dense set in E^∞.

Exercise 3.4.8. Prove directly that E^∞ is compact.

Hint: Since E^∞ is metrizable, it suffices to prove that E^∞ is sequentially compact. That is, it is sufficient to prove that every sequence in E^∞ admits a convergent subsequence.

Exercise 3.4.9. Describe the subshift of finite type E_A^∞ generated by the incidence matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

This shift is called the *golden mean shift*. (We will explain why in Chapter 7.)

Exercise 3.4.10. Find a closed shift-invariant subset F of $\{0,1\}^\infty$ such that $\sigma|_F : F \to F$ is not open.

Exercise 3.4.11. Prove that the shift map $\sigma : E_A^\infty \to E_A^\infty$ is injective if and only if every column of the incidence matrix A contains at most one 1.

Exercise 3.4.12. Show that the shift map $\sigma : E_A^\infty \to E_A^\infty$ may be transitive even if the incidence matrix A contains a row of zeros. (Thus, Theorem 3.2.14 does not hold if one does not assume that every row of A contains at least one 1.)

Exercise 3.4.13. Construct an irreducible matrix which is not primitive.

Exercise 3.4.14. Let $(a_n)_{n=1}^\infty$ and $(b_n)_{n=1}^\infty$ be sequences of real numbers. Prove that

$$\limsup_{n\to\infty} \max\{a_n, b_n\} = \max\Big\{\limsup_{n\to\infty} a_n, \limsup_{n\to\infty} b_n\Big\}.$$

Show a similar result for lim inf. Also, show that max can be replaced by min and that the corresponding statements hold. Finally, show that the statements do not necessarily hold for lim.
Note: Though this exercise has been stated with two sequences only, these statements hold for any finite number of sequences.

Exercise 3.4.15. In this exercise, you will prove that the shift $\sigma : E_A^\infty \to E_A^\infty$ is minimal if and only if A is a permutation matrix with a unique class of communicating letters.

(a) Relying upon our standing assumption that A has a 1 in each of its rows, prove that E_A^∞ admits a periodic point. (In fact, it is possible to show that the set of eventually periodic points is dense in E_A^∞.)
(b) Suppose that $\sigma : E_A^\infty \to E_A^\infty$ is minimal. Deduce from the minimality of σ that E_A^∞ coincides with the orbit of a periodic point.
(c) Deduce that σ is a bijection.
(d) Deduce that A is a permutation matrix.
(e) Show that A has a unique communicating class.

To prove the converse, suppose that A is a permutation matrix with a unique class of communicating letters.

(f) Prove that E_A^∞ coincides with the orbit of a periodic point.
(g) Deduce that $\sigma : E_A^\infty \to E_A^\infty$ is minimal.

Exercise 3.4.16. Let C and C' be classes of communicating letters for a matrix A. Class C is said to *lead* to class C' if one of the letters in C leads to one of the letters in C'. Prove that the set of periodic points of E_A^∞ is dense if and only if A consists of classes of communicating letters, none of which leads to another (in particular, A does not have any noncommunicating letter).

In other words, the set of periodic points of E_A^∞ is dense if and only if E_A^∞ is a disjoint union of irreducible subsystems.

Exercise 3.4.17. Let C be the middle-third Cantor set. Show that the map $f : [0,1] \to [0,1]$ defined by

$$f(x) = 3x \pmod 1$$

restricted to C is continuous. Show that $f|_C : C \to C$ is topologically conjugate to the full shift map on two symbols.

Exercise 3.4.18. Show that if the shift map $\sigma : E_A^\infty \to E_A^\infty$ is such that $\sigma([e] \cap E_A^\infty) = E_A^\infty \neq \emptyset$ for every $e \in E$, then $E_A^\infty = E^\infty$.

Exercise 3.4.19. Show that the full shift map $\sigma : \{0,1\}^\infty \to \{0,1\}^\infty$ has uncountably many points that are not transitive (i. e., with a nondense orbit).

Exercise 3.4.20. Construct a transitive subshift of finite type whose periodic points have even periods.

4 Distance expanding maps

In this chapter, we first define and give some examples of distance expanding maps, which, as their name suggests, expand distances between points. On a compact metric space, this behavior may only be observed locally. Accordingly, the definition of an expanding map involves two constants: a constant describing the magnitude of the expansion of the system under scrutiny, and a constant delimiting the neighborhoods on which the expansion can be observed.

Distance expanding maps were introduced in [61]. A fairly complete account of them can be found in [57]. Our approach stems from that work, but in many instances is much more detailed. Moreover, the proof of the existence of Markov partitions in Section 4.4 is substantially simplified.

In Section 4.2, we introduce the notion and study the properties of inverse branches of a distance expanding map. This is a way of dealing with the noninvertibility of these maps.

In Section 4.3, we describe two new concepts: pseudo-orbit and shadowing. The latter makes precise the fact that, given a measuring device of some prescribed accuracy, sequences of points which remain sufficiently close to one another cannot be distinguished by the said device.

Sections 4.4 and 4.5 are crucial. In the former, we introduce the concept of Markov partitions and their existence for open, distance expanding systems, while in the latter we show exactly how to use them to represent the dynamics of such systems by means of the symbolic dynamics studied in Chapter 3. The final theorem of the chapter describes the properties of the coding map between the underlying compact metric space (the phase space) and some subshift of finite type (a symbolic space).

The concept of Markov partition was introduced to dynamical systems by Adler, Konheim, and McAndrew in the paper [2] in 1965. It achieved its full significance in Rufus Bowen's book [11]. It is in this book that the existence of Markov partitions was proved for Axiom A diffeomorphisms and the corresponding symbolic representation/dynamics along with thermodynamic formalism were developed. Our approach, via the book [57], traces back to Bowen's work. Markov partitions, in their various forms, play an enormous role in the modern (that is, after Bowen) theory of dynamical systems.

4.1 Definition and examples

Definition 4.1.1. A continuous map $T : X \to X$ of a compact metric space (X, d) is called *distance expanding* provided that there exist two constants $\lambda > 1$ and $\delta > 0$ such that

$$d(x,y) < 2\delta \Longrightarrow d(T(x),T(y)) \geq \lambda d(x,y).$$

https://doi.org/10.1515/9783110702682-004

The use of 2δ in the above definition, as opposed to simply δ, is only to make the forthcoming expressions and calculations simpler.

Remark 4.1.2.
(a) If $T : X \to X$ is a distance expanding map, then for every forward T-invariant closed set $F \subseteq X$ the map $T|_F : F \to F$ is also distance expanding.
(b) If $d(x,y) < 2\delta$, then $d(T(x),T(y)) \geq \lambda d(x,y)$. Therefore, if $x \neq y$, we have that $T(x) \neq T(y)$. Thus, if $T(x) = T(y)$ and $x \neq y$, then $d(x,y) \geq 2\delta$. In particular, this demonstrates that any distance expanding map is locally injective.

Example 4.1.3 (Subshifts of finite type). The full shift map $\sigma : E^\infty \to E^\infty$ and all of its subshifts of finite type $\sigma : E_A^\infty \to E_A^\infty$ are expanding whichever metric d_s is used. More precisely, σ is expanding with $\lambda = s^{-1}$ and any $0 < \delta \leq 1/2$. To see this, let $\omega, \tau \in E_A^\infty$ with $d_s(\omega,\tau) < 1$. This means that $|\omega \wedge \tau| \geq 1$ and, therefore, $|\sigma(\omega) \wedge \sigma(\tau)| = |\omega \wedge \tau| - 1$. So,

$$d_s\big(\sigma(\omega),\sigma(\tau)\big) = s^{|\sigma(\omega)\wedge\sigma(\tau)|} = s^{|\omega\wedge\tau|-1} = s^{-1}s^{|\omega\wedge\tau|} = s^{-1}d_s(\omega,\tau).$$

4.1.1 Expanding repellers

A large class of distance expanding maps are the expanding repellers, which we define and study in this subsection.

Definition 4.1.4. Let U be a nonempty open subset of $\mathbb{R}^d$ and $T : U \to \mathbb{R}^d$ a C^1 map (that is, T is continuously differentiable on U). Let X be a nonempty compact subset of U. The triple (X,U,T) is called an *expanding repeller* provided that the following conditions are satisfied:
(a) $T(X) = X$.
(b) There exists $\lambda > 1$ such that $\|T'(x)v\| \geq \lambda\|v\|$ for all $v \in \mathbb{R}^d$ and all $x \in X$.
(c) $\bigcap_{n=0}^\infty T^{-n}(U) = X$.

The set X is sometimes called the *limit set* of the repeller.

Recall that $T'(x) = D_xT : \mathbb{R}^d \to \mathbb{R}^d$ is, by definition, the unique bounded linear operator such that

$$\lim_{y\to x} \frac{\|T(y) - T(x) - D_xT(y-x)\|}{\|y-x\|} = 0,$$

where $\|v\| = \sum_{i=1}^d v_i^2$ is the standard Euclidean norm on $\mathbb{R}^d$. To shorten notation, we write $T'(x)v$ instead of $T'(x)(v)$ or $(T'(x))(v)$. Alternatively, we write $D_xT(v)$. In what follows, $|T'(x)|$ denotes the operator norm of $T'(x)$, that is,

$$\big|T'(x)\big| = \sup\big\{\big\|T'(x)v\big\| : \|v\| \leq 1\big\},$$

while $\|T'\|$ denotes the supremum of those norms on U, that is,

$$\|T'\| = \sup_{x\in U}\big|T'(x)\big|.$$

Condition (c) states that the points whose orbits are confined to U forever are exactly the points of X. Thus each point which is not in X eventually escapes from U under iteration by T. Usually, the closer such a point is to X the longer the escape from U takes. Such points are said to be *repelled* from X. We will now show that every expanding repeller, when restricted to the set X, is a distance expanding map. In fact, we will show that an expanding repeller is distance expanding on an open κ-neighborhood of X denoted by $B(X,\kappa) := \{y \in \mathbb{R}^d : d(y,X) < \kappa\}$. Note that the proof here uses many of the ideas we will present more generally in the next section. In particular, we use the following topological fact, which we state here without proof: For every open cover $\mathcal{U}$ of a compact metric space X, there exists a positive number ϵ, called a *Lebesgue number*, such that every subset of X of diameter less than ϵ is contained entirely in some element of the cover $\mathcal{U}$.

Theorem 4.1.5. *If (X,U,T) is an expanding repeller, then there exists $\kappa > 0$ such that the map $T|_{B(X,\kappa)} : B(X,\kappa) \to \mathbb{R}^d$ is distance expanding. In particular, the system $T|_X : X \to X$ is distance expanding.*

Proof. As this property depends solely on the first iterate of T, the proof relies solely on condition (b) of the definition of a repeller and on the compactness of X.

Since T is differentiable on U, its derivative $T'(x)$ exists at every point $x \in U$. Furthermore, as $T \in C^1(U)$, its derivative T' is continuous on U. Condition (b) guarantees that for each $x \in X$ there exists $r'_x > 0$ such that

$$\big\|T'(z)v\big\| \geq \frac{1+\lambda}{2}\|v\|, \quad \forall v \in \mathbb{R}^d,\ \forall z \in B\big(x, r'_x\big).$$

In particular, this implies that $T'(z)$ is one-to-one, and is therefore a linear isomorphism of $\mathbb{R}^d$, for every $z \in \bigcup_{x\in X} B(x,r'_x) \supseteq X$. Therefore, the inverse function theorem (Theorem A.2.1) asserts that for every $x \in X$ there exists $r''_x > 0$ such that $T : B(x,r''_x) \to T(B(x,r''_x))$ is a diffeomorphism, and $T : \bigcup_{x\in X} B(x,r''_x) \to \mathbb{R}^d$ is a local diffeomorphism. For every $x \in X$, let $r_x = \min\{r'_x, r''_x\}$. The family of open balls $\{B(x,r_x) : x \in X\}$ forms an open cover of X and hence admits a Lebesgue number $\delta > 0$. That is, for every $x \in X$ there is $\tilde{x} \in X$ such that $B(x,\delta/2) \subseteq B(\tilde{x}, r_{\tilde{x}})$. Setting $r = \delta/2$, it follows that

$$\big\|T'(z)v\big\| \geq \frac{1+\lambda}{2}\|v\|, \quad \forall v \in \mathbb{R}^d,\ \forall z \in B(X,r) := \bigcup_{x\in X} B(x,r) \tag{4.1}$$

and $T : B(x,r) \to T(B(x,r))$ is a diffeomorphism for all $x \in X$. Denote the inverse of the diffeomorphism $T : B(x,r) \to T(B(x,r))$ by

$$T_x^{-1} : T\big(B(x,r)\big) \to B(x,r). \tag{4.2}$$

As the set $T(B(x,r))$ is open for all $x \in X$, let q_x be the largest radius $Q > 0$ such that $B(T(x),Q) \subseteq T(B(x,r))$. As X is compact, we have $q := \inf_{x \in X} q_x > 0$. (Take the fact that $q > 0$ for granted for the moment; the proof of Lemma 4.2.2 below applies here.) Then

$$B(T(x),q) \subseteq T(B(x,r)), \quad \forall x \in X. \tag{4.3}$$

Furthermore, since it is the image of an open set under a diffeomorphism, the set $T_x^{-1}(B(T(x),q))$ is open for all $x \in X$. Let p_x be the largest $0 < P \leq r$ such that $B(x,P) \subseteq T_x^{-1}(B(T(x),q))$. As X is compact, we have $p := \inf_{x \in X} p(x) > 0$. (Take the fact that $p > 0$ for granted for the moment; a variation of the proof of Lemma 4.2.2 below can also be applied here.) Note that $p \leq r$ by definition. In addition,

$$B(x,p) \subseteq T_x^{-1}(B(T(x),q)), \quad \forall x \in X. \tag{4.4}$$

Let $y_1, y_2 \in B(X, p/2)$ be such that $\|y_1 - y_2\| < p/2$. Then there exists $x \in X$ such that $y_1, y_2 \in B(x,p)$. Therefore, $T(y_1), T(y_2) \in B(T(x),q)$ according to (4.4). Let $S = [T(y_1), T(y_2)]$ be the line segment joining $T(y_1)$ and $T(y_2)$. Due to the convexity of balls in $\mathbb{R}^d$, we know that $S \subseteq B(T(x),q)$. Moreover, $T_x^{-1}(T(y_1)) = y_1$ and $T_x^{-1}(T(y_2)) = y_2$. Therefore, the curve $T_x^{-1}(S)$ joins the points y_1 to y_2, and thus

$$\|y_1 - y_2\| \leq \ell(T_x^{-1}(S)) = \int_S \|(T_x^{-1})'(w)u\| \, dw, \tag{4.5}$$

where $\ell(T_x^{-1}(S))$ stands for the length of the curve $T_x^{-1}(S)$ and u is the unit vector in the direction from $T(y_1)$ to $T(y_2)$. Since $T_x^{-1}(S) \subseteq T_x^{-1}(B(T(x),q)) \subseteq B(x,r)$, for any $w \in S$ inequality (4.1) can be applied with $z = T_x^{-1}(w)$ and $v = (T_x^{-1})'(w)u$ to yield

$$1 = \|u\| = \|T'(T_x^{-1}(w))((T_x^{-1})'(w)u)\| \geq \frac{1+\lambda}{2}\|(T_x^{-1})'(w)u\|.$$

Consequently,

$$\|(T_x^{-1})'(w)u\| \leq \frac{2}{1+\lambda}, \quad \forall w \in S,$$

and hence (4.5) gives

$$\|y_1 - y_2\| \leq \int_S \frac{2}{1+\lambda} dz = \frac{2}{1+\lambda}\ell(S) = \frac{2}{1+\lambda}\|T(y_1) - T(y_2)\|.$$

In other words,

$$d(T(y_1), T(y_2)) \geq \frac{1+\lambda}{2} d(y_1, y_2), \quad \forall y_1, y_2 \in B(X, p/2) \text{ with } d(y_1, y_2) < p/2.$$

Letting $\kappa = p/2$ completes the proof. □

4.1.2 Hyperbolic Cantor sets

In this section, we introduce and study in detail one special class of expanding repellers (and so of distance expanding maps), namely, hyperbolic Cantor sets. Their construction is a prototype of conformal iterated function systems and conformal graph directed Markov systems, whose systematic account will be given in Chapter 19.

Recall that a *similarity map* $S : \mathbb{R}^d \to \mathbb{R}^d$ is a bijection that multiplies all distances by the same positive real number r called *similarity ratio*, that is,

$$\|S(x) - S(y)\| = r\|x - y\|, \quad \forall x, y \in \mathbb{R}^d.$$

When $r = 1$, a similarity is called an isometry. Two sets are called *similar* if one is the image of the other under a similarity. A similarity $S : \mathbb{R}^d \to \mathbb{R}^d$ with ratio r takes the form

$$S(x) = rA(x) + b,$$

where $A : \mathbb{R}^d \to \mathbb{R}^d$ is an $d \times d$ orthogonal matrix and $b \in \mathbb{R}^d$ is a translation vector. Similarities preserve shapes, including line segments, lines, planes, parallelism, and perpendicularity. Similarities preserve angles but do not necessarily preserve orientation (in fact, S and A preserve orientation if and only if $\det(A) > 0$).

Note also that $S'(x) = rA$ for all $x \in \mathbb{R}^d$. Therefore, $\|S'(x)v\| = r\|v\|$ for all $v \in \mathbb{R}^d$. Consequently, $|S'(x)| = r$ for all $x \in \mathbb{R}^d$, and hence $\|S'\| = r$.

Let E be a finite set such that $\#E \geq 2$. Let $\varphi_e : \mathbb{R}^d \to \mathbb{R}^d$, $e \in E$, be similarities for which there exists a compact set $X_0 \subseteq \mathbb{R}^d$ with the following properties:

(i) $0 < \|\varphi_e'\| < 1$ for all $e \in E$.
(ii) $\varphi_e(X_0) \subseteq X_0$ for all $e \in E$.
(iii) $\varphi_e(X_0) \cap \varphi_f(X_0) = \emptyset$ for all $e, f \in E$ with $e \neq f$.

Construction of the limit set X.

We now define the *limit set X* by constructing a descending sequence of compact sets $(X_n)_{n=1}^{\infty}$, all of which are subsets of $\bigcup_{e\in E} \varphi_e(X_0)$. The set X_n is called the nth *level set* of the construction. The limit set X will be the intersection of the level sets.

We will use symbolic dynamics notation. For every $\omega \in E^*$, define

$$\varphi_\omega := \varphi_{\omega_1} \circ \varphi_{\omega_2} \circ \cdots \circ \varphi_{\omega_{|\omega|}}.$$

The maps φ_e, $e \in E$, are said to be the *generators* of the construction, and so we say that the map φ_ω is generated by the word ω. The nth level set X_n is the disjoint union of the images of X_0 under all maps generated by words of length n, namely

$$X_n := \bigcup_{\omega\in E^n} \varphi_\omega(X_0). \tag{4.6}$$

As a finite union of compact sets, each level set X_n is compact. Moreover, by condition (ii),

$$X_{n+1} = \bigcup_{\omega \in E^{n+1}} \varphi_{\omega|_n}(\varphi_{\omega_{n+1}}(X_0)) \subseteq \bigcup_{\tau \in E^n} \varphi_\tau(X_0) = X_n \tag{4.7}$$

for all $n \in \mathbb{N}$. As the intersection of a descending sequence of nonempty compact sets, the limit set

$$X := \bigcap_{n=1}^{\infty} X_n \tag{4.8}$$

is a nonempty compact set.

The limit set X is a topological Cantor set (the proof of this fact is postponed) with the property that

$$\bigcup_{e \in E} \varphi_e(X) = X.$$

Indeed, we first observe that the generators map the nth level set down to the $(n+1)$th level set since

$$\begin{aligned}
\bigcup_{e \in E} \varphi_e(X_n) &= \bigcup_{e \in E} \bigcup_{\omega \in E^n} \varphi_e(\varphi_\omega(X_0)) \\
&= \bigcup_{e \in E} \bigcup_{\omega \in E^n} \varphi_{e\omega}(X_0) \\
&= \bigcup_{\tau \in E^{n+1}} \varphi_\tau(X_0) \\
&= X_{n+1}
\end{aligned} \tag{4.9}$$

for all $n \in \mathbb{N}$.

Moreover, $x \in \bigcap_{n=1}^{\infty}[\bigcup_{e \in E} \varphi_e(X_n)]$ if and only if for every $n \in \mathbb{N}$ there exists $e_n \in E$ such that $x \in \varphi_{e_n}(X_n)$. Since $\varphi_e(X_n) \subseteq \varphi_e(X_0)$ for all $e \in E$, the e_n's are unique according to condition (iii). By this very same condition, since $x \in \varphi_{e_n}(X_n) \cap \varphi_{e_{n+1}}(X_{n+1}) \subseteq \varphi_{e_n}(X_0) \cap \varphi_{e_{n+1}}(X_0)$, it turns out that $e_n = e_{n+1}$ for all $n \in \mathbb{N}$. In summary, $x \in \bigcap_{n=1}^{\infty}[\bigcup_{e \in E} \varphi_e(X_n)]$ if and only if there is a unique $e \in E$ such that $x \in \varphi_e(X_n)$ for all $n \in \mathbb{N}$. In other words,

$$\bigcup_{e \in E} \left[\bigcap_{n=1}^{\infty} \varphi_e(X_n)\right] = \bigcap_{n=1}^{\infty} \left[\bigcup_{e \in E} \varphi_e(X_n)\right]. \tag{4.10}$$

It follows from (4.8), (4.9), (4.10), and the injectivity of the generators that

$$\begin{aligned}
\bigcup_{e \in E} \varphi_e(X) = \bigcup_{e \in E} \varphi_e\left(\bigcap_{n=1}^{\infty} X_n\right) &= \bigcup_{e \in E} \left[\bigcap_{n=1}^{\infty} \varphi_e(X_n)\right] \\
&= \bigcap_{n=1}^{\infty} \left[\bigcup_{e \in E} \varphi_e(X_n)\right] = \bigcap_{n=1}^{\infty} X_{n+1} = X.
\end{aligned}$$

By induction, we have that

$$\bigcup_{\omega\in E^n} \varphi_\omega(X) = X, \quad \forall n \in \mathbb{N}. \tag{4.11}$$

Construction of a neighborhood U of X.

Since $\varphi_e(X_0) \cap \varphi_f(X_0) = \emptyset$ for all $e \neq f$ and since there are finitely many compact sets $\varphi_e(X_0)$, $e \in E$, the continuity of the generators ensures the existence of an open ε-neighborhood $B(X_0, \varepsilon) := \{x \in \mathbb{R}^d : d(x, X_0) < \varepsilon\}$ of X_0 such that

$$\varphi_e\big(B(X_0,\varepsilon)\big) \cap \varphi_f\big(B(X_0,\varepsilon)\big) = \emptyset, \quad \forall e \neq f.$$

Let

$$U = \bigcup_{e\in E} \varphi_e\big(B(X_0,\varepsilon)\big). \tag{4.12}$$

Construction of a map $T : U \to \mathbb{R}^d$.

Finally, we define a map $T : U \to \mathbb{R}^d$ by

$$T|_{\varphi_e(B(X_0,\varepsilon))} = \varphi_e^{-1}. \tag{4.13}$$

This piecewise-similar map is well-defined since the sets $\varphi_e(B(X_0,\varepsilon))$, $e \in E$, are mutually disjoint.

Proof that the triple (X, U, T) is an expanding repeller.

Condition (a) for a repeller is rather easy to check. First, note that T maps the nth level set up to the $(n-1)$th level set, that is,

$$T(X_n) = X_{n-1}, \quad \forall n \in \mathbb{N}. \tag{4.14}$$

Indeed, let $n \in \mathbb{N}$. Since $\varphi_e(X_{n-1}) \subseteq \varphi_e(B(X_0,\varepsilon))$ and since $T|_{\varphi_e(B(X_0,\varepsilon))} = \varphi_e^{-1}$ for every $e \in E$, we have

$$\begin{aligned}
T(X_n) &= \bigcup_{\omega\in E^n} T\big(\varphi_\omega(X_0)\big) \\
&= \bigcup_{\tau\in E^{n-1}} \bigcup_{e\in E} T\big(\varphi_{e\tau}(X_0)\big) \\
&= \bigcup_{\tau\in E^{n-1}} \bigcup_{e\in E} T \circ \varphi_e\big(\varphi_\tau(X_0)\big) \\
&= \bigcup_{\tau\in E^{n-1}} \varphi_\tau(X_0) \\
&= X_{n-1}.
\end{aligned}$$

Then

$$T(X) = T\left(\bigcap_{n=1}^{\infty} X_n\right) \subseteq \bigcap_{n=1}^{\infty} T(X_n) = \bigcap_{n=1}^{\infty} X_{n-1} = X.$$

This establishes that $T(X) \subseteq X$. To prove the reverse inclusion, pick $x \in X$. Then $x \in X_n$ for every $n \in \mathbb{N}$. Fix an arbitrary $e \in E$. Then $\varphi_e(x) \in X_{n+1}$ for every $n \in \mathbb{N}$. Thus $\varphi_e(x) \in X$. It follows that $x = \varphi_e^{-1}(\varphi_e(x)) = T(\varphi_e(x)) \in T(X)$. Hence $X \subseteq T(X)$. Since both inclusions hold, we conclude that

$$T(X) = X. \tag{4.15}$$

Condition (b) for a repeller is also straightforward to verify. Indeed, let $x \in U$. There exists a unique $e_x \in E$ such that $x \in \varphi_{e_x}(B(X_0, \varepsilon))$. Then, for all $v \in \mathbb{R}^d$,

$$\begin{aligned} \|T'(x)v\| &= \|(T|_{\varphi_{e_x}(B(X_0,\varepsilon))})'(x)v\| = \|(\varphi_{e_x}^{-1})'(x)v\| = \|\varphi_{e_x}'\|^{-1}\|v\| \\ &\geq \min_{e\in E} \frac{1}{\|\varphi_e'\|}\|v\| = \frac{1}{M}\|v\|, \end{aligned} \tag{4.16}$$

where $M := \max_{e\in E} \|\varphi_e'\| < 1$ by condition (i).

Finally, we show that condition (c) for a repeller is fulfilled. First, we observe that

$$T^{-n}(U) = \bigcup_{\omega\in E^{n+1}} \varphi_\omega(B(X_0, \varepsilon)), \quad \forall n \in \mathbb{Z}_+. \tag{4.17}$$

Indeed, by definition of U, the relationship holds when $n = 0$. For the inductive step, let $x \in U$ and $n \in \mathbb{N}$. Then

$$\begin{aligned} x \in T^{-n}(U) &\Longleftrightarrow T(x) \in T^{-(n-1)}(U) \\ &\Longleftrightarrow T(x) \in \bigcup_{\omega\in E^n} \varphi_\omega(B(X_0,\varepsilon)) \\ &\Longleftrightarrow \varphi_f^{-1}(x) \in \bigcup_{\omega\in E^n} \varphi_\omega(B(X_0,\varepsilon)), \text{ if } x \in \varphi_f(B(X_0,\varepsilon)) \\ &\Longleftrightarrow x \in \bigcup_{\tau\in E^{n+1}} \varphi_\tau(B(X_0,\varepsilon)). \end{aligned}$$

By induction, (4.17) holds.

Now, observe that the similarity map φ_e enjoys the property that

$$\varphi_e(B(X_0,\varepsilon)) = B(\varphi_e(X_0), \|\varphi_e'\|\varepsilon) \subseteq B(\varphi_e(X_0), M\varepsilon).$$

By an induction argument, we deduce that

$$\varphi_\omega(B(X_0,\varepsilon)) = B(\varphi_\omega(X_0), \|\varphi_\omega'\|\varepsilon) \subseteq B(\varphi_\omega(X_0), M^{|\omega|}\varepsilon), \quad \forall \omega \in E^*. \tag{4.18}$$

Therefore,

$$\begin{aligned} \bigcup_{\omega\in E^n} \varphi_\omega(B(X_0,\varepsilon)) &\subseteq \bigcup_{\omega\in E^n} B(\varphi_\omega(X_0), M^n\varepsilon) \\ &= B\Big(\bigcup_{\omega\in E^n} \varphi_\omega(X_0), M^n\varepsilon\Big) \\ &= B(X_n, M^n\varepsilon), \quad \forall n \in \mathbb{N}. \end{aligned} \tag{4.19}$$

It follows that (cf. Exercise 4.6.4)

$$X = \bigcap_{n=1}^{\infty} X_n = \bigcap_{n=1}^{\infty} \bigcup_{\omega\in E^n} \varphi_\omega(X_0) \subseteq \bigcap_{n=1}^{\infty} \bigcup_{\omega\in E^n} \varphi_\omega(B(X_0,\varepsilon)) \subseteq \bigcap_{n=1}^{\infty} B(X_n, M^n\varepsilon) = X.$$

From this and (4.17), we conclude that

$$X = \bigcap_{n=1}^{\infty} \bigcup_{\omega\in E^n} \varphi_\omega(B(X_0,\varepsilon)) = \bigcap_{n=0}^{\infty} T^{-n}(U). \tag{4.20}$$

This establishes condition (c), and completes the proof that the triple (X, U, T) is an expanding repeller.

Alternative construction of the limit set X and proof that X is a topological Cantor set.

The limit set X can also be constructed in a slightly different way. Let $\omega \in E^\infty$. Since (by condition (ii))

$$\varphi_{\omega|_{n+1}}(X_0) = \varphi_{\omega|_n}(\varphi_{\omega_{n+1}}(X_0)) \subseteq \varphi_{\omega|_n}(X_0) \tag{4.21}$$

and since

$$\operatorname{diam}(\varphi_{\omega|_n}(X_0)) = \|\varphi'_{\omega|_n}\| \operatorname{diam}(X_0) \le M^n \operatorname{diam}(X_0) \tag{4.22}$$

for all $n \in \mathbb{N}$, the sets $(\varphi_{\omega|_n}(X_0))_{n=1}^{\infty}$ form a descending sequence of nonempty compact sets whose diameters tend to 0 (by condition (i)). Therefore, $\bigcap_{n=1}^{\infty} \varphi_{\omega|_n}(X_0)$ is a singleton. Define the *coding map* $\pi : E^\infty \to \mathbb{R}^d$ by

$$\{\pi(\omega)\} := \bigcap_{n=1}^{\infty} \varphi_{\omega|_n}(X_0).$$

This map is injective. Indeed, if $\omega \neq \tau \in E^\infty$, then there is a smallest $n \in \mathbb{N}$ such that $\omega_n \neq \tau_n$. It follows from the injectivity of the generators and condition (iii) that

$$\begin{aligned}
\{\pi(\omega)\} \cap \{\pi(\tau)\} &\subseteq \varphi_{\omega|_n}(X_0) \cap \varphi_{\tau|_n}(X_0) \\
&= [\varphi_{\omega|_{n-1}}(\varphi_{\omega_n}(X_0))] \cap [\varphi_{\omega|_{n-1}}(\varphi_{\tau_n}(X_0))] \\
&= \varphi_{\omega|_{n-1}}(\varphi_{\omega_n}(X_0) \cap \varphi_{\tau_n}(X_0)) \\
&= \emptyset.
\end{aligned}$$

We claim that

$$X = \pi(E^\infty). \tag{4.23}$$

Indeed, let $\omega \in E^\infty$. Since $\varphi_{\omega|_n}(X_0) \subseteq X_n$ for all $n \in \mathbb{N}$, we have $\{\pi(\omega)\} = \bigcap_{n=1}^\infty \varphi_{\omega|_n}(X_0) \subseteq \bigcap_{n=1}^\infty X_n = X$. Therefore, $\pi(E^\infty) \subseteq X$. Now, let $x \in X$. Then $x \in X_n$ for every $n \in \mathbb{N}$. This means that there exists a unique $\omega^{(n)} \in E^n$ such that $x \in \varphi_{\omega^{(n)}}(X_0)$. The uniqueness of the $\omega^{(n)}$'s implies that $\omega^{(n+1)}|_n = \omega^{(n)}$. Define $\omega \in E^\infty$ to be such that $\omega|_n = \omega^{(n)}$ for all $n \in \mathbb{N}$. Then $x \in \varphi_{\omega^{(n)}}(X_0) = \varphi_{\omega|_n}(X_0)$ for all $n \in \mathbb{N}$. Thus $x = \pi(\omega)$. Hence $X \subseteq \pi(E^\infty)$. Since both inclusions hold, the claim has been shown.

Furthermore, the map π is continuous. This ensues from the fact that, for every $\rho \in E^*$,

$$\operatorname{diam}(\pi([\rho])) \le \operatorname{diam}(\varphi_\rho(X_0)) \le M^{|\rho|} \operatorname{diam}(X_0). \tag{4.24}$$

In summary, the map $\pi : E^\infty \to X$ is a continuous bijection between two compact metrizable spaces. Thus $\pi : E^\infty \to X$ is a homeomorphism and X is a homeomorphic image of E^∞, that is, X is a topological Cantor set.

For later purposes, we further note that $\varphi_{\omega_1} \circ \pi \circ \sigma(\omega) = \pi(\omega)$ for all $\omega \in E^\infty$. In light of (4.13), this means that

$$\pi \circ \sigma = T \circ \pi, \tag{4.25}$$

that is, the symbolic system (E^∞, σ) is topologically conjugate to the dynamical system (X, T) via the coding map π.

One final word: Condition (i) is the reason for calling X an hyperbolic Cantor set. In differentiable dynamical systems, hyperbolicity takes the shape of a derivative that stays away from 1.

Example 4.1.6. Let $\varphi_e : \mathbb{R} \to \mathbb{R}$, $e \in E := \{0, 2\}$, be the two contracting similarities defined by

$$\varphi_e(x) = \frac{x + e}{3}.$$

Let $X_0 = I := [0, 1]$ be the unit interval. The limit set X defined in Subsection 4.1.2 is called the *middle-third Cantor set* and is usually denoted by C. Let $\varepsilon = 1/3$. Then $B(X_0, \varepsilon) = (-1/3, 4/3)$ and

$$U = \varphi_0\big((-1/3, 4/3)\big) \cup \varphi_2\big((-1/3, 4/3)\big) = (-1/9, 4/9) \cup (5/9, 10/9).$$

Moreover, $T : U \to \mathbb{R}$ is defined by $T(x) = 3x - e$ if $x \in \varphi_e(B(X_0, \varepsilon))$, that is,

$$T(x) = \begin{cases} 3x & \text{if } x \in (-1/9, 4/9) \\ 3x - 2 & \text{if } x \in (5/9, 10/9). \end{cases}$$

In two dimensions, the most classic examples of hyperbolic Cantor sets are Sierpiński triangles (also called Sierpiński gaskets) and Sierpiński carpets. The following example is a natural generalization of the middle-third Cantor set.

Example 4.1.7. Let $\varphi_e : \mathbb{R}^2 \to \mathbb{R}^2$, $e \in E := \{0, 1, 2\}$, be the three contracting similarities defined by

$$\varphi_0(x,y) = \left(\frac{x}{3}, \frac{y}{3}\right)$$
$$\varphi_1(x,y) = \left(\frac{x+1}{3}, \frac{y+\sqrt{3}}{3}\right)$$
$$\varphi_2(x,y) = \left(\frac{x+2}{3}, \frac{y}{3}\right).$$

Let X_0 be the filled-in equilateral triangle with vertices $(0,0)$, $(1/2, \sqrt{3}/2)$ and $(1,0)$. The limit set X defined in Subsection 4.1.2 is a totally disconnected *Sierpinski triangle*. Let $\varepsilon = 1/3$. Then $U = \bigcup_{e \in E} \varphi_e(B(X_0, \varepsilon))$ and the map $T : U \to \mathbb{R}^2$ is defined by

$$T(x,y) := \varphi_e^{-1}(x,y), \quad \forall (x,y) \in \varphi_e\big(B(X_0,\varepsilon)\big), \quad \forall e \in E.$$

That is,

$$T(x,y) = \begin{cases} (3x, 3y) & \text{if } (x,y) \in \varphi_0(B(X_0,\varepsilon)) \\ (3x-1, 3y-\sqrt{3}) & \text{if } (x,y) \in \varphi_1(B(X_0,\varepsilon)) \\ (3x-2, 3y) & \text{if } (x,y) \in \varphi_2(B(X_0,\varepsilon)). \end{cases}$$

Example 4.1.8. Let $\varphi_e : \mathbb{R}^2 \to \mathbb{R}^2$, $e \in E := \{0, 1, 2, 3\}$, be the following four contracting similarities:

$$\varphi_0(x,y) = \left(\frac{x}{4}, \frac{y}{4}\right), \quad \varphi_1(x,y) = \left(\frac{x+3}{4}, \frac{y}{4}\right),$$
$$\varphi_2(x,y) = \left(\frac{x+3}{4}, \frac{y+3}{4}\right), \quad \varphi_3(x,y) = \left(\frac{x}{4}, \frac{y+3}{4}\right).$$

Let $X_0 = I^2$ be the unit square. The limit set X defined in Subsection 4.1.2 is a totally disconnected *Sierpinski carpet* (see Figure 4.1). Let $\varepsilon = 1/2$. Then $U = \bigcup_{e \in E} \varphi_e(B(X_0, \varepsilon))$ and the map $T : U \to \mathbb{R}^2$ is defined by

$$T(x,y) := \varphi_e^{-1}(x,y), \quad \forall (x,y) \in \varphi_e\big(B(X_0,\varepsilon)\big), \quad \forall e \in E.$$

That is,

$$T(x,y) = \begin{cases} (4x, 4y) & \text{if } (x,y) \in \varphi_0(B(X_0,\varepsilon)) \\ (4x-3, 4y) & \text{if } (x,y) \in \varphi_1(B(X_0,\varepsilon)) \\ (4x-3, 4y-3) & \text{if } (x,y) \in \varphi_2(B(X_0,\varepsilon)) \\ (4x, 4y-3) & \text{if } (x,y) \in \varphi_3(B(X_0,\varepsilon)). \end{cases}$$

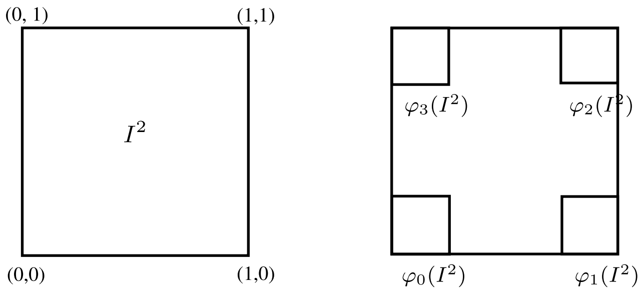

Figure 4.1: The action of the four contracting similarities $\varphi_0, \dots, \varphi_3$ on the closed unit square I^2.

4.2 Inverse branches

In order for a map to have a properly defined inverse, it is necessary that the map be injective. Nonetheless, we can get around the noninjectivity of a map by defining its *inverse branches* as long as the map is locally injective. The following proposition is the first step in the construction of the inverse branches of a distance expanding map.

Proposition 4.2.1. *Let $T : X \to X$ be a distance expanding map. For all $x \in X$, the restriction $T|_{B(x,\delta)}$ is injective.*

Proof. For each $x \in X$, apply Remark 4.1.2(b) to $B(x, \delta)$. □

We shall assume from this point on that $T : X \to X$ is an **open**, distance expanding map of a compact metric space X. Note that the restriction $T|_F$ of T to a closed forward T-invariant subset F of X need not be open (see Exercise 4.6.5). However, since T is open, for every $x \in X$ and $r > 0$ the set $T(B(x, r))$ is open and, therefore, contains a nonempty open ball centered at $T(x)$, say $B(T(x), s(r))$. Accordingly, we define

$$R(x, r) := \sup\{s > 0 : B\big(T(x), s\big) \subseteq T\big(B(x, r)\big)\} > 0.$$

In fact, $R(x, r)$ is the radius of the largest ball centered at $T(x)$ which is contained in $T(B(x, r))$. In the following lemma, we investigate the greatest lower bound of the radii $R(x, r)$ for a fixed $r > 0$.

Lemma 4.2.2. *For every $r > 0$, we have $R(r) := \inf\{R(x, r) : x \in X\} > 0$.*

Proof. We shall prove this lemma by contradiction. Suppose that there exists some $r > 0$ for which $R(r) = 0$. This means that there exists a sequence $(x_n)_{n=1}^{\infty}$ in X such that $\lim_{n\to\infty} R(x_n, r) = 0$. Since X is compact, the sequence $(x_n)_{n=1}^{\infty}$ has a convergent subsequence, say $(x_{n_k})_{k=1}^{\infty}$. Define $x := \lim_{k\to\infty} x_{n_k}$. Then there exists $K \in \mathbb{N}$ such that $d(x_{n_k}, x) < r/2$ for all $k \geq K$. In particular, this implies that

$$B(x, r/2) \subseteq B(x_{n_k}, r), \quad \forall k \geq K. \tag{4.26}$$

Now, since T is an open map, there exists $\eta > 0$ such that

$$B\big(T(x), \eta\big) \subseteq T\big(B(x, r/2)\big). \tag{4.27}$$

Moreover, $T(x) = \lim_{k\to\infty} T(x_{n_k})$ since T is continuous. Thus there exists $K' \in \mathbb{N}$ such that $d(T(x_{n_k}), T(x)) < \eta/2$ for all $k \geq K'$. In particular, this implies that

$$B\big(T(x_{n_k}), \eta/2\big) \subseteq B\big(T(x), \eta\big), \quad \forall k \geq K'. \tag{4.28}$$

From (4.26), (4.27), and (4.28), it follows that for all $k \geq \max\{K, K'\}$,

$$B\big(T(x_{n_k}), \eta/2\big) \subseteq B\big(T(x), \eta\big) \subseteq T\big(B(x, r/2)\big) \subseteq T\big(B(x_{n_k}, r)\big).$$

Therefore, $R(x_{n_k}, r) \geq \eta/2$ for all $k \geq \max\{K, K'\}$, which contradicts our assumption that $\lim_{n\to\infty} R(x_n, r) = 0$. Thus $\inf\{R(x, r) : x \in X\} > 0$. □

Remark 4.2.3. Notice that we did not use the distance expanding property of T in the proof, so this result holds for all open, continuous maps of a compact metric space. The compactness of the space X ensures that for any fixed r, the radii $R(x, r)$ have a positive greatest lower bound $R(r)$. Thus the image of any ball of radius r contains a ball of radius $R(r)$.

Lemma 4.2.2 shows that for each $r > 0$ the image of the ball of radius r centered at the point x contains the ball of radius $R(r)$ centered at $T(x)$. For an open, distance expanding map T, the quantity

$$\xi := \min\big\{\delta, R(\delta)\big\} > 0, \tag{4.29}$$

where δ is a constant delimiting the neighborhoods of expansion of the map, is of particular interest. Indeed, given that, according to Proposition 4.2.1, the restricted map $T|_{B(x,\delta)}$ is injective for every $x \in X$, we can define its inverse

$$T|_{B(x,\delta)}^{-1} : T\big(B(x, \delta)\big) \to B(x, \delta).$$

By the definition of ξ given above, we have that

$$B\big(T(x), \xi\big) \subseteq T\big(B(x, \delta)\big)$$

for every $x \in X$. This inclusion is illustrated in Figure 4.2. We denote the restriction of the inverse of the map $T|_{B(x,\delta)}$ to $B(T(x), \xi)$ by

$$T_x^{-1} := \big(T|_{B(x,\delta)}^{-1}\big)\big|_{B(T(x),\xi)} : B\big(T(x), \xi\big) \to B(x, \delta). \tag{4.30}$$

Note that T_x^{-1} is injective but not necessarily surjective. The map T_x^{-1} is the local inverse branch of T that maps $T(x)$ to x. As T expands distances by a factor $\lambda > 1$, one naturally expects the local inverse branches T_x^{-1}, $x \in X$, to contract distances by a factor λ^{-1}. This is indeed the case.

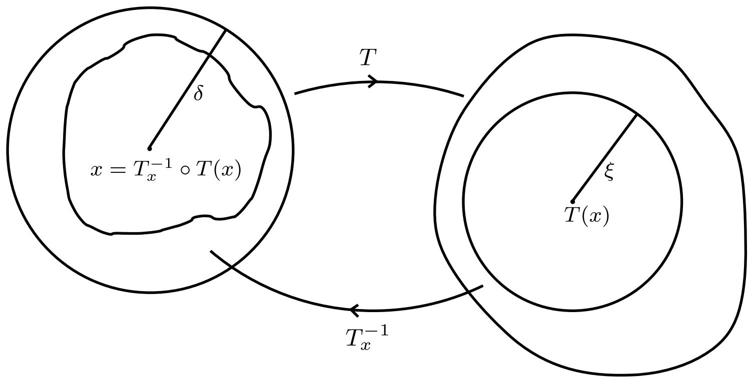

Figure 4.2: Illustration of the ball $B(T(x), \xi)$ mapped under the inverse branch T_x^{-1} inside the ball $B(x, \delta)$.

Proposition 4.2.4. *The local inverse branches T_x^{-1}, for each $x \in X$, are contractions with (contraction) ratio λ^{-1}.*

Proof. Fix $x \in X$. We aim to prove that if $y, z \in B(T(x), \xi)$, then

$$d(T_x^{-1}(y), T_x^{-1}(z)) \leq \lambda^{-1} d(y, z),$$

where λ is a constant of expansion for T. Since $T_x^{-1}(B(T(x), \xi)) \subseteq B(x, \delta)$, both $T_x^{-1}(y)$ and $T_x^{-1}(z)$ lie in $B(x, \delta)$, and hence

$$d(T_x^{-1}(y), T_x^{-1}(z)) < 2\delta.$$

Therefore, the expanding property of T guarantees that

$$d(y, z) = d(T \circ T_x^{-1}(y), T \circ T_x^{-1}(z)) \geq \lambda d(T_x^{-1}(y), T_x^{-1}(z)).$$

Consequently,

$$d(T_x^{-1}(y), T_x^{-1}(z)) \leq \lambda^{-1} d(y, z).$$ □

Now, let $w \in B(T(x), \xi)$. Proposition 4.2.4 implies that

$$d(T_x^{-1}(w), x) = d(T_x^{-1}(w), T_x^{-1}(T(x))) \leq \lambda^{-1} d(w, T(x)) < \lambda^{-1} \xi.$$

Thus

$$T_x^{-1}(B(T(x), \xi)) \subseteq B(x, \lambda^{-1}\xi) \subseteq B(x, \xi).$$

Thanks to this property, we can define the local inverse branches of the iterates of T. Let $x \in X$ and $n \in \mathbb{N}$. The local inverse branch of T^n that maps $T^n(x)$ to x is defined to be

$$T_x^{-n} := T_x^{-1} \circ T_{T(x)}^{-1} \circ \cdots \circ T_{T^{n-1}(x)}^{-1} : B(T^n(x), \xi) \to B(x, \xi). \tag{4.31}$$

This composition will henceforth be called the *inverse branch of T^n determined by the point x*. See Figure 4.3.

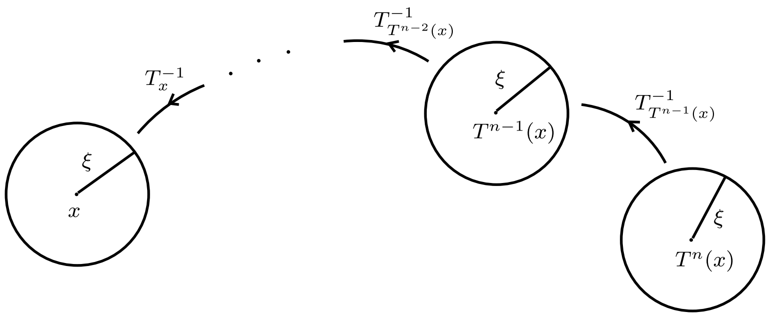

Figure 4.3: The map $T^{-1}_{T^{n-1}(x)}$ sends the ball $B(T^n(x),\xi)$ into the ball $B(T^{n-1}(x),\xi)$, which is in turn mapped into the ball $B(T^{n-2}(x),\xi)$ by $T^{-1}_{T^{n-2}(x)}$ and so on, until finally T_x^{-1} sends us back inside $B(x,\xi)$.

Remark 4.2.5. Let $x \in X$ and $y,z \in B(T^n(x),\xi)$. A successive application of Proposition 4.2.4 at the points $T^{n-1}(x)$, $T^{n-2}(x), \ldots, T(x)$ and x establishes that

$$d\big(T_x^{-n}(y), T_x^{-n}(z)\big) \le \lambda^{-n}\, d(y,z).$$

In particular,

$$T_x^{-n}\big(B(T^n(x),\xi')\big) \subseteq B(x,\lambda^{-n}\xi') \subseteq B(x,\xi'), \quad \forall 0 < \xi' \le \xi,\ \forall n \in \mathbb{N}, \tag{4.32}$$

and hence

$$B(T^n(x),\xi') \subseteq T^n\big(B(x,\lambda^{-n}\xi')\big) \subseteq T^n\big(B(x,\xi')\big), \quad \forall 0 < \xi' \le \xi,\ \forall n \in \mathbb{N}. \tag{4.33}$$

The inverse branches of an open, distance expanding map are easiest to grasp with the aid of an example. Below, we first calculate the inverse branches for the map $T(x) := 2x \pmod 1$, and secondly give the example of a subshift of finite type.

Example 4.2.6. Consider the map $T : \mathbb{S}^1 \to \mathbb{S}^1$ defined by

$$T(x) := \begin{cases} 2x & \text{if } x \in [0,1/2] \\ 2x-1 & \text{if } x \in [1/2,1]. \end{cases}$$

Note that T is the map T_m we have seen in Example 1.1.3(b), with $m = 2$. The subscript 2 has been dropped to simplify notation in what follows. We have that T is distance expanding on neighborhoods of size $\delta = 1/4$ with expanding constant $\lambda = 2$. More precisely, if $x,y \in \mathbb{S}^1$ with $d(x,y) < 1/2$, then $d(T(x),T(y)) = 2d(x,y)$. Here, the metric d is the usual Euclidean metric on the circle. So, for instance, $d(1/8,7/8) = 1/4$, whereas

$d(T(1/8), T(7/8)) = d(1/4, 3/4) = 1/2$. Also, T is easily seen to be an open map. We can now consider the inverse branches of T. Let us start by investigating the inverse branches determined by the points 1/4 and 3/4. We know, by Proposition 4.2.1 (and, in this case, by inspection), that T is injective on any open subinterval of $\mathbb{S}^1$ with radius at most 1/4. In particular, T is injective on the intervals $(0, 1/2)$ and $(1/2, 1)$. Notice that $T(1/4) = 1/2 = T(3/4)$. Moreover,

$$T\Big(B\Big(\tfrac{1}{4}, \tfrac{1}{4}\Big)\Big) = T\Big(\Big(0, \tfrac{1}{2}\Big)\Big) = (0,1) = B\Big(\tfrac{1}{2}, \tfrac{1}{2}\Big) = B\Big(T\Big(\tfrac{1}{4}\Big), \tfrac{1}{2}\Big)$$

and

$$T\Big(B\Big(\tfrac{3}{4}, \tfrac{1}{4}\Big)\Big) = T\Big(\Big(\tfrac{1}{2}, 1\Big)\Big) = (0,1) = B\Big(\tfrac{1}{2}, \tfrac{1}{2}\Big) = B\Big(T\Big(\tfrac{3}{4}\Big), \tfrac{1}{2}\Big).$$

Therefore, $R(1/4, 1/4) = R(3/4, 1/4) = 1/2$. In fact, the image under T of any ball $B(x, 1/4)$ contains a ball of radius 1/2 about the point $T(x)$. Thus, in this case, $R(\delta) = R(1/4) = 1/2$ and so $\xi := \min\{\delta, R(\delta)\} = 1/4$. Hence, we obtain inverse branches $T_x^{-1} : B(T(x), 1/4) \to B(x, 1/4)$. Note that every interval $B(T(x), 1/4)$ has two inverse branches defined upon it, one taking points back to an interval around the preimage of $T(x)$ lying in $(0, 1/2)$ and the other sending points to an interval around the preimage of $T(x)$ lying in $(1/2, 1)$. For example, the two inverse branches defined on the interval $B(1/2, 1/4) = (1/4, 3/4)$ are

$$T_{\frac{1}{4}}^{-1} : \Big(\tfrac{1}{4}, \tfrac{3}{4}\Big) \to \Big(0, \tfrac{1}{2}\Big), \quad \text{defined by } T_{\frac{1}{4}}^{-1}(y) := \tfrac{y}{2}$$

and

$$T_{\frac{3}{4}}^{-1} : \Big(\tfrac{1}{4}, \tfrac{3}{4}\Big) \to \Big(\tfrac{1}{2}, 1\Big), \quad \text{defined by } T_{\frac{3}{4}}^{-1}(y) := \tfrac{y+1}{2}.$$

Let us now consider the inverse branches of the iterates of T. For each point $x \in \mathbb{S}^1$, we have the inverse branch of T^n determined by x:

$$T_x^{-n} : B\Big(T^n(x), \tfrac{1}{4}\Big) \to B\Big(x, \tfrac{1}{4}\Big).$$

Recall that this map is injective but not surjective. In particular, if $x = 2^{-(n+1)}$, the map $T_{2^{-(n+1)}}^{-n} : B(1/2, 1/4) \to B(2^{-(n+1)}, 1/4)$ turns out to be $T_{2^{-(n+1)}}^{-n}(y) = y/2^n$. For the map T, every interval $B(T^n(x), 1/4)$ has 2^n inverse branches defined upon it.

Example 4.2.7 (Subshifts of finite type). It was shown in Example 4.1.3 that the shift map $\sigma : E_A^\infty \to E_A^\infty$ is distance expanding, with $\lambda = s^{-1}$ and $\delta = 1/2$ when E_A^∞ is endowed with the metric d_s. Also, by Theorem 3.2.12, we know that the shift map is an open map. Now, let $\omega \in E_A^\infty$ and observe that $\sigma^{-1}(\omega) = \{i\omega : A_{i\omega_1} = 1\}$. We want to find the inverse branches of σ. Since $2\delta = 1$, we first describe the open ball $B(\omega, 1)$ upon which we know the shift map to be injective, due to Proposition 4.2.1:

$$B(\omega,1) = \{\tau \in E_A^\infty : s^{|\omega\wedge\tau|} < 1\} = \{\tau \in E_A^\infty : |\omega \wedge \tau| \geq 1\} = [\omega_1].$$

In other words, $B(\omega, 1)$ is the initial cylinder set determined by the first letter of ω. Then the inverse branch of σ determined by the word ω is the map

$$\sigma_\omega^{-1} : B(\sigma(\omega), 1) \to B(\omega, 1)$$

with the property that $\sigma_\omega^{-1}(\sigma(\omega)) = \omega$, that is,

$$\begin{array}{ccccc} \sigma_\omega^{-1} & : & [\omega_2] & \longrightarrow & [\omega_1] \\ & & \tau & \longmapsto & \omega_1\tau. \end{array}$$

In fact, note that $\sigma_\omega^{-1} : [\omega_2] \to [\omega_1\omega_2]$. Similarly, the inverse branch of the nth iterate of σ determined by the word ω is the map which adds to each word $\tau \in [\omega_{n+1}]$ a prefix (or initial block) consisting of the first n letters of ω, that is,

$$\begin{array}{ccccc} \sigma_\omega^{-n} & : & [\omega_{n+1}] & \longrightarrow & [\omega_1\omega_2\ldots\omega_n\omega_{n+1}] \\ & & \tau & \longmapsto & \omega_1\omega_2\ldots\omega_n\tau. \end{array}$$

The next lemma states that inverse branches determined by distinct points have disjoint images.

Lemma 4.2.8. *Let $z \in X$. Let $n \in \mathbb{N}$ and $x, y \in T^{-n}(z)$ with $x \neq y$. Then*

$$T_x^{-n}(B(z,\xi)) \cap T_y^{-n}(B(z,\xi)) = \emptyset.$$

Proof. Let $1 \leq k \leq n$ be the smallest integer such that $T^k(x) = T^k(y)$ and further let $w := T^k(x) = T^k(y)$. By Remark 4.2.5, we deduce that

$$\begin{aligned} T_x^{-n}(B(z,\xi)) &= T_x^{-k} \circ T_{T^k(x)}^{-(n-k)}(B(z,\xi)) \\ &\subseteq T_x^{-k}(B(T^k(x),\xi)) = T_x^{-k}(B(w,\xi)). \end{aligned}$$

Likewise,

$$T_y^{-n}(B(z,\xi)) \subseteq T_y^{-k}(B(w,\xi)).$$

It is thus sufficient to show that the sets $T_x^{-k}(B(w,\xi))$ and $T_y^{-k}(B(w,\xi))$ are disjoint. To shorten notation in what follows, set

$$w_x := T^{k-1}(x) \text{ and } w_y := T^{k-1}(y).$$

Then

$$T_x^{-k} = T_x^{-(k-1)} \circ T_{w_x}^{-1} \text{ and } T_y^{-k} = T_y^{-(k-1)} \circ T_{w_y}^{-1}. \tag{4.34}$$

Moreover, by the definition of k and the hypothesis that $x \neq y$, we have that $T(w_x) = T(w_y) = w$ and also that $w_x \neq w_y$. Consequently, in light of Remark 4.1.2(b), it follows that $d(w_x, w_y) \geq 2\delta$. Recalling that $\xi \leq \delta$, we deduce that

$$B(w_x, \xi) \cap B(w_y, \xi) = \emptyset.$$

According to Remark 4.2.5, we know that

$$T_{w_x}^{-1}(B(w,\xi)) \subseteq B(w_x,\xi) \quad \text{and} \quad T_{w_y}^{-1}(B(w,\xi)) \subseteq B(w_y,\xi).$$

Therefore,

$$T_{w_x}^{-1}(B(w,\xi)) \cap T_{w_y}^{-1}(B(w,\xi)) = \emptyset.$$

It then follows from (4.34) that

$$\begin{aligned} T_x^{-k}(B(w,\xi)) \cap T_y^{-k}(B(w,\xi)) &= T_x^{-(k-1)}(T_{w_x}^{-1}(B(w,\xi))) \cap T_y^{-(k-1)}(T_{w_y}^{-1}(B(w,\xi))) \\ &\subseteq T^{-(k-1)}(T_{w_x}^{-1}(B(w,\xi))) \cap T^{-(k-1)}(T_{w_y}^{-1}(B(w,\xi))) \\ &= T^{-(k-1)}(T_{w_x}^{-1}(B(w,\xi)) \cap T_{w_y}^{-1}(B(w,\xi))) \\ &= T^{-(k-1)}(\emptyset) = \emptyset. \end{aligned}$$

Thus $T_x^{-n}(B(z,\xi)) \cap T_y^{-n}(B(z,\xi)) \subseteq T_x^{-k}(B(w,\xi)) \cap T_y^{-k}(B(w,\xi)) = \emptyset$. □

We now give a description of the preimage of any set of small diameter in terms of the local inverse branches of T or of one of its iterates.

Lemma 4.2.9. *For all $z \in X$, for all $A \subseteq B(z,\xi)$ and for all $n \in \mathbb{N}$, we have that*

$$T^{-n}(A) = \bigcup_{x \in T^{-n}(z)} T_x^{-n}(A).$$

Proof. Fix $z \in X$ and $A \subseteq B(z,\xi)$. Since $T^{-n}(A) \supseteq \bigcup_{x\in T^{-n}(z)} T_x^{-n}(A)$ for all $n \in \mathbb{N}$, we only need to prove the opposite inclusion. We shall do this by induction. As the basis of induction, we first do it for $n = 1$. So, let $w \in T^{-1}(A)$. We aim to show that $w \in \bigcup_{x\in T^{-1}(z)} T_x^{-1}(A)$. Since $T(w) \in A \subseteq B(z,\xi)$, we have that $z \in B(T(w),\xi)$. Now define $x := T_w^{-1}(z) \in T^{-1}(z)$. We shall show that $w \in T_x^{-1}(A)$. Recall that $T_x^{-1} : B(T(x),\xi) \to B(x,\xi)$. Since $T(x) = T(T_w^{-1}(z)) = z$, we then have that $T_x^{-1} : B(z,\xi) \to B(x,\xi)$. As $T(w) \in B(z,\xi)$, the point $w' := T_x^{-1}(T(w))$ is well-defined. Moreover, $w' \in T_x^{-1}(A)$, as $T(w) \in A$. Thus, to see that $w \in T_x^{-1}(A)$, it only remains to show that $w' = w$. We know that $T(w') = T(w)$ by definition of w'. So, according to Remark 4.1.2(b), it suffices to show that $d(w', w) < 2\delta$. Using Proposition 4.2.4, observe that

$$d(w',x) = d(T_x^{-1}(T(w)), T_x^{-1}(T(x))) \leq \lambda^{-1} d(T(w),T(x)) = \lambda^{-1} d(T(w),z) < \lambda^{-1}\xi < \delta$$

and

$$d(w,x) = d(T_w^{-1}(T(w)), T_w^{-1}(z)) \leq \lambda^{-1} d(T(w),z) < \lambda^{-1}\xi < \delta.$$

These last two inequalities combine to give

$$d(w', w) \le d(w', x) + d(x, w) < \delta + \delta = 2\delta.$$

So $w = w' := T_x^{-1}(T(w)) \in T_x^{-1}(A)$. We have thus shown that

$$T^{-1}(A) = \bigcup_{x \in T^{-1}(z)} T_x^{-1}(A).$$

For the sake of the inductive step, suppose that the assertion of our lemma holds for all $n = 1, \ldots, k$. Then

$$\begin{aligned} T^{-(k+1)}(A) = T^{-k}(T^{-1}(A)) = T^{-k}\Big(\bigcup_{x \in T^{-1}(z)} T_x^{-1}(A)\Big) \\ &= \bigcup_{x \in T^{-1}(z)} T^{-k}(T_x^{-1}(A)) \\ &= \bigcup_{x \in T^{-1}(z)} \bigcup_{y \in T^{-k}(x)} T_y^{-k}(T_x^{-1}(A)) \\ &= \bigcup_{v \in T^{-(k+1)}(z)} T_v^{-(k+1)}(A). \end{aligned}$$

This completes the proof. □

We now describe conditions under which a transitive system is very strongly transitive (see Definitions 1.5.14–1.5.15).

Lemma 4.2.10. *Every open, distance expanding and transitive dynamical system $T : X \to X$ is very strongly transitive.*

Proof. Given that T is open and X compact, it suffices to show that T is strongly transitive. Let U be an open subset of X. Let also ξ be as in (4.29). According to Theorem 1.5.11, there exists a point $x \in U$ and a number $0 < \xi' \le \xi$ such that $B(x, \xi') \subseteq U$ and $\overline{\mathcal{O}_+(x)} = X$. From (4.33), we deduce that

$$X = \overline{\mathcal{O}_+(x)} = \bigcup_{n=0}^{\infty} B(T^n(x), \xi') \subseteq \bigcup_{n=0}^{\infty} T^n(B(x, \xi')) \subseteq \bigcup_{n=0}^{\infty} T^n(U).$$

Since U is an arbitrary open set, we conclude that T is strongly transitive. □

4.3 Shadowing

Imagine that you observe the dynamics of a system $T : X \to X$ by means of some instrument which is only accurate up to a given $\alpha \ge 0$. In other words, assume that your instrument can only locate the position of a point with a precision at best α.

Then, with your instrument, you will not be able to distinguish points that are within a distance α from each other. In particular, this means that if a point x_0 lands under the map T within a distance α of a point x_1, then you will not be able to distinguish x_1 from the image of x_0 under T. Similarly, if x_1 lands under the map T within a distance α of a point x_2, then you will not be able to distinguish x_2 from the image of x_1 under T, and so on. To summarize, this sequence (x_i) can be mistaken for the orbit of the point x_0, although, in reality, it is not the orbit of x_0 and, in fact, it is not necessarily an orbit at all. The following definition and terminology make this precise.

Definition 4.3.1. Let $\alpha \geq 0$. A sequence $(x_i)_{i=0}^{n}$, where n can be finite or infinite, is said to be an *α-pseudo-orbit* if

$$d\big(T(x_i), x_{i+1}\big) < \alpha, \quad \forall 0 \leq i < n.$$

In particular, notice that the orbit $\mathcal{O}_+(x)$ of a point $x \in X$ can be written as $\mathcal{O}_+(x) = \{x_0 = x, x_1 = T(x), x_2 = T^2(x), \ldots\}$, which precisely means that $d(T(x_i), x_{i+1}) = 0$ for all $i \in \mathbb{Z}_+$. Thus, an orbit, when converted to the sequence of the iterates of a point, is a 0-pseudo-orbit. Inversely, a 0-pseudo-orbit is merely a sequence of successive iterates of a point.

In the following definition, we come to the important concept of *shadowing* an α-pseudo-orbit, as advertised in the title of this section.

Definition 4.3.2. A point $x \in X$ is said to *β-shadow* a pseudo-orbit $(x_i)_{i=0}^{n}$ if

$$d\big(T^i(x), x_i\big) < \beta, \quad \forall 0 \leq i \leq n.$$

That is, the orbit of x lies within a distance β of the pseudo-orbit $(x_i)_{i=0}^{n}$.

Pseudo-orbits and shadowing (along with the forthcoming closing lemma), form a long lived, important, and convenient way of studying dynamical systems exhibiting some sort of hyperbolic or expanding behavior. At the very least, they can be traced back to the breakthrough work of Anosov and Sinai (see [4, 5]). They found a mature, elegant form in [11]. Our approach follows [57], which in turn is based upon [11].

We shall now prove that any infinite sequence in X can be δ-shadowed by at most one point of the space if the dynamical system $T : X \to X$ under consideration is a map expanding balls of radius δ.

Proposition 4.3.3. *Let $T : X \to X$ be a distance expanding map with δ as a constant delimiting the neighborhoods of expansion. Then every infinite sequence of points $(x_i)_{i=0}^{\infty}$ in X can be δ-shadowed by at most one point of X.*

Proof. Suppose that y and z are two points which each δ-shadow the same sequence $(x_i)_{i=0}^{\infty}$. For all $i \geq 0$, we then have

$$d\big(T^i(y), x_i\big) < \delta \text{ and } d\big(T^i(z), x_i\big) < \delta.$$

Then, by the triangle inequality, for all $i \geq 0$ we have that

$$d(T^i(y), T^i(z)) < 2\delta.$$

By the expanding property of T, we deduce, for all $i \geq 0$, that

$$d(T^{i+1}(y), T^{i+1}(z)) \geq \lambda\, d(T^i(y), T^i(z)).$$

So, by induction, we conclude that

$$d(T^n(y), T^n(z)) \geq \lambda^n d(y,z)$$

for all $n \geq 0$. However, since the compact space X has finite diameter, this can only happen when $y = z$. □

More generally, we have the following result on the existence and uniqueness of shadowing.

Proposition 4.3.4. *Let $T : X \to X$ be an open, distance expanding map. Let $0 < \beta < \xi$, where ξ is as defined in (4.29). Let $\alpha = \min\{\xi, (\lambda - 1)\beta/2\}$ and let $(x_i)_{i=0}^n$ be an α-pseudo-orbit (where n can be finite or infinite). For each $0 \leq i < n$, let $x_i' = T_{x_i}^{-1}(x_{i+1})$. Then:*

(a) *For all $0 \leq i < n$, we have that*

$$T_{x_i'}^{-1}(\overline{B(x_{i+1}, \beta/2)}) \subseteq \overline{B(x_i, \beta/2)}$$

and thus, by induction, the composite map

$$T_{x_0'}^{-1} \circ \cdots \circ T_{x_i'}^{-1} : \overline{B(x_{i+1}, \beta/2)} \to \overline{B(x_0, \beta/2)}$$

is well-defined. Henceforth, we denote this composition by T_i^{-1}.

(b) *$(T_i^{-1}(\overline{B(x_{i+1}, \beta/2)}))_{i=0}^{n-1}$ is a descending sequence of nonempty compact sets.*

(c) *The intersection $\bigcap_{i=0}^{n-1} T_i^{-1}(\overline{B(x_{i+1}, \beta/2)})$ is nonempty and all of its elements β-shadow the α-pseudo-orbit $(x_i)_{i=0}^n$.*

(d) *If $n = \infty$, then $\bigcap_{i=0}^{\infty} T_i^{-1}(\overline{B(x_{i+1}, \beta/2)})$ consists of the unique point that β-shadows the infinite α-pseudo-orbit $(x_i)_{i=0}^{\infty}$.*

Proof. Toward part (a), let $x \in \overline{B(x_{i+1}, \beta/2)}$. Since $x_i' = T_{x_i}^{-1}(x_{i+1})$, we have $T(x_i') = x_{i+1}$ and hence $x_i' = T_{x_i'}^{-1}(T(x_i')) = T_{x_i'}^{-1}(x_{i+1})$. Using Proposition 4.2.4, it follows that

$$\begin{aligned}
d(T_{x_i'}^{-1}(x), x_i) &\leq d(T_{x_i'}^{-1}(x), x_i') + d(x_i', x_i) \\
&= d(T_{x_i'}^{-1}(x), T_{x_i'}^{-1}(x_{i+1})) + d(T_{x_i}^{-1}(x_{i+1}), T_{x_i}^{-1}(T(x_i))) \\
&\leq \lambda^{-1} d(x, x_{i+1}) + \lambda^{-1} d(x_{i+1}, T(x_i)) \\
&< \lambda^{-1}(\beta/2 + \alpha) \leq \lambda^{-1}(\beta/2 + (\lambda - 1)\beta/2) = \beta/2.
\end{aligned}$$

Hence $T_{x_i'}^{-1}(x) \in \overline{B(x_i, \beta/2)}$, and this proves the first assertion.

To prove part (b), notice that $T_i^{-1} = T_{i-1}^{-1} \circ T_{x_i'}^{-1}$ for every $1 \le i < n$. Using part (a), we deduce that

$$T_i^{-1}(\overline{B(x_{i+1}, \beta/2)}) = T_{i-1}^{-1} \circ T_{x_i'}^{-1}(\overline{B(x_{i+1}, \beta/2)}) \subseteq T_{i-1}^{-1}(\overline{B(x_i, \beta/2)}).$$

This proves the second assertion.

To prove part (c), recall that the intersection of a descending sequence of nonempty compact sets is a nonempty compact set. From part (b), we readily obtain that $\bigcap_{i=0}^{n-1} T_i^{-1}(\overline{B(x_{i+1}, \beta/2)}) \neq \emptyset$. Moreover, for all $0 \le j < n$, observe that

$$T^j\left(\bigcap_{i=0}^{n-1} T_i^{-1}(\overline{B(x_{i+1}, \beta/2)})\right) \subseteq T^j(T_j^{-1}(\overline{B(x_{j+1}, \beta/2)})) \subseteq \overline{B(x_{j+1}, \beta/2)}.$$

This implies that $d(T^j(x), x_{j+1}) < \beta$ for all $0 \le j < n$ and $x \in \bigcap_{i=0}^{n-1} T_i^{-1}(\overline{B(x_{i+1}, \beta/2)})$. This proves that every such x β-shadows the α-pseudo-orbit $(x_i)_{i=0}^{n}$. If $n = \infty$, Proposition 4.3.3 guarantees that only one such x exists since $\beta < \xi \le \delta$, and part (d) follows. □

We will deduce several important facts from the preceding proposition. Before doing so, we need another definition.

Definition 4.3.5. A map $T : X \to X$ satisfies the *shadowing property* if for all $\beta > 0$ there exists an $\alpha > 0$ such that every infinite α-pseudo-orbit is β-shadowed by a point of the space X.

Corollary 4.3.6 (Existence and uniqueness of shadowing). *Every open, distance expanding map satisfies the shadowing property. Moreover, if β is small enough (namely, if $\beta < \xi$), then one can choose α so that every infinite α-pseudo-orbit is β-shadowed by one and only one point of the space. In fact, α can be chosen as in Proposition* 4.3.4.

In light of the above corollary, we will say that every open, distance expanding map satisfies the *unique-shadowing property*.

Corollary 4.3.7 (Closing lemma). *For every $\beta > 0$, there exists an $\alpha > 0$ with the following property: If x is a point such that $d(T^n(x), x) < \alpha$ for some $n \in \mathbb{N}$, then there exists a periodic point of period n which β-shadows the orbit $(T^i(x))_{i=0}^{n-1}$. In fact, if $\beta < \xi$ then α can be chosen as in Proposition* 4.3.4.

Proof. First, note that if the property holds for some $\tilde{\beta} > 0$ and a corresponding $\alpha(\tilde{\beta})$, then it also holds for any $\beta \ge \tilde{\beta}$ and $\alpha(\beta) = \alpha(\tilde{\beta})$. Thus we may assume without loss of generality that $0 < \beta < \xi$. Choose α as in Proposition 4.3.4. Now, consider the infinite sequence

$$x, T(x), T^2(x), \ldots, T^{n-1}(x), x, T(x), T^2(x), \ldots, T^{n-1}(x), x, \ldots$$

This sequence can be expressed as $(x_i)_{i=0}^{\infty}$, where $x_{kn+j} = T^j(x)$ for all $k \in \mathbb{Z}_+$ and $0 \le j < n$. We claim that this sequence constitutes an α-pseudo-orbit. Indeed, when $0 \le j < n-1$, we have

$$d(T(x_{kn+j}), x_{kn+j+1}) = d(T(T^j(x)), T^{j+1}(x)) = 0 < \alpha,$$

while when $j = n-1$, we have

$$\begin{aligned} d(T(x_{kn+j}), x_{kn+j+1}) &= d(T(x_{kn+n-1}), x_{kn+n}) = d(T(x_{kn+n-1}), x_{(k+1)n}) \\ &= d(T(T^{n-1}(x)), x) = d(T^n(x), x) < \alpha. \end{aligned}$$

Thus the sequence $(x, T(x), \ldots, T^{n-1}(x), x, T(x), \ldots, T^{n-1}(x), x, T(x), \ldots)$ is an α-pseudo-orbit, and by Corollary 4.3.6 there exists a unique point y which β-shadows it. We also notice that the point $T^n(y)$ β-shadows this infinite sequence, since $d(T^j(T^n(y)), x_j) = d(T^{n+j}(y), x_j) = d(T^{n+j}(y), x_{n+j}) < \beta$. As β-shadowing is unique, we conclude that $T^n(y) = y$. Hence y is a periodic point of period n which β-shadows the orbit $(T^i(x))_{i=0}^{n-1}$. □

From this result, we can infer that any open, distance expanding map has at least one periodic point. Let $\mathrm{Per}(T)$ denote the set of periodic points of T. We shall prove the following.

Corollary 4.3.8 (Closing lemma, existence of a periodic point). *Every open, distance expanding map of a compact metric space has a periodic point. More precisely,* $\mathrm{Per}(T) \subseteq \bigcup_{x\in X} \omega(x) \subseteq \overline{\mathrm{Per}(T)}$, *and as the middle set is nonempty, so is* $\mathrm{Per}(T)$.

Proof. The left-hand side inclusion is immediate as $x \in \omega(x)$ for all $x \in \mathrm{Per}(T)$. In order to prove the right-hand side one, choose any $x \in X$. Recall that the set $\omega(x)$, which is nonempty since X is compact, is the set of accumulation points of the sequence $(T^n(x))_{n=0}^{\infty}$ of iterates of x. Let $y \in \omega(x)$. Fix momentarily an arbitrary $\beta > 0$ and let $\alpha := \alpha(\beta) > 0$ be as in the closing lemma. Then there exists a subsequence $(T^{n_k}(x))_{k=0}^{\infty}$ such that $d(T^{n_k}(x), y) < \alpha/2$ for all $k \in \mathbb{N}$. Therefore, $d(T^{n_k}(x), T^{n_j}(x)) < \alpha$ for all $j, k \in \mathbb{N}$. Fix k and let $j := k+1$. Further, define $z := T^{n_k}(x)$. Then

$$d(T^{n_{k+1}-n_k}(z), z) = d(T^{n_{k+1}}(x), T^{n_k}(x)) < \alpha.$$

According to the closing lemma, there then exists a periodic point w of period $n_{k+1} - n_k$ which β-shadows the orbit $(T^i(z))_{i=0}^{n_{k+1}-n_k-1} = (T^i(x))_{i=n_k}^{n_{k+1}-1}$. Then

$$d(w, y) \le d(w, z) + d(z, y) \le \beta + \alpha/2.$$

This means that there is a periodic point at a distance at most $\beta + \alpha/2$ from y. As β tends to zero, we also have that α tends to zero. Hence, the point y belongs to the closure of the set of periodic points. □

As an immediate consequence of this corollary, we obtain the following result.

Corollary 4.3.9 (Density of periodic points). *The set of periodic points of an open, distance expanding map $T : X \to X$ of a compact metric space X is dense if and only if $\overline{\bigcup_{x\in X} \omega(x)} = X$.*

From the definition of transitivity (cf. Definition 1.5.6), we also obtain the following.

Corollary 4.3.10 (Density of periodic points for transitive maps). *For every transitive, open, distance expanding map of a compact metric space, the set of periodic points is dense.*

In the previous two results, we imposed some restriction on the dynamics of the map. This time we impose some conditions on the space on which the system lives.

Corollary 4.3.11 (Density of periodic points on a connected space). *For every open distance expanding map of a connected compact metric space, the set of periodic points is dense.*

Proof. Fix an arbitrary $x \in X$. We aim to demonstrate that there are periodic points arbitrarily close to x. Let $0 < \beta < \xi$, where ξ was defined in (4.29), and let $\alpha := \alpha(\beta) > 0$ be as in the closing lemma and Proposition 4.3.4. Let $\{U_1, U_2, \ldots, U_p\}$ be a finite open cover of X of diameter less than β (that is, the diameter of each U_i is less than β). Choose any $n \in \mathbb{N}$ such that $(p+1)\lambda^{-n}\beta < \alpha$. Since X is connected, there exists a β-chain of length at most $p+1$ joining x to $T^n(x)$. In other words, there exists a finite sequence

$$x =: y_0, y_1, \ldots, y_{k-1}, y_k := T^n(x)$$

such that $d(y_j, y_{j+1}) < \beta$ for each $0 \le j < k$, where $k \le p$. The elements of the β-chain are chosen to be such that $y_j, y_{j+1} \in U_{i_j}$ for all $0 \le j < k$. By applying an appropriately chosen inverse branch of T^n to this chain, we can construct a $(\lambda^{-n}\beta)$-chain of length at most $p+1$ ending at x. Indeed, let $y_k^{(n)} = T_x^{-n}(T^n(x)) = x$. By recursion on j from $k-1$ to 0, we define $y_j^{(n)} = T^{-n}_{y_{j+1}^{(n)}}(y_j)$. This results in the finite sequence

$$y_0^{(n)} = T^{-n}_{y_1^{(n)}}(y_0), \ldots, y_{k-1}^{(n)} = T^{-n}_{y_k^{(n)}}(y_{k-1}) = T_x^{-n}(y_{k-1}), y_k^{(n)} = x.$$

Observe that for all $0 \le j < k$ we have

$$\begin{aligned} d(y_j^{(n)}, y_{j+1}^{(n)}) &= d\big(T^{-n}_{y_{j+1}^{(n)}}(y_j), T^{-n}_{y_{j+1}^{(n)}}(T^n(y_{j+1}^{(n)}))\big) \\ &\le \lambda^{-n} d(y_j, T^n(y_{j+1}^{(n)})) \\ &= \lambda^{-n} d(y_j, y_{j+1}) \\ &< \lambda^{-n}\beta. \end{aligned}$$

Thus we have defined a $(\lambda^{-n}\beta)$-chain of length at most $p+1$ ending at x. Consequently, by the triangle inequality, we deduce that

$$d(y_0^{(n)}, x) = d(y_0^{(n)}, y_k^{(n)}) \leq (k+1)\lambda^{-n}\beta \leq (p+1)\lambda^{-n}\beta < \alpha.$$

Note also that $T^n(y_0^{(n)}) = y_0 = x$. It follows from these last two facts that the infinite sequence

$$y_0^{(n)}, T(y_0^{(n)}), \ldots, T^{n-1}(y_0^{(n)}), T^n(y_0^{(n)}) = x, y_0^{(n)}, T(y_0^{(n)}), \ldots, T^{n-1}(y_0^{(n)}), x, \ldots$$

is an infinite α-pseudo-orbit. Then, according to Proposition 4.3.4(d), there is a unique point z that β-shadows this pseudo-orbit. However, $T^n(z)$ also β-shadows this pseudo-orbit. Thus z is a periodic point of period n. This implies in particular that there exists a periodic point which is β-close to x. As $0 < \beta < \xi$ was chosen arbitrarily, we deduce that the point x is a periodic point or a point of accumulation of periodic points. As x was chosen arbitrarily in X, we conclude that the periodic points of T are dense in X. □

Note that there exist open distance expanding maps defined upon disconnected compact metric spaces whose set of periodic points is not dense (see Exercise 4.6.6).

4.4 Markov partitions

As was alluded to in Chapter 3, symbolic dynamical systems are often used to “represent” other dynamical systems. In the remainder of this chapter, we shall show that an open, expanding map $T : X \to X$ of a compact metric space X can be represented by a subshift of finite type $\sigma : F \to F$, where $F \subseteq E^\infty$ for some finite set E.

In general, one cannot expect that T and σ be topologically conjugate. For instance, T might act on a connected space X, whereas σ always acts on a totally disconnected subshift $F \subseteq E^\infty$. As continuous maps preserve connectedness, it is then out of the question that $\sigma : F \to F$ be a factor of $T : X \to X$, let alone that σ and T be topologically conjugate. In general, the best we may hope for is that T be a factor of σ and that most points of X, ideally points which form a dense G_δ-subset of X, be represented by a unique symbolic point ω in F. Ideally, F would be a subshift of finite type, that is, F would be of the form E_A^∞ for some incidence/transition matrix A. This turns out to be possible.

The construction of such representations can be roughly described as follows. Cover the space X with some special finite collection $\mathcal{R} = \{R_1, R_2, \ldots, R_p\}$ of subsets of X. The collection $\mathcal{R}$ will be called a *Markov “partition”*. We will shortly give some justification for the conditions imposed on a Markov “partition”, but first we outline how the “partition” can be used to generate symbolic representations of points in X. The orbit of any point $x \in X$ may be tracked by recording the members of $\mathcal{R}$ in which each iterate of x lands. We may thereby associate to each point x at least one symbolic point $\omega = (\omega_i)_{i=0}^{\infty} \in E^{\infty} := \{1, 2, \ldots, p\}^{\infty}$ such that

$$T^i(x) \in R_{\omega_i}, \quad \forall i \geq 0.$$

Equivalently, this can be expressed by requiring that

$$x \in \bigcap_{i=0}^{\infty} T^{-i}(R_{\omega_i}).$$

However, it is often possible to associate more than one symbolic point to a given point x in this way. For instance, this occurs whenever the orbit of x falls in the nonempty intersection of two members of the Markov "partition" $\mathcal{R}$. We immediately obtain at least two representatives for such an x. In order to achieve a one-to-one association on as large a subset of X as possible (ideally on a dense G_δ-subset of X), we require that the sets R_j intersect as little as possible. Namely, we require that they only intersect within their boundaries, if they intersect at all. Recall from topology that the boundary of a closed set is a nowhere dense set. This justifies condition (b) in the definition of a Markov partition below.

On the other hand, in order that sets of the form $\bigcap_{i=0}^{\infty} T^{-i}(R_{\omega_i})$ each generate at most one point of X, we require that the sets R_j be "small," in some sense. For open expanding maps, this means that the diameters of the R_j should be small enough that the inverse branches of T^i be defined on them (i. e., they should be of diameter less than ξ), so that these inverse branches contract the R_j by a factor λ^{-i}. Moreover, to track the entire orbit of a point x, we usually track its first n iterates and then "take the limit" as n tends to infinity to track the entire orbit. This means that, should the finite intersection $\bigcap_{i=0}^{n} T^{-i}(R_{\omega_i})$ be nonempty for each $n \geq 0$, we would like the infinite intersection $\bigcap_{i=0}^{\infty} T^{-i}(R_{\omega_i})$ to be nonempty. This can be guaranteed by requiring that the R_j be closed or, equivalently, compact.

All of the above requirements can be fulfilled in any compact metric space X. Indeed, it is not too difficult to construct a finite cover $\mathcal{R} = \{R_1, R_2, \ldots, R_p\}$ consisting of closed sets of diameters as small as desired and which intersect only in their boundaries (see Exercise 4.6.10). Although the association $x \mapsto \omega$ is one-to-one on a dense set in X and the closure F of the symbolic points hence generated is a subshift of $E^{\infty} := \{1, 2, \ldots, p\}^{\infty}$, this subshift is generally not of finite type. To ensure that F be of finite type, we impose condition (c) in the definition of a Markov partition. Moreover, condition (a) ensures that closed sets R_j with the property that $R_j = \partial R_j$ are not added to the partition, since such sets only provide information about the dynamics of a negligible set of points of X.

Note that Markov partitions are generally not partitions of the space in the usual sense of disjoint sets, as this would imply that the space is disconnected.

Definition 4.4.1. A finite collection of closed sets $\mathcal{R} = \{R_1, R_2, \ldots, R_p\}$, which covers the space X is called a *Markov partition* for a dynamical system $T : X \to X$ if it satisfies the following three conditions:
(a) $R_i = \overline{\mathrm{Int}(R_i)}$ for all $1 \le i \le p$.
(b) $\mathrm{Int}(R_i) \cap \mathrm{Int}(R_j) = \emptyset$ for all $i \neq j$.
(c) If $T(\mathrm{Int}(R_i)) \cap \mathrm{Int}(R_j) \neq \emptyset$, then $T(R_i) \supseteq R_j$.

In other words, condition (a) requires that each element of a Markov partition be the closure of its interior, condition (b) states that the elements of a Markov partition can only intersect on their boundaries, and condition (c) states that if the image of the interior of an element R_i intersects the interior of an element R_j, then the image of R_i completely covers R_j. Since T is an open map, note that condition (b) is equivalent to
(b′) $R_i \cap \mathrm{Int}(R_j) = \emptyset$ for all $i \neq j$.

For the same reason, condition (c) can be replaced by either of the following conditions:
(c′) If $T(R_i) \cap \mathrm{Int}(R_j) \neq \emptyset$, then $T(R_i) \supseteq R_j$.
(c″) If $T(\mathrm{Int}(R_i)) \cap R_j \neq \emptyset$, then $T(R_i) \supseteq R_j$.

Example 4.4.2. Let T be the shift map $\sigma : E_A^\infty \to E_A^\infty$. Then the one-cylinders $\{[e]\}_{e \in E}$ form a Markov partition for σ. Indeed, every one-cylinder is both open and closed, and hence satisfies condition (a) of the definition of a Markov partition. Condition (b) is clearly satisfied, since words which begin with different letters are distinct. Finally, $\sigma([f]) \cap [e] \neq \emptyset$ means that $A_{fe} = 1$, that is, fe is an admissible word. Therefore, $\sigma([f]) \supseteq \sigma([fe]) = [e]$. Thus condition (c) is satisfied.

Example 4.4.3. Once again, let T be the shift map $\sigma : E_A^\infty \to E_A^\infty$. Fix any $n \in \mathbb{N}$. Then the n-cylinders $\{[\omega]\}_{\omega \in E_A^n}$ form a Markov partition for σ. The proof of this fact is similar to the one given in Example 4.4.2.

Example 4.4.4. Fix $m \in \mathbb{N}$. Recall the map $T_m(x) := mx \pmod 1$ from Example 1.1.3(b). The collection of closed intervals

$$\left\{ R_i = \left[\frac{i}{m}, \frac{i+1}{m} \right] : 0 \le i < m \right\}$$

is a Markov partition for T_m. Indeed, one can immediately verify that the first two conditions are satisfied. Concerning condition (c'), observe that $T_m(R_i) = \mathbb{S}^1$ for all i, and thus condition (c') is fulfilled. Indeed, $T_m(R_i) \supseteq R_j$ for all $0 \le i, j < m$.

Example 4.4.5. Fix $m \in \mathbb{N}$ and consider again the map $T_m(x) := mx \pmod 1$. Now, fix $k \in \mathbb{N}$. The collection of closed intervals

$$\left\{ R_i = \left[\frac{i}{m^k}, \frac{i+1}{m^k} \right] : 0 \le i < m^k \right\}$$

is another Markov partition for T_m. Exactly as in the previous example, the first two conditions are clearly satisfied. Concerning condition (c'), observe that

$$T_m(R_i) = \left[\frac{mi}{m^k}, \frac{m(i+1)}{m^k}\right] = \left[\frac{mi}{m^k}, \frac{mi+m}{m^k}\right] = \bigcup_{j=mi}^{mi+m-1}\left[\frac{j}{m^k}, \frac{j+1}{m^k}\right]$$

for all i, and thus condition (c') is fulfilled.

We now present the main result of this section. Examples 4.4.3 and 4.4.5 show that the shift map and the m-times maps T_m, for all $m \in \mathbb{N}$, admit arbitrarily small Markov partitions. This is the case for all open, distance expanding maps, as we show in the next theorem. Part of the proof given here is due to David Simmons.

Theorem 4.4.6 (Existence of Markov partitions). *Every open, distance expanding map $T : X \to X$ of a compact metric space X admits Markov partitions of arbitrarily small diameters.*

Proof. Since T is an open, distance expanding map, it follows from Corollary 4.3.6 that T has the unique-shadowing property. Choose $0 < \beta < \xi/8$. Then there exists $\alpha > 0$ such that every α-pseudo-orbit is β-shadowed by exactly one point of X. As T is continuous on a compact metric space, it is uniformly continuous. Therefore, we can choose $0 < \gamma < \min(\beta, \alpha/2)$ such that for all $x_1, x_2 \in X$ with $d(x_1, x_2) < \gamma$, we know that

$$d\big(T(x_1), T(x_2)\big) < \alpha/2.$$

Step 1. Establishment of a factor map φ between a subshift (Ω, σ) and (T, X).

The collection $\{B(x,\gamma) : x \in X\}$ is an open cover of the compact space X. Therefore, there exists a finite set $E \subseteq X$ such that

$$X = \bigcup_{a \in E} B(a, \gamma).$$

Define the space Ω by

$$\Omega := \big\{\omega = (\omega_i)_{i=0}^{\infty} \in E^{\infty} : d\big(T(\omega_i), \omega_{i+1}\big) < \alpha \text{ for all } i \geq 0\big\}.$$

Observe that $\sigma(\Omega) \subseteq \Omega$. Hence, (Ω, σ) is a subshift of E^{∞} according to Theorem 3.2.4. By definition, each element of the space Ω is an α-pseudo-orbit and, therefore, for each $\omega \in \Omega$ there exists a unique point whose orbit β-shadows ω. Let us call this point $\varphi(\omega)$. In this way, we define a map $\varphi : \Omega \to X$, and by uniqueness of shadowing we have that

$$\varphi \circ \sigma = T \circ \varphi. \tag{4.35}$$

In order for φ to be a factor map, we hope that φ is continuous and surjective. Let us first show that φ is continuous. Let $\omega, \tau \in \Omega$. As $\varphi(\omega)$ β-shadows ω, we have that

$d(T^i(\varphi(\omega)), \omega_i) < \beta$ for all $i \geq 0$. Similarly, $d(T^i(\varphi(\tau)), \tau_i) < \beta$ for all $i \geq 0$. Since $\omega_i = \tau_i$ for all $0 \leq i < |\omega \wedge \tau|$, we can apply the triangle inequality to obtain that

$$d\big(T^i(\varphi(\omega)), T^i(\varphi(\tau))\big) < 2\beta < \xi \leq \delta, \quad \forall 0 \leq i < |\omega \wedge \tau|,$$

where $\delta > 0$ comes from the definition of T being expanding. So,

$$d\big(T^{i+1}(\varphi(\omega)), T^{i+1}(\varphi(\tau))\big) \geq \lambda\, d\big(T^i(\varphi(\omega)), T^i(\varphi(\tau))\big), \quad \forall 0 \leq i < |\omega \wedge \tau|.$$

It follows by a straightforward argument that

$$\begin{aligned} d\big(\varphi(\omega), \varphi(\tau)\big) &\leq \lambda^{-|\omega\wedge\tau|} \operatorname{diam}(X) = \big(s^{|\omega\wedge\tau|}\big)^{-\frac{\log\lambda}{\log s}} \operatorname{diam}(X) \\ &= \operatorname{diam}(X)\big(d_s(\omega, \tau)\big)^{-\frac{\log\lambda}{\log s}}. \end{aligned}$$

Thus φ is Hölder continuous with exponent $-\log\lambda/\log s$ and is therefore continuous.

In order to show that φ is surjective, let $x \in X$. Then, for all $i \geq 0$, we have that $T^i(x) \in B(\omega_i, \gamma)$ for some $\omega_i \in E$. As $d(T^i(x), \omega_i) < \gamma$, it follows from the choice of γ that $d(T^{i+1}(x), T(\omega_i)) < \alpha/2$ for all $i \geq 0$. We deduce that

$$d\big(T(\omega_i), \omega_{i+1}\big) \leq d\big(T(\omega_i), T^{i+1}(x)\big) + d\big(T^{i+1}(x), \omega_{i+1}\big) < \alpha/2 + \gamma < \alpha$$

for all $i \geq 0$. Thus $\omega = \omega_0\omega_1\omega_2\ldots \in \Omega$ and, by construction, x γ-shadows ω, that is, $\varphi(\omega) = x$. The proof of the surjectivity of φ is complete.

Step 2. A property of the images of one-cylinders.

For each $a \in E$, define the sets

$$P_a := \varphi\big([a]\big) = \varphi\big(\{\omega \in \Omega : \omega_0 = a\}\big).$$

All sets P_a are closed in X since they are the images under the factor map φ of the one-cylinder sets $[a]$, which are themselves closed in the compact space Ω. For each $a \in E$, set

$$W(a) := \big\{b \in E : d\big(T(a), b\big) < \alpha\big\}.$$

We claim that the following property is satisfied:

$$T(P_a) = \bigcup_{b \in W(a)} P_b. \tag{4.36}$$

Indeed, if $x \in P_a$, then $x = \varphi(\omega)$ for some $\omega \in \Omega$ with $\omega_0 = a$. By the definition of Ω, it follows that $\omega_1 \in W(a)$. Thus, invoking (4.35), we obtain that $T(x) = T(\varphi(\omega)) = \varphi(\sigma(\omega)) \in \varphi([\omega_1])$, and hence $T(x) \in P_{\omega_1} \subseteq \bigcup_{b\in W(a)} P_b$. Consequently, $T(P_a) \subseteq \bigcup_{b\in W(a)} P_b$. Conversely, let $y \in P_b$ for some $b \in W(a)$. Then $y = \varphi(\omega)$ for some

$\omega \in \Omega$ with $\omega_0 = b$. By the definition of $W(a)$, the concatenation $a\omega$ belongs to the set Ω and therefore, using (4.35) again, we get that

$$y = \varphi(\omega) = \varphi(\sigma(a\omega)) = T(\varphi(a\omega)) \in T(P_a).$$

Consequently, $\bigcup_{b \in W(a)} P_b \subseteq T(P_a)$. Relation (4.36) has been proved. This relation can be expressed by means of an incidence/transition matrix $A : E \times E \to \{0,1\}$ by setting

$$A_{ab} := \begin{cases} 1 & \text{if } T(P_a) \supseteq P_b \\ 0 & \text{if } T(P_a) \not\supseteq P_b. \end{cases}$$

Then (4.36) means that

$$T(P_a) = \bigcup_{\{b \in E : A_{ab} = 1\}} P_b. \tag{4.37}$$

It is also worth observing that $P_a \subseteq B(a, \beta)$ for every $a \in E$. Thus, if $P_a \cap P_b \neq \emptyset$ for some $a, b \in E$, then $P_a \cup P_b \subseteq B(a, 4\beta) \cap B(b, 4\beta)$. Since $4\beta < \xi/2$, the restriction $T : B(a, 4\beta) \to T(B(a, 4\beta))$ is a homeomorphism. In particular, T is injective on $P_a \cup P_b$.

Step 3. Construction of the elements of a Markov partition.

For each nonempty subset S of E, define B_S to be

$$B_S := \Big[\bigcap_{a \in S} P_a\Big] \cap \Big[\bigcap_{b \in E \setminus S} (X \setminus P_b)\Big].$$

We claim that the family

$$\mathcal{R} := \{R_S := \overline{\operatorname{Int}(\overline{B_S})} : S \in \mathcal{P}_+(E)\}$$

forms a Markov partition, where $\mathcal{P}_+(E) := \{S \subseteq E : R_S \neq \emptyset\}$.

First, we shall show that for each nonempty subset $S \subseteq E$, the set $T(B_S)$ is a union of elements of $\mathcal{R}$. Toward this end, fix $S \in \mathcal{P}_+(E)$, pick $a_S \in S$ and define

$$E_{a_S} := \{e \in E : P_e \cap P_{a_S} \neq \emptyset\}.$$

Note that $S \subseteq E_{a_S}$. Indeed, $S \in \mathcal{P}_+(E)$ means that $R_S \neq \emptyset$. This implies that $B_S \neq \emptyset$, which in particular implies that $\bigcap_{a \in S} P_a \neq \emptyset$. It ensues that $P_a \cap P_{a_S} \neq \emptyset$ for every $a \in S$.

Moreover, recall that T is injective on $P_a \cup P_b$ whenever $P_a \cap P_b \neq \emptyset$. Then

$$\begin{aligned} T(B_S) &= T\bigg(\Big[\bigcap_{a \in S} P_a\Big] \cap \Big[\bigcap_{b \in E \setminus S} (X \setminus P_b)\Big]\bigg) \\ &= T\bigg(\Big[P_{a_S} \cap \bigcap_{a \in S} P_a\Big] \cap \Big[P_{a_S} \cap \bigcap_{b \in E \setminus S} (X \setminus P_b)\Big]\bigg) \end{aligned}$$

$$
\begin{aligned}
&= T\Big(P_{a_S} \cap \bigcap_{a\in S} P_a\Big) \cap T\Big(P_{a_S} \cap \bigcap_{b\in E\setminus S} (X \setminus P_b)\Big) \\
&= T\Big(\bigcap_{a\in S} P_{a_S} \cap P_a\Big) \cap T\Big(\bigcap_{b\in E\setminus S} P_{a_S} \cap (X \setminus P_b)\Big) \\
&= T\Big(\bigcap_{a\in S} P_{a_S} \cap P_a\Big) \cap T\Big(\bigcap_{b\in E\setminus S} (P_{a_S} \setminus P_b)\Big) \\
&= \Big[\bigcap_{a\in S} T(P_{a_S} \cap P_a)\Big] \cap \Big[\bigcap_{b\in E_{a_S}\setminus S} T(P_{a_S} \setminus P_b)\Big] \\
&\quad \cap \Big[\bigcap_{b\in E\setminus E_{a_S}} T(P_{a_S} \setminus P_b)\Big] \\
&= \Big[\bigcap_{a\in S} \big(T(P_{a_S}) \cap T(P_a)\big)\Big] \cap \Big[\bigcap_{b\in E_{a_S}\setminus S} \big(T(P_{a_S}) \setminus T(P_b)\big)\Big] \\
&\quad \cap \Big[\bigcap_{b\in E\setminus E_{a_S}} T(P_{a_S})\Big] \\
&= \Big[\bigcap_{a\in S} T(P_a)\Big] \cap \Big[\bigcap_{b\in E_{a_S}\setminus S} \big(X \setminus T(P_b)\big)\Big] \\
&= \Big[\bigcap_{a\in S} \bigcup_{\{c\in E\,:\,A_{ac}=1\}} P_c\Big] \cap \Big[\bigcap_{b\in E_{a_S}\setminus S} \bigcap_{\{d\in E\,:\,A_{bd}=1\}} (X \setminus P_d)\Big] \\
&= \Big[\bigcap_{a\in S} \bigcup_{\{c\in E\,:\,A_{ac}=1\}} P_c\Big] \cap \Big[\bigcap_{b\in \widehat{S}^c} (X \setminus P_b)\Big],
\end{aligned}
\tag{4.38}
$$

where $\widehat{S}^c := \bigcup_{b\in E_{a_S}\setminus S}\{d \in E : A_{bd} = 1\}$.

Now, let $x \in T(B_S)$ be arbitrary. Define

$$S(x) := \{b \in E \setminus \widehat{S}^c : x \in P_b\}.$$

Observe that if $e \in S(x)$, then $x \in P_e$. However, if $e \notin S(x)$ then $e \in \widehat{S}^c$ or $x \in X \setminus P_e$. In the former case, there exists $a \in E_{a_S} \setminus S$ such that $A_{ae} = 1$. This implies that $T(P_a) \supseteq P_e$. We will now show that $x \in X \setminus P_e$ also in this case. By way of contradiction, suppose that $x \in P_e$. Then there exists $y \in P_a$ such that $T(y) = x$. On the other hand, since $x \in T(B_S)$, there exists $z \in \bigcap_{i\in S} P_i \cap \bigcap_{j\in E\setminus S}(X \setminus P_j)$ such that $T(z) = x$. As $a_S \in S$, we know that $z \in P_{a_S}$. Thus we have $y, z \in P_a \cup P_{a_S}$ with $T(y) = x = T(z)$. As $a \in E_{a_S}$, we get $P_a \cap P_{a_S} \neq \emptyset$, and hence T is injective on $P_a \cup P_{a_S}$. We deduce that $y = z$. As $a \notin S$, we have $z \in X \setminus P_a$ by definition of z. So $y = z \in P_a \cap (X \setminus P_a) = \emptyset$. This contradiction shows that $x \in X \setminus P_e$. Thus, in either case, if $e \notin S(x)$, then $x \in X \setminus P_e$.

In summary, if $e \in S(x)$ then $x \in P_e$ whereas if $e \notin S(x)$ then $x \in X \setminus P_e$. Consequently,

$$x \in \left[\bigcap_{i \in S(x)} P_i\right] \cap \left[\bigcap_{j \in E \setminus S(x)} (X \setminus P_j)\right] = B_{S(x)}. \tag{4.39}$$

Next, we claim that $B_{S(x)} \subseteq T(B_S)$. Indeed, since $x \in T(B_S)$, it follows from (4.38) that for every $i \in S$, there exists $j_i \in E$ such that $A_{ij_i} = 1$ and $x \in P_{j_i}$. Since $x \in B_{S(x)}$ by (4.39), we deduce that $j_i \in S(x)$ and $P_{j_i} \supseteq B_{S(x)}$. Hence,

$$\bigcap_{i \in S} P_{j_i} \supseteq B_{S(x)}. \tag{4.40}$$

Since $\widehat{S}^c \subseteq E \setminus S(x)$ (by definition of $S(x)$), we have that

$$B_{S(x)} \subseteq \bigcap_{j \in E \setminus S(x)} (X \setminus P_j) \subseteq \bigcap_{j \in \widehat{S}^c} (X \setminus P_j).$$

In conjunction with (4.40), we therefore obtain that

$$B_{S(x)} \subseteq \left[\bigcap_{i \in S} P_{j_i}\right] \cap \left[\bigcap_{j \in \widehat{S}^c} (X \setminus P_j)\right].$$

Thus, according to (4.38),

$$B_{S(x)} \subseteq T(B_S),$$

proving the claim made. It follows immediately that

$$T(B_S) = \bigcup_{x \in T(B_S)} B_{S(x)}.$$

Keep in mind that, though the set $T(B_S)$ generally contains infinitely many points x, the sets $S(x)$ are all subsets of the finite set E and, therefore, there are only finitely many different subsets $S(x)$. Let $\widetilde{S} \subseteq \mathcal{P}(E)$ be the finite set consisting of all different subsets $S(x)$, $x \in T(B_S)$. Then

$$T(B_S) = \bigcup_{Q \in \widetilde{S}} B_Q. \tag{4.41}$$

Since $\overline{\mathrm{Int}(C)} \subseteq \overline{C} = C$ for any closed set C, we have

$$\overline{\mathrm{Int}(R_S)} \subseteq R_S.$$

On the other hand, as $\overline{\mathrm{Int}(Y)} \subseteq \overline{\mathrm{Int}(\overline{\mathrm{Int}(Y)})}$ for any set Y, we have

$$R_S = \overline{\mathrm{Int}(\overline{B_S})} \subseteq \overline{\mathrm{Int}(\overline{\mathrm{Int}(\overline{B_S})})} = \overline{\mathrm{Int}(R_S)}.$$

So, condition (a) of Definition 4.4.1 is satisfied.

To verify condition (b), assume that S_1, S_2 are two nonempty subsets of E such that $S_1 \neq S_2$. Without loss of generality, say there exists $e \in S_1 \setminus S_2$. Then $B_{S_1} \subseteq P_e$ and $B_{S_2} \subseteq X \setminus P_e \subseteq X \setminus \operatorname{Int}(P_e)$. Hence, $R_{S_1} \subseteq \overline{B_{S_1}} \subseteq P_e$, and thus

$$\operatorname{Int}(R_{S_1}) \subseteq \operatorname{Int}(P_e).$$

Also,

$$R_{S_2} \subseteq \overline{B_{S_2}} \subseteq \overline{X \setminus \operatorname{Int}(P_e)} = X \setminus \operatorname{Int}(P_e).$$

Therefore,

$$\operatorname{Int}(R_{S_1}) \cap R_{S_2} = \emptyset, \tag{4.42}$$

which is more than enough to prove condition (b).

Aiming now to prove that condition (c) holds, we will first show that $T(R_S) = \bigcup_{Q \in \widetilde{S}} R_Q$. Using (4.41), the fact that T is a homeomorphism on $B(a_S, \xi) \supseteq P_{a_S} \supseteq \overline{B_S}$ and the fact that $\widetilde{S}$ is a finite set, we obtain that

$$\begin{aligned}
T(R_S) &= T\big(\overline{\operatorname{Int}(\overline{B_S})}\big) = \overline{T\big(\operatorname{Int}(\overline{B_S})\big)} \\
&= \overline{\operatorname{Int}\big(T(\overline{B_S})\big)} = \overline{\operatorname{Int}\big(\overline{T(B_S)}\big)} = \overline{\operatorname{Int}\Big(\overline{\bigcup_{Q \in \widetilde{S}} B_Q}\Big)} \\
&= \overline{\operatorname{Int}\Big(\bigcup_{Q \in \widetilde{S}} \overline{B_Q}\Big)} && (4.43) \\
&\supseteq \overline{\bigcup_{Q \in \widetilde{S}} \operatorname{Int}(\overline{B_Q})} = \bigcup_{Q \in \widetilde{S}} \overline{\operatorname{Int}(\overline{B_Q})} \\
&= \bigcup_{Q \in \widetilde{S}} R_Q. && (4.44)
\end{aligned}$$

On the other hand, if $x \in \overline{\operatorname{Int}(\bigcup_{Q \in \widetilde{S}} \overline{B_Q})}$, then for every open set G containing x, we have $H := G \cap \operatorname{Int}(\bigcup_{Q \in \widetilde{S}} \overline{B_Q}) \neq \emptyset$. This means that H is a nonempty open subset of $\bigcup_{Q \in \widetilde{S}} \overline{B_Q}$. Therefore, by virtue of the Baire category theorem, there exists $Q \in \widetilde{S}$ such that $H \cap \operatorname{Int}(\overline{B_Q}) \neq \emptyset$. Thus $G \cap \operatorname{Int}(\overline{B_Q}) \neq \emptyset$. Taking now the sets G to be open balls centered at x with radii converging to zero and recalling that the set $\widetilde{S}$ is finite, we conclude that there exists $Q_x \in \widetilde{S}$ such that $x \in \overline{\operatorname{Int}(\overline{B_{Q_x}})} = R_{Q_x}$. Hence we have shown that

$$\overline{\operatorname{Int}\Big(\bigcup_{Q \in \widetilde{S}} \overline{B_Q}\Big)} \subseteq \bigcup_{Q \in \widetilde{S}} R_Q.$$

Along with (4.43) and (4.44), this yields

$$T(R_S) = \bigcup_{Q \in \widetilde{S}} R_Q. \tag{4.45}$$

So, if $T(R_S) \cap \mathrm{Int}(R_Z) \neq \emptyset$, there exists $Q \subseteq \widetilde{S}$ such that $R_Q \cap \mathrm{Int}(R_Z) \neq \emptyset$, which, by invoking (4.42), yields that $Q = Z$. Employing (4.45), this gives that

$$T(R_S) \supseteq R_Z.$$

This establishes condition (c′).

It only remains to demonstrate that $\mathcal{R}$ is a cover of X. Indeed, $\{B_S : S \subseteq E\}$ is obviously a cover of X. Hence, $\{\overline{B_S} : S \subseteq E\}$ is also a cover of X. Thus, by the same argument as the one above based on the Baire category theorem, we get that

$$\bigcup_{S\subseteq E} R_S = \bigcup_{S\subseteq E} \overline{\mathrm{Int}(\overline{B_S})} = \overline{\bigcup_{S\subseteq E} \mathrm{Int}(\overline{B_S})} = X.$$

Therefore, $\mathcal{R}$ covers X and we are done. □

4.5 Symbolic representation generated by a Markov partition

Let $T : X \to X$ be an open, distance expanding map of a compact metric space X with constants λ and δ. Let $\mathcal{R} = \{R_1, \ldots, R_p\}$ be a Markov partition with $\mathrm{diam}(\mathcal{R}) < \delta$. This partition induces the alphabet $E := \{1, \ldots, p\}$ and an incidence/transition matrix $A : E \times E \to \{0, 1\}$ defined by

$$A_{ij} := \begin{cases} 1 & \text{if } T(\mathrm{Int}(R_i)) \cap \mathrm{Int}(R_j) \neq \emptyset \\ 0 & \text{otherwise.} \end{cases} \tag{4.46}$$

Let $\sigma : E_A^\infty \to E_A^\infty$ be the subshift of finite type induced by A.

Lemma 4.5.1. *If $\omega \in E_A^\infty$, then $\bigcap_{n=0}^\infty T^{-n}(R_{\omega_n})$ is a singleton.*

Proof. For every $i \in E$, the restriction $T|_{R_i} : R_i \to T(R_i)$ is injective since $\mathrm{diam}(R_i) < \delta$ (cf. Proposition 4.2.1). So the inverse map $T_i^{-1} : T(R_i) \to R_i$ is well-defined and is a contraction with ratio λ^{-1}. Note also that if $A_{ij} = 1$ then $T(\mathrm{Int}(R_i)) \cap \mathrm{Int}(R_j) \neq \emptyset$, and as $\mathcal{R}$ is a Markov partition, $R_j \subseteq T(R_i)$. Then $T^{-1}(R_j) \cap R_i \neq \emptyset$. Consequently, for any set $B \subseteq R_j$, we have

$$T^{-1}(B) \cap R_i = T_i^{-1}(B). \tag{4.47}$$

Now, let $\omega \in E_A^\infty$. We claim that

$$\bigcap_{k=0}^{n+1} T^{-k}(R_{\omega_k}) = T_{\omega_0}^{-1} \circ T_{\omega_1}^{-1} \circ \cdots \circ T_{\omega_n}^{-1}(R_{\omega_{n+1}}), \quad \forall n \geq 0.$$

We shall prove this claim by induction. For the case $n = 0$, we have that $R_{\omega_0} \cap T^{-1}(R_{\omega_1}) = T_{\omega_0}^{-1}(R_{\omega_1})$ using (4.47). Suppose now that the claim holds for $n = 0, \ldots, m$. Using (4.47) again, we obtain

$$\begin{aligned}
\bigcap_{k=0}^{(m+1)+1} T^{-k}(R_{\omega_k}) &= R_{\omega_0} \cap \bigcap_{k=1}^{m+2} T^{-k}(R_{\omega_k}) \\
&= R_{\omega_0} \cap T^{-1}\left(\bigcap_{j=0}^{m+1} T^{-j}(R_{\omega_{j+1}})\right) \\
&= R_{\omega_0} \cap T^{-1}\left(\bigcap_{j=0}^{m+1} T^{-j}(R_{(\sigma(\omega))_j})\right) \\
&= R_{\omega_0} \cap T^{-1}\big(T^{-1}_{(\sigma(\omega))_0} \circ \cdots \circ T^{-1}_{(\sigma(\omega))_m}(R_{(\sigma(\omega))_{m+1}})\big) \\
&= R_{\omega_0} \cap T^{-1}\big(T^{-1}_{\omega_1} \circ \cdots \circ T^{-1}_{\omega_{m+1}}(R_{\omega_{m+2}})\big) \\
&= T^{-1}_{\omega_0}\big(T^{-1}_{\omega_1} \circ \cdots \circ T^{-1}_{\omega_{m+1}}(R_{\omega_{m+2}})\big) \\
&= T^{-1}_{\omega_0} \circ T^{-1}_{\omega_1} \circ \cdots \circ T^{-1}_{\omega_{m+1}}(R_{\omega_{(m+1)+1}}).
\end{aligned}$$

So the claim is proved. The claim also shows that $(\bigcap_{k=0}^{n+1} T^{-k}(R_{\omega_k}))_{n=0}^{\infty}$ is a descending sequence of nonempty compact sets. Using the fact that each T_i^{-1} is contracting with ratio λ^{-1}, we obtain

$$\operatorname{diam}\left(\bigcap_{k=0}^{n+1} T^{-k}(R_{\omega_k})\right) \le \lambda^{-(n+1)} \operatorname{diam}(R_{\omega_{n+1}}) \le \lambda^{-(n+1)}\delta.$$

Since $\lim_{n\to\infty} \lambda^{-(n+1)}\delta = 0$, the set $\bigcap_{n=0}^{\infty} T^{-n}(R_{\omega_n})$ is a singleton. □

Thanks to this lemma, the *coding map* $\pi : E_A^{\infty} \to X$, where $\pi(\omega)$ is defined to be the singleton point in the set

$$\bigcap_{n=0}^{\infty} T^{-n}(R_{\omega_n}), \tag{4.48}$$

is well-defined.

Amongst other properties, we want to show that the coding map is Hölder continuous. Recall that on a compact metric space, a map is Hölder continuous if and only if it is locally Hölder continuous. That is, for a map $f : (Y, d_Y) \to (Z, d_Z)$ with Y compact, it is sufficient to show that there exist constants $\delta > 0$ and $C \ge 0$ such that for every $x, y \in Y$ with $d_Y(x, y) < \delta$, we have that $d_Z(f(x), f(y)) \le C(d_Y(x, y))^{\alpha}$.

Theorem 4.5.2. *The coding map $\pi : (E_A^{\infty}, d_s) \to (X, d)$ satisfies the following properties:*
(a) *The map π is Hölder continuous.*
(b) *The map π is surjective.*
(c) *The restriction of π to $\pi^{-1}(X \setminus \bigcup_{n=0}^{\infty} T^{-n}(\bigcup_{i=1}^{p} \partial R_i))$ is injective. So every point of the forward T-invariant, dense G_δ-set $X \setminus \bigcup_{n=0}^{\infty} T^{-n}(\bigcup_{i=1}^{p} \partial R_i)$ has a unique preimage under π.*

(d) *The map π makes the following diagram commutative:*

$$\begin{tikzcd} E_A^\infty \arrow[r, "\sigma"] \arrow[d, "\pi"'] & E_A^\infty \arrow[d, "\pi"] \\ X \arrow[r, "T"'] & X \end{tikzcd}$$

That is, $\pi \circ \sigma = T \circ \pi$.

In particular, π is a factor map between the symbolic system/representation $\sigma : E_A^\infty \to E_A^\infty$ and the original dynamical system $T : X \to X$.

Note that $X \setminus \bigcup_{n=0}^\infty T^{-n}(\bigcup_{i=1}^p \partial R_i)$ is the set of all points in X whose orbit under T never encounters the boundary of the elements of the Markov partition $\mathcal{R} = \{R_1, R_2, \ldots, R_p\}$.

Proof. In order to shorten the notation, in the following proof we write:

$$Z := X \setminus \bigcup_{n=0}^{\infty} T^{-n}\left(\bigcup_{i=1}^{p} \partial R_i\right).$$

(a) We will prove that π is (locally) Lipschitz continuous with respect to the metric $d_{\lambda^{-1}}$. Recall that $d_{\lambda^{-1}}(\omega, \tau) = \lambda^{-|\omega\wedge\tau|}$. Choose $\omega, \tau \in E_A^\infty$ to be such that $|\omega \wedge \tau| \geq 1$. Therefore,

$$\pi(\omega) \in \bigcap_{n=0}^{|\omega\wedge\tau|-1} T^{-n}(R_{\omega_n}) = \bigcap_{n=0}^{|\omega\wedge\tau|-1} T^{-n}(R_{\tau_n}) \ni \pi(\tau).$$

Thus,

$$\begin{aligned} d(\pi(\omega), \pi(\tau)) &\leq \operatorname{diam}\left(\bigcap_{n=0}^{|\omega\wedge\tau|-1} T^{-n}(R_{\omega_n})\right) \\ &\leq \lambda^{-(|\omega\wedge\tau|-1)} \operatorname{diam}(R_{\omega_{|\omega\wedge\tau|}}) \\ &\leq \lambda^{-|\omega\wedge\tau|} \cdot \lambda \operatorname{diam}(X) = (\lambda \operatorname{diam}(X)) d_{\lambda^{-1}}(\omega, \tau). \end{aligned}$$

So π is Lipschitz continuous (i. e., Hölder continuous with exponent $\alpha = 1$) when E_A^∞ is endowed with the metric $d_{\lambda^{-1}}$. Since the metrics d_s, $s \in (0, 1)$, are Hölder equivalent (see Exercise 3.4.5), we deduce that π is Hölder continuous with respect to any metric d_s.

(b) We now show that π is surjective. For this, it suffices to show that $Z \subseteq \pi(E_A^\infty)$ and that $\overline{Z} = X$. This is because the map $\pi : E_A^\infty \to X$ is continuous, E_A^∞ is compact and so $\pi(E_A^\infty)$ is compact and thereby closed. We shall first demonstrate that Z is a dense G_δ-set in X. Notice that

$$Z = \bigcap_{n=0}^{\infty}\left[X \setminus T^{-n}\left(\bigcup_{i=1}^{p}\partial R_i\right)\right] = \bigcap_{n=0}^{\infty} T^{-n}\left(X \setminus \bigcup_{i=1}^{p}\partial R_i\right) = \bigcap_{n=0}^{\infty} T^{-n}\left(\bigcap_{i=1}^{p}(X \setminus \partial R_i)\right).$$

Now, the boundary of any closed set is a closed, nowhere dense set (for a closed set, "nowhere dense" amounts to saying that the set has empty interior). Thus the complement of any boundary is an open, dense subset of the space. This means that the sets $X \setminus \partial R_i$, $1 \le i \le p$, are all open, dense subsets of X. Consequently, their finite intersection $\bigcap_{i=1}^{p}(X \setminus \partial R_i)$ is an open, dense subset of X. As the preimage of an open, dense set under a continuous map is open and dense, we deduce that $T^{-n}(\bigcap_{i=1}^{p}(X \setminus \partial R_i))$ is an open, dense subset of X for every $n \ge 0$. Hence, Z is a countable intersection of open sets, that is, Z is a G_δ-set. Moreover, as Z is a countable intersection of open, dense subsets of X (a complete metric space), it follows from the Baire category theorem that Z is dense in X.

Second, let us show that $Z \subseteq \pi(E_A^\infty)$. Let $x \in Z$. We shall find an A-admissible word $\rho = (\rho_n)_{n=0}^{\infty}$ such that $\pi(\rho) = x$. For each $n \ge 0$, the letter ρ_n is selected among those letters of the alphabet E in such a way that $x \in T^{-n}(\mathrm{Int}(R_{\rho_n}))$. Thus

$$x \in \bigcap_{n=0}^{\infty} T^{-n}\big(\mathrm{Int}(R_{\rho_n})\big).$$

We show that $\rho \in E_A^\infty$, that is, that $A_{\rho_n\rho_{n+1}} = 1$ for each $n \ge 0$. Indeed,

$$\begin{aligned} A_{\rho_n\rho_{n+1}} = 1 &\Longleftrightarrow T\big(\mathrm{Int}(R_{\rho_n})\big) \cap \mathrm{Int}(R_{\rho_{n+1}}) \neq \emptyset \\ &\Longleftrightarrow \mathrm{Int}(R_{\rho_n}) \cap T^{-1}\big(\mathrm{Int}(R_{\rho_{n+1}})\big) \neq \emptyset. \end{aligned}$$

As $x \in \bigcap_{n=0}^{\infty} T^{-n}(\mathrm{Int}(R_{\rho_n}))$, we know that

$$\begin{aligned} x &\in T^{-n}\big(\mathrm{Int}(R_{\rho_n})\big) \cap T^{-(n+1)}\big(\mathrm{Int}(R_{\rho_{n+1}})\big) \\ &= T^{-n}\big(\mathrm{Int}(R_{\rho_n}) \cap T^{-1}\big(\mathrm{Int}(R_{\rho_{n+1}})\big)\big) \end{aligned}$$

for each $n \ge 0$. Then

$$T^n(x) \in \mathrm{Int}(R_{\rho_n}) \cap T^{-1}\big(\mathrm{Int}(R_{\rho_{n+1}})\big)$$

for all $n \ge 0$. In particular, this intersection is nonempty. So $A_{\rho_n\rho_{n+1}} = 1$ for all $n \ge 0$, and $x = \pi(\rho)$ for some $\rho \in E_A^\infty$.

(c) In the course of the proof of (b), we demonstrated that Z is a dense G_δ-set. The fact that Z is forward T-invariant is obvious since this set consists of all points in X whose orbit under T never encounters the boundary of the elements of the

Markov partition. Thus $T(Z) \subseteq Z$. We now show that $\pi^{-1}(x)$ is a singleton for every $x \in Z$. Suppose that $\omega, \tau \in \pi^{-1}(Z)$ are such that $\pi(\omega) = \pi(\tau)$. As $\{\pi(\omega)\} = \bigcap_{n=0}^{\infty} T^{-n}(R_{\omega_n})$, we have that $\pi(\omega) \in T^{-n}(R_{\omega_n})$ for every $n \geq 0$. On the other hand, since $\pi(\omega) \in Z$, we have $\pi(\omega) \notin T^{-n}(\partial R_{\omega_n})$ for every $n \geq 0$. It therefore follows that $\pi(\omega) \in T^{-n}(\mathrm{Int}(R_{\omega_n}))$ for every $n \geq 0$. Similarly, $\pi(\tau) \in T^{-n}(\mathrm{Int}(R_{\tau_n}))$ for every $n \geq 0$. Since $\pi(\omega) = \pi(\tau) =: x$, we deduce that $T^n(x) \in \mathrm{Int}(R_{\omega_n}) \cap \mathrm{Int}(R_{\tau_n})$ for every $n \geq 0$. Condition (b) imposed on the Markov partition forces $\omega_n = \tau_n$ for each $n \geq 0$, that is, $\omega = \tau$.

(d) Finally, we show that the diagram in statement (d) does indeed commute. Let $\omega \in E_A^\infty$. Then

$$\begin{aligned} T(\{\pi(\omega)\}) = T\Bigg(\bigcap_{n=0}^{\infty} T^{-n}(R_{\omega_n})\Bigg) &\subseteq \bigcap_{n=0}^{\infty} T^{-(n-1)}(R_{\omega_n}) \\ &= T(R_{\omega_0}) \cap \bigcap_{n=1}^{\infty} T^{-(n-1)}(R_{(\sigma(\omega))_{n-1}}) \\ &\subseteq \bigcap_{m=0}^{\infty} T^{-m}(R_{(\sigma(\omega))_m}) = \{\pi(\sigma(\omega))\}. \end{aligned}$$

Since both the left- and the right-hand sides are singletons, equality follows, and $T(\pi(\omega)) = \pi(\sigma(\omega))$ for all $\omega \in E_A^\infty$. That is, $T \circ \pi = \pi \circ \sigma$. □

Example 4.5.3. Consider again the map $T_m : \mathbb{S}^1 \to \mathbb{S}^1$, which is given by $T_m(x) := mx \pmod 1$. We have seen in Example 4.4.4 that the family of closed intervals $\{[\frac{i}{m}, \frac{i+1}{m}] : 0 \leq i < m\}$ forms a Markov partition for T_m. One can then show that the coding map generated by this partition is given, for every $\omega := (\omega_k)_{k=0}^\infty \in \{0, \ldots, m-1\}^\infty$, by

$$\pi(\omega) = \sum_{k=0}^{\infty} \frac{\omega_k}{m^{k+1}}.$$

In particular, if $m = 2$, we obtain the binary coding of each point $x \in \mathbb{S}^1$.

To end this chapter, we express basic properties of Markov partitions in symbolic terms and show that σ inherits some dynamical properties from T.

Lemma 4.5.4. *Let $\mathcal{R} = \{R_1, R_2, \ldots, R_p\}$ be a Markov partition for an open, distance expanding dynamical system $T : X \to X$. Let $E = \{1, 2, \ldots, p\}$, A as in (4.46) and $\sigma : E_A^\infty \to E_A^\infty$ be the symbolic representation of T induced by $\mathcal{R}$. Let also $n \in \mathbb{N}$.*
(a) *For every $i \in E$, it holds that $T^n(R_i) = \bigcup_{j \in E : A_{ij}^n \neq 0} R_j$.*
(b) *$\mathcal{R}$ is a Markov partition for T^n.*
(c) *If T is topologically transitive, then so is σ.*
(d) *If T is topologically exact, then so is σ.*

(e) *Let $\partial\mathcal{R} := \bigcup_{e\in E}\partial R_e$. If T is a local homeomorphism on a neighborhood of each element of $\mathcal{R}$, then the set $\partial\mathcal{R}$ is forward T-invariant while the set $\pi^{-1}(X\setminus\partial\mathcal{R})$ is backward σ-invariant.*

Proof.

(a) Fix $i\in E$. First, assume that $n=1$. Recall that

$$R_\ell = \overline{\mathrm{Int}(R_\ell)}, \quad \forall \ell\in E. \tag{4.49}$$

Since T is continuous and X a compact metric space, it follows that

$$T(R_\ell) = T\big(\overline{\mathrm{Int}(R_\ell)}\big) = \overline{T\big(\mathrm{Int}(R_\ell)\big)}, \quad \forall \ell\in E. \tag{4.50}$$

If $A_{ij}\neq 0$, that is, if $T(\mathrm{Int}(R_i))\cap\mathrm{Int}(R_j)\neq\emptyset$, then $T(R_i)\supseteq R_j$ since $\mathcal{R}$ is a Markov partition. Therefore, $T(R_i)\supseteq\bigcup_{j\in E:A_{ij}\neq 0}R_j$. If it turned out that $T(R_i)\neq\bigcup_{j\in E:A_{ij}\neq 0}R_j$ then we would have $T(R_i)\cap[X\setminus\bigcup_{j\in E:A_{ij}\neq 0}R_j]\neq\emptyset$. By (4.50) and the openness of $X\setminus\bigcup_{j\in E:A_{ij}\neq 0}R_j$, this would imply that $T(\mathrm{Int}(R_i))\cap[X\setminus\bigcup_{j\in E:A_{ij}\neq 0}R_j]\neq\emptyset$. Since $X\setminus\bigcup_{j\in E:A_{ij}\neq 0}R_j\subseteq\bigcup_{k\in E:A_{ik}=0}R_k$, we would deduce that $T(\mathrm{Int}(R_i))\cap[\bigcup_{k\in E:A_{ik}=0}R_k]\neq\emptyset$. By (4.49) and the openness of $T(\mathrm{Int}(R_i))$, it would ensue that $T(\mathrm{Int}(R_i))\cap\mathrm{Int}(R_{\tilde{k}})\neq\emptyset$ for some $\tilde{k}$ such that $A_{i\tilde{k}}=0$. But $T(\mathrm{Int}(R_i))\cap\mathrm{Int}(R_{\tilde{k}})\neq\emptyset$ means that $A_{i\tilde{k}}=1$. This contradiction imposes that

$$T(R_i) = \bigcup_{j\in E\,:\,A_{ij}\neq 0} R_j.$$

This is the basic step in this proof by induction. For the inductive step, suppose that the statement holds for some $n\in\mathbb{N}$, that is, $T^n(R_i)=\bigcup_{j\in E\,:\,A^n_{ij}\neq 0}R_j$. Then

$$\begin{aligned}
T^{n+1}(R_i) &= T(T^n(R_i)) = T\Big(\bigcup_{j\in E\,:\,A^n_{ij}\neq 0} R_j\Big)\\
&= \bigcup_{j\in E\,:\,A^n_{ij}\neq 0} T(R_j) = \bigcup_{j\in E\,:\,A^n_{ij}\neq 0}\ \bigcup_{k\in E\,:\,A_{jk}\neq 0} R_k \qquad (4.51)\\
&= \bigcup_{k\in E\,:\,A^{n+1}_{ik}\neq 0} R_k. \qquad (4.52)
\end{aligned}$$

(b) If $T^n(\mathrm{Int}(R_i))\cap\mathrm{Int}(R_j)\neq\emptyset$, then by (a) there is $k\in E$ such that $A^n_{ik}\neq 0$ and $R_k\cap\mathrm{Int}(R_j)\neq\emptyset$. Since $\mathcal{R}$ is a Markov partition, we infer that $R_k=R_j$, and thus $k=j$. Therefore, $A^n_{ij}\neq 0$ and, by (a) again, we conclude that $T^n(R_i)\supseteq R_j$. Consequently, $\mathcal{R}$ is a Markov partition for T^n.

(c) Let $i,j\in E$. If T is topologically transitive, then there exists $n\in\mathbb{N}$ such that $T^n(\mathrm{Int}(R_i))\cap\mathrm{Int}(R_j)\neq\emptyset$. By (a) and (b), it follows that $A^n_{ij}\neq 0$. Since $i,j\in E$ are arbitrary, this means that the matrix A is irreducible, that is, the shift map σ is transitive according to Theorem 3.2.14.

(d) If T is topologically exact, then for every open subset U of X there exists $N(U) \in \mathbb{N}$ such that $T^{N(U)}(U) = X$. In particular, for every $i \in E$ there is $N(i) \in \mathbb{N}$ such that $T^{N(i)}(\mathrm{Int}(R_i)) = X$. Let $N = \max\{N(i) : i \in E\}$. Then $T^N(\mathrm{Int}(R_i)) = X$ for all $i \in E$. So $T^N(\mathrm{Int}(R_i)) \cap \mathrm{Int}(R_j) \neq \emptyset$ for all $i, j \in E$. By (a) and (b), it follows that $A^N_{ij} \neq 0$ for all $i, j \in E$. This precisely means that $A^N > 0$, and thus the matrix A is primitive, that is, the shift map σ is topologically exact according to Theorem 3.2.16.

(e) For any $i \in E$, given that T is by hypothesis a local homeomorphism on a neighborhood of R_i, it follows from (a) that

$$T(\partial R_i) = \partial T(R_i) = \partial\Big(\bigcup_{j\in E\,:\,A_{ij}\neq 0} R_j\Big) \subseteq \bigcup_{j\in E\,:\,A_{ij}\neq 0} \partial R_j.$$

Consequently,

$$T(\partial\mathcal{R}) = \bigcup_{i\in E} T(\partial R_i) \subseteq \bigcup_{j\in E} \partial R_j = \partial\mathcal{R}.$$

That is, the set $\partial\mathcal{R}$ is forward T-invariant. According to Theorem 4.5.2, the coding map π is a factor map between $\sigma : E^\infty_A \to E^\infty_A$ and $T : X \to X$. By Remark 1.4.4, the set $\pi^{-1}(\partial\mathcal{R})$ is forward σ-invariant. Thus its complement $E^\infty_A \setminus \pi^{-1}(\partial\mathcal{R}) = \pi^{-1}(X \setminus \partial\mathcal{R})$ is backward σ-invariant. □

4.6 Exercises

Exercise 4.6.1. Let $S, T : X \to X$ be two distance expanding maps on a compact metric space (X, d). Show that $S \circ T : X \to X$ is a distance expanding map. Deduce from this that every iterate T^n, $n \in \mathbb{N}$, of T is distance expanding.

Exercise 4.6.2. Suppose that $T : X \to X$ is a continuous map on a compact metric space (X, d) whose nth iterate T^n is distance expanding with constant of expansion λ and constant delimiting the neighborhoods of expansion δ. Prove that T is distance expanding with the same constants λ and δ when X is endowed with the metric d' defined by

$$d'(x, y) = \sum_{k=0}^{n-1} \frac{1}{\lambda^k}\, d(T^k(x), T^k(y)).$$

Show also that the metrics d and d' are topologically equivalent. (Exercises 4.6.1 and 4.6.2 thus prove that T is distance expanding if and only if all its iterates are distance expanding. However, they may be expanding distances with respect to different, though topologically equivalent, metrics.)

Exercise 4.6.3. Inspiring yourself from the proof of Lemma 4.2.2, show that $p > 0$ in the proof of Theorem 4.1.5.

Exercise 4.6.4. Let $(X_n)_{n=0}^{\infty}$ be a descending sequence of compact sets in a metric space (X,d). Let $(\delta_n)_{n=0}^{\infty}$ be a sequence of positive numbers that converges to 0. Show that

$$\bigcap_{n=0}^{\infty} X_n = \bigcap_{n=0}^{\infty} B(X_n,\delta_n).$$

Exercise 4.6.5. Find an open map $T : X \to X$ that admits a subsystem $T|_F : F \to F$ which is not open, where F is a closed forward T-invariant subset of X.

Exercise 4.6.6. Find an open, distance expanding dynamical system $T : X \to X$ defined upon a disconnected, compact metric space X which does not have a dense set of periodic points.

Exercise 4.6.7. Prove that the set of conditions (a), (b), and (c') defines a Markov partition.

Exercise 4.6.8. Let T be the shift map $\sigma : E_A^{\infty} \to E_A^{\infty}$. Fix any $n \in \mathbb{N}$. Prove that the n-cylinders $\{[\omega]\}_{\omega\in E_A^n}$ form a Markov partition for σ.

Exercise 4.6.9. Show that the preimage of an open, dense set under a continuous map is open and dense.

Exercise 4.6.10. Let X be a compact metric space and let $\delta > 0$. Construct a finite cover of X consisting of closed sets of diameters less than δ which intersect only in their boundaries.

Exercise 4.6.11. Let X be an infinite compact metric space and $T : X \to X$ be a transitive, open, distance expanding map. Show that T is not minimal.

Exercise 4.6.12. Let (X,d) be a compact metric space. Define a metric ρ on $X \times \{0,1\}$ by setting

$$\rho\big((x_1,x_2),(y_1,y_2)\big) := d(x_1,y_1) + |x_2 - y_2|.$$

Show that if $T : X \to X$ is a distance expanding map, then the map $\widehat{T} : X \times \{0,1\} \to X \times \{0,1\}$ given by the formula

$$\widehat{T}(x_1,x_2) := \big(T(x_1), x_2 +_2 1\big)$$

is also distance expanding, where $+_2 : \{0,1\} \to \{0,1\}$ denotes addition modulo 2.

Exercise 4.6.13. For every $n \in \mathbb{N}$, show that a distance expanding map can have only finitely many periodic points of period n.

Exercise 4.6.14. Let $T : X \to X$ be an open distance expanding map. Show that the function $x \mapsto \#(T^{-1}(x))$ is locally constant.

Exercise 4.6.15. Suppose that $S, T : X \to X$ are two arbitrary dynamical systems. Define the distance between S and T by the formula

$$d_\infty(S, T) := \sup\{d(S(x), T(x)) : x \in X\}.$$

Now suppose that S and T are both open distance expanding maps with the same parameters δ, λ and ξ. Show that if $d_\infty(S, T) < \min\{\xi, \xi(\lambda - 1)\}$, then S and T are topologically conjugate (cf. the discussion on structural stability at the end of Section 1.2).

5 (Positively) expansive maps

There are various concepts of expansion of a map which have aroused the interest of a great many mathematicians. We have already encountered distance expanding maps; in the present chapter we will introduce positively expansive maps. The study of this class of maps goes back to the 1960s. Such maps are abundant. In particular, all distance expanding maps are expansive, and so, more particularly, all subshifts over a finite alphabet are expansive.

Amidst the large variety of dynamical behaviors which can be thought of as expansion in some sense, expansiveness has turned out to be a rather weak but nevertheless useful mathematical notion. It is a topological concept, in the sense that it is a topological conjugacy invariant. Expansive maps are important for many reasons. One of them is that the entropy function is upper semi-continuous within this class. In particular, all expansive maps admit a measure of maximal entropy and, more generally, have equilibrium states under all continuous potentials.

In Section 5.1, we introduce the concept of expansiveness. In Section 5.2, we define the notion of uniform expansiveness and prove that expansiveness and uniform expansiveness are one and the same notion on compact metrizable spaces. In Section 5.3, we demonstrate that every expansive system is in fact expanding with respect to some metric compatible with the topology on the underlying space. This important fact is due to Coven and Reddy [17] (cf. [18]). It signifies that many of the results proved in Chapter 4, such as the existence of Markov partitions and of a nice symbolic representation, the density of periodic points, the closing lemma, and shadowing, hold for all positively expansive maps. Finally, in Section 5.4 we provide a class of examples of expansive maps that are not expanding. They generate what are called parabolic Cantor sets.

5.1 Expansiveness

The notion of expansiveness was introduced by Utz [73] in 1950 for homeomorphisms. He used the term *unstable* homeomorphisms to describe these maps. Gottschalk and Hedlund [25] later suggested the term *expansive* homeomorphisms, which has been used ever since. Five years after Utz, Williams [78] investigated positive expansiveness of maps. Among other early contributors are Bryant [14], Keynes and Robertson [37], Sears [63], and Coven and Reddy [17].

Definition 5.1.1. A topological dynamical system $T : (X, d) \to (X, d)$ is said to be *(positively) expansive* provided that there exists a constant $\delta > 0$ such that for every $x \neq y$ there is $n = n(x, y) \geq 0$ with

$$d(T^n(x), T^n(y)) > \delta.$$

https://doi.org/10.1515/9783110702682-005

The constant δ is called an *expansive constant* for T, and T is then said to be δ-*expansive*. Equivalently, T is δ-expansive if

$$\sup_{n\geq 0} d\big(T^n(x), T^n(y)\big) \leq \delta \quad \Longrightarrow \quad x = y.$$

In other words, δ-expansiveness means that two forward T-orbits that remain forever within a distance δ from each other originate from the same point, and are therefore only one orbit.

We now note some important and interesting properties of expansiveness.

Remark 5.1.2.
(a) If T is δ-expansive, then T is δ'-expansive for any $0 < \delta' < \delta$.
(b) The expansiveness of T is independent of topologically equivalent metrics, although particular expansive constants generally depend on the metric chosen. See Exercise 5.5.1.
(c) In light of (b), the concept of expansiveness can be defined solely in topological terms, meaning without a reference to a metric. See Exercise 5.5.2.
(d) Unlike the expanding property examined in Chapter 4, expansiveness is a topological conjugacy invariant. See Exercise 5.5.3.

The expansiveness of a system can also be expressed in terms of the following "dynamical" metrics. These metrics are sometimes called Bowen's metrics, since Bowen [10] used them extensively in defining topological entropy for noncompact dynamical systems.

Definition 5.1.3. Let $T : (X, d) \to (X, d)$ be a dynamical system. For every $n \in \mathbb{N}$, let $d_n : X \times X \to [0, \infty)$ be the metric

$$d_n(x, y) := \max\{d\big(T^j(x), T^j(y)\big) : 0 \leq j < n\}.$$

Although the notation does not make explicit the dependence on T, it is crucial to remember that the metrics d_n arise from the dynamics of the system T. It is in this sense that they are dynamical metrics. The corresponding balls will be denoted by

$$B_n(x, \varepsilon) := \{y \in X : d_n(x, y) < \varepsilon\}.$$

Observe that $d_1 = d$ and that for each $x, y \in X$ we have $d_n(x, y) \geq d_m(x, y)$ whenever $n \geq m$. Moreover, it is worth noticing that the metrics d_n, $n \in \mathbb{N}$, are topologically equivalent. Indeed, given a sequence $(x_k)_{k=1}^{\infty}$ in X, the continuity of T ensures that

$$\begin{aligned}\lim_{k\to\infty} d(x_k, y) = 0 &\Longleftrightarrow \lim_{k\to\infty} d\big(T^j(x_k), T^j(y)\big) = 0, \quad \forall 0 \leq j < n,\ \forall n \in \mathbb{N} \\ &\Longleftrightarrow \lim_{k\to\infty} d_n(x_k, y) = 0, \quad \forall n \in \mathbb{N}.\end{aligned}$$

Furthermore, it is easy to see that a dynamical system $T : (X,d) \to (X,d)$ is δ-expansive if and only if

$$\sup_{n\in\mathbb{N}} d_n(x,y) \le \delta \quad \Longrightarrow \quad x = y.$$

Example 5.1.4. Every subshift of finite type $\sigma : E_A^\infty \to E_A^\infty$ is expansive, with any $\delta \in (0,1)$ as an expansive constant. Indeed, note that if $\omega = \omega_1\omega_2\omega_3\ldots$ and $\tau = \tau_1\tau_2\tau_3\ldots$ are any two distinct elements of E_A^∞, then there exists $n \in \mathbb{N}$ such that $\omega_n \ne \tau_n$. Hence,

$$d_s(\sigma^{n-1}(\omega), \sigma^{n-1}(\tau)) = s^{-|\omega_n\omega_{n+1}\ldots\wedge\tau_n\tau_{n+1}\ldots|} = s^0 = 1 > \delta.$$

This example is an instance of the fact that any distance expanding dynamical system is expansive.

Proposition 5.1.5. *Every distance expanding dynamical system is expansive.*

Proof. Let $T : X \to X$ be a distance expanding dynamical system on a compact metric space (X,d), and let δ and λ be constants determining neighborhoods of expansion and magnitude of that expansion, respectively, per Definition 4.1.1. We will show that any $0 < \delta' < 2\delta$ is an expansive constant for T. Let $x, y \in X$ be such that

$$\sup_{n\ge 0} d(T^n(x), T^n(y)) \le \delta'. \tag{5.1}$$

Then

$$d(T^n(x), T^n(y)) \ge \lambda\, d(T^{n-1}(x), T^{n-1}(y)), \quad \forall n \in \mathbb{N}.$$

By induction, it follows that

$$d(T^n(x), T^n(y)) \ge \lambda^n d(x,y), \quad \forall n \ge 0.$$

Therefore,

$$0 \le d(x,y) \le \limsup_{n\to\infty} \lambda^{-n} \mathrm{diam}(X) = 0.$$

Hence $d(x,y) = 0$, that is, $x = y$. So T is δ'-expansive for all $0 < \delta' < 2\delta$. □

5.2 Uniform expansiveness

In this section, we introduce a certain type of uniformity for expansiveness. In 1962, Bryant [14] remarked on the uniformity in the expansiveness of compact dynamical systems. This uniform expansiveness was formalized and studied by Sears [63] eleven years later.

Definition 5.2.1. A topological dynamical system $T : (X,d) \to (X,d)$ is said to be *(positively) uniformly expansive* if there exists $\delta > 0$ with the property that for every

$0 < \zeta < \delta$ there is $N = N(\zeta) \in \mathbb{N}$ such that

$$d(x,y) > \zeta \Longrightarrow d_N(x,y) > \delta.$$

The constant δ is called a *uniformly expansive constant* for T, and T is then said to be *uniformly δ-expansive*.

It is easy to check that every uniformly δ-expansive system is δ-expansive (this is left to the reader). It turns out that the converse is also true for systems that are defined on a compact metric space.

Proposition 5.2.2. *A topological dynamical system $T : (X,d) \to (X,d)$ is δ-expansive if and only if it is uniformly δ-expansive.*

Proof. As mentioned above, it is straightforward to check that every uniformly δ-expansive system is δ-expansive. To prove the converse, suppose by way of contradiction that $T : (X,d) \to (X,d)$ is a δ-expansive system that is not uniformly δ-expansive. Then there exist $0 < \delta' < \delta$ and sequences $(x_n)_{n=0}^{\infty}$ and $(y_n)_{n=0}^{\infty}$ in X such that $d(x_n,y_n) > \delta'$ but $d_n(x_n,y_n) \leq \delta$ for all $n \geq 0$. Since X is compact, we may assume (by passing to subsequences if necessary) that the sequences $(x_n)_{n=0}^{\infty}$ and $(y_n)_{n=0}^{\infty}$ converge to, say, x and y, respectively. On one hand, this implies that

$$d(x,y) = \lim_{n\to\infty} d(x_n,y_n) \geq \delta' > 0,$$

and hence $x \neq y$. On the other hand, if we fix momentarily $N \in \mathbb{N}$, for all $n \geq N$, we have that

$$d_N(x_n,y_n) \leq d_n(x_n,y_n) \leq \delta.$$

Therefore,

$$d_N(x,y) = \lim_{n\to\infty} d_N(x_n,y_n) \leq \delta.$$

Since $d_N(x,y) \leq \delta$ for every $N \in \mathbb{N}$, the δ-expansiveness of the system implies that $x = y$. This is, of course, in contradiction with our previous deduction that $x \neq y$. Thus, T is uniformly δ-expansive. □

Remark 5.2.3. Note that expansiveness and uniform expansiveness are distinct concepts in the realm of noncompact dynamical systems (see [63]).

Finally, we record the following observation, which is interesting on its own but will also be used in Section 9.6 on Brin–Katok's local entropy formula.

Observation 5.2.4. Let $T : X \to X$ be an expansive topological dynamical system and let d be a metric compatible with the topology on X. If $\delta > 0$ is an expansive constant for T corresponding to this metric, then for every $x \in X$, every $\zeta \in (0,\delta]$ and every integer $n > N(\zeta)$, note that

$$B_n(x,\delta) \subseteq B_{n-N(\zeta/2)}(x,\zeta).$$

Proof. Let $y \in X \setminus B_{n-N(\zeta/2)}(x,\zeta)$. Then there exists a least integer $0 \le k < n - N(\zeta/2)$ such that $T^k(y) \notin B(T^k(x),\zeta)$. This means that $d(T^k(y),T^k(x)) \ge \zeta > \zeta/2$. It follows from Definition 5.2.1 that $d_{N(\zeta/2)}(T^k(y),T^k(x)) > \delta$. Thus $d_{k+N(\zeta/2)}(y,x) > \delta$. Hence,

$$y \notin B_{k+N(\zeta/2)}(x,\delta) \supseteq B_{n-N(\zeta/2)+N(\zeta/2)}(x,\delta) = B_n(x,\delta).$$

So $y \in X \setminus B_n(x,\delta)$. That is, $X \setminus B_{n-N(\zeta/2)}(x,\zeta) \subseteq X \setminus B_n(x,\delta)$. □

5.3 Expansive maps are expanding with respect to an equivalent metric

The aim in this section is to provide a partial converse to Proposition 5.1.5, where it was shown that every distance expanding map is expansive.

Theorem 5.3.1. *If a topological dynamical system $T : X \to X$ is expansive, then there exists a metric, compatible with the topology of X, with respect to which T is distance expanding.*

The original proof that an expansive map defined upon a compact metric space is expanding with respect to a topologically equivalent metric is due to Coven and Reddy [17]. The proof we now present differs slightly by using uniform expansiveness. Like that of Coven and Reddy, the proof relies on a topological lemma, which we state here without proof (cf. [23]).

Frink's Metrization Lemma. *Let X be a metrizable space and let $(U_n)_{n=0}^{\infty}$ be a sequence of open neighborhoods of the diagonal $\triangle := \{(x,x) : x \in X\}$ of $X \times X$ having the following three properties:*

(a) $U_0 = X \times X$.
(b) $\bigcap_{n=0}^{\infty} U_n = \triangle$.
(c) $U_n \circ U_n \circ U_n \subseteq U_{n-1}$ *for every $n \in \mathbb{N}$, where*

$$R \circ S := \big\{(x,y) \in X \times X : \exists z \in X \text{ with } (x,z) \in R \text{ and } (z,y) \in S\big\}.$$

Then there exists a metric ρ, compatible with the topology of X, such that

$$U_n \subseteq \big\{(x,y) \in X \times X : \rho(x,y) < 2^{-n}\big\} \subseteq U_{n-1}$$

for every $n \in \mathbb{N}$.

Proof of Theorem 5.3.1. We shall construct a family of sets that satisfies the hypotheses of Frink's lemma, and then show that some iterate of T is expanding with respect to Frink's metric.

Since T is expansive, Proposition 5.2.2 implies that it is uniformly expansive. Let $3\theta > 0$ be a uniformly expansive constant for T with respect to a metric d compatible with the topology of X. For all $n \geq 0$ and all $\varepsilon > 0$, let

$$V_n(\varepsilon) := \{(x,y) \in X \times X : d_n(x,y) < \varepsilon\}.$$

Each set $V_n(\varepsilon)$ is an open neighborhood of the diagonal $\triangle$ in $X \times X$. The set $V_n(\varepsilon)$ is the collection of all couples of points whose forward orbits stay within a distance ε from each other up to and including time $n-1$. Let $M \geq 0$ be such that $d_M(x,y) > 3\theta$ whenever $d(x,y) > \theta/2$ (cf. Definition 5.2.1). Then no couple (x,y) such that $d(x,y) \geq \theta$ belongs to $V_M(3\theta)$. Hence,

$$V_M(3\theta) \subseteq \{(x,y) \in X \times X : d(x,y) < \theta\} = V_0(\theta). \tag{5.2}$$

Now, let $U_0 = X \times X$ and define

$$U_n := V_{Mn}(\theta) = \{(x,y) \in X \times X : d_{Mn}(x,y) < \theta\}$$

for each $n \in \mathbb{N}$. We shall show that the sets $(U_n)_{n=0}^{\infty}$ satisfy the three conditions of Frink's lemma. The first condition is satisfied by definition of U_0. Regarding the second condition, it is clear that $\triangle \subseteq \bigcap_{n=0}^{\infty} U_n$ because $\triangle \subseteq U_n$ for each $n \geq 0$. For the opposite inclusion, let $(x,y) \in \bigcap_{n=0}^{\infty} U_n$. Then $(x,y) \in U_n := V_{Mn}(\theta)$ for all $n \in \mathbb{N}$. Hence, $d_{Mn}(x,y) < \theta$ for all $n \in \mathbb{N}$, or, in other words, $d(T^j(x), T^j(y)) < \theta$ for all $j \geq 0$. As θ is an expansive constant for T, we deduce that $x = y$. So $\bigcap_{n=0}^{\infty} U_n \subseteq \triangle$. Since both inclusions hold, we conclude that

$$\bigcap_{n=0}^{\infty} U_n = \triangle.$$

It only remains to show that the third condition is satisfied, namely that

$$U_n \circ U_n \circ U_n \subseteq U_{n-1}, \quad \forall n \in \mathbb{N}.$$

For this, fix $n \in \mathbb{N}$ and let $(x,y) \in U_n \circ U_n \circ U_n$. Then there exist points $u, v \in X$ such that $(x,u), (u,v), (v,y) \in U_n$. Therefore, by the triangle inequality, $d_{Mn}(x,y) < 3\theta$. This means that

$$d_M\big(T^j(x), T^j(y)\big) < 3\theta$$

for all $0 \leq j \leq M(n-1)$. Thus, $(T^j(x), T^j(y)) \in V_M(3\theta) \subseteq V_0(\theta)$ for all $0 \leq j \leq M(n-1)$, by (5.2). Hence, $(x,y) \in V_{M(n-1)}(\theta) = U_{n-1}$, and we have proved that $U_n \circ U_n \circ U_n \subseteq U_{n-1}$ for any $n \in \mathbb{N}$. To summarize, the family $(U_n)_{n=0}^{\infty}$ satisfies all three hypotheses of Frink's lemma.

Therefore, we can now apply Frink's lemma to obtain a metric ρ, compatible with the topology of X, such that

$$U_n \subseteq \{(x,y) \in X \times X : \rho(x,y) < 2^{-n}\} \subseteq U_{n-1}$$

for all $n \in \mathbb{N}$.

We now show that $T^{3M} : X \to X$ is distance expanding with respect to the metric ρ. To this end, choose $x \neq y$ such that $\rho(x,y) < 2^{-4}$. Then there exists a unique $n \geq 0$ such that $(x,y) \in U_n \backslash U_{n+1}$, because $(U_n)_{n=0}^{\infty}$ is a descending sequence of sets such that $\bigcap_{n=0}^{\infty} U_n = \triangle$ and because $U_0 = X \times X$. By Frink's lemma, since $\rho(x,y) < 2^{-4}$, we have that $(x,y) \in U_3$. So $n \geq 3$.

On one hand, since (x,y) belongs to U_n, we have that

$$\rho(x,y) < 2^{-n}.$$

On the other hand, given that $(x,y) \in U_n \backslash U_{n+1} := V_{Mn}(\theta) \backslash V_{M(n+1)}(\theta)$, there exists $Mn \leq j < M(n+1)$ such that $d(T^j(x), T^j(y)) \geq \theta$. Write j in the form $j = i + 3M$. Then we have that $0 \leq M(n-3) \leq i < M(n-2)$ and

$$d(T^i(T^{3M}(x)), T^i(T^{3M}(y))) \geq \theta.$$

From this, we obtain that

$$(T^{3M}(x), T^{3M}(y)) \notin V_{M(n-2)} = U_{n-2} \supseteq \{(x,y) \in X \times X : \rho(x,y) < 2^{-(n-1)}\}.$$

Then it follows that whenever $\rho(x,y) < 2^{-4}$,

$$\rho(T^{3M}(x), T^{3M}(y)) \geq 2^{-(n-1)} = 2 \cdot 2^{-n} \geq 2\rho(x,y).$$

Hence, T^{3M} is expanding with respect to Frink's metric. It follows from Exercise 4.6.2 that the map T is expanding with respect to a metric topologically equivalent to Frink's metric, which is in turn topologically equivalent to the original metric d. □

We now draw some important conclusions from the fact that an expansive dynamical system is expanding with respect to a topologically equivalent metric. Namely, we can infer the existence of Markov partitions and of a symbolic representation.

Corollary 5.3.2. *Every open, expansive dynamical system $T : X \to X$ admits Markov partitions of arbitrarily small diameters.*

Proof. This follows immediately from Theorems 5.3.1 and 4.4.6. □

Corollary 5.3.3. *Every open, expansive dynamical system $T : X \to X$ admits a symbolic representation. More precisely, every open, expansive system is a factor of a subshift of finite type $\sigma : E_A^{\infty} \to E_A^{\infty}$ with a coding map $\pi : E_A^{\infty} \to X$ such that every point in a forward T-invariant dense G_δ-subset of X admits a unique symbolic representation.*

Proof. This follows from Theorems 5.3.1 and 4.5.2. □

Yet another repercussion of the fact that an expansive dynamical system is expanding with respect to a topologically equivalent metric is the following.

Corollary 5.3.4. *Every open, expansive, and transitive dynamical system $T : X \to X$ is very strongly transitive.*

Proof. This is a direct consequence of Theorems 5.3.1 and 4.2.10. □

5.4 Parabolic Cantor sets

In this section, we describe a family of C^1 maps defined on topological Cantor subspaces of $\mathbb{R}$, that are expansive but not distance expanding with respect to the standard Euclidean metric on $\mathbb{R}$. Of course, bearing in mind what we have just proved in the previous section, for each of these maps there exists a metric, compatible with the Euclidean topology on $\mathbb{R}$, with respect to which the map is distance expanding.

Let $I := [0,1]$. Let E be a finite set, say $E = \{0, 1, \dots, k-1\}$, and let $\varphi_e : I \to I$, $e \in E$, be C^1 maps with the following properties:

(I1) $\varphi_0'(0) = 1$.
(I2) $0 < \varphi_0'(x) < 1$ for all $x \in I \setminus \{0\}$.
(I3) $0 < |\varphi_e'(x)| < 1$ for all $x \in I$ and $e \in E \setminus \{0\}$.
(II) $\varphi_0(0) = 0$.
(III) $\varphi_e(I) \cap \varphi_f(I) = \emptyset$ for all $e \neq f$.

This setting is reminiscent of hyperbolic Cantor sets in Subsection 4.1.2, with $X_0 = I$. Indeed, conditions (I1)–(I3) concern the derivative of the generators and are the counterparts for parabolic Cantor sets of condition (i) for hyperbolic Cantor sets. In particular, they imply that the generators are invertible functions on I, so each generator is either strictly increasing or strictly decreasing. Condition (III) is identical to condition (iii). Condition (ii) is automatically fulfilled.

Conditions (I1) and (II) are the reason for calling the limit set X a *parabolic Cantor set*, as opposed to an hyperbolic one. These conditions ensure that the derivative of one of the generators is equal to 1 at a fixed point.

Under these conditions, the limit set X is constructed through the same procedure as for hyperbolic Cantor sets (see (4.8) and (4.6)). The neighborhood U can be constructed in precisely the same way as for hyperbolic Cantor sets (see (4.12)), provided that we first extend and replace the generators with C^1 diffeomorphisms of $\mathbb{R}$ as follows:

$$\overline{\varphi}_0(x) = \begin{cases} -\varphi_0(1) + \varphi_0'(1) \cdot (x+1) & \text{if } x \leq -1 \\ -\varphi_0(-x) & \text{if } x \in -I \\ \varphi_0(x) & \text{if } x \in I \\ \varphi_0(1) + \varphi_0'(1) \cdot (x-1) & \text{if } x \geq 1, \end{cases}$$

whereas, for all $e \in E \setminus \{0\}$,

$$\overline{\varphi}_e(x) = \begin{cases} \varphi_e(0) + \varphi_e'(0) \cdot x & \text{if } x \leq 0 \\ \varphi_e(x) & \text{if } x \in I \\ \varphi_e(1) + \varphi_e'(1) \cdot (x-1) & \text{if } x \geq 1. \end{cases}$$

The map $T : U \to \mathbb{R}$ is then defined in exactly the same manner as for hyperbolic Cantor sets (see (4.13)).

All definitions, relations, and proofs for hyperbolic Cantor sets which do not involve the constant M hold for parabolic Cantor sets. Observe that $(\varphi_e^{-1})'(\varphi_e(x)) = 1/\varphi_e'(x)$ for all $x \in I$ and $e \in E$. Condition (I1) implies that $T'(0) = 1$. The triple (X, U, T) is not an expanding repeller for parabolic Cantor sets since it does not satisfy condition (b) of the definition of a repeller (cf. Definition 4.1.4 and (4.16)). Condition (II) implies that the point 0 lies in the limit set X and prevents any iterate of T of being an expanding repeller. However, conditions (a) and (c) of a repeller are satisfied.

Now, define

$$\lambda := \min_{e\in E\setminus\{0\}} \min_{x\in\varphi_e(I)} |T'(x)| = \frac{1}{M} > 1, \quad \text{where} \quad M := \max_{e\in E\setminus\{0\}} \max_{x\in I} |\varphi_e'(x)|.$$

Theorem 5.4.1. *The map $T : X \to X$ is expansive.*

Proof. Consider $x \in \varphi_0(I) \setminus \{0\}$. By the mean value theorem, there exists $y \in \mathrm{Int}(\varphi_0(I))$ such that

$$\frac{T(x)}{x} = \frac{T(x) - T(0)}{x - 0} = T'(y) > 1. \tag{5.3}$$

Therefore, $T(x) > x$ for all $x \in \varphi_0(I) \setminus \{0\}$ and the map $T|_{\varphi_0(I)} : \varphi_0(I) \to I$ has no fixed point other than 0. Let $\overline{\delta} > 0$ be the length of the smallest gap between all $\varphi_e(I)$'s. We claim that any $0 < \delta < \overline{\delta}$ is an expansive constant for $T : X \to X$. So suppose that there exist points $z, w \in X$ such that

$$|T^n(z) - T^n(w)| \le \delta, \quad \forall n \ge 0.$$

Let $\triangle$ be the closed interval joining z and w. Then for every $n \ge 0$, there exists a unique $e \in E$ such that $T^n(\triangle) \subseteq \varphi_e(I)$. It follows from the fact that $T'(x) > 1$ for all $x \in U \setminus \{0\}$ and from the definition of λ that for all $n \ge 0$,

$$|T^{n+1}(\triangle)| > |T^n(\triangle)| \quad \text{whenever } T^n(\triangle) \subseteq \varphi_0(I) \tag{5.4}$$

and

$$|T^{n+1}(\triangle)| \ge \lambda |T^n(\triangle)| \quad \text{whenever } T^n(\triangle) \subseteq \varphi_e(I) \text{ for some } e \in E \setminus \{0\}. \tag{5.5}$$

Since $\lambda > 1$ and $|T^n(\triangle)| \le 1$ for all $n \ge 0$, relation (5.5) can only hold for finitely many n. That is, there exists $N \ge 0$ such that for all $n \ge 0$,

$$T^{n+N}(\triangle) = T^n(T^N(\triangle)) \subseteq \varphi_0(I).$$

Fix $x \in T^N(\triangle)$. If $x \neq 0$, then $T^n(x) \in \varphi_0(I) \setminus \{0\}$ for all $n \ge 0$. It follows from (5.3) that the sequence of iterates $(T^n(x))_{n=0}^{\infty}$ is (strictly) increasing. Thus, it has a limit point which, according to Lemma 1.1.4, is a fixed point for T. This contradicts the fact that $T|_{\varphi_0(I)} : \varphi_0(I) \to I$ has no fixed point but 0. Therefore, $T^N(\triangle) = \{0\}$. Hence, $\triangle = \{0\}$ and $z = w (= 0)$. $\square$

Since T is expansive on the limit set X, according to Theorem 5.3.1 there exists a metric, compatible with the topology of X, with respect to which T is distance expanding on the limit set X. However, that metric is not the usual Euclidean metric.

Proposition 5.4.2. *The map $T : X \to X$ is not expanding with respect to the Euclidean metric.*

Proof. Let $\varepsilon > 0$. Since T is a C^1 map and $T'(0) = 1$, there exists $\eta > 0$ such that $T'(x) < 1+\varepsilon$ for all $x \in [0,\eta)$. Fix $y \in X \cap (0,\eta)$. It follows from the mean value theorem that for some $z \in (0,y)$,

$$\frac{|T(y) - T(0)|}{|y - 0|} = T'(z) < 1 + \varepsilon.$$

Therefore, T is not expanding with respect to the metric $d(x,y) = |x - y|$. □

5.5 Exercises

Exercise 5.5.1. Prove that the expansiveness of $T : X \to X$ is independent of the metric on X (though expansive constants generally depend on the metric chosen). That is, show that, given two metrics d and d', which generate the topology of the compact metrizable space X, the map T is expansive when X is equipped with the metric d if and only if T is expansive when X is endowed with the metric d'.

Exercise 5.5.2. A dynamical system $T : X \to X$ on a topological space X is said to be expansive if there exists a base $\mathcal{B}$ for the topology such that for every $x \neq y$ there is $n = n(x,y) \geq 0$ with

$$\{T^n(x), T^n(y)\} \not\subseteq U, \quad \forall U \in \mathcal{B}.$$

Note that if X is second-countable, then expansiveness is equivalent to the existence of a countable base with the above property. Show that this definition is equivalent to Definition 5.1.1 when X is a compact metrizable space.

Exercise 5.5.3. Prove that expansiveness is a topological conjugacy invariant.

Exercise 5.5.4. Prove that the metrics d_n, $n \in \mathbb{N}$, given in Definition 5.1.3 induce the same topology.

Exercise 5.5.5. Let $T : (X,d) \to (X,d)$ be a dynamical system. For every $n \geq 0$, let $d_\infty : X \times X \to [0,\infty)$ be the function

$$d_\infty(x,y) := \sup_{0 \leq j < \infty} d(T^j(x), T^j(y)) = \sup_{n \in \mathbb{N}} d_n(x,y).$$

Show that d_∞ defines a metric on X. Prove that if T is expansive on X then d_∞ generates the discrete topology on X. In particular, if X has infinite cardinality and T is expansive, then d_∞ is not topologically equivalent to any d_n, $n \in \mathbb{N}$.

Exercise 5.5.6. Let $T : X \to X$ be a topological dynamical system. Prove that the following conditions are equivalent:

(a) T is expansive.
(b) T^n is expansive for some $n \in \mathbb{N}$.
(c) T^n is expansive for all $n \in \mathbb{N}$.

Exercise 5.5.7. Show that the Cartesian product of finitely many expansive dynamical systems is an expansive system.

Exercise 5.5.8. Find two expansive maps on the same compact metric space whose composition is not expansive.

Exercise 5.5.9. Recall that the unit circle $\mathbb{S}^1$ is homeomorphic to the closed interval $[0, 1]$ when 0 and 1 are identified. Define the map $T : \mathbb{S}^1 \to \mathbb{S}^1$ by the following formula:

$$T(x) = \begin{cases} x + 2x^2 & \text{if } 0 \le x \le 1/2 \\ 2x - 1 & \text{if } 1/2 \le x \le 1. \end{cases}$$

Show that T is not distance expanding with respect to the linear Euclidean metric on $\mathbb{S}^1$, but that T is an expansive map. Recall that the linear Euclidean metric on $\mathbb{S}^1$ is given by

$$d(x, y) = \min\{|x - y|, |1 + x - y|, |1 + y - x|\}.$$

Exercise 5.5.10. Suppose that for all $n \in \mathbb{N}$, the map $T_n : X_n \to X_n$ is a continuous map of a compact metric space X_n. Let X be the one-point (Alexandroff) compactification of the disjoint union of all spaces X_n, $n \in \mathbb{N}$. Denote the added point by ω. Define the map $T : X \to X$ by the formula

$$T(x) = \begin{cases} T_n(x) & \text{if } x \in X_n \\ \omega & \text{if } x = \omega. \end{cases}$$

Show that T is never expansive.

6 Shub expanding endomorphisms

In Section 6.2 of this chapter, we give a systematic account of Shub expanding endomorphisms. These maps are far-reaching generalizations of the expanding endomorphisms of the circle which we first introduce in Section 6.1. They constitute a large subclass of distance expanding maps. Their origins lie in the seminal papers of Epstein and Shub [22], Shub [66], and Krzyżewski and Szlenk [42]. Our exposition stems from the chapter on expanding endomorphisms in Szlenk's book [71].

Basic knowledge of algebraic geometry/topology is assumed. The first chapter of the book by Hatcher [28] is an engaging source for the reader unfamiliar with notions such as lift, deck transformation, homotopy, and the fundamental group of a topological space, notions which will be used throughout this chapter, especially in our digression into algebraic topology in Section 6.3.

Finally, in Section 6.4 we establish that Shub's expanding endomorphisms are structurally stable, form an open set in an appropriate topology of smooth maps, are topologically exact (and hence transitive), have at least one fixed point as well as a dense set of periodic points, and their universal covering space is diffeomorphic to $\mathbb{R}^n$.

6.1 Shub expanding endomorphisms of the circle

In this section, we study a special class of maps of the unit circle, the Shub expanding endomorphisms of $\mathbb{S}^1$.

Let $\gamma : \mathbb{S}^1 \to (0,\infty)$ be a C^1 function on $\mathbb{S}^1$ (recall that this means that the first derivative of γ exists and is continuous). The function γ induces the Riemannian metric $\rho_\gamma = \gamma|dx|$ on $\mathbb{S}^1$. If $\triangle$ is an arc of $\mathbb{S}^1$ and $\varphi : \triangle \to \mathbb{S}^1$ is a C^1 curve on $\mathbb{S}^1$, then the length $\ell_\gamma(\varphi)$ of φ is defined to be

$$\ell_\gamma(\varphi) := \int_\triangle |\varphi'(x)|\,\gamma(x)\,dx. \tag{6.1}$$

The Riemannian metric ρ_γ induces a distance (which, in somewhat of an abuse of notation, we will also denote by ρ_γ) on $\mathbb{S}^1$ as follows. Let $a, b \in \mathbb{S}^1$, and let $\triangle_1$ and $\triangle_2$ be the two arcs of $\mathbb{S}^1$ joining a and b. Let $\mathrm{Id}_{\triangle_i} : \triangle_i \to \triangle_i$, $i = 1, 2$, be the identity curves on these respective arcs. We define

$$\rho_\gamma(a,b) := \min_{i=1,2} \ell_\gamma(\mathrm{Id}_{\triangle_i}) = \min_{i=1,2} \int_{\triangle_i} \gamma(x)\,dx.$$

If $v \in T_x\mathbb{S}^1$, that is, if v is a vector in $\mathbb{R}^2$ tangent to $\mathbb{S}^1$ at the point $x \in \mathbb{S}^1$, then the norm of v relative to the metric ρ_γ is given by

$$\|v\|_\gamma := \gamma(x)\|v\|,$$

https://doi.org/10.1515/9783110702682-006

where $\|\cdot\|$ is the standard Euclidean norm in $\mathbb{R}^2$. We can thus rewrite (6.1) as

$$\ell_\gamma(\varphi) := \int_\triangle \|\varphi'(x)\|_\gamma \, dx.$$

If $T : \mathbb{S}^1 \to \mathbb{S}^1$ is a differentiable map, then for every $x \in \mathbb{S}^1$ the derivative map $T'(x)$ sends $T_x\mathbb{S}^1$ into $T_{T(x)}\mathbb{S}^1$ and

$$\|T'(x)(v)\| = \|T'(x)\| \cdot \|v\|, \quad \forall v \in T_x\mathbb{S}^1.$$

Hence,

$$\begin{aligned}\|T'(x)(v)\|_\gamma &= \gamma\big(T(x)\big)\|T'(x)(v)\| = \gamma\big(T(x)\big)\|T'(x)\| \cdot \|v\| \\ &= \frac{\gamma(T(x))}{\gamma(x)}\|T'(x)\| \cdot \|v\|_\gamma.\end{aligned}$$

We naturally set

$$\|T'(x)\|_\gamma := \frac{\|T'(x)(v)\|_\gamma}{\|v\|_\gamma} = \frac{\gamma(T(x))}{\gamma(x)}\|T'(x)\| \tag{6.2}$$

and call this quantity the *norm of* $T'(x)$ *with respect to the metric* ρ_γ. We call a C^1 endomorphism $T : \mathbb{S}^1 \to \mathbb{S}^1$ *Shub expanding* with respect to the metric ρ_γ if there exists some $\lambda > 1$ such that

$$\|T'(x)(v)\|_\gamma \geq \lambda\|v\|_\gamma, \quad \forall v \in T_x\mathbb{S}^1, \ \forall x \in \mathbb{S}^1.$$

Equivalently, $T : \mathbb{S}^1 \to \mathbb{S}^1$ is Shub expanding if

$$\|T'(x)\|_\gamma \geq \lambda, \quad \forall x \in \mathbb{S}^1. \tag{6.3}$$

Example 6.1.1. For every integer $|k| \geq 2$, the map $E_k : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $E_k(x) := kx \pmod 1$ is Shub expanding with respect to the standard Riemannian metric (that is, when $\gamma \equiv 1$). Indeed, $\|E_k'(x)\| = |k| \geq 2$ for every $x \in \mathbb{S}^1$.

We now show that each Shub expanding map of the circle is, up to a C^1 conjugacy, Shub expanding with respect to the standard Euclidean metric on $\mathbb{S}^1$. Indeed, in light of (6.2), we may multiply γ by a constant factor without changing $\|T'(x)\|_\gamma$ and in such a way that

$$\int_{\mathbb{S}^1} \gamma(x)\, dx = 1.$$

We say that such a Riemannian metric is normalized. Then the map $H : \mathbb{S}^1 \to \mathbb{S}^1$ given by the formula

$$H(x) = \int_0^x \gamma(t)\, dt$$

defines a C^1 diffeomorphism of $\mathbb{S}^1$ such that

$$H'(x) = \gamma(x), \quad \forall x \in \mathbb{S}^1. \tag{6.4}$$

We define the C^1 endomorphism

$$\overline{T} := H \circ T \circ H^{-1} : \mathbb{S}^1 \to \mathbb{S}^1. \tag{6.5}$$

We then obtain the following important result.

Theorem 6.1.2. *Each Shub expanding map $T : \mathbb{S}^1 \to \mathbb{S}^1$ of the unit circle is C^1 conjugate to the map $\overline{T} : \mathbb{S}^1 \to \mathbb{S}^1$, which is Shub expanding with respect to the standard Euclidean metric on $\mathbb{S}^1$.*

Proof. Suppose that $T : \mathbb{S}^1 \to \mathbb{S}^1$ is a Shub expanding map. As argued above, we may assume without loss of generality that the corresponding function $\gamma : \mathbb{S}^1 \to (0, \infty)$ is normalized. Let $\overline{T} : \mathbb{S}^1 \to \mathbb{S}^1$ be given by (6.5). Using the chain rule, (6.4) and (6.3), we obtain that

$$\begin{aligned}
\|\overline{T}\,'(x)\| &= H'(T(H^{-1}(x))) \cdot \|T'(H^{-1}(x))\| \cdot (H^{-1})'(x) \\
&= \gamma(T(H^{-1}(x)))\|T'(H^{-1}(x))\|(H'(H^{-1}(x)))^{-1} \\
&= \gamma(T(H^{-1}(x)))\|T'(H^{-1}(x))\|(\gamma(H^{-1}(x)))^{-1} \\
&= \|T'(H^{-1}(x))\|_\gamma \geq \lambda > 1.
\end{aligned}$$

Thus, $\overline{T}$ is Shub expanding with respect to the standard Euclidean metric on $\mathbb{S}^1$. □

Our goal now is to prove a structure theorem for Shub expanding maps of the unit circle and to demonstrate the structural stability of the maps E_k, $|k| \geq 2$, from Example 6.1.1 (see Section 1.2 for more on structural stability). Before stating that theorem, let us add one more piece of notation. Given $k \in \mathbb{Z}$, let $\mathcal{E}^k(\mathbb{S}^1)$ be the space of all Shub expanding endomorphisms of $\mathbb{S}^1$ with degree equal to k. For a review of the notions of lift and degree of a circle map, see Section 2.1.

Theorem 6.1.3. *Every Shub expanding map $T \in \mathcal{E}^k(\mathbb{S}^1)$, where $|k| \geq 2$, is topologically conjugate to the map $E_k : \mathbb{S}^1 \to \mathbb{S}^1$. In addition, the map $E_k : \mathbb{S}^1 \to \mathbb{S}^1$ is strongly structurally stable when the space $\mathcal{E}^k(\mathbb{S}^1)$ is endowed with the topology of uniform convergence.*

Proof. In light of Theorem 6.1.2, we may assume without loss of generality that T is Shub expanding with respect to the standard Euclidean metric.

The following classical argument is essentially the proof of Theorem 2.4.6 in [33]. We give the proof for any positive k and mention afterwards the modifications necessary for a negative k. Consider the arcs

$$\triangle_n^m = \pi\left(\left[\frac{m}{k^n}, \frac{m+1}{k^n}\right]\right)$$

for all $n \in \mathbb{Z}_+$ and $0 \le m < k^n$, where $\mathbb{R} \ni x \mapsto \pi(x) = x (\mathrm{mod}\ 1) \in \mathbb{S}^1$. For each $n \in \mathbb{Z}_+$, the family

$$\xi_n = \{\triangle_n^0, \ldots, \triangle_n^{k^n-1}\}$$

is the "partition" of $\mathbb{S}^1$ into the k^n arcs whose endpoints are consecutive rational numbers with denominator k^n. These arcs are such that

$$E_k(\triangle_n^m) = \triangle_{n-1}^{m'}, \tag{6.6}$$

where m' is the unique integer between 0 and $k^{n-1} - 1$ such that $m' = m\ (\mathrm{mod}\ k^{n-1})$.

We now construct a nested sequence of "partitions"

$$\zeta_n = \{\pi(\Gamma_n^0), \ldots, \pi(\Gamma_n^{k^n-1})\}$$

of $\mathbb{S}^1$ into arcs which will be in a natural, order-preserving correspondence with the standard sequence ξ_n. Let p be a fixed point for a lift $\tilde{T}$ of T as in Lemma 2.1.14. If T is close to E_k, pick p close to 0. Since $\tilde{T}(p) = p$ and T is of degree k, we know that $\tilde{T}(p+1) = p + k$. Moreover, since T is locally injective, its lift $\tilde{T}$ is a strictly monotone continuous function. Then there are unique real numbers

$$p = a_1^0 < a_1^1 < a_1^2 < \cdots < a_1^{k-1} < a_1^k = p + 1$$

such that $\tilde{T}(a_1^m) = p + m$ for each $0 \le m \le k$. Let $\Gamma_1^m = [a_1^m, a_1^{m+1}]$ for every $0 \le m < k$. Then

$$T(\pi(\Gamma_1^m)) = \pi \circ \tilde{T}([a_1^m, a_1^{m+1}]) = \pi([p+m, p+m+1]) = \mathbb{S}^1,$$

and T is injective on the arc $\pi(\Gamma_1^m)$ up to identification of the endpoints of Γ_1^m. If T is close to E_k, then clearly each number a_1^m is close to m/k.

Furthermore, since $\tilde{T}(a_1^m) = p + m$ and $\tilde{T}(a_1^{m+1}) = p + m + 1$, and since $\tilde{T}$ is a strictly monotone continuous function, there are unique real numbers

$$a_1^m = a_2^{km} < a_2^{km+1} < \cdots < a_2^{km+k-1} < a_2^{k(m+1)} = a_1^{m+1}$$

such that $\tilde{T}(a_2^{km+i}) = a_1^i\ (\mathrm{mod}\ 1)$ for $0 \le i \le k$. Again, a_2^{km+i} is close to $(km+i)/k^2$ if T is close to E_k. Let $\Gamma_2^m = [a_2^m, a_2^{m+1}]$ for $0 \le m < k^2$, so that $T(\pi(\Gamma_2^m)) = \pi \circ \tilde{T}([a_2^m, a_2^{m+1}]) = \pi([a_1^{m'}, a_1^{m'+1}]) = \pi(\Gamma_1^{m'})$, where m' is the unique integer between 0 and $k-1$ such that $m' = m\ (\mathrm{mod}\ k)$.

We continue inductively and for each $n \in \mathbb{N}$ we define points a_n^{km+i} for $0 \le m < k^{n-1}$ and $0 \le i \le k$ such that

$$a_{n-1}^m = a_n^{km} < a_n^{km+1} < \cdots < a_n^{km+k-1} < a_n^{k(m+1)} = a_{n-1}^{m+1}$$

and

$$\widetilde{T}(a_n^{km+i}) = a_{n-1}^{m'} \pmod 1, \tag{6.7}$$

where $0 \le m' < k^{n-1}$ and $m' = km + i \pmod{k^{n-1}}$. Let $\Gamma_n^m = [a_n^m, a_n^{m+1}]$ for $0 \le m < k^n$. Then $T(\pi(\Gamma_n^m)) = \pi(\Gamma_{n-1}^{m'})$, where $0 \le m' < k^{n-1}$ and $m = m' \pmod{k^{n-1}}$. By induction, $T^n(\pi(\Gamma_n^m)) = \mathbb{S}^1$ and T^n is injective on $\pi(\Gamma_n^m)$ up to identification of the endpoints of Γ_n^m.

So far, we have only used the facts that T is locally injective (and thus its lift $\widetilde{T}$ is strictly monotone) and that T has degree k. If T is Shub expanding, that is, if $\|T'(x)\| \ge \lambda > 1$ for all $x \in \mathbb{S}^1$, then the length of each arc $\pi(\Gamma_n^m)$ does not exceed λ^{-n}, so the set of points $\{\pi(a_n^m)\}_{n\in\mathbb{N},0\le m<k^n}$ is dense in $\mathbb{S}^1$ while $\{a_n^m\}_{n\in\mathbb{N},0\le m<k^n}$ is dense in the interval $[p, p+1]$. This is the only place in the proof where the fact that T is an expanding map is used. (In fact, the use of differentiability could be easily avoided.)

Furthermore, for any $N \in \mathbb{N}$ and $\varepsilon > 0$ one can find $\delta > 0$ such that if T is δ-close to E_k in the uniform topology, then

$$\left|a_n^m - \frac{m}{k^n}\right| < \frac{\varepsilon}{3}, \quad \forall 1 \le n \le N,\ \forall 0 \le m < k^n. \tag{6.8}$$

We define a correspondence h between the set $\{a_n^m\}_{n\in\mathbb{N},0\le m<k^n}$ and all k-ary rationals, that is, the rational numbers whose denominators are powers of k, by setting

$$h(a_n^m) = \frac{m}{k^n}.$$

This correspondence is monotone and since the set $\{a_n^m\}_{n\in\mathbb{N},0\le m<k^n}$ is dense in the interval $[p, p+1]$, it can be uniquely extended to a homeomorphism $h : [p, p+1] \to [0, 1]$. Since $h(\Gamma_n^m) = \triangle_n^m$ for all $n \in \mathbb{N}$ and $0 \le m < k^n$, relations (6.6) and (6.7) imply that T is topologically conjugate to E_k via h, that is,

$$E_k \circ h = h \circ T. \tag{6.9}$$

Assuming under the conditions of (6.8) that N and ε are chosen such that $1/k^N < \varepsilon/3$, one sees in addition that $|a_n^m - h(a_n^m)| < \varepsilon$ for $n \in \mathbb{N}$ and $0 \le m < k^n$, and hence $|h(x) - x| < \varepsilon$ for all x, that is, $h \in B_{\rho_\infty}(\mathrm{Id}_{\mathbb{S}^1}, \varepsilon)$. Recall also that if T is close enough to E_k in the topology of uniform convergence, then the degree of T is k, by Lemma 2.1.12. Thus, E_k is strongly structurally stable in the space $\mathcal{E}^k(\mathbb{S}^1)$.

The case of a negative k differs primarily in notation. The order of the real numbers a_n^{km+i} between a_{n-1}^m and a_{n-1}^{m+1} will be increasing for even n's and decreasing for odd n's, the same as the corresponding structure imposed by the map E_k on the k-ary rationals. □

Let $C^1(\mathbb{S}^1, \mathbb{S}^1)$ denote the space of all C^1 endomorphisms of $\mathbb{S}^1$ endowed with the C^1 topology.

Corollary 6.1.4. *Every Shub expanding map of the unit circle is structurally stable in* $C^1(\mathbb{S}^1, \mathbb{S}^1)$ *and every map* $E_k : \mathbb{S}^1 \to \mathbb{S}^1$, $|k| \geq 2$, *is strongly structurally stable in this class.*

Proof. This is an immediate consequence of Theorem 6.1.3 once one observes that each element of $\mathcal{E}^k(\mathbb{S}^1)$ has a neighborhood U in $C^1(\mathbb{S}^1, \mathbb{S}^1)$ such that $U \subseteq \mathcal{E}^k(\mathbb{S}^1)$. For more information, see Theorem 6.2.5. □

6.2 Definition, characterization, and properties of general Shub expanding endomorphisms

Unless stated otherwise, we shall let M be a compact connected smooth (i. e., C^∞) manifold, and ρ a Riemannian metric on M.

If $\gamma : I \to M$ is a smooth curve defined on an interval $I \subseteq \mathbb{R}$, then the length of γ with respect to the Riemannian metric ρ is defined to be

$$\ell_\rho(\gamma) := \int_I \|D_t\gamma(e_t)\|\, dt = \int_I \|D_t\gamma\|\, dt,$$

where e_t is the unit tangent vector to I at the point t. Given $x, y \in M$, let $\Gamma(x, y)$ be the collection of all smooth curves on M whose endpoints are x and y. The distance $\rho(x, y)$ between x and y is defined as

$$\rho(x, y) := \inf\{\ell_\rho(\gamma) \mid \gamma \in \Gamma(x, y)\},$$

where, as above, we shall use the same symbol ρ to denote the original Riemannian metric and the distance it induces on M. A curve γ joining x to y whose length is equal to $\rho(x, y)$ is called a *geodesic* from x to y. Although we will not rely on this fact, a geodesic joining x and y always exists. In fact, it is unique if the points x and y are sufficiently close.

Let us begin by defining the class of transformations of M that we will study.

Definition 6.2.1. A C^1 endomorphism $T : M \to M$ is *Shub expanding* if there exists $k \in \mathbb{N}$ such that

$$\|D_xT^k(v)\|_{T^k(x)} \geq 2\|v\|_x, \quad \forall x \in M,\ \forall v \in T_xM.$$

We immediately present a characterization of these maps. We invite the reader to envision the implications it will have on the dynamics of these maps.

Proposition 6.2.2. *If* $T : M \to M$ *is a* C^1 *endomorphism, then the following statements are equivalent:*

(a) *The map* $T : M \to M$ *is Shub expanding.*
(b) *There exist constants* $\mu > 1$ *and* $C > 0$ *such that for all* $n \in \mathbb{N}$,

$$\|D_xT^n(v)\|_{T^n(x)} \geq C\mu^n\|v\|_x, \quad \forall x \in M,\ \forall v \in T_xM.$$

(c) *There exist* $\lambda > 1$ *and a Riemannian metric* ρ' *on* M *such that*

$$\|D_xT(v)\|_{T(x),\rho'} \geq \lambda\|v\|_{x,\rho'}, \quad \forall x \in M,\ \forall v \in T_xM. \tag{6.10}$$

Proof. Let us first prove that (c)⇒(a). Since M is compact, the Riemannian norms ρ and ρ' are equivalent in the sense that there exists a constant $L \geq 1$ such that

$$L^{-1}\|v\|_{x,\rho} \leq \|v\|_{x,\rho'} \leq L\|v\|_{x,\rho}, \quad \forall x \in M,\ \forall v \in T_xM.$$

Using the chain rule repeatedly, for every $k \in \mathbb{N}$ we obtain that

$$\begin{aligned}
\|D_xT^k(v)\|_{T^k(x),\rho} &\geq L^{-1}\|D_xT^k(v)\|_{T^k(x),\rho'} \\
&= L^{-1}\|D_{T^{k-1}(x)}T(D_xT^{k-1}(v))\|_{T(T^{k-1}(x)),\rho'} \\
&\geq L^{-1}\lambda\|D_xT^{k-1}(v)\|_{T^{k-1}(x),\rho'} \\
&\geq \ldots \geq L^{-1}\lambda^k\|v\|_{x,\rho'} \geq \lambda^k L^{-2}\|v\|_{x,\rho}.
\end{aligned}$$

It suffices to take $k \in \mathbb{N}$ so large that $\lambda^k \geq 2L^2$.

We now prove that (a)⇒(b). To begin, it follows from the definition of a Shub expanding endomorphism that $\mathrm{Ker}(D_xT^k) = \{0\}$ for every $x \in M$. Hence, $\mathrm{Ker}(D_xT^n) = \{0\}$ for all $x \in M$ and all $n \in \mathbb{N}$ (first, use the chain rule in the form $D_xT^k = D_{T(x)}T^{k-1} \circ D_xT$ to establish the statement for $n = 1$ and then deduce it for any n). Since the tangent spaces T_xM and $T_{T^n(x)}M$ are of finite dimension equal to $\dim(M)$, all the maps $D_xT^n : T_xM \to T_{T^n(x)}M$ are linear isomorphisms and thereby invertible. In particular, $\|(D_xT)^{-1}\| < \infty$ for each $x \in M$. Moreover, observe that the determinant function $x \mapsto \det(D_xT)$ is continuous on M since $T \in C^1(M, M)$, and does not vanish anywhere on M since D_xT is invertible for every $x \in M$. As the entries of the inverse matrix A^{-1} of an invertible matrix A are polynomial functions of the entries of A divided by $\det(A)$, the entries of the matrix $(D_xT)^{-1}$ depend continuously on $x \in M$. Consequently, the function $x \mapsto \|(D_xT)^{-1}\|$ is continuous on M, and thus $\|(DT)^{-1}\|_\infty := \max_{x\in M} \|(D_xT)^{-1}\| < \infty$ since M is compact. Let

$$\alpha := \max\{1, \|(DT)^{-1}\|_\infty\} < \infty.$$

Fix an arbitrary $n \in \mathbb{N}$. Write $n = qk + r$, where q and r are integers such that $q \geq 0$ and $0 \leq r < k$. For every $x \in M$ and every $v \in T_xM$, we have

$$\begin{aligned}
\|v\|_x &= \|(D_xT^n)^{-1}(D_xT^n(v))\|_x \leq \|(D_xT^n)^{-1}\| \cdot \|D_xT^n(v)\|_{T^n(x)} \\
&= \|(D_{T^{qk}(x)}T^r \circ D_xT^{qk})^{-1}\| \cdot \|D_xT^n(v)\|_{T^n(x)} \\
&\leq \|(D_{T^{qk}(x)}T^r)^{-1}\| \cdot \|(D_xT^{qk})^{-1}\| \cdot \|D_xT^n(v)\|_{T^n(x)} \\
&\leq \prod_{i=0}^{r-1}\|(D_{T^{qk+i}(x)}T)^{-1}\| \cdot \prod_{j=0}^{q-1}\|(D_{T^{jk}(x)}T^k)^{-1}\| \cdot \|D_xT^n(v)\|_{T^n(x)}
\end{aligned}$$

$$\le \alpha^r 2^{-q} \|D_x T^n(v)\|_{T^n(x)}$$
$$\le 2\alpha^{k-1} 2^{-(q+1)} \|D_x T^n(v)\|_{T^n(x)}.$$

Given that $k(q+1) \ge n$, it follows from the above estimate that

$$\|D_x T^n(v)\|_{T^n(x)} \ge \frac{1}{2}\alpha^{1-k}(2^{1/k})^n \|v\|_x.$$

Thus, part (b) is proved with $C = \alpha^{1-k}/2 > 0$ and $\mu = 2^{1/k} > 1$.

Since the implication (b)⇒(a) is obvious, to complete the proof it suffices to show that (a)⇒(c). Define on M a new metric ρ' with scalar product on the tangent spaces given by

$$\langle v, w\rangle'_x := \sum_{j=0}^{k-1} \langle D_x T^j(v), D_x T^j(w)\rangle_{T^j(x)}.$$

Then

$$\|v\|^2_{x,\rho'} = \langle v, v\rangle'_x = \sum_{j=0}^{k-1} \|D_x T^j(v)\|^2_{T^j(x),\rho} \tag{6.11}$$

whereas

$$\|D_x T(v)\|^2_{T(x),\rho'} = \sum_{j=1}^{k} \|D_x T^j(v)\|^2_{T^j(x),\rho}. \tag{6.12}$$

For all $j = 0, 1, \ldots, k$, we have that

$$\|D_x T^j(v)\|_{T^j(x),\rho} \le \|DT\|^j_{\infty,\rho} \|v\|_{x,\rho} \le \max\{1, \|DT\|^k_{\infty,\rho}\} \|v\|_{x,\rho}.$$

Write $\beta := \max\{1, \|DT\|^k_{\infty,\rho}\}$. By (a), it then follows that

$$\|D_x T^k(v)\|^2_{T^k(x),\rho} \ge 4\|v\|^2_{x,\rho} = 2\|v\|^2_{x,\rho} + \frac{2}{k}\sum_{j=0}^{k-1} \|v\|^2_{x,\rho}$$
$$\ge 2\|v\|^2_{x,\rho} + \frac{2}{k}\sum_{j=0}^{k-1} \beta^{-2} \|D_x T^j(v)\|^2_{T^j(x),\rho}.$$

From this, (6.11) and (6.12), we deduce that

$$\|D_x T(v)\|^2_{T(x),\rho'} = \sum_{j=1}^{k-1} \|D_x T^j(v)\|^2_{T^j(x),\rho} + \|D_x T^k(v)\|^2_{T^k(x),\rho}$$
$$\ge 2\|v\|^2_{x,\rho} + (1 + 2\beta^{-2}k^{-1}) \sum_{j=1}^{k-1} \|D_x T^j(v)\|^2_{T^j(x),\rho}$$

$$
\begin{aligned}
&\geq \left(1+\min\{1,2\beta^{-2}k^{-1}\}\right)\sum_{j=0}^{k-1}\|D_xT^j(v)\|^2_{T^j(x),\rho}\\
&= \left(1+\min\{1,2\beta^{-2}k^{-1}\}\right)\|v\|^2_{x,\rho'}.
\end{aligned}
$$

Taking $\lambda = (1+\min\{1,2\beta^{-2}k^{-1}\})^{1/2} > 1$ completes the proof that (a)⇒(c) and thereby completes the proof of the proposition. □

A Riemannian metric satisfying condition (6.10) is said to be *e-adapted* to T while a corresponding number λ is called an *expanding factor* for this metric.

Corollary 6.2.3. *Being a Shub expanding endomorphism of a compact connected smooth Riemannian manifold M is independent of the Riemannian metric that M is endowed with. More precisely, a map $T : M \to M$ which is a Shub expanding endomorphism with respect to some Riemannian metric on M is a Shub expanding endomorphism with respect to all Riemannian metrics on M.*

As observed in the proof of Proposition 6.2.2 (or as may be readily deduced from part (c) of that proposition), all the maps $D_xT : T_xM \to T_{T(x)}M$, $x \in M$, are linear isomorphisms. Therefore, by virtue of the inverse function theorem, the map T is a local diffeomorphism at every point of M. Since M is compact, T is a covering map. We have thus obtained the following important fact.

Theorem 6.2.4. *Every Shub expanding endomorphism is a covering map.*

We now turn our attention to topological properties of sets of Shub expanding endomorphisms. For every $r \geq 1$, we denote by $\mathcal{E}^r(M)$ the set of all C^r Shub expanding endomorphisms of the manifold M. This set has the remarkable property of being open.

Theorem 6.2.5. *For each $r \geq 1$, the set $\mathcal{E}^r(M)$ is an open subset of the space $C^r(M,M)$ endowed with the C^1 topology.*

Proof. Let

$$
\begin{aligned}
\mathcal{I}^r(M) &:= \{T \in C^r(M,M) \mid D_xT \text{ is invertible}, \forall x \in M\}\\
&= \{T \in C^r(M,M) \mid \det(D_xT) \neq 0, \forall x \in M\}.
\end{aligned}
$$

Since the determinant function $x \mapsto \det(D_xT)$ is continuous on M for any $T \in C^r(M,M)$ and since M is compact, we deduce that

$$
\mathcal{I}^r(M) = \left\{T \in C^r(M,M) : \min_{x\in M}|\det(D_xT)| \neq 0\right\}.
$$

Moreover, since the determinant function $(x,T) \mapsto \det(D_xT)$ is continuous on $M \times C^r(M,M)$, the function $T \mapsto \min_{x\in M}|\det(D_xT)|$ is continuous on $C^r(M,M)$. This guarantees that $\mathcal{I}^r(M)$ is an open subset of $C^r(M,M)$. Notice also that the continuous function $(x,T) \mapsto \det(D_xT)$ does not vanish on $M \times \mathcal{I}^r(M)$. Since the entries of the inverse matrix A^{-1} of an invertible matrix A are polynomials of the entries of A divided by $\det(A)$, the entries of the matrix $(D_xT)^{-1}$ depend continuously on $(x,T) \in M \times \mathcal{I}^r(M)$. Consequently, the function $(x,T) \mapsto \|(D_xT)^{-1}\|$ is continuous on $M \times \mathcal{I}^r(M)$. Thus, the function $T \mapsto \max_{x\in M}\|(D_xT)^{-1}\|$ is continuous on $\mathcal{I}^r(M)$. As observed in (the proof of) Proposition 6.2.2, $\mathcal{E}^r(M) \subseteq \mathcal{I}^r(M)$. Fix $S \in \mathcal{E}^r(M)$. Let ρ be a metric e-adapted to S and λ an expanding factor for this metric. Formula (6.10) implies that $\|(D_xS)^{-1}\| \leq \lambda^{-1}$ for all $x \in M$. The continuity of the function $T \mapsto \max_{x\in M}\|(D_xT)^{-1}\|$ on $\mathcal{I}^r(M)$ ensures the existence of a neighborhood U of S in $\mathcal{I}^r(M)$ such that

$$\max_{x\in M}\|(D_xT)^{-1}\| \leq \frac{\lambda^{-1}+1}{2}$$

for all $T \in U$. But

$$\|v\|_x = \|(D_xT)^{-1}(D_xT(v))\|_x \leq \|(D_xT)^{-1}\| \cdot \|D_xT(v)\|_{T(x)}$$

for every $x \in M$. Therefore,

$$\|D_xT(v)\|_{T(x)} \geq \|(D_xT)^{-1}\|^{-1}\|v\|_x \geq \frac{2}{\lambda^{-1}+1}\|v\|_x$$

for every $x \in M$. Since $2/(\lambda^{-1}+1) > 1$, we conclude that each $T \in U$ is Shub expanding. □

One fundamental fact about continuous maps on a compact connected smooth manifold is that they are homotopic if they are sufficiently close.

Theorem 6.2.6. *If M is a compact connected smooth manifold, then any two continuous maps sufficiently close in $C(M,M)$ are homotopic.*

Proof. Let ρ be a Riemannian metric on M and let $\exp : TM \to M$ be the corresponding exponential map. Let $\exp_x := \exp|_{T_xM}$ for each $x \in M$. Since M is compact, there exists a radius $\delta > 0$ such that for every $x \in M$, the inverse map $\exp_x^{-1} : B_\rho(x,\delta) \to T_xM$ is well-defined and so diffeomorphic. Take any two elements $f, g \in C(M,M)$ such that $\rho_\infty(f,g) < \delta$. Define a map $F : M \times [0,1] \to M$ as follows:

$$F(x,t) := \exp_{f(x)}\big(t\exp_{f(x)}^{-1}(g(x))\big).$$

As a composition of continuous maps, the map F is continuous. Also, $F(x,0) = \exp_{f(x)}(0) = f(x)$ and $F(x,1) = \exp_{f(x)}(\exp_{f(x)}^{-1}(g(x))) = g(x)$. Thus, F is a homotopy from f to g. □

According to Theorems 6.2.5 and 6.2.6, in order to establish the structural stability of Shub expanding endomorphisms, it suffices to prove that any two homotopic Shub expanding endomorphisms are topologically conjugate. This feat will be achieved at the very end of this chapter. Theorem 6.2.6 also partly explains the involvement of algebraic topology, which we will now briefly investigate.

6.3 A digression into algebraic topology

In this section, we develop some algebraically topological results that will be relied upon in the rest of the chapter.

6.3.1 Deck transformations

Let M be a compact connected smooth manifold with Riemannian metric ρ. Let $\widetilde{M}$ be the universal covering space of M and $\pi : \widetilde{M} \to M$ be the canonical projection from $\widetilde{M}$ to M. Every continuous map $\overline{G} : \widetilde{M} \to \widetilde{M}$ such that the diagram

$$\begin{tikzcd} & \widetilde{M} \arrow[d, "\pi"] \\ \widetilde{M} \arrow[ur, "\overline{G}"] \arrow[r, "\pi"'] & M \end{tikzcd}$$

commutes, that is, such that

$$\pi \circ \overline{G} = \pi,$$

is called a *deck transformation* of the manifold M. We adopt the convention of denoting deck transformations with an overline. By the unique lifting property (cf. Proposition 1.34 in [28]), a deck transformation is uniquely determined by its value at any point of $\widetilde{M}$.

Moreover, given any two points $\tilde{x}, \tilde{y} \in \widetilde{M}$ such that $\pi(\tilde{x}) = \pi(\tilde{y})$, by the unique lifting property there exist unique deck transformations $\overline{G}_{\tilde{x},\tilde{y}} : \widetilde{M} \to \widetilde{M}$ and $\overline{G}_{\tilde{y},\tilde{x}} : \widetilde{M} \to \widetilde{M}$ such that

$$\overline{G}_{\tilde{x},\tilde{y}}(\tilde{x}) = \tilde{y} \quad \text{and} \quad \overline{G}_{\tilde{y},\tilde{x}}(\tilde{y}) = \tilde{x}.$$

Consequently, $\overline{G}_{\tilde{y},\tilde{x}} \circ \overline{G}_{\tilde{x},\tilde{y}}$ is a deck transformation such that $\overline{G}_{\tilde{y},\tilde{x}} \circ \overline{G}_{\tilde{x},\tilde{y}}(\tilde{x}) = \tilde{x}$. It follows from the unique lifting property that $\overline{G}_{\tilde{y},\tilde{x}} \circ \overline{G}_{\tilde{x},\tilde{y}} = \mathrm{Id}_{\widetilde{M}}$, and, by the same token, $\overline{G}_{\tilde{x},\tilde{y}} \circ \overline{G}_{\tilde{y},\tilde{x}} = \mathrm{Id}_{\widetilde{M}}$. Therefore, $\overline{G}_{\tilde{x},\tilde{y}}$ is a diffeomorphism of $\widetilde{M}$.

Furthermore, since $\pi : \widetilde{M} \to M$ is a local diffeomorphism, it induces a Riemannian metric $\tilde{\rho}$ on $\widetilde{M}$ defined as follows:

$$\langle w, v \rangle_{\tilde{x},\tilde{\rho}} := \big\langle D_{\tilde{x}}\pi(w), D_{\tilde{x}}\pi(v) \big\rangle_{\pi(\tilde{x}),\rho}, \quad \forall \tilde{x} \in \widetilde{M},\ \forall w, v \in T_{\tilde{x}}\widetilde{M}.$$

With this Riemannian metric on $\widetilde{M}$, the projection map $\pi : \widetilde{M} \to M$ is an infinitesimal and local isometry and all deck transformations of M are infinitesimal and global isometries with respect to the metric $\tilde{\rho}$.

Proposition 6.3.1. *If M is a compact connected smooth manifold with Riemannian metric ρ, then the set D_M of all deck transformations of M is a group of diffeomorphisms (with composition as group action) acting transitively on each fiber of π. Each element of D_M is uniquely determined by its value at any point of $\widetilde{M}$, is an infinitesimal and global $\widetilde{\rho}$-isometry, where $\widetilde{\rho}$ is the metric induced by π and ρ.*

Proof. The only part that remains to be proved is the transitivity. For this, let $\overline{G}, \overline{H} \in D_M$. Then

$$\pi \circ (\overline{G} \circ \overline{H}) = (\pi \circ \overline{G}) \circ \overline{H} = \pi \circ \overline{H} = \pi.$$

That is, the group D_M acts transitively on each fiber of π. □

Later on, we will also need the following result.

Proposition 6.3.2. *If X is a connected Hausdorff topological space and $f, g : X \to \widetilde{M}$ are two continuous maps such that $\pi \circ f = \pi \circ g$, then there exists a unique deck transformation $\overline{G} \in D_M$ such that $\overline{G} \circ f = g$.*

Proof. Fix $x_0 \in M$ and let $\overline{G}$ be the unique element of D_M such that $\overline{G} \circ f(x_0) = g(x_0)$. Let

$$E = \{x \in X \mid \overline{G} \circ f(x) = g(x)\}.$$

Obviously, E is nonempty. It is also closed since its complement is open in X. We shall prove that E is also open. Indeed, let $x \in E$. Since the projection $\pi : \widetilde{M} \to M$ is a local homeomorphism, there exists an open neighborhood $\widetilde{V}$ of $g(x)$ in $\widetilde{M}$ such that the map $\pi|_{\widetilde{V}}$ is one-to-one. As the maps $\overline{G} \circ f$ and g are continuous, there is an open neighborhood U of x in X such that

$$\overline{G} \circ f(U) \subseteq \widetilde{V} \quad \text{and} \quad g(U) \subseteq \widetilde{V}.$$

Let $y \in U$. Then $\overline{G} \circ f(y)$ and $g(y)$ belong to $\widetilde{V}$. Moreover, $\pi(\overline{G} \circ f(y)) = \pi(f(y)) = \pi(g(y))$. Thus $\overline{G} \circ f(y) = g(y)$ by the injectivity of $\pi|_{\widetilde{V}}$. This shows that $y \in E$ and hence $U \subseteq E$, thereby proving that the nonempty, closed set E is also open. Since X is connected, we therefore conclude that $E = X$. The uniqueness of $\overline{G}$ follows immediately from Proposition 6.3.1 since $\overline{G}$ must be the unique deck transformation satisfying $\overline{G}(f(x_0)) = g(x_0)$. □

We now point out a fascinating characterization of the convergence of sequences of deck transformations at any point of the universal covering space.

Lemma 6.3.3. *Every sequence $(\overline{G}_n)_{n=1}^{\infty}$ in D_M that converges at one point of $\widetilde{M}$ is eventually constant.*

Proof. Suppose that there exists $\tilde{x} \in \widetilde{M}$ such that $(\overline{G}_n(\tilde{x}))_{n=1}^{\infty}$ is a convergent sequence in $\widetilde{M}$. Let $x := \pi(\tilde{x}) \in M$. There is $r > 0$ such that the balls $(B_{\tilde{\rho}}(\tilde{y}, r))_{\tilde{y} \in \pi^{-1}(x)}$ are mutually disjoint. Since $\overline{G}_n \in D_M$ for all $n \in \mathbb{N}$, we have $\pi(\overline{G}_n(\tilde{x})) = \pi(\tilde{x}) = x$, that is, $\overline{G}_n(\tilde{x}) \in \pi^{-1}(x)$ for all $n \in \mathbb{N}$. Thus, if $k \in \mathbb{N}$ is so large that $\tilde{\rho}(\overline{G}_i(\tilde{x}), \overline{G}_j(\tilde{x})) < r$ for all $i, j \geq k$, then $\overline{G}_i(\tilde{x}) = \overline{G}_j(\tilde{x})$. Since deck transformations are uniquely determined by their value at any point according to Proposition 6.3.1, we conclude that $\overline{G}_i = \overline{G}_j$ for all $i, j \geq k$. □

As an immediate consequence of this lemma, we obtain the following powerful result.

Corollary 6.3.4. *A sequence $(\overline{G}_n)_{n=1}^{\infty}$ in D_M converges uniformly on compact subsets of $\widetilde{M}$ if and only if it is eventually constant.*

Proof. If a sequence $(\overline{G}_n)_{n=1}^{\infty}$ in D_M converges uniformly on compact subsets of $\widetilde{M}$, then it converges pointwise on $\widetilde{M}$. Therefore, it is eventually constant according to Lemma 6.3.3. Obviously, any eventually constant sequence converges uniformly on compact subsets. □

Theoretically, establishing uniform convergence on compact subsets may prove to be difficult. Fortunately, there exists a simpler, pointwise criterion for sequences of deck transformations.

Lemma 6.3.5. *If $(\overline{G}_n)_{n=1}^{\infty}$ is a sequence in D_M and if $(\tilde{z}_n)_{n=1}^{\infty}$ is a sequence of points in $\widetilde{M}$ converging to some point $\tilde{z} \in \widetilde{M}$ such that $\tilde{w} := \lim_{n\to\infty} \overline{G}_n(\tilde{z}_n)$ exists, then the sequence $(\overline{G}_n)_{n=1}^{\infty}$ converges uniformly on compact subsets of $\widetilde{M}$ to an element $\overline{G} \in D_M$, which is uniquely determined by the requirement that $\overline{G}(\tilde{z}) = \tilde{w}$. In fact, the sequence $(\overline{G}_n)_{n=1}^{\infty}$ is eventually constant. More precisely, its terms eventually coincide with the unique element $\overline{G} \in D_M$ such that $\overline{G}(\tilde{z}) = \tilde{w}$.*

Proof. Let

$$\xi = \max\Big\{\sup\{\tilde{\rho}(\tilde{z}_n, \tilde{z}) : n \in \mathbb{N}\}, \sup\{\tilde{\rho}(\overline{G}_n(\tilde{z}_n), \tilde{w}) : n \in \mathbb{N}\}\Big\} < \infty.$$

Fix $r > 0$ and take $\tilde{x} \in \overline{B}_{\tilde{\rho}}(\tilde{z}, r)$. Then for every $n \in \mathbb{N}$, we have that

$$\begin{aligned}\tilde{\rho}(\overline{G}_n(\tilde{x}), \tilde{w}) &\leq \tilde{\rho}(\overline{G}_n(\tilde{x}), \overline{G}_n(\tilde{z})) + \tilde{\rho}(\overline{G}_n(\tilde{z}), \overline{G}_n(\tilde{z}_n)) + \tilde{\rho}(\overline{G}_n(\tilde{z}_n), \tilde{w})\\ &= \tilde{\rho}(\tilde{x}, \tilde{z}) + \tilde{\rho}(\tilde{z}, \tilde{z}_n) + \tilde{\rho}(\overline{G}_n(\tilde{z}_n), \tilde{w}) \leq r + \xi + \xi = 2\xi + r.\end{aligned}$$

This means that for all $n \in \mathbb{N}$,

$$\overline{G}_n(\overline{B}_{\tilde{\rho}}(\tilde{z}, r)) \subseteq \overline{B}_{\tilde{\rho}}(\tilde{w}, 2\xi + r). \tag{6.13}$$

Now let $(n_k)_{k=1}^{\infty}$ be a strictly increasing sequence of positive integers. Fix $j \in \mathbb{N}$ and suppose that we have extracted from $(n_k)_{k=1}^{\infty}$ a subsequence $(n_k^{(j)})_{k=1}^{\infty}$ such that the sequence $(\overline{G}_{n_k^{(j)}})_{k=1}^{\infty}$ converges uniformly on $\overline{B}_{\tilde{\rho}}(\tilde{z}, j)$ to some continuous map

$\overline{G}^{(j)} : \overline{B}_{\widetilde{\rho}}(\widetilde{z}, j) \to \overline{B}_{\widetilde{\rho}}(\widetilde{w}, 2\xi + j)$. The inductive step, which also provides the basis of the induction, is as follows. Since both balls $\overline{B}_{\widetilde{\rho}}(\widetilde{z}, j+1)$ and $\overline{B}_{\widetilde{\rho}}(\widetilde{w}, 2\xi + j + 1)$ are compact and since the sequence of $\widetilde{\rho}$-isometries $(\overline{G}_n)_{n=1}^{\infty}$ forms an equicontinuous family of maps, Arzelà–Ascoli's theorem permits us to extract from $(n_k^{(j)})_{k=1}^{\infty}$ a subsequence $(n_k^{(j+1)})_{k=1}^{\infty}$ such that the sequence $(\overline{G}_{n_k^{(j+1)}})_{k=1}^{\infty}$ converges uniformly on $\overline{B}_{\widetilde{\rho}}(\widetilde{z}, j+1)$ to some continuous map $\overline{G}^{(j+1)} : \overline{B}_{\widetilde{\rho}}(\widetilde{z}, j+1) \to \overline{B}_{\widetilde{\rho}}(\widetilde{w}, 2\xi + j + 1)$.

Obviously, $\overline{G}^{(j+1)}|_{\overline{B}_{\widetilde{\rho}}(\widetilde{z},j)} = \overline{G}^{(j)}$ and gluing all the maps $(\overline{G}^{(j)})_{j=1}^{\infty}$ together results in a map $\overline{G} : \widetilde{M} \to \widetilde{M}$ defined as $\overline{G}(\widetilde{x}) = \overline{G}^{(j)}(\widetilde{x})$ if $\widetilde{x} \in \overline{B}_{\widetilde{\rho}}(\widetilde{z}, j)$. The sequence $(\overline{G}_{n_j^{(j)}})_{j=1}^{\infty}$ is a subsequence of $(\overline{G}_n)_{n=1}^{\infty}$ that converges uniformly to $\overline{G}$ on every compact ball $\overline{B}_{\widetilde{\rho}}(\widetilde{z}, i)$, $i \in \mathbb{N}$. This means that $(\overline{G}_{n_j^{(j)}})_{j=1}^{\infty}$ converges to $\overline{G}$ uniformly on compact subsets of $\widetilde{M}$. So $\overline{G}$ is a continuous map from $\widetilde{M}$ to $\widetilde{M}$ and

$$\pi \circ \overline{G}(\widetilde{x}) = \pi\Big(\lim_{j\to\infty} \overline{G}_{n_j^{(j)}}(\widetilde{x})\Big) = \lim_{j\to\infty} \big(\pi \circ \overline{G}_{n_j^{(j)}}(\widetilde{x})\big) = \lim_{j\to\infty} \pi(\widetilde{x}) = \pi(\widetilde{x}).$$

Hence $\overline{G} \in D_M$ and

$$\overline{G}(\widetilde{z}) = \lim_{j\to\infty} \overline{G}_{n_j^{(j)}}(\widetilde{z}) = \lim_{j\to\infty} \overline{G}_{n_j^{(j)}}(\widetilde{z}_{n_j^{(j)}}) = \lim_{n\to\infty} \overline{G}_n(\widetilde{z}_n) = \widetilde{w}.$$

The uniqueness of $\overline{G}$ follows from Proposition 6.3.1. The rest of the proposition follows from Corollary 6.3.4. □

6.3.2 Lifts

We now study the concept of the *lift of a map*. We keep with the convention adopted in Chapter 2 of denoting a lift of a given map with a tilde above the map.

Proposition 6.3.6. *Let N and M be two compact connected smooth manifolds. For any continuous map $S : N \to M$ there exists a continuous map $\widetilde{S} : \widetilde{N} \to \widetilde{M}$ such that the following diagram commutes:*

$$\begin{tikzcd} \widetilde{N} \arrow[r, "\widetilde{S}"] \arrow[d, "\pi_N"'] & \widetilde{M} \arrow[d, "\pi_M"] \\ N \arrow[r, "S"'] & M \end{tikzcd}$$

That is,

$$S \circ \pi_N = \pi_M \circ \widetilde{S}.$$

All such maps $\widetilde{S}$ are called lifts *of S.*

Proof. Denote by $\pi_1(X)$ be the fundamental group of a path-connected space X and let $(\pi_X)_* : \pi_1(\widetilde{X}) \to \pi_1(X)$ be the homomorphism induced by the canonical projection $\pi_X : \widetilde{X} \to X$, where $\widetilde{X}$ is the universal covering space of X.

Consider the diagram

$$\begin{tikzcd} & & \widetilde{M} \arrow[d, "\pi_M"] \\ \widetilde{N} \arrow[r, "\pi_N"'] & N \arrow[r, "S"'] & M \end{tikzcd}$$

Notice that $(S \circ \pi_N)_*(\pi_1(\widetilde{N})) = \{0\}_M = (\pi_M)_*(\{0\}_M) = (\pi_M)_*(\pi_1(\widetilde{M}))$. From the lifting criterion (cf. Proposition 1.33 in [28]), there thus exists a continuous map $\widetilde{S} : \widetilde{N} \to \widetilde{M}$ such that the following diagram commutes:

$$\begin{tikzcd} & & \widetilde{M} \arrow[d, "\pi_M"] \\ \widetilde{N} \arrow[urr, "\widetilde{S}"] \arrow[r, "\pi_N"'] & N \arrow[r, "S"'] & M \end{tikzcd}$$

□

One particularly interesting case is the lifting of covering maps.

Proposition 6.3.7. *If $S : N \to M$ is a covering map, then all of its lifts $\widetilde{S} : \widetilde{N} \to \widetilde{M}$ are homeomorphisms.*

Proof. Recall from the proof of Proposition 6.3.6 that

$$(S \circ \pi_N)_*\big(\pi_1(\widetilde{N})\big) = (\pi_M)_*\big(\pi_1(\widetilde{M})\big).$$

From the lifting criterion (cf. Proposition 1.33 in [28]), there then exists a continuous map $\widehat{S} : \widetilde{M} \to \widetilde{N}$ such that the following diagram commutes:

$$\begin{tikzcd} & \widetilde{N} \arrow[d, "\pi_N"] \\ & N \arrow[d, "S"] \\ \widetilde{M} \arrow[uur, "\widehat{S}"] \arrow[r, "\pi_M"'] & M \end{tikzcd}$$

Using Proposition 6.3.6, let $\widetilde{S} : \widetilde{N} \to \widetilde{M}$ be a lift of $S : N \to M$. Then

$$\pi_M \circ (\widetilde{S} \circ \widehat{S}) = (\pi_M \circ \widetilde{S}) \circ \widehat{S} = (S \circ \pi_N) \circ \widehat{S} = \pi_M.$$

Hence, in view of Proposition 6.3.1, the map $\widetilde{S} \circ \widehat{S}$ is a homeomorphism. On the other hand, $(S \circ \pi_N) \circ (\widehat{S} \circ \widetilde{S}) = (S \circ \pi_N \circ \widehat{S}) \circ \widetilde{S} = \pi_M \circ \widetilde{S} = S \circ \pi_N$. Since $S \circ \pi_N : \widetilde{N} \to M$ is a covering map, an argument similar to the one yielding Proposition 6.3.1 certifies that $\widehat{S} \circ \widetilde{S}$ is a homeomorphism. As $\widetilde{S} \circ \widehat{S}$ and $\widehat{S} \circ \widetilde{S}$ are homeomorphisms, so are $\widehat{S}$ and $\widetilde{S}$. □

Note that deck transformations of M are simply lifts of the identity covering map $\mathrm{Id}_M : M \to M$.

We now provide a characterization of lifts.

Proposition 6.3.8. *Let N and M be two compact connected smooth manifolds. A continuous map $\widetilde{S} : \widetilde{N} \to \widetilde{M}$ is a lift of some continuous map from N to M if and only if there exists a (necessarily unique) map $h : D_N \to D_M$ such that the following diagram commutes for all $\overline{F} \in D_N$:*

$$\begin{tikzcd}
\widetilde{N} \arrow[r, "\overline{F}"] \arrow[d, "\widetilde{S}"'] & \widetilde{N} \arrow[d, "\widetilde{S}"] \\
\widetilde{M} \arrow[r, "h(\overline{F})"'] & \widetilde{M}
\end{tikzcd}$$

In other words,

$$\widetilde{S} \circ \overline{F} = h(\overline{F}) \circ \widetilde{S}, \quad \forall \overline{F} \in D_N.$$

Moreover, $h : D_N \to D_M$ is a group homomorphism. This induced homomorphism of the groups of deck transformations will also be denoted by $\widetilde{S}^*$.

Proof. First, suppose that $S : N \to M$ is a continuous map which has for lift $\widetilde{S} : \widetilde{N} \to \widetilde{M}$. For any $\overline{F} \in D_N$, we then have that

$$\pi_M \circ (\widetilde{S} \circ \overline{F}) = (\pi_M \circ \widetilde{S}) \circ \overline{F} = (S \circ \pi_N) \circ \overline{F} = S \circ (\pi_N \circ \overline{F}) = S \circ \pi_N = \pi_M \circ \widetilde{S}.$$

It follows from Proposition 6.3.2 that there exists a unique $h(\overline{F}) \in D_M$ such that $h(\overline{F}) \circ \widetilde{S} = \widetilde{S} \circ \overline{F}$.

For the converse implication, suppose that $\widetilde{S} : \widetilde{N} \to \widetilde{M}$ is a continuous map and that there is a map $h : D_N \to D_M$ such that $\widetilde{S} \circ \overline{F} = h(\overline{F}) \circ \widetilde{S}$ for all $\overline{F} \in D_N$. Define the map $S : N \to M$ by setting

$$S(x) := \pi_M \circ \widetilde{S}(\widetilde{x}),$$

where $\widetilde{x}$ is any element of $\pi_N^{-1}(x)$. To be well-defined, we must show that $S(x)$ is independent of the element $\widetilde{x}$ chosen in $\pi_N^{-1}(x)$. For this, let $\widetilde{z} \in \pi_N^{-1}(x)$. Since $\pi_N(\widetilde{x}) = \pi_N(\widetilde{z})$, there is a unique $\overline{F} \in D_N$ such that $\widetilde{z} = \overline{F}(\widetilde{x})$. Then

$$\pi_M \circ \widetilde{S}(\widetilde{z}) = \pi_M \circ \widetilde{S} \circ \overline{F}(\widetilde{x}) = \pi_M \circ h(\overline{F}) \circ \widetilde{S}(\widetilde{x}) = \pi_M \circ \widetilde{S}(\widetilde{x}).$$

The map $S : N \to M$ is thus well-defined. It is continuous since $\pi_M \circ \widetilde{S}$ is continuous and the projection $\pi_N : \widetilde{N} \to N$ is a covering map. Furthermore, $S \circ \pi_N(\widetilde{x}) = S(x) = \pi_M \circ \widetilde{S}(\widetilde{x})$. So $\widetilde{S}$ is a lift of S.

Regarding the last assertion, for any $\overline{F}_1, \overline{F}_2 \in D_N$ we have that $\overline{F}_1 \circ \overline{F}_2 \in D_N$. As $h : D_N \to D_M$ is the unique map such that $h(\overline{F}) \circ \widetilde{S} = \widetilde{S} \circ \overline{F}$ for all $\overline{F} \in D_N$, it ensues that

$$\big(h(\overline{F}_1 \circ \overline{F}_2)\big) \circ \widetilde{S} = \widetilde{S} \circ \overline{F}_1 \circ \overline{F}_2 = h(\overline{F}_1) \circ \widetilde{S} \circ \overline{F}_2 = h(\overline{F}_1) \circ h(\overline{F}_2) \circ \widetilde{S} = \big(h(\overline{F}_1) \circ h(\overline{F}_2)\big) \circ \widetilde{S}.$$

The uniqueness of h implies that $h(\overline{F}_1 \circ \overline{F}_2) = h(\overline{F}_1) \circ h(\overline{F}_2)$. So h is a homomorphism between the groups D_N and D_M. $\square$

Given a certain lift, we next show that the set of all lifts that share the same induced homomorphism as the given lift naturally forms a complete metric space.

Let N and M be two compact connected smooth manifolds and let $\widetilde{\alpha} : \widetilde{N} \to \widetilde{M}$ be a lift of some continuous map from N to M. Let $V_{\widetilde{\alpha}}$ denote the set of all continuous maps $\widetilde{S} : \widetilde{N} \to \widetilde{M}$ that are lifts of some continuous map from N to M and such that $\widetilde{S}^* = \widetilde{\alpha}^*$. Let $\widetilde{\rho}_\infty : V_{\widetilde{\alpha}} \times V_{\widetilde{\alpha}} \to [0, \infty]$ be the function defined by

$$\widetilde{\rho}_\infty(\widetilde{S}_1, \widetilde{S}_2) := \sup\{\widetilde{\rho}_M(\widetilde{S}_1(\widetilde{x}), \widetilde{S}_2(\widetilde{x})) \mid \widetilde{x} \in \widetilde{N}\}.$$

Lemma 6.3.9. *The function $\widetilde{\rho}_\infty$ constitutes a metric on the set $V_{\widetilde{\alpha}}$ and the metric space $(V_{\widetilde{\alpha}}, \widetilde{\rho}_\infty)$ is complete.*

Proof. The symmetry of $\widetilde{\rho}_\infty$ and the triangle inequality being obvious, we shall only demonstrate the finiteness of $\widetilde{\rho}_\infty$ to establish that $\widetilde{\rho}_\infty$ is a metric on $V_{\widetilde{\alpha}}$.

Let $R > \operatorname{diam}_{\rho_N}(N)$. We first show that

$$\pi_N(B_{\widetilde{\rho}_N}(\widetilde{z}, R)) = N, \quad \forall\, \widetilde{z} \in \widetilde{N}. \tag{6.14}$$

To that end, fix $\widetilde{z} \in \widetilde{N}$ and $x \in N$. Let γ be a smooth curve in N joining $\pi_N(\widetilde{z})$ and x whose ρ_N-length is smaller than R. Let $\widetilde{\gamma}$ be a lift of γ to $\widetilde{N}$ whose initial point is $\widetilde{z}$. Let $\widetilde{w}$ denote the other endpoint of the curve γ. Since $\pi_N : \widetilde{N} \to N$ is a local (and thus infinitesimal) isometry, we deduce that

$$\widetilde{\rho}_N(\widetilde{z}, \widetilde{w}) \le \ell_{\widetilde{\rho}_N}(\widetilde{\gamma}) = \ell_{\rho_N}(\gamma) < R.$$

Since $x = \pi_N(\widetilde{w})$, it follows that $x \in \pi_N(B_{\widetilde{\rho}_N}(\widetilde{z}, R))$, and thus (6.14) holds.

Now, let $\widetilde{S}_1, \widetilde{S}_2 \in V_{\widetilde{\alpha}}$. Since $\overline{B}_{\widetilde{\rho}_N}(\widetilde{z}, R)$ is a compact subset of $\widetilde{N}$, we have

$$A := \sup\{\widetilde{\rho}_M(\widetilde{S}_1(\widetilde{x}), \widetilde{S}_2(\widetilde{x})) \mid \widetilde{x} \in \overline{B}_{\widetilde{\rho}_N}(\widetilde{z}, R)\} < \infty.$$

Take an arbitrary point $\widetilde{w} \in \widetilde{N}$. In light of (6.14), there exists a point $\widetilde{x} \in B_{\widetilde{\rho}_N}(\widetilde{z}, R)$ such that $\pi_N(\widetilde{x}) = \pi_N(\widetilde{w})$. Hence, there exists a deck transformation $\overline{F} \in D_N$ such that $\overline{F}(\widetilde{x}) = \widetilde{w}$. Consequently,

$$\begin{aligned}
\widetilde{\rho}_M(\widetilde{S}_1(\widetilde{w}), \widetilde{S}_2(\widetilde{w})) &= \widetilde{\rho}_M(\widetilde{S}_1 \circ \overline{F}(\widetilde{x}), \widetilde{S}_2 \circ \overline{F}(\widetilde{x})) \\
&= \widetilde{\rho}_M(\widetilde{S}_1^*(\overline{F}) \circ \widetilde{S}_1(\widetilde{x}), \widetilde{S}_2^*(\overline{F}) \circ \widetilde{S}_2(\widetilde{x})) \\
&= \widetilde{\rho}_M(\widetilde{\alpha}^*(\overline{F}) \circ \widetilde{S}_1(\widetilde{x}), \widetilde{\alpha}^*(\overline{F}) \circ \widetilde{S}_2(\widetilde{x})) \\
&= \widetilde{\rho}_M(\widetilde{S}_1(\widetilde{x}), \widetilde{S}_2(\widetilde{x})) \le A.
\end{aligned}$$

Thus, $\widetilde{\rho}_\infty(\widetilde{S}_1, \widetilde{S}_2) \le A < \infty$ and thereby $\widetilde{\rho}_\infty(V_{\widetilde{\alpha}} \times V_{\widetilde{\alpha}}) \subseteq [0, \infty)$. So $\widetilde{\rho}_\infty$ is a metric.

Let us now show that $\widetilde{\rho}_\infty$ is complete. Let $(\widetilde{S}_n)_{n=1}^\infty$ be a Cauchy sequence in $V_{\widetilde{\alpha}}$. Then this sequence is also a Cauchy sequence with respect to the topology of uniform convergence on compact subsets of $\widetilde{N}$. Let $\widehat{S} : \widetilde{N} \to \widetilde{M}$ be the limit of that sequence. Let $\overline{F} \in D_N$. Since $\widetilde{S}_n \circ \overline{F} = \widetilde{\alpha}^*(\overline{F}) \circ \widetilde{S}_n$ for all $n \in \mathbb{N}$, we infer that $\widehat{S} \circ \overline{F} = \widetilde{\alpha}^*(\overline{F}) \circ \widehat{S}$. By Proposition 6.3.8, we deduce that $\widehat{S} \in V_{\widetilde{\alpha}}$. Let $\widetilde{z} \in \widetilde{N}$ and $R > \operatorname{diam}_{\rho_N}(N)$. As established above,

$$\widetilde{\rho}_\infty(\widetilde{S}_n, \widehat{S}) \leq \sup\{\widetilde{\rho}_M(\widetilde{S}_n(\widetilde{x}), \widehat{S}(\widetilde{x})) \mid \widetilde{x} \in \overline{B}_{\widetilde{\rho}_N}(\widetilde{z}, R)\}$$

for all $n \in \mathbb{N}$, whence the uniform convergence of the sequence $(\widetilde{S}_n)_{n=1}^{\infty}$ to $\widehat{S}$ on the compact ball $\overline{B}_{\widetilde{\rho}_N}(\widetilde{z}, R)$ implies that it also converges to $\widehat{S}$ with respect to the metric $\widetilde{\rho}_\infty$ on $V_{\widetilde{\alpha}}$. □

6.4 Dynamical properties

In the last section of this chapter, we will discover the dynamical properties of Shub expanding endomorphisms.

6.4.1 Expanding property

Shub expanding endomorphisms are distance expanding in the following sense.

Theorem 6.4.1. *Every Shub expanding endomorphism $T : M \to M$ is distance expanding with respect to the distance ρ induced on M by any Riemannian metric ρ e-adapted to T.*

Proof. According to Theorem 6.2.4, every Shub expanding endomorphism T is a covering map, and thus a local homeomorphism. Thanks to the compactness of M, there then exists some $\delta_T > 0$ such that the map $T|_{B_\rho(x,2\delta_T)} : B_\rho(x, 2\delta_T) \to M$ is injective for all $x \in M$. Fix two points $x_1, x_2 \in M$ such that $\rho(x_1, x_2) < 2\delta_T$ and pick any smooth curve $\gamma : I \to M$ joining $T(x_1)$ and $T(x_2)$, that is, $\gamma(a) = T(x_1)$ and $\gamma(b) = T(x_2)$, where $I = [a, b] \subseteq \mathbb{R}$. Since $T : M \to M$ is a covering map, there exists a smooth curve $\widehat{\gamma} : I \to M$ such that $\widehat{\gamma}(a) = x_1$ and $T \circ \widehat{\gamma} = \gamma$. In particular, $T(\widehat{\gamma}(b)) = T(x_2)$. So if $\widehat{\gamma}(b) \notin B_\rho(x_1, 2\delta_T)$, then

$$\rho(\widehat{\gamma}(a), \widehat{\gamma}(b)) = \rho(x_1, \widehat{\gamma}(b)) \geq 2\delta_T > \rho(x_1, x_2).$$

On the other hand, if $\widehat{\gamma}(b) \in B_\rho(x_1, 2\delta_T)$, then $\widehat{\gamma}(b) = x_2$ since the map $T|_{B_\rho(x_1,2\delta_T)}$ is injective, whence $\rho(\widehat{\gamma}(a), \widehat{\gamma}(b)) = \rho(x_1, x_2)$. In either case,

$$\ell_\rho(\widehat{\gamma}) \geq \rho(\widehat{\gamma}(a), \widehat{\gamma}(b)) \geq \rho(x_1, x_2).$$

Therefore,

$$\begin{aligned} \ell_\rho(\gamma) &= \int_I \|D_t\gamma(e_t)\|\,dt = \int_I \|D_t(T \circ \widehat{\gamma})(e_t)\|\,dt \\ &= \int_I \|D_{\widehat{\gamma}(t)}T(D_t\widehat{\gamma}(e_t))\|\,dt \\ &\geq \lambda \int_I \|D_t\widehat{\gamma}(e_t)\|\,dt = \lambda\ell_\rho(\widehat{\gamma}) \geq \lambda\rho(x_1, x_2). \end{aligned}$$

So, taking the infimum over all curves $\gamma \in \Gamma(T(x_1), T(x_2))$, we conclude that $\rho(T(x_1), T(x_2)) \geq \lambda\rho(x_1, x_2)$ for all $x_1, x_2 \in M$ such that $\rho(x_1, x_2) < 2\delta_T$. □

The previous theorem shows that Shub expanding endomorphisms provide a large class of distance expanding maps.

6.4.2 Topological exactness and density of periodic points

The next theorem reveals several additional dynamical properties of Shub expanding maps.

Theorem 6.4.2. *Let M be a compact connected smooth manifold. If $T : M \to M$ is a Shub expanding endomorphism, then:*
(a) *T has a fixed point.*
(b) *The universal covering manifold $\widetilde{M}$ is diffeomorphic to $\mathbb{R}^k$, where $k = \dim(M)$.*
(c) *T is topologically exact.*
(d) *The set of periodic points of T is dense in M.*

Proof. Let ρ be a Riemannian metric on M e-adapted to T, and $\widetilde{\rho}$ the Riemannian metric induced by π and ρ on $\widetilde{M}$. Recall that with these metrics, the projection map $\pi : \widetilde{M} \to M$ is an infinitesimal and local isometry and all maps in D_M are infinitesimal and global isometries with respect to the metric $\widetilde{\rho}$. Let $\widetilde{T} : \widetilde{M} \to \widetilde{M}$ be a lift of T to $\widetilde{M}$. Such a lift exists according to Proposition 6.3.6. By Theorem 6.2.4 and Proposition 6.3.7, the map $\widetilde{T}$ is a diffeomorphism. With a calculation analogous to that in the proof of Theorem 6.4.1, we can prove the following claim.

Claim. *The diffeomorphism $\widetilde{T}^{-1} : \widetilde{M} \to \widetilde{M}$ is a global contraction with respect to the metric $\widetilde{\rho}$. More precisely,*

$$\widetilde{\rho}\big(\widetilde{T}^{-1}(\widetilde{x}), \widetilde{T}^{-1}(\widetilde{y})\big) \leq \lambda^{-1}\widetilde{\rho}(\widetilde{x}, \widetilde{y}), \quad \forall \widetilde{x}, \widetilde{y} \in \widetilde{M}. \tag{6.15}$$

(a) By the Banach contraction principle, the map $\widetilde{T}^{-1} : \widetilde{M} \to \widetilde{M}$ has a unique fixed point $\widetilde{w} \in \widetilde{M}$. That is, $\widetilde{T}^{-1}(\widetilde{w}) = \widetilde{w}$, or, equivalently, $\widetilde{T}(\widetilde{w}) = \widetilde{w}$. Then $T(\pi(\widetilde{w})) = \pi(\widetilde{T}(\widetilde{w})) = \pi(\widetilde{w})$, that is, $\pi(\widetilde{w})$ is a fixed point of $T : M \to M$.

(b) Let $\widetilde{w} \in \widetilde{M}$ be the fixed point of the maps $\widetilde{T}, \widetilde{T}^{-1} : \widetilde{M} \to \widetilde{M}$. Since $\widetilde{M}$ is a smooth manifold, there exist $r > 0$ and a smooth diffeomorphism $\varphi : V \to B_{\widetilde{\rho}}(\widetilde{w}, r)$ from an open neighborhood V of the origin in $\mathbb{R}^k$ onto $B_{\widetilde{\rho}}(\widetilde{w}, r)$ and such that $\varphi(0) = \widetilde{w}$ and $\varphi'(0) : \mathbb{R}^k \to T_{\widetilde{w}}\widetilde{M}$ is an isometry. Since

$$\widetilde{T}^{-1}\big(B_{\widetilde{\rho}}(\widetilde{w}, r)\big) \subseteq B_{\widetilde{\rho}}\big(\widetilde{w}, \lambda^{-1}r\big) \subseteq B_{\widetilde{\rho}}(\widetilde{w}, r),$$

the conjugate of $\widetilde{T}^{-1}$ via φ is a well-defined diffeomorphism, namely

$$G := \varphi^{-1} \circ \widetilde{T}^{-1} \circ \varphi : V \to V.$$

Hence, for all $k \geq 0$,

$$G^k = \varphi^{-1} \circ \widetilde{T}^{-k} \circ \varphi : V \to V. \tag{6.16}$$

Notice that

$$G(0) = \varphi^{-1} \circ \widetilde{T}^{-1} \circ \varphi(0) = \varphi^{-1}\big(\widetilde{T}^{-1}(\widetilde{w})\big) = \varphi^{-1}(\widetilde{w}) = 0$$

and

$$\|G'(0)\| = \big\|(\varphi^{-1})'(\widetilde{w}) \circ (\widetilde{T}^{-1})'(\widetilde{w}) \circ \varphi'(0)\big\| = \big\|(\widetilde{T}^{-1})'(\widetilde{w})\big\| \leq \lambda^{-1} < 1.$$

Therefore, there exists $R > 0$ so small that $\overline{B}(0, R) \subseteq V$,

$$G\big(B(0,R)\big) \subseteq B\left(0, \frac{\lambda^{-1}+1}{2}R\right) \subseteq B(0,R), \tag{6.17}$$

and

$$\|G'(x)\| \leq \frac{\lambda^{-1}+1}{2}, \quad \forall x \in B(0,R). \tag{6.18}$$

In view of Exercise 6.5.5, there then exists a diffeomorphism $\widehat{G} : \mathbb{R}^k \to \mathbb{R}^k$ such that

$$\widehat{G}|_{B(0,R)} = G \quad \text{and} \quad \big\|\widehat{G}'(x)\big\| \leq \frac{\lambda^{-1}+1}{2}, \quad \forall x \in \mathbb{R}^k. \tag{6.19}$$

Let us now define a map $H : \widetilde{M} \to \mathbb{R}^k$ in the following way. For each $\widetilde{x} \in \widetilde{M}$, choose $n = n(\widetilde{x}) \geq 0$ such that $\widetilde{T}^{-n}(\widetilde{x}) \in \varphi(B(0,R))$ and declare

$$H(\widetilde{x}) = \widehat{G}^{-n} \circ \varphi^{-1} \circ \widetilde{T}^{-n}(\widetilde{x}). \tag{6.20}$$

We shall first show that $H(\widetilde{x})$ is well-defined by establishing that its definition is independent of the choice of n. Then we will proceed on showing that $H : \widetilde{M} \to \mathbb{R}^k$ is a diffeomorphism. So assume that, in addition to $\widetilde{T}^{-n}(\widetilde{x})$, the iterate $\widetilde{T}^{-j}(\widetilde{x})$ is in $\varphi(B(0,R))$. We may assume without loss of generality that $0 \leq j \leq n$. Write $\widetilde{T}^{-j}(\widetilde{x}) = \varphi(x')$, where $x' \in B(0,R) \subseteq V$. Using (6.16) with $k = n - j$, (6.17) and (6.19), we get

$$\begin{aligned}
\widehat{G}^{-n} \circ \varphi^{-1} \circ \widetilde{T}^{-n}(\widetilde{x}) &= \widehat{G}^{-n} \circ \varphi^{-1} \circ \widetilde{T}^{-(n-j)} \circ \widetilde{T}^{-j}(\widetilde{x}) \\
&= \widehat{G}^{-n} \circ \varphi^{-1} \circ \widetilde{T}^{-(n-j)} \circ \varphi(x') \\
&= \widehat{G}^{-n} \circ G^{n-j}(x') = \widehat{G}^{-j}(x') = \widehat{G}^{-j} \circ \varphi^{-1} \circ \widetilde{T}^{-j}(\widetilde{x}).
\end{aligned}$$

Thus, the map $H : \widetilde{M} \to \mathbb{R}^k$ is well-defined. Since the same n used to define H at $\widetilde{x}$ works for any point $\widetilde{y} \in B_{\widetilde{\rho}}(\widetilde{x}, \varepsilon)$ if $\varepsilon > 0$ is sufficiently small, it follows from (6.20) that the map H is smooth as a composition of smooth maps. As a composition of local diffeomorphisms, it is further a local diffeomorphism. It only remains to show that H

is globally bijective. To prove injectivity, assume that $H(\tilde{x}) = H(\tilde{y})$. Since $\tilde{T}^{-1} : \widetilde{M} \to \widetilde{M}$ is a global contraction fixing $\tilde{w} = \varphi(0)$, there exists $n \geq 0$ so large that both $\tilde{T}^{-n}(\tilde{x})$ and $\tilde{T}^{-n}(\tilde{y})$ lie in $\varphi(B(0, R))$. Then

$$\widehat{G}^{-n} \circ \varphi^{-1} \circ \tilde{T}^{-n}(\tilde{x}) = \widehat{G}^{-n} \circ \varphi^{-1} \circ \tilde{T}^{-n}(\tilde{y}).$$

Applying to this equality $\widehat{G}^{n}$, φ and $\tilde{T}^{n}$ successively, we conclude that $\tilde{x} = \tilde{y}$, thereby establishing the injectivity of H. To prove its surjectivity, take an arbitrary $y \in \mathbb{R}^k$. Since $\widehat{G} : \mathbb{R}^k \to \mathbb{R}^k$ is a global contraction according to (6.19) and has 0 for fixed point, there exists $n \geq 0$ so large that $\widehat{G}^n(y) \in B(0, R)$. Then $\varphi(\widehat{G}^n(y)) \in \varphi(B(0, R))$. It follows that

$$H(\tilde{T}^n(\varphi(\widehat{G}^n(y)))) = \widehat{G}^{-n} \circ \varphi^{-1} \circ \tilde{T}^{-n}(\tilde{T}^n \circ \varphi \circ \widehat{G}^n(y)) = y.$$

Thus, H is surjective. Because $H : \widetilde{M} \to \mathbb{R}^k$ is a diffeomorphism, $k = \dim(\widetilde{M}) = \dim(M)$.

(c) Let $R > \operatorname{diam}_\rho(M)$ and recall that (6.14) then holds with $N = M$. Let U be a nonempty, open subset of M. Fix an arbitrary $x \in U$ and any $\tilde{x} \in \pi^{-1}(x)$. Since $\pi^{-1}(U)$ is an open subset of $\widetilde{M}$ containing $\tilde{x}$, there exists some $r > 0$ such that $B_{\tilde{\rho}}(\tilde{x}, r) \subseteq \pi^{-1}(U)$. Choose $n \geq 0$ so large that $\lambda^n r \geq R$. By (6.15), we observe that

$$\tilde{T}^n(B_{\tilde{\rho}}(\tilde{x}, r)) \supseteq B_{\tilde{\rho}}(\tilde{T}^n(\tilde{x}), \lambda^n r) \supseteq B_{\tilde{\rho}}(\tilde{T}^n(\tilde{x}), R).$$

It follows from (6.14) that

$$T^n(U) \supseteq T^n(\pi(B_{\tilde{\rho}}(\tilde{x}, r))) = \pi(\tilde{T}^n(B_{\tilde{\rho}}(\tilde{x}, r))) \supseteq \pi(B_{\tilde{\rho}}(\tilde{T}^n(\tilde{x}), R)) = M.$$

Thus, T is topologically exact.

(d) As in part (c), let $R > \operatorname{diam}_\rho(M)$ and recall that (6.14) then holds. Let also U be a nonempty, open subset of M. Fix an arbitrary $x \in U$ and $\tilde{x} \in \pi^{-1}(x)$. Since $\pi^{-1}(U)$ is an open subset of $\widetilde{M}$ containing $\tilde{x}$, there exists some $0 < r \leq R$ such that $\overline{B}_{\tilde{\rho}}(\tilde{x}, r) \subseteq \pi^{-1}(U)$. Choose $n \geq 0$ so large that $\lambda^n r \geq 2R$. By (6.14), there exists $\tilde{y} \in B_{\tilde{\rho}}(\tilde{T}^n(\tilde{x}), R)$ such that $\pi(\tilde{y}) = x = \pi(\tilde{x})$ and hence there is $\overline{G} \in D_M$ such that $\overline{G}(\tilde{x}) = \tilde{y} \in B_{\tilde{\rho}}(\tilde{T}^n(\tilde{x}), R)$. Using (6.15), it follows that

$$\tilde{T}^n(B_{\tilde{\rho}}(\tilde{x}, r)) \supseteq B_{\tilde{\rho}}(\tilde{T}^n(\tilde{x}), \lambda^n r) \supseteq B_{\tilde{\rho}}(\tilde{T}^n(\tilde{x}), 2R) \supseteq B_{\tilde{\rho}}(\overline{G}(\tilde{x}), R).$$

Since $\overline{G} : \widetilde{M} \to \widetilde{M}$ is a $\tilde{\rho}$-isometry, we deduce that

$$\tilde{T}^n(B_{\tilde{\rho}}(\tilde{x}, r)) \supseteq \overline{G}(B_{\tilde{\rho}}(\tilde{x}, R)) \supseteq \overline{G}(B_{\tilde{\rho}}(\tilde{x}, r)).$$

Consequently,

$$\tilde{T}^{-n} \circ \overline{G}(\overline{B}_{\tilde{\rho}}(\tilde{x}, r)) \subseteq \overline{B}_{\tilde{\rho}}(\tilde{x}, r).$$

As $\widetilde{T}^{-n}$ is a contraction and $\overline{G}$ is an isometry, the map $\widetilde{T}^{-n} \circ \overline{G} : \overline{B}_{\widetilde{\rho}}(\widetilde{x}, r) \to \overline{B}_{\widetilde{\rho}}(\widetilde{x}, r)$ is a contraction and the Banach contraction principle asserts that $\widetilde{T}^{-n} \circ \overline{G}$ has a fixed point $\widetilde{w} \in \overline{B}_{\widetilde{\rho}}(\widetilde{x}, r)$. Therefore, $\widetilde{T}^n(\widetilde{w}) = \overline{G}(\widetilde{w})$, and hence

$$T^n(\pi(\widetilde{w})) = \pi(\widetilde{T}^n(\widetilde{w})) = \pi(\overline{G}(\widetilde{w})) = \pi(\widetilde{w}).$$

Furthermore,

$$\pi(\widetilde{w}) \in \pi(\overline{B}_{\widetilde{\rho}}(\widetilde{x}, r)) \subseteq U.$$

Thus, T^n has a fixed point in U. Since U is an arbitrary open set in M, we conclude that the set of periodic points of T is dense in M. □

6.4.3 Topological conjugacy and structural stability

In order to establish a topological conjugacy between any two Shub expanding endomorphisms that are homotopic, we shall first show that the existence of a semiconjugacy between the induced homomorphisms of lifts of two Shub expanding maps implies the existence of a semiconjugacy between the lifts themselves.

Lemma 6.4.3. *Let N and M be compact connected smooth manifolds. Let $S : N \to N$ and $T : M \to M$ be Shub expanding endomorphisms. Finally, let $\alpha : N \to M$ be a continuous map. If there exist lifts $\widetilde{S} : \widetilde{N} \to \widetilde{N}$ of S, $\widetilde{T} : \widetilde{M} \to \widetilde{M}$ of T, and $\widetilde{\alpha} : \widetilde{N} \to \widetilde{M}$ of α, such that*

$$\widetilde{T}^* \circ \widetilde{\alpha}^* = \widetilde{\alpha}^* \circ \widetilde{S}^*, \tag{6.21}$$

then there exists a unique map $\widetilde{H} \in V_{\widetilde{\alpha}}$ such that

$$\widetilde{T} \circ \widetilde{H} = \widetilde{H} \circ \widetilde{S},$$

where $V_{\widetilde{\alpha}}$ is the set of all continuous maps $\widetilde{A} : \widetilde{N} \to \widetilde{M}$ that are lifts of some continuous map from N to M and such that $\widetilde{A}^* = \widetilde{\alpha}^*$.

Proof. For every $\widetilde{A} \in V_{\widetilde{\alpha}}$, define

$$\theta(\widetilde{A}) := \widetilde{T}^{-1} \circ \widetilde{A} \circ \widetilde{S}.$$

Claim 1. *The transformation $\theta(\widetilde{A})$ is a lift of some continuous map from N to M.*

Proof of Claim 1. According to Proposition 6.3.8, for every $\overline{G} \in D_M$ we have $\widetilde{T}^*(\overline{G}) \circ \widetilde{T} = \widetilde{T} \circ \overline{G}$. Since $\widetilde{T}$ is a homeomorphism by Proposition 6.3.7 and Theorem 6.2.4, we obtain that $\overline{G} = \widetilde{T}^{-1} \circ \widetilde{T}^*(\overline{G}) \circ \widetilde{T}$. So, if $\overline{G}_1, \overline{G}_2 \in D_M$ and $\widetilde{T}^*(\overline{G}_1) = \widetilde{T}^*(\overline{G}_2)$, then $\overline{G}_1 = \overline{G}_2$. This means that the homomorphism $\widetilde{T}^* : D_M \to D_M$ is injective. Since $\widetilde{\alpha}^* \circ \widetilde{S}^* = \widetilde{T}^* \circ \widetilde{\alpha}^*$, the range of the map $\widetilde{\alpha}^* \circ \widetilde{S}^*$ is contained in the range of $\widetilde{T}^*$ and, therefore, the map $\beta := (\widetilde{T}^*)^{-1} \circ \widetilde{\alpha}^* \circ \widetilde{S}^* : D_N \to D_M$ is well-defined. We will show that

$$\theta(\widetilde{A}) \circ \overline{G} = \beta(\overline{G}) \circ \theta(\widetilde{A}), \quad \forall \widetilde{A} \in V_{\widetilde{\alpha}}, \forall \overline{G} \in D_N.$$

Indeed, for all $\widetilde{A} \in V_{\widetilde{\alpha}}$, all $\overline{G}_1 \in D_N$ and all $\overline{G}_2 \in D_M$, we have

$$(\widetilde{A} \circ \widetilde{S}) \circ \overline{G}_1 = (\widetilde{A}^* \circ \widetilde{S}^*)(\overline{G}_1) \circ \widetilde{A} \circ \widetilde{S} \quad \text{and} \quad \widetilde{T}^{-1} \circ \widetilde{T}^*(\overline{G}_2) \circ \widetilde{T} = \overline{G}_2.$$

Recalling that $\widetilde{A}^* = \widetilde{\alpha}^*$ by hypothesis, it follows that for all $\overline{G} \in D_N$,

$$\begin{aligned}
\theta(\widetilde{A}) \circ \overline{G} &= \widetilde{T}^{-1} \circ \widetilde{A} \circ \widetilde{S} \circ \overline{G} = \widetilde{T}^{-1} \circ (\widetilde{A}^* \circ \widetilde{S}^*)(\overline{G}) \circ \widetilde{A} \circ \widetilde{S} \\
&= \widetilde{T}^{-1} \circ \widetilde{T}^* \circ (\widetilde{T}^*)^{-1} \circ \widetilde{A}^* \circ \widetilde{S}^*(\overline{G}) \circ \widetilde{A} \circ \widetilde{S} \\
&= \widetilde{T}^{-1} \circ \widetilde{T}^*(((\widetilde{T}^*)^{-1} \circ \widetilde{A}^* \circ \widetilde{S}^*)(\overline{G})) \circ \widetilde{T} \circ \widetilde{T}^{-1} \circ \widetilde{A} \circ \widetilde{S} \\
&= ((\widetilde{T}^*)^{-1} \circ \widetilde{A}^* \circ \widetilde{S}^*)(\overline{G}) \circ \widetilde{T}^{-1} \circ \widetilde{A} \circ \widetilde{S} \\
&= \beta(\overline{G}) \circ \theta(\widetilde{A}).
\end{aligned}$$

As $\beta(\overline{G}) \in D_M$, by virtue of Proposition 6.3.8, the above equality implies that $\theta(\widetilde{A})$ is a lift of some continuous map from N to M. Thus, the proof of Claim 1 is complete. □

Claim 2. $\theta(V_{\widetilde{\alpha}}) \subseteq V_{\widetilde{\alpha}}$.

Proof of Claim 2. We aim to show that if $\widetilde{A} \in V_{\widetilde{\alpha}}$, then $(\theta(\widetilde{A}))^*(\overline{G}) = \widetilde{\alpha}^*(\overline{G})$ for all $\overline{G} \in D_N$. Recall from Proposition 6.3.8 that the map $^* : D_N \to D_M$ is a homomorphism. From this fact and from (6.21), we obtain that

$$\begin{aligned}
(\theta(\widetilde{A}))^*(\overline{G}) &= (\widetilde{T}^{-1} \circ \widetilde{A} \circ \widetilde{S})^*(\overline{G}) \\
&= (\widetilde{T}^{-1})^* \circ \widetilde{A}^* \circ \widetilde{S}^*(\overline{G}) = (\widetilde{T}^{-1})^* \circ \widetilde{\alpha}^* \circ \widetilde{S}^*(\overline{G}) \\
&= (\widetilde{T}^{-1})^* \circ \widetilde{T}^* \circ \widetilde{\alpha}^*(\overline{G}) = (\widetilde{T}^{-1} \circ \widetilde{T})^* \circ \widetilde{\alpha}^*(\overline{G}) = \widetilde{\alpha}^*(\overline{G}).
\end{aligned}$$

This establishes Claim 2. □

Claim 3. *The map $\theta : V_{\widetilde{\alpha}} \to V_{\widetilde{\alpha}}$ is a contraction with respect to the metric $\widetilde{\rho}_\infty$ on $V_{\widetilde{\alpha}}$.*

Proof of Claim 3. Let $\widetilde{A}, \widetilde{B} \in V_{\widetilde{\alpha}}$. Using (6.15), we get

$$\begin{aligned}
\widetilde{\rho}_\infty(\theta(\widetilde{A}), \theta(\widetilde{B})) &= \widetilde{\rho}_\infty(\widetilde{T}^{-1} \circ \widetilde{A} \circ \widetilde{S}, \widetilde{T}^{-1} \circ \widetilde{B} \circ \widetilde{S}) \\
&= \sup\{\widetilde{\rho}_M(\widetilde{T}^{-1} \circ \widetilde{A} \circ \widetilde{S}(\widetilde{x}), \widetilde{T}^{-1} \circ \widetilde{B} \circ \widetilde{S}(\widetilde{x})) : \widetilde{x} \in \widetilde{N}\} \\
&\le \lambda^{-1} \sup\{\widetilde{\rho}_M(\widetilde{A} \circ \widetilde{S}(\widetilde{x}), \widetilde{B} \circ \widetilde{S}(\widetilde{x})) : \widetilde{x} \in \widetilde{N}\} \\
&= \lambda^{-1} \sup\{\widetilde{\rho}_M(\widetilde{A}(\widetilde{y}), \widetilde{B}(\widetilde{y})) : \widetilde{y} \in \widetilde{N}\} \\
&= \lambda^{-1} \widetilde{\rho}_\infty(\widetilde{A}, \widetilde{B}).
\end{aligned}$$

This substantiates Claim 3.

In light of Claim 3 and Lemma 6.3.9, Banach's contraction principle affirms that the map $\theta : V_{\widetilde{\alpha}} \to V_{\widetilde{\alpha}}$ has a unique fixed point $\widetilde{H} \in V_{\widetilde{\alpha}}$. The equality $\theta(\widetilde{H}) = \widetilde{H}$ is equivalent to the equality $\widetilde{T} \circ \widetilde{H} = \widetilde{H} \circ \widetilde{S}$. □

We now demonstrate that homotopic Shub expanding endomorphisms exhibit conjugate dynamics. This generalizes Theorem 6.1.3.

Theorem 6.4.4. *Let M be a compact connected smooth manifold. If $T, S : M \to M$ are two homotopic Shub expanding endomorphisms, then T and S are topologically conjugate.*

Proof. Let $(F_t)_{0\le t\le 1}$ be a homotopy from T to S in M. Thus, $F_0 = T$ while $F_1 = S$. Let $(\widetilde{F}_t)_{0\le t\le 1}$ be a lift of $(F_t)_{0\le t\le 1}$ to $\widetilde{M}$. In particular, $\widetilde{F}_0$ is a lift of T and $\widetilde{F}_1$ is a lift of S. In light of Proposition 6.3.8, we have for every $t \in [0,1]$ that

$$\widetilde{F}_t \circ \overline{G} = \widetilde{F}_t^*(\overline{G}) \circ \widetilde{F}_t, \quad \forall \overline{G} \in D_M. \tag{6.22}$$

Claim. *The function $[0,1] \ni t \mapsto \widetilde{F}_t^*(\overline{G}) \in D_M$ is constant for every $\overline{G} \in D_M$.*

Proof of the claim. Fix $\overline{G} \in D_M$. Let $s \in [0,1]$. Choose any sequence $(s_n)_{n=1}^\infty$ in $[0,1]$ converging to s. Fix $\widetilde{x} \in \widetilde{M}$. Let $\widetilde{z} := \widetilde{F}_s(\widetilde{x})$ and $\widetilde{z}_n := \widetilde{F}_{s_n}(\widetilde{x})$ for all $n \in \mathbb{N}$. Then

$$\lim_{n\to\infty} \widetilde{z}_n = \lim_{n\to\infty} \widetilde{F}_{s_n}(\widetilde{x}) = \widetilde{F}_s(\widetilde{x}) = \widetilde{z}$$

and, by (6.22),

$$\lim_{n\to\infty} (\widetilde{F}_{s_n}^*(\overline{G}))(\widetilde{z}_n) = \lim_{n\to\infty} \widetilde{F}_{s_n}^*(\overline{G}) \circ \widetilde{F}_{s_n}(\widetilde{x}) = \lim_{n\to\infty} \widetilde{F}_{s_n} \circ \overline{G}(\widetilde{x}) = \widetilde{F}_s(\overline{G}(\widetilde{x})).$$

Therefore, Lemma 6.3.5 asserts that the sequence $(\widetilde{F}_{s_n}^*(\overline{G}))_{n=1}^\infty$ eventually coincides with the unique deck transformation $\overline{\Gamma} \in D_M$ determined by the condition $\overline{\Gamma}(\widetilde{z}) = \widetilde{F}_s(\overline{G}(\widetilde{x}))$. But $\widetilde{z} = \widetilde{F}_s(\widetilde{x})$, so

$$\overline{\Gamma}(\widetilde{F}_s(\widetilde{x})) = \widetilde{F}_s(\overline{G}(\widetilde{x})) = \widetilde{F}_s^*(\overline{G})(\widetilde{F}_s(\widetilde{x})).$$

Hence, $\overline{\Gamma} = \widetilde{F}_s^*(\overline{G})$. In summary, the sequence $(\widetilde{F}_{s_n}^*(\overline{G}))_{n=1}^\infty$ is eventually equal to $\widetilde{F}_s^*(\overline{G})$. Since this is true for any convergent sequence $(s_n)_{n=1}^\infty$ in $[0,1]$, we conclude that the function $[0,1] \ni t \mapsto \widetilde{F}_t^*(\overline{G}) \in D_M$ is constant. This confirms the claim.

Setting $\widetilde{F}_0 = \widetilde{T}$ and $\widetilde{F}_1 = \widetilde{S}$ and letting $(\widetilde{F}_t)_{0\le t\le 1}$ be a lift of $(F_t)_{0\le t\le 1}$ to $\widetilde{M}$, it follows from the claim that

$$\widetilde{T}^* = \widetilde{S}^*.$$

So we may apply Lemma 6.4.3 with $M = N$, with $\alpha = \mathrm{Id}_M$ and with $\widetilde{\alpha} = \mathrm{Id}_{\widetilde{M}}$, to obtain a unique element $\widetilde{A} \in V_{\mathrm{Id}_{\widetilde{M}}}$ such that

$$\widetilde{T} \circ \widetilde{A} = \widetilde{A} \circ \widetilde{S}. \tag{6.23}$$

By the symmetry between $\widetilde{T}$ and $\widetilde{S}$, there is also an element $\widetilde{B} \in V_{\mathrm{Id}_{\widetilde{M}}}$ such that

$$\widetilde{S} \circ \widetilde{B} = \widetilde{B} \circ \widetilde{T}.$$

Hence,

$$\widetilde{S} \circ (\widetilde{B} \circ \widetilde{A}) = (\widetilde{S} \circ \widetilde{B}) \circ \widetilde{A} = (\widetilde{B} \circ \widetilde{T}) \circ \widetilde{A} = \widetilde{B} \circ (\widetilde{T} \circ \widetilde{A}) = \widetilde{B} \circ (\widetilde{A} \circ \widetilde{S}) = (\widetilde{B} \circ \widetilde{A}) \circ \widetilde{S}.$$

Moreover,

$$\widetilde{S} \circ \mathrm{Id}_{\widetilde{M}} = \mathrm{Id}_{\widetilde{M}} \circ \widetilde{S}.$$

Therefore, the uniqueness part of Lemma 6.4.3, applied with $M = N$, $T = S$, $\alpha = \mathrm{Id}_M$ and $\widetilde{\alpha} = \mathrm{Id}_{\widetilde{M}}$, yields $\widetilde{B} \circ \widetilde{A} = \mathrm{Id}_{\widetilde{M}}$. Likewise, symmetrically, $\widetilde{A} \circ \widetilde{B} = \mathrm{Id}_{\widetilde{M}}$. Let $x \in M$ and choose an arbitrary $\widetilde{x} \in \pi^{-1}(x)$. Given that $\widetilde{A}$ and $\widetilde{B}$, as elements of $V_{\mathrm{Id}_{\widetilde{M}}}$, are lifts of some continuous maps $A : M \to M$ and $B : M \to M$, respectively, we then have that

$$A \circ B(x) = A \circ B \circ \pi(\widetilde{x}) = A \circ \pi \circ \widetilde{B}(\widetilde{x}) = \pi \circ \widetilde{A} \circ \widetilde{B}(\widetilde{x}) = \pi(\widetilde{x}) = x.$$

So, $A \circ B = \mathrm{Id}_M$ and, likewise, $B \circ A = \mathrm{Id}_M$. Thus, A and B are homeomorphisms. Furthermore, due to (6.23), we have that

$$\begin{aligned} T \circ A(x) &= T \circ A \circ \pi(\widetilde{x}) = T \circ \pi \circ \widetilde{A}(\widetilde{x}) = \pi \circ \widetilde{T} \circ \widetilde{A}(\widetilde{x}) = \pi \circ \widetilde{A} \circ \widetilde{S}(\widetilde{x}) \\ &= A \circ \pi \circ \widetilde{S}(\widetilde{x}) = A \circ S \circ \pi(\widetilde{x}) = A \circ S(x). \end{aligned}$$

This means that $T \circ A = A \circ S$ for some homeomorphism $A : M \to M$, that is, T and S are topologically conjugate. □

The crowning statement of this chapter pertains to the structural stability of Shub expanding endomorphisms. Recall that structural stability was defined in Section 1.2.

Theorem 6.4.5. *Every Shub expanding endomorphism of a compact connected smooth manifold M is structurally stable in $\mathcal{E}^1(M)$, the space of all C^1 endomorphisms of M.*

Proof. This is an immediate consequence of Theorems 6.2.5, 6.2.6, and 6.4.4. □

In Chapter 13, we will develop the theory of Gibbs states for open distance expanding systems. In conjunction with the theory of Shub expanding endomorphisms described here, we will derive in Section 13.7 the following theorem, which was first proved for C^2 maps by Krzyżewski and Szlenk [42]. It is also in this paper that the appropriate transfer (also called Ruelle or Perron–Frobenius) operator was for the first time explicitly used in dynamical systems. Our proof will be different, based on the theory of Gibbs states developed in Chapter 13; nevertheless, there will be significant similarities with that of Krzyżewski and Szlenk.

Each Riemannian metric ρ on a compact connected smooth manifold M induces a unique volume (Lebesgue) measure λ_ρ on M and the volume measures induced by various Riemannian metrics are mutually equivalent. Call this class of measures the Lebesgue measure class on M.

Theorem 6.4.6. *If $T : M \to M$ is a $C^{1+\varepsilon}$ Shub expanding endomorphism on a compact connected smooth manifold M, then there exists a unique T-invariant Borel probability measure μ on M which is absolutely continuous with respect to the Lebesgue measure class on M. In fact, μ is equivalent to the Lebesgue measure class on M, and μ is ergodic.*

6.5 Exercises

Exercise 6.5.1. Let $T : \mathbb{S}^1 \to \mathbb{S}^1$ be a Shub expanding map. Prove that $|\deg(T)| \geq 2$.

Exercise 6.5.2. Let M be a compact connected smooth manifold and $f, g : M \to M$ be Shub expanding endomorphisms with respect to some Riemannian metric ρ on M. Show that $f \circ g$ is also a Shub expanding endomorphism with respect to ρ.

Exercise 6.5.3. Suppose that $f_1, f_2 : M \to M$ are Shub expanding endomorphisms with respect to Riemannian metrics ρ_1 and ρ_2, respectively. Is there always a Riemannian metric ρ such that $f_1 \circ f_2$ is expanding with respect to ρ? You may assume that $M = \mathbb{S}^1$.

Exercise 6.5.4. Prove that the Cartesian product of finitely many Shub expanding endomorphisms is a Shub expanding endomorphism if the product manifold is endowed with the standard L^1 product metric

$$\langle v, w\rangle_x := \sum_{k=1}^{n} \langle v_k, w_k\rangle_{x_k}.$$

Exercise 6.5.5. Suppose that V is an open neighborhood of the origin in a Euclidean space $\mathbb{R}^k$ and that $G : V \to V$ is a diffeomorphism. Let $R > 0$ be such that $\overline{B}(0, R) \subseteq V$. Show that the map $\widehat{G} : \mathbb{R}^k \to \mathbb{R}^k$, where

$$\widehat{G}(x) = \begin{cases} G(x) & \text{if } x \in \overline{B}(0, R) \\ G(\frac{Rx}{\|x\|}) + [G'(\frac{Rx}{\|x\|})](x - \frac{Rx}{\|x\|}) & \text{if } x \notin \overline{B}(0, R) \end{cases}$$

is a diffeomorphic extension of $G|_{\overline{B}(0,R)}$.

7 Topological entropy

In this chapter, we study the notion of topological entropy, one of the most useful and widely-applicable topological invariant thus far discovered. It was introduced to dynamical systems by Adler, Konheim, and McAndrew [2] in 1965. Their definition was motivated by Kolmogorov and Sinai's definition of metric/measure-theoretic entropy introduced in [67] less than a decade earlier. In this book, we do not follow the historical order of discovery of these notions. It is more suitable to present topological entropy first.

Metric and topological entropies not only have related origins and similar names. There are truly significant mathematical relations between them, particularly the one given by the variational principle, which is treated at length in Chapter 12.

The topological entropy of a dynamical system $T : X \to X$, which we introduce in Section 7.2 and shall be denoted by $\mathrm{h}_{\mathrm{top}}(T)$, is a nonnegative extended real number that measures the complexity of the system. Somewhat more precisely, $\mathrm{h}_{\mathrm{top}}(T)$ is the exponential growth rate of the number of orbits separated under T. The topological entropy of a dynamical system is defined in three stages. First, we define the entropy of a cover of the underlying space. Second, we define the entropy of the system with respect to any given cover. Third, the entropy of the system is defined to be the supremum, over all covers, of the entropy of the system with respect to each of those.

Recall from Chapter 1 that a mathematical property is said to be a topological invariant for the category of topological dynamical systems if it is shared by any pair of topologically conjugate systems. For topological entropy, being an invariant means that if $T : X \to X$ and $S : Y \to Y$ are two topologically conjugate dynamical systems, then $\mathrm{h}_{\mathrm{top}}(T) = \mathrm{h}_{\mathrm{top}}(S)$. However, the converse is generally not true. That is, if $T : X \to X$ and $S : Y \to Y$ are two dynamical systems with equal topological entropy, then T and S may not be topologically conjugate. You are asked to provide such an example in Exercise 7.6.15. Therefore, topological entropy is not a complete invariant.

In Section 7.3, we treat at length Bowen's characterization of topological entropy in terms of separated and spanning sets.

In Chapter 11, we will introduce and deal with topological pressure, which is a substantial generalization of topological entropy. Our approach to topological pressure will stem from and extend that for topological entropy. In this sense, the present chapter can be viewed as a preparation to Chapter 11.

7.1 Covers of a set

Definition 7.1.1. Let X be a nonempty set. A family $\mathcal{U}$ of subsets of X is said to form a *cover* of X if

$$X \subseteq \bigcup_{U\in\mathcal{U}} U.$$

Furthermore, $\mathcal{V}$ is said to be a *subcover* of $\mathcal{U}$ if $\mathcal{V}$ is itself a cover and $\mathcal{V} \subseteq \mathcal{U}$.

https://doi.org/10.1515/9783110702682-007

We will always denote covers by calligraphic letters, $\mathcal{U}, \mathcal{V}, \mathcal{W}$, and so on.

Let us begin by introducing a useful way of obtaining a new cover from two existing covers.

Definition 7.1.2. If $\mathcal{U}$ and $\mathcal{V}$ are covers of X, then their *join*, denoted $\mathcal{U} \vee \mathcal{V}$, is the cover

$$\mathcal{U} \vee \mathcal{V} := \{U \cap V : U \in \mathcal{U}, V \in \mathcal{V}\}.$$

Remark 7.1.3. The join operation is commutative (i. e., $\mathcal{U} \vee \mathcal{V} = \mathcal{V} \vee \mathcal{U}$) and associative (in other words, $(\mathcal{U} \vee \mathcal{V}) \vee \mathcal{W} = \mathcal{U} \vee (\mathcal{V} \vee \mathcal{W})$). Thanks to this associativity, the join operation extends naturally to any finite collection $\{\mathcal{U}_j\}_{j=0}^{n-1}$ of covers of X:

$$\bigvee_{j=0}^{n-1} \mathcal{U}_j := \mathcal{U}_0 \vee \cdots \vee \mathcal{U}_{n-1} = \left\{ \bigcap_{j=0}^{n-1} U_j : U_j \in \mathcal{U}_j, \forall 0 \le j \le n-1 \right\}.$$

It is also useful to be able to compare covers. For this purpose, we introduce the following relation on the collection of all covers of a set.

Definition 7.1.4. Let $\mathcal{U}$ and $\mathcal{V}$ be covers of a set X. We say that $\mathcal{V}$ is *finer* than, or a *refinement* of, $\mathcal{U}$, and denote this by $\mathcal{U} \prec \mathcal{V}$, if every element of $\mathcal{V}$ is a subset of an element of $\mathcal{U}$. That is, for every set $V \in \mathcal{V}$ there exists a set $U \in \mathcal{U}$ such that $V \subseteq U$. It is also sometimes said that $\mathcal{V}$ is *inscribed* in $\mathcal{U}$, or that $\mathcal{U}$ is coarser than $\mathcal{V}$.

Lemma 7.1.5. *Let $\mathcal{U}$, $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{X}$ be covers of a set X. Then:*

(a) *The refinement relation $\prec$ is reflexive (i. e., $\mathcal{U} \prec \mathcal{U}$) and transitive (i. e., if $\mathcal{U} \prec \mathcal{V}$ and $\mathcal{V} \prec \mathcal{W}$, then $\mathcal{U} \prec \mathcal{W}$).*

(b) $\mathcal{U} \prec \mathcal{U} \vee \mathcal{V}$.

(c) *If $\mathcal{V}$ is a subcover of $\mathcal{U}$, then $\mathcal{U} \prec \mathcal{V}$.*

(d) *$\mathcal{U}$ is a subcover of $\mathcal{U} \vee \mathcal{U}$. Hence, from (c) and (b), we deduce that*

$$\mathcal{U} \prec \mathcal{U} \vee \mathcal{U} \prec \mathcal{U}.$$

Nevertheless, $\mathcal{U}$ is not equal to $\mathcal{U} \vee \mathcal{U}$ in general.

(e) *If $\mathcal{U} \prec \mathcal{V}$ <u>or</u> $\mathcal{U} \prec \mathcal{W}$, then $\mathcal{U} \prec \mathcal{V} \vee \mathcal{W}$.*

(f) *If $\mathcal{U} \prec \mathcal{W}$ <u>and</u> $\mathcal{V} \prec \mathcal{W}$, then $\mathcal{U} \vee \mathcal{V} \prec \mathcal{W}$.*

(g) *If $\mathcal{U} \prec \mathcal{W}$ <u>and</u> $\mathcal{V} \prec \mathcal{X}$, then $\mathcal{U} \vee \mathcal{V} \prec \mathcal{W} \vee \mathcal{X}$.*

Proof. All of these properties can be proved directly and are left to the reader. As a hint, observe that property (e) is a consequence of (b) and the transitivity part of (a), while property (g) follows upon combining (e) and (f). □

Remark 7.1.6. Although it is reflexive and transitive, the relation $\prec$ is not antisymmetric (i. e., $\mathcal{U} \prec \mathcal{V} \prec \mathcal{U}$ does not necessarily imply $\mathcal{U} = \mathcal{V}$; see Lemma 7.1.5(d)). Therefore, $\prec$ does not generate a partial order on the collection of all covers of a set X.

If X is a metric space, then it makes sense to talk about the diameter of a cover in terms of the diameter of its elements. This is the purpose of the next definition.

Definition 7.1.7. If (X, d) is a metric space, then the *diameter* of a cover $\mathcal{U}$ of X is defined by

$$\operatorname{diam}(\mathcal{U}) := \sup\{\operatorname{diam}(U) : U \in \mathcal{U}\},$$

where

$$\operatorname{diam}(U) := \sup\{d(x, y) : x, y \in U\}.$$

It is also often of interest to know that all sets of some specified diameter are each contained in at least one element of a given cover. This is made precise in the following definition.

Definition 7.1.8. A number $\delta > 0$ is said to be a *Lebesgue number* for a cover $\mathcal{U}$ of a metric space (X, d) if every set of diameter at most δ is contained in an element of $\mathcal{U}$.

It is clear that if δ_0 is a Lebesgue number for a cover $\mathcal{U}$, then so is any δ with $0 < \delta < \delta_0$. One can easily prove by contradiction that every open cover of a compact metric space admits such a number. By an open cover, we simply mean a cover whose elements are all open subsets of the space.

7.1.1 Dynamical covers

In this subsection, we add a dynamical aspect to the above discussion. Let X be a nonempty set and let $T : X \to X$ be a map. We will define covers that are induced by the dynamics of the map T. First, let us define the preimage of a cover under a map.

Definition 7.1.9. Let X and Y be nonempty sets. Let $h : X \to Y$ be a map and $\mathcal{V}$ be a cover of Y. The *preimage* of $\mathcal{V}$ under the map h is the cover of X consisting of all the preimages of the elements of $\mathcal{V}$ under h, that is,

$$h^{-1}(\mathcal{V}) := \{h^{-1}(V) : V \in \mathcal{V}\}.$$

We now show that, as far as set operations go, the operator h^{-1} behaves well with respect to cover operations.

Lemma 7.1.10. *Let $h : X \to Y$ be a map, and $\mathcal{U}$ and $\mathcal{V}$ be covers of Y. The following assertions hold:*

(a) *The map h^{-1} preserves the refinement relation, that is,*

$$\mathcal{U} \prec \mathcal{V} \implies h^{-1}(\mathcal{U}) \prec h^{-1}(\mathcal{V}).$$

Moreover, if $\mathcal{V}$ is a subcover of $\mathcal{U}$ then $h^{-1}(\mathcal{V})$ is a subcover of $h^{-1}(\mathcal{U})$.

(b) *The map h^{-1} respects the join operation, that is,*

$$h^{-1}(\mathcal{U} \vee \mathcal{V}) = h^{-1}(\mathcal{U}) \vee h^{-1}(\mathcal{V}).$$

Note that if $Y = X$, then by induction h^{-n} enjoys these properties for any $n \in \mathbb{N}$.

Proof. These assertions are straightforward to prove and are thus left to the reader. □

We now introduce covers that follow the orbits of a given map by indicating to which elements of a given cover the successive iterates of the map belong.

Definition 7.1.11. Let $T : X \to X$ be a map and $\mathcal{U}$ be a cover of X. For every $n \in \mathbb{N}$ and $0 \leq m < n$, define the *dynamical cover*

$$\mathcal{U}_m^n := \bigvee_{j=m}^{n-1} T^{-j}(\mathcal{U}) = T^{-m}(\mathcal{U}) \vee T^{-(m+1)}(\mathcal{U}) \vee \cdots \vee T^{-(n-1)}(\mathcal{U}).$$

To lighten notation, we will write $\mathcal{U}^n$ in lieu of $\mathcal{U}_0^n$.

A typical element of $\mathcal{U}^n$ is of the form $U_0 \cap T^{-1}(U_1) \cap T^{-2}(U_2) \cap \ldots \cap T^{-(n-1)}(U_{n-1})$ for some $U_0, U_1, U_2, \ldots, U_{n-1} \in \mathcal{U}$. This element is the set of all points of X whose iterates under T fall successively into the elements U_0, U_1, U_2, and so on, up to U_{n-1}.

Lemma 7.1.12. *Let $\mathcal{U}$ and $\mathcal{V}$ be covers of a set X. Let $T : X \to X$ be a map. For every $k, m, n \in \mathbb{N}$, the following statements hold:*
(a) *If $\mathcal{U} \prec \mathcal{V}$, then $\mathcal{U}^n \prec \mathcal{V}^n$.*
(b) $(\mathcal{U} \vee \mathcal{V})^n = \mathcal{U}^n \vee \mathcal{V}^n$.
(c) $\mathcal{U}^n \prec \mathcal{U}^{n+1}$.
(d) $(\mathcal{U}^k)^n \prec \mathcal{U}^{n+k-1} \prec (\mathcal{U}^k)^n$.
(e) $\mathcal{U}_m^n = T^{-k}(\mathcal{U}_{m-k}^{n-k})$ *for all $k \leq m < n$. In particular, $\mathcal{U}_m^n = T^{-m}(\mathcal{U}^{n-m})$.*

Proof. Property (a) follows directly from Lemmas 7.1.10(a) and 7.1.5(g). Property (b) is a consequence of Lemma 7.1.10(b) and the associativity of the join operation. As $\mathcal{U}^{n+1} = \mathcal{U}^n \vee T^{-n}(\mathcal{U})$, property (c) follows from an application of Lemma 7.1.5(b). Property (d) is a little more intricate to prove. Using Lemma 7.1.10(b) and Remark 7.1.3, we obtain that

$$\begin{aligned}
(\mathcal{U}^k)^n &= \mathcal{U}^k \vee T^{-1}(\mathcal{U}^k) \vee \cdots \vee T^{-(n-1)}(\mathcal{U}^k) \\
&= (\mathcal{U} \vee \cdots \vee T^{-(k-1)}(\mathcal{U})) \vee T^{-1}(\mathcal{U} \vee \cdots \vee T^{-(k-1)}(\mathcal{U})) \vee \cdots \\
&\quad \cdots \vee T^{-(n-1)}(\mathcal{U} \vee \cdots \vee T^{-(k-1)}(\mathcal{U})) \\
&= \mathcal{U} \vee (T^{-1}(\mathcal{U}) \vee T^{-1}(\mathcal{U})) \vee (T^{-2}(\mathcal{U}) \vee T^{-2}(\mathcal{U}) \vee T^{-2}(\mathcal{U})) \vee \cdots \\
&\quad \cdots \vee (T^{-(n+k-3)}(\mathcal{U}) \vee T^{-(n+k-3)}(\mathcal{U})) \vee T^{-(n+k-2)}(\mathcal{U}).
\end{aligned}$$

Now, according to Lemma 7.1.5(d), $T^{-j}(\mathcal{U}) \prec T^{-j}(\mathcal{U}) \vee T^{-j}(\mathcal{U}) \prec T^{-j}(\mathcal{U})$ for all $j \in \mathbb{N}$. We deduce from a repeated application of Lemma 7.1.5(g) that

$$T^{-j}(\mathcal{U}) \prec \bigvee_{l=1}^{m} T^{-j}(\mathcal{U}) \prec T^{-j}(\mathcal{U}), \quad \forall m \in \mathbb{N}.$$

Another round of repeated applications of Lemma 7.1.5(g) allows us to conclude that

$$(\mathcal{U}^k)^n \prec \mathcal{U} \vee T^{-1}(\mathcal{U}) \vee T^{-2}(\mathcal{U}) \vee \cdots \vee T^{-(n+k-3)}(\mathcal{U}) \vee T^{-(n+k-2)}(\mathcal{U}) \prec (\mathcal{U}^k)^n.$$

That is,

$$(\mathcal{U}^k)^n \prec \mathcal{U}^{n+k-1} \prec (\mathcal{U}^k)^n.$$

Finally, property (e) follows from Lemma 7.1.10(b) with $h = T^k$ since

$$\mathcal{U}_m^n = \bigvee_{j=m}^{n-1} T^{-j}(\mathcal{U}) = T^{-k}\left(\bigvee_{j=m-k}^{n-k-1} T^{-j}(\mathcal{U})\right) = T^{-k}(\mathcal{U}_{m-k}^{n-k}).$$ □

7.2 Definition of topological entropy via open covers

The definition of topological entropy via open covers only requires the underlying space to be a topological space. It need not be a metrizable space. The topological entropy of a dynamical system $T : X \to X$ is defined in three stages, which, for clarity of exposition, we split into the following three subsections.

7.2.1 First stage: entropy of an open cover

At this stage, the dynamics of the system T are not in consideration. We simply look at the difficulty of covering the underlying compact space X with open covers.

Definition 7.2.1. Let $\mathcal{U}$ be an open cover of X. Define

$$Z_1(\mathcal{U}) := \min\{\#\mathcal{V} : \mathcal{V} \text{ is a subcover of } \mathcal{U}\}.$$

That is, $Z_1(\mathcal{U})$ denotes the minimum number of elements of $\mathcal{U}$ necessary to cover X. A subcover of $\mathcal{U}$ whose cardinality equals this minimum number is called a *minimal subcover* of $\mathcal{U}$.

Every open cover admits at least one minimal subcover and any such subcover is finite since X is compact. Thus $1 \le Z_1(\mathcal{U}) < \infty$ for all open covers $\mathcal{U}$ of X.

We now observe that the function $Z_1(\cdot)$ acts as desired with respect to the refinement relation. In other words, the finer the cover, the larger the minimum number of elements required to cover the space, that is, the more difficult it is to cover the space.

Lemma 7.2.2. *If $\mathcal{U} \prec \mathcal{V}$, then $Z_1(\mathcal{U}) \le Z_1(\mathcal{V})$. In particular, this holds if $\mathcal{V}$ is a subcover of $\mathcal{U}$.*

Proof. Let $\mathcal{U} \prec \mathcal{V}$. For every $V \in \mathcal{V}$, there exists a set $i(V) \in \mathcal{U}$ such that $V \subseteq i(V)$. This defines a function $i : \mathcal{V} \to \mathcal{U}$. Let $\mathcal{W}$ be a minimal subcover of $\mathcal{V}$. Then $i(\mathcal{W}) := \{i(W) : W \in \mathcal{W}\} \subseteq \mathcal{U}$ is a cover of X since

$$X \subseteq \bigcup_{W\in\mathcal{W}} W \subseteq \bigcup_{W\in\mathcal{W}} i(W).$$

Thus $i(\mathcal{W})$ is a subcover of $\mathcal{U}$, and hence

$$Z_1(\mathcal{U}) \le \#i(\mathcal{W}) \le \#\mathcal{W} = Z_1(\mathcal{V}).$$ □

Another fundamental property of the function $Z_1(\cdot)$ is that it is submultiplicative with respect to the join operation. Recall that a sequence $(a_n)_{n=1}^{\infty}$ of real numbers is said to be submultiplicative if

$$a_{m+n} \le a_m a_n, \quad \forall m, n \in \mathbb{N}.$$

Lemma 7.2.3. *Let $\mathcal{U}$ and $\mathcal{V}$ be open covers of X. Then*

$$Z_1(\mathcal{U} \vee \mathcal{V}) \le Z_1(\mathcal{U}) \cdot Z_1(\mathcal{V}).$$

Proof. Let $\underline{\mathcal{U}}$ be a minimal subcover of $\mathcal{U}$ and $\underline{\mathcal{V}}$ be a minimal subcover of $\mathcal{V}$. Then $\underline{\mathcal{U}} \vee \underline{\mathcal{V}}$ is a subcover of $\mathcal{U} \vee \mathcal{V}$. Therefore,

$$Z_1(\mathcal{U} \vee \mathcal{V}) \le \#(\underline{\mathcal{U}} \vee \underline{\mathcal{V}}) \le \#\underline{\mathcal{U}} \cdot \#\underline{\mathcal{V}} = Z_1(\mathcal{U}) \cdot Z_1(\mathcal{V}).$$ □

We can now define the entropy of a cover.

Definition 7.2.4. Let $\mathcal{U}$ be an open cover of X. The *entropy* of $\mathcal{U}$ is defined to be

$$H(\mathcal{U}) := \log Z_1(\mathcal{U}).$$

So, the entropy of an open cover is simply the logarithm of the minimum number of elements of that cover needed to cover the space. The presence of the logarithm function shall be explained shortly. If the entropy of a given cover is to accurately reflect the complexity of that cover, that is, the number of elements necessary for covering the space, then the finer the cover, the larger its entropy should be. In other words, entropy of covers should be increasing with respect to the refinement relation. This, along with other basic properties of the entropy of covers, is shown to hold in the following lemma.

Lemma 7.2.5. *Let $\mathcal{U}$ and $\mathcal{V}$ be open covers of X. Entropy of covers satisfies the following properties:*

(a) $0 \le H(\mathcal{U}) < \infty$.
(b) $H(\mathcal{U}) = 0$ *if and only if* $X \in \mathcal{U}$.
(c) *If* $\mathcal{U} \prec \mathcal{V}$, *then* $H(\mathcal{U}) \le H(\mathcal{V})$.
(d) $H(\mathcal{U} \vee \mathcal{V}) \le H(\mathcal{U}) + H(\mathcal{V})$.

Proof. The first two properties come directly from entropy's definition. The third follows from Lemma 7.2.2 and the fact that the logarithm is an increasing function. The fourth and final property follows from Lemma 7.2.3. □

7.2.2 Second stage: entropy of a system relative to an open cover

In this second stage, we will take into account the dynamics of the topological dynamical system $T : X \to X$. Since T is continuous, every open cover $\mathcal{U}$ of X generates the sequence of dynamical covers $(\mathcal{U}^n)_{n=1}^{\infty}$, all of which are also open.

Definition 7.2.6. Let $\mathcal{U}$ be an open cover of X. For every $n \in \mathbb{N}$, define $Z_n(\mathcal{U})$ to be

$$Z_n(\mathcal{U}) := Z_1(\mathcal{U}^n) = \min\{\#\mathcal{V} : \mathcal{V} \text{ is a subcover of } \mathcal{U}^n\}.$$

Thus $Z_n(\mathcal{U})$ is the minimum number of elements of $\mathcal{U}^n$ needed to cover X. This number describes the complexity of the dynamics of T with respect to $\mathcal{U}$ from time 0 until time $n-1$. Observe also that

$$Z_n(\mathcal{U}) = \exp\big(H(\mathcal{U}^n)\big).$$

For a given open cover, the sequence $(Z_n(\cdot))_{n=1}^{\infty}$ has an interesting property.

Lemma 7.2.7. *For any open cover $\mathcal{U}$ of X, the sequence $(Z_n(\mathcal{U}))_{n=1}^{\infty}$ is nondecreasing.*

Proof. Since $\mathcal{U}^n \prec \mathcal{U}^{n+1}$ for all $n \in \mathbb{N}$ according to Lemma 7.1.12(c), the sequence $(Z_n(\mathcal{U}))_{n=1}^{\infty} = (Z_1(\mathcal{U}^n))_{n=1}^{\infty}$ is nondecreasing by Lemma 7.2.2. □

As the next lemma shows, like the function $Z_1(\cdot)$, the functions $Z_n(\cdot)$ respect the refinement relation.

Lemma 7.2.8. *If $\mathcal{U} \prec \mathcal{V}$, then $Z_n(\mathcal{U}) \le Z_n(\mathcal{V})$, and thus $H(\mathcal{U}^n) \le H(\mathcal{V}^n)$ for every $n \in \mathbb{N}$. In particular, these inequalities hold if $\mathcal{V}$ is a subcover of $\mathcal{U}$.*

Proof. If $\mathcal{U} \prec \mathcal{V}$, then Lemma 7.1.12(a) states that $\mathcal{U}^n \prec \mathcal{V}^n$ for every $n \in \mathbb{N}$. It follows from Lemma 7.2.2 that

$$Z_n(\mathcal{U}) = Z_1\big(\mathcal{U}^n\big) \le Z_1\big(\mathcal{V}^n\big) = Z_n(\mathcal{V}).$$

Since the logarithm is an increasing function, it ensues that

$$H\big(\mathcal{U}^n\big) = \log Z_n(\mathcal{U}) \le \log Z_n(\mathcal{V}) = H\big(\mathcal{V}^n\big).$$ □

Similar to the function $Z_1(\cdot)$, the functions $Z_n(\cdot)$ are submultiplicative with respect to the join operation.

Lemma 7.2.9. *Let $\mathcal{U}$ and $\mathcal{V}$ be open covers of X and let $n \in \mathbb{N}$. Then*

$$Z_n(\mathcal{U} \vee \mathcal{V}) \le Z_n(\mathcal{U})Z_n(\mathcal{V}),$$

and thus

$$H\big((\mathcal{U} \vee \mathcal{V})^n\big) \le H(\mathcal{U}^n) + H(\mathcal{V}^n).$$

Proof. Using Lemmas 7.1.12(b) and 7.2.3, we obtain that

$$\begin{aligned} Z_n(\mathcal{U} \vee \mathcal{V}) &= Z_1\big((\mathcal{U} \vee \mathcal{V})^n\big) = Z_1(\mathcal{U}^n \vee \mathcal{V}^n) \\ &\le Z_1(\mathcal{U}^n)Z_1(\mathcal{V}^n) = Z_n(\mathcal{U})Z_n(\mathcal{V}). \end{aligned}$$

Taking the logarithm of both sides gives $H((\mathcal{U} \vee \mathcal{V})^n) \le H(\mathcal{U}^n) + H(\mathcal{V}^n)$. □

We now refocus our attention on the sequence $(Z_n(\mathcal{U}))_{n=1}^{\infty}$ for a given open cover $\mathcal{U}$ of X. We have already established in Lemma 7.2.7 that this sequence is nondecreasing. Lemma 7.2.3 suggests that this sequence might be submultiplicative and might even grow exponentially with n. This explains the use of the logarithm function. By working in a logarithmic scale, we study the exponential growth rate of the numbers $(Z_n(\mathcal{U}))_{n=1}^{\infty}$. This will further ensure that the entropy of the system with respect to any specific open cover is finite.

Lemma 7.2.10. *For any open cover $\mathcal{U}$ of X, the sequence $(Z_n(\mathcal{U}))_{n=1}^{\infty}$ is submultiplicative.*

Proof. Let $m, n \in \mathbb{N}$. Choose a minimal subcover $\mathcal{A}$ of $\mathcal{U}^m$ and a minimal subcover $\mathcal{B}$ of $\mathcal{U}^n$. Using Lemma 7.1.10, we obtain that the open cover $\mathcal{A} \vee T^{-m}(\mathcal{B})$ satisfies

$$\begin{aligned} \mathcal{A} \vee T^{-m}(\mathcal{B}) &\subseteq \mathcal{U}^m \vee T^{-m}(\mathcal{U}^n) \\ &= \big(\mathcal{U} \vee \cdots \vee T^{-(m-1)}(\mathcal{U})\big) \vee T^{-m}\big(\mathcal{U} \vee \cdots \vee T^{-(n-1)}(\mathcal{U})\big) \\ &= \mathcal{U} \vee \cdots \vee T^{-(m-1)}(\mathcal{U}) \vee T^{-m}(\mathcal{U}) \vee \cdots \vee T^{-(m+n-1)}(\mathcal{U}) \\ &= \mathcal{U}^{m+n}. \end{aligned}$$

That is, $\mathcal{A} \vee T^{-m}(\mathcal{B})$ is a subcover of $\mathcal{U}^{m+n}$. Consequently,

$$Z_{m+n}(\mathcal{U}) \le \#\big(\mathcal{A} \vee T^{-m}(\mathcal{B})\big) \le \#\mathcal{A} \cdot \#\big(T^{-m}(\mathcal{B})\big) \le \#\mathcal{A} \cdot \#\mathcal{B} = Z_m(\mathcal{U})Z_n(\mathcal{U}).$$

This establishes the submultiplicativity of the sequence $(Z_n(\mathcal{U}))_{n=1}^{\infty}$. □

We immediately deduce the following.

Corollary 7.2.11. *For any open cover $\mathcal{U}$ of X, the sequence $(H(\mathcal{U}^n))_{n=1}^{\infty}$ is subadditive.*

Proof. Since $Z_{m+n}(\mathcal{U}) \le Z_m(\mathcal{U})Z_n(\mathcal{U})$ for all $m, n \in \mathbb{N}$ according to Lemma 7.2.10, taking the logarithm of both sides yields that

$$H(\mathcal{U}^{m+n}) = \log Z_{m+n}(\mathcal{U}) \le \log Z_m(\mathcal{U}) + \log Z_n(\mathcal{U}) = H(\mathcal{U}^m) + H(\mathcal{U}^n). \qquad \square$$

We are now ready to take the second step in the definition of the topological entropy of a system.

Definition 7.2.12. Let $T : X \to X$ be a dynamical system and let $\mathcal{U}$ be an open cover of X. The *topological entropy of T with respect to $\mathcal{U}$* is defined as

$$\mathrm{h_{top}}(T, \mathcal{U}) := \lim_{n\to\infty} \frac{1}{n} H(\mathcal{U}^n) = \inf_{n\in\mathbb{N}} \frac{1}{n} H(\mathcal{U}^n).$$

The existence of the limit and its equality with the infimum follow directly from combining Corollary 7.2.11 and Lemma 3.2.17. Furthermore, note that

$$\mathrm{h_{top}}(T, \mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(\mathcal{U}) = \inf_{n\in\mathbb{N}} \frac{1}{n} \log Z_n(\mathcal{U}).$$

Remark 7.2.13. Since $1 \le Z_n(\mathcal{U}) < \infty$ for all $n \in \mathbb{N}$, we readily see that

$$0 \le \mathrm{h_{top}}(T, \mathcal{U}) \le H(\mathcal{U}) < \infty.$$

Similar to the functions $Z_n(\cdot)$, the topological entropy with respect to covers respects the refinement relation. It is also subadditive with respect to the join operation, as the following proposition shows.

Proposition 7.2.14. *Let $\mathcal{U}$ and $\mathcal{V}$ be open covers of X.*
(a) *If $\mathcal{U} \prec \mathcal{V}$, then $\mathrm{h_{top}}(T, \mathcal{U}) \le \mathrm{h_{top}}(T, \mathcal{V})$. In particular, if $\mathcal{V}$ is a subcover of $\mathcal{U}$ then $\mathrm{h_{top}}(T, \mathcal{U}) \le \mathrm{h_{top}}(T, \mathcal{V})$.*
(b) $\mathrm{h_{top}}(T, \mathcal{U} \vee \mathcal{V}) \le \mathrm{h_{top}}(T, \mathcal{U}) + \mathrm{h_{top}}(T, \mathcal{V})$.

Proof. (a) By Lemma 7.2.8, we have $Z_n(\mathcal{U}) \le Z_n(\mathcal{V})$ for every $n \in \mathbb{N}$. Taking the logarithm of both sides, dividing by n and passing to the limit as n tends to infinity, we deduce that $\mathrm{h_{top}}(T, \mathcal{U}) \le \mathrm{h_{top}}(T, \mathcal{V})$ whenever $\mathcal{U} \prec \mathcal{V}$.

(b) By Lemma 7.2.9, we have $Z_n(\mathcal{U} \vee \mathcal{V}) \le Z_n(\mathcal{U})Z_n(\mathcal{V})$ for every $n \in \mathbb{N}$. Taking the logarithm of both sides, dividing by n and passing to the limit as n tends to infinity, we conclude that $\mathrm{h_{top}}(T, \mathcal{U} \vee \mathcal{V}) \le \mathrm{h_{top}}(T, \mathcal{U}) + \mathrm{h_{top}}(T, \mathcal{V})$. $\square$

An interesting property of the entropy of a system with respect to a given cover is that it remains the same for all dynamical covers generated by that cover.

Lemma 7.2.15. $\mathrm{h_{top}}(T, \mathcal{U}^k) = \mathrm{h_{top}}(T, \mathcal{U})$ *for each $k \in \mathbb{N}$.*

Proof. The case $k = 1$ is trivial. So suppose that $k \geq 2$. Let $n \in \mathbb{N}$. Lemma 7.1.12(d) asserts that $(\mathcal{U}^k)^n \prec \mathcal{U}^{n+k-1} \prec (\mathcal{U}^k)^n$. From Lemma 7.2.2, we infer that $Z_1((\mathcal{U}^k)^n) = Z_1(\mathcal{U}^{n+k-1})$, and thus

$$Z_n(\mathcal{U}^k) = Z_1((\mathcal{U}^k)^n) = Z_1(\mathcal{U}^{n+k-1}) = Z_{n+k-1}(\mathcal{U}).$$

Therefore,

$$\begin{aligned} \mathrm{h_{top}}(T,\mathcal{U}^k) &= \lim_{n\to\infty} \frac{1}{n} \log Z_n(\mathcal{U}^k) = \lim_{n\to\infty} \frac{n+k-1}{n(n+k-1)} \log Z_{n+k-1}(\mathcal{U}) \\ &= \lim_{n\to\infty} \frac{n+k-1}{n} \cdot \lim_{n\to\infty} \frac{1}{n+k-1} \log Z_{n+k-1}(\mathcal{U}) = \mathrm{h_{top}}(T,\mathcal{U}). \end{aligned}$$ □

7.2.3 Third and final stage: entropy of a system

At this point, we are in a position to give the definition of the topological entropy of a dynamical system $T : X \to X$. The topological entropy of T is defined to be the supremum over all open covers, of the entropy of the system with respect to each of these covers.

Definition 7.2.16. The *topological entropy* of T is defined to be

$$\mathrm{h_{top}}(T) := \sup\{\mathrm{h_{top}}(T,\mathcal{U}) : \mathcal{U} \text{ is an open cover of } X\}.$$

Remark 7.2.17.
(a) In view of Remark 7.2.13, we have that $0 \leq \mathrm{h_{top}}(T) \leq \infty$.
(b) The topological entropy of the identity map $\mathrm{Id}(x) = x$ is zero. Indeed, for any open cover $\mathcal{U}$ of X we have that $\mathcal{U}^n = \mathcal{U}$, and hence $Z_n(\mathcal{U}) = Z_1(\mathcal{U})$, for every $n \in \mathbb{N}$. Thus $\mathrm{h_{top}}(\mathrm{Id},\mathcal{U}) = 0$ for all open covers $\mathcal{U}$ of X, and thereby $\mathrm{h_{top}}(\mathrm{Id}) = 0$.
(c) Despite the fact that $\mathrm{h_{top}}(T,\mathcal{U}) < \infty$ for every open cover $\mathcal{U}$ of X, there exist dynamical systems T that have infinite topological entropy.
(d) As every open cover of a compact space admits a finite subcover, it follows from Proposition 7.2.14(a) that the supremum in the definition of topological entropy can be restricted to finite open covers.

Our next aim is to address the most important and natural question: Is topological entropy a topological conjugacy invariant? Before answering this question, the reader might like to recall from Chapter 1 that if $T : X \to X$ and $S : Y \to Y$ are two dynamical systems for which there exists a continuous surjection $h : X \to Y$ such that $h \circ T = S \circ h$, then S is said to be a *factor* of T. In such a situation, it is intuitively clear that $\mathrm{h_{top}}(S) \leq \mathrm{h_{top}}(T)$ since every orbit of T is projected onto an orbit of S. Thus T may have "more" orbits (in some sense) than S and is therefore at least as complex as S.

Proposition 7.2.18. *If $T : X \to X$ and $S : Y \to Y$ are two dynamical systems such that S is a factor of T, then*

$$\mathrm{h_{top}}(S) \le \mathrm{h_{top}}(T).$$

In particular, if S and T are topologically conjugate, then $\mathrm{h_{top}}(S) = \mathrm{h_{top}}(T)$.

Proof. Let $h : X \to Y$ be a factor map, so that $h \circ T = S \circ h$. Since h is a continuous surjection, every open cover $\mathcal{V}$ of Y can be lifted to the open cover $h^{-1}(\mathcal{V})$ of X. We shall prove that $\mathrm{h_{top}}(T, h^{-1}(\mathcal{V})) = \mathrm{h_{top}}(S, \mathcal{V})$ for every open cover $\mathcal{V}$ of Y. From this, we shall conclude that

$$\begin{aligned}
\mathrm{h_{top}}(T) &= \sup\{\mathrm{h_{top}}(T, \mathcal{U}) : \mathcal{U} \text{ open cover of } X\} \\
&\ge \sup\{\mathrm{h_{top}}(T, h^{-1}(\mathcal{V})) : \mathcal{V} \text{ open cover of } Y\} \\
&= \sup\{\mathrm{h_{top}}(S, \mathcal{V}) : \mathcal{V} \text{ open cover of } Y\} \\
&= \mathrm{h_{top}}(S).
\end{aligned}$$

In particular, if S and T are topologically conjugate, then S is a factor of T and T is a factor of S. So $\mathrm{h_{top}}(T) \ge \mathrm{h_{top}}(S)$ and $\mathrm{h_{top}}(S) \ge \mathrm{h_{top}}(T)$, that is, $\mathrm{h_{top}}(S) = \mathrm{h_{top}}(T)$.

It remains to prove that $\mathrm{h_{top}}(T, h^{-1}(\mathcal{V})) = \mathrm{h_{top}}(S, \mathcal{V})$ for each open cover $\mathcal{V}$ of Y. Fix $n \in \mathbb{N}$ momentarily. The respective actions of the maps S and T on $\mathcal{V}$ and $h^{-1}(\mathcal{V})$ until time $n-1$ will be denoted by $\mathcal{V}_S^n$ and $(h^{-1}(\mathcal{V}))_T^n$. Then

$$\begin{aligned}
(h^{-1}(\mathcal{V}))_T^n &= h^{-1}(\mathcal{V}) \vee T^{-1}(h^{-1}(\mathcal{V})) \vee \cdots \vee T^{-(n-1)}(h^{-1}(\mathcal{V})) \\
&= h^{-1}(\mathcal{V}) \vee (h \circ T)^{-1}(\mathcal{V}) \vee \cdots \vee (h \circ T^{n-1})^{-1}(\mathcal{V}) \\
&= h^{-1}(\mathcal{V}) \vee (S \circ h)^{-1}(\mathcal{V}) \vee \cdots \vee (S^{n-1} \circ h)^{-1}(\mathcal{V}) \\
&= h^{-1}(\mathcal{V}) \vee h^{-1}(S^{-1}(\mathcal{V})) \vee \cdots \vee h^{-1}(S^{-(n-1)}(\mathcal{V})) \\
&= h^{-1}(\mathcal{V} \vee S^{-1}(\mathcal{V}) \vee \cdots \vee S^{-(n-1)}(\mathcal{V})) \\
&= h^{-1}(\mathcal{V}_S^n).
\end{aligned}$$

Therefore,

$$Z_n(T, h^{-1}(\mathcal{V})) = Z_1((h^{-1}(\mathcal{V}))_T^n) = Z_1(h^{-1}(\mathcal{V}_S^n)) \le Z_1(\mathcal{V}_S^n) = Z_n(S, \mathcal{V}).$$

Since h is surjective, $(h^{-1}(\mathcal{V}))_T^n = h^{-1}(\mathcal{V}_S^n)$ implies that $h((h^{-1}(\mathcal{V}))_T^n) = \mathcal{V}_S^n$. Thus,

$$Z_n(S, \mathcal{V}) = Z_1(\mathcal{V}_S^n) = Z_1(h((h^{-1}(\mathcal{V}))_T^n)) \le Z_1((h^{-1}(\mathcal{V}))_T^n) = Z_n(T, h^{-1}(\mathcal{V})).$$

Combining the previous two inequalities, we obtain that

$$Z_n(T, h^{-1}(\mathcal{V})) = Z_n(S, \mathcal{V}).$$

Since n was chosen arbitrarily, by successively taking the logarithm of both sides, dividing by n and passing to the limit as n tends to infinity, we conclude that

$$\mathrm{h_{top}}(T, h^{-1}(\mathcal{V})) = \mathrm{h_{top}}(S, \mathcal{V}). \qquad \square$$

We have now shown that topological entropy is indeed a topological conjugacy invariant. Let us now study its behavior with respect to the iterates of the system.

Theorem 7.2.19. *For every $k \in \mathbb{N}$, we have* $\mathrm{h}_{\mathrm{top}}(T^k) = k\,\mathrm{h}_{\mathrm{top}}(T)$.

Proof. Fix $k \in \mathbb{N}$. Let $\mathcal{U}$ be an open cover of X. The action of the map T^k on $\mathcal{U}$ until time $n-1$ will be denoted by $\mathcal{U}^n_{T^k}$. For every $n \in \mathbb{N}$, we have that

$$\begin{aligned}
\mathcal{U}^n_{T^k} &= \mathcal{U} \vee T^{-k}(\mathcal{U}) \vee \cdots \vee T^{-(n-1)k}(\mathcal{U}) \\
&\prec \mathcal{U} \vee T^{-1}(\mathcal{U}) \vee T^{-2}(\mathcal{U}) \vee \cdots \vee T^{-k}(\mathcal{U}) \vee \cdots \vee T^{-(n-1)k}(\mathcal{U}) \\
&= \mathcal{U}^{(n-1)k+1}.
\end{aligned}$$

By Lemma 7.2.2, it ensues that

$$Z_n(T^k,\mathcal{U}) = Z_1(\mathcal{U}^n_{T^k}) \le Z_1(\mathcal{U}^{(n-1)k+1}) = Z_{(n-1)k+1}(T,\mathcal{U}).$$

Consequently,

$$\begin{aligned}
\mathrm{h}_{\mathrm{top}}(T^k,\mathcal{U}) &= \lim_{n\to\infty} \frac{1}{n} \log Z_n(T^k,\mathcal{U}) \\
&\le \lim_{n\to\infty} \frac{(n-1)k+1}{n} \frac{1}{(n-1)k+1} \log Z_{(n-1)k+1}(T,\mathcal{U}) \\
&= \lim_{n\to\infty} \frac{(n-1)k+1}{n} \cdot \lim_{n\to\infty} \frac{1}{(n-1)k+1} \log Z_{(n-1)k+1}(T,\mathcal{U}) \\
&= k \cdot \mathrm{h}_{\mathrm{top}}(T,\mathcal{U}).
\end{aligned}$$

The inequality arises from the fact that there were gaps in the iterates of T in the cover $\mathcal{U}^n_{T^k}$ that are not present in $\mathcal{U}^{(n-1)k+1}$. We can fill in those gaps by considering $\mathcal{U}^k$ rather than $\mathcal{U}$. Indeed,

$$\begin{aligned}
(\mathcal{U}^k)^n_{T^k} &= \mathcal{U}^k \vee T^{-k}(\mathcal{U}^k) \vee \cdots \vee T^{-(n-1)k}(\mathcal{U}^k) \\
&= (\mathcal{U} \vee T^{-1}(\mathcal{U}) \vee T^{-2}(\mathcal{U}) \vee \cdots \vee T^{-(k-1)}(\mathcal{U})) \\
&\quad \vee T^{-k}(\mathcal{U} \vee T^{-1}(\mathcal{U}) \vee T^{-2}(\mathcal{U}) \vee \cdots \vee T^{-(k-1)}(\mathcal{U})) \\
&\quad \vee \cdots \\
&\quad \vee T^{-(n-1)k}(\mathcal{U} \vee T^{-1}(\mathcal{U}) \vee T^{-2}(\mathcal{U}) \vee \cdots \vee T^{-(k-1)}(\mathcal{U})) \\
&= \mathcal{U} \vee T^{-1}(\mathcal{U}) \vee T^{-2}(\mathcal{U}) \vee \cdots \vee T^{-(k-1)}(\mathcal{U}) \\
&\quad \vee T^{-k}(\mathcal{U}) \vee T^{-(k+1)}(\mathcal{U}) \vee T^{-(k+2)}(\mathcal{U}) \vee \cdots \vee T^{-(2k-1)}(\mathcal{U}) \\
&\quad \vee \cdots \\
&\quad \vee T^{-(n-1)k}(\mathcal{U}) \vee T^{-((n-1)k+1)}(\mathcal{U}) \vee \cdots \vee T^{-(nk-1)}(\mathcal{U}) \\
&= \mathcal{U}^{nk}.
\end{aligned}$$

Therefore,

$$Z_n(T^k,\mathcal{U}^k) = Z_1((\mathcal{U}^k)^n_{T^k}) = Z_1(\mathcal{U}^{nk}) = Z_{nk}(T,\mathcal{U}).$$

Consequently,

$$\mathrm{h_{top}}(T^k,\mathcal{U}^k) = \lim_{n\to\infty}\frac{1}{n}\log Z_n(T^k,\mathcal{U}^k) = k\cdot\lim_{n\to\infty}\frac{1}{nk}\log Z_{nk}(T,\mathcal{U}) = k\cdot\mathrm{h_{top}}(T,\mathcal{U}).$$

Using Lemma 7.2.15, we conclude that

$$\begin{aligned}\mathrm{h_{top}}(T^k) &= \sup\{\mathrm{h_{top}}(T^k,\mathcal{U}) : \mathcal{U}\text{ open cover of }X\}\\ &= \sup\{\mathrm{h_{top}}(T^k,\mathcal{U}^k) : \mathcal{U}\text{ open cover of }X\}\\ &= k\sup\{\mathrm{h_{top}}(T,\mathcal{U}) : \mathcal{U}\text{ open cover of }X\} = k\,\mathrm{h_{top}}(T). \qquad\square\end{aligned}$$

Taking a supremum over the collection of all (finite) open covers can be inconvenient, to say the least. Indeed, this collection is usually uncountable. We would thus like to identify situations in which the topological entropy of a system is determined by the topological entropy of the system with respect to a countable family of covers, that is, with respect to a sequence of covers. If topological entropy really provides a good description of the complexity of the dynamics of the system, then it is natural to request that this sequence of covers eventually become finer and finer, and that it encompass the structure of the underlying space at increasingly small scales. In a metrizable space, this suggests looking at the diameter of the covers.

The forthcoming lemma is the first result that requires the underlying space to be metrizable. In this lemma, note that $\mathcal{U}_n$ is a general cover. In particular, it is typically not equal to the dynamical cover $\mathcal{U}^n$. So, take care not to confuse the two.

Lemma 7.2.20. *The following quantities are all equal:*
(a) $\mathrm{h_{top}}(T)$.
(b) $\sup\{\mathrm{h_{top}}(T,\mathcal{U}) : \mathcal{U}$ *open cover with* $\mathrm{diam}(\mathcal{U}) \le \delta\}$ *for any* $\delta > 0$.
(c) $\lim_{\varepsilon\to 0}\mathrm{h_{top}}(T,\mathcal{U}_\varepsilon)$ *for any open covers* $(\mathcal{U}_\varepsilon)_{\varepsilon\in(0,\infty)}$ *such that* $\lim_{\varepsilon\to 0}\mathrm{diam}(\mathcal{U}_\varepsilon) = 0$.
(d) $\lim_{n\to\infty}\mathrm{h_{top}}(T,\mathcal{U}_n)$ *for any open covers* $(\mathcal{U}_n)_{n=1}^\infty$ *such that* $\lim_{n\to\infty}\mathrm{diam}(\mathcal{U}_n) = 0$.

Proof. Clearly, (a)≥(b). It is also easy to see that (b)≥(c) for any $\delta > 0$ and any family $(\mathcal{U}_\varepsilon)_{\varepsilon\in(0,\infty)}$ as described, and that (b)≥(d) for any sequence $(\mathcal{U}_n)_{n=1}^\infty$ as specified. It thus suffices to prove that (c)≥(a) and (d)≥(a). Actually, we will prove that (d)=(a) and leave to the reader the task of adapting that proof to establish that (c)=(a).

Let $\mathcal{V}$ be any open cover of X, and let $\delta(\mathcal{V})$ be a Lebesgue number for $\mathcal{V}$. As $\lim_{n\to\infty}\mathrm{diam}(\mathcal{U}_n) = 0$, there exists $N\in\mathbb{N}$ such that $\mathrm{diam}(\mathcal{U}_n) < \delta(\mathcal{V})$ for each $n \ge N$. Fix $n \ge N$ momentarily. As $\mathrm{diam}(\mathcal{U}_n) < \delta(\mathcal{V})$, every member of $\mathcal{U}_n$ is contained in a member of $\mathcal{V}$. Thus $\mathcal{V} \prec \mathcal{U}_n$. By Proposition 7.2.14(a), we obtain that

$$\mathrm{h_{top}}(T,\mathcal{V}) \le \mathrm{h_{top}}(T,\mathcal{U}_n).$$

Since this is true for all $n \geq N$, we deduce that

$$h_{top}(T, \mathcal{V}) \leq \inf_{n \geq N} h_{top}(T, \mathcal{U}_n) \leq \liminf_{n \to \infty} h_{top}(T, \mathcal{U}_n).$$

As the open cover $\mathcal{V}$ was chosen arbitrarily, we conclude that

$$\begin{aligned} h_{top}(T) = \sup_{\mathcal{V}} h_{top}(T, \mathcal{V}) &\leq \liminf_{n \to \infty} h_{top}(T, \mathcal{U}_n) \\ &\leq \limsup_{n \to \infty} h_{top}(T, \mathcal{U}_n) \\ &\leq \sup_{\mathcal{U}} h_{top}(T, \mathcal{U}) = h_{top}(T), \end{aligned}$$

which implies that $h_{top}(T) = \lim_{n \to \infty} h_{top}(T, \mathcal{U}_n)$. □

Part (d) of Lemma 7.2.20 characterized the topological entropy of a system as the limit of the topological entropy of the system relative to a sequence of covers. An even better result would be the characterization of the topological entropy as the topological entropy with respect to a single cover. This quest suggests introducing the following notion.

Definition 7.2.21. An open cover $\mathcal{U}$ of a metric space (X, d) is said to be a *generator* for a topological dynamical system $T : X \to X$ if

$$\lim_{n \to \infty} \operatorname{diam}(\mathcal{U}^n) = 0.$$

Lemma 7.2.22. *If a system $T : X \to X$ has a generator $\mathcal{U}$, then*

$$h_{top}(T) = h_{top}(T, \mathcal{U}).$$

Proof. It follows from Lemma 7.2.20 (with $\mathcal{U}_n = \mathcal{U}^n$) and Lemma 7.2.15 that

$$h_{top}(T) = \lim_{n \to \infty} h_{top}(T, \mathcal{U}^n) = \lim_{n \to \infty} h_{top}(T, \mathcal{U}) = h_{top}(T, \mathcal{U}).$$ □

It is natural to wonder about the class(es) of systems that admit a generator.

Lemma 7.2.23. *A topological dynamical system $T : X \to X$ admits a generator if and only if it is expansive. In fact, if T is δ-expansive when the compact metric space X is equipped with a metric d, then every open cover $\mathcal{U}$ of X such that* $\operatorname{diam}(\mathcal{U}) \leq \delta$ *is a generator for T.*

Proof. Suppose that an open cover $\mathcal{V}$ is a generator for T. Let δ be a Lebesgue number for $\mathcal{V}$. If $d(T^n(x), T^n(y)) \leq \delta$ for all $n \geq 0$, then for every n there exists $V_n \in \mathcal{V}$ such that $T^n(x), T^n(y) \in V_n$. Therefore $x, y \in \bigcap_{n=0}^{\infty} T^{-n}(V_n)$. This implies that x, y lie in a common member of $\mathcal{V}^N$ for all $N \in \mathbb{N}$. Since $\lim_{N \to \infty} \operatorname{diam}(\mathcal{V}^N) = 0$, we conclude that $x = y$. So δ is an expansive constant for T.

The converse implication can be seen as an immediate consequence of uniform expansiveness. See Proposition 5.2.2. Nevertheless, we provide a direct proof. Let $\mathcal{U} = \{U_e : e \in E\}$ be a finite open cover of X with $\operatorname{diam}(\mathcal{U}) \le \delta$. This means that E is a finite index set (an alphabet), and E^∞ is compact when endowed with any of the metrics $d_s(\omega, \tau) = s^{|\omega\wedge\tau|}$ (cf. Definition 3.1.10 and Lemma 3.1.7). Let $\varepsilon > 0$. We must show that there exists $N \in \mathbb{N}$ such that $\operatorname{diam}(\mathcal{U}^n) < \varepsilon$ for all $n \ge N$. First, we claim that for every $\omega \in E^\infty$, the set

$$\bigcap_{j=0}^{\infty} T^{-j}(\overline{U_{\omega_j}})$$

comprises at most one point. To show this, let $x, y \in \bigcap_{j=0}^{\infty} T^{-j}(\overline{U_{\omega_j}})$. Then $T^j(x), T^j(y) \in \overline{U_{\omega_j}}$ for all $j \ge 0$. So, for all $j \ge 0$,

$$d\big(T^j(x), T^j(y)\big) \le \operatorname{diam}(\overline{U_{\omega_j}}) = \operatorname{diam}(U_{\omega_j}) \le \operatorname{diam}(\mathcal{U}) \le \delta.$$

By the δ-expansiveness of T, we deduce that $x = y$ and the claim is proved. It follows from this fact that

$$\lim_{n\to\infty} \operatorname{diam}\left(\bigcap_{j=0}^{n-1} T^{-j}(\overline{U_{\omega_j}})\right) = 0, \quad \forall \omega \in E^\infty.$$

Otherwise, a compactness argument shows that the set $\bigcap_{j=0}^{\infty} T^{-j}(\overline{U_{\omega_j}})$ would contain at least two points (see Exercise 7.6.6), which would contradict the claim established above.

Thus, for every $\omega \in E^\infty$ there exists a smallest $N(\omega) \in \mathbb{N}$ such that

$$\operatorname{diam}\left(\bigcap_{j=0}^{N(\omega)-1} T^{-j}(\overline{U_{\omega_j}})\right) < \varepsilon.$$

This defines a function $N : E^\infty \to \mathbb{N}$. We claim that this function is locally constant and thereby continuous. Indeed, let $\omega \in E^\infty$ and pick any $\tau \in [\omega|_{N(\omega)}] = [\omega_0\omega_1\omega_2\ldots\omega_{N(\omega)-1}]$. Then $\tau_j = \omega_j$ for all $0 \le j < N(\omega)$. Therefore, for all $1 \le n \le N(\omega)$, we have that

$$\bigcap_{j=0}^{n-1} T^{-j}(\overline{U_{\tau_j}}) = \bigcap_{j=0}^{n-1} T^{-j}(\overline{U_{\omega_j}}).$$

This implies that for all $1 \le n < N(\omega)$,

$$\operatorname{diam}\left(\bigcap_{j=0}^{n-1} T^{-j}(\overline{U_{\tau_j}})\right) = \operatorname{diam}\left(\bigcap_{j=0}^{n-1} T^{-j}(\overline{U_{\omega_j}})\right) \ge \varepsilon,$$

whereas

$$\operatorname{diam}\left(\bigcap_{j=0}^{N(\omega)-1} T^{-j}(\overline{U_{\tau_j}})\right) = \operatorname{diam}\left(\bigcap_{j=0}^{N(\omega)-1} T^{-j}(\overline{U_{\omega_j}})\right) < \varepsilon.$$

Thus $N(\tau) = N(\omega)$ for every $\tau \in [\omega|_{N(\omega)}]$. This proves that N is a locally constant function.

Since N is continuous and E^∞ is compact, the image $N(E^\infty)$ is a compact subset of $\mathbb{N}$ and is hence bounded. Set $N_{\max} := \max\{N(\omega) : \omega \in E^\infty\} < \infty$. Then for every $n \geq N_{\max}$ and for every $\omega \in E^\infty$, we have

$$\operatorname{diam}\left(\bigcap_{j=0}^{n-1} T^{-j}(U_{\omega_j})\right) \leq \operatorname{diam}\left(\bigcap_{j=0}^{N_{\max}-1} T^{-j}(\overline{U_{\omega_j}})\right) \leq \operatorname{diam}\left(\bigcap_{j=0}^{N(\omega)-1} T^{-j}(\overline{U_{\omega_j}})\right) < \varepsilon.$$

So $\operatorname{diam}(\mathcal{U}^n) < \varepsilon$ for all $n \geq N_{\max}$. As $\varepsilon > 0$ was chosen arbitrarily, we conclude that $\mathcal{U}$ is a generator for T. □

In light of the previous two results, the topological entropy of an expansive system can be characterized as the topological entropy of that system with respect to a single cover.

Theorem 7.2.24. *If $T : X \to X$ is a δ-expansive dynamical system on a compact metric space (X, d), then*

$$\mathrm{h_{top}}(T) = \mathrm{h_{top}}(T, \mathcal{U})$$

for any open cover $\mathcal{U}$ of X with $\operatorname{diam}(\mathcal{U}) \leq \delta$.

Proof. This is an immediate consequence of Lemmas 7.2.22 and 7.2.23. □

Remark 7.2.25. Notice that Theorem 7.2.24 immediately implies that the topological entropy of an expansive dynamical system on a compact metric space is finite.

Example 7.2.26. Let E be a finite alphabet and A be an incidence/transition matrix. Let $\sigma : E_A^\infty \to E_A^\infty$ be the corresponding subshift of finite type. We will show that

$$\mathrm{h_{top}}(\sigma) = \log r(A),$$

where $r(A)$ is the spectral radius of A. Recall from Example 5.1.4 that the shift map σ is expansive and has for an expansive constant any $0 < \delta < 1$ when E_A^∞ is endowed with any metric $d_s(\omega, \tau) = s^{|\omega \wedge \tau|}$, where $0 < s < 1$. Choose $\mathcal{U} := \{[e] : e \in E\}$ as a (finite) open cover of E_A^∞. So $\mathcal{U}$ is the open partition of E_A^∞ into initial 1-cylinders. Since $\operatorname{diam}(\mathcal{U}) = s < 1$, in light of Theorem 7.2.24 we know that $\mathrm{h_{top}}(\sigma) = \mathrm{h_{top}}(\sigma, \mathcal{U})$. In order to compute $\mathrm{h_{top}}(\sigma, \mathcal{U})$, notice that for each $n \in \mathbb{N}$ we have

$$\mathcal{U}^n = \sigma^{-(n-1)}(\mathcal{U}) = \{[\omega] : \omega \in E_A^n\}.$$

That is to say, $\mathcal{U}^n$ is the open partition of E_A^∞ into initial cylinders of length n. Since the only subcover that a partition admits is itself, we obtain that

$$Z_n(\mathcal{U}) = \# E_A^n.$$

Consequently,

$$\mathrm{h}_{\mathrm{top}}(\sigma) = \mathrm{h}_{\mathrm{top}}(\sigma, \mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(\mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} \log \# E_A^n = \log r(A).$$

The last equality follows from Theorem 3.2.24.

Example 7.2.27. If E is a finite alphabet, then the topological entropy of the full E-shift is equal to $\log \# E$. This is a special case of the previous example with A as the matrix that consists only of 1's. Alternatively, notice that $\#(E^n) = (\# E)^n$.

We shall now compute the topological entropy of a particular subshift of finite type, the well-known *golden mean shift*, which was introduced in Exercise 3.4.9. Perhaps not surprisingly, given its name, it will turn out that the topological entropy of the golden mean shift is equal to the logarithm of the golden mean.

Example 7.2.28. Let $\sigma : E_A^\infty \to E_A^\infty$ be the golden mean shift, that is, the subshift of finite type induced by the incidence matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

By Example 7.2.26, the computation of the topological entropy of a subshift of finite type boils down to counting the number of initial n-cylinders for all but finitely many n's, or more simply to computing the largest eigenvalue (in absolute value) of the transition matrix A.

The reader can calculate that the golden mean $\gamma := (1+\sqrt{5})/2$ and its conjugate $\gamma^* := (1-\sqrt{5})/2$ are the eigenvalues of the matrix A. Therefore,

$$\mathrm{h}_{\mathrm{top}}(\sigma) = \log r(A) = \log \max\{|\gamma|, |\gamma^*|\} = \log \gamma.$$

Alternatively, one can prove by induction that $\#(E_A^n) = f_{n+2}$, where f_n is the nth Fibonacci number (see Exercise 7.6.7). Then one can verify by induction that

$$f_n = \frac{1}{\sqrt{5}}[\gamma^n - (\gamma^*)^n] = \frac{\gamma^n}{\sqrt{5}}[1 - (\gamma^*/\gamma)^n].$$

It follows immediately that

$$\begin{aligned}
\mathrm{h}_{\mathrm{top}}(\sigma) &= \lim_{n\to\infty} \frac{1}{n} \log \#(E_A^n) = \lim_{n\to\infty} \frac{1}{n} \log f_{n+2} \\
&= \lim_{n\to\infty} \frac{1}{n} \log\left(\frac{\gamma^{n+2}}{\sqrt{5}} [1 - (\gamma^*/\gamma)^{n+2}] \right) \\
&= \lim_{n\to\infty} \frac{1}{n} \left((n+2)\log\gamma - \log\sqrt{5} + \log[1 - (\gamma^*/\gamma)^{n+2}] \right) \\
&= \log \gamma.
\end{aligned}$$

7.3 Bowen's definition of topological entropy

In this section, we present an alternative definition of topological entropy due to Rufus Bowen. In contrast with the definition via open covers, this definition is only valid in a metric space X. Although Bowen's definition does make sense in a noncompact metric space, we will as usual assume that X is compact throughout.

Let $T : X \to X$ be a dynamical system on a compact metric space (X, d). Recall from Definition 5.1.3 the dynamical metrics (also called Bowen's metrics), d_n, for each $n \in \mathbb{N}$:

$$d_n(x, y) = \max\{d(T^j(x), T^j(y)) : 0 \leq j < n\}.$$

The open ball centered at x of radius r induced by the metric d_n is denoted by $B_n(x, r)$ and is called the dynamical (n, r)-ball at x. As $d_1 = d$, we shall denote $B_1(x, r)$ simply by $B(x, r)$. Finally, observe that

$$B_n(x, r) = \{y \in X : d_n(x, y) < r\} = \bigcap_{j=0}^{n-1} T^{-j}(B(T^j(x), r)).$$

In other words, the ball $B_n(x, r)$ consists of those points whose iterates stay within a distance r from the corresponding iterates of x until time $n-1$ at least. In the language of Chapter 4, the ball $B_n(x, r)$ is the set of all those points whose orbits are r-shadowed by the orbit of x until time $n - 1$ at least.

Definition 7.3.1. A subset E of X is said to be (n, ε)-*separated* if E is ε-separated with respect to the metric d_n, which is to say that $d_n(x, y) \geq \varepsilon$ for all $x, y \in E$ with $x \neq y$.

Remark 7.3.2.
(a) If E is an (m, ε)-separated set and $m < n$, then E is also (n, ε)-separated.
(b) If E is an (n, ε')-separated set and $\varepsilon < \varepsilon'$, then E is also (n, ε)-separated.
(c) Given that the underlying space X is compact, any (n, ε)-separated set is finite. Indeed, let E be an (n, ε)-separated set, and consider the family of balls $\{B_n(x, \varepsilon/2) : x \in E\}$. If the intersection of $B_n(x, \varepsilon/2)$ and $B_n(y, \varepsilon/2)$ is nonempty for some $x, y \in E$, then there exists $z \in B_n(x, \varepsilon/2) \cap B_n(y, \varepsilon/2)$ and it follows that

$$d_n(x, y) \leq d_n(x, z) + d_n(z, y) < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

As E is an (n, ε)-separated set, this inequality implies that $x = y$. This means that the balls $\{B_n(x, \varepsilon/2) : x \in E\}$ are mutually disjoint. Hence, as X is compact, there can only be finitely many points in E.

The largest separated sets will be especially useful in describing the complexity of the dynamics that the system exhibits.

Definition 7.3.3. A subset E of X is called a *maximal* (n, ε)*-separated set* if for any (n, ε)-separated set E' with $E \subseteq E'$, we have $E = E'$. In other words, no strict superset of E is (n, ε)-separated.

The counterpart of the notion of separated set is the concept of spanning set.

Definition 7.3.4. A subset F of X is said to be an (n, ε)*-spanning set* if

$$\bigcup_{x \in F} B_n(x, \varepsilon) = X.$$

That is, the orbit of every point in the space is ε-shadowed by the orbit of a point of F until time $n - 1$ at least.

The smallest spanning sets play a special role in describing the complexity of the dynamics that the system possesses. They constitute the counterpart of the maximal separated sets.

Definition 7.3.5. A subset F of X is called a *minimal* (n, ε)*-spanning set* if for any (n, ε)-spanning set F' with $F \supseteq F'$, we have $F = F'$. In other words, no strict subset of F is (n, ε)-spanning.

Remark 7.3.6.
(a) If F is an (n, ε)-spanning set and $m < n$, then F is also (m, ε)-spanning.
(b) If F is an (n, ε)-spanning set and $\varepsilon < \varepsilon'$, then F is also (n, ε')-spanning.
(c) Any minimal (n, ε)-spanning set is finite since the open cover $\{B_n(x, \varepsilon) : x \in X\}$ of the compact metric space X admits a finite subcover.

The next lemma describes two useful relations between separated and spanning sets.

Lemma 7.3.7. *The following statements hold:*
(a) *Every maximal* (n, ε)*-separated set is a minimal* (n, ε)*-spanning set.*
(b) *Every* $(n, 2\varepsilon)$*-separated set is embedded into any* (n, ε)*-spanning set.*

Proof. (a) Let E be a maximal (n, ε)-separated set. First, suppose that E is not (n, ε)-spanning. Then there exists a point $y \in X \backslash \bigcup_{x \in E} B_n(x, \varepsilon)$. Observe then that the set $E \cup \{y\}$ is (n, ε)-separated, which contradicts the maximality of E. Therefore, E is (n, ε)-spanning. Suppose now that E is not a minimal (n, ε)-spanning set. Then there exists an (n, ε)-spanning set $E' \subsetneq E$. Let $y \in E \backslash E'$. Since $\bigcup_{y' \in E'} B_n(y', \varepsilon) = X$, there is $y' \in E' \subseteq E$ such that $d_n(y', y) < \varepsilon$. This implies that E is not (n, ε)-separated, a contradiction. Consequently, if E is a maximal (n, ε)-separated set, then E is a minimal (n, ε)-spanning set.

(b) Let E be an $(n, 2\varepsilon)$-separated set and F an (n, ε)-spanning set. For each $x \in E$, choose $i(x) \in F$ such that $x \in B_n(i(x), \varepsilon)$. We claim that the map $i : E \to F$ is injective. To show this, let $x, y \in E$ be such that $i(x) = i(y) =: z$. Then $x, y \in B_n(z, \varepsilon)$. Therefore

$d_n(x,y) < 2\varepsilon$. Since E is a $(n, 2\varepsilon)$-separated set, we deduce that $x = y$, that is, the map i is injective and so E is embedded into F. □

The next theorem is the main result of this section. It gives us another way of defining the topological entropy of a system.

Theorem 7.3.8. *For all $\varepsilon > 0$ and $n \in \mathbb{N}$, let $E_n(\varepsilon)$ be a maximal (n,ε)-separated set in X and $F_n(\varepsilon)$ be a minimal (n,ε)-spanning set in X. Then*

$$\begin{aligned}\mathrm{h_{top}}(T) = \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\#E_n(\varepsilon) &= \lim_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\#E_n(\varepsilon)\\ &\le \liminf_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\#F_n(\varepsilon).\end{aligned}$$

Proof. Fix $\varepsilon > 0$ and $n \in \mathbb{N}$ and let $E_n(\varepsilon)$ be a maximal (n,ε)-separated set in X. Let $\mathcal{U}$ be an open cover of X consisting of balls of radius $\varepsilon/2$. Let $\underline{\mathcal{U}}$ be a minimal subcover of $\mathcal{U}^n$, so that $Z_n(\mathcal{U}) = \#\underline{\mathcal{U}}$. For each $x \in E_n(\varepsilon)$, let $\underline{U}(x)$ be an element of the cover $\underline{\mathcal{U}}$ which contains x and define the function $i : E_n(\varepsilon) \to \underline{\mathcal{U}}$ by setting $i(x) = \underline{U}(x)$. We claim that this function is an injection. Indeed, suppose that $x, y \in E_n(\varepsilon)$ are such that $\underline{U}(x) = \underline{U}(y)$. Then, by the definition of $\mathcal{U}^n$, we have that

$$x, y \in \bigcap_{j=0}^{n-1} T^{-j}(U_j),$$

where $U_j = B(z_j, \varepsilon/2)$ for some $z_j \in X$. This means that both $T^j(x)$ and $T^j(y)$ belong to $B(x_j, \varepsilon/2)$ for each $0 \le j < n$. So, $d_n(x,y) < \varepsilon$, and thus $x = y$ since $E_n(\varepsilon)$ is (n,ε)-separated. This establishes that $i : E_n(\varepsilon) \to \underline{\mathcal{U}}$ is injective. Therefore, $Z_n(\mathcal{U}) = \#\underline{\mathcal{U}} \ge \#E_n(\varepsilon)$. Since $\mathcal{U}$ does not depend on n and the inequality $Z_n(\mathcal{U}) \ge \#E_n(\varepsilon)$ holds for all $n \in \mathbb{N}$, we deduce that

$$\mathrm{h_{top}}(T) \ge \mathrm{h_{top}}(T,\mathcal{U}) = \lim_{n\to\infty}\frac{1}{n}\log Z_n(\mathcal{U}) \ge \limsup_{n\to\infty}\frac{1}{n}\log\#E_n(\varepsilon).$$

Consequently,

$$\mathrm{h_{top}}(T) \ge \limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\#E_n(\varepsilon). \tag{7.1}$$

On the other hand, if $\mathcal{V}$ is an arbitrary open cover of X, if $\delta(\mathcal{V})$ is a Lebesgue number for $\mathcal{V}$, if $0 < \varepsilon < \delta(\mathcal{V})/2$ and if $n \in \mathbb{N}$, then for all $0 \le k < n$ and all $x \in E_n(\varepsilon)$, we have that

$$T^k(B_n(x,\varepsilon)) \subseteq B(T^k(x),\varepsilon) \implies \mathrm{diam}(T^k(B_n(x,\varepsilon))) \le 2\varepsilon < \delta(\mathcal{V}).$$

Hence, for all $0 \le k < n$, the set $T^k(B_n(x,\varepsilon))$ is contained in at least one element of the cover $\mathcal{V}$. Denote one of these elements by $V_k(x)$. It follows that $B_n(x,\varepsilon) \subseteq T^{-k}(V_k(x))$

for each $0 \le k < n$. In other words, we have that $B_n(x,\varepsilon) \subseteq \bigcap_{k=0}^{n-1} T^{-k}(V_k(x))$. But this intersection is an element of $\mathcal{V}^n$. Let us denote it by $V(x)$.

Since $E_n(\varepsilon)$ is a maximal (n,ε)-separated set, Lemma 7.3.7(a) asserts that it is also (n,ε)-spanning. Thus the family $\{B_n(x,\varepsilon)\}_{x\in E_n(\varepsilon)}$ is an open cover of X. Each one of these balls is contained in the corresponding element $V(x)$ of $\mathcal{V}^n$. Hence, the family $\{V(x)\}_{x\in E_n(\varepsilon)}$ is also an open cover of X, and thus a subcover of $\mathcal{V}^n$. Consequently,

$$Z_n(\mathcal{V}) \le \#\{V(x)\}_{x\in E_n(\varepsilon)} \le \#E_n(\varepsilon).$$

Since this is true for all $n \in \mathbb{N}$, we deduce that

$$\mathrm{h}_{\mathrm{top}}(T,\mathcal{V}) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(\mathcal{V}) \le \liminf_{n\to\infty} \frac{1}{n} \log \#E_n(\varepsilon).$$

As this inequality holds for all $0 < \varepsilon < \delta(\mathcal{V})/2$, we obtain that

$$\mathrm{h}_{\mathrm{top}}(T,\mathcal{V}) \le \liminf_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \#E_n(\varepsilon).$$

Because $\mathcal{V}$ was chosen to be an arbitrary open cover of X, we conclude that

$$\mathrm{h}_{\mathrm{top}}(T) \le \liminf_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \#E_n(\varepsilon). \tag{7.2}$$

The inequalities (7.1) and (7.2) combined are sufficient to establish the result for the separated sets.

Now, since every $(n,2\varepsilon)$-separated set is embedded into any (n,ε)-spanning set according to Lemma 7.3.7(b), we have that

$$\#E_n(2\varepsilon) \le \#F_n(\varepsilon). \tag{7.3}$$

The inequalities (7.2) and (7.3) suffice to deduce the result for the spanning sets. □

In Theorem 7.3.8, the topological entropy of the system is expressed in terms of a specific family of maximal separated (resp., minimal spanning) sets. However, to derive theoretical results, it is sometimes simpler to use the following quantities.

Definition 7.3.9. For all $n \in \mathbb{N}$ and $\varepsilon > 0$, let

$$r_n(\varepsilon) = \sup\{\#E_n(\varepsilon) : E_n(\varepsilon) \text{ maximal } (n,\varepsilon)\text{-separated set}\}$$

and

$$s_n(\varepsilon) = \inf\{\#F_n(\varepsilon) : F_n(\varepsilon) \text{ minimal } (n,\varepsilon)\text{-spanning set}\}.$$

Thereafter, let

$$\underline{r}(\varepsilon) = \liminf_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon) \quad \text{while} \quad \overline{r}(\varepsilon) = \limsup_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon)$$

and

$$\underline{s}(\varepsilon) = \liminf_{n\to\infty} \frac{1}{n} \log s_n(\varepsilon) \quad \text{whereas} \quad \overline{s}(\varepsilon) = \limsup_{n\to\infty} \frac{1}{n} \log s_n(\varepsilon).$$

The following are simple but key observations about these quantities.

Remark 7.3.10. For all $m < n \in \mathbb{N}$ and $0 < \varepsilon < \varepsilon'$, the following relations hold.
(a) $r_m(\varepsilon) \le r_n(\varepsilon)$ and $s_m(\varepsilon) \le s_n(\varepsilon)$ by Remarks 7.3.2 and 7.3.6.
(b) $r_n(\varepsilon) \ge r_n(\varepsilon')$ and $s_n(\varepsilon) \ge s_n(\varepsilon')$ by Remarks 7.3.2 and 7.3.6.
(c) $0 < s_n(\varepsilon) \le r_n(\varepsilon) \le s_n(\varepsilon/2) < \infty$ by Lemma 7.3.7.
(d) $\underline{r}(\varepsilon) \le \overline{r}(\varepsilon)$ and $\underline{s}(\varepsilon) \le \overline{s}(\varepsilon)$.
(e) $\underline{r}(\varepsilon) \ge \underline{r}(\varepsilon')$ and $\overline{r}(\varepsilon) \ge \overline{r}(\varepsilon')$ by (b).
(f) $\underline{s}(\varepsilon) \ge \underline{s}(\varepsilon')$ and $\overline{s}(\varepsilon) \ge \overline{s}(\varepsilon')$ by (b).
(g) $0 \le \overline{s}(\varepsilon) \le \overline{r}(\varepsilon) \le \overline{s}(\varepsilon/2) \le \infty$ by (c).
(h) $0 \le \underline{s}(\varepsilon) \le \underline{r}(\varepsilon) \le \underline{s}(\varepsilon/2) \le \infty$ by (c).

We will now prove two properties that relate r_n's, s_n's, and Z_n's.

Lemma 7.3.11. *The following relations hold:*
(a) *If $\mathcal{U}$ is an open cover of X with Lebesgue number 2δ, then*

$$Z_n(\mathcal{U}) \le s_n(\delta) \le r_n(\delta), \quad \forall n \in \mathbb{N}.$$

(b) *If $\varepsilon > 0$ and $\mathcal{V}$ is an open cover of X with* $\operatorname{diam}(\mathcal{V}) \le \varepsilon$, *then*

$$s_n(\varepsilon) \le r_n(\varepsilon) \le Z_n(\mathcal{V}), \quad \forall n \in \mathbb{N}.$$

Proof. Let $n \in \mathbb{N}$. We already know that $s_n(\delta) \le r_n(\delta)$.

(a) Let $\mathcal{U}$ be an open cover with Lebesgue number 2δ and let F be an (n,δ)-spanning set. Then the dynamic balls $\{B_n(x,\delta) : x \in F\}$ form a cover of X. For every $0 \le i < n$ the ball $B(T^i(x),\delta)$, which has diameter at most 2δ, is contained in an element of $\mathcal{U}$. Therefore, $B_n(x,\delta) = \bigcap_{i=0}^{n-1} T^{-i}(B(T^i(x),\delta))$ is contained in an element of $\mathcal{U}^n = \bigvee_{i=0}^{n-1} T^{-i}(\mathcal{U})$. That is, $\mathcal{U}^n \prec \{B_n(x,\delta) : x \in F\}$. Thus $Z_n(\mathcal{U}) = Z_1(\mathcal{U}^n) \le \#F$. Since F is an arbitrary (n,δ)-spanning set, it ensues that $Z_n(\mathcal{U}) \le s_n(\delta)$.

(b) Let $\mathcal{V}$ be an open cover with $\operatorname{diam}(\mathcal{V}) \le \varepsilon$ and let E be an (n,ε)-separated set. Then no element of the cover $\mathcal{V}^n$ contains more than one element of E. Hence, $\#E \le Z_n(\mathcal{V})$. Since E is an arbitrary (n,ε)-separated set, it follows that $r_n(\varepsilon) \le Z_n(\mathcal{V})$. □

Together, Lemmas 7.3.11 and 7.2.20 have the following immediate corollary. Unlike Theorem 7.3.8, this result is symmetric with respect to separated and spanning sets. It is the advantage of using spanning sets of minimal cardinality, rather than spanning sets that are minimal in terms of inclusion.

Corollary 7.3.12. *The following equalities hold:*

$$\mathrm{h_{top}}(T) = \lim_{\varepsilon\to0} \underline{r}(\varepsilon) = \lim_{\varepsilon\to0} \overline{r}(\varepsilon) = \lim_{\varepsilon\to0} \underline{s}(\varepsilon) = \lim_{\varepsilon\to0} \overline{s}(\varepsilon).$$

Corollary 7.3.12 is useful to derive theoretical results. Nevertheless, in practice, Theorem 7.3.8 is simpler to use, as only one family (in essence, a double sequence) of sets is needed. Sometimes a single sequence is enough.

Theorem 7.3.13. *If a topological dynamical system $T : X \to X$ admits a generator with Lebesgue number 2δ, then the following statements hold for all $0 < \varepsilon \le \delta$:*
(a) *If $(E_n(\varepsilon))_{n=1}^{\infty}$ is a sequence of maximal (n,ε)-separated sets in X, then*

$$\mathrm{h_{top}}(T) = \lim_{n\to\infty} \frac{1}{n} \log \#E_n(\varepsilon).$$

(b) *If $(F_n(\varepsilon))_{n=1}^{\infty}$ is a sequence of minimal (n,ε)-spanning sets in X, then*

$$\mathrm{h_{top}}(T) \le \liminf_{n\to\infty} \frac{1}{n} \log \#F_n(\varepsilon).$$

(c) $\mathrm{h_{top}}(T) = \lim_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon)$.
(d) $\mathrm{h_{top}}(T) = \lim_{n\to\infty} \frac{1}{n} \log s_n(\varepsilon)$.

Proof. We will prove (a) and leave it to the reader to show the other parts using similar arguments.

Let $\mathcal{U}$ be a generator for T with Lebesgue number 2δ. Set $0 < \varepsilon \le \delta$. Observe that 2ε is also a Lebesgue number for $\mathcal{U}$. Choose any sequence $(E_n(\varepsilon))_{n=1}^{\infty}$ of maximal (n,ε)-separated sets. Since maximal (n,ε)-separated sets are (n,ε)-spanning sets, it follows from Lemma 7.3.11(a) that $Z_n(\mathcal{U}) \le s_n(\varepsilon) \le \#E_n(\varepsilon)$. Therefore,

$$\mathrm{h_{top}}(T,\mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(\mathcal{U}) \le \liminf_{n\to\infty} \frac{1}{n} \log \#E_n(\varepsilon). \tag{7.4}$$

On the other hand, since $\mathcal{U}$ is a generator, there exists $k \in \mathbb{N}$ such that $\mathrm{diam}(\mathcal{U}^k) \le \varepsilon$. It ensues from Lemma 7.3.11(b) that $\#E_n(\varepsilon) \le r_n(\varepsilon) \le Z_n(\mathcal{U}^k)$. Consequently,

$$\limsup_{n\to\infty} \frac{1}{n} \log \#E_n(\varepsilon) \le \lim_{n\to\infty} \frac{1}{n} \log Z_n(\mathcal{U}^k) = \mathrm{h_{top}}(T,\mathcal{U}^k) = \mathrm{h_{top}}(T,\mathcal{U}), \tag{7.5}$$

where the last equality follows from Lemma 7.2.15. Combining (7.4) and (7.5) gives

$$\mathrm{h_{top}}(T,\mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} \log \#E_n(\varepsilon).$$

As $\mathcal{U}$ is a generator, $\mathrm{h_{top}}(T) = \mathrm{h_{top}}(T,\mathcal{U})$ by Lemma 7.2.22. □

For expansive systems, the Lebesgue number can be expressed in terms of the expansive constant.

Theorem 7.3.14. *If $T : X \to X$ is a δ_0-expansive dynamical system on a compact metric space (X,d), then Theorem 7.3.13 applies with any $0 < \delta < \delta_0/4$.*

Proof. According to Theorem 7.2.23, any open cover $\mathcal{U}$ of X with $\mathrm{diam}(\mathcal{U}) \le \delta_0$ is a generator for T. In particular, a cover composed of open balls works. Let $x_1,\dots,x_n \in X$ be such that $X = \bigcup_{i=1}^{n} B(x_i, \delta_0/2 - 2\delta)$. Then the cover $\{B(x_i,\delta_0/2) : 1 \le i \le n\}$ has diameter at most δ_0 and admits 2δ as Lebesgue number. □

7.4 Topological degree

In this section, which in the context of this book is a preparation for the subsequent section where we will give a very useful lower bound for the topological entropy of a C^1 endomorphism, we introduce the concept of topological degree for maps of smooth compact orientable manifolds. Although this is in fact a topological concept well-defined for continuous maps of topological manifolds, we concentrate on differentiable maps. This is somewhat easier and perfectly fits the needs of the proof of the entropy bound mentioned above. For more information on these notions, please see Hirsch [30].

Definition 7.4.1. Let M and N be smooth compact orientable d-dimensional manifolds. Let $f : M \to N$ be a C^1 map. A point $x \in M$ is called a regular point for f if $D_x f$ is invertible. A point $y \in N$ is called a regular value of f if $f^{-1}(y)$ consists of regular points. Otherwise, y is called a singular value of f.

It is obvious that the set of regular values is open. It is also easy to see that the preimage of any regular value is of finite cardinality.

Lemma 7.4.2. *If $y \in N$ is a regular value of a C^1 map $f : M \to N$, then $\# f^{-1}(y) < \infty$.*

Proof. Suppose that $\# f^{-1}(y) = \infty$. Then there exists a sequence $(x_n)_{n=1}^{\infty}$ in M such that $\lim_{n\to\infty} x_n = x$ for some $x \in M$ and $f(x_n) = y$ for all $n \in \mathbb{N}$. The continuity of f imposes that $f(x) = y$, and thus f is not injective on any neighborhood of x. By the implicit function theorem, it follows that $D_x f$ is not invertible, that is, x is not a regular point. So y is a singular value. □

Definition 7.4.3. Suppose that $y \in N$ is a regular value of a C^1 map $f : M \to N$. For each $x \in f^{-1}(y)$, let $\epsilon_x = \pm 1$ depending on whether $D_x f$ preserves or reverses orientation. Then the degree of f at y is defined by

$$\deg_y(f) = \sum_{x \in f^{-1}(y)} \epsilon_x.$$

The preimage of any regular value being of finite cardinality, this sum is a well-defined integer. The degree of f at y measures how many times f covers N near y, counted with appropriate positive and negative multiplicities. In fact, the degree is independent of the choice of the regular value in N. In order to show that, there exists an alternative definition based on integration.

Definition 7.4.4. A positive normalized volume element on N is a continuous d-form ω that is positive on positively oriented frames and such that $\int_M \omega = 1$. The pullback $f^*\omega$ of ω under a C^1 map $f : M \to N$ is the d-form on M given by

$$(f^*\omega)(v_1, \ldots, v_d) = \omega(Df(v_1), \ldots, Df(v_d)).$$

Definition 7.4.5. If ω is a positive normalized volume element on N, then the degree of a C^1 map $f : M \to N$ with respect to ω is defined by

$$\deg_\omega(f) = \int_M f^*\omega.$$

We now show that the two aforementioned definitions of degree are independent of y and ω, respectively.

Lemma 7.4.6. *Let $y \in N$ be a regular value of a C^1 map $f : M \to N$ and ω a positive normalized volume element on N. Then* $\deg_y(f) = \deg_\omega(f)$.

Proof. As y is a regular value, there are disjoint open neighborhoods $U_1, \ldots, U_k \subseteq M$ of the points $x_1, \ldots, x_k$ of $f^{-1}(y)$ such that $\bigcup_{i=1}^k U_i$ is the preimage of a neighborhood V of y and $f|_{U_i}$ is a diffeomorphism for all i. If ν is an n-form supported in V such that $\int_V \nu = \int_N \nu = 1$, then $\omega = \nu + d\alpha$ for some $(n-1)$-form α, so $\int_M f^*\omega = \int_M (f^*\nu + f^*d\alpha) = \int_M f^*\nu = \sum_{i=1}^k \int (f|_{U_i})^*\nu = \sum_{i=1}^k \int_{U_i} f^*\nu$. By the transformation rule, each of the latter integrals is ± 1 according to whether $f|_{U_i}$, or equivalently $D_{x_i}f$, preserves or reverses orientation. Consequently, $\deg_\omega(f) = \int_M f^*\omega = \deg_y(f)$. □

So we can now make the following definition.

Definition 7.4.7. The degree of a C^1 map $f : M \to N$ is defined by $\deg(f) := \deg_y(f)$ for any regular value $y \in N$.

7.5 Misiurewicz–Przytycki theorem

In this section, we shall provide a very effective lower bound for the topological entropy of C^1 endomorphisms. Its attractiveness lies in it being expressed in relatively simple terms. The much stronger theorem of Yomdin [79], commonly referred to as the *entropy conjecture* and which gives the lower bound on topological entropy in terms of the logarithm of the spectral radius of the map induced on the full homology ring, is incomparably harder to prove and, often, harder to apply. The proof below is a slight modification of the one given in [33].

Theorem 7.5.1 (Misiurewicz–Przytycki theorem). *If M is a smooth compact orientable manifold and $T : M \to M$ a C^1 endomorphism, then* $\mathrm{h_{top}}(T) \geq \log|\deg(T)|$.

Proof. Fix a volume element ω on M and $\alpha \in (0,1)$. Let $L := \sup_{x\in M} \|D_xT\|$, and ϵ be such that $2\epsilon^{1-\alpha}L^\alpha = 1$. Set $B := \{x \in M : \|D_xT\| \geq \epsilon\}$. Pick a cover of B by open sets on which T is injective and let δ be a Lebesgue number for the cover. Thus, if $x, y \in B$ and $d(x,y) \leq \delta$ then $T(x) \neq T(y)$.

For every $n \in \mathbb{N}$, let

$$A := \{x \in M : \#(B \cap \{x, T(x), \ldots, T^{n-1}(x)\}) \leq [\alpha n]\},$$

where $[\cdot]$ denotes the integer part function. Observe that $\lim_{n\to\infty}\frac{[\alpha n]}{n} = \alpha$. If $x \in A$ and n is so large that $\epsilon^{1-\frac{[\alpha n]}{n}} L^{\frac{[\alpha n]}{n}} \leq 2\epsilon^{1-\alpha}L^{\alpha}$, then

$$\|D_x T^n\| = \prod_{j=0}^{n-1}\|D_{T^j(x)}T\| < \epsilon^{n-[\alpha n]}L^{[\alpha n]} = \left(\epsilon^{1-\frac{[\alpha n]}{n}} L^{\frac{[\alpha n]}{n}}\right)^n \leq \left(2\epsilon^{1-\alpha}L^{\alpha}\right)^n = 1.$$

Hence, the volume of $T^n(A)$ is less than that of M. But Sard's theorem asserts that the set of singular values has Lebesgue measure zero (for more information, see [30]). Therefore, there exists a regular value x of T^n that lies in $M \setminus T^n(A)$.

We will now extract an (n, δ)-separated set from $T^{-n}(x)$. Since x is regular for T, it has at least $N := |\deg(T)|$ preimages. If at least N of them are in B (a "good transition") then take Q_1 to consist of N such preimages. Otherwise (a "bad transition"), take Q_1 to be a single preimage outside B. Either way, $Q_1 \subseteq T^{-1}(x)$ consists of regular values of T since x is a regular value of T^n. Thus we can apply the same procedure to every $y \in Q_1$ and by collecting all of the points chosen that way obtain $Q_2 \subseteq T^{-2}(x)$, and so on. The set $Q_n \subseteq T^{-n}(x)$ we hence obtain is (n, δ)-separated. Indeed, suppose that $y_1, y_2 \in Q_n$ and $d(T^k(y_1), T^k(y_2)) < \delta$ for all $k \in \{0, \ldots, n-1\}$. Then $T^{n-1}(y_1), T^{n-1}(y_2) \in Q_1$. If $T^{n-1}(y_1) \neq T^{n-1}(y_2)$, then by construction of Q_1 we know that $T^{n-1}(y_1), T^{n-1}(y_2) \in B$. Moreover, $T(T^{n-1}(y_1)) = x = T(T^{n-1}(y_2))$ and $d(T^{n-1}(y_1), T^{n-1}(y_2)) < \delta$. By definition of δ, we deduce that $T^{n-1}(y_1) = T^{n-1}(y_2)$. This contradicts the assumption that $T^{n-1}(y_1) \neq T^{n-1}(y_2)$. So $T^{n-1}(y_1) = T^{n-1}(y_2)$. Likewise $T^{n-2}(y_1) = T^{n-2}(y_2)$, and so forth, so $y_1 = y_2$ and Q_n is (n, δ)-separated.

Now $Q_n \subseteq T^{-n}(x) \subseteq T^{-n}(M \setminus T^n(A)) \subseteq M \setminus A$, that is, $Q_n \cap A = \emptyset$. Thus, for any $y \in Q_n$ there are by definition of A more than αn numbers $k \in \{0, \ldots, n-1\}$ for which $T^k(y) \in B$. So in passing from x to any $y \in Q_n$ there are at least $m := [\alpha n] + 1$ "good transitions," and hence $\#Q_n \geq N^m \geq N^{\alpha n}$. Therefore, the maximal cardinality of an (n, δ)-separated set is at least $N^{\alpha n}$, and thus $\mathrm{h_{top}}(T) \geq \alpha \log N$ by Theorem 7.3.8. Since this holds for all $\alpha \in (0, 1)$, it ensues that $\mathrm{h_{top}}(T) \geq \log N = \log|\deg(T)|$. □

The two properties of smoothness that make the preceding proof work are boundedness of the derivative together with the fact that a smooth map is a local homeomorphism near any point where the derivative is nonzero.

There are certain classes of systems where the inequality given in Theorem 7.5.1 becomes an equality. Expanding maps of a compact manifold (e. g., the unit circle) form one such class. So do rational functions of the Riemann sphere

$$T(z) = \frac{P(z)}{Q(z)},$$

where P and Q are relatively prime polynomials. Since these maps are orientation preserving, their degree is equal to the number of preimages of a regular point $w \in \widehat{\mathbb{C}}$, that is, the number of solutions to the equation $T(z) = w$. The degree of T is therefore equal to the maximum of the algebraic degrees of P and Q.

7.6 Exercises

Exercise 7.6.1. Prove Remark 7.1.3.

Exercise 7.6.2. Prove Lemma 7.1.5.

Exercise 7.6.3. Prove Lemma 7.1.10.

Exercise 7.6.4. Fill in the details of the proof of Lemma 7.1.12.

Exercise 7.6.5. Using Lemmas 7.3.11 and 7.2.20, prove Corollary 7.3.12. Then prove Theorem 7.3.13(b,c,d).

Exercise 7.6.6. Let X be a metric space. Let $(X_n)_{n=1}^{\infty}$ be a descending sequence of nonempty compact subsets of X. Prove that $\bigcap_{n=1}^{\infty} X_n$ is a singleton if and only if

$$\lim_{n\to\infty} \operatorname{diam}(X_n) = 0.$$

Furthermore, show that this result does not generally hold if the sequence is not descending.

Exercise 7.6.7. Let $E = \{0,1\}$. Let $\sigma : E_A^{\infty} \to E_A^{\infty}$ be the golden mean shift, that is, the subshift of finite type induced by the incidence matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

In this exercise, you will prove by induction that $\#(E_A^n) = f_{n+2}$ for all $n \in \mathbb{N}$, where f_n is the nth Fibonacci number.

(a) Show that $E_A^{n+1} = (E_A^n \times \{0\}) \bigcup (E_A^{n,0} \times \{1\})$ for all $n \in \mathbb{N}$, where $E_A^{n,0}$ is the set of all words in E_A^n whose last letter is 0.
(b) Prove that $E_A^{n,0} = E_A^{n-1} \times \{0\}$ for all $n \in \mathbb{N}$.
(c) Deduce that $\#(E_A^{n+1}) = \#(E_A^n) + \#(E_A^{n-1})$ for all $n \in \mathbb{N}$.
(d) Conclude that $\#(E_A^n) = f_{n+2}$ for all $n \in \mathbb{N}$.

Exercise 7.6.8. Prove that the topological entropy of any isometry of a compact metric space is equal to zero.

Exercise 7.6.9. Show that $\mathrm{h_{top}}(T^{-1}) = \mathrm{h_{top}}(T)$ for any homeomorphism $T : X \to X$ of a compact metrizable space X.

Exercise 7.6.10. Prove that the topological entropy of any homeomorphism of the unit circle is equal to zero.

Exercise 7.6.11. Let X be a countable compact metrizable space and $T : X \to X$ a dynamical system. Show that $\mathrm{h_{top}}(T) = 0$.

Exercise 7.6.12. Prove that the topological entropy of every transitive, open, expansive dynamical system is positive.

Exercise 7.6.13. Let $T : X \to X$ be a topological dynamical system and F be a closed forward T-invariant subset of X. Show that the entropy of the subsystem $T|_F : F \to F$ satisfies $\mathrm{h_{top}}(T|_F) \le \mathrm{h_{top}}(T)$.

Exercise 7.6.14. Let $T : X \to X$ be a topological dynamical system and $F_1, \ldots, F_n$ be finitely many closed forward T-invariant subsets of X covering X. Prove that $\mathrm{h_{top}}(T) = \max\{\mathrm{h_{top}}(T|_{F_i}) : 1 \le i \le n\}$.

Exercise 7.6.15. Find two dynamical systems which have the same topological entropy but are not topologically conjugate.

Exercise 7.6.16. The formula $\mathrm{h_{top}}(T^n) = n\,\mathrm{h_{top}}(T)$ may suggest that $\mathrm{h_{top}}(T \circ S) = \mathrm{h_{top}}(T) + \mathrm{h_{top}}(S)$. Show that this is not true in general even if S and T commute.

Exercise 7.6.17. Let $d \in \mathbb{N}$. Prove that the topological entropy of the map of the unit circle $z \mapsto z^d$ is equal to $\log d$.

Exercise 7.6.18. For every dynamical system $T : X \to X$, let

$$\deg(T) := \min_{x \in X} \#T^{-1}(x).$$

Show that if T is a local homeomorphism, then $\mathrm{h_{top}}(T) \ge \log \deg(T)$.

8 Ergodic theory

In this chapter, we move away from the study of purely topological dynamical systems to consider instead dynamical systems that come equipped with a measure. That is, instead of self-maps acting on compact metrizable spaces, we now ask that the self-maps act upon measure spaces.

The etymology of the word *ergodic* is found in the amalgamation of the two Greek words *ergon* (meaning "work") and *odos* (meaning "path"). This term was coined by the great physicist Ludwig Boltzmann while carrying out research in statistical mechanics. The goal of ergodic theory is to study the temporal and spatial long-term behavior of a measure-preserving dynamical system. Given such a system and a measurable subset of the space it acts on, it is natural to ask with which frequency the orbits of "typical" points visit that subset. One way to think about ergodic systems is that they are systems such that the visiting frequency of orbits is equal to the measure of the subset visited. In other words, "time averages" are equal to "space averages" for these systems. This will all be made precise shortly.

The chapter is organized as follows. Section 8.1 introduces the basic object of study in ergodic theory, namely, invariant measures. In brief, a measure is said to be invariant under a measurable self-map if the measure of the set of points that are mapped to a measurable subset is equal to the measure of that subset. Section 8.2 presents the notion of ergodicity and comprises a demonstration of Birkhoff's ergodic theorem, proved by G. D. Birkhoff [9] in 1931. This theorem is the most fundamental result in ergodic theory. It is extremely useful in numerous applications. Birkhoff's original proof was very involved and complex. Over time some simplifications were brought by several authors. The simple proof we provide here originates from the short and elegant one due to Katok and Hasselblatt [33]. The class of ergodic measures for a given transformation is then studied in more detail. The penultimate Section 8.3 contains an introduction to various measure-theoretic mixing properties that a system may have (which ought to be compared to the topological mixing introduced in Chapter 1). It shows that ergodicity is a very weak form of mixing. In the final Section 8.4, Rokhlin's construction of an invertible system from any given dynamical system is described and the mixing properties of this natural extension are investigated.

The reader who is not familiar with, or desires a refresher on, measure theory is encouraged to consult Appendix A. We will repeatedly refer to it in this and subsequent chapters.

8.1 Measure-preserving transformations

Throughout, $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ will denote measurable spaces, and the transformation $T : (X, \mathcal{A}) \to (Y, \mathcal{B})$ will be measurable.

https://doi.org/10.1515/9783110702682-008

If the domain of T is endowed with a measure, then the measurable transformation T induces a measure on its codomain.

Definition 8.1.1. Let $T : (X, \mathcal{A}) \to (Y, \mathcal{B})$ be a measurable transformation and let μ be a measure on $(X, \mathcal{A})$. The set function $\mu \circ T^{-1} : \mathcal{B} \to [0, \infty]$, defined by

$$(\mu \circ T^{-1})(B) := \mu(T^{-1}(B)), \quad \forall B \in \mathcal{B},$$

is a measure on $(Y, \mathcal{B})$ called the *push-down or push-forward* of the measure μ under the transformation T.

The integration of a measurable function $f : (Y, \mathcal{B}) \to \mathbb{R}$ with respect to the measure $\mu \circ T^{-1}$ can be carried out by integrating the composition of f and T with respect to the measure μ.

Lemma 8.1.2. *If $T : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B})$ is a measurable transformation, then*

$$\int_Y f \, d(\mu \circ T^{-1}) = \int_X f \circ T \, d\mu$$

for all measurable functions $f : (Y, \mathcal{B}) \to \mathbb{R}$ such that the integral $\int_X f \circ T \, d\mu$ is defined.

Proof. It is easy to see that the equality holds for characteristic functions and, by linearity of the integral, for nonnegative measurable simple functions. The result then follows for any nonnegative measurable function by approaching it pointwise via an increasing sequence of nonnegative measurable simple functions (see Theorem A.1.17 in Appendix A) and calling upon the monotone convergence theorem (Theorem A.1.35). Finally, any measurable function can be expressed as the difference between its positive and negative parts, which are both nonnegative measurable functions. □

Measure-preserving transformations are transformations between measure spaces for which the push down of the measure on the domain coincides with the measure on the codomain.

Definition 8.1.3. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measure spaces. A measurable transformation $T : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ is said to be *measure-preserving* if $\mu \circ T^{-1} = \nu$.

Proving measure preservation for all the elements of a σ-algebra is generally an onerous task. As for equality of measures, when the measures under consideration are finite, it suffices to prove measure preservation on a π-system that generates the σ-algebra on the codomain.

Lemma 8.1.4. *Let $T : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ be a measurable transformation between probability spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$. If $\mathcal{B} = \sigma(\mathcal{P})$ is a σ-algebra generated by a π-system $\mathcal{P}$ on Y, then*

$$T \text{ is measure-preserving} \quad \Longleftrightarrow \quad \mu \circ T^{-1}(P) = \nu(P), \ \forall P \in \mathcal{P}.$$

Proof. This follows immediately from Lemma A.1.26. □

Let us now consider self-transformations, that is, transformations whose codomain coincides with their domain.

Definition 8.1.5. A measure-preserving self-transformation $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$, that is, a measurable self-transformation such that $\mu \circ T^{-1} = \mu$, is called a *measure-preserving dynamical system*. Alternatively, μ is said to be *T-invariant* or *invariant with respect to T*.

Note that if a measurable transformation $T : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ is invertible and its inverse T^{-1} is measurable, then $\mu(T^{-1}(B)) = \nu(B)$ for every $B \in \mathcal{B}$ if and only if $\mu(A) = \nu(T(A))$ for every $A \in \mathcal{A}$. In particular, if $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$, then μ is T-invariant if and only if μ is T^{-1}-invariant. This justifies the following definitions.

Definition 8.1.6. A measure-preserving transformation $T : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$, which is invertible and whose inverse is measurable, is called a *measure-preserving isomorphism*.

Definition 8.1.7. A measure-preserving dynamical system $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$, which is invertible and whose inverse is measurable, is called a *measure-preserving automorphism*.

8.1.1 Examples of invariant measures

In this section, we give several examples of invariant measures for various transformations.

Example 8.1.8. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation with a fixed point x_0, that is, $T(x_0) = x_0$. Let δ_{x_0} be the Dirac point mass supported at x_0 (cf. Example A.1.21). Then δ_{x_0} is T-invariant, that is, $\delta_{x_0}(T^{-1}(A)) = \delta_{x_0}(A)$ for each $A \in \mathcal{A}$, since $x_0 \in T^{-1}(A)$ if and only if $x_0 \in A$. This example easily generalizes to invariant measures supported on periodic orbits.

Example 8.1.9. Let $\mathbb{S}^1 = [0, 2\pi] \pmod{2\pi}$. Let $\alpha \in \mathbb{R}$ and define the map $T_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ by

$$T_\alpha(x) = x + 2\pi\alpha \pmod{2\pi}.$$

Thus T_α is the rotation of the unit circle by the angle $2\pi\alpha$. The topological dynamics of T_α are radically different depending on whether the number α is rational or irrational (see Theorem 1.5.12). So will be the ergodicity of T_α with respect to the Lebesgue measure λ. However, it is fairly easy to foresee that T_α preserves λ, irrespective of the nature of α. Indeed, $T_\alpha^{-1}(x) = x - 2\pi\alpha \pmod{2\pi}$ and so $|\det DT_\alpha^{-1}(x)| = 1$ for all $x \in \mathbb{S}^1$.

Therefore,

$$\lambda(T_\alpha^{-1}(B)) = \int_B |\det DT_\alpha^{-1}(x)|\, d\lambda(x) = \int_B d\lambda(x) = \lambda(B)$$

for all $B \in \mathcal{B}(\mathbb{S}^1)$, that is, T_α preserves λ. Since T_α is invertible and its inverse is measurable, T_α is a Lebesgue measure-preserving automorphism.

Example 8.1.10. Fix $n \in \mathbb{N}$ and consider the map $T_n : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $T_n(x) := nx \pmod 1$, where $\mathbb{S}^1$ is equipped with the Borel σ-algebra $\mathcal{B}(\mathbb{S}^1)$ and the Lebesgue measure λ. We claim that λ is T-invariant. Let I be a proper subinterval of $\mathbb{S}^1$. Then $T_n^{-1}(I)$ consists of n mutually disjoint intervals (arcs) of length $\frac{1}{n}\lambda(I)$. Consequently,

$$\lambda(T_n^{-1}(I)) = n \cdot \frac{1}{n}\lambda(I) = \lambda(I).$$

Since the family of all proper subintervals of $\mathbb{S}^1$ forms a π-system which generates $\mathcal{B}(\mathbb{S}^1)$ and since T_n preserves the Lebesgue measure of all proper subintervals, Lemma 8.1.4 asserts that T_n preserves λ.

Example 8.1.11.

(a) Recall the tent map $T : [0,1] \to [0,1]$ from Example 1.1.3:

$$T(x) := \begin{cases} 2x & \text{if } x \in [0,1/2] \\ 2-2x & \text{if } x \in [1/2,1]. \end{cases}$$

The family of all intervals $\{[a,b),(a,b) : 0 < a < b < 1\}$ forms a π-system that generates the Borel σ-algebra $\mathcal{B}([0,1])$. Since the preimage of any such interval consists of 2 disjoint subintervals (one on each side of the tent) of half the length of the original interval, one readily sees from Lemma 8.1.4 that the Lebesgue measure on $[0,1]$ is invariant under the tent map.

(b) In fact, the previous example generalizes to a much larger family of maps. Let $T : [0,1] \to [0,1]$ be a piecewise linear map of the unit interval that admits a “partition” $\mathcal{P} = \{p_j\}_{j=0}^q$, where $1 \le q < \infty$ and $0 = p_0 < p_1 < \cdots < p_{q-1} < p_q = 1$, with the following properties:

(1) $[0,1] = I_1 \cup \cdots \cup I_q$, where $I_j = [p_{j-1}, p_j]$’s are the successive intervals of monotonicity of T.

(2) $T(I_j) = [0,1]$ for all $1 \le j \le q$.

(3) T is linear on I_j for all $1 \le j \le q$.

Such a map T will be called a *full Markov map*. We claim such a T preserves the Lebesgue measure λ. Indeed, it is easy to see that the absolute value of the slope of the restriction of T to the interval I_j is $1/(p_j - p_{j-1})$. Therefore, the absolute value of the slope of the corresponding inverse branch of T is $p_j - p_{j-1}$. Let $I \subseteq (0,1)$ be any interval. Then $T^{-1}(I) = \bigcup_{j=1}^q T|_{I_j}^{-1}(I)$, where $T|_{I_j}^{-1}(I)$ is a subinterval of $\mathrm{Int}(I_j)$ of length $(p_j - p_{j-1}) \cdot \lambda(I)$. Since $\mathrm{Int}(I_j) \cap \mathrm{Int}(I_k) = \emptyset$ for all $1 \le j < k \le q$, it ensues from that set disjointness (see Lemma A.1.19(g) if necessary) that

$$\lambda(T^{-1}(I)) = \sum_{j=1}^{q} \lambda(T|_{I_j}^{-1}(I)) = \sum_{j=1}^{q} (p_j - p_{j-1}) \cdot \lambda(I) = (p_q - p_0)\lambda(I) = \lambda(I).$$

Moreover, $0 \le \lambda(T^{-1}(\{0\})) \le \lambda(\{p_j : 0 \le j \le q\}) = 0$. So $\lambda(T^{-1}(\{0\})) = 0 = \lambda(\{0\})$. Similarly, $\lambda(T^{-1}(\{1\})) = 0 = \lambda(\{1\})$. It follows that $\lambda(T^{-1}(J)) = \lambda(J)$ for every interval $J \subseteq [0,1]$. Since the family of all intervals in $[0,1]$ forms a π-system that generates $\mathcal{B}([0,1])$, Lemma 8.1.4 asserts that the Lebesgue measure is invariant under any full Markov map.

Example 8.1.12. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ and $S : (Y, \mathcal{B}) \to (Y, \mathcal{B})$ be measurable transformations for which there exists a measurable transformation $h : (X, \mathcal{A}) \to (Y, \mathcal{B})$ such that $h \circ T = S \circ h$. We will show that every T-invariant measure generates an S-invariant push down under h. Let μ be a T-invariant measure on $(X, \mathcal{A})$. Recall that the push down of μ under h is the measure $\mu \circ h^{-1}$ on $(Y, \mathcal{B})$. It follows from the T-invariance of μ that

$$(\mu \circ h^{-1}) \circ S^{-1} = \mu \circ (S \circ h)^{-1} = \mu \circ (h \circ T)^{-1} = (\mu \circ T^{-1}) \circ h^{-1} = \mu \circ h^{-1}.$$

That is, the push down $\mu \circ h^{-1}$ is S-invariant.

Example 8.1.13. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two probability spaces, and let $T : X \to X$ and $S : Y \to Y$ be two measure-preserving dynamical systems. The direct product of T and S is the map $T \times S : X \times Y \to X \times Y$ defined by

$$(T \times S)(x, y) = (T(x), S(y)).$$

The direct product σ-algebra $\sigma(\mathcal{A} \times \mathcal{B})$ on $X \times Y$ is the σ-algebra generated by the semialgebra of measurable rectangles

$$\mathcal{A} \times \mathcal{B} := \{A \times B : A \in \mathcal{A},\ B \in \mathcal{B}\}.$$

The direct product measure $\mu \times \nu$ on $(X \times Y, \sigma(\mathcal{A} \times \mathcal{B}))$ is uniquely determined by its values on the generating semialgebra, values which are naturally given by

$$(\mu \times \nu)(A \times B) := \mu(A)\nu(B).$$

The existence and uniqueness of this product measure can be established using Theorem A.1.27, Lemma A.1.29, and Theorem A.1.28. For more information, see Halmos [27] (pp. 157–158) or Taylor [72] (Chapter III, Section 4). We claim that the product map $T \times S : (X \times Y, \sigma(\mathcal{A} \times \mathcal{B}), \mu \times \nu) \to (X \times Y, \sigma(\mathcal{A} \times \mathcal{B}), \mu \times \nu)$ is measure-preserving.

Thanks to Lemma 8.1.4, it suffices to show that $(\mu\times\nu)\circ(T\times S)^{-1}(A\times B) = (\mu\times\nu)(A\times B)$ for all $A\times B\in\mathcal{A}\times\mathcal{B}$. And indeed,

$$\begin{aligned}(\mu\times\nu)\circ(T\times S)^{-1}(A\times B) &= (\mu\times\nu)\big(T^{-1}(A)\times S^{-1}(B)\big)\\ &= \mu(T^{-1}(A))\nu(S^{-1}(B))\\ &= \mu(A)\nu(B)\\ &= (\mu\times\nu)(A\times B).\end{aligned}$$

The final example pertains to the shift map introduced in Chapter 3 on symbolic dynamics. In fact, we look at this map in a more general context.

Example 8.1.14. Let $(E,\mathcal{F},P)$ be a probability space. Consider the one-sided product set $E^{\mathbb{N}} = E^{\infty} := \prod_{k=1}^{\infty} E$. The product σ-algebra $\mathcal{F}_\infty$ on E^∞ is the σ-algebra generated by the semialgebra $\mathcal{S}$ of all (finite) cylinders (also called rectangles), that is, sets of the form

$$\prod_{k=1}^{n} E_k \times \prod_{l=n+1}^{\infty} E = \big\{\omega = (\omega_j)_{j=1}^{\infty}\in E^\infty : \omega_k\in E_k,\ \forall 1\le k\le n\big\},$$

where $n\in\mathbb{N}$ and $E_k\in\mathcal{F}$ for all $1\le k\le n$. The product measure μ_P on $\mathcal{F}_\infty$ is the unique probability measure which confers to a cylinder the value

$$\mu_P\left(\prod_{k=1}^{n} E_k\times\prod_{l=n+1}^{\infty} E\right) = \prod_{k=1}^{n} P(E_k). \tag{8.1}$$

The existence and uniqueness of this measure can be established using Theorem A.1.27, Lemma A.1.29, and Theorem A.1.28. For more information, see Halmos [27] (pp. 157–158) or Taylor [72] (Chapter III, Section 4).

As in Chapter 3, let $\sigma : E^\infty \to E^\infty$ be the left shift map, which is defined by $\sigma((\omega_n)_{n=1}^{\infty}) := (\omega_{n+1})_{n=1}^{\infty}$. The product measure μ_P is σ-invariant. Indeed, since the cylinder sets form a semialgebra which generates the product σ-algebra, in light of Lemma 8.1.4 it is sufficient to show that $\mu_P(\sigma^{-1}(S)) = \mu_P(S)$ for all cylinder sets $S\in\mathcal{S}$. And we have

$$\begin{aligned}\mu_P\circ\sigma^{-1}\left(\prod_{k=1}^{n} E_k\times\prod_{l=n+1}^{\infty} E\right) &= \mu_P\left(E\times\prod_{k=1}^{n} E_k\times\prod_{l=n+2}^{\infty} E\right)\\ &= P(E)\prod_{k=1}^{n} P(E_k)\\ &= \prod_{k=1}^{n} P(E_k)\\ &= \mu_P\left(\prod_{k=1}^{n} E_k\times\prod_{l=n+1}^{\infty} E\right).\end{aligned}$$

This completes the proof that the product measure is shift-invariant.

The measure-preserving dynamical system $(\sigma : E^\infty \to E^\infty, \mu_P)$ is commonly referred to as a *one-sided Bernoulli shift* with set of states E. Of particular importance is the case when E is a finite set having at least two elements. Also, the case of a countably infinite set of states E will be of special importance in this book. This will be particularly transparent in Chapters 13 and 16 onward in the second volume, where we will consider Gibbs states of Hölder continuous potentials for which Bernoulli measures are very special cases.

More examples of invariant measures can be found in Exercises 8.5.22 and 8.5.25–8.5.33.

8.1.2 Poincaré's recurrence theorem

We now present one of the fundamental results of finite ergodic theory, namely, Poincaré's recurrence theorem. This theorem states that, in a finite measure space, almost all points of a given set return infinitely often to that set under iteration. It is worth pointing out that Poincaré's recurrence theorem is striking (and, as we will see, unusual), in that its hypotheses are so completely general.

But, first, we show that the points from a measurable set that never return to that set under iteration are negligible. That is, they form a subset of measure zero.

Lemma 8.1.15. *If $T : X \to X$ is a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, then for every set $A \in \mathcal{A}$ we have*

$$\mu(\{x \in A : T^n(x) \notin A,\ \forall n \in \mathbb{N}\}) = 0.$$

Proof. Let $A \in \mathcal{A}$ and

$$N = N(T, A) := \{x \in A : T^n(x) \notin A,\ \forall n \in \mathbb{N}\}.$$

To show that $\mu(N) = 0$, let $x \in N$. Then $T^n(x) \notin A$ for every $n \in \mathbb{N}$. Therefore, $T^n(x) \notin N$ for all $n \in \mathbb{N}$. Thus $N \cap T^{-n}(N) = \emptyset$ for all $n \in \mathbb{N}$. Now, fix $k \in \mathbb{N}$ and let $1 \le j < k$. Then

$$T^{-j}(N) \cap T^{-k}(N) = T^{-j}(N \cap T^{-(k-j)}(N)) = T^{-j}(\emptyset) = \emptyset.$$

So the preimages $\{T^{-n}(N)\}_{n=0}^{\infty}$ of N under the iterates of T form a pairwise disjoint family of sets. It follows that

$$1 = \mu(X) \ge \mu\left(\bigcup_{n=0}^{\infty} T^{-n}(N)\right) = \sum_{n=0}^{\infty} \mu(T^{-n}(N)) = \sum_{n=0}^{\infty} \mu(N),$$

where the last equality follows from the T-invariance of μ. Hence, $\mu(N) = 0$. □

Knowing this, we can now demonstrate that, in a finite measure space, almost all points of a given set return infinitely often to that set under iteration.

Theorem 8.1.16 (Poincaré's recurrence theorem). *If $T : X \to X$ is a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, then*

$$\mu(\{x \in A : T^n(x) \in A \text{ for infinitely many } n \in \mathbb{N}\}) = \mu(A)$$

for every set $A \in \mathcal{A}$.

Proof. Let

$$N(T, A) := \{x \in A : T^n(x) \notin A, \ \forall n \in \mathbb{N}\}.$$

For each $k \in \mathbb{N}$, let

$$\begin{aligned} N_k &:= \{x \in A : T^n(x) \notin A, \ \forall n \geq k\} \\ &\subseteq \{x \in A : T^{kj}(x) \notin A, \ \forall j \in \mathbb{N}\} = N(T^k, A). \end{aligned}$$

Replacing T by T^k in Lemma 8.1.15, we obtain that $\mu(N(T^k, A)) = 0$ and thereby $\mu(N_k) = 0$ for all $k \in \mathbb{N}$. It follows that

$$\mu\left(\bigcup_{k=1}^{\infty} N_k\right) = 0.$$

Observe also that

$$\{x \in A : T^n(x) \in A \text{ for infinitely many } n \in \mathbb{N}\} = A \setminus \bigcup_{k=1}^{\infty} N_k.$$

Consequently,

$$\mu(\{x \in A : T^n(x) \in A \text{ for infinitely many } n\}) = \mu(A) - \mu\left(\bigcup_{k=1}^{\infty} N_k\right) = \mu(A). \qquad \square$$

Note that this result does not generally hold in infinite measure spaces. Indeed, the simplest counterexample is a translation of the real line. For instance, take the transformation of the real line $T(x) = x + 1$, which certainly preserves the Lebesgue measure, and let $A = (0, 1)$. No point of A ever comes back to A under iteration, although the Lebesgue measure of A is evidently not equal to zero.

8.1.3 Existence of invariant measures

In general, a measurable transformation $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ may not admit any invariant measure (see Exercise 8.5.29). There are also measurable transformations that admit infinite invariant measures but not finite ones (see Exercise 8.5.30). However, if

X is a compact metrizable space and $\mathcal{A}$ is the Borel σ-algebra $\mathcal{B}(X)$, then every continuous transformation does have an invariant Borel probability measure. In other words, every topological dynamical system admits an invariant Borel probability measure. Before proving this, we will study the properties of the set of invariant probability measures.

Definition 8.1.17. Let $(X, \mathcal{A})$ be a measurable space. The set of all probability measures on $(X, \mathcal{A})$ is denoted by $M(X, \mathcal{A})$. Given a measurable transformation $T : (X, \mathcal{A}) \to (X, \mathcal{A})$, the subset of all T-invariant probability measures on $(X, \mathcal{A})$ is denoted by $M(T, \mathcal{A})$.

In particular, if $\mathcal{A}$ is the Borel σ-algebra on a topological space X, then the set of all Borel probability measures on X is simply denoted by $M(X)$ instead of $M(X, \mathcal{B}(X))$, while its subset of T-invariant measures is denoted by $M(T)$ rather than $M(T, \mathcal{B}(X))$. For more information about $M(X)$, please see Subsection A.1.8.

For topological dynamical systems, there exists a characterization of invariant Borel probability measures in terms of the way they integrate continuous functions.

Theorem 8.1.18. *Let $T : X \to X$ be a topological dynamical system. Then*

$$\mu \in M(T) \quad \Longleftrightarrow \quad \int_X f \circ T \, d\mu = \int_X f \, d\mu, \quad \forall f \in C(X).$$

Proof. This follows immediately from Lemma 8.1.2 and Corollary A.1.54. □

We will now show that the set $M(T)$ is a compact and convex subset of $M(X)$ whenever T is a topological dynamical system.

In fact, the convexity holds for all measurable transformations.

Lemma 8.1.19. *For any measurable transformation $T : (X, \mathcal{A}) \to (X, \mathcal{A})$, the set $M(T, \mathcal{A})$ is a convex subset of $M(X, \mathcal{A})$.*

Proof. Let $\mu, \nu \in M(T, \mathcal{A})$. Let m be a convex combination of μ and ν, that is, let $\alpha \in [0, 1]$ and let $m = \alpha\mu + (1 - \alpha)\nu$. By the obvious convexity of $M(X, \mathcal{A})$, we already know that $m \in M(X, \mathcal{A})$. Let $A \in \mathcal{A}$. Observe that

$$m(T^{-1}(A)) = \alpha\mu(T^{-1}(A)) + (1 - \alpha)\nu(T^{-1}(A)) = \alpha\mu(A) + (1 - \alpha)\nu(A) = m(A).$$

Thus, m is T-invariant, and hence $m \in M(T, \mathcal{A})$. □

Theorem 8.1.20. *Let $T : X \to X$ be a topological dynamical system. The set $M(T)$ is a compact convex subset of the compact convex space $M(X)$ in the* weak* *topology.*

Proof. The convexity of $M(T)$ has been established in Lemma 8.1.19. Thus, we can concentrate on demonstrating the compactness of $M(T)$. According to Theorem A.1.58, the set $M(X)$ is compact in the weak* topology of $C(X)^*$. Hence, it suffices to show that

$M(T)$ is closed in $M(X)$. As described in Subsection A.1.8, the space $M(X)$ admits a metric compatible with the weak* topology. Therefore, a set is closed in that space if and only if it is sequentially closed. Let $(\mu_n)_{n=1}^{\infty}$ be a sequence in $M(T)$ which converges to a measure μ in $M(X)$. We aim to show that $\mu \in M(T)$. To that end, let $f \in C(X)$. According to Theorem 8.1.18, it suffices to show that $\int_X f \circ T \, d\mu = \int_X f \, d\mu$. Since $(\mu_n)_{n=1}^{\infty}$ converges to μ in the weak* topology and $\mu_n \in M(T)$ for all n, we deduce that

$$\int\limits_X f \circ T \, d\mu = \lim_{n\to\infty} \int\limits_X f \circ T \, d\mu_n = \lim_{n\to\infty} \int\limits_X f \, d\mu_n = \int\limits_X f \, d\mu.$$

Thus $\mu \in M(T)$ and $M(T)$ is closed in $M(X)$. As a closed subset of the compact set $M(X)$, the set $M(T)$ is compact as well. □

We now briefly examine the map $\mu \mapsto \mu \circ T^{-1}$, which will be helpful in proving the existence of invariant measures.

Lemma 8.1.21. *Let $T : X \to X$ be a topological dynamical system. The map $S : M(X) \to M(X)$, where $S(\mu) = \mu \circ T^{-1}$, is continuous and affine.*

Proof. The proof of affinity is left to the reader. We concentrate on the continuity of S. Let $(\mu_n)_{n=1}^{\infty}$ be a sequence in $M(X)$ which weak* converges to μ. Then, for any $f \in C(X)$,

$$\begin{aligned}
\lim_{n\to\infty} \int\limits_X f \, d(S(\mu_n)) &= \lim_{n\to\infty} \int\limits_X f \, d(\mu_n \circ T^{-1}) \\
&= \lim_{n\to\infty} \int\limits_X f \circ T \, d\mu_n \\
&= \int\limits_X f \circ T \, d\mu = \int\limits_X f \, d(\mu \circ T^{-1}) \\
&= \int\limits_X f \, d(S(\mu)).
\end{aligned}$$

Since f was chosen arbitrarily in $C(X)$, the sequence $(S(\mu_n))_{n=1}^{\infty}$ weak* converges to $S(\mu)$. Thus S is continuous. □

We come now to the main result of this section, namely, showing that every topological dynamical system admits at least one invariant Borel probability measure. This theorem is not very difficult to prove, but it is obviously important. For this reason, we provide two different proofs. The first involves functional analysis, whereas the second is rather more constructive.

Theorem 8.1.22 (Krylov–Bogolyubov theorem). *Let $T : X \to X$ be a topological dynamical system. Then $M(T) \neq \emptyset$.*

Proof. By the Riesz representation theorem (Theorem A.1.53), the set $M(X)$ can be identified with a subset of the Banach space $C(X)^*$. According to Theorem A.1.58, the set $M(X)$ is compact and convex in the weak* topology of $C(X)^*$. By Lemma 8.1.21, we also know that the map $S(\mu) := \mu \circ T^{-1}$ is a continuous affine self-map of $M(X)$. Thus, by Schauder–Tychonoff's fixed-point theorem (cf. Theorem V.10.5 in Dunford and Schwartz [20]) the map S has a fixed point. In other words, there exists $\mu \in M(X)$ such that $\mu \circ T^{-1} = \mu$. Note: Alternatively, since S is affine, one may use Markov-Kakutani's fixed-point theorem. It is more elementary and proved in Theorem V.10.6 of [20].

Alternative proof. Let $\mu_0 \in M(X)$ (for example, a Dirac point mass supported at a point of X). Construct the sequence of Borel probability measures $(\mu_n)_{n=1}^{\infty}$, where

$$\mu_n = \frac{1}{n}\sum_{j=0}^{n-1} \mu_0 \circ T^{-j}.$$

Since $M(X)$ is compact in the weak* topology, the sequence $(\mu_n)_{n=1}^{\infty}$ has at least one weak* limit point. Denote such a point by μ_∞. We claim that $\mu_\infty \in M(T)$. To show this, let $(\mu_{n_k})_{k=1}^{\infty}$ be a subsequence of the sequence $(\mu_n)_{n=1}^{\infty}$ which weak* converges to μ_∞. The weak* convergence of the subsequence means that

$$\int_X f\,d\mu_\infty = \lim_{k\to\infty}\int_X f\,d\mu_{n_k}, \quad \forall f \in C(X).$$

Moreover,

$$\begin{aligned}
\left|\int_X f\circ T\,d\mu_{n_k} - \int_X f\,d\mu_{n_k}\right| &= \left|\frac{1}{n_k}\sum_{j=0}^{n_k-1}\int_X f\circ T\,d(\mu_0\circ T^{-j}) - \frac{1}{n_k}\sum_{j=0}^{n_k-1}\int_X f\,d(\mu_0\circ T^{-j})\right| \\
&= \frac{1}{n_k}\left|\sum_{j=0}^{n_k-1}\left(\int_X f\circ T^{j+1}\,d\mu_0 - \int_X f\circ T^j\,d\mu_0\right)\right| \\
&= \frac{1}{n_k}\left|\int_X f\circ T^{n_k}\,d\mu_0 - \int_X f\,d\mu_0\right| \\
&\le \frac{2}{n_k}\|f\|_\infty.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left|\int_X f\circ T\,d\mu_\infty - \int_X f\,d\mu_\infty\right| &= \left|\lim_{k\to\infty}\int_X f\circ T\,d\mu_{n_k} - \lim_{k\to\infty}\int_X f\,d\mu_{n_k}\right| \\
&= \lim_{k\to\infty}\left|\int_X f\circ T\,d\mu_{n_k} - \int_X f\,d\mu_{n_k}\right| \\
&\le 2\|f\|_\infty \lim_{k\to\infty}\frac{1}{n_k} = 0.
\end{aligned}$$

Hence,

$$\int_X f\,d(\mu_\infty \circ T^{-1}) = \int_X f \circ T\,d\mu_\infty = \int_X f\,d\mu_\infty, \quad \forall f \in C(X).$$

By Corollary A.1.54, we conclude that $\mu_\infty \circ T^{-1} = \mu_\infty$. □

The Krylov–Bogolyubov theorem can be restated as follows: Every topological dynamical system induces at least one measure-preserving dynamical system.

8.2 Ergodic transformations

One of the aims of the present section is to state and demonstrate the first published ergodic theorem, originally proved at the outset of the 1930s by George David Birkhoff. Of course, before setting out to prove an ergodic theorem, we must first define and investigate the notion of ergodicity.

Definition 8.2.1. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. Then T is said to be *ergodic* with respect to a measure μ on $(X, \mathcal{A})$ if all completely T-invariant sets $A \in \mathcal{A}$, that is, such that $T^{-1}(A) = A$, have the property that $\mu(A) = 0$ or $\mu(X \backslash A) = 0$. Alternatively, μ is said to be T-ergodic or ergodic with respect to T.

A system is ergodic if and only if it does not admit any nontrivial subsystem. See alternative definitions in Exercise 8.5.36.

The following is a simple but important observation.

Lemma 8.2.2. *If a measure μ is ergodic with respect to a measurable transformation $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ and if a measure ν on $(X, \mathcal{A})$ is absolutely continuous with respect to μ, then ν is also ergodic. That is,*

$$\mu \text{ ergodic } \;\&\; \nu \ll \mu \quad \Longrightarrow \quad \nu \text{ ergodic.}$$

Proof. The proof is left to the reader as an exercise. □

While complete invariance of a set is an appropriate concept for topological dynamical systems, we will see that a more suitable notion for measure-preserving dynamical systems is that of almost invariance of a set.

Definition 8.2.3. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and let μ be a measure on $(X, \mathcal{A})$. A set $A \in \mathcal{A}$ is said to be *μ-almost T-invariant* if $\mu(T^{-1}(A) \triangle A) = 0$.

Of course, any completely T-invariant set is μ-almost T-invariant.

Our next goal is to show that a measure μ is ergodic if and only if all μ-almost T-invariant sets are trivial in a measure-theoretic sense, that is, have measure zero or full measure. The proof of this characterization of ergodic measures boils down to constructing a completely T-invariant set from an almost T-invariant one. This raises the more general question: Given an arbitrary set S, how can we construct from that set a completely T-invariant one?

If a set R is forward T-invariant, that is, if $T^{-1}(R) \supseteq R$, then a related completely T-invariant set is the union of all the preimages of R, that is, the set $\bigcup_{k=0}^{\infty} T^{-k}(R)$ of all points whose orbits eventually hit R. This reduces our question to the following one: Given a set S, how can we construct from it a forward T-invariant set R? One possibility is the intersection of all the preimages of S, that is, the set $R = \bigcap_{n=0}^{\infty} T^{-n}(S)$ of all points whose orbits are trapped within S. Hence, the set

$$\bigcup_{k=0}^{\infty} T^{-k}\left(\bigcap_{n=0}^{\infty} T^{-n}(S)\right) = \bigcup_{k=0}^{\infty}\bigcap_{n=0}^{\infty} T^{-(k+n)}(S) = \bigcup_{k=0}^{\infty}\bigcap_{n=k}^{\infty} T^{-n}(S)$$

is completely T-invariant. This is the set of all points whose iterates eventually fall into S and remain trapped there forever.

Similarly, if a set R is backward T-invariant, that is, if $T^{-1}(R) \subseteq R$, then an obvious candidate for a completely T-invariant set is the intersection of all the preimages of R, namely $\bigcap_{k=0}^{\infty} T^{-k}(R)$. This reduces our original question to the following: Given a set S, how can we construct from it a backward T-invariant set R? The union of all the preimages of S, namely $R = \bigcup_{n=0}^{\infty} T^{-n}(S)$, is such a set. Hence, the set

$$\bigcap_{k=0}^{\infty} T^{-k}\left(\bigcup_{n=0}^{\infty} T^{-n}(S)\right) = \bigcap_{k=0}^{\infty}\bigcup_{n=0}^{\infty} T^{-(k+n)}(S) = \bigcap_{k=0}^{\infty}\bigcup_{n=k}^{\infty} T^{-n}(S)$$

is completely T-invariant. This is the set of all points whose orbits visit S infinitely many times.

Proposition 8.2.4. *Let $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ be a measure-preserving dynamical system. Then T is ergodic with respect to μ if and only if all μ-almost T-invariant sets $A \in \mathcal{A}$ satisfy $\mu(A) = 0$ or $\mu(X\backslash A) = 0$.*

Proof. Suppose that all μ-almost T-invariant sets $A \in \mathcal{A}$ satisfy $\mu(A) = 0$ or $\mu(X\backslash A) = 0$. Let $A' \in \mathcal{A}$ be any completely T-invariant set. Since every completely T-invariant set is μ-almost T-invariant, it ensues that $\mu(A') = 0$ or $\mu(X\backslash A) = 0$. Thus T is ergodic.

We shall now prove the converse implication. Though μ is T-invariant, it is sufficient that $\mu \circ T^{-1} \ll \mu$ in the following proof (μ is then said to be quasi-T-invariant; see Definition 10.1.1). Suppose that T is ergodic and let $A \in \mathcal{A}$ be a μ-almost T-invariant set, that is, A is such that $\mu(T^{-1}(A) \triangle A) = 0$. We must show that $\mu(A) = 0$ or $\mu(X\backslash A) = 0$.

Claim 1. $\mu(T^{-n}(A) \triangle A) = 0$ *for all* $n \geq 0$.

Proof of Claim 1. Since $\mu(T^{-1}(A) \triangle A) = 0$, since μ is T-invariant and since $f^{-1}(C \triangle D) = f^{-1}(C) \triangle f^{-1}(D)$ for any map f and any sets C and D, we have for all $k \in \mathbb{N}$ that

$$\mu\big(T^{-(k+1)}(A) \triangle T^{-k}(A)\big) = \mu\big(T^{-k}\big(T^{-1}(A) \triangle A\big)\big) = \mu\big(T^{-1}(A) \triangle A\big) = 0.$$

As $C \triangle D \subseteq (C \triangle E) \cup (E \triangle D)$ for any sets C, D and E, it follows for all $n \in \mathbb{N}$ that

$$\mu(T^{-n}(A) \triangle A) \leq \mu\left(\bigcup_{k=0}^{n-1}[T^{-(k+1)}(A) \triangle T^{-k}(A)]\right)$$
$$\leq \sum_{k=0}^{n-1} \mu(T^{-(k+1)}(A) \triangle T^{-k}(A)) = 0. \qquad \square$$

Claim 2. *The set*

$$B := \bigcup_{k=0}^{\infty}\bigcap_{n=k}^{\infty} T^{-n}(A)$$

is completely T-invariant, and thus $\mu(B) = 0$ or $\mu(X\backslash B) = 0$.

Proof of Claim 2. Indeed,

$$T^{-1}(B) = \bigcup_{k=0}^{\infty}\bigcap_{n=k}^{\infty} T^{-(n+1)}(A) = \bigcup_{k=0}^{\infty}\bigcap_{n=k+1}^{\infty} T^{-n}(A) = \bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty} T^{-n}(A) = B.$$

Since T is ergodic, we deduce that $\mu(B) = 0$ or $\mu(X\backslash B) = 0$. $\square$

Claim 3. $\mu(B \triangle A) = 0 = \mu((X\backslash B) \triangle (X\backslash A))$.

Proof of Claim 3. To prove this, we will use Claim 1 and two properties of the symmetric difference operation: $(\bigcup_{i\in I} C_i) \triangle D \subseteq \bigcup_{i\in I}(C_i \triangle D)$ and $(\bigcap_{i\in I} C_i) \triangle D \subseteq \bigcup_{i\in I}(C_i \triangle D)$. Indeed,

$$\begin{aligned}
\mu(B \triangle A) &= \mu\left(\left[\bigcup_{k=0}^{\infty}\bigcap_{n=k}^{\infty} T^{-n}(A)\right] \triangle A\right) \\
&\leq \mu\left(\bigcup_{k=0}^{\infty}\left[\left(\bigcap_{n=k}^{\infty} T^{-n}(A)\right) \triangle A\right]\right) \\
&\leq \mu\left(\bigcup_{k=0}^{\infty}\bigcup_{n=k}^{\infty}(T^{-n}(A) \triangle A)\right) \\
&= \mu\left(\bigcup_{n=0}^{\infty}(T^{-n}(A) \triangle A)\right) \\
&\leq \sum_{n=0}^{\infty} \mu(T^{-n}(A) \triangle A) \\
&= 0.
\end{aligned}$$

Since $(X\backslash B) \triangle (X\backslash A) = B \triangle A$, Claim 3 is proved. $\square$

Claim 4. $\mu(A) = \mu(B)$ *and* $\mu(X\backslash A) = \mu(X\backslash B)$.

Proof of Claim 4. Since $B \triangle A = (B\backslash A) \cup (A\backslash B)$, it immediately follows from Claim 3 that $\mu(B\backslash A) = 0 = \mu(A\backslash B)$. Therefore,

$$\mu(B) = \mu\big((B\backslash A) \cup (B \cap A)\big) = \mu(B\backslash A) + \mu(B \cap A) = \mu(B \cap A) \le \mu(A).$$

Likewise, $\mu(A) \le \mu(B)$. Thus $\mu(A) = \mu(B)$. Similarly, $\mu(X\backslash A) = \mu(X\backslash B)$. □

To conclude the proof of the proposition, Claims 2 and 4 assert that $\mu(A) = \mu(B) = 0$ or $\mu(X\backslash A) = \mu(X\backslash B) = 0$. Finally, note that we could equally well have chosen the set $B = \bigcap_{k=0}^{\infty} \bigcup_{n=k}^{\infty} T^{-n}(A)$ in Claim 2. □

It is not difficult to check that the family $\{A \in \mathcal{A} \mid T^{-1}(A) = A\}$ of all completely T-invariant sets forms a sub-σ-algebra of $\mathcal{A}$. So does the family $\{A \mid \mu(T^{-1}(A) \triangle A) = 0\}$ of all μ-almost T-invariant sets, a fact which we shall prove shortly. Of course, the former is a smaller σ-algebra. However, it is not sufficiently flexible for our measure-theoretic purposes, as it is only defined set-theoretically. For this reason, we shall usually work with the latter.

Definition 8.2.5. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and let μ be a measure on $(X, \mathcal{A})$. The collection of all μ-almost T-invariant sets shall be denoted by

$$\mathcal{I}_\mu := \{A \in \mathcal{A} \,|\, \mu(T^{-1}(A) \triangle A) = 0\}.$$

Proposition 8.2.6. *The family $\mathcal{I}_\mu$ is a sub-σ-algebra of $\mathcal{A}$. Furthermore, if μ is T-invariant then $T^{-1}(\mathcal{I}_\mu) \subseteq \mathcal{I}_\mu$.*

Proof. It is clear that $\emptyset \in \mathcal{I}_\mu$.

We now show that $\mathcal{I}_\mu$ is closed under the operation of complementation. Let $A \in \mathcal{I}_\mu$. Then

$$\begin{aligned} T^{-1}(X\backslash A) \triangle (X\backslash A) &= \big(X\backslash T^{-1}(A)\big) \triangle (X\backslash A) \\ &= \big((X\backslash T^{-1}(A))\backslash(X\backslash A)\big) \cup \big((X\backslash A)\backslash(X\backslash T^{-1}(A))\big) \\ &= \big(A\backslash T^{-1}(A)\big) \cup \big(T^{-1}(A)\backslash A\big) \\ &= T^{-1}(A) \triangle A. \end{aligned}$$

Thus $\mu(T^{-1}(X\backslash A) \triangle (X\backslash A)) = \mu(T^{-1}(A) \triangle A) = 0$, and hence $X\backslash A \in \mathcal{I}_\mu$.

It only remains to show that $\mathcal{I}_\mu$ is closed under countable unions, that is, we must show that if $(A_n)_{n=1}^{\infty}$ is a sequence in $\mathcal{I}_\mu$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{I}_\mu$. For this, observe that

$$\begin{aligned} \left[T^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right)\right] \triangle \left[\bigcup_{n=1}^{\infty} A_n\right] &= \left[\bigcup_{n=1}^{\infty} T^{-1}(A_n)\right] \triangle \left[\bigcup_{n=1}^{\infty} A_n\right] \\ &= \left[\bigcup_{n=1}^{\infty} T^{-1}(A_n) \setminus \bigcup_{n=1}^{\infty} A_n\right] \bigcup \left[\bigcup_{n=1}^{\infty} A_n \setminus \bigcup_{n=1}^{\infty} T^{-1}(A_n)\right] \end{aligned}$$

$$\subseteq \left[\bigcup_{n=1}^{\infty}(T^{-1}(A_n)\backslash A_n)\right] \bigcup \left[\bigcup_{n=1}^{\infty}(A_n\backslash T^{-1}(A_n))\right]$$
$$= \bigcup_{n=1}^{\infty}(T^{-1}(A_n) \bigtriangleup A_n).$$

Consequently,

$$\mu\left(\left[T^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right)\right] \bigtriangleup \left[\bigcup_{n=1}^{\infty} A_n\right]\right) \le \sum_{n=1}^{\infty} \mu(T^{-1}(A_n) \bigtriangleup A_n) = \sum_{n=1}^{\infty} 0 = 0.$$

So $\bigcup_{n=1}^{\infty} A_n \in \mathcal{I}_\mu$ and $\mathcal{I}_\mu$ is a σ-algebra.

If μ is T-invariant (in fact, it suffices that μ be quasi-T-invariant, i. e. $\mu \circ T^{-1} \ll \mu$) and if $A \in \mathcal{I}_\mu$, then

$$\mu(T^{-1}(T^{-1}(A)) \bigtriangleup T^{-1}(A)) = \mu(T^{-1}(T^{-1}(A) \bigtriangleup A)) = \mu(T^{-1}(A) \bigtriangleup A) = 0.$$

That is, $T^{-1}(A) \in \mathcal{I}_\mu$. Thus $T^{-1}(\mathcal{I}_\mu) \subseteq \mathcal{I}_\mu$. □

We have already discussed invariant sets, measure-theoretically invariant sets and invariant measures. We now introduce invariant and measure-theoretically invariant functions.

Definition 8.2.7. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation, let μ be a measure on $(X, \mathcal{A})$ and let $\varphi : (X, \mathcal{A}) \to \mathbb{R}$ be a measurable function.

(a) The function φ is said to be *T-invariant* if $\varphi \circ T = \varphi$.

(b) The function φ is called *μ-a. e. T-invariant* if $\varphi \circ T = \varphi$ μ-almost everywhere. In other words, φ is μ-a. e. T-invariant if the measurable set

$$D_\varphi := \{x \in X \mid \varphi(T(x)) \neq \varphi(x)\}$$

is a null set.

Lemma 8.2.8. *Let $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ be a measure-preserving dynamical system. A measurable function $\varphi : X \to \mathbb{R}$ is μ-a. e. T-invariant if and only if φ is constant on the forward orbit of μ-a. e. $x \in X$.*

Proof. If φ is constant on the forward orbit of μ-a. e. $x \in X$, then $\varphi(T(x)) = \varphi(x)$ for μ-a. e. $x \in X$, that is, φ is μ-a. e. T-invariant.

To prove the converse, suppose that φ is μ-a. e. T-invariant. For every $n \in \mathbb{N}$, we have

$$\{y \in X : \varphi(T^n(y)) \neq \varphi(T^{n-1}(y))\} = T^{-(n-1)}(\{x \in X : \varphi(T(x)) \neq \varphi(x)\}).$$

Since μ is T-invariant (in fact, it suffices that μ be quasi-T-invariant) and since φ is μ-a. e. T-invariant, this implies that for every $n \in \mathbb{N}$,

$$\mu(\{y \in X : \varphi(T^n(y)) \neq \varphi(T^{n-1}(y))\}) = 0.$$

Therefore,

$$\begin{aligned}\mu\big(\{x \in X : \varphi \text{ not constant over } \mathcal{O}_+(x)\}\big) &= \mu\bigg(\bigcup_{n=1}^{\infty}\{x \in X : \varphi(T^n(x)) \neq \varphi(T^{n-1}(x))\}\bigg)\\ &\leq \sum_{n=1}^{\infty}\mu\big(\{x \in X : \varphi(T^n(x)) \neq \varphi(T^{n-1}(x))\}\big)\\ &= 0.\end{aligned}$$

So φ is constant on the forward orbit of μ-a. e. $x \in X$. □

The following lemma shows that measure-theoretically invariant functions are characterized by the fact that they are measurable with respect to the σ-algebra of measure-theoretically invariant sets.

Lemma 8.2.9. *Let $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ be a measure-preserving dynamical system. A function $\varphi : X \to \mathbb{R}$ is μ-a. e. T-invariant if and only if φ is measurable with respect to the σ-algebra $\mathcal{I}_\mu$.*

Proof. First, suppose that φ is μ-a. e. T-invariant. Let $B \subseteq \mathbb{R}$ be a Borel set. In order for φ to be $\mathcal{I}_\mu$-measurable, we need to show that $\varphi^{-1}(B) \in \mathcal{I}_\mu$. To begin, notice that if

$$x \in T^{-1}(\varphi^{-1}(B)) \triangle (\varphi^{-1}(B)) = (\varphi \circ T)^{-1}(B) \triangle (\varphi^{-1}(B)),$$

then only one of the real numbers $\varphi(T(x))$ or $\varphi(x)$ belongs to the set B. Thus $\varphi(T(x)) \neq \varphi(x)$ and $x \in D_\varphi$. This means that

$$T^{-1}(\varphi^{-1}(B)) \triangle (\varphi^{-1}(B)) \subseteq D_\varphi.$$

Consequently, $\mu(T^{-1}(\varphi^{-1}(B)) \triangle (\varphi^{-1}(B))) \leq \mu(D_\varphi) = 0$. So $\varphi^{-1}(B) \in \mathcal{I}_\mu$.

To prove the converse implication, suppose by way of contradiction that φ is $\mathcal{I}_\mu$-measurable but that $\mu(D_\varphi) > 0$. We can always write

$$D_\varphi = \bigcup_{a \in \mathbb{Q}}\Big(\{x \in X : \varphi(x) < a < \varphi(T(x))\} \cup \{x \in X : \varphi(x) > a > \varphi(T(x))\}\Big).$$

This is a countable union of $\mathcal{A}$-measurable sets with positive total measure. Hence, without loss of generality, there exists some $a \in \mathbb{Q}$ such that the set

$$B_a := \{x \in X \mid \varphi(x) < a < \varphi(T(x))\}$$

is of positive measure (if not, replace φ by $-\varphi$). Observe that

$$B_a = \varphi^{-1}((-\infty, a)) \cap (\varphi \circ T)^{-1}((a, \infty)).$$

Note that $\varphi^{-1}((-\infty, a)) \in \mathcal{I}_\mu$ since φ is $\mathcal{I}_\mu$-measurable by assumption. Moreover, $(\varphi \circ T)^{-1}((a, \infty)) = T^{-1}(\varphi^{-1}((a, \infty))) \in \mathcal{I}_\mu$ since φ is $\mathcal{I}_\mu$-measurable and $T^{-1}(\mathcal{I}_\mu) \subseteq \mathcal{I}_\mu$ per

Proposition 8.2.6. Thus $B_a \in \mathcal{I}_\mu$. Now, notice that

$$T^{-1}(B_a) = \{x \in X : T(x) \in B_a\} \subseteq \{x \in X : \varphi(T(x)) < a\} \subseteq X \backslash B_a.$$

So $T^{-1}(B_a) \cap B_a = \emptyset$ and, as $B_a \in \mathcal{I}_\mu$, we deduce that

$$0 = \mu(T^{-1}(B_a) \triangle B_a) = \mu(T^{-1}(B_a) \cup B_a) \geq \mu(B_a).$$

Consequently, $\mu(B_a) = 0$ and we have reached a contradiction. Therefore, $\mu(D_\varphi) = 0$ and hence φ is μ-a. e. T-invariant. □

In other terms, Lemma 8.2.9 asserts that φ is μ-a. e. T-invariant if and only if $E(\varphi|\mathcal{I}_\mu) = \varphi$, where $E(\varphi|\mathcal{I}_\mu)$ is the conditional expectation of φ with respect to μ.

This conditional expectation function is an intrinsic part of the most important result in ergodic theory: Birkhoff's ergodic theorem. For more information on this function, see Subsection A.1.9.

8.2.1 Birkhoff's ergodic theorem

We are almost ready to state the main result of this chapter. Before doing so, we must introduce one more notation and terminology.

Definition 8.2.10. Let $T : X \to X$ be a map and let $\varphi : X \to \mathbb{R}$ be a real-valued function. Let $n \in \mathbb{N}$. The nth *Birkhoff sum* of φ at a point $x \in X$ is defined to be

$$S_n\varphi(x) = \sum_{j=0}^{n-1} \varphi(T^j(x)).$$

In other words, $S_n\varphi(x)$ is the sum of the values of the function φ at the first n points in the orbit of x. Sometimes this is referred to as the nth *ergodic sum*. As we will see in a moment, it is also convenient to define $S_0\varphi(x) = 0$.

It is easy to see that the following recurrence formula holds:

$$S_n\varphi(x) = S_k\varphi(x) + S_{n-k}\varphi(T^k(x)), \quad \forall x \in X, \forall k, n \in \mathbb{N} \text{ with } k \leq n. \tag{8.2}$$

We now come to the most important result in ergodic theory. This theorem was originally proved by George David Birkhoff in 1931. There exists now a variety of proofs. The simple one given here originates from Katok and Hasselblatt [33].

Theorem 8.2.11 (Birkhoff's ergodic theorem). *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. If $\varphi \in L^1(X, \mathcal{A}, \mu)$, then*

$$\lim_{n\to\infty} \left\| \frac{1}{n} S_n\varphi - E(\varphi|\mathcal{I}_\mu) \right\|_1 = 0 \quad \textit{and} \quad \lim_{n\to\infty} \frac{1}{n} S_n\varphi(x) = E(\varphi|\mathcal{I}_\mu)(x) \quad \textit{for } \mu\textit{-a. e. } x \in X.$$

Proof. For the μ-a. e. pointwise convergence, it suffices to prove that

$$\limsup_{n\to\infty} \frac{1}{n} S_n\varphi(x) \leq E(\varphi|\mathcal{I}_\mu)(x) \quad \text{for } \mu\text{-a. e. } x \in X. \tag{8.3}$$

Indeed, replacing φ by $-\varphi$ in (8.3), it follows that for μ-a. e. $x \in X$,

$$\liminf_{n\to\infty} \frac{1}{n} S_n\varphi(x) = -\limsup_{n\to\infty} \frac{1}{n} S_n(-\varphi)(x) \geq -E(-\varphi|\mathcal{I}_\mu)(x) = E(\varphi|\mathcal{I}_\mu)(x). \tag{8.4}$$

If (8.3) and consequently (8.4) hold, we can conclude for μ-a. e. $x \in X$ that

$$E(\varphi|\mathcal{I}_\mu)(x) \leq \liminf_{n\to\infty} \frac{1}{n} S_n\varphi(x) \leq \limsup_{n\to\infty} \frac{1}{n} S_n\varphi(x) \leq E(\varphi|\mathcal{I}_\mu)(x),$$

and this would complete the proof. In order to prove (8.3), it is sufficient to show that for every $\varepsilon > 0$ we have

$$\limsup_{n\to\infty} \frac{1}{n} S_n\varphi(x) \leq E(\varphi|\mathcal{I}_\mu)(x) + \varepsilon \quad \text{for } \mu\text{-a. e. } x \in X. \tag{8.5}$$

Indeed, if for each $\varepsilon > 0$ relation (8.5) holds everywhere except on a set X_ε of measure zero, then relation (8.3) holds everywhere except on the set $\bigcup_{k=1}^{\infty} X_{1/k}$ and $\mu(\bigcup_{k=1}^{\infty} X_{1/k}) = 0$. So fix $\varepsilon > 0$. We claim that proving (8.5) is equivalent to showing that

$$\limsup_{n\to\infty} \frac{1}{n} S_n\varphi_\varepsilon(x) \leq 0 \quad \text{for } \mu\text{-a. e. } x \in X, \tag{8.6}$$

where

$$\varphi_\varepsilon := \varphi - E(\varphi|\mathcal{I}_\mu) - \varepsilon.$$

Indeed, since $E(\varphi|\mathcal{I}_\mu)$ is $\mathcal{I}_\mu$-measurable by definition, Lemma 8.2.9 implies that $E(\varphi|\mathcal{I}_\mu)$ is μ-a. e. T-invariant, that is, $E(\varphi|\mathcal{I}_\mu)\circ T(x) = E(\varphi|\mathcal{I}_\mu)(x)$ for μ-a. e. $x \in X$. It then follows that for μ-a. e. $x \in X$,

$$\begin{aligned}
\frac{1}{n} S_n\varphi_\varepsilon(x) &= \frac{1}{n} S_n\varphi(x) - \frac{1}{n} S_n E(\varphi|\mathcal{I}_\mu)(x) - \frac{1}{n} S_n\varepsilon(x) \\
&= \frac{1}{n} S_n\varphi(x) - \frac{1}{n} \sum_{j=0}^{n-1} E(\varphi|\mathcal{I}_\mu) \circ T^j(x) - \frac{1}{n} \sum_{j=0}^{n-1} \varepsilon \circ T^j(x) \\
&= \frac{1}{n} S_n\varphi(x) - E(\varphi|\mathcal{I}_\mu)(x) - \varepsilon.
\end{aligned}$$

Thus

$$\limsup_{n\to\infty} \frac{1}{n} S_n\varphi(x) = E(\varphi|\mathcal{I}_\mu)(x) + \varepsilon + \limsup_{n\to\infty} \frac{1}{n} S_n\varphi_\varepsilon(x)$$

for μ-a. e. $x \in X$. However, in order to prove (8.6), it suffices to show that

$$\mu(A_\varepsilon) = 0, \tag{8.7}$$

where

$$A_\varepsilon := \left\{x \in X \mid \sup_{n\in\mathbb{N}} S_n\varphi_\varepsilon(x) = \infty\right\},$$

since for any $x \notin A_\varepsilon$ we have that $\sup_{n\in\mathbb{N}} S_n\varphi_\varepsilon(x) < \infty$ and it follows that $\limsup_{n\to\infty} \frac{1}{n} S_n\varphi_\varepsilon(x) \le 0$. Now we make a critical observation:

$$E(\varphi_\varepsilon|\mathcal{I}_\mu) = E(\varphi|\mathcal{I}_\mu) - E\big(E(\varphi|\mathcal{I}_\mu)|\mathcal{I}_\mu\big) - E(\varepsilon|\mathcal{I}_\mu) = E(\varphi|\mathcal{I}_\mu) - E(\varphi|\mathcal{I}_\mu) - \varepsilon = -\varepsilon < 0.$$

Rather than restricting our attention to φ_ε, we will prove that (8.7) holds for all $f \in L^1(X, \mathcal{A}, \mu)$ such that $E(f|\mathcal{I}_\mu) < 0$, with φ_ε being one such f. The T-invariance of μ implies that $f \circ T^k \in L^1(X, \mathcal{A}, \mu)$ for all $k \in \mathbb{N}$. It immediately follows that $S_k f \in L^1(X, \mathcal{A}, \mu)$ for all $k \in \mathbb{N}$. For each $n \in \mathbb{N}$ and each $x \in X$ define

$$M_n f(x) = \max_{1\le k\le n} S_k f(x).$$

It is easy to deduce that $M_n f \in L^1(X, \mathcal{A}, \mu)$ for all $n \in \mathbb{N}$. It is also obvious that the sequence $(M_n f(x))_{n=1}^\infty$ is nondecreasing for all $x \in X$. Moreover, the recurrence formula (8.2) between successive ergodic sums $(S_n\varphi(x) = \varphi(x) + S_{n-1}\varphi(T(x)))$ suggests the existence of a recurrence formula for their successive maxima. Indeed, for all $x \in X$,

$$\begin{aligned}
M_{n+1} f(x) &= \max_{1\le k\le n+1} S_k f(x) \\
&= \max_{1\le k\le n+1} \big[f(x) + S_{k-1} f\big(T(x)\big)\big] \\
&= f(x) + \max_{0\le l\le n} S_l f\big(T(x)\big) \\
&= f(x) + \max\Big\{0, \max_{1\le l\le n} S_l f\big(T(x)\big)\Big\} \\
&= f(x) + \max\big\{0, M_n f\big(T(x)\big)\big\}.
\end{aligned}$$

Therefore, for all $x \in X$,

$$M_{n+1} f(x) - M_n f\big(T(x)\big) = f(x) + \max\big\{-M_n f\big(T(x)\big), 0\big\}. \tag{8.8}$$

Since the sequence $(M_n f(T(x)))_{n=1}^\infty$ is nondecreasing for all $x \in X$, the sequence $(\max\{-M_n f(T(x)), 0\})_{n=1}^\infty$ is nonincreasing for all $x \in X$. By (8.8), the sequence $(M_{n+1} f(x) - M_n f(T(x)))_{n=1}^\infty$ is therefore nonincreasing for all $x \in X$. In order to prove (8.7) for the function f, we will investigate the limit of this latter sequence on the set

$$A = \Big\{x \in X : \sup_{n\in\mathbb{N}} S_n f(x) = \infty\Big\} = \Big\{x \in X : \lim_{n\to\infty} M_n f(x) = \infty\Big\}.$$

Using the recurrence formula (8.2), it is easy to see that $T^{-1}(A) = A$. In particular, this implies that $A \in \mathcal{I}_\mu$. Also, if $x \in A$ then $T(x) \in A$, and thus $\lim_{n\to\infty} M_n f(T(x)) = \infty$.

According to (8.8), it ensues that

$$\lim_{n\to\infty}\big(M_{n+1}f(x) - M_nf(T(x))\big) = f(x), \quad \forall x \in A. \tag{8.9}$$

Knowing the pointwise limit of this sequence on A, we further show that this non-increasing sequence is uniformly bounded by an integrable function. For all $n \in \mathbb{N}$ and $x \in X$, we have

$$f(x) \le M_{n+1}f(x) - M_nf(T(x)) \le M_2f(x) - M_1f(T(x)) = f(x) + \max\{-f(T(x)), 0\}.$$

For all $n \in \mathbb{N}$ and $x \in X$, it follows that

$$|M_{n+1}f(x) - M_nf(T(x))| \le |f(x)| + |f(T(x))|.$$

Since $|f| + |f \circ T| \in L^1(X, \mathcal{A}, \mu)$, Lebesgue's dominated convergence theorem (Theorem A.1.38) applies. We deduce from the facts that the sequence $(M_nf)_{n=1}^{\infty}$ is non-decreasing, that μ is T-invariant and that (8.9) holds on A, that

$$\begin{aligned}
0 &\le \int_A (M_{n+1}f - M_nf)\,d\mu \\
&= \int_A M_{n+1}f\,d\mu - \int_A M_nf\,d\mu \\
&= \int_A M_{n+1}f\,d\mu - \int_A M_nf\,d(\mu\circ T^{-1}) \\
&= \int_A M_{n+1}f\,d\mu - \int_A M_nf\circ T\,d\mu \\
&= \int_A \big(M_{n+1}f(x) - M_nf(T(x))\big)\,d\mu(x) \to \int_A f(x)\,d\mu(x).
\end{aligned}$$

Hence, $\int_A f\,d\mu \ge 0$. Recall that $A \in \mathcal{I}_\mu$ and $E(f|\mathcal{I}_\mu) < 0$. If it were the case that $\mu(A) > 0$, it would follow from the definition of $E(f|\mathcal{I}_\mu)$ that

$$0 \le \int_A f\,d\mu = \int_A E(f|\mathcal{I}_\mu)\,d\mu < 0,$$

which would result in a contradiction. Thus $\mu(A) = 0$. Setting $f = \varphi_\varepsilon$, we conclude that $\mu(A_\varepsilon) = 0$, hence establishing (8.7) and the μ-a. e. pointwise convergence of the sequence $(\frac{1}{n}S_n\varphi)_{n=1}^{\infty}$ to $E(\varphi|\mathcal{I}_\mu)$.

Now, if φ is bounded then

$$\left\|\frac{1}{n}S_n\varphi\right\|_\infty \le \|\varphi\|_\infty, \quad \forall n \in \mathbb{N},$$

and thus Lebesgue's dominated convergence theorem (Theorem A.1.38) asserts that the sequence $(\frac{1}{n}S_n\varphi)_{n=1}^{\infty}$ converges in $L^1(\mu)$ to $E(\varphi|\mathcal{I}_\mu)$.

In general, since the set of bounded measurable functions is dense in $L^1(\mu)$, for every $\varepsilon > 0$ there exists a bounded measurable function $\varphi_\varepsilon : X \to \mathbb{R}$ such that

$$\|\varphi - \varphi_\varepsilon\|_1 < \frac{\varepsilon}{3}.$$

By the already proven part, there then exists $N_\varepsilon \in \mathbb{N}$ such that

$$\left\|\frac{1}{n}S_n\varphi_\varepsilon - E(\varphi_\varepsilon|\mathcal{I}_\mu)\right\|_1 < \frac{\varepsilon}{3}, \quad \forall n \geq N_\varepsilon.$$

For all $n \geq N_\varepsilon$, we then deduce that

$$\begin{aligned}\left\|\frac{1}{n}S_n\varphi - E(\varphi|\mathcal{I}_\mu)\right\|_1 &\leq \left\|\frac{1}{n}S_n(\varphi - \varphi_\varepsilon)\right\|_1 + \left\|\frac{1}{n}S_n\varphi_\varepsilon - E(\varphi_\varepsilon|\mathcal{I}_\mu)\right\|_1 \\ &\quad + \left\|E(\varphi_\varepsilon|\mathcal{I}_\mu) - E(\varphi|\mathcal{I}_\mu)\right\|_1 \\ &\leq \frac{1}{n}\sum_{j=0}^{n-1}\left\|(\varphi - \varphi_\varepsilon)\circ T^j\right\|_1 + \frac{\varepsilon}{3} + \left\|E(|\varphi_\varepsilon - \varphi|\,\big|\,\mathcal{I}_\mu)\right\|_1 \\ &= \frac{1}{n}\sum_{j=0}^{n-1}\|\varphi - \varphi_\varepsilon\|_1 + \frac{\varepsilon}{3} + \|\varphi_\varepsilon - \varphi\|_1 < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.\end{aligned}$$

So $\lim_{n\to\infty}\|\frac{1}{n}S_n\varphi - E(\varphi|\mathcal{I}_\mu)]\|_1 = 0$. □

Let $p \geq 1$. The set $L^p(X, \mathcal{A}, \mu)$ is the set of $\mathcal{A}$-measurable functions $\varphi : X \to \mathbb{R}$ such that $\varphi^p \in L^1(X, \mathcal{A}, \mu)$. It is well known that $L^p(X, \mathcal{A}, \mu) \subseteq L^1(X, \mathcal{A}, \mu)$. Theorem 8.2.11, along with the last part of its proof where one would use this time the density of bounded measurable functions in L^p, yield the following slight generalization.

Theorem 8.2.12 (Birkhoff's ergodic theorem in L^p). *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. If $\varphi \in L^p(X, \mathcal{A}, \mu)$, then*

$$\lim_{n\to\infty}\left\|\frac{1}{n}S_n\varphi - E(\varphi|\mathcal{I}_\mu)\right\|_p = 0 \quad \textit{and} \quad \lim_{n\to\infty}\frac{1}{n}S_n\varphi(x) = E(\varphi|\mathcal{I}_\mu)(x) \quad \textit{for } \mu\textit{-a.e. } x \in X.$$

Remark 8.2.13. If $p = 2$, then the L^p part of Theorem 8.2.12 is commonly referred to as von Neumann's ergodic theorem, proved for the first time in [52].

If a measure-preserving dynamical system on a probability space is ergodic, then Birkhoff's ergodic theorem implies the following.

Corollary 8.2.14 (Ergodic case of Birkhoff's ergodic theorem). *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. If T is ergodic with respect to μ and $\varphi \in L^1(X, \mathcal{A}, \mu)$, then $E(\varphi|\mathcal{I}_\mu) = \int_X \varphi\, d\mu$ and*

$$\lim_{n\to\infty}\left\|\frac{1}{n}S_n\varphi - \int_X \varphi\, d\mu\right\|_1 = 0 \quad \textit{and} \quad \lim_{n\to\infty}\frac{1}{n}S_n\varphi(x) = \int_X \varphi\, d\mu \quad \textit{for } \mu\textit{-a.e. } x \in X.$$

Proof. According to Example A.1.61, the collection of sets

$$\mathcal{N} := \{A \in \mathcal{A} : \mu(A) = 0 \text{ or } \mu(A) = 1\}$$

is a σ-algebra and $E(\varphi|\mathcal{N}) = \int_X \varphi\, d\mu$. As T is ergodic with respect to μ, we have that $\mathcal{I}_\mu \subseteq \mathcal{N}$. By Proposition A.1.60(f,e), it ensues that

$$E(\varphi|\mathcal{I}_\mu) = E\big(E(\varphi|\mathcal{N})\,\big|\,\mathcal{I}_\mu\big) = E\left(\int_X \varphi\, d\mu \,\middle|\, \mathcal{I}_\mu\right) = \int_X \varphi\, d\mu.$$

The result follows from Birkhoff's ergodic theorem. □

When T is ergodic with respect to an invariant probability measure μ, Birkhoff's ergodic theorem asserts that the average of φ along the forward orbit of μ-almost every $x \in X$ is asymptotically equal to the average of φ over the entire space. In other words, for any "typical" point the "time average" of a μ-integrable function is equal to its "space average."

In particular, if φ is the characteristic function of a measurable set, Corollary 8.2.14 guarantees the following.

Corollary 8.2.15. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. If T is ergodic with respect to μ, then for every $A \in \mathcal{A}$,*

$$\lim_{n\to\infty} \frac{1}{n} \#\{0 \le j < n : T^j(x) \in A\} = \mu(A) \quad \textit{for } \mu\text{-a.e. } x \in X.$$

In other terms, the average time the forward orbit of a "typical point" spends in a measurable set is asymptotically equal to the measure of that set. This provides more information than Poincaré's recurrence theorem (Theorem 8.1.16).

Birkhoff's ergodic theorem is a terrifically useful tool. It has had many applications in different areas of mathematics. In particular, it is very useful in number theory. In Exercises 8.5.42–8.5.43, you will use it to prove in a simple way various statements about real numbers whose original, nonergodic proofs were quite involved.

One intuitive understanding of ergodicity is that an ergodic system is one in which for every pair of measurable sets A and B, the sets $T^{-n}(A)$ become independent of B on average.

Lemma 8.2.16. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. Then T is ergodic with respect to μ if and only if*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu\big(T^{-j}(A) \cap B\big) = \mu(A)\mu(B), \quad \forall A, B \in \mathcal{A}. \tag{8.10}$$

Equivalently,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \big[\mu\big(T^{-j}(A) \cap B\big) - \mu(A)\mu(B)\big] = 0, \quad \forall A, B \in \mathcal{A}.$$

Proof. First, suppose that T is ergodic and let $A, B \in \mathcal{A}$. For every $n \in \mathbb{N}$, we have

$$\begin{aligned}\frac{1}{n}\sum_{j=0}^{n-1}\mu(T^{-j}(A)\cap B) &= \frac{1}{n}\sum_{j=0}^{n-1}\int_B \mathbb{1}_{T^{-j}(A)}\,d\mu \\ &= \frac{1}{n}\sum_{j=0}^{n-1}\int_B \mathbb{1}_A\circ T^j\,d\mu \\ &= \int_B \frac{1}{n}\sum_{j=0}^{n-1}\mathbb{1}_A\circ T^j\,d\mu \\ &= \int_B \frac{1}{n}\#\{0\le j<n : T^j(x)\in A\}\,d\mu(x).\end{aligned}$$

Passing to the limit and using Lebesgue's dominated convergence theorem (Theorem A.1.38) and Corollary 8.2.15, we get

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\mu(T^{-j}(A)\cap B) = \int_B \lim_{n\to\infty}\frac{1}{n}\#\{0\le j<n : T^j(x)\in A\}\,d\mu(x) = \mu(A)\mu(B).$$

For the converse implication, suppose that relation (8.10) holds true for all $A, B \in \mathcal{A}$. Let $E \in \mathcal{A}$ be a completely T-invariant set. Setting $A = B = E$ in (8.10), we obtain $\mu(E) = (\mu(E))^2$. So $\mu(E) \in \{0, 1\}$ and T is ergodic with respect to μ. □

To determine whether a measure-preserving dynamical system is ergodic, it suffices to check ergodicity on a semialgebra that generates the σ-algebra.

Lemma 8.2.17. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. If $\mathcal{B} = \sigma(\mathcal{S})$ for some semialgebra $\mathcal{S}$, then T is ergodic if and only if relation* (8.10) *holds for all $A, B \in \mathcal{S}$.*

Proof. If T is ergodic, then relation (8.10) holds for all $A, B \in \mathcal{B} \supseteq \mathcal{S}$.

For the converse implication, suppose that relation (8.10) holds for all $A, B \in \mathcal{S}$. Since each member of the algebra $\mathcal{A}(\mathcal{S})$ generated by $\mathcal{S}$ can be written as a finite disjoint union of elements of $\mathcal{S}$, a straightforward calculation shows that relation (8.10) also holds for all elements of $\mathcal{A}(\mathcal{S})$.

So, let $\varepsilon > 0$ and $A, B \in \mathcal{B} = \sigma(\mathcal{S}) = \sigma(\mathcal{A}(\mathcal{S}))$. By virtue of Lemma A.1.32, there are $A_0, B_0 \in \mathcal{A}(\mathcal{S})$ such that $\mu(A \triangle A_0) < \varepsilon$ and $\mu(B \triangle B_0) < \varepsilon$. By Exercise 8.5.11, it follows that

$$|\mu(A) - \mu(A_0)| < \varepsilon \quad \text{and} \quad |\mu(B) - \mu(B_0)| < \varepsilon. \tag{8.11}$$

Using Exercise 8.5.10, notice also that for every $j \ge 0$,

$$\begin{aligned}(T^{-j}(A)\cap B)\triangle(T^{-j}(A_0)\cap B_0) &\subseteq (T^{-j}(A)\triangle T^{-j}(A_0))\cup(B\triangle B_0) \\ &= T^{-j}(A\triangle A_0)\cup(B\triangle B_0).\end{aligned}$$

Therefore,

$$\mu\big((T^{-j}(A) \cap B) \triangle (T^{-j}(A_0) \cap B_0)\big) \le \mu(A \triangle A_0) + \mu(B \triangle B_0) < 2\varepsilon, \quad \forall j \ge 0.$$

By Exercise 8.5.11 again, we deduce that

$$\big|\mu\big(T^{-j}(A) \cap B\big) - \mu\big(T^{-j}(A_0) \cap B_0\big)\big| < 2\varepsilon, \quad \forall j \ge 0. \tag{8.12}$$

Using (8.12) and (8.11), we obtain for all $j \ge 0$ that

$$\begin{aligned}
\mu\big(T^{-j}(A) \cap B\big) - \mu(A)\mu(B) &\le \big[\mu\big(T^{-j}(A) \cap B\big) - \mu\big(T^{-j}(A_0) \cap B_0\big)\big] \\
&\quad + \big[\mu\big(T^{-j}(A_0) \cap B_0\big) - \mu(A_0)\mu(B_0)\big] \\
&\quad + \big[\mu(A_0)\mu(B_0) - \mu(A)\mu(B_0)\big] \\
&\quad + \big[\mu(A)\mu(B_0) - \mu(A)\mu(B)\big] \\
&\le \big|\mu\big(T^{-j}(A) \cap B\big) - \mu\big(T^{-j}(A_0) \cap B_0\big)\big| \\
&\quad + \big[\mu\big(T^{-j}(A_0) \cap B_0\big) - \mu(A_0)\mu(B_0)\big] \\
&\quad + \big|\mu(A_0) - \mu(A)\big|\mu(B_0) \\
&\quad + \mu(A)\big|\mu(B_0) - \mu(B)\big| \\
&< 4\varepsilon + \big[\mu\big(T^{-j}(A_0) \cap B_0\big) - \mu(A_0)\mu(B_0)\big]
\end{aligned}$$

and it follows that

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \big[\mu\big(T^{-j}(A) \cap B\big) - \mu(A)\mu(B)\big] \\
&\qquad \le 4\varepsilon + \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \big[\mu\big(T^{-j}(A_0) \cap B_0\big) - \mu(A_0)\mu(B_0)\big] = 4\varepsilon,
\end{aligned} \tag{8.13}$$

where the above limit is 0 as a consequence of (8.10) holding for $A_0, B_0 \in \mathcal{A}(S)$.

Similarly, for all $j \ge 0$,

$$\mu(A)\mu(B) - \mu\big(T^{-j}(A) \cap B\big) < 4\varepsilon + \big[\mu(A_0)\mu(B_0) - \mu\big(T^{-j}(A_0) \cap B_0\big)\big]$$

and it ensues that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \big[\mu(A)\mu(B) - \mu\big(T^{-j}(A) \cap B\big)\big] \le 4\varepsilon.$$

Therefore,

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \big[\mu\big(T^{-j}(A) \cap B\big) - \mu(A)\mu(B)\big] \ge -4\varepsilon. \tag{8.14}$$

By (8.13) and (8.14), the limsup is at most 4ε while the liminf is at least -4ε. Since $\varepsilon > 0$ was chosen arbitrarily, we deduce that the limit exists and is 0. Thus, relation (8.10) holds for all elements of $\mathcal{B}$, and T is ergodic by Lemma 8.2.16. □

We end this section with a characterization of ergodicity in terms of invariant and measure-theoretically invariant functions.

Theorem 8.2.18. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. The following statements are equivalent:*
(a) *T is ergodic with respect to μ.*
(b) *If φ is a T-invariant $L^1(X, \mathcal{A}, \mu)$-function, then φ is μ-a. e. constant.*
(c) *If φ is a μ-a. e. T-invariant $L^1(X, \mathcal{A}, \mu)$-function, then φ is μ-a. e. constant.*
(d) *If φ is a T-invariant measurable function, then φ is μ-a. e. constant.*
(e) *If φ is a μ-a. e. T-invariant measurable function, then φ is μ-a. e. constant.*

Proof. We shall first prove the chain of implications (a)⇒(c)⇒(b)⇒(a) and then show that (e)⇒(d)⇒(a)⇒(e).

To begin, suppose that T is ergodic and let φ be a μ-a. e. T-invariant $L^1(X, \mathcal{A}, \mu)$-function. As φ is μ-a. e. T-invariant and μ is T-invariant, it follows from Lemma 8.2.8 that φ is constant over the forward orbit of μ-a. e. $x \in X$. This implies that $S_n\varphi(x) = n\varphi(x)$ for all $n \in \mathbb{N}$ for μ-a. e. $x \in X$. Using this and the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14), it follows that

$$\varphi(x) = \lim_{n\to\infty} \frac{1}{n} S_n\varphi(x) = \int_X \varphi \, d\mu \quad \text{for } \mu\text{-a. e. } x \in X.$$

So φ is constant μ-almost everywhere. This proves that (a)⇒(c).

Since every T-invariant function is μ-a. e. T-invariant, it is clear that (c)⇒(b).

We now want to show that (b)⇒(a). Suppose that every T-invariant $L^1(X, \mathcal{A}, \mu)$-function is constant μ-a. e. and assume by way of contradiction that T is not ergodic with respect to μ. Then there exists a set $A \in \mathcal{A}$ such that $T^{-1}(A) = A$ with $\mu(A) > 0$ and $\mu(X\backslash A) > 0$. Since $T^{-1}(A) = A$, we have that $\mathbb{1}_A \circ T = \mathbb{1}_A$. However, $\mathbb{1}_A$ is not constant μ-a. e. since $\mu(A) > 0$ and $\mu(X\backslash A) > 0$. Thus, $\mathbb{1}_A$ is a T-invariant $L^1(X, \mathcal{A}, \mu)$-function which is not constant μ-almost everywhere. This contradiction shows that T must be ergodic.

This completes the proof of the first chain (a)⇒(c)⇒(b)⇒(a).

Accordingly, let us turn our attention to the second chain. It is clear that (e)⇒(d). The above proof that (b)⇒(a) carries over directly to show that (d)⇒(a) by simply replacing "$L^1(X, \mathcal{A}, \mu)$" by "measurable." All that is left is to establish that (a)⇒(e). So, suppose that T is ergodic and let φ be a μ-a. e. T-invariant measurable function. Then $\mu(\{x \in X \mid \varphi(T(x)) \neq \varphi(x)\}) = 0$. Assume by way of contradiction that φ is not μ-a. e. constant. Then there exists $r \in \mathbb{R}$ such that $\mu(S_r) > 0$ and $\mu(X\backslash S_r) > 0$, where $S_r := \{x \mid \varphi(x) < r\}$. The set S_r is μ-a. e. T-invariant since

$$\begin{aligned} T^{-1}(S_r) \triangle S_r &= (T^{-1}(S_r)\backslash S_r) \cup (S_r\backslash T^{-1}(S_r)) \\ &= \{x : \varphi(T(x)) < r \leq \varphi(x)\} \cup \{x : \varphi(x) < r \leq \varphi(T(x))\} \\ &\subseteq \{x : \varphi(T(x)) \neq \varphi(x)\} \end{aligned}$$

and hence

$$\mu(T^{-1}(S_r) \triangle S_r) \le \mu(\{x \in X : \varphi(T(x)) \neq \varphi(x)\}) = 0.$$

Thus $\mu(T^{-1}(S_r) \triangle S_r) = 0$. Since μ is ergodic, we deduce that either $\mu(S_r) = 0$ or $\mu(X\backslash S_r) = 0$. This contradiction implies that φ must be μ-a. e. constant. □

Remark 8.2.19. It is possible to prove that the property "$L^1(X, \mathcal{A}, \mu)$" can be replaced by "$L^p(X, \mathcal{A}, \mu)$" for any $1 \le p < \infty$.

8.2.2 Existence of ergodic measures

We will shortly embark on a study of the set of all ergodic measures for a given measurable transformation.

Definition 8.2.20. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. The set of all T-invariant probability measures that are ergodic with respect to T is denoted by $E(T, \mathcal{A})$. If $\mathcal{A}$ is the Borel σ-algebra on a topological space X, in line with the notation for the T-invariant Borel probability measures, we simply write $E(T) := E(T, \mathcal{B}(X))$.

We saw in Lemma 8.1.19 that the set of invariant probability measures $M(T, \mathcal{A})$ is convex. We shall soon prove that the ergodic measures $E(T, \mathcal{A})$ form the extreme points of $M(T, \mathcal{A})$. First, we show that any two ergodic measures are either equal or mutually singular (for more information on mutual singularity, see Subsection A.1.7).

Theorem 8.2.21. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. If $\mu_1, \mu_2 \in E(T, \mathcal{A})$ and $\mu_1 \neq \mu_2$, then $\mu_1 \perp \mu_2$.*

Proof. Since $\mu_1 \neq \mu_2$, there exists some set $A \in \mathcal{A}$ with $\mu_1(A) \neq \mu_2(A)$. By Corollary 8.2.15 of Birkhoff's ergodic theorem, for each $i = 1, 2$ there exists a set X_i of full μ_i-measure such that

$$\lim_{n\to\infty} \frac{1}{n} \#\{0 \le j < n : T^j(x) \in A\} = \mu_i(A), \quad \forall x \in X_i.$$

Consequently, $\mu_1(A) = \mu_2(A)$ on $X_1 \cap X_2$. As we know that $\mu_1(A) \neq \mu_2(A)$, we deduce that $X_1 \cap X_2 = \emptyset$. Thus $\mu_1(X_1) = 1$, $\mu_2(X_2) = 1$ and $X_1 \cap X_2 = \emptyset$. Therefore, $\mu_1 \perp \mu_2$. □

We use the above theorem to give a characterization of ergodic measures as those invariant probability measures with respect to which no other invariant probability measure is absolutely continuous.

Theorem 8.2.22. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and let $\mu \in M(T, \mathcal{A})$. Then $\mu \in E(T, \mathcal{A})$ if and only if there is no $\nu \in M(T, \mathcal{A})$ such that $\nu \ll \mu$ and $\nu \neq \mu$.*

Proof. First, suppose that $\mu \in E(T, \mathcal{A})$. Let $\nu \in M(T, \mathcal{A})$ be such that $\nu \ll \mu$. We claim that $\nu \in E(T, \mathcal{A})$, too. Indeed, suppose by way of contradiction that there exists $A \in \mathcal{A}$ such that $T^{-1}(A) = A$ with $\nu(A) > 0$ and $\nu(X \backslash A) > 0$. Since $\nu \ll \mu$, it follows that $\mu(A) > 0$ and $\mu(X \backslash A) > 0$. This contradicts the ergodicity of μ. So $\nu \in E(T, \mathcal{A})$. Now, if $\nu \neq \mu$ then Theorem 8.2.21 affirms that $\nu \perp \mu$. This contradicts the hypothesis that $\nu \ll \mu$. Hence, $\nu = \mu$.

For the converse implication, suppose that μ is not ergodic (but still T-invariant by hypothesis). Then there exists some $A \in \mathcal{A}$ such that $T^{-1}(A) = A$ with $\mu(A) > 0$ and $\mu(X \backslash A) > 0$. Let μ_A be the conditional measure of μ on A, as expressed in Definition A.1.70. Then one immediately verifies that μ_A is a T-invariant probability measure such that $\mu_A \neq \mu$ and $\mu_A \ll \mu$. □

Recall that in a vector space the extreme points of a convex set are those points which cannot be represented as a nontrivial convex combination of two distinct points of the set. In concrete terms, let V be a vector space and C be a convex subset of V. A vector $v \in C$ is an *extreme point* of C if the only combination of distinct vectors $v_1, v_2 \in C$ such that $v = \alpha v_1 + (1 - \alpha) v_2$ for some $\alpha \in [0, 1]$ is a combination with $\alpha = 0$ or $\alpha = 1$.

Theorem 8.2.23. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. The ergodic measures $E(T, \mathcal{A})$ are the extreme points of the set of invariant probability measures $M(T, \mathcal{A})$.*

Proof. Suppose that $\mu \in E(T, \mathcal{A})$ is not an extreme point of $M(T, \mathcal{A})$. Then there exist measures $\mu_1 \neq \mu_2$ in $M(T, \mathcal{A})$ and $0 < \alpha < 1$ such that $\mu = \alpha\mu_1 + (1 - \alpha)\mu_2$. It follows immediately that $\mu_1 \ll \mu$ and $\mu_2 \ll \mu$. By Theorem 8.2.22, we deduce from the ergodicity of μ that $\mu_1 = \mu = \mu_2$. This contradicts the fact that $\mu_1 \neq \mu_2$. Thus μ is an extreme point of $M(T, \mathcal{A})$.

To prove the converse implication, let $\mu \in M(T, \mathcal{A}) \backslash E(T, \mathcal{A})$. We want to show that μ is not an extreme point of $M(T, \mathcal{A})$. Since μ is not ergodic, there exists a set $A \in \mathcal{A}$ such that $T^{-1}(A) = A$ with $\mu(A) > 0$ and $\mu(X \backslash A) > 0$. Observe that μ can be written as the following nontrivial convex combination of the T-invariant conditional measures μ_A and $\mu_{X \backslash A}$: for every $B \in \mathcal{A}$,

$$\begin{aligned}\mu(B) = \mu(A \cap B) + \mu\big((X \backslash A) \cap B\big) &= \mu(A)\,\mu_A(B) + \mu(X \backslash A)\,\mu_{X \backslash A}(B)\\ &= \mu(A)\,\mu_A(B) + \big(1 - \mu(A)\big)\,\mu_{X \backslash A}(B).\end{aligned}$$

Hence, μ is a nontrivial convex combination of two distinct T-invariant probability measures and thus μ is not an extreme point of $M(T, \mathcal{A})$. □

We now invoke Krein–Milman's theorem to deduce that every topological dynamical system admits an ergodic and invariant measure. Recall that the *convex hull* of a subset S of a vector space V is the set of all convex combinations of vectors of S.

Theorem 8.2.24 (Krein–Milman's theorem). *If K is a compact subset of a locally convex topological vector space V and E is the set of its extremal points, then $\overline{\mathrm{co}}(E) \supseteq K$,* *where* $\overline{\mathrm{co}}(E)$ *is the closed convex hull of* E. *Consequently,* $\overline{\mathrm{co}}(E) = \overline{\mathrm{co}}(K)$. *In particular, if* K *is convex then* $\overline{\mathrm{co}}(E) = K$.

Proof. See Theorem V.8.4 in Dunford and Schwartz [20]. □

Corollary 8.2.25. *Let* $T : X \to X$ *be a topological dynamical system. Then*

$$\overline{\mathrm{co}}\big(E(T)\big) = M(T) \neq \emptyset.$$

In particular, $E(T) \neq \emptyset$.

Proof. In Theorems 8.1.20 and 8.1.22, we saw that whenever $T : X \to X$ is a topological dynamical system, the set $M(T)$ is a nonempty compact convex subset of the (compact) convex space $M(X)$. Moreover, Theorem 8.2.23 established that $E(T)$ is the set of extreme points of $M(T)$. The result then follows from the application of Krein–Milman's theorem with $K = M(T)$, $V = M(X)$ and $E = E(T)$. □

This corollary can be restated as follows: Every topological dynamical system induces at least one ergodic measure-preserving dynamical system.

The compactness and convexity of $M(T)$ as a subset of the convex space $M(X)$ (equipped with the weak* topology) further allows us to use Choquet's representation theorem to express each element of $M(T)$ in terms of the elements of its set of extreme points $E(T)$. In fact, this decomposition holds in a more general case.

Theorem 8.2.26 (Ergodic decomposition). *Let* $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ *be a measure-preserving transformation of a Borel probability space* $(X, \mathcal{B}, \mu)$. *Then there is a Borel probability space* $(Y, \mathcal{B}(Y), \nu)$ *and a measurable map* $Y \ni y \mapsto \mu_y \in M(X)$ *such that*

(a) μ_y *is an ergodic* T*-invariant Borel probability measure on* X *for* ν*-almost every* $y \in Y$; *and*
(b) $\mu = \int_Y \mu_y \, d\nu(y)$.

Moreover, one may require that the map $y \mapsto \mu_y$ *be injective, or alternatively set*

$$\big(Y, \mathcal{B}(Y), \nu\big) = (X, \mathcal{B}, \mu) \quad \textit{and} \quad \mu_x = \mu_x^{\mathcal{I}_\mu},$$

where $\mathcal{I}_\mu$ *is the* σ*-algebra of* μ*-almost* T*-invariant sets (see Definition* 8.2.5*) and* $\mu_x^{\mathcal{I}}$ *is a Borel probability measure on* X *for which*

$$E(\varphi|\mathcal{I}_\mu)(x) = \int_X \varphi(z)\, d\mu_x^{\mathcal{I}_\mu}(z) \quad \textit{for } \mu\textit{-a.e. } x \in X$$

for all $\varphi \in L^1(X, \mathcal{B}, \mu)$.

Proof. The interested reader is invited to consult Theorem 6.2 in [21]. □

So, for any topological dynamical system, every invariant measure can be uniquely written as a generalized convex combination of ergodic invariant measures.

We end our theoretical discussion of ergodic measures with the following result. We already know from Theorem 1.5.11 that the set of transitive points for any transitive map is a dense set, which can be thought of as "topologically full." The forthcoming result asserts that a dynamical system which admits an ergodic invariant measure supported on the entire space is transitive, and its set of transitive points is "full" not only topologically but also measure-theoretically.

Theorem 8.2.27. *Let $T : X \to X$ be a topological dynamical system. If $\mu \in E(T)$ and* $\operatorname{supp}(\mu) = X$*, then μ-almost every $x \in X$ is a transitive point for T. In particular, T is transitive.*

Proof. Let $\{U_k\}_{k=1}^{\infty}$ be a base for the topology of X. For each $k \in \mathbb{N}$, let X_k be the set of points whose orbits visit U_k infinitely often; in other words,

$$X_k = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} T^{-n}(U_k).$$

We observed in the discussion preceding Proposition 8.2.4 that the set X_k is completely T-invariant. Since μ is ergodic, we deduce that $\mu(X_k) = 0$ or $\mu(X \backslash X_k) = 0$. However,

$$\begin{aligned}\mu(X_k) &= \mu\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} T^{-n}(U_k)\right) = \lim_{m\to\infty} \mu\left(\bigcup_{n=m}^{\infty} T^{-n}(U_k)\right) \\ &\geq \lim_{m\to\infty} \mu(T^{-m}(U_k)) = \lim_{m\to\infty} \mu(U_k) = \mu(U_k) > 0,\end{aligned}$$

where the last strict inequality is due to the fact that the support of the measure μ is X. Therefore, $\mu(X \backslash X_k) = 0$. Since this is true for all $k \in \mathbb{N}$, we conclude that $\mu(X \backslash \bigcap_{k=1}^{\infty} X_k) = 0$. But all points in the set $\bigcap_{k=1}^{\infty} X_k$ have an orbit that visits each basic open set U_k infinitely often. Thus all points of $\bigcap_{k=1}^{\infty} X_k$ are transitive. Hence, μ-a. e. $x \in X$ is a transitive point for T. □

8.2.3 Examples of ergodic measures

We begin this section with a simple example.

Example 8.2.28. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation with a fixed point x_0. Let δ_{x_0} be the Dirac point mass supported at x_0. We saw in Example 8.1.8 that δ_{x_0} is T-invariant. The measure δ_{x_0} is also trivially ergodic since any measurable set is of measure 0 or 1.

We now revisit the rotations of the unit circle. For a comparative perspective of the topological dynamics of these maps, see Theorem 1.5.12.

Proposition 8.2.29. *Let $T_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ be the map defined by $T_\alpha(x) := x + \alpha \pmod 1$. Then T_α is ergodic with respect to the Lebesgue measure if and only if $\alpha \in \mathbb{R}\backslash\mathbb{Q}$.*

Proof. We demonstrated in Example 8.1.9 that the Lebesgue measure λ on $\mathbb{S}^1$ is T_α-invariant for any $\alpha \in \mathbb{R}$. First, assume that $\alpha \notin \mathbb{Q}$. We want to show that λ is ergodic with respect to T_α. For this, we will show that if $f \circ T_\alpha = f$ and $f \in L^2(\lambda)$, then f is λ-a. e. constant. It will then result from Theorem 8.2.18 and Remark 8.2.19 that λ is ergodic. Consider the Fourier series representation of f, which is given by

$$f(x) = \sum_{k\in\mathbb{Z}} a_k e^{2\pi ikx}.$$

Then

$$f \circ T_\alpha(x) = \sum_{k\in\mathbb{Z}} a_k e^{2\pi ik(x+\alpha)} = \sum_{k\in\mathbb{Z}} a_k e^{2\pi ik\alpha} e^{2\pi ikx}.$$

Since we assumed that $f \circ T_\alpha = f$, we deduce from the uniqueness of the Fourier series representation that $a_k e^{2\pi ik\alpha} = a_k$ for all $k \in \mathbb{Z}$. Hence, for each k we have $a_k = 0$ or $e^{2\pi ik\alpha} = 1$. The latter equality holds if and only if $k\alpha \in \mathbb{Z}$. As $\alpha \notin \mathbb{Q}$, this occurs only when $k = 0$. Thus $f(x) = a_0$ for λ-a. e. $x \in \mathbb{S}^1$, that is, f is λ-a. e. constant. This implies that λ is ergodic.

Now, suppose that $\alpha = p/q \in \mathbb{Q}$. We may assume without loss of generality that $q > p \geq 0$. In what follows, all sets must be interpreted modulo 1. Let

$$A := \bigcup_{n=0}^{q-1} \left[\frac{n}{q}, \left(n + \frac{1}{2}\right)\frac{1}{q}\right].$$

Then

$$\begin{aligned}
T_\alpha^{-1}(A) &= \bigcup_{n=0}^{q-1} \left[\frac{n-p}{q}, \left(n - p + \frac{1}{2}\right)\frac{1}{q}\right] \\
&= \bigcup_{n=0}^{p-1} \left[\frac{n-p}{q}, \left(n - p + \frac{1}{2}\right)\frac{1}{q}\right] \cup \bigcup_{n=p}^{q-1} \left[\frac{n-p}{q}, \left(n - p + \frac{1}{2}\right)\frac{1}{q}\right] \\
&= \bigcup_{n=0}^{p-1} \left[\frac{n+q-p}{q}, \left(n + q - p + \frac{1}{2}\right)\frac{1}{q}\right] \cup \bigcup_{k=0}^{q-(p+1)} \left[\frac{k}{q}, \left(k + \frac{1}{2}\right)\frac{1}{q}\right] \\
&= \bigcup_{k=q-p}^{q-1} \left[\frac{k}{q}, \left(k + \frac{1}{2}\right)\frac{1}{q}\right] \cup \bigcup_{k=0}^{q-(p+1)} \left[\frac{k}{q}, \left(k + \frac{1}{2}\right)\frac{1}{q}\right] \\
&= \bigcup_{k=0}^{q-1} \left[\frac{k}{q}, \left(k + \frac{1}{2}\right)\frac{1}{q}\right] = A.
\end{aligned}$$

Also, one immediately verifies that $\lambda(A) = q \cdot (1/2)(1/q) = 1/2$. In summary, $T_\alpha^{-1}(A) = A$ and $\lambda(A) \notin \{0, 1\}$. Thus, λ is not ergodic with respect to T_α when $\alpha \in \mathbb{Q}$. □

We now return to the doubling map and its generalizations.

Example 8.2.30. Fix $n > 1$. Recall once more the map $T_n : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $T_n(x) = nx \pmod 1$. We claim that T_n is ergodic with respect to the Lebesgue measure λ. We saw in Example 8.1.10 that T_n preserves λ. It is possible to demonstrate the ergodicity of T_n with respect to λ in a similar way that we did for T_α in Proposition 8.2.29. However, in this example we will provide a different proof. Let $A \in \mathcal{B}(\mathbb{S}^1)$ be a set such that $T_n^{-1}(A) = A$ and $\lambda(A) > 0$. To establish ergodicity, we need to show that $\lambda(A) = 1$.

Recall that Lebesgue's density theorem (see Corollary 2.14 in Mattila [46]) states that for any Lebesgue measurable set $A \subseteq \mathbb{R}^n$, the density of A is 0 or 1 at λ-almost every point of $\mathbb{R}^n$. Moreover, the density of A is 1 at λ-almost every point of A. The density of A at $x \in \mathbb{R}^n$ is defined as

$$\lim_{r\to 0} \frac{\lambda(A \cap B(x,r))}{\lambda(B(x,r))}.$$

Given that $\lambda(A) > 0$, take x to be a Lebesgue density point of A, that is, a point where the density of A is 1, that is,

$$\lim_{r\to 0} \frac{\lambda(A \cap B(x,r))}{2r} = 1. \tag{8.15}$$

Set

$$r_k := 1/(2n^k).$$

Then T_n^k is injective on each arc of length less than $2r_k$. So $T_n^k|_{B(x,r_k)}$ is injective for each $x \in \mathbb{S}^1$. On the other hand,

$$T_n^k(B(x,r_k)) = \mathbb{S}^1 \backslash \{T_n^k(x + r_k)\}.$$

Thus

$$\lambda(T_n^k(B(x,r_k))) = 1.$$

Therefore,

$$\lambda(A) = \lambda(T_n^k(A)) \geq \frac{\lambda(T_n^k(A \cap B(x,r_k)))}{\lambda(T_n^k(B(x,r_k)))} = \frac{n^k \lambda(A \cap B(x,r_k))}{n^k \lambda(B(x,r_k))} = \frac{\lambda(A \cap B(x,r_k))}{2r_k} \underset{k\to\infty}{\longrightarrow} 1$$

by (8.15). Consequently, $\lambda(A) = 1$. This proves the ergodicity of λ.

Example 8.2.31. Recall the full Markov maps from Example 8.1.11. We claim that any such T is ergodic with respect to the Lebesgue measure λ. As we did in the previous example, we would like to use Lebesgue's density theorem to prove this. However, in contradistinction with the preceding example, for all $r > 0$ and all $k \in \mathbb{N}$, the restriction of T^k to the ball $B(p_j, r)$ is not one-to-one when p_j is a point of continuity for a full Markov map T; for example, the point 1/2 for the tent map. Despite that potential lack of injectivity, let us try to use Lebesgue's density theorem.

For each $n \in \mathbb{N}$, let $\mathcal{P}_n := \{I_j^{(n)} \mid 1 \le j \le q^n\}$ be the "partition" of $[0,1]$ into the successive intervals of monotonicity of T^n. In particular, $I_j^{(1)} = I_j$ for all $1 \le j \le q$. For each $1 \le j < q^n$, let $p_j^{(n)}$ be the unique point in $I_j^{(n)} \cap I_{j+1}^{(n)}$. For all $x \in [0,1]\backslash\{p_j^{(n)} : 1 \le j < q^n\}$, let $I^{(n)}(x)$ be the unique element of $\mathcal{P}_n$ containing x. We will need two claims.

Claim 1. *For every $n \in \mathbb{N}$, the map $T^n : [0,1] \to [0,1]$ is a full Markov map under the "partition" $\mathcal{P}_n$, and $\mathcal{P}_{n+1}$ is finer than $\mathcal{P}_n$.*

Claim 2. *If $A \in \mathcal{B}([0,1])$, then*

$$\lim_{n\to\infty} \frac{\lambda(A \cap I^{(n)}(x))}{\lambda(I^{(n)}(x))} = 1 \textit{ for } \lambda\text{-}a.e.\ x \in A.$$

Proof of ergodicity of T. For the time being, suppose that both claims hold. Let A be a Borel subset of $[0,1]$ such that $T^{-1}(A) = A$ and $\lambda(A) > 0$. By the surjectivity of T, we know that $T(A) = A$. Fix any $x \in A$ satisfying Claim 2. For each $n \in \mathbb{N}$, let m_n be the slope of $T^n|_{I^{(n)}(x)}$. Using both claims, we obtain that

$$\lambda(A) = \lambda(T^n(A)) \ge \frac{\lambda(T^n(A \cap I^{(n)}(x)))}{\lambda(T^n(I^{(n)}(x)))} = \frac{m_n \lambda(A \cap I^{(n)}(x))}{m_n \lambda(I^{(n)}(x))} = \frac{\lambda(A \cap I^{(n)}(x))}{\lambda(I^{(n)}(x))} \underset{n\to\infty}{\longrightarrow} 1.$$

Consequently, $\lambda(A) = 1$. This proves the ergodicity of λ. □

Proof of Claim 1. We proceed by induction. Suppose that T^n is a full Markov map under the "partition" $\mathcal{P}_n$. It is obvious that T^{n+1} is piecewise linear. Fix $I_j^{(n)} \in \mathcal{P}_n$. For all $1 \le i \le q$, consider

$$I_{j,i}^{(n+1)} := T|_{I_i}^{-1}(I_j^{(n)}).$$

Define

$$\widetilde{\mathcal{P}}_{n+1} := \{I_{j,i}^{(n+1)} \mid 1 \le j \le q^n, 1 \le i \le q\}.$$

Then

$$\begin{aligned}
\bigcup_{j=1}^{q^n}\bigcup_{i=1}^{q} I_{j,i}^{(n+1)} &= \bigcup_{j=1}^{q^n}\bigcup_{i=1}^{q} T|_{I_i}^{-1}(I_j^{(n)}) = \bigcup_{i=1}^{q}\bigcup_{j=1}^{q^n} T|_{I_i}^{-1}(I_j^{(n)}) \\
&= \bigcup_{i=1}^{q} T|_{I_i}^{-1}\left(\bigcup_{j=1}^{q^n} I_j^{(n)}\right) = \bigcup_{i=1}^{q} T|_{I_i}^{-1}([0,1]) \\
&= \bigcup_{i=1}^{q} I_i = [0,1].
\end{aligned}$$

Thus $\widetilde{\mathcal{P}}_{n+1}$ is a cover of $[0,1]$. Obviously, the interiors of the intervals in $\widetilde{\mathcal{P}}_{n+1}$ are mutually disjoint and $\widetilde{\mathcal{P}}_{n+1}$ is finer than $\widetilde{\mathcal{P}}_n$. For all $1 \le j \le q^n$ and all $1 \le i \le q$ we further have

$$T^{n+1}(I_{j,i}^{(n+1)}) = T^n(T(I_{j,i}^{(n+1)})) = T^n(T(T|_{I_i}^{-1}(I_j^{(n)}))) = T^n(I_j^{(n)}) = [0,1].$$

So $\mathcal{P}_{n+1} = \widetilde{\mathcal{P}}_{n+1}$ and T^{n+1} is a full Markov map under the "partition" $\mathcal{P}_{n+1}$, which is finer than $\mathcal{P}_n$. This completes the inductive step. Since the claim clearly holds when $n = 1$, Claim 1 has been established for all $n \in \mathbb{N}$. □

Proof of Claim 2. Let $\mathcal{A}_n := \sigma(\mathcal{P}_n)$ be the σ-algebra generated by $\mathcal{P}_n$. By Claim 1, $\mathcal{P}_{n+1}$ is finer than $\mathcal{P}_n$ for all $n \in \mathbb{N}$ and thus the sequence of σ-algebras $(\mathcal{A}_n)_{n=1}^{\infty}$ is ascending. Now, let

$$m := \min\{|\mathrm{slope}(T|_{I_i})| : 1 \le i \le q\} = \{|(T|_{\mathrm{Int}(I_i)})'| : 1 \le i \le q\} > 1.$$

Then

$$\mathrm{diam}(\mathcal{P}_n) := \sup\{\mathrm{diam}(I_j^{(n)}) : 1 \le j \le q^n\} \le m^{-n} \underset{n\to\infty}{\longrightarrow} 0.$$

Therefore, $\mathcal{A}_\infty := \sigma(\bigcup_{n=1}^{\infty} \mathcal{A}_n)$ contains all the subintervals of $[0,1]$. Hence, $\mathcal{A}_\infty = \mathcal{B}([0,1])$. Let $A \in \mathcal{B}([0,1])$. According to Theorem A.1.67,

$$\lim_{n\to\infty} E(\mathbb{1}_A|\mathcal{A}_n)(x) = E(\mathbb{1}_A|\mathcal{B}([0,1]))(x) = \mathbb{1}_A(x) \quad \text{for } \lambda\text{-a. e. } x \in [0,1]. \tag{8.16}$$

Now, recall that $\mathcal{P}_n := \{I_j^{(n)} : 1 \le j \le q^n\}$ is the "partition" of $[0,1]$ into the successive intervals of monotonicity of T^n. Moreover, $p_j^{(n)}$ is the unique point in $I_j^{(n)} \cap I_{j+1}^{(n)}$ for every $1 \le j < q^n$. Define

$$\mathcal{P}_n' := \{I_1^{(n)} \backslash \{p_1^{(n)}\}\} \cup \{\mathrm{Int}(I_j^{(n)}) : 1 < j < q^n\} \cup \{I_{q^n}^{(n)} \backslash \{p_{q^n-1}^{(n)}\}\} \cup \{\{p_j^{(n)}\} : 1 \le j < q^n\}.$$

Though $\mathcal{P}_n$ is not a partition per se, the family $\mathcal{P}_n'$ is a finite partition of $[0,1]$. Moreover, it is easy to see that $\mathcal{A}_n := \sigma(\mathcal{P}_n) = \sigma(\mathcal{P}_n')$. By Example A.1.62, the conditional expectation function $E(\mathbb{1}_A|\mathcal{A}_n)$ is constant on each element of the partition $\mathcal{P}_n'$. For all x except the points $p_j^{(n)}$, by definition of the conditional expectation function we must also have

$$\int_{I^{(n)}(x)} E(\mathbb{1}_A|\mathcal{A}_n)\, d\lambda = \int_{I^{(n)}(x)} \mathbb{1}_A\, d\lambda = \lambda(A \cap I^{(n)}(x)).$$

Therefore,

$$E(\mathbb{1}_A|\mathcal{A}_n)(y) = \frac{\lambda(A \cap I^{(n)}(x))}{\lambda(I^{(n)}(x))}, \quad \forall y \in \mathrm{Int}(I^{(n)}(x)).$$

Since $I^{(n)}(y) = I^{(n)}(x)$ for all $y \in \mathrm{Int}(I^{(n)}(x))$, it ensues that

$$E(\mathbb{1}_A|\mathcal{A}_n)(z) = \frac{\lambda(A \cap I^{(n)}(z))}{\lambda(I^{(n)}(z))} \quad \text{for } \lambda\text{-a.e. } z \in [0,1].$$

It follows from (8.16) that

$$\lim_{n\to\infty} \frac{\lambda(A \cap I^{(n)}(x))}{\lambda(I^{(n)}(x))} = 1 \quad \text{for } \lambda\text{-a.e. } x \in A.$$

Claim 2 is proved and this completes this example. □

The next example concerns the shift map.

Example 8.2.32. Recall the one-sided Bernoulli shift from Example 8.1.14 where, given a probability space $(E, \mathcal{F}, P)$, the product space $(E^\infty, \mathcal{F}_\infty, \mu_P)$ is a probability space and the product measure μ_P is invariant under the left shift map σ. We now demonstrate that μ_P is also ergodic with respect to σ. To prove this, we use Lemmas 8.2.16–8.2.17. Let A, B be cylinder sets of length M and N, respectively. Since cylinder A depends on the first M coordinates, cylinder $\sigma^{-j}(A)$ depends on coordinates $j+1$ to $j+M$. Consequently, cylinders $\sigma^{-j}(A)$ and B depend on different coordinates as soon as $j \geq N$. Since μ_P is a product measure and is σ-invariant, we deduce that

$$\mu_P(\sigma^{-j}(A) \cap B) = \mu_P(\sigma^{-j}(A))\mu_P(B) = \mu_P(A)\mu_P(B), \quad \forall j \geq N.$$

It follows immediately that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu_P(\sigma^{-j}(A) \cap B) = \mu_P(A)\mu_P(B).$$

The ergodicity of σ ensues from Lemmas 8.2.16 and 8.2.17. In Example 8.3.13, we shall see that the shift map enjoys an even stronger property than ergodicity.

Our final example pertains to ordinary normal numbers.

Example 8.2.33. Let $n \geq 2$. On one hand, consider the probability space $(E, \mathcal{F}, P)$, with set $E = \{0, 1, \ldots, n-1\}$, σ-algebra $\mathcal{F} = \mathcal{P}(E)$ and probability measure $P = (1/n)\sum_{k=0}^{n-1} \delta_k$. According to Example 8.2.32, the product space $(E^\infty, \mathcal{F}_\infty, \mu_P)$ is a probability space and the measure μ_P is ergodic with respect to the shift map σ.

On the other hand, consider the map $T_n : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $T_n(x) = nx \pmod 1$. In Example 8.2.30, we learned that T_n is ergodic with respect to the Lebesgue measure λ. This map is also distance expanding, and thus admits a Markov partition. In Examples 4.4.4 and 4.4.5, explicit partitions were given. In Example 4.5.3, the coding map generated by the partition $\mathcal{R} = \{R_i = [\frac{i}{n}, \frac{i+1}{n}] : 0 \leq i < n\}$ was identified as

$$\begin{array}{cccc} \pi : & E^\infty & \longrightarrow & \mathbb{S}^1 \\ & \omega = (\omega_k)_{k=1}^\infty & \longmapsto & \pi(\omega) = \displaystyle\sum_{k=1}^\infty \frac{\omega_k}{n^k}. \end{array}$$

Properties of π were given in Theorem 4.5.2. Among others, π is continuous, surjective, and its restriction to the set $Z := \pi^{-1}(\mathbb{S}^1 \backslash \bigcup_{k=0}^{\infty} T_n^{-k}(\bigcup_{i=0}^{n-1} \partial R_i))$ is injective. The set Z consists of all $\omega \in E^{\infty}$ whose coordinates are not eventually constant and equal to 0 or $n-1$, i. e.

$$E^{\infty} \backslash Z = \bigcup_{\tau \in E^*} \{\tau 0^{\infty}, \tau(n-1)^{\infty}\}.$$

The coding map π is two-to-one on $E^{\infty}\backslash Z$, which is a countable set and thus has μ_P-measure zero. So $\mu_P(Z) = 1$.

The coding map $\pi : (E^{\infty}, \mathcal{F}_{\infty}, \mu_P) \to (\mathbb{S}^1, \mathcal{B}(\mathbb{S}^1), \lambda)$ is measure-preserving. Indeed, it is easy to show that the family

$$\mathcal{P} := \left\{ \left[\frac{i}{n^k}, \frac{i+1}{n^k} \right] : 0 \le i < n^k,\ k \in \mathbb{N} \right\} \bigcup \left\{ \frac{i}{n^k} : 0 \le i < n^k,\ k \in \mathbb{N} \right\}$$

is a π-system that generates $\mathcal{B}(\mathbb{S}^1)$ and $\mu_P \circ \pi^{-1}(P) = \lambda(P)$ for all $P \in \mathcal{P}$. Hence the coding map is measure-preserving according to Lemma 8.1.4.

Let $\varphi : E^{\infty} \to E \subseteq \mathbb{R}$ be the function $\varphi(\omega) = \omega_1$. Clearly, $\varphi \in L^1(E^{\infty}, \mathcal{F}_{\infty}, \mu_P)$. Furthermore, $\varphi \circ \sigma^j(\omega) = \omega_{j+1}$ for all $j \ge 0$. Since μ_P is ergodic with respect to the shift map σ, the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14) asserts that there is $M \in \mathcal{F}_{\infty}$ such that $\mu_P(M) = 1$ and

$$\begin{aligned} \lim_{k\to\infty} \frac{1}{k} \sum_{j=1}^{k} \omega_j &= \lim_{k\to\infty} \frac{1}{k} S_k\varphi(\omega) = \int_{E^{\infty}} \varphi(\omega)\, d\mu_P(\omega) \\ &= \sum_{i=0}^{n-1} \int_{[i]} i\, d\mu_P = \sum_{i=0}^{n-1} i\, \mu_P([i]) = \sum_{i=0}^{n-1} i \cdot \frac{1}{n} \\ &= \frac{n-1}{2}, \quad \forall \omega \in M. \end{aligned}$$

As π is continuous, it is Borel measurable and thus $\pi(M)$ is Lebesgue measurable. Since π is measure-preserving, we have that $\lambda(\pi(M)) = \mu_P \circ \pi^{-1}(\pi(M)) \ge \mu_P(M) = 1$. We infer that the digits of the n-adic expansion of λ-almost every number between 0 and 1 average $(n-1)/2$ asymptotically.

Now, fix any $0 \le i < n$. Since μ_P is ergodic with respect to σ, Corollary 8.2.15 of Birkhoff's ergodic theorem affirms that for μ_P-a. e. $\omega \in E^{\infty}$,

$$\lim_{k\to\infty} \frac{1}{k} \#\{1 \le j \le k : \omega_j = i\} = \lim_{k\to\infty} \frac{1}{k} \#\{0 \le j < k : \sigma^j(\omega) \in [i]\} = \mu_P([i]) = \frac{1}{n}.$$

Since the coding map is measure-preserving, we deduce that λ-almost every number between 0 and 1 has a n-adic expansion whose digits are equal to i with a frequency of $1/n$. This frequency is independent of the digit i, as one naturally expects.

8.2.4 Uniquely ergodic transformations

As mentioned in Subsection 8.1.3, there are measurable transformations that do not admit any invariant measure. However, in Theorem 8.1.22 we showed that every topological dynamical system carries invariant probability measures. Among those systems, some have only one such measure. Per Corollary 8.2.25, this happens precisely when there is a unique ergodic invariant measure. These maps deserve a special name.

Definition 8.2.34. A measurable transformation $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ is said to be *uniquely ergodic* if $E(T, \mathcal{A})$ is a singleton.

Lemma 8.2.35. *A topological dynamical system $T : X \to X$ is uniquely ergodic if and only if $M(T)$ is a singleton.*

Proof. According to Corollary 8.2.25, $M(T)$ is the closed convex hull of $E(T)$. Therefore, $M(T)$ is a singleton precisely when $E(T)$ is. □

Recall that by Riesz' representation theorem (Theorem A.1.53), whenever X is a compact metrizable space, every $\mu \in M(X)$ is uniquely determined by a normalized positive continuous linear functional $\ell_\mu \in C(X)^*$, namely

$$\ell_\mu(f) = \int_X f\, d\mu, \quad \forall f \in C(X).$$

If $T : X \to X$ is a topological dynamical system, then by Theorem 8.1.18 each $\mu \in M(T)$ corresponds to a T-invariant functional in the sense that

$$\ell_\mu(f \circ T) = \ell_\mu(f), \quad \forall f \in C(X).$$

We will show that if T is uniquely ergodic, then all T-invariant continuous linear functionals are scalar multiples of ℓ_{μ_0}, where μ_0 is the unique ergodic T-invariant measure.

First, we introduce the variation of a functional.

Definition 8.2.36. Let $\ell \in C(X)^*$. The *variation* of ℓ is the functional $\mathrm{var}(\ell) : C(X) \to \mathbb{R}$ defined as follows: For any function $f \in C(X)$, $f \geq 0$, set

$$\mathrm{var}(\ell)(f) := \sup_{\substack{g \in C(X) \\ 0 \leq g \leq f}} \ell(g),$$

and for any other $f \in C(X)$ let

$$\mathrm{var}(\ell)(f) := \mathrm{var}(\ell)(f_+) - \mathrm{var}(\ell)(f_-).$$

Lemma 8.2.37. *Let $\ell \in C(X)^*$. Set $\Delta\ell := \mathrm{var}(\ell) - \ell$. Then $\mathrm{var}(\ell), \Delta\ell \in C(X)^*$ and both are positive. In addition, if ℓ is T-invariant then so are $\mathrm{var}(\ell)$ and $\Delta\ell$.*

Proof. Let $f \in C(X), f \geq 0$. The positivity of $\mathrm{var}(\ell)$ is obvious since $\mathrm{var}(\ell)(f) \geq \ell(0) = 0$. It is also easy to see that $\mathrm{var}(\ell)(cf) = c \cdot \mathrm{var}(\ell)(f)$ for all $c \geq 0$. Now, let $f_1, f_2 \in C(X)$, $f_1, f_2 \geq 0$. On one hand, if $g_1, g_2 \in C(X)$ satisfy $0 \leq g_1 \leq f_1$ and $0 \leq g_2 \leq f_2$, then $0 \leq g_1 + g_2 \leq f_1 + f_2$, and hence

$$\ell(g_1) + \ell(g_2) = \ell(g_1 + g_2) \leq \mathrm{var}(\ell)(f_1 + f_2).$$

Taking the supremum over all such g_1, g_2, we get

$$\mathrm{var}(\ell)(f_1) + \mathrm{var}(\ell)(f_2) \leq \mathrm{var}(\ell)(f_1 + f_2). \tag{8.17}$$

On the other hand, let $g \in C(X)$ be such that $0 \leq g \leq f_1 + f_2$. Define $g_1 = \min\{f_1, g\}$ and $g_2 = g - g_1$. Then $g_1, g_2 \in C(X)$ and satisfy $0 \leq g_1 \leq f_1$ and $0 \leq g_2 \leq f_2$. It follows that

$$\ell(g) = \ell(g_1 + g_2) = \ell(g_1) + \ell(g_2) \leq \mathrm{var}(\ell)(f_1) + \mathrm{var}(\ell)(f_2).$$

Taking the supremum over all such g, we get

$$\mathrm{var}(\ell)(f_1 + f_2) \leq \mathrm{var}(\ell)(f_1) + \mathrm{var}(\ell)(f_2). \tag{8.18}$$

By (8.17) and (8.18),

$$\mathrm{var}(\ell)(f_1 + f_2) = \mathrm{var}(\ell)(f_1) + \mathrm{var}(\ell)(f_2).$$

This proves the linearity for nonnegative functions. The linearity for other functions follows directly from the definition of $\mathrm{var}(\ell)$ for such functions.

Now, suppose that ℓ is T-invariant. Given $f \in C(X), f \geq 0$, first notice that

$$\mathrm{var}(\ell)(f \circ T) = \sup_{\substack{g \in C(X) \\ 0 \leq g \leq f \circ T}} \ell(g) \geq \sup_{\substack{h \in C(X) \\ 0 \leq h \leq f}} \ell(h \circ T) = \sup_{\substack{h \in C(X) \\ 0 \leq h \leq f}} \ell(h) = \mathrm{var}(\ell)(f). \tag{8.19}$$

We shall prove that this inequality implies the desired equality. To do this, let us pass from functionals to measures. Let ν be the Borel measure corresponding to $\mathrm{var}(\ell)$. We shall show that (8.19) implies that $\nu \circ T^{-1}(B) \geq \nu(B)$ for all Borel sets B. Since ν is a regular measure (by Theorem A.1.24), it suffices to prove the inequality for closed sets. Let F be a closed set in X. By Urysohn's lemma (see 15.6 in Willard [77]), there is a descending sequence of open sets $(U_n)_{n=1}^{\infty}$ whose intersection is F and to which corresponds a sequence $(f_n)_{n=1}^{\infty}$ of continuous functions on X such that $0 \leq f_n \leq 1$, $f_n = 0$ on $X \backslash U_n$ and $f_n = 1$ on F. Then, for all $n \in \mathbb{N}$,

$$\int_X f_n \, d(\nu \circ T^{-1}) = \int_X f_n \circ T \, d\nu = \mathrm{var}(\ell)(f_n \circ T) \geq \mathrm{var}(\ell)(f_n) = \int_X f_n \, d\nu.$$

Since $\lim_{n\to\infty} f_n = \mathbb{1}_F$ and $\|f_n\|_\infty \leq 1$ for all $n \in \mathbb{N}$, Lebesgue's dominated convergence theorem (Theorem A.1.38) implies that

$$\nu \circ T^{-1}(F) = \int_X \mathbb{1}_F \, d(\nu \circ T^{-1}) = \lim_{n\to\infty} \int_X f_n \, d(\nu \circ T^{-1}) \geq \lim_{n\to\infty} \int_X f_n \, d\nu = \int_X \mathbb{1}_F \, d\nu = \nu(F).$$

As the closed set F was arbitrarily chosen and the Borel measure ν is regular, the measure of any Borel set B is equal to the supremum of the measures of all closed sets contained in B. It immediately follows that

$$\nu \circ T^{-1}(B) \geq \nu(B), \quad \forall B \in \mathcal{B}(X). \tag{8.20}$$

Replacing B by $X \backslash B$ in (8.20), we obtain that

$$\nu \circ T^{-1}(X \backslash B) \geq \nu(X \backslash B), \quad \forall B \in \mathcal{B}(X). \tag{8.21}$$

Since $\nu(B) + \nu(X\backslash B) = \nu(X) = \nu \circ T^{-1}(B) + \nu \circ T^{-1}(X \backslash B)$, inequality (8.21) implies that

$$\nu \circ T^{-1}(B) \leq \nu(B), \quad \forall B \in \mathcal{B}(X). \tag{8.22}$$

From (8.20) and (8.22), we deduce that $\nu \circ T^{-1}(B) = \nu(B)$ for all $B \in \mathcal{B}(X)$, that is, ν is T-invariant. It follows that $\operatorname{var}(\ell)$ is T-invariant since

$$\operatorname{var}(\ell)(f \circ T) = \int_X f \circ T \, d\nu = \int_X f \, d(\nu \circ T^{-1}) = \int_X f \, d\nu = \operatorname{var}(\ell)(f).$$

For any $f \in C(X)$, it ensues that

$$\operatorname{var}(\ell)(f \circ T) = \operatorname{var}(\ell)(f_+ \circ T) - \operatorname{var}(\ell)(f_- \circ T) = \operatorname{var}(\ell)(f_+) - \operatorname{var}(\ell)(f_-) = \operatorname{var}(\ell)(f).$$

The proof of the statements on $\Delta\ell$ are left to the reader. □

We use the variation functional to demonstrate that uniquely ergodic systems admit only T-invariant functionals that are multiples of the ergodic T-invariant measure.

Lemma 8.2.38. *Let $T : X \to X$ be a uniquely ergodic topological dynamical system. Let μ_0 be the unique ergodic T-invariant measure and ℓ_{μ_0} its corresponding T-invariant normalized positive continuous linear functional. Then any (not necessarily positive or normalized) T-invariant $\ell \in C(X)^*$ is of the form*

$$\ell = c \cdot \ell_{\mu_0},$$

where $c \in \mathbb{R}$.

Proof. Assume that $\ell \in C(X)^*$ is T-invariant. Lemma 8.2.37 then says that $\operatorname{var}(\ell)$ and $\Delta\ell$ are T-invariant positive continuous linear functionals on $C(X)$. Since T is uniquely ergodic, Lemma 8.2.35 implies that there must exist $C, \widetilde{C} \geq 0$ such that $\operatorname{var}(\ell) = C\,\ell_{\mu_0}$ and $\Delta\ell = \widetilde{C}\,\ell_{\mu_0}$. It then follows that $\ell = \operatorname{var}(\ell) - \Delta\ell = (C - \widetilde{C})\ell_{\mu_0}$. □

We aim to show that a stronger variant of Birkhoff's ergodic theorem (Theorem 8.2.11) holds for uniquely ergodic dynamical systems. More precisely, the Birkhoff averages converge uniformly and thereby everywhere. The proof relies upon a deep result in functional analysis called the Hahn–Banach theorem. The statement and proof of this theorem can be found as Theorem II.3.10 in Dunford and Schwartz [20].

In light of the existence of an invariant Borel probability measure for every topological dynamical system, we will first introduce a set of functions which will play an important role on multiple occasions in the sequel.

Definition 8.2.39. Let $T : X \to X$ be a topological dynamical system. A function $f \in C(X)$ is said to be *cohomologous to zero* in the additive group $C(X)$ if

$$f = g \circ T - g$$

for some $g \in C(X)$. The set of such functions will be denoted by $C_0(T)$.

These functions have the property that their integral with respect to any invariant measure is equal to zero. Moreover, when the system is uniquely ergodic, every function whose integral is equal to zero can be approximated by functions that are cohomologous to zero, as the following lemma shows.

Lemma 8.2.40. *Let $T : X \to X$ be a topological dynamical system and $\mu \in M(T)$. Let*

$$C_0(\mu) := \left\{ f \in C(X) \;\Big|\; \int_X f \, d\mu = 0 \right\}.$$

Then $C_0(T)$ and $C_0(\mu)$ are vector subspaces of $C(X)$. Moreover, $C_0(\mu)$ is closed in $C(X)$ and $\overline{C_0(T)} \subseteq C_0(\mu)$. In addition, if T is uniquely ergodic then $\overline{C_0(T)} = C_0(\mu)$.

Proof. It is easy to see that $C_0(T)$ and $C_0(\mu)$ are vector subspaces of $C(X)$, and that $C_0(\mu)$ is closed in $C(X)$ (recall that this latter is endowed with the topology of uniform convergence). Let $f \in C_0(T)$. Then $f = g \circ T - g$ for some $g \in C(X)$. The T-invariance of μ yields

$$\int_X f \, d\mu = \int_X g \circ T \, d\mu - \int_X g \, d\mu = \int_X g \, d(\mu \circ T^{-1}) - \int_X g \, d\mu = 0.$$

So $C_0(T) \subseteq C_0(\mu)$, and hence $\overline{C_0(T)} \subseteq \overline{C_0(\mu)} = C_0(\mu)$.

Assume now that T is uniquely ergodic. Suppose by way of contradiction that there exists $f_0 \in C_0(\mu) \backslash \overline{C_0(T)}$. According to the Hahn–Banach theorem, there is $\ell \in C(X)^*$ such that $\ell(f) = 0$ for all $f \in \overline{C_0(T)}$ whereas $\ell(f_0) = 1$. By Lemma 8.2.38, there exists $c \in \mathbb{R}$ such that $\ell = c \cdot \ell_\mu$. But this is impossible since $\ell_\mu(f_0) = 0$ while $\ell(f_0) = 1$. This contradiction implies that $\overline{C_0(T)} = C_0(\mu)$ when T is uniquely ergodic. □

We can now state a stronger version of the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14) for uniquely ergodic dynamical systems.

Theorem 8.2.41. *Let $T : X \to X$ be a topological dynamical system and $\mu \in M(T)$. The following statements are equivalent:*

(a) *T is uniquely ergodic.*
(b) *For every $f \in C(X)$, the Birkhoff averages $\frac{1}{n}S_nf(x)$ converge to $\int_X f\,d\mu$ for all $x \in X$.*
(c) *For every $f \in C(X)$, the Birkhoff averages $\frac{1}{n}S_nf$ converge uniformly to $\int_X f\,d\mu$.*

Proof. The structure of the proof will be the following sequence of implications: (a)⇒(c)⇒(b)⇒(a).

[(a)⇒(c)] Suppose that T is uniquely ergodic. For any $f \in C_0(T)$ and any $x \in X$ we have

$$\left|\frac{1}{n}S_nf(x)\right| = \left|\frac{1}{n}S_ng(T(x)) - \frac{1}{n}S_ng(x)\right| = \frac{1}{n}\left|g(T^n(x)) - g(x)\right| \le \frac{2}{n}\|g\|_\infty.$$

Thus

$$\lim_{n\to\infty}\left\|\frac{1}{n}S_nf - 0\right\|_\infty = 0, \quad \forall f \in C_0(T). \tag{8.23}$$

In Lemma 8.2.40, we observed that $C_0(T) \subseteq C_0(\mu)$. Therefore, the sequence $(\frac{1}{n}S_nf)_{n=1}^\infty$ converges uniformly on X to $\int_X f\,d\mu$ for every $f \in C_0(T)$.

Now, suppose that $f \in C_0(\mu)$. Since T is uniquely ergodic, Lemma 8.2.40 asserts that $\overline{C_0(T)} = C_0(\mu)$. Let $\varepsilon > 0$ and choose $f_\varepsilon \in C_0(T)$ such that $\|f - f_\varepsilon\|_\infty \le \varepsilon$. Then

$$\begin{aligned}
\left\|\frac{1}{n}S_nf - \int_X f\,d\mu\right\|_\infty &\le \left\|\frac{1}{n}S_nf - \frac{1}{n}S_nf_\varepsilon\right\|_\infty + \left\|\frac{1}{n}S_nf_\varepsilon - \int_X f_\varepsilon\,d\mu\right\|_\infty + \left\|\int_X f_\varepsilon\,d\mu - \int_X f\,d\mu\right\|_\infty \\
&\le \frac{1}{n}\sum_{k=0}^{n-1}\|f\circ T^k - f_\varepsilon\circ T^k\|_\infty + \left\|\frac{1}{n}S_nf_\varepsilon - \int_X f_\varepsilon\,d\mu\right\|_\infty + 0 \\
&\le \|f - f_\varepsilon\|_\infty + \left\|\frac{1}{n}S_nf_\varepsilon - \int_X f_\varepsilon\,d\mu\right\|_\infty \\
&\le \varepsilon + \left\|\frac{1}{n}S_nf_\varepsilon - \int_X f_\varepsilon\,d\mu\right\|_\infty.
\end{aligned}$$

As $f_\varepsilon \in C_0(T)$, relation (8.23) guarantees that $\lim_{n\to\infty}\|\frac{1}{n}S_nf_\varepsilon - \int_X f_\varepsilon\,d\mu\|_\infty = 0$ and we deduce that

$$\limsup_{n\to\infty}\left\|\frac{1}{n}S_nf - \int_X f\,d\mu\right\|_\infty \le \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the implication (a)⇒(c) is proved for any function $f \in C_0(\mu)$ and, therefore, for any $f \in C(X)$ by replacing f by $f - \int_X f\,d\mu$.

[(c)⇒(b)] This is obvious.

[(b)⇒(a)] Let $\nu \in M(T)$ and $f \in C(X)$. Since $\lim_{n\to\infty}\frac{1}{n}S_nf(x) = \int_X f\,d\mu$ for all $x \in X$ and $\|\frac{1}{n}S_nf\|_\infty \le \|f\|_\infty < \infty$ for all $n \in \mathbb{N}$, Lebesgue's dominated convergence theorem (Theorem A.1.38) asserts that

$$\lim_{n\to\infty}\int_X \frac{1}{n}S_nf(x)\,d\nu(x) = \int_X\left(\int_X f\,d\mu\right)d\nu(x) = \int_X f\,d\mu.$$

On the other hand, the T-invariance of ν means that $\nu = \nu \circ T^{-k}$ for all $k \in \mathbb{N}$. Thus for all $n \in \mathbb{N}$ we have that $\nu = \frac{1}{n}\sum_{k=0}^{n-1} \nu \circ T^{-k}$, and hence

$$\int_X \frac{1}{n} S_n f(x)\, d\nu(x) = \int_X \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \, d\nu = \frac{1}{n} \sum_{k=0}^{n-1} \int_X f \, d(\nu \circ T^{-k}) = \int_X f \, d\nu.$$

From the last two formulas, it follows that $\int_X f \, d\nu = \int_X f \, d\mu$ for all $f \in C(X)$. By Corollary A.1.54, we conclude that $\nu = \mu$. So μ is the unique T-invariant measure. □

The previous theorem has the following consequence in topological dynamics. Compare this with Theorem 8.2.27.

Corollary 8.2.42. *Let $T : X \to X$ be a uniquely ergodic topological dynamical system. If* $\operatorname{supp}(\mu) = X$ *for the unique* $\mu \in M(T) = E(T)$, *then T is minimal.*

Proof. Let U be a nonempty open set in X. Then $\mu(U) > 0$ because $\operatorname{supp}(\mu) = X$. By Theorem A.1.24, μ is regular. Hence, $\mu(U) = \sup\{\mu(F) : F \subseteq U, F \text{ closed}\} > 0$ and thus there is a closed set $F_0 \subseteq U$ such that $\mu(F_0) > 0$. The sets F_0 and $X \backslash U$ are disjoint closed sets and Urysohn's lemma (see 15.6 in Willard [77]) states that there is a nonnegative function $f \in C(X)$ with the properties that $f = 1$ on F_0 and $f = 0$ on $X \backslash U$. Notice that $\int_X f \, d\mu \geq \mu(F_0) > 0$. Consequently, for any $x \in X$, Theorem 8.2.41 asserts that

$$\lim_{n\to\infty} \frac{1}{n} S_n f(x) = \int_X f \, d\mu > 0.$$

Hence, there exists $n \in \mathbb{N}$ such that $f(T^n(x)) > 0$ and so $T^n(x) \in U$. But since U was chosen arbitrarily, we conclude that the orbit of x visits every open set in X; in other words, the orbit of x is dense in X. Since x was chosen arbitrarily, every orbit is dense, and thus the system T is minimal according to Theorem 1.5.4. □

We now revisit the rotations (sometimes called translations) of the torus. Recall that these rotations are a case of rotations of topological groups (see Subsection 1.6.1). According to Haar's theorem, there is, up to a positive multiplicative constant, a unique measure μ on the Borel subsets of a topological group G satisfying the following properties:

(a) μ is left-translation-invariant: $\mu(gS) = \mu(S)$ for every $g \in G$ and all Borel sets $S \subseteq G$.
(b) μ is finite on every compact set: $\mu(K) < \infty$ for all compact $K \subseteq G$.
(c) μ is outer regular on Borel sets $S \subseteq G$: $\mu(S) = \inf\{\mu(U) : S \subseteq U, U \text{ open}\}$.
(d) μ is inner regular on open sets $U \subseteq G$: $\mu(U) = \sup\{\mu(K) : K \subseteq U, K \text{ compact}\}$.

Such a measure is called a left Haar measure. As a consequence of the above properties, it also turns out that $\mu(U) > 0$ for every nonempty open subset $U \subseteq G$. In particular, if G is compact then $0 < \mu(G) < \infty$. Thus, we can uniquely specify a left Haar measure on G by adding the normalization condition $\mu(G) = 1$. This is obviously the case for the n-dimensional torus, where μ will be denoted by λ_n. This is the n-dimensional Lebesgue measure on the torus.

Proposition 8.2.43. *Let $L_\gamma : \mathbb{T}^n \to \mathbb{T}^n$ be a translation of the torus, where $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n) \in \mathbb{T}^n$. The following statements are equivalent:*
(a) *The numbers $1, \gamma_1, \gamma_2, \ldots, \gamma_n$ are linearly independent over $\mathbb{Q}$.*
(b) *L_γ is minimal.*
(c) *L_γ is transitive.*
(d) *L_γ is ergodic with respect to λ_n.*
(e) *L_γ is uniquely ergodic.*

Proof. Theorem 1.6.3 already asserted the equivalencies (a)⇔(b)⇔(c). It is also obvious that (e)⇒(d). Implication (d)⇒(c) follows from Theorem 8.2.27. It remains to prove that (a)⇒(e). To that end, assume that $1, \gamma_1, \ldots, \gamma_n$ are linearly independent over $\mathbb{Q}$. We will first show that if $f \circ L_\gamma = f$ for some $f \in L^2(\lambda_n)$, then f is λ_n-a. e. constant. It will then follow from Theorem 8.2.18 and Remark 8.2.19 that λ_n is ergodic with respect to L_γ. Consider the Fourier series representation of f:

$$f(x) = \sum_{k\in\mathbb{Z}^n} a_k e^{2\pi i\langle k,x\rangle}, \quad \text{where } \langle k,x\rangle = \sum_{j=1}^{n} k_j x_j.$$

Then

$$f \circ L_\gamma(x) = \sum_{k\in\mathbb{Z}^n} a_k e^{2\pi i\langle k,x+\gamma\rangle} = \sum_{k\in\mathbb{Z}^n} a_k e^{2\pi i\langle k,\gamma\rangle} e^{2\pi i\langle k,x\rangle}.$$

The above equalities are understood to hold in $L^2(\lambda_n)$ and hold only for λ_n-a. e. $x \in \mathbb{T}^n$. As $f \circ L_\gamma = f$, we deduce from the uniqueness of the Fourier series representation that

$$a_k e^{2\pi i\langle k,\gamma\rangle} = a_k, \quad \forall k \in \mathbb{Z}^n.$$

Hence, for each $k \in \mathbb{Z}^n$ we have $a_k = 0$ or $e^{2\pi i\langle k,\gamma\rangle} = 1$. The latter condition holds if and only if $\langle k,\gamma\rangle \in \mathbb{Z}$. As $1, \gamma_1, \ldots, \gamma_n$ are linearly independent over $\mathbb{Q}$, this happens if and only if $k_1 = k_2 = \cdots = k_n = 0$. Thus $f(x) = a_{(0,\ldots,0)}$ for λ_n-a. e. $x \in \mathbb{T}^n$. That is, f is λ_n-a. e. constant. This implies that λ_n is ergodic. It only remains to show that λ_n is the unique ergodic L_γ-invariant measure. Let $\varphi \in C(\mathbb{T}^n)$. By the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14), for λ_n-a. e. $x \in \mathbb{T}^n$ we have

$$\lim_{k\to\infty} \frac{1}{k} S_k\varphi(x) = \int_{\mathbb{T}^n} \varphi \, d\lambda_n. \tag{8.24}$$

Let $x_0 \in \mathbb{T}^n$ and $\varepsilon > 0$. Since φ is uniformly continuous on the compact metric space $\mathbb{T}^n$, there exists $\delta > 0$ such that

$$\|x' - x\| < \delta \Longrightarrow |\varphi(x') - \varphi(x)| < \varepsilon.$$

Furthermore, as $\operatorname{supp}(\lambda_n) = \mathbb{T}^n$, there exists $x_1 \in \mathbb{T}^n$ such that (8.24) holds for x_1 and $\|x_0 - x_1\| < \delta$. Bearing in mind that L_γ is an isometry, for all $k \in \mathbb{N}$ it follows that

$$\begin{aligned}\left|\frac{1}{k}S_k\varphi(x_0) - \int_{\mathbb{T}^n} \varphi\, d\lambda_n\right| &\le \left|\frac{1}{k}S_k\varphi(x_0) - \frac{1}{k}S_k\varphi(x_1)\right| + \left|\frac{1}{k}S_k\varphi(x_1) - \int_{\mathbb{T}^n} \varphi\, d\lambda_n\right| \\ &\le \frac{1}{k}\sum_{j=0}^{k-1}\left|\varphi(L_\gamma^j(x_0)) - \varphi(L_\gamma^j(x_1))\right| + \left|\frac{1}{k}S_k\varphi(x_1) - \int_{\mathbb{T}^n} \varphi\, d\lambda_n\right| \\ &< \frac{1}{k}\sum_{j=0}^{k-1}\varepsilon + \left|\frac{1}{k}S_k\varphi(x_1) - \int_{\mathbb{T}^n} \varphi\, d\lambda_n\right| \\ &= \varepsilon + \left|\frac{1}{k}S_k\varphi(x_1) - \int_{\mathbb{T}^n} \varphi\, d\lambda_n\right|.\end{aligned}$$

Letting $k \to \infty$, we deduce that $\left|\lim_{k\to\infty} \frac{1}{k}S_k\varphi(x_0) - \int_{\mathbb{T}^n} \varphi\, d\lambda_n\right| \le \varepsilon$. As $\varepsilon > 0$ was arbitrary, we conclude that $\lim_{k\to\infty} \frac{1}{k}S_k\varphi(x_0) = \int_{\mathbb{T}^n} \varphi\, d\lambda_n$. But x_0 was chosen arbitrarily in $\mathbb{T}^n$ and so L_γ is uniquely ergodic by Theorem 8.2.41. □

Remark 8.2.44.
(a) This proof in fact shows that any isometry T on a compact metrizable space X that admits an invariant probability measure which is ergodic and of full topological support, is uniquely ergodic.
(b) Being an isometry can be weakened by requiring only that the iterates $\{T^n\}_{n=1}^{\infty}$ form an equicontinuous family.

8.3 Mixing transformations

In the penultimate section of this chapter, we introduce various notions of mixing for measure-preserving dynamical systems. These should be contrasted with topological mixing, which was introduced in Section 1.5. These measure-theoretical mixing forms are stronger than ergodicity in the sense that they all imply ergodicity, and are important from a statistical viewpoint (for instance, for decay of correlations and similar questions).

8.3.1 Weak mixing

Definition 8.3.1. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. The system T is said to be *weakly mixing* if

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1}\left|\mu(T^{-j}(A) \cap B) - \mu(A)\mu(B)\right| = 0, \quad \forall A, B \in \mathcal{B}. \tag{8.25}$$

Like for ergodicity, to find out if a system is weakly mixing it suffices to check weak mixing on a semialgebra that generates the σ-algebra.

Lemma 8.3.2. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. If $\mathcal{B} = \sigma(\mathcal{S})$ for some semialgebra $\mathcal{S}$, then T is weakly mixing if and only if relation* (8.25) *holds for all $A, B \in \mathcal{S}$.*

Proof. The proof is nearly identical to that of Lemma 8.2.17. Simply replace the square brackets by absolute values. □

Weak mixing is a stronger property than ergodicity. This is not surprising if you compare the definition of weak mixing with the characterization of ergodicity given in Lemma 8.2.16. Nevertheless, we will give a more direct proof of that fact.

Lemma 8.3.3. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. If T is weakly mixing, then T is ergodic.*

Proof. Suppose by way of contradiction that T is weakly mixing but not ergodic. Thus there exists a completely T-invariant set $E \in \mathcal{B}$ with $\mu(E) > 0$ and $\mu(X \backslash E) > 0$. Then $T^{-j}(E) \cap (X \backslash E) = E \cap (X \backslash E) = \emptyset$ for all $j \in \mathbb{N}$. Setting $A = E$ and $B = X \backslash E$ in (8.25), we deduce that $\mu(E)\mu(X \backslash E) = 0$. So $\mu(E) = 0$ or $\mu(X \backslash E) = 0$. This contradiction shows that T is ergodic. □

The converse of this lemma is not true; that is to say, there exist dynamical systems that are ergodic but not weakly mixing. We now provide such an example.

Example 8.3.4. Let $\alpha \in \mathbb{R} \backslash \mathbb{Q}$ and consider again the map $T_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $T_\alpha(x) := \langle x + \alpha \rangle$, where $\langle r \rangle$ denotes the fractional part of r. This is the rotation of the unit circle $\mathbb{S}^1$ by the angle $2\pi\alpha$. We saw in Proposition 8.2.29 that this map is ergodic with respect to the Lebesgue measure λ on $\mathbb{S}^1$. We shall now show that it is not weakly mixing.

By Corollary 8.2.15, for any interval $I \subseteq \mathbb{S}^1$ and λ-almost every $x \in \mathbb{S}^1$,

$$\lim_{M \to \infty} \frac{1}{M} \#\{0 \le n < M : T_\alpha^n(x) \in I\} = \lambda(I).$$

In other words, for λ-almost every $x \in \mathbb{S}^1$, the sequence $(\langle x + n\alpha \rangle)_{n=1}^\infty$ is uniformly distributed in $\mathbb{S}^1$. It follows, upon rotating by $-x$, that the sequence $(\langle n\alpha \rangle)_{n=1}^\infty$ is uniformly distributed in $\mathbb{S}^1$. Let $A = B = (0, 1/2)$ and let $(n_i)_{i=1}^\infty$ be the subsequence of $\mathbb{N}$ such that $\langle n_i \alpha \rangle \in (0, 1/10)$. Then $T_\alpha^{-n_i}(A) \supseteq [0, 4/10]$ and hence

$$\lambda\big(T_\alpha^{-n_i}(A) \cap B\big) - \lambda(A)\lambda(B) \ge \lambda\big((0, 4/10]\big) - \big(\lambda\big((0, 1/2)\big)\big)^2 = \frac{4}{10} - \frac{1}{4} = \frac{3}{20}.$$

Consequently,

$$\begin{aligned}
\liminf_{N \to \infty} \frac{1}{n_N} \sum_{k=0}^{n_N} \big|\lambda\big(T_\alpha^{-k}(A) \cap B\big) - \lambda(A)\lambda(B)\big| &\ge \liminf_{N \to \infty} \frac{1}{n_N} \sum_{i=1}^{N} \big|\lambda\big(T_\alpha^{-n_i}(A) \cap B\big) - \lambda(A)\lambda(B)\big| \\
&\ge \liminf_{N \to \infty} \frac{1}{n_N} \cdot N \cdot \frac{3}{20}
\end{aligned}$$

$$\geq \frac{3}{20} \liminf_{M\to\infty} \frac{1}{M} \#\{0 \leq n < M : \langle n\alpha\rangle \in (0, 1/10)\}$$
$$= \frac{3}{20} \cdot \frac{1}{10} > 0.$$

Therefore, T_α cannot be weakly mixing.

The following lemma provides a characterization of weakly mixing dynamical systems.

Lemma 8.3.5. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. The following statements are equivalent:*
(a) *T is weakly mixing.*
(b) *For all $f, g \in L^2(\mu)$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \left| \int_X (f \circ T^j) \cdot g \, d\mu - \int_X f \, d\mu \int_X g \, d\mu \right| = 0.$$

Proof. That (b) implies (a) follows upon choosing the functions $f = \mathbb{1}_A$ and $g = \mathbb{1}_B$. For the converse, a straightforward argument involving approximation by simple functions is enough to complete the proof. We leave the details as an exercise. □

The following result will be used to give alternative formulations of weak mixing. A subset J of $\mathbb{Z}_+$ is said to have *density zero* if

$$\lim_{n\to\infty} \frac{\#(J \cap \{0, 1, \ldots, n-1\})}{n} = 0.$$

Theorem 8.3.6. *If $(a_n)_{n=0}^\infty$ is a bounded sequence in $\mathbb{R}$, then the following statements are equivalent:*
(a) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0$.
(b) *There exists a set $J \subseteq \mathbb{Z}_+$ of density zero such that* $\lim_{J\not\ni n\to\infty} a_n = 0$.
(c) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} a_i^2 = 0$.

Proof. We shall first prove that (a)⇔(b) and then (b)⇔(c).

[(a)⇒(b)] Suppose that $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0$. To lighten notation, let $c_J(n) = \#(J \cap \{0, 1, \ldots, n-1\})$. For each $k \in \mathbb{N}$, define

$$J_k := \left\{ i \geq 0 : |a_i| \geq \frac{1}{k} \right\}.$$

Then $(J_k)_{k=1}^\infty$ is an ascending sequence of sets. We claim that each J_k has density zero. Indeed, for each $k \in \mathbb{N}$,

$$\frac{1}{n} \sum_{i=0}^{n-1} |a_i| \geq \frac{1}{n} \sum_{\substack{i=0 \\ i \in J_k}}^{n-1} |a_i| \geq \frac{1}{n} \frac{1}{k} c_{J_k}(n).$$

Thus $\lim_{n\to\infty}\frac{1}{n}\frac{1}{k}\,c_{J_k}(n)=0$. This implies that $\lim_{n\to\infty}\frac{1}{n}\,c_{J_k}(n)=0$, that is, each J_k has density zero. Therefore, there exists a strictly increasing sequence $(\ell_k)_{k=1}^{\infty}$ in $\mathbb{N}$ such that for every $k\in\mathbb{N}$,

$$\frac{1}{n}c_{J_k}(n)<\frac{1}{k},\quad \forall n\geq \ell_k.$$

Set

$$J:=\bigcup_{k=1}^{\infty}J_k\cap[\ell_k,\ell_{k+1}).$$

We claim that J has density zero. Indeed, since the sets $(J_k)_{k=1}^{\infty}$ form an ascending sequence, for every $\ell_k\leq n<\ell_{k+1}$ we have

$$J\cap[0,n)\subseteq J_k\cap[0,n)$$

and so

$$\frac{1}{n}\,c_J(n)\leq\frac{1}{n}c_{J_k}(n)<\frac{1}{k}.$$

Letting $n\to\infty$ imposes $k\to\infty$ and hence $\lim_{n\to\infty}\frac{1}{n}c_J(n)=0$. So J has density zero, as claimed.

Moreover, if $n\geq\ell_k$ and $n\notin J$, then $n\notin J_k$ and thus $|a_n|<1/k$. Therefore,

$$\lim_{J\not\ni n\to\infty}|a_n|=0.$$

[(b)⇒(a)] For the opposite implication, suppose that $\lim_{J\not\ni n\to\infty}|a_n|=0$ for some set $J\subseteq\mathbb{Z}_+$ of density zero. Since $(a_n)_{n=0}^{\infty}$ is bounded, let $B\geq 0$ be such that $|a_n|\leq B$ for all $n\geq 0$. Fix $\varepsilon>0$. There exists $N(\varepsilon)\in\mathbb{N}$ such that $|a_n|<\varepsilon$ whenever $n\geq N(\varepsilon)$ and $n\notin J$, and such that $c_J(n)/n<\varepsilon$ for all $n\geq N(\varepsilon)$. Then, for all $n\geq N(\varepsilon)$, we have that

$$\frac{1}{n}\sum_{i=0}^{n-1}|a_i|=\frac{1}{n}\Bigg[\sum_{i\in J\cap\{0,1,\dots,n-1\}}|a_i|\quad+\quad\sum_{i\in\{0,1,\dots,n-1\}\setminus J}|a_i|\Bigg]<\frac{1}{n}\big[c_J(n)B+n\varepsilon\big]<(B+1)\varepsilon.$$

Thus

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}|a_i|=0.$$

[(b)⇔(c)] Using the fact that (b)⇔(a), it suffices to note that $\lim_{J\not\ni n\to\infty}a_i=0$ if and only if $\lim_{J\not\ni n\to\infty}a_i^2=0$. □

This theorem allows us to reformulate weak mixing in the following alternative ways.

Corollary 8.3.7. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. The following statements are equivalent:*
(a) *T is weakly mixing.*
(b) *For every $A, B \in \mathcal{B}$, there is a set $J(A, B) \subseteq \mathbb{Z}_+$ of density zero such that*

$$\lim_{J(A,B)\not\ni n\to\infty} \mu\big(T^{-n}(A) \cap B\big) = \mu(A)\mu(B).$$

(c) *For every $A, B \in \mathcal{B}$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \big[\mu\big(T^{-j}(A) \cap B\big) - \mu(A)\mu(B)\big]^2 = 0.$$

Proof. Apply Theorem 8.3.6 with $a_n = \mu(T^{-n}(A) \cap B) - \mu(A)\mu(B)$. □

Corollary 8.3.7 offers an intuitive view of weakly mixing systems: a system is weakly mixing if for every measurable set A, the events $T^{-n}(A)$, $n \in \mathbb{N}$, become asymptotically independent of any other measurable set B, as long as we overlook a few instances of time. The avoided times naturally depend on both A and B as well as T and μ.

Let us finish with the relation between a weakly mixing system and the product of that system with itself.

Theorem 8.3.8. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. The following statements are equivalent:*
(a) *T is weakly mixing.*
(b) *$T \times T$ is ergodic.*
(c) *$T \times T$ is weakly mixing.*

Proof. Let us first show that (a)⇒(c). To that end, let $A, B, C, D \in \mathcal{B}$ and, using Corollary 8.3.7, let J_1 and J_2 be sets of density zero such that

$$\lim_{J_1\not\ni n\to\infty} \mu\big(T^{-n}(A \cap B)\big) = \mu(A)\mu(B) \quad \text{and} \quad \lim_{J_2\not\ni n\to\infty} \mu\big(T^{-n}(C \cap D)\big) = \mu(C)\mu(D).$$

Then

$$\begin{aligned}
&\lim_{J_1\cup J_2\not\ni n\to\infty} (\mu \times \mu)\big((T \times T)^{-n}(A \times C) \cap (B \times D)\big) \\
&\quad = \lim_{J_1\cup J_2\not\ni n\to\infty} (\mu \times \mu)\big(\big(T^{-n}(A) \times T^{-n}(C)\big) \cap (B \times D)\big) \\
&\quad = \lim_{J_1\cup J_2\not\ni n\to\infty} (\mu \times \mu)\big(\big(T^{-n}(A) \cap B\big) \times \big(T^{-n}(C) \cap D\big)\big) \\
&\quad = \lim_{J_1\cup J_2\not\ni n\to\infty} \mu\big(T^{-n}(A) \cap B\big) \cdot \mu\big(T^{-n}(C) \cap D\big) \\
&\quad = \mu(A)\mu(B) \cdot \mu(C)\mu(D) = \mu(A)\mu(C) \cdot \mu(B)\mu(D) \\
&\quad = (\mu \times \mu)(A \times C) \cdot (\mu \times \mu)(B \times D).
\end{aligned}$$

Thanks to Theorem 8.3.6, we deduce that

$$\lim_{n\to\infty}\sum_{j=0}^{n-1}\big|(\mu\times\mu)\big((T\times T)^{-j}(A\times C)\cap(B\times D)\big)\\ -(\mu\times\mu)(A\times C)\cdot(\mu\times\mu)(B\times D)\big| = 0.$$

Since the collection of measurable rectangles $\{E\times F : E,F\in\mathcal{B}\}$ forms a semialgebra that generates $\mathcal{B}\times\mathcal{B}$, Lemma 8.3.2 allows us to conclude that $T\times T$ is weakly mixing.

That (c)⇒(b) is an immediate consequence of Lemma 8.3.3.

It only remains to show that (b)⇒(a). To that end, let $A,B\in\mathcal{B}$. We aim to show that $\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}[\mu(T^{-j}(A)\cap B)-\mu(A)\mu(B)]^2 = 0$. Applying Lemma 8.2.16 to $T\times T$ and the rectangles $A\times X$ and $B\times X$, we get

$$\begin{aligned}\frac{1}{n}\sum_{j=0}^{n-1}\mu\big(T^{-j}(A)\cap B\big) &= \frac{1}{n}\sum_{j=0}^{n-1}(\mu\times\mu)\big((T\times T)^{-j}(A\times X)\cap(B\times X)\big)\\ &\underset{n\to\infty}{\longrightarrow}(\mu\times\mu)(A\times X)\cdot(\mu\times\mu)(B\times X)=\mu(A)\mu(B).\end{aligned}$$

Applying the same lemma to the rectangles $A\times A$ and $B\times B$, we obtain

$$\begin{aligned}\frac{1}{n}\sum_{j=0}^{n-1}\big[\mu\big(T^{-j}(A)\cap B\big)\big]^2 &= \frac{1}{n}\sum_{j=0}^{n-1}(\mu\times\mu)\big((T\times T)^{-j}(A\times A)\cap(B\times B)\big)\\ &\underset{n\to\infty}{\longrightarrow}(\mu\times\mu)(A\times A)\cdot(\mu\times\mu)(B\times B)=\mu(A)^2\mu(B)^2.\end{aligned}$$

Thus

$$\begin{aligned}&\frac{1}{n}\sum_{j=0}^{n-1}\big[\mu\big(T^{-j}(A)\cap B\big)-\mu(A)\mu(B)\big]^2\\ &\quad=\frac{1}{n}\sum_{j=0}^{n-1}\Big(\big[\mu\big(T^{-j}(A)\cap B\big)\big]^2-2\mu\big(T^{-j}(A)\cap B\big)\mu(A)\mu(B)+\mu(A)^2\mu(B)^2\Big)\\ &\quad\underset{n\to\infty}{\longrightarrow}\mu(A)^2\mu(B)^2-2\mu(A)^2\mu(B)^2+\mu(A)^2\mu(B)^2=0.\end{aligned}$$

Therefore, T is weakly mixing according to Corollary 8.3.7. □

8.3.2 Mixing

We now investigate a stronger mixing form.

Definition 8.3.9. Let $T : X\to X$ be a measure-preserving dynamical system on a probability space $(X,\mathcal{B},\mu)$. The system T is said to be *mixing* if

$$\lim_{n\to\infty}\mu\big(T^{-n}(A)\cap B\big)=\mu(A)\mu(B),\quad \forall A,B\in\mathcal{B}. \tag{8.26}$$

Like for ergodicity and weak mixing, to ascertain whether a system is mixing it suffices to check that it is mixing on a semialgebra that generates the σ-algebra.

Lemma 8.3.10. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. If $\mathcal{B} = \sigma(\mathcal{S})$ for some semialgebra $\mathcal{S}$, then T is mixing if and only if relation* (8.26) *holds for all $A, B \in \mathcal{S}$.*

Proof. The proof, which goes along similar lines to that of Lemma 8.2.17, is left as an exercise. □

Lemma 8.3.11. *If $T : X \to X$ is a mixing transformation on a probability space $(X, \mathcal{B}, \mu)$, then T is weakly mixing (and therefore ergodic).*

Proof. This is immediate from the definitions of weak mixing and mixing. □

Just as was the case for weakly mixing systems, we have the following characterization of mixing systems.

Lemma 8.3.12. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. The following statements are equivalent:*

(a) *T is mixing.*

(b) *For all $f, g \in L^2(\mu)$,*

$$\lim_{n\to\infty} \int_X (f \circ T^n) \cdot g \, d\mu = \int_X f \, d\mu \int_X g \, d\mu.$$

Proof. The proof is almost identical to that of Lemma 8.3.5 and is left to the reader. □

Example 8.3.13. In Example 8.2.32, we proved that the shift map σ is ergodic with respect to the product measure μ_P. In particular, we showed that for every pair A, B of cylinder sets, $\mu_P(\sigma^{-n}(A) \cap B) = \mu_P(A)\mu_P(B)$ as long as n is large enough. It is thus clear that σ is mixing, according to Lemma 8.3.10.

Remark 8.3.14. There are several examples of dynamical systems which are weakly mixing but not mixing. For instance, Katok [32] showed that all *interval exchange transformations* are not mixing, whereas Avila and Forni [6] later proved that almost all of these transformations are weakly mixing. Interval exchange transformations are a very nice class of examples which were first introduced by Ja. G. Sinai in a series of lectures in Russian at Erivan State University in 1973 and were introduced in a published paper in English by Keane in [34]. These maps are simple to define, exhibit interesting ergodic properties, and turn up in many seemingly surprising areas of mathematics. The basic idea is the following: Partition a bounded interval $I \subseteq \mathbb{R}$ into finitely many subintervals and define a bijective map from I to I that is a translation on each subinterval. The idea is best grasped with the aid of an illustration. See Figure 8.1.

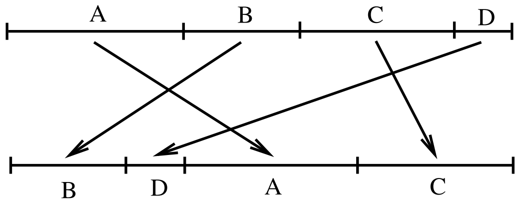

Figure 8.1: An example of an interval exchange transformation.

These maps are discontinuous at finitely many points. We will not delve into their dynamical properties. In addition to the papers mentioned above, the interested reader might consult [24, 45, 53, 74] and the references therein.

8.3.3 K-mixing

Before defining K-mixing, the reader who needs a quick refresher about the conditional expectation function with respect to the σ-algebra generated by a countable measurable partition is invited to consult Example A.1.62. As partitions are covers, the concepts, operations, and properties outlined in Section 7.1 will all be relevant here. It is worth noticing that all the operations introduced in that section result in countable measurable partitions. For instance, the join of two countable measurable partitions is a countable measurable partition; likewise, the preimage of a countable measurable partition is a countable measurable partition. Furthermore, note that for partitions, the relation $\prec$ is antisymmetric, that is, $\alpha \prec \beta \prec \alpha \iff \alpha = \beta$ (cf. Remark 7.1.6).

Definition 8.3.15. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and let α be a countable measurable partition of X. For every $n \in \mathbb{N}$, define

$$\alpha_n^\infty := \bigcup_{m>n} \alpha_n^m.$$

Given a probability measure μ on $(X, \mathcal{A})$, denote by $\sigma_c(\alpha_n^\infty)$ the completed σ-algebra generated by α_n^∞. (For more information about the completion of a σ-algebra, see Exercises 8.5.7–8.5.8.) The tail σ-algebra of α with respect to T is defined as

$$\mathrm{Tail}_T(\alpha) := \bigcap_{n=0}^{\infty} \sigma_c(\alpha_n^\infty).$$

Definition 8.3.16. A measurable transformation $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ is said to be K-*mixing* if

$$\lim_{n\to\infty} \sup_{A \in \sigma_c(\alpha_n^\infty)} \left|\mu(A \cap B) - \mu(A)\mu(B)\right| = 0$$

for every set $B \in \mathcal{A}$ and every finite measurable partition α of X.

The letter K is in honor of Kolmogorov, who first introduced this concept. K-mixing systems are simply referred to as K-*systems*. K-mixing is the strongest of all mixing properties discussed in this chapter.

Theorem 8.3.17. *Each* K-*system is mixing, and hence weakly mixing and ergodic.*

Proof. Suppose that $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ is a K-mixing transformation. Fix $A, B \in \mathcal{A}$ and consider the measurable partition $\alpha = \{A, X\backslash A\}$. Then $T^{-n}(A) \in \alpha_n^\infty$ for every $n \in \mathbb{N}$ and, therefore, by the K-mixing property,

$$\lim_{n\to\infty} \left|\mu(T^{-n}(A) \cap B) - \mu(A)\mu(B)\right| \le \lim_{n\to\infty} \sup_{F\in\alpha_n^\infty} \left|\mu(F \cap B) - \mu(F)\mu(B)\right| = 0.$$

Hence, T is mixing. □

We now give a characterization of K-systems which, as well as being interesting in its own right, will be used later to show that Rokhlin's natural extension of every metrically exact system is K-mixing.

Theorem 8.3.18. *Let $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ be a measurable transformation and $\mathcal{N}$ the σ-algebra of all sets of null or full μ-measure. Then T is* K-*mixing if and only if* $\mathrm{Tail}_T(\alpha) \subseteq \mathcal{N}$ *for every finite measurable partition α of X.*

Proof. Let $B \in \mathcal{A}$ and α a finite measurable partition of X. Suppose that $\mathrm{Tail}_T(\alpha) \subseteq \mathcal{N}$. Combining Example A.1.61 with Proposition A.1.60(f,e) reveals that

$$E(\mathbb{1}_B|\mathrm{Tail}_T(\alpha)) = E(E(\mathbb{1}_B|\mathcal{N}) \mid \mathrm{Tail}_T(\alpha)) = E(\mu(B) \mid \mathrm{Tail}_T(\alpha)) = \mu(B).$$

Fix $n \ge 0$. For every $A \in \sigma_c(\alpha_n^\infty)$, we have

$$\begin{aligned}
\left|\mu(A \cap B) - \mu(A)\mu(B)\right| &= \left|\int_A \mathbb{1}_B \, d\mu - \int_A \mu(B) \, d\mu\right| \\
&= \left|\int_A \left[E(\mathbb{1}_B|\sigma_c(\alpha_n^\infty)) - \mu(B)\right] d\mu\right| \\
&\le \int_X \left|E(\mathbb{1}_B|\sigma_c(\alpha_n^\infty)) - \mu(B)\right| d\mu.
\end{aligned}$$

So

$$\sup_{A\in\sigma_c(\alpha_n^\infty)} \left|\mu(A \cap B) - \mu(A)\mu(B)\right| \le \int_X \left|E(\mathbb{1}_B|\sigma_c(\alpha_n^\infty)) - \mu(B)\right| d\mu. \tag{8.27}$$

As $(\sigma_c(\alpha_n^\infty))_{n=0}^\infty$ is a descending sequence of σ-algebras whose intersection is $\mathrm{Tail}_T(\alpha)$, Theorem A.1.68 affirms that the sequence $(E(\mathbb{1}_B|\sigma_c(\alpha_n^\infty)))_{n=0}^\infty$ converges in $L^1(\mu)$ and pointwise μ-a. e. to $E(\mathbb{1}_B|\mathrm{Tail}_T(\alpha))$, which equals $\mu(B)$. It follows from (8.27) and Lebesgue's dominated convergence theorem (Theorem A.1.38) that

$$\begin{aligned}\lim_{n\to\infty}\sup_{A\in\sigma_c(\alpha_n^\infty)}\left|\mu(A\cap B)-\mu(A)\mu(B)\right| &\le \lim_{n\to\infty}\int_X\left|E(\mathbb{1}_B|\sigma_c(\alpha_n^\infty))-\mu(B)\right|d\mu\\ &=\int_X\left|\lim_{n\to\infty}E(\mathbb{1}_B|\sigma_c(\alpha_n^\infty))-\mu(B)\right|d\mu\\ &=0.\end{aligned}$$

Since B and α are arbitrary, T is K-mixing and one implication is proved.

In order to prove the converse, fix $C\in\mathrm{Tail}_T(\alpha)$. Then $C\in\sigma_c(\alpha_n^\infty)$ for every $n\ge 0$ and employing the definition of K-mixing with $B=C$, we obtain that

$$\left|\mu(C)-\mu(C)^2\right|=\left|\mu(C\cap C)-\mu(C)\mu(C)\right|\le\sup_{A\in\sigma_c(\alpha_n^\infty)}\left|\mu(A\cap C)-\mu(A)\mu(C)\right|\underset{n\to\infty}{\longrightarrow}0.$$

Hence, $\mu(C)=\mu(C)^2$ and it follows that either $\mu(C)=0$ or $\mu(C)=1$. So $\mathrm{Tail}_T(\alpha)\subseteq\mathcal{N}$. □

We will use this theorem in a moment. First, we recall the notion of Lebesgue space.

Definition 8.3.19. A Borel probability space $(X,\mathcal{B}(X),\mu)$ is said to be a *Lebesgue space* if X is a *Polish space* (i. e., X is completely metrizable and separable) and if μ is a complete measure.

Proposition 8.3.20. *Let $T : X\to X$ be a measure-preserving automorphism on a Lebesgue space $(X,\mathcal{B}(X),\mu)$. Assume that there exists a σ-algebra $\mathcal{B}\subseteq\mathcal{B}(X)$ with the following properties:*
(a) $T^{-1}(\mathcal{B})\subseteq\mathcal{B}$.
(b) $\sigma_c\big(\bigcup_{n=0}^\infty T^n(\mathcal{B})\big)=\mathcal{B}(X)$.
(c) $\bigcap_{n=0}^\infty T^{-n}(\mathcal{B})\subseteq\mathcal{N}$, *where $\mathcal{N}$ is the σ-algebra of all sets of null or full μ-measure.*

Then T is K*-mixing.*

Before proving this proposition, we need the following fairly nontrivial lemma, which can be found as Lemma 1 in Section 8 of Cornfeld, Fomin, and Sinai [16].

Lemma 8.3.21. *Let $T : X\to X$ be a measure-preserving automorphism of a Lebesgue space $(X,\mathcal{B}(X),\mu)$. If $\mathcal{A}$ and $\mathcal{B}$ are sub-σ-algebras of $\mathcal{B}(X)$ such that $\mathcal{A}\subseteq\sigma_c\big(\bigcup_{n=-\infty}^\infty T^n(\mathcal{B})\big)$, then* $\mathrm{Tail}_T(\mathcal{A})\subseteq\mathrm{Tail}_T(\mathcal{B})$, *where*

$$\mathrm{Tail}_T(\mathcal{C}):=\bigcap_{n=0}^\infty\sigma_c\left(\bigcup_{k=n}^\infty T^{-k}(\mathcal{C})\right)$$

for any sub-σ-algebra $\mathcal{C}$ of $\mathcal{B}(X)$.

Proof of Proposition 8.3.20. We will use Theorem 8.3.18 to establish the K-mixing property of T. Let α be a finite Borel partition of X. Then α generates a finite sub-σ-algebra $\mathcal{A}$ of $\mathcal{B}(X)$. By hypotheses (b) and (a), we have $\mathcal{A} \subseteq \sigma_c(\bigcup_{n=0}^{\infty} T^n(\mathcal{B})) = \sigma_c(\bigcup_{n=-\infty}^{\infty} T^n(\mathcal{B}))$. Using Lemma 8.3.21 and hypotheses (a) and (c), we obtain that

$$\mathrm{Tail}_T(\mathcal{A}) \subseteq \mathrm{Tail}_T(\mathcal{B}) = \bigcap_{n=0}^{\infty} \sigma_c\left(\bigcup_{k=n}^{\infty} T^{-k}(\mathcal{B})\right) = \bigcap_{n=0}^{\infty} \sigma_c(T^{-n}(\mathcal{B})) \subseteq \mathcal{N}.$$

The result follows from Theorem 8.3.18. □

We will see later that all two-sided Bernoulli shifts are K-mixing (cf. Example 8.1.14 and Exercise 8.5.22). This will be shown in Subsection 13.9.5, where we will treat the much more general case of Gibbs measures for Hölder continuous potentials. The first step in this direction is provided in the following section. In the meantime, see Exercise 8.5.23 for a direct proof based on Proposition 8.3.20.

8.4 Rokhlin's natural extension

Let $T : X \to X$ be a surjective measure-preserving dynamical system on a Lebesgue space $(X, \mathcal{F}, \mu)$ (see Definition 8.3.19). Consider the set of sequences

$$\widetilde{X} := \{(x_n)_{n=0}^{\infty} \in X^{\infty} : T(x_{n+1}) = x_n, \ \forall n \geq 0\} \subseteq X^{\infty}.$$

For every $k \geq 0$, let $\pi_k : \widetilde{X} \to X$ denote the projection onto the kth coordinate of $\widetilde{X}$, that is,

$$\pi_k((x_n)_{n=0}^{\infty}) := x_k.$$

Observe that $T \circ \pi_{k+1} = \pi_k$ for all $k \geq 0$. Equip the set $\widetilde{X}$ with the smallest σ-algebra $\widetilde{\mathcal{F}}$ that makes every projection $\pi_k : \widetilde{X} \to X$ continuous.

Note that in this construction the surjectivity of T is not really an essential assumption. Indeed, since X is a Lebesgue space, the sets $T^n(X)$, $n \geq 0$, are measurable. Because μ is T-invariant, it turns out that $\mu(T^n(X)) = 1$ for all $n \geq 0$. As the sets $T^n(X)$, $n \geq 0$, form a descending sequence, it follows that $\mu(\bigcap_{n=0}^{\infty} T^n(X)) = 1$. Finally, the map $T : \bigcap_{n=0}^{\infty} T^n(X) \to \bigcap_{n=0}^{\infty} T^n(X)$ is clearly surjective.

Definition 8.4.1. *Rokhlin's natural extension of* T is the measurable transformation $\widetilde{T} : \widetilde{X} \to \widetilde{X}$ defined by

$$\widetilde{T}((x_n)_{n=0}^{\infty}) := (T(x_0), x_0, x_1, x_2, \ldots).$$

Theorem 8.4.2. *Rokhlin's natural extension has the following properties:*

(a) *The transformation* $\widetilde{T} : \widetilde{X} \to \widetilde{X}$ *is invertible and its inverse* $\widetilde{T}^{-1} : \widetilde{X} \to \widetilde{X}$ *is (the restriction of) the left shift map*

$$\widetilde{T}^{-1}\big((x_n)_{n=0}^{\infty}\big) := (x_{n+1})_{n=0}^{\infty}.$$

(b) *For each* $n \geq 0$, *the following diagram commutes:*

$$\begin{array}{ccc} \widetilde{X} & \xrightarrow{\widetilde{T}} & \widetilde{X} \\ {\scriptstyle \pi_n}\big\downarrow & & \big\downarrow{\scriptstyle \pi_n} \\ X & \xrightarrow[T]{} & X \end{array}$$

(c) *There exists a unique probability measure* $\widetilde{\mu}$ *on the space* $(\widetilde{X}, \widetilde{\mathcal{F}})$ *such that*

$$\widetilde{\mu} \circ \pi_n^{-1} = \mu, \quad \forall n \geq 0.$$

(d) *The probability measure* $\widetilde{\mu}$ *is* $\widetilde{T}$*-invariant.*

Proof. Proof of properties (a) and (b) is left as an exercise. Property (c) follows directly from the Daniel–Kolmogorov consistency theorem (see Theorem 3.6.4 in Parthasarathy [55]). Regarding property (d), let $A \in \mathcal{F}$. For every $n \geq 0$, it follows from (b) and (c) that

$$\begin{aligned} \widetilde{\mu} \circ \widetilde{T}^{-1}\big(\pi_n^{-1}(A)\big) &= \widetilde{\mu} \circ (\pi_n \circ \widetilde{T})^{-1}(A) = \widetilde{\mu} \circ (T \circ \pi_n)^{-1}(A) \\ &= \widetilde{\mu} \circ \pi_n^{-1} \circ T^{-1}(A) = \mu \circ T^{-1}(A) \\ &= \mu(A) = \widetilde{\mu}\big(\pi_n^{-1}(A)\big). \end{aligned}$$

The family $\{\pi_n^{-1}(A) : A \in \mathcal{F}, n \geq 0\}$ forms a π-system that generates $\widetilde{\mathcal{F}}$. It ensues from Lemma A.1.26 that $\widetilde{\mu} \circ \widetilde{T}^{-1} = \widetilde{\mu}$. □

This theorem sometimes allows us to replace the μ-measure-preserving dynamical system T, which is not necessarily invertible, with the $\widetilde{\mu}$-measure-preserving automorphism $\widetilde{T} : \widetilde{X} \to \widetilde{X}$. This turns out to be of great advantage in some proofs, since dealing with invertible transformations is frequently easier than dealing with noninvertible ones. Natural extensions share many properties with their original maps. An example of this is given in the following theorem.

Theorem 8.4.3. *The natural extension measure* $\widetilde{\mu}$ *on* $\widetilde{X}$ *from Theorem* 8.4.2 *is ergodic with respect to* $\widetilde{T}$ *if and only if the measure* μ *is ergodic with respect to* T.

Proof. Suppose first that μ is not ergodic with respect to $T : X \to X$. Then there exists a set $A \in \mathcal{F}$ such that $T^{-1}(A) = A$ and $0 < \mu(A) < 1$. It follows from Theorem 8.4.2(c) that $\widetilde{\mu}(\pi_0^{-1}(A)) = \mu(A) \in (0, 1)$. Furthermore, it ensues from Theorem 8.4.2(b) that

$$\widetilde{T}^{-1}\big(\pi_0^{-1}(A)\big) = (\pi_0 \circ \widetilde{T})^{-1}(A) = (T \circ \pi_0)^{-1}(A) = \pi_0^{-1}\big(T^{-1}(A)\big) = \pi_0^{-1}(A).$$

Therefore, $\widetilde{\mu}$ is not ergodic with respect to $\widetilde{T}$.

Now assume that $T : X \to X$ is ergodic with respect to μ. We want to show that $\widetilde{T} : \widetilde{X} \to \widetilde{X}$ is then ergodic with respect to $\widetilde{\mu}$. Let $F \in L^1(\widetilde{X}, \widetilde{\mathcal{F}}, \widetilde{\mu})$ be a $\widetilde{T}$-invariant function. According to Theorem 8.2.18, it suffices to demonstrate that F is $\widetilde{\mu}$-a. e. constant.

As T is ergodic, the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14) yields that (to lighten notation, we use $\mu(g) := \int_X g\, d\mu$)

$$\lim_{k\to\infty} \left\| \frac{1}{k} \sum_{j=0}^{k-1} g \circ T^j - \mu(g) \right\|_{L^1(\mu)} = 0$$

for every function $g \in L^1(X, \mathcal{F}, \mu)$. Invoking Theorem 8.4.2, this implies that

$$\lim_{k\to\infty} \left\| \frac{1}{k} \sum_{j=0}^{k-1} G \circ \widetilde{T}^j - \widetilde{\mu}(G) \right\|_{L^1(\widetilde{\mu})} = 0 \tag{8.28}$$

for every $G \in L^1(\widetilde{X}, \widetilde{\mathcal{F}}, \widetilde{\mu})$ of the form $g \circ \pi_n$, where $g \in L^1(X, \mathcal{F}, \mu)$ and $n \ge 0$.

Fix $n \ge 0$ momentarily. The function $E(F|\widetilde{\mathcal{F}}_n)$ depends only on the nth coordinate of a point in $\widetilde{X}$ and can thus be expressed as $f_n \circ \pi_n$ for some $f_n \in L^1(X, \mathcal{F}, \mu)$. Setting $g = f_n$ and $G = f_n \circ \pi_n = E(F|\widetilde{\mathcal{F}}_n)$, it follows that $\widetilde{\mu}(G) = \widetilde{\mu}(F)$ and from (8.28) that

$$\lim_{k\to\infty} \left\| \frac{1}{k} \sum_{j=0}^{k-1} E(F|\widetilde{\mathcal{F}}_n) \circ \widetilde{T}^j - \widetilde{\mu}(F) \right\|_{L^1(\widetilde{\mu})} = 0. \tag{8.29}$$

Now, for every $n \ge 0$, let $\widetilde{\mathcal{F}}_n := \pi_n^{-1}(\mathcal{F})$. Since $T \circ \pi_{n+1} = \pi_n$, it turns out that

$$\widetilde{\mathcal{F}}_{n+1} = \pi_{n+1}^{-1}(\mathcal{F}) \supseteq \pi_{n+1}^{-1}\big(T^{-1}(\mathcal{F})\big) = (T \circ \pi_{n+1})^{-1}(\mathcal{F}) = \pi_n^{-1}(\mathcal{F}) = \widetilde{\mathcal{F}}_n.$$

Thus $(\widetilde{\mathcal{F}}_n)_{n=0}^{\infty}$ is an ascending sequence of sub-σ-algebras of $\widetilde{\mathcal{F}}$. By definition, $\widetilde{\mathcal{F}}$ is the σ-algebra generated by that sequence and we know that $F = E(F|\widetilde{\mathcal{F}})$. The martingale convergence theorem (Theorem A.1.67) then affirms that

$$\lim_{n\to\infty} \|F - E(F|\widetilde{\mathcal{F}}_n)\|_{L^1(\widetilde{\mu})} = 0. \tag{8.30}$$

But for every $j \ge 0$ and every $n \in \mathbb{N}$, it ensues from the $\widetilde{T}$-invariance of $\widetilde{\mu}$ that

$$\|F \circ \widetilde{T}^j - E(F|\widetilde{\mathcal{F}}_n) \circ \widetilde{T}^j\|_{L^1(\widetilde{\mu})} = \|F - E(F|\widetilde{\mathcal{F}}_n)\|_{L^1(\widetilde{\mu})}. \tag{8.31}$$

Fix $\varepsilon > 0$. By virtue of (8.30)–(8.31), there exists $N \in \mathbb{N}$ such that

$$\|F \circ \widetilde{T}^j - E(F|\widetilde{\mathcal{F}}_n) \circ \widetilde{T}^j\|_{L^1(\widetilde{\mu})} \le \varepsilon$$

for every $j \ge 0$ and every $n \ge N$. Therefore, by the triangle inequality,

$$\left\| \frac{1}{k} \sum_{j=0}^{k-1} F \circ \widetilde{T}^j - \frac{1}{k} \sum_{j=0}^{k-1} E(F|\widetilde{\mathcal{F}}_n) \circ \widetilde{T}^j \right\|_{L^1(\widetilde{\mu})} \le \varepsilon$$

for every $k \in \mathbb{N}$ and every $n \geq N$. Given that F is $\widetilde{T}$-invariant, this reduces to

$$\left\| F - \frac{1}{k} \sum_{j=0}^{k-1} E(F|\widetilde{\mathcal{F}}_n) \circ \widetilde{T}^j \right\|_{L^1(\widetilde{\mu})} \leq \varepsilon$$

for every $k \in \mathbb{N}$ and every $n \geq N$.

Fixing $n \geq N$ and letting $k \to \infty$, we deduce from this and (8.29) that $\|F - \widetilde{\mu}(F)\|_{L^1(\widetilde{\mu})} \leq \varepsilon$. Therefore $F = \widetilde{\mu}(F)$ in $L^1(\widetilde{\mu})$. Consequently, $F = \widetilde{\mu}(F)$ $\widetilde{\mu}$-almost everywhere. That is, F is $\widetilde{\mu}$-a. e. constant. □

We now introduce the concept of metric exactness.

Definition 8.4.4. A measure-preserving dynamical system $T : X \to X$ on a Lebesgue space $(X, \mathcal{F}, \mu)$ is said to be *metrically exact* if for each $A \in \mathcal{F}$ such that $\mu(A) > 0$ we have

$$\lim_{n\to\infty} \mu\big(T^n(A)\big) = 1.$$

Note that each set $T^n(A)$ is measurable since T is a measurable transformation of a Lebesgue space.

Metric exactness of a system can be characterized in terms of the tail σ-algebra of the system.

Proposition 8.4.5. *Let $T : X \to X$ be a measure-preserving dynamical system on a Lebesgue space $(X, \mathcal{F}, \mu)$. Then T is metrically exact if and only if the tail σ-algebra* $\mathrm{Tail}_T(\mathcal{F}) = \bigcap_{n=0}^{\infty} T^{-n}(\mathcal{F})$ *is contained in the σ-algebra $\mathcal{N}$ of all sets of null or full μ-measure.*

Proof. Suppose that T is metrically exact and let $F \in \mathrm{Tail}_T(\mathcal{F})$. By definition of the tail σ-algebra, there exists a sequence of sets $(F_n)_{n=0}^{\infty}$ in $\mathcal{F}$ such that $F = T^{-n}(F_n)$ for each $n \geq 0$. Suppose that $\mu(F) > 0$. Then

$$1 = \lim_{n\to\infty} \mu\big(T^n(F)\big) = \lim_{n\to\infty} \mu\big(T^n\big(T^{-n}(F_n)\big)\big) = \lim_{n\to\infty} \mu(F_n) = \lim_{n\to\infty} \mu \circ T^{-n}(F_n) = \mu(F).$$

Thus $\mathrm{Tail}_T(\mathcal{F})$ consists only of sets of measure zero and one. This proves one implication.

Now suppose that $\mathrm{Tail}_T(\mathcal{F}) \subseteq \mathcal{N}$. Fix $F \in \mathcal{F}$ with $\mu(F) > 0$. For every $k \geq 0$, consider the measurable sets

$$F_k := \bigcup_{n=k}^{\infty} T^{-n}\big(T^n(F)\big) \supseteq F \quad \text{and} \quad F_\infty := \bigcap_{k=0}^{\infty} F_k \supseteq F.$$

We claim that $F_\infty \in \mathrm{Tail}_T(\mathcal{F})$. Indeed, by definition, the sequence of sets $(F_k)_{k=0}^{\infty}$ is descending and, therefore,

$$\mu(F_\infty) = \lim_{k\to\infty} \mu(F_k) \quad \text{and} \quad F_\infty = \bigcap_{k=j}^{\infty} F_k, \quad \forall j \geq 0.$$

If $k \geq j \geq 0$, then

$$F_k = \bigcup_{n=k}^{\infty} T^{-n}(T^n(F)) = \bigcup_{i=0}^{\infty} T^{-k}(T^{-i}(T^{k+i}(F))) = T^{-j}\left(T^{-(k-j)}\left(\bigcup_{i=0}^{\infty} T^{-i}(T^{k+i}(F))\right)\right).$$

To shorten notation, let $A_{j,k} := T^{-(k-j)}(\bigcup_{i=0}^{\infty} T^{-i}(T^{k+i}(F)))$. Then

$$F_\infty = \bigcap_{k=j}^{\infty} T^{-j}(A_{j,k}) = T^{-j}\left(\bigcap_{k=j}^{\infty} A_{j,k}\right).$$

Hence, for every $j \geq 0$, it follows that $F_\infty \in T^{-j}(\mathcal{F})$, or, equivalently, $F_\infty \in \mathrm{Tail}_T(\mathcal{F})$. Since $F_\infty \supseteq F$ and $\mu(F) > 0$, our hypothesis that $\mathrm{Tail}_T(\mathcal{F}) \subseteq \mathcal{N}$ implies $\mu(F_\infty) = 1$. So

$$\mu(F_0) = 1. \tag{8.32}$$

However, as $T^{-1}(T(A)) \supseteq A$ for every subset A of X, we observe that

$$T^{-(n+1)}(T^{n+1}(F)) = T^{-n}(T^{-1}(T(T^n(F)))) \supseteq T^{-n}(T^n(F)),$$

that is, the sequence of sets $(T^{-n}(T^n(F)))_{n=0}^{\infty}$ is ascending to their union F_0. Using this, (8.32) and the T-invariance of μ, we deduce that

$$1 = \mu(F_0) = \lim_{n\to\infty} \mu(T^{-n}(T^n(F))) = \lim_{n\to\infty} \mu(T^n(F)).$$

As $F \in \mathcal{F}$ was chosen arbitrarily, the transformation T is metrically exact. □

We can now demonstrate that Rokhlin's natural extension of any metrically exact system is K-mixing.

Theorem 8.4.6. *Let $T : X \to X$ be a measure-preserving dynamical system on a Lebesgue space $(X, \mathcal{F}, \mu)$. If T is metrically exact, then Rokhlin's natural extension $\widetilde{T} : \widetilde{X} \to \widetilde{X}$ is* K*-mixing.*

Proof. Let

$$\mathcal{B} := \pi_0^{-1}(\mathcal{F}) \subseteq \widetilde{\mathcal{F}}.$$

We shall verify that all the hypotheses of Proposition 8.3.20 hold. First,

$$\widetilde{T}^{-1}(\mathcal{B}) = \widetilde{T}^{-1}(\pi_0^{-1}(\mathcal{F})) = (\pi_0 \circ \widetilde{T})^{-1}(\mathcal{F}) = (T \circ \pi_0)^{-1}(\mathcal{F}) = \pi_0^{-1}(T^{-1}(\mathcal{F})) \subseteq \pi_0^{-1}(\mathcal{F}) = \mathcal{B}.$$

So condition (a) of Proposition 8.3.20 is satisfied. In order to show that condition (b) holds, note that $\widetilde{T}^n(\mathcal{B}) = \widetilde{T}^n \circ \pi_0^{-1}(\mathcal{F}) = \pi_n^{-1}(\mathcal{F})$ for all $n \geq 0$. Recall also that $\widetilde{\mathcal{F}}$ is the smallest σ-algebra containing all of the σ-algebras $\pi_n^{-1}(\mathcal{F})$. Thus condition (b) is satisfied. Finally, with $\mathcal{N}$ (resp., $\widetilde{\mathcal{N}}$) denoting the sub-σ-algebra consisting of the null and full μ-measure sets (resp., $\tilde{\mu}$-measure sets), we obtain that

$$\begin{aligned}\bigcap_{n=0}^{\infty} \widetilde{T}^{-n}(\mathcal{B}) &= \bigcap_{n=0}^{\infty} \widetilde{T}^{-n}(\pi_0^{-1}(\mathcal{F})) = \bigcap_{n=0}^{\infty} (\pi_0 \circ \widetilde{T}^n)^{-1}(\mathcal{F}) \\ &= \bigcap_{n=0}^{\infty} (T^n \circ \pi_0)^{-1}(\mathcal{F}) = \bigcap_{n=0}^{\infty} \pi_0^{-1} \circ T^{-n}(\mathcal{F}) \\ &= \pi_0^{-1}\left(\bigcap_{n=0}^{\infty} T^{-n}(\mathcal{F})\right) = \pi_0^{-1}(\mathrm{Tail}_T(\mathcal{F})) \subseteq \pi_0^{-1}(\mathcal{N}) \subseteq \widetilde{\mathcal{N}},\end{aligned}$$

where the first set inclusion comes from Proposition 8.4.5 and the second from Theorem 8.4.2(c). So condition (c) holds. Apply Proposition 8.3.20 to conclude. □

Finally, we provide an explicit description of the Rokhlin's natural extensions of a very important class of noninvertible measure-preserving dynamical systems, namely the one-sided Bernoulli shifts with finite sets of states, which were introduced in Example 8.1.14.

Let E be a finite set. Define the map $h : \widetilde{E^{\mathbb{N}}} \to E^{\mathbb{Z}}$ by

$$(h(\omega))_n = \begin{cases} (\omega_0)_n & \text{if } n \geq 0 \\ (\omega_{-n})_0 & \text{if } n < 0 \end{cases}$$

A straightforward inspection shows that h is bijective and that the following diagram commutes:

$$\begin{array}{ccc} \widetilde{E^{\mathbb{N}}} & \xrightarrow{\tilde{\sigma}} & \widetilde{E^{\mathbb{N}}} \\ {\scriptstyle h}\big\downarrow & & \big\downarrow{\scriptstyle h} \\ E^{\mathbb{Z}} & \xrightarrow[\sigma]{} & E^{\mathbb{Z}} \end{array}$$

In addition, if $\widetilde{E^{\mathbb{N}}}$ and $E^{\mathbb{Z}}$ are endowed with their respective product (Tychonov) topologies, then the map h is a homeomorphism, and thus is a measurable isomorphism if $\widetilde{E^{\mathbb{N}}}$ and $E^{\mathbb{Z}}$ are equipped with the corresponding Borel σ-algebras.

Furthermore, if E has at least two elements and $P : E \to [0, 1]$ is a probability vector, let μ_P^+ be the corresponding one-sided Bernoulli measure on $E^{\mathbb{N}}$ introduced in Example 8.1.14 and denoted there just by μ_P. Likewise, let μ_P be the two-sided Bernoulli measure on $E^{\mathbb{Z}}$ introduced in Exercise 8.5.22.

For every $k \leq 0$ and $n \geq 0$, the cylinder $[\omega_k\omega_{k+1} \dots \omega_{-1}\omega_0\omega_1 \dots \omega_n] \subseteq E^{\mathbb{Z}}$ satisfies

$$h^{-1}([\omega_k \dots \omega_n]) = [\omega_0 \dots \omega_n] \times [\omega_{-1}] \times [\omega_{-2}] \times \dots \times [\omega_k] \times \prod_{j=-k+1}^{\infty} E.$$

Consequently,

$$\widetilde{\mu_P^+}(h^{-1}([\omega_k \dots \omega_n])) = \prod_{i=k}^{n} P_i = \mu_P([\omega_k \dots \omega_n]).$$

Since all such cylinders form a π-system generating the Borel σ-algebra on $E^{\mathbb{Z}}$, we conclude that

$$\widetilde{\mu_P^+} \circ h^{-1} = \mu_P.$$

We have therefore proved the following.

Theorem 8.4.7. *If E is a finite set having at least two elements and $P : E \to [0,1]$ is a probability vector, then the Rokhlin's natural extension of the one-sided Bernoulli shift $(\sigma : E^{\mathbb{N}} \to E^{\mathbb{N}}, \mu_P)$ is metrically (i. e. measure-theoretically) isomorphic to the two-sided Bernoulli shift $(\sigma : E^{\mathbb{Z}} \to E^{\mathbb{Z}}, \mu_P)$.*

8.5 Exercises

Note: Exercises 8.5.1–8.5.19 pertain to measure theory. They use the terminology and notation introduced in Appendix A. The reader who would rather just concentrate on ergodic theory may skip those exercises.

Exercise 8.5.1. Let X be a set and $\mathcal{C}$ be a finite collection of subsets of X.

(a) Show that the algebra $\mathcal{A}(\mathcal{C})$ generated by $\mathcal{C}$ is finite.

Hint: First, identify the elements of $\mathcal{A}(\mathcal{C})$ when $\mathcal{C}$ consists of two disjoint sets. Generalize your argument to the case in which $\mathcal{C}$ consists of a finite number of disjoint sets. Then reduce the general case of a finite collection $\mathcal{C}$ to the case of an equivalent finite collection of disjoint sets.

(b) Deduce that $\sigma(\mathcal{C}) = \mathcal{A}(\mathcal{C})$.

Exercise 8.5.2. Let X be a set and $Y \subseteq X$. What is $\mathcal{A}(\{Y\})$? And $\sigma(\{Y\})$?

Exercise 8.5.3. Let X be an infinite set. Show that

$$\{Y \subseteq X : \text{either } Y \text{ or } X \setminus Y \text{ is finite}\}$$

is an algebra but not a σ-algebra on X.

Exercise 8.5.4. Let X be an uncountable set. Show that

$$\{Y \subseteq X : \text{either } Y \text{ or } X \setminus Y \text{ is countable}\}$$

is a σ-algebra on X. It is often called the *countable-cocountable* σ-algebra.

Exercise 8.5.5. Let X be a set and $Y \subseteq X$. If $\mathcal{B}$ is a σ-algebra on X, prove that

$$\mathcal{B}|_Y = \{B \cap Y : B \in \mathcal{B}\}$$

is a σ-algebra on Y.

Exercise 8.5.6. Let $(X, \mathcal{A}, \mu)$ be a measure space. Show that

$$\{A \in \mathcal{A} : \mu(A) = 0 \text{ or } \mu(X \backslash A) = 0\}$$

is a sub-σ-algebra of $\mathcal{A}$.

Exercise 8.5.7. Let $(X, \mathcal{B}, \mu)$ be a measure space and

$$\mathcal{N} := \{N \in \mathcal{B} \mid \mu(N) = 0\}$$

be the collection of all sets of measure zero, sometimes called null sets. Define

$$\overline{\mathcal{N}} := \{\overline{N} \subseteq X \mid \exists\, N \in \mathcal{N} \text{ such that } \overline{N} \subseteq N\} = \bigcup_{N \in \mathcal{N}} \mathcal{P}(N).$$

The completion of $(X, \mathcal{B}, \mu)$ is the measure space $(X, \overline{\mathcal{B}}, \overline{\mu})$, where

$$\overline{\mathcal{B}} := \{B \cup \overline{N} \,|B \in \mathcal{B},\ \overline{N} \in \overline{\mathcal{N}}\}$$

and

$$\overline{\mu}(\overline{B}) = \mu(B) \text{ whenever } \overline{B} = B \cup \overline{N} \text{ for some } B \in \mathcal{B} \text{ and } \overline{N} \in \overline{\mathcal{N}}.$$

(a) Prove that the space $(X, \overline{\mathcal{B}}, \overline{\mu})$ is well-defined (namely, that $\overline{\mathcal{B}}$ is a σ-algebra on X and that $\overline{\mu}$ is well-defined).
(b) Show that $(X, \overline{\mathcal{B}}, \overline{\mu})$ is an extension of the space $(X, \mathcal{B}, \mu)$ (i. e., $\overline{\mathcal{B}} \supseteq \mathcal{B}$ and $\overline{\mu} = \mu$ on $\mathcal{B}$).
(c) Observe that the space $(X, \overline{\mathcal{B}}, \overline{\mu})$ is complete.

Exercise 8.5.8. Let $(X, \mathcal{B}, \mu)$ be a measure space and let

$$\mathcal{B}^* := \{E \subseteq X \mid \exists\, A, B \in \mathcal{B} \text{ such that } A \subseteq E \subseteq B \text{ and } \mu(B \backslash A) = 0\}$$

be the collection of all subsets of X that are squeezed by some measurable sets whose difference is of measure zero. Define

$$\mu^*(E) = \mu(A) \text{ whenever } \exists\, A, B \in \mathcal{B} \text{ such that } A \subseteq E \subseteq B \text{ and } \mu(B \backslash A) = 0.$$

Prove that $(X, \mathcal{B}^*, \mu^*)$ is the completion of $(X, \mathcal{B}, \mu)$ (cf. Exercise 8.5.7).

Exercise 8.5.9. Show that Lemma A.1.26 does not hold for infinite measures in general.

Hint: Consider the family of all Borel subsets of $\mathbb{R}$ which do not have 0 for element.

Exercise 8.5.10. In this exercise, we look at the set-theoretic properties of the symmetric difference operation. Let X be a set. Let $A, B, C \subseteq X$. Let $T : X \to X$ be a map. Prove the following statements:

(a) $A \triangle B = B \triangle A$.
(b) If $A \cap B = \emptyset$ then $A \triangle B = A \cup B$.
(c) $(X \backslash A) \triangle (X \backslash B) = A \triangle B$.
(d) $T^{-1}(A \triangle B) = T^{-1}(A) \triangle T^{-1}(B)$.
(e) $A \triangle C \subseteq (A \triangle B) \cup (B \triangle C)$.
This statement generalizes to any finite number of intermediaries, that is, $A_n \triangle A_0 \subseteq \bigcup_{k=0}^{n-1}(A_{k+1} \triangle A_k)$.
(f) $A \triangle (B \cup C) \subseteq (A \triangle B) \cup (A \triangle C)$.
More generally, $(\bigcup_{i\in I} A_i) \triangle (\bigcup_{i\in I} B_i) \subseteq \bigcup_{i\in I}(A_i \triangle B_i)$ for any index set I.
(g) $A \triangle (B \cap C) \subseteq (A \triangle B) \cup (A \triangle C)$.
More generally, $(\bigcap_{i\in I} A_i) \triangle (\bigcap_{i\in I} B_i) \subseteq \bigcup_{i\in I}(A_i \triangle B_i)$ for any index set I.

Exercise 8.5.11. Let $(X, \mathcal{A}, \mu)$ be a probability space and $A, B \in \mathcal{A}$. Prove the following statements:

(a) $|\mu(A) - \mu(B)| \leq \mu(A \triangle B)$.
(b) If $\mu(A \triangle B) = 0$ then $\mu(A) = \mu(B)$.

Exercise 8.5.12. Let $(X, \mathcal{A}, \mu)$ be a probability space and $A, B \in \mathcal{A}$. Prove that $\mu(A \cap B) \geq \mu(A) + \mu(B) - 1$.

Exercise 8.5.13. Let $(X, \mathcal{A}, \mu)$ be a measure space. Show that if $(f_n)_{n=1}^{\infty}$ is a sequence of nonnegative measurable functions, then

$$\int_X \sum_{n=1}^{\infty} f_n \, d\mu = \sum_{n=1}^{\infty} \int_X f_n \, d\mu.$$

Exercise 8.5.14. The purpose of this exercise is to establish that a sequence of L^1 functions that converges pointwise need not converge in L^1. Construct a sequence of functions $f_n : [0,1] \to [0,\infty)$, $n \in \mathbb{N}$, with the following properties:

(a) Each function is continuous.
(b) The sequence converges pointwise to the constant function 0.
(c) $\int_{[0,1]} f_n \, d\lambda = 1$ for all $n \in \mathbb{N}$, where λ is the Lebesgue measure on $[0,1]$.

Deduce that the sequence does not converge in $L^1([0,1], \mathcal{B}([0,1]), \lambda)$.

Exercise 8.5.15. Find a sequence $(f_n)_{n=1}^{\infty}$ of $L^1(X, \mathcal{A}, \mu)$ functions with the following properties:

(a) The sequence converges pointwise to a function f.
(b) $\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu$.
(c) $\lim_{n\to\infty} \int_X |f_n| \, d\mu \neq \int_X |f| \, d\mu$.

Exercise 8.5.16. Show that on finite measure spaces Theorem A.1.41 is a generalization of Lebesgue's dominated convergence theorem (Theorem A.1.38).

Exercise 8.5.17. Construct a sequence of integrable functions that converges in L^1 to an integrable function but that does not converge pointwise almost everywhere.
Hint: Use indicator functions of carefully selected subintervals of $[0, 1]$.

Exercise 8.5.18. Let $(X, \mathcal{A}, \mu)$ be a probability space, and let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Let $\overline{\mathcal{B}}$ be the completion of $\mathcal{B}$ in $\mathcal{A}$, that is,

$$\overline{\mathcal{B}} := \{A \in \mathcal{A} \mid \exists B \in \mathcal{B} \text{ such that } \mu(A \triangle B) = 0\}.$$

Show that $\overline{\mathcal{B}}$ is a sub-σ-algebra of $\mathcal{A}$ for which the following two properties hold:
(a) $\overline{\mathcal{B}} \supseteq \mathcal{B}$.
(b) $E(\varphi|\overline{\mathcal{B}}) = E(\varphi|\mathcal{B})$.

Exercise 8.5.19. Let $(X, \mathcal{A}, \mu)$ be a probability space, and let $\mathcal{B}$ and $\mathcal{C}$ be sub-σ-algebras of $\mathcal{A}$. Say that $\mathcal{B} \approx \mathcal{C}$ if $\mathcal{B} \subseteq \overline{\mathcal{C}}$ and $\mathcal{C} \subseteq \overline{\mathcal{B}}$ (cf. Exercise 8.5.18). Show that $E(\varphi|\mathcal{B}) = E(\varphi|\mathcal{C})$ if $\mathcal{B} \approx \mathcal{C}$.

Exercise 8.5.20. Let $\mathcal{B}(\mathbb{R})$ denote the Borel σ-algebra of $\mathbb{R}$. The collection

$$\mathcal{B} := \{B \in \mathcal{B}(\mathbb{R}) \mid B = -B\}$$

of all Borel sets that are symmetric with respect to the origin forms a sub-σ-algebra of $\mathcal{B}(\mathbb{R})$. Let λ denote the Lebesgue measure on $\mathbb{R}$ and let $\varphi \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. Prove that

$$E(\varphi|\mathcal{B})(x) = \frac{1}{2}[\varphi(x) + \varphi(-x)], \quad \forall x \in \mathbb{R}.$$

Hint: First show that $E(\varphi|\mathcal{B})$ must be an even function. Then use the fact that the transformation $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = -x$ is λ-invariant.

Exercise 8.5.21. Let $T : (X, \mathcal{A}) \to (Y, \mathcal{B})$ be a measurable transformation and μ be a measure on $(X, \mathcal{A})$. Show that the set function $\mu \circ T^{-1}$ is a measure on $(Y, \mathcal{B})$.

Exercise 8.5.22. This exercise pertains to a two-sided version of Example 8.1.14. Let $(E, \mathcal{F}, P)$ be a probability space. The product set $E^{\mathbb{Z}} := \prod_{k=-\infty}^{\infty} E$ is commonly equipped with the product σ-algebra $\mathcal{F}_{\mathbb{Z}}$ generated by the semialgebra of all (finite) cylinders (also called rectangles), namely, given $m, n \in \mathbb{Z}$, with $m \leq n$, and $E_m, E_{m+1}, \ldots, E_n \in \mathcal{F}$, the set

$$\prod_{k=-\infty}^{m-1} E \times E_m \times E_{m+1} \times \cdots \times E_n \times \prod_{k=n+1}^{\infty} E = \{\tau \in E^{\mathbb{Z}} : \tau_k \in E_k, \forall m \leq k \leq n\}$$

is called a cylinder. The product measure μ_P on $\mathcal{F}_{\mathbb{Z}}$ is the unique probability measure which confers to a cylinder the value

$$\mu_P\left(\prod_{k=-\infty}^{m-1} E \times E_m \times E_{m+1} \times \cdots \times E_n \times \prod_{k=n+1}^{\infty} E\right) := \prod_{k=m}^{n} P(E_k). \tag{8.33}$$

The existence and uniqueness of this measure can be established using Theorem A.1.27, Lemma A.1.29 and Theorem A.1.28 successively. For more information, see Halmos [27] (pp. 157–158) or Taylor [72] (Chapter III, Section 4).

It is easy to show that the left shift map $\sigma : E^{\mathbb{Z}} \to E^{\mathbb{Z}}$, which is defined by $\sigma\big((\tau_n)_{n=-\infty}^{\infty}\big) := (\tau_{n+1})_{n=-\infty}^{\infty}$, preserves the product measure μ_P. The measure-preserving dynamical system $(\sigma : E^{\mathbb{Z}} \to E^{\mathbb{Z}}, \mu_P)$ is commonly referred to as a *two-sided Bernoulli shift* with set of states E. In this book, primarily focused on noninvertible dynamical systems, one-sided Bernoulli shifts will be of primer importance. However, in abstract ergodic theory, two-sided Bernoulli shifts seem to have played a more prominent role and they are encountered several times in this book, particularly in Sections 8.3.3, 8.4, and 13.9.5.

From this point on, we assume that the original probability space is of the form $(E, \mathcal{F}, P)$, where E is countable and $\mathcal{F}$ is the σ-algebra $\mathcal{P}(E)$ of all subsets of E.

Let $E^{\mathbb{Z}_-} = \prod_{k=-\infty}^{-1} E$ and $E^{\mathbb{Z}_+} = \prod_{k=0}^{\infty} E$, so $E^{\mathbb{Z}} = E^{\mathbb{Z}_-} \times E^{\mathbb{Z}_+}$. Accordingly, for any $\tau \in E^{\mathbb{Z}}$, write $\tau_- := \cdots \tau_{-2}\tau_{-1} \in E^{\mathbb{Z}_-}$ and $\tau_+ := \tau_0\tau_1\tau_2\ldots \in E^{\mathbb{Z}_+}$.

(a) Show that the family $C = \{E^{\mathbb{Z}_-} \times \{\omega\}\}_{\omega \in E^{\mathbb{Z}_+}}$, that is, the family of sets consisting of double-sided sequences having a common positive part, constitutes an uncountable measurable partition of $E^{\mathbb{Z}}$.

Let $\mathcal{C} := \sigma(C)$ be the sub-σ-algebra of $\mathcal{F}_{\mathbb{Z}}$ generated by the partition C. We aim to calculate $E(\varphi|\mathcal{C})$ for any function $\varphi \in L^1(E^{\mathbb{Z}}, \mathcal{F}_{\mathbb{Z}}, \mu_P)$.

For each $n \geq 0$, consider the family of all n-cylinders

$$C_n := \left\{E^{\mathbb{Z}_-} \times \{e_0\} \times \cdots \times \{e_{n-1}\} \times \prod_{k=n}^{\infty} E : e_k \in E,\ \forall 0 \leq k \leq n-1\right\}.$$

(b) Prove that $(C_n)_{n=0}^{\infty}$ is an ascending sequence of countable measurable partitions of $E^{\mathbb{Z}}$.

(c) For every $n \geq 0$, let $\mathcal{C}_n := \sigma(C_n)$. Demonstrate that $(\mathcal{C}_n)_{n=0}^{\infty}$ is an ascending sequence of sub-σ-algebras of $\mathcal{F}_{\mathbb{Z}}$.

(d) Let $\mathcal{C}_\infty := \sigma(\bigcup_{n=0}^{\infty} C_n)$. Show that $\mathcal{C} = \mathcal{C}_\infty$.

(e) Deduce that $E(\varphi|\mathcal{C}) = \lim_{n\to\infty} E(\varphi|\mathcal{C}_n)$.

For any given $\tau \in E^{\mathbb{Z}}$ and $m, n \in \mathbb{Z}$ such that $m \leq n$, let

$$[\tau]_m^n = \prod_{k=-\infty}^{m-1} E \times \{\tau_m\} \times \{\tau_{m+1}\} \times \cdots \times \{\tau_n\} \times \prod_{k=n+1}^{\infty} E.$$

(f) Establish that $E(\varphi|\mathcal{C})(\tau) = \lim_{n\to\infty} E(\varphi|[\tau]_0^{n-1})$ for every $\tau \in E^{\mathbb{Z}}$.
(g) Prove that the sequence of cylinder sets $([\tau]_0^{n-1})_{n=1}^{\infty}$ is descending and that $\bigcap_{n=1}^{\infty}[\tau]_0^{n-1} = E^{\mathbb{Z}_-} \times \{\tau_+\}$.
(h) Deduce that $\lim_{n\to\infty} \mu_P([\tau]_0^{n-1}) = \mu_P(E^{\mathbb{Z}_-} \times \{\tau_+\})$.
(i) Deduce further that $\lim_{n\to\infty} \int_{[\tau]_0^{n-1}} \varphi\, d\mu_P = \int_{E^{\mathbb{Z}_-}\times\{\tau_+\}} \varphi\, d\mu_P$.
(j) Conclude that $E(\varphi|\mathcal{C})(\tau) = E(\varphi|E^{\mathbb{Z}_-} \times \{\tau_+\})$ for all $\tau \in E^{\mathbb{Z}}$.

That is, you have showed that $E(\varphi|\mathcal{C})$ is constant, and is equal to the mean value of φ, on each set of the form $E^{\mathbb{Z}_-} \times \{\omega\}$, where $\omega \in E^{\mathbb{Z}_+}$.

Exercise 8.5.23. Let E be a finite set having at least two elements, let $P : E \to [0,1]$ be a probability vector, and let $(\sigma : E^{\mathbb{Z}} \to E^{\mathbb{Z}}, \mu_P)$ be the corresponding two-sided Bernoulli shift described in Exercise 8.5.22. Provide a direct proof, based on Proposition 8.3.20, that the automorphism $(\sigma : E^{\mathbb{Z}} \to E^{\mathbb{Z}}, \mu_P)$ is K-mixing.

Hint: Consider the projection $\pi_+ : E^{\mathbb{Z}} \to E^{\mathbb{N}}$ defined by the formula $\pi_+(\omega) = \omega|_0^{\infty}$.

Exercise 8.5.24. Let $\mathcal{B}_+$ be the standard Borel σ-algebra on $E^{\infty} = E^{\mathbb{N}}$. Show that the σ-algebra $\mathcal{B} := \pi_+^{-1}(\mathcal{B}_+)$ on $E^{\mathbb{Z}}$ satisfies all the hypotheses of Proposition 8.3.20.

Exercise 8.5.25. The map $G : [0,1] \to [0,1]$ defined by

$$G(x) := \begin{cases} 0 & \text{if } x = 0 \\ \langle \frac{1}{x} \rangle & \text{if } x > 0, \end{cases}$$

where $\langle r \rangle$ denotes the fractional part of r, is called the *Gauss map*.
(a) Show that this map is not invariant under the Lebesgue measure λ on $[0,1]$.
(b) Prove that the Borel probability measure

$$\mu_G(B) := \frac{1}{\log 2} \int_B \frac{1}{1+x}\, dx$$

is G-invariant. The measure μ_G is known as the *Gauss measure*.

Exercise 8.5.26. Recall the Farey map $F : [0,1] \to [0,1]$ from Example 1.2.3. Show that the Borel probability measure

$$\mu_F(B) := \int_B \frac{1}{x}\, dx$$

is F-invariant, while the Lebesgue measure λ is not.

Exercise 8.5.27. Show that the Dirac point-mass δ_0 is the only T-invariant Borel probability measure for the doubling map on the entire real line, that is, for $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = 2x$.

Exercise 8.5.28. Find all T-invariant Borel probability measures for the squaring map on the entire real line, that is, for $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = x^2$.

Exercise 8.5.29. Prove that the continuous transformation $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = x^2 + 1$ does not admit any (finite or infinite) T-invariant Borel measure.

Exercise 8.5.30. Let $b \neq 0$. Prove that the translation $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = x+b$ admits a σ-finite T-invariant Borel measure but not a finite one.

Exercise 8.5.31. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. Suppose that $B \in \mathcal{A}$ is forward T-invariant, that is, $T(B) \subseteq B$. Let $\mathcal{A}|_B := \{A \cap B : A \in \mathcal{A}\}$ be the projection of $\mathcal{A}$ onto B. Suppose that μ is a probability measure on $(B, \mathcal{A}|_B)$. Define the set function $\widehat{\mu} : \mathcal{A} \to [0, 1]$ by the formula

$$\widehat{\mu}(A) := \mu(A \cap B).$$

Show that $\widehat{\mu}$ is a probability measure on $(X, \mathcal{A})$. Furthermore, prove that if μ is $T|_B$-invariant then $\widehat{\mu}$ is T-invariant.

Exercise 8.5.32. Find a nontrivial measure-preserving dynamical system $T : X \to X$ on a probability space $(X, \mathcal{B}, \mu)$ for which there exist at least three completely T-invariant measurable sets of positive μ-measure.

Exercise 8.5.33. Identify a nontrivial measure-preserving dynamical system $T : X \to X$ on a probability space $(X, \mathcal{B}, \mu)$ for which there exist uncountably many measurable sets of positive measure and the symmetric difference of any two of these sets has positive measure.

Exercise 8.5.34. Show that the inverse of a measure-preserving isomorphism is a measure-preserving isomorphism.

Exercise 8.5.35. Let $T : X \to X$ be a map and let $\varphi : X \to \mathbb{R}$ be a real-valued function. Let $x \in X$. If $k, n \in \mathbb{N}$ are such that $k < n$, prove that

$$S_n\varphi(x) = S_k\varphi(x) + S_{n-k}\varphi(T^k(x)).$$

Exercise 8.5.36. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. Prove the following statements:
(a) If $Y \subseteq X$ is forward T-invariant, then $\mu(Y) = \mu(\bigcup_{n=0}^{\infty} T^{-n}(Y))$.
(b) If $Z \subseteq X$ is backward T-invariant, then $\mu(Z) = \mu(\bigcap_{n=0}^{\infty} T^{-n}(Z))$.
(c) If T is ergodic with respect to μ and $W \subseteq X$ is forward or backward T-invariant, then $\mu(W) \in \{0, 1\}$.

N. B.: Part (c) means that the concept of ergodicity might alternatively be defined in terms of forward invariant sets or in terms of backward invariant sets.

Exercise 8.5.37. Find a topological dynamical system that admits a measure which is ergodic but not invariant.

Exercise 8.5.38. Fix $n > 1$. Consider the map $T_n : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $T_n(x) = nx \pmod 1$. Using the uniqueness of the Fourier series representation of functions in $L^2(\lambda)$ (like in the proof of Proposition 8.2.29), show that T_n is ergodic with respect to the Lebesgue measure λ.

Exercise 8.5.39. Recalling Exercise 8.5.25, show that the Gauss measure is ergodic for the Gauss map.

Exercise 8.5.40. Suppose that X is a countable set, $\mathcal{P}(X)$ is the σ-algebra of all subsets of X and μ is a probability measure on $(X, \mathcal{P}(X))$. Show that if $T : X \to X$ is ergodic with respect to μ then there exists a periodic point y of T such that $\mu(\{T^n(y) : n \geq 0\}) = 1$.

Exercise 8.5.41. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. Let also $f \in L^1(\mu)$. Show that

$$\lim_{n\to\infty} \frac{1}{n} f(T^n(x)) = 0 \quad \text{for } \mu\text{-a. e. } x \in X.$$

Exercise 8.5.42. Let $n \geq 2$. Every number $x \in [0, 1]$ has a n-adic expansion, that is,

$$x = \sum_{i=1}^{\infty} \frac{\omega_i}{n^i}$$

for some $\omega = (\omega_i)_{i=1}^{\infty} \in \{0, 1, \dots, n-1\}^{\infty}$. Let $p, q \in \{0, 1, \dots, n-1\}$. Show that

$$\lim_{k\to\infty} \frac{1}{k} \#\{1 \leq j \leq k : \omega_j = p \text{ and } \omega_{j+3} = q\} = \frac{1}{n^2} \quad \text{for } \lambda\text{-a. e. } x \in [0, 1].$$

Exercise 8.5.43. Let $n = 2$. Using the same notation as in Exercise 8.5.42, show that

$$\lim_{k\to\infty} \frac{1}{k} \sum_{j=0}^{k-1} (\omega_j^2 + \omega_{j+1}^2) = 1 \quad \text{for } \lambda\text{-a. e. } x \in [0, 1].$$

Exercise 8.5.44. Let $\ell \in C(X)^*$. Set $\Delta\ell := \mathrm{var}(\ell) - \ell$, where $\mathrm{var}(\ell)$ comes from Definition 8.2.36. Prove that $\Delta\ell \in C(X)^*$ and is positive. In addition, show that if ℓ is T-invariant then so is $\Delta\ell$.

Exercise 8.5.45. Prove Theorem 8.4.2(a,b,c).

Exercise 8.5.46. Referring back to the proof of Theorem 8.4.3, show that $\widetilde{\mathcal{F}}_n = \widetilde{T}^{-1}(\widetilde{\mathcal{F}}_{n+1})$ for all $n \geq 0$.

Exercise 8.5.47. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. A measurable function $g : X \to \mathbb{R}$ is said to be μ-a. e. T-subinvariant if $g(T(x)) \leq g(x)$ for μ-a. e. $x \in X$. Show that if T is ergodic and $g \in L^1(\mu)$ is μ-a. e. T-subinvariant, then g is μ-a. e. constant (cf. Theorem 8.2.18).

Exercise 8.5.48. In this exercise, we introduce and discuss Markov chains, which are in a sense a generalization of the Bernoulli shifts studied in Examples 8.1.14 and 8.2.32. Let E be a countable alphabet with at least two letters. Let $A : E \times E \to (0,1)$ be a stochastic matrix, that is, a matrix such that

$$\sum_{j \in E} A_{ij} = 1, \quad \forall i \in E.$$

(a) Prove (you may use the classical Perron–Frobenius theorem for positive matrices) that there exists a unique probability vector $P : E \to [0,1]$ such that $PA = P$, that is,

$$\sum_{i \in E} P_i A_{ij} = P_j, \quad \forall j \in E.$$

(b) Further show that $P_j \in (0,1)$ for all $j \in E$.

For every $\omega \in E^*$, say $\omega \in E^n$, set

$$\mu_A([\omega]) = P_{\omega_1} \prod_{k=1}^{n-1} A_{\omega_k \omega_{k+1}}.$$

In a similar way to Examples 8.1.14 and 8.2.32:

(c) Prove that μ_A uniquely extends to a Borel probability measure on $E^{\mathbb{N}}$. (In the sequel, we keep the same symbol μ_A for this measure.)

(d) Show that the measure μ_A is shift-invariant and ergodic.

The dynamical system $(\sigma : E^{\mathbb{N}} \to E^{\mathbb{N}}, \mu_A)$ is called the one-sided *Markov chain* generated by the stochastic matrix A.

Assuming that the alphabet E is finite:

(e) Prove that the one-cylinder partition $\{[e]\}_{e \in E}$ is a weak Bernoulli generator for the measure-preserving dynamical system $(\sigma : E^{\mathbb{N}} \to E^{\mathbb{N}}, \mu_A)$. Conclude that this system is weak Bernoulli and that its Rokhlin's natural extension is isomorphic to a two-sided Bernoulli shift.

(f) Wanting an explicit representation of the Rokhlin's natural extension of the system $(\sigma : E^{\mathbb{N}} \to E^{\mathbb{N}}, \mu_A)$, prove that, as in the case of Bernoulli shifts, this extension is isomorphic to the naturally defined two-sided Markov chain $(\sigma : E^{\mathbb{Z}} \to E^{\mathbb{Z}}, \mu_A)$ generated by the stochastic matrix A and the unique probability vector P.

Note that when E is a countable set, every Bernoulli shift generated by a probability vector $P : E \to (0,1)$ is the Markov chain generated by the stochastic matrix A whose columns are all equal to that vector P. In this sense, Markov chains are generalizations of Bernoulli shifts. We established this for countable alphabets E. But in fact, even more generally, when E is an arbitrary set as in Examples 8.1.14 and 8.2.32, one can define appropriate Markov chains that generalize the Bernoulli shifts considered in those examples.

Finally, as for Bernoulli shifts, Markov chains will be shown to be Gibbs and equilibrium states for appropriately chosen Hölder continuous potentials in Exercises 13.11.11–13.11.12 and 17.9.13–17.9.14.

9 Measure-theoretic entropy

In Chapter 7, we studied the topological entropy of a topological dynamical system. We now study its measure-theoretic counterpart. Measure-theoretic entropy is also sometimes known as metric entropy or Kolmogorov–Sinai metric entropy. It was introduced by A. Kolmogorov and Ya. Sinai in the late 1950s; see [67]. Since then, its account has been presented in virtually every textbook on ergodic theory. Its introduction to dynamical systems was motivated by Ludwig Boltzmann's concept of entropy in statistical mechanics and Claude Shannon's work on information theory; see [64, 65].

As for topological entropy, there are three stages in the definition of metric entropy. Recall that topological entropy is defined by **covering** the underlying topological space with basic sets in that space, that is, open sets; metric entropy, on the other hand, is defined by **partitioning** the underlying measurable space with basic sets in that space, namely, measurable sets. Indeed, whereas one cannot generally partition a topological space into open sets (this is only possible in a disconnected space), it is generally possible to partition a measurable space into measurable sets. Accordingly, we first study measurable partitions in Section 9.2. Then we examine the concepts of information and conditional information in Section 9.3. In Section 9.4, we finally define metric entropy. And in Section 9.5, we formulate and prove the full version of Shannon–McMillan–Breiman's characterization of metric entropy. This characterization depicts what metric entropy really is. Finally, in Section 9.6 we shed further light on the nature of entropy by proving the Brin–Katok local entropy formula. Like the Shannon–McMillan–Breiman theorem, the Brin–Katok local entropy formula is very useful in applications.

9.1 An excursion into the origins of entropy

This exposition is inspired by [80].

The concept of metric entropy arose from the creation of information theory by Shannon [64, 65]. That notion was adapted from Boltzmann's advances on entropy in statistical mechanics.

Contemplate the conduct of a random experiment (for instance, the rolling of a die) with a finite number of possible outcomes $(A_1, A_2, \dots, A_n)$ with respective probabilities $(p_1, p_2, \dots, p_n)$. Naturally, we would like to ascribe to this experiment a quantity (a number) that indicates the level of uncertainty associated with the outcome of the experiment. For example, if a six-face unfair die has the outcomes $(1, 2, 3, 4, 5, 6)$ with associated probabilities $(95\,\%, 1\,\%, 1\,\%, 1\,\%, 1\,\%, 1\,\%)$, then the level of uncertainty of the outcome is much smaller than the level of uncertainty in the throwing of a fair die, that is, a die with an equal probability of 1/6 of falling on any of its 6 faces. We aim at

https://doi.org/10.1515/9783110702682-009

finding a real-valued function

$$\mathrm{H}(p_1, p_2, \dots, p_n)$$

that describes the level of uncertainty. Obviously, this nonnegative function is defined on the n-tuples $(p_1, p_2, \dots, p_n)$ such that $0 \le p_i \le 1$ for all $1 \le i \le n$ and $\sum_{i=1}^{n} p_i = 1$. We now provide a rationale for the type of function that naturally emerges in this context.

Intuitively, the level of uncertainty of the outcome reaches...

- a minimum of 0 when one of the outcomes is absolutely certain, that is, has a probability of 100 % = 1 of occurring; this means that $\mathrm{H}(p_1, p_2, \dots, p_n) = 0$ when $p_i = 1$ for some i (and thus $p_j = 0$ for all $j \neq i$).
- a maximum when the n outcomes have equal probability $1/n$ of taking place, that is, $\max \mathrm{H}(p_1, p_2, \dots, p_n) = \mathrm{H}(1/n, 1/n, \dots, 1/n)$.

Speaking of the equiprobable case, it is interesting and even crucial to look at the behavior of the function

$$f(n) = \mathrm{H}(1/n, 1/n, \dots, 1/n).$$

As the number n of outcomes grows, the level of uncertainty grows: it is easier to predict the outcome of casting a six-side fair die than a twenty-side fair die. So we posit that:

(A1) The function f must be strictly increasing.

Consider now two independent experiments, a first one with n_1 equiprobable outcomes and a second one with n_2 equiprobable outcomes. Running both experiments simultaneously results in a “product” experiment which consists of $n_1 n_2$ equiprobable outcomes. Knowledge of the outcome of the first experiment does not affect the uncertainty surrounding the outcome of the second one, and vice versa. Accordingly, it is natural to expect that once the uncertainty of the outcome of the first experiment, $f(n_1)$, is subtracted from the uncertainty on the entire experiment, $f(n_1 n_2)$, the remaining uncertainty coincides with that of the conduct of the second experiment, $f(n_2)$, that is,

$$f(n_1 n_2) - f(n_1) = f(n_2).$$

Equivalently, we posit that

(A2) $f(n_1 n_2) = f(n_1) + f(n_2), \forall n_1, n_2 \in \mathbb{N}$.

Let us return now to a general experiment with n outcomes. Partition these outcomes into two subsets A and B, with respective total probabilities $p_A = p_1 + \dots + p_k$ and $p_B = p_{k+1} + \dots + p_n$. For instance, in the experiment of throwing a six-face die, we might be interested not in the face number the die falls on but rather in whether it settles on an even face or an odd face. Naturally, we would like to relate the uncertainty of the original experiment with that of the "simplified" experiment where the outcome is perceived as A or B. If the outcome of the simplified experiment is A, then the remaining uncertainty about the outcome of the original experiment is $\mathrm{H}(p_1/p_A, \dots, p_k/p_A)$. Similarly, if the outcome of the simplified experiment is B, then the remaining uncertainty about the outcome of the original experiment is $\mathrm{H}(p_{k+1}/p_B, \dots, p_n/p_B)$. Since the two outcomes in the simplified experiment occur with probabilities p_A and p_B respectively, we posit that the level of uncertainty about the original experiment can be expressed as

(A3) $\mathrm{H}(p_1, \dots, p_n) = \mathrm{H}(p_A, p_B) + p_A \mathrm{H}(\frac{p_1}{p_A}, \dots, \frac{p_k}{p_A}) + p_B \mathrm{H}(\frac{p_{k+1}}{p_B}, \dots, \frac{p_n}{p_B})$.

Finally, it is reasonable to assume that the function H is continuous, that is, a small change in the probabilities of the outcomes of an experiment, results in a small change in the level of uncertainty besetting the experiment. Because of axiom (A3), it suffices to make this assumption in the case of a binary outcome experiment:

(A4) The function $p \mapsto \mathrm{H}(p, 1-p)$ is continuous on $(0, 1)$.

Theorem 9.1.1. *The only functions satisfying axioms* (A1)–(A4) *are the functions of the form*

$$\mathrm{H}(p_1, \dots, p_n) = -C \sum_{i=1}^{n} p_i \log p_i$$

for some constant $C > 0$.

Proof. See Exercise 9.7.1. □

This will be the form of the entropy function. Mathematically, the various events that can be witnessed in an experiment constitute a measurable partition of the measurable space of all outcomes of the experiment. This explains why we study measurable partitions in the next section.

9.2 Partitions of a measurable space

Definition 9.2.1. Let $(X, \mathcal{A})$ be a measurable space. A countable measurable *partition* of X is a family $\alpha = \{A_k\}_{k=1}^{\infty}$ such that:
(a) $A_k \in \mathcal{A}$ for all $k \in \mathbb{N}$;
(b) $A_i \cap A_j = \emptyset$ for all $i \neq j$; and
(c) $\bigcup_{k=1}^{\infty} A_k = X$.

The individual sets A_k, $k \in \mathbb{N}$, making up the partition α are called *atoms* of α. For each $x \in X$ the unique atom of the partition α which contains the point x will be denoted by $\alpha(x)$. Finally, we shall denote the set of all countable measurable partitions on the space $(X, \mathcal{A})$ by $\mathrm{Part}(X, \mathcal{A})$.

In the sequel, it will always be implicitly understood that partitions are countable (finite or infinite) and measurable.

In Definition 7.1.4, we introduced a refinement relation for covers of a space. As partitions of a space constitute a special class of covers of that space, it is natural to examine the restriction of the refinement relation to partitions. In contrast with the relation for covers, the restriction turns out to be a relation of partial order on the set of all partitions.

Definition 9.2.2. Let $(X, \mathcal{A})$ be a measurable space and $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. We say that partition β is *finer* than partition α, or that α is *coarser* than β, which will be denoted by $\alpha \le \beta$, if for every atom $B \in \beta$ there exists some atom $A \in \alpha$ such that $B \subseteq A$. In other words, each atom of α is a union of atoms of β.

Equivalently, β is a refinement of α if $\beta(x) \subseteq \alpha(x)$ for all $x \in X$. See also Exercises 9.7.2–9.7.5.

We now introduce for partitions the analogue of the join of two covers.

Definition 9.2.3. Given $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$, the partition

$$\alpha \vee \beta := \{A \cap B \mid A \in \alpha, B \in \beta\}$$

is called the *join* of α and β.

The basic properties of the join are given in the following lemma. Their proofs are left to the reader as an exercise.

Lemma 9.2.4. *Since partitions are covers, the relation $\le$ and the operation $\vee$ enjoy all the properties of $\prec$ and $\vee$ described in Remark* 7.1.3 *and Lemma* 7.1.5. *Moreover, if $\alpha, \beta, \gamma \in$* $\mathrm{Part}(X, \mathcal{A})$, *then:*

(a) $\le$ *is a relation of partial order on* $\mathrm{Part}(X, \mathcal{A})$, *that is, it is...*
  - *reflexive ($\alpha \le \alpha$);*
  - *transitive ($\alpha \le \beta$ and $\beta \le \gamma \Longrightarrow \alpha \le \gamma$); and*
  - *antisymmetric ($\alpha \le \beta \le \alpha \Longleftrightarrow \alpha = \beta$).*

(b) $\alpha \le \beta \iff \alpha \vee \beta = \beta$.

(c) $\alpha \vee \alpha = \alpha$.

(d) $\alpha \vee \{X\} = \alpha$.

9.3 Information and conditional information functions

Let $(X, \mathcal{A}, \mu)$ be a probability space. The set X may be construed as the set of all possible states (or outcomes) of an experiment, while the σ-algebra $\mathcal{A}$ is the set of all possible events, and $\mu(A)$ is the probability that event $A \in \mathcal{A}$ take place. Imagine that this experiment is conducted using an instrument which, due to some limitation, can only provide measurements accurate up to the atoms of a partition $\alpha = \{A_k\}_{k=1}^{\infty} \in \mathrm{Part}(X, \mathcal{A})$. In other words, this instrument can only tell us which atom of α the outcome of the experiment falls into. Any observation made through this instrument will therefore be of the form A_k for a unique k. If the experiment were conducted today, the probability that its outcome belongs to A_k, that is, the probability that the experiment results in observing event A_k with our instrument, would be $\mu(A_k)$.

We would like to introduce a function that describes the information that our instrument would give us about the outcome of the experiment. So, let $x \in X$. Intuitively, the smaller the atom of the partition to which x belongs, the more information our instrument provides us about x. In particular, if x lies in an atom of full measure, then our instrument gives us essentially no information about x. Moreover, because our instrument cannot distinguish points which belong to a common atom of the partition, the sought-after information function must be constant on every atom. In light of Theorem 9.1.1, the following definition is natural (think about the relation between information and uncertainty on the outcome of an experiment).

Definition 9.3.1. Let $(X, \mathcal{A}, \mu)$ be a probability space and $\alpha \in \mathrm{Part}(X, \mathcal{A})$. The nonnegative function $I_\mu(\alpha) : X \to [0, \infty]$ defined by

$$I_\mu(\alpha)(x) := -\log \mu\big(\alpha(x)\big)$$

is called the *information function* of the partition α. By convention, $\log 0 = -\infty$.

As the function $t \mapsto -\log t$ is strictly decreasing, for any $x \in X$ the smaller $\mu(\alpha(x))$ is, the larger $I_\mu(\alpha)(x)$ is, that is, the smaller the measure of the atom $\alpha(x)$ is, the more information the partition α gives us about x. In particular, the finer the partition, the more information it gives us about every point in the space.

In the next lemma, we collect some of the basic properties of the information function. Their proofs are straightforward and are left to the reader.

Lemma 9.3.2. *Let $(X, \mathcal{A}, \mu)$ be a probability space and $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. Then:*
(a) *If $\alpha \le \beta$, then $I_\mu(\alpha) \le I_\mu(\beta)$.*
(b) *$I_\mu(\alpha)(x) = 0$ if and only if $\mu(\alpha(x)) = 1$.*
(c) *$I_\mu(\alpha)(x) = \infty$ if and only if $\mu(\alpha(x)) = 0$.*
(d) *If $\alpha(x) = \alpha(y)$, then $I_\mu(\alpha)(x) = I_\mu(\alpha)(y)$, that is, $I_\mu(\alpha)$ is constant over each atom of α.*

More advanced properties of the information function will be presented below. Meanwhile, we introduce a function which describes the information gathered from a partition α given that a partition β has already been applied.

Definition 9.3.3. Let $(X, \mathcal{A}, \mu)$ be a probability space and $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. The *conditional information function* of partition α given partition β is the nonnegative function

$I_\mu(\alpha|\beta) : X \to [0,\infty]$ defined by

$$I_\mu(\alpha|\beta)(x) := -\log \mu_{\beta(x)}(\alpha(x)),$$

where μ_B is the conditional measure from Definition A.1.70 in Appendix A. Observe that

$$I_\mu(\alpha|\beta)(x) := -\log \frac{\mu(\alpha(x) \cap \beta(x))}{\mu(\beta(x))} = -\log \frac{\mu((\alpha \vee \beta)(x))}{\mu(\beta(x))} = I_\mu(\alpha \vee \beta)(x) - I_\mu(\beta)(x).$$

By convention, $\frac{0}{0} = 0$ and $\infty - \infty = \infty$.

For any partition α, notice that $I_\mu(\alpha|\{X\}) = I_\mu(\alpha)$, that is, the information function coincides with the conditional information function with respect to the trivial partition $\{X\}$. Note further that $I_\mu(\alpha|\beta)$ is constant over each atom of $\alpha \vee \beta$.

Our next aim is to give some advanced properties of the conditional information function. Notice that some of these properties hold pointwise, while others hold atomwise only, that is, after integrating over atoms. In particular, the reader should compare statements (e–h) in the next theorem. First, though, we make one further definition, which is related to our excursion in Section 9.1.

Definition 9.3.4. Let the function $k : [0,1] \to [0,1]$ be defined by

$$k(t) = -t \log t,$$

where it is understood that $0 \cdot (-\infty) = 0$.

The function k is continuous, strictly increasing on the interval $[0, e^{-1}]$, strictly decreasing on the interval $[e^{-1}, 1]$, and concave. See Figure 9.1. Recall that a function $k : I \to \mathbb{R}$, where $I \subseteq \mathbb{R}$ is an interval, is *concave* on I if

$$k(tx + (1-t)y) \ge tk(x) + (1-t)k(y), \quad \forall t \in [0,1], \forall x, y \in I,$$

that is, the line segment joining any two points on the curve lies under the curve.

Theorem 9.3.5. *Let $(X, \mathcal{A}, \mu)$ be a probability space and $\alpha, \beta, \gamma \in \mathrm{Part}(X, \mathcal{A})$. The following statements hold:*

(a) $I_\mu(\alpha \vee \beta|\gamma) = I_\mu(\alpha|\gamma) + I_\mu(\beta|\alpha \vee \gamma)$.
(b) $I_\mu(\alpha \vee \beta) = I_\mu(\alpha) + I_\mu(\beta|\alpha)$.
(c) *If $\alpha \le \beta$, then $I_\mu(\alpha|\gamma) \le I_\mu(\beta|\gamma)$.*
(d) *If $\alpha \le \beta$, then $I_\mu(\alpha) \le I_\mu(\beta)$.*
(e) *If $\beta \le \gamma$, then for all $A \in \alpha$ and all $B \in \beta$,*

$$\int_{A\cap B} I_\mu(\alpha|\beta)\, d\mu \ge \int_{A\cap B} I_\mu(\alpha|\gamma)\, d\mu.$$

Note: In general, $\beta \le \gamma \not\Rightarrow I_\mu(\alpha|\beta) \ge I_\mu(\alpha|\gamma)$.

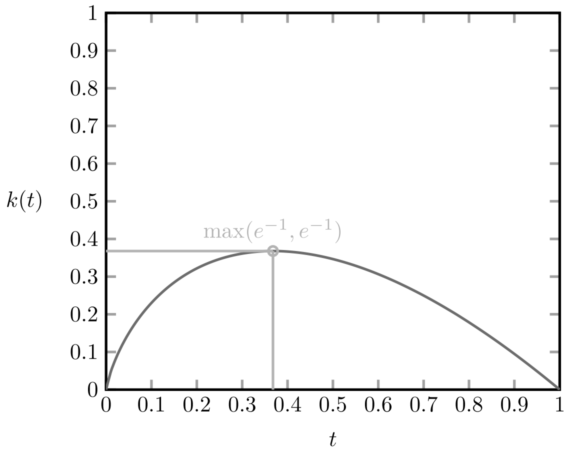

Figure 9.1: The function underlying entropy: $k(t) = -t \log t$.

(f) *For all $C \in \gamma$,*

$$\int_C I_\mu(\alpha \vee \beta|\gamma)\, d\mu \leq \int_C I_\mu(\alpha|\gamma)\, d\mu + \int_C I_\mu(\beta|\gamma)\, d\mu.$$

Note: In general, $I_\mu(\alpha \vee \beta|\gamma) \not\leq I_\mu(\alpha|\gamma) + I_\mu(\beta|\gamma)$.

(g)
$$\int_X I_\mu(\alpha \vee \beta)\, d\mu \leq \int_X I_\mu(\alpha)\, d\mu + \int_X I_\mu(\beta)\, d\mu.$$

(h) *For all $A \in \alpha$ and all $B \in \beta$,*

$$\int_{A\cap B} I_\mu(\alpha|\gamma)\, d\mu \leq \int_{A\cap B} I_\mu(\alpha|\beta)\, d\mu + \int_{A\cap B} I_\mu(\beta|\gamma)\, d\mu.$$

Note: In general, $I_\mu(\alpha|\gamma) \not\leq I_\mu(\alpha|\beta) + I_\mu(\beta|\gamma)$.

(i) $I_\mu(\alpha) \leq I_\mu(\alpha|\beta) + I_\mu(\beta)$.

Proof. (a) Let $x \in X$. Then

$$\begin{aligned}
I_\mu(\alpha \vee \beta|\gamma)(x) &= -\log \frac{\mu\big((\alpha \vee \beta \vee \gamma)(x)\big)}{\mu(\gamma(x))} \\
&= -\log \frac{\mu\big(\beta(x) \cap (\alpha \vee \gamma)(x)\big)}{\mu(\gamma(x))} \\
&= -\log\left(\frac{\mu\big(\beta(x) \cap (\alpha \vee \gamma)(x)\big)}{\mu\big((\alpha \vee \gamma)(x)\big)} \cdot \frac{\mu\big((\alpha \vee \gamma)(x)\big)}{\mu(\gamma(x))} \right) \\
&= -\log \frac{\mu\big(\beta(x) \cap (\alpha \vee \gamma)(x)\big)}{\mu\big((\alpha \vee \gamma)(x)\big)} - \log \frac{\mu\big((\alpha \vee \gamma)(x)\big)}{\mu(\gamma(x))} \\
&= I_\mu(\beta|\alpha \vee \gamma)(x) + I_\mu(\alpha|\gamma)(x).
\end{aligned}$$

(b) Setting $\gamma = \{X\}$ in (a) results in

$$I_\mu(\alpha \vee \beta)(x) = I_\mu(\alpha \vee \beta|\{X\})(x) = I_\mu(\beta|\alpha \vee \{X\})(x) + I_\mu(\alpha|\{X\})(x) = I_\mu(\beta|\alpha)(x) + I_\mu(\alpha)(x).$$

(c) Notice that if $\alpha \le \beta$, then $\alpha \vee \beta = \beta$. It then follows from (a) that

$$I_\mu(\beta|\gamma) = I_\mu(\alpha \vee \beta|\gamma) = I_\mu(\alpha|\gamma) + I_\mu(\beta|\alpha \vee \gamma) \ge I_\mu(\alpha|\gamma).$$

(d) Setting $\gamma = \{X\}$ in (c) leads to (d).
(e) Suppose that $\beta \le \gamma$. Let $A \in \alpha$ and $B \in \beta$. The downward concavity of the function k from Definition 9.3.4 means that

$$k\left(\sum_{n=1}^{\infty} a_n b_n\right) \ge \sum_{n=1}^{\infty} a_n k(b_n)$$

whenever $a_n, b_n \in [0,1]$ for all $n \in \mathbb{N}$ and $\sum_{n=1}^{\infty} a_n = 1$. Therefore,

$$k\left(\sum_{C\in\gamma} \mu_B(C)\frac{\mu(A\cap C)}{\mu(C)}\right) \ge \sum_{C\in\gamma} \mu_B(C) k\left(\frac{\mu(A\cap C)}{\mu(C)}\right). \tag{9.1}$$

Since $\beta \le \gamma$, each atom of β is a union of atoms of γ. So, either $C \cap B = C$ or $C \cap B = \emptyset$. Thus, either $\mu_B(C) = \frac{\mu(C)}{\mu(B)}$ or $\mu_B(C) = 0$, and the left-hand side of (9.1) simplifies to

$$k\left(\sum_{C\in\gamma} \mu_B(C)\frac{\mu(A\cap C)}{\mu(C)}\right) = k\left(\sum_{C\subseteq B} \frac{\mu(A\cap C)}{\mu(B)}\right) = k\left(\frac{\mu(A\cap B)}{\mu(B)}\right) = -\frac{\mu(A\cap B)}{\mu(B)}\log\frac{\mu(A\cap B)}{\mu(B)}.$$

The right-hand side of (9.1) reduces to

$$\sum_{C\in\gamma} \mu_B(C) k\left(\frac{\mu(A\cap C)}{\mu(C)}\right) = \sum_{C\subseteq B} \frac{\mu(C)}{\mu(B)} k\left(\frac{\mu(A\cap C)}{\mu(C)}\right) = \sum_{C\subseteq B} -\frac{\mu(A\cap C)}{\mu(B)}\log\frac{\mu(A\cap C)}{\mu(C)}.$$

Hence inequality (9.1) becomes

$$-\frac{\mu(A\cap B)}{\mu(B)}\log\frac{\mu(A\cap B)}{\mu(B)} \ge \sum_{C\subseteq B} -\frac{\mu(A\cap C)}{\mu(B)}\log\frac{\mu(A\cap C)}{\mu(C)}.$$

Multiplying both sides by $\mu(B)$ yields

$$-\mu(A\cap B)\log\frac{\mu(A\cap B)}{\mu(B)} \ge \sum_{C\subseteq B} -\mu(A\cap C)\log\frac{\mu(A\cap C)}{\mu(C)}.$$

Then

$$\begin{aligned}\int_{A\cap B} I_\mu(\alpha|\beta)\, d\mu &= -\mu(A\cap B)\log\frac{\mu(A\cap B)}{\mu(B)} \\ &\ge \sum_{C\subseteq B} -\mu(A\cap C)\log\frac{\mu(A\cap C)}{\mu(C)}\end{aligned}$$

$$= \sum_{C \subseteq B} \int_{A \cap C} I_\mu(\alpha|\gamma)\, d\mu$$
$$= \int_{A \cap B} I_\mu(\alpha|\gamma)\, d\mu.$$

(f) Since $\gamma \leq \alpha \vee \gamma$, this statement follows directly from combining (a) and (e).
(g) Setting $\gamma = \{X\}$ in (f) gives (g).
(h) Using part (c) and then part (a), we obtain

$$I_\mu(\alpha|\gamma) \leq I_\mu(\alpha \vee \beta|\gamma) = I_\mu(\beta|\gamma) + I_\mu(\alpha|\beta \vee \gamma).$$

Since $\beta \leq \beta \vee \gamma$, part (e) ensures that for all $A \in \alpha$ and all $B \in \beta$,

$$\int_{A \cap B} I_\mu(\alpha|\beta)\, d\mu \geq \int_{A \cap B} I_\mu(\alpha|\beta \vee \gamma)\, d\mu.$$

Therefore, for all $A \in \alpha$ and $B \in \beta$, we have that

$$\int_{A \cap B} I_\mu(\alpha|\gamma)\, d\mu \leq \int_{A \cap B} I_\mu(\beta|\gamma)\, d\mu + \int_{A \cap B} I_\mu(\alpha|\beta)\, d\mu.$$

(i) Using parts (d) and (b) in succession, we deduce that

$$I_\mu(\alpha) \leq I_\mu(\alpha \vee \beta) = I_\mu(\alpha|\beta) + I_\mu(\beta).$$

□

9.4 Definition of measure-theoretic entropy

The entropy of a measure-preserving dynamical system $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ is defined in three stages, which, for clarity of exposition, we split into the following three subsections.

9.4.1 First stage: entropy and conditional entropy for partitions

At this stage, the behavior of the system T is not in consideration. We solely look at the absolute and relative information provided by partitions.

The information function associated with a partition gives us the amount of information that can be gathered from the partition about each and every outcome of the experiment. It is useful to encompass the information given by a partition within a single number rather than a function. A natural way to achieve this is to calculate the average information given by the partition. This means integrating the information function over the entire space. The resulting integral is called the entropy of the partition.

Definition 9.4.1. Let $(X, \mathcal{A}, \mu)$ be a probability space and $\alpha \in \operatorname{Part}(X, \mathcal{A})$. The *entropy of* α with respect to the measure μ is defined to be the nonnegative extended number

$$\mathrm{H}_\mu(\alpha) := \int_X I_\mu(\alpha)\, d\mu = \sum_{A\in\alpha} -\mu(A)\log\mu(A),$$

where it is still understood that $0 \cdot (-\infty) = 0$, since null sets do not contribute to the integral.

The entropy of a partition is equal to zero if and only if the partition has an atom of full measure (which implies that all other atoms are of null measure). In particular, $\mathrm{H}_\mu(\{X\}) = 0$. Moreover, the entropy of a partition is small if the partition contains one atom with nearly full measure (so all other atoms have small measure). If the partition α is finite, it is possible, using calculus, to show that

$$0 \le \mathrm{H}_\mu(\alpha) \le \log \#\alpha \tag{9.2}$$

and that

$$\mathrm{H}_\mu(\alpha) = \log \#\alpha \quad \Longleftrightarrow \quad \mu(A) = \frac{1}{\#\alpha}, \quad \forall A \in \alpha.$$

In other words, on average we gain the most information from carrying out an experiment when the potential events are equiprobable of occurring.

Definition 9.4.2. Let $(X, \mathcal{A}, \mu)$ be a probability space and $\alpha, \beta \in \operatorname{Part}(X, \mathcal{A})$. The *conditional entropy of* α *given* β is defined to be

$$\mathrm{H}_\mu(\alpha|\beta) := \int_X I_\mu(\alpha|\beta)\, d\mu = \sum_{A\in\alpha}\sum_{B\in\beta} -\mu(A\cap B)\log\frac{\mu(A\cap B)}{\mu(B)}.$$

Note that $\mathrm{H}_\mu(\alpha) = \mathrm{H}_\mu(\alpha|\{X\})$. Recalling the measure μ_B from Definition A.1.70 and defining a partition $\alpha|_B$ of B by $\alpha|_B := \{A \cap B : A \in \alpha\}$, it follows that

$$\begin{aligned}
\mathrm{H}_\mu(\alpha|\beta) &= \sum_{B\in\beta}\sum_{A\in\alpha} -\frac{\mu(A\cap B)}{\mu(B)}\log\frac{\mu(A\cap B)}{\mu(B)}\cdot\mu(B) \\
&= \sum_{B\in\beta}\sum_{A\in\alpha} -\mu_B(A)\log\mu_B(A)\cdot\mu(B) \\
&= \sum_{B\in\beta} \mathrm{H}_{\mu_B}(\alpha|_B)\cdot\mu(B).
\end{aligned}$$

Hence, the conditional entropy of α given β is the weighted average of the entropies of the partitions of each atom $B \in \beta$ into the sets $\{A \cap B : A \in \alpha\}$.

Of course, the properties of entropy (resp. conditional entropy) are inherited from the properties of the information function (resp., the conditional information function) via integration, as the following theorem shows.

Theorem 9.4.3. *Let* $(X, \mathcal{A}, \mu)$ *be a probability space and* $\alpha, \beta, \gamma \in \mathrm{Part}(X, \mathcal{A})$. *The following statements hold:*
(a) $\mathrm{H}_\mu(\alpha \vee \beta|\gamma) = \mathrm{H}_\mu(\alpha|\gamma) + \mathrm{H}_\mu(\beta|\alpha \vee \gamma)$.
(b) $\mathrm{H}_\mu(\alpha \vee \beta) = \mathrm{H}_\mu(\alpha) + \mathrm{H}_\mu(\beta|\alpha)$.
(c) *If* $\alpha \leq \beta$, *then* $\mathrm{H}_\mu(\alpha|\gamma) \leq \mathrm{H}_\mu(\beta|\gamma)$.
(d) *If* $\alpha \leq \beta$, *then* $\mathrm{H}_\mu(\alpha) \leq \mathrm{H}_\mu(\beta)$.
(e) *If* $\beta \leq \gamma$, *then* $\mathrm{H}_\mu(\alpha|\beta) \geq \mathrm{H}_\mu(\alpha|\gamma)$.
(f) $\mathrm{H}_\mu(\alpha \vee \beta|\gamma) \leq \mathrm{H}_\mu(\alpha|\gamma) + \mathrm{H}_\mu(\beta|\gamma)$.
(g) $\mathrm{H}_\mu(\alpha \vee \beta) \leq \mathrm{H}_\mu(\alpha) + \mathrm{H}_\mu(\beta)$.
(h) $\mathrm{H}_\mu(\alpha|\gamma) \leq \mathrm{H}_\mu(\alpha|\beta) + \mathrm{H}_\mu(\beta|\gamma)$.
(i) $\mathrm{H}_\mu(\alpha) \leq \mathrm{H}_\mu(\alpha|\beta) + \mathrm{H}_\mu(\beta)$.

Proof. All the statements follow from their counterparts in Theorem 9.3.5 after integration or summation over atoms. For instance, let us prove (e). If $\beta \leq \gamma$, then it follows from Theorem 9.3.5(e) that

$$\mathrm{H}_\mu(\alpha|\beta) = \int_X I_\mu(\alpha|\beta)\,d\mu = \sum_{A\in\alpha}\sum_{B\in\beta}\int_{A\cap B} I_\mu(\alpha|\beta)\,d\mu \geq \sum_{A\in\alpha}\sum_{B\in\beta}\int_{A\cap B} I_\mu(\alpha|\gamma)\,d\mu = \mathrm{H}_\mu(\alpha|\gamma). \qquad \square$$

9.4.2 Second stage: entropy of a system relative to a partition

In this second stage, we take into account the behavior of a measure-preserving dynamical system relative to a given partition. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$ and $\alpha \in \mathrm{Part}(X, \mathcal{A})$. Observe that $T^{-1}\alpha := \{T^{-1}(A) : A \in \alpha\} \in \mathrm{Part}(X, \mathcal{A})$, too.

Recall that the set X represents the set of all possible outcomes (or states) of an experiment, while the σ-algebra $\mathcal{A}$ consists of the set of all possible events, and $\mu(A)$ is the probability that event A happen. Recall also that a partition $\alpha = \{A_k\}$ can be thought of as the set of all observations that can be made with a given instrument. The action of T on $(X, \mathcal{A}, \mu)$ may be conceived as the passage of one unit of time (for instance, a day). Today would naturally be taken as reference point for time 0. Suppose that we conduct the experiment with our instrument tomorrow. The resulting observation would be one of the atoms of α, say A_{k_1}, on day 1. Due to the passage of time (in other words, one iteration of T), in order to make observation A_{k_1} at time 1, our measure-preserving system would have to be, today, in one of the states of $T^{-1}(A_{k_1})$. The probability of making observation A_{k_1} on day 1 is thus $\mu(T^{-1}(A_{k_1}))$. Assume now that we conduct the same experiment for n consecutive days, starting today. What is the probability that we make the sequence of observations $A_{k_0}, A_{k_1}, \ldots, A_{k_{n-1}}$ on those successive days? We would make those observations precisely if our system is, today, in one of the states of $\bigcap_{m=0}^{n-1} T^{-m}(A_{k_m})$. Therefore, the probability that our observations are respectively $A_{k_0}, A_{k_1}, \ldots, A_{k_{n-1}}$ on n successive days starting today is $\mu(\bigcap_{m=0}^{n-1} T^{-m}(A_{k_m}))$. It is thus natural to consider for all $0 \leq m < n$ the partitions

$$\alpha_m^n := \bigvee_{i=m}^{n-1} T^{-i}\alpha = T^{-m}\alpha \vee \cdots \vee T^{-(n-1)}\alpha.$$

If $m \geq n$, we define α_m^n to be the trivial partition $\{X\}$. To shorten notation, we shall write α^n in lieu of α_0^n and $T^{-i}\alpha$ rather than $T^{-i}(\alpha)$. In the following lemma, we list some of the basic properties of the operator T^{-1} on partitions.

Lemma 9.4.4. *Let $T : X \to X$ be a measurable transformation of a measurable space $(X, \mathcal{A})$, and let $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. The following statements hold:*
(a) $T^{-1}(\alpha \vee \beta) = (T^{-1}\alpha) \vee (T^{-1}\beta)$.
(b) $T^{-1}(\alpha_m^n) = (T^{-1}\alpha)_m^n$ *for all* $m, n \geq 0$.
(c) $(\alpha \vee \beta)_m^n = \alpha_m^n \vee \beta_m^n$ *for all* $m, n \geq 0$.
(d) $(\alpha_k^l)_m^n = \alpha_{k+m}^{l+n-1}$.
(e) T^{-1} *preserves the partial order* $\leq$, *that is, if* $\alpha \leq \beta$ *then* $T^{-1}\alpha \leq T^{-1}\beta$.
(f) *More generally, if* $\alpha \leq \beta$ *then* $\alpha_m^n \leq \beta_m^n$ *for all* $m, n \geq 0$.
(g) $(T^{-1}\alpha)(x) = T^{-1}(\alpha(T(x)))$ *for all* $x \in X$.

Proof. The proof of assertions (a) and (e) are left to the reader.

(b) Using (a) repeatedly, we have that

$$T^{-1}(\alpha_m^n) = T^{-1}\left(\bigvee_{i=m}^{n-1} T^{-i}\alpha\right) = \bigvee_{i=m}^{n-1} T^{-1}(T^{-i}\alpha) = \bigvee_{i=m}^{n-1} T^{-i}(T^{-1}\alpha) = (T^{-1}\alpha)_m^n.$$

(c) Again by using (a) repeatedly, we obtain that

$$(\alpha \vee \beta)_m^n = \bigvee_{i=m}^{n-1} T^{-i}(\alpha \vee \beta) = \bigvee_{i=m}^{n-1} (T^{-i}\alpha \vee T^{-i}\beta) = \left(\bigvee_{i=m}^{n-1} T^{-i}\alpha\right) \vee \left(\bigvee_{i=m}^{n-1} T^{-i}\beta\right) = \alpha_m^n \vee \beta_m^n.$$

(d) Using (a), it follows that

$$(\alpha_k^l)_m^n = \bigvee_{j=m}^{n-1} T^{-j}(\alpha_k^l) = \bigvee_{j=m}^{n-1} T^{-j}\left(\bigvee_{i=k}^{l-1} T^{-i}\alpha\right) = \bigvee_{j=m}^{n-1}\bigvee_{i=k}^{l-1} T^{-(i+j)}\alpha = \bigvee_{s=k+m}^{l+n-2} T^{-s}\alpha = \alpha_{k+m}^{l+n-1}.$$

(f) Suppose that $\alpha \leq \beta$. Using (e) repeatedly and Lemma 7.1.5(g), we obtain that

$$\alpha_m^n = \bigvee_{i=m}^{n-1} T^{-i}\alpha \leq \bigvee_{i=m}^{n-1} T^{-i}\beta = \beta_m^n.$$

(g) Let $x \in X$. Choose $A \in \alpha$ such that $x \in T^{-1}(A)$. Then $T(x) \in A$, that is, $A = \alpha(T(x))$. Hence, $(T^{-1}\alpha)(x) = T^{-1}(A) = T^{-1}(\alpha(T(x)))$. □

We now describe the behavior of the operator T^{-1} with respect to the information function for any measure-preserving dynamical system T.

Lemma 9.4.5. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. Then*

$$I_\mu(T^{-1}\alpha | T^{-1}\beta) = I_\mu(\alpha|\beta) \circ T.$$

In particular,

$$I_\mu(T^{-1}\alpha) = I_\mu(\alpha) \circ T.$$

Proof. Let $x \in X$. By Lemma 9.4.4(a) and (g) and the assumption that μ is T-invariant, we have that

$$\begin{aligned}
I_\mu(T^{-1}\alpha|T^{-1}\beta)(x) &= -\log \frac{\mu\big((T^{-1}\alpha \vee T^{-1}\beta)(x)\big)}{\mu((T^{-1}\beta)(x))} \\
&= -\log \frac{\mu\big((T^{-1}(\alpha \vee \beta))(x)\big)}{\mu((T^{-1}\beta)(x))} \\
&= -\log \frac{\mu\big(T^{-1}((\alpha \vee \beta)(T(x)))\big)}{\mu\big(T^{-1}(\beta(T(x)))\big)} \\
&= -\log \frac{\mu\big((\alpha \vee \beta)(T(x))\big)}{\mu(\beta(T(x)))} \\
&= I_\mu(\alpha|\beta)\big(T(x)\big) = I_\mu(\alpha|\beta) \circ T(x).
\end{aligned}$$

Set $\beta = \{X\}$ to get the particular, unconditional case. $\square$

A more intricate property of the information function is the following.

Lemma 9.4.6. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha \in \mathrm{Part}(X, \mathcal{A})$. For all $n \in \mathbb{N}$,*

$$I_\mu(\alpha^n) = \sum_{j=1}^{n} I_\mu(\alpha|\alpha_1^j) \circ T^{n-j}.$$

Proof. We will prove this lemma by induction. For $n = 1$, since α_1^1 is by definition equal to the trivial partition $\{X\}$, we have

$$I_\mu(\alpha^1) = I_\mu(\alpha) = I_\mu(\alpha|\{X\}) = I_\mu(\alpha|\alpha_1^1) = I_\mu(\alpha|\alpha_1^1) \circ T^{1-1}.$$

Now suppose that the lemma holds for some $n \in \mathbb{N}$. Then, in light of Theorem 9.3.5(b) and Lemma 9.4.5, we obtain that

$$\begin{aligned}
I_\mu(\alpha^{n+1}) &= I_\mu(\alpha \vee \alpha_1^{n+1}) = I_\mu(\alpha_1^{n+1}) + I_\mu(\alpha|\alpha_1^{n+1}) \\
&= I_\mu(T^{-1}(\alpha^n)) + I_\mu(\alpha|\alpha_1^{n+1}) = I_\mu(\alpha^n) \circ T + I_\mu(\alpha|\alpha_1^{n+1})
\end{aligned}$$

$$= \sum_{j=1}^{n} I_\mu(\alpha|\alpha_1^j) \circ T^{n-j} \circ T + I_\mu(\alpha|\alpha_1^{n+1})$$
$$= \sum_{j=1}^{n} I_\mu(\alpha|\alpha_1^j) \circ T^{n+1-j} + I_\mu(\alpha|\alpha_1^{n+1}) \circ T^{n+1-(n+1)}$$
$$= \sum_{j=1}^{n+1} I_\mu(\alpha|\alpha_1^j) \circ T^{n+1-j}. \qquad \square$$

We now turn our attention to the effect that a measure-preserving dynamical system has on entropy. In particular, observe that because the system is measure-preserving, conducting the experiment today or tomorrow (or at any time in the future) gives us the same amount of average information about the outcome. This is the meaning of the second of the following properties of entropy.

Lemma 9.4.7. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. The following statements hold:*
(a) $\mathrm{H}_\mu(T^{-1}\alpha|T^{-1}\beta) = \mathrm{H}_\mu(\alpha|\beta)$.
(b) $\mathrm{H}_\mu(T^{-1}\alpha) = \mathrm{H}_\mu(\alpha)$.
(c) $\mathrm{H}_\mu(\alpha^n|\beta^n) \le n\,\mathrm{H}_\mu(\alpha|\beta)$ *for all* $n \in \mathbb{N}$.

Proof. (a) Using Lemma 9.4.5 and the T-invariance of μ, we have that

$$\mathrm{H}_\mu(T^{-1}\alpha|T^{-1}\beta) = \int_X I_\mu(T^{-1}\alpha|T^{-1}\beta)\,d\mu = \int_X I_\mu(\alpha|\beta) \circ T\,d\mu = \int_X I_\mu(\alpha|\beta)\,d\mu = \mathrm{H}_\mu(\alpha|\beta).$$

(b) Set $\beta = \{X\}$ in (a) to obtain (b).
(c) We first prove that $\mathrm{H}_\mu(\alpha^n|\beta^n) \le \sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-j}\alpha|T^{-j}\beta)$. This statement clearly holds when $n = 1$. Suppose that it holds for some $n \in \mathbb{N}$. Using Theorem 9.4.3(a) and (e), we have that

$$\begin{aligned}
\mathrm{H}_\mu(\alpha^{n+1}|\beta^{n+1}) &= \mathrm{H}_\mu(\alpha^n \vee T^{-n}\alpha|\beta^n \vee T^{-n}\beta) \\
&= \mathrm{H}_\mu(\alpha^n|\beta^n \vee T^{-n}\beta) + \mathrm{H}_\mu(T^{-n}\alpha|\alpha^n \vee \beta^n \vee T^{-n}\beta) \\
&\le \mathrm{H}_\mu(\alpha^n|\beta^n) + \mathrm{H}_\mu(T^{-n}\alpha|T^{-n}\beta) \\
&\le \sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-j}\alpha|T^{-j}\beta) + \mathrm{H}_\mu(T^{-n}\alpha|T^{-n}\beta) \\
&= \sum_{j=0}^{n} \mathrm{H}_\mu(T^{-j}\alpha|T^{-j}\beta).
\end{aligned}$$

By induction, the above statement holds for all $n \in \mathbb{N}$. By (a), we obtain that

$$\mathrm{H}_\mu(\alpha^n|\beta^n) \le \sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-j}\alpha|T^{-j}\beta) = \sum_{j=0}^{n-1} \mathrm{H}_\mu(\alpha|\beta) = n\,\mathrm{H}_\mu(\alpha|\beta). \qquad \square$$

The average information gained by conducting an experiment on n consecutive days using the partition α is given by the entropy $\mathrm{H}_\mu(\alpha^n)$ since α^n has for atoms the sets $\bigcap_{m=0}^{n-1} T^{-m}(A_{k_m})$, where $A_{k_m} \in \alpha$ for all m. Not surprisingly, the average information gained by conducting the experiment on n consecutive days using the partition α is equal to the sum of the average conditional information gained by performing α on day $j+1$ given that the outcome of performing α over the previous j days is known, summing from the first day to the last day. This is formalized in the next lemma.

Lemma 9.4.8. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha \in \mathrm{Part}(X, \mathcal{A})$. Then for all $n \in \mathbb{N}$,*

$$\mathrm{H}_\mu(\alpha^n) = \sum_{j=1}^{n} \mathrm{H}_\mu(\alpha|\alpha_1^j).$$

Proof. We deduce from Lemma 9.4.6 and the T-invariance of μ that

$$\mathrm{H}_\mu(\alpha^n) = \int_X I_\mu(\alpha^n)\, d\mu = \sum_{j=1}^{n} \int_X I_\mu(\alpha|\alpha_1^j) \circ T^{n-j}\, d\mu = \sum_{j=1}^{n} \int_X I_\mu(\alpha|\alpha_1^j)\, d\mu = \sum_{j=1}^{n} \mathrm{H}_\mu(\alpha|\alpha_1^j). \qquad \square$$

Below is an alternative expression for the entropy $\mathrm{H}_\mu(\alpha^n)$.

Lemma 9.4.9. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha \in \mathrm{Part}(X, \mathcal{A})$. Then for all $n \in \mathbb{N}$,*

$$\mathrm{H}_\mu(\alpha^n) = \sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-j}\alpha|\alpha^j).$$

Proof. We prove this result by induction. The statement is trivial when $n = 1$. Suppose that it holds for $n-1$. Using Theorem 9.4.3(b), we get

$$\begin{aligned}\mathrm{H}_\mu(\alpha^n) &= \mathrm{H}_\mu(\alpha^{n-1} \vee T^{-(n-1)}\alpha) = \mathrm{H}_\mu(\alpha^{n-1}) + \mathrm{H}_\mu(T^{-(n-1)}\alpha|\alpha^{n-1}) \\ &= \sum_{j=0}^{n-2} \mathrm{H}_\mu(T^{-j}\alpha|\alpha^j) + \mathrm{H}_\mu(T^{-(n-1)}\alpha|\alpha^{n-1}) = \sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-j}\alpha|\alpha^j).\end{aligned}$$

So the statement holds for all $n \in \mathbb{N}$. $\square$

Returning to Lemma 9.4.8, since $\alpha_1^{j+1} \geq \alpha_1^j$ observe that $\mathrm{H}_\mu(\alpha|\alpha_1^{j+1}) \leq \mathrm{H}_\mu(\alpha|\alpha_1^j)$ by Theorem 9.4.3(e). So the sequence $\big(\mathrm{H}_\mu(\alpha|\alpha_1^j)\big)_{j=1}^\infty$ decreases to some limit which we shall denote by $\mathrm{h}_\mu(T, \alpha)$. Consequently, the corresponding sequence of Cesàro averages $\big(\frac{1}{n}\sum_{j=1}^n \mathrm{H}_\mu(\alpha|\alpha_1^j)\big)_{n=1}^\infty = \big(\frac{1}{n}\mathrm{H}_\mu(\alpha^n)\big)_{n=1}^\infty$ decreases to the same limit. Thus the following definition makes sense. This is the second step in the definition of the measure-theoretic entropy of a system.

Definition 9.4.10. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha \in \mathrm{Part}(X, \mathcal{A})$. The *entropy of T with respect to* α, denoted $\mathrm{h}_\mu(T, \alpha)$, is defined by

$$\begin{aligned}\mathrm{h}_\mu(T,\alpha) &:= \lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\alpha_1^n) = \lim_{n\to\infty} \frac{1}{n}\mathrm{H}_\mu(\alpha^n)\\ &= \inf_{n\to\infty} \mathrm{H}_\mu(\alpha|\alpha_1^n) = \inf_{n\to\infty} \frac{1}{n}\mathrm{H}_\mu(\alpha^n).\end{aligned}$$

The following theorem lists some of the basic properties of $\mathrm{h}_\mu(T, \cdot)$.

Theorem 9.4.11. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. The following statements hold:*
(a) $\mathrm{h}_\mu(T, \alpha) \le \mathrm{H}_\mu(\alpha)$.
(b) $\mathrm{h}_\mu(T, \alpha) \le \mathrm{H}_\mu(T^{-1}\alpha|\alpha)$.
(c) $\mathrm{h}_\mu(T, \alpha \vee \beta) \le \mathrm{h}_\mu(T, \alpha) + \mathrm{h}_\mu(T, \beta)$.
(d) *If $\alpha \le \beta$, then* $\mathrm{h}_\mu(T, \alpha) \le \mathrm{h}_\mu(T, \beta)$.
(e) $\mathrm{h}_\mu(T, \alpha) \le \mathrm{h}_\mu(T, \beta) + \mathrm{H}_\mu(\alpha|\beta)$.
(f) $\mathrm{h}_\mu(T, T^{-1}\alpha) = \mathrm{h}_\mu(T, \alpha)$.
(g) $\mathrm{h}_\mu(T, \alpha^k) = \mathrm{h}_\mu(T, \alpha)$ *for all* $k \in \mathbb{N}$.
(h) $\mathrm{h}_\mu(T^k, \alpha^k) = k \cdot \mathrm{h}_\mu(T, \alpha)$ *for all* $k \in \mathbb{N}$.
(i) *If T is invertible, then* $\mathrm{h}_\mu(T, \alpha) = \mathrm{h}_\mu(T, \bigvee_{i=-k}^{k} T^i\alpha)$ *for all* $k \in \mathbb{N}$.
(j) *If* $(\beta_n)_{n=1}^\infty$ *is a sequence in* $\mathrm{Part}(X, \mathcal{A})$ *such that* $\lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\beta_n) = 0$, *then*

$$\mathrm{h}_\mu(T, \alpha) \le \liminf_{n\to\infty} \mathrm{h}_\mu(T, \beta_n).$$

(k) *If* $\lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\beta^n) = 0$, *then* $\mathrm{h}_\mu(T, \alpha) \le \mathrm{h}_\mu(T, \beta)$.

Proof. (a) This follows from the fact that $\mathrm{h}_\mu(T, \alpha) = \inf_{n\in\mathbb{N}} \frac{1}{n}\mathrm{H}_\mu(\alpha^n)$.
(b) Using Lemmas 9.4.9 and 9.4.7(a) and Theorem 9.4.3(e), we have

$$\begin{aligned}\mathrm{h}_\mu(T,\alpha) &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-j}\alpha|\alpha^j)\\ &\le \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-j}\alpha\big|T^{-(j-1)}\alpha)\\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-(j-1)}(T^{-1}\alpha)\big|T^{-(j-1)}\alpha)\\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-1}\alpha|\alpha) = \mathrm{H}_\mu(T^{-1}\alpha|\alpha).\end{aligned}$$

(c) Using Lemma 9.4.4(c) and Theorem 9.4.3(g), we get

$$\begin{aligned}\mathrm{h}_\mu(T,\alpha\vee\beta) &= \lim_{n\to\infty} \frac{1}{n}\mathrm{H}_\mu((\alpha\vee\beta)^n)\\ &= \lim_{n\to\infty} \frac{1}{n}\mathrm{H}_\mu(\alpha^n\vee\beta^n)\end{aligned}$$

$$\begin{aligned}&\leq \lim_{n\to\infty}\frac{1}{n}\big[\mathrm{H}_\mu(\alpha^n)+\mathrm{H}_\mu(\beta^n)\big]\\&=\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\alpha^n)+\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\beta^n)\\&=\mathrm{h}_\mu(T,\alpha)+\mathrm{h}_\mu(T,\beta).\end{aligned}$$

(d) If $\alpha \leq \beta$, then $\alpha^n \leq \beta^n$ and hence $\mathrm{H}_\mu(\alpha^n) \leq \mathrm{H}_\mu(\beta^n)$ for all $n \in \mathbb{N}$. Therefore,

$$\mathrm{h}_\mu(T,\alpha)=\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\alpha^n)\leq \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\beta^n)=\mathrm{h}_\mu(T,\beta).$$

(e) Calling upon Theorem 9.4.3(i) and Lemma 9.4.7(c), we obtain that

$$\begin{aligned}\mathrm{h}_\mu(T,\alpha)&=\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\alpha^n)\\&\leq \liminf_{n\to\infty}\frac{1}{n}\big[\mathrm{H}_\mu(\alpha^n|\beta^n)+\mathrm{H}_\mu(\beta^n)\big]\\&=\liminf_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\alpha^n|\beta^n)+\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\beta^n)\\&\leq \mathrm{H}_\mu(\alpha|\beta)+\mathrm{h}_\mu(T,\beta).\end{aligned}$$

(f) By Lemma 9.4.4(b), we know that $(T^{-1}\alpha)^n = T^{-1}(\alpha^n)$ for all $n \in \mathbb{N}$. Then, using Lemma 9.4.7(b), we deduce that

$$\mathrm{h}_\mu(T,T^{-1}\alpha)=\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu\big((T^{-1}\alpha)^n\big)=\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(T^{-1}(\alpha^n))=\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\alpha^n)=\mathrm{h}_\mu(T,\alpha).$$

(g) By Lemma 9.4.4(d), we know that $(\alpha^k)^n = \alpha^{n+k-1}$ and hence

$$\begin{aligned}\mathrm{h}_\mu(T,\alpha^k)&=\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu\big((\alpha^k)^n\big)=\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\alpha^{n+k-1})\\&=\lim_{n\to\infty}\frac{n+k-1}{n}\cdot\frac{1}{n+k-1}\mathrm{H}_\mu(\alpha^{n+k-1})\\&=\lim_{n\to\infty}\frac{n+k-1}{n}\cdot\lim_{n\to\infty}\frac{1}{n+k-1}\mathrm{H}_\mu(\alpha^{n+k-1})\\&=\lim_{m\to\infty}\frac{1}{m}\mathrm{H}_\mu(\alpha^m)=\mathrm{h}_\mu(T,\alpha).\end{aligned}$$

(h) Let $k \in \mathbb{N}$. Then

$$\begin{aligned}\mathrm{h}_\mu(T^k,\alpha^k)&=\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu\left(\bigvee_{j=0}^{n-1}T^{-kj}(\alpha^k)\right)\\&=\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu\left(\bigvee_{j=0}^{n-1}T^{-kj}\left(\bigvee_{i=0}^{k-1}T^{-i}\alpha\right)\right)\\&=\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu\left(\bigvee_{l=0}^{kn-1}T^{-l}\alpha\right)\\&=k\lim_{n\to\infty}\frac{1}{kn}\mathrm{H}_\mu(\alpha^{kn})=k\,\mathrm{h}_\mu(T,\alpha).\end{aligned}$$

(i) The proof is similar to that of part (f) and is thus left to the reader.
(j) Let $(\beta_n)_{n=1}^\infty$ be a sequence of partitions such that $\lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\beta_n) = 0$. By part (e),

$$\mathrm{h}_\mu(T,\alpha) \le \mathrm{h}_\mu(T,\beta_n) + \mathrm{H}_\mu(\alpha|\beta_n), \quad \forall n \in \mathbb{N}.$$

Consequently,

$$\begin{aligned}\mathrm{h}_\mu(T,\alpha) &\le \liminf_{n\to\infty}[\mathrm{h}_\mu(T,\beta_n) + \mathrm{H}_\mu(\alpha|\beta_n)] \\ &= \liminf_{n\to\infty} \mathrm{h}_\mu(T,\beta_n) + \lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\beta_n) = \liminf_{n\to\infty} \mathrm{h}_\mu(T,\beta_n).\end{aligned}$$

(k) Suppose that $\lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\beta^n) = 0$. By parts (j) and (g), we have

$$\mathrm{h}_\mu(T,\alpha) \le \liminf_{n\to\infty} \mathrm{h}_\mu(T,\beta^n) = \mathrm{h}_\mu(T,\beta). \qquad \square$$

9.4.3 Third and final stage: entropy of a system

The measure-theoretic entropy of a system is defined in a similar way to topological entropy. The third and last step in the definition consists in passing to a supremum.

Definition 9.4.12. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. The *measure-theoretic entropy* of T, denoted $\mathrm{h}_\mu(T)$, is defined by

$$\mathrm{h}_\mu(T) := \sup\{\mathrm{h}_\mu(T,\alpha) : \alpha \in \mathrm{Part}_{\mathrm{Fin}}(X,\mathcal{A})\},$$

where

$$\mathrm{Part}_{\mathrm{Fin}}(X,\mathcal{A}) := \{\alpha \in \mathrm{Part}(X,\mathcal{A}) : \#\alpha < \infty\}.$$

The following theorem is a useful tool for calculating the measure-theoretic entropy of a system. The first part is analogous to Theorem 7.2.19.

Theorem 9.4.13. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. The following statements hold:*
(a) $\mathrm{h}_\mu(T^k) = k \cdot \mathrm{h}_\mu(T)$ *for all* $k \in \mathbb{N}$.
(b) *If T is invertible, then* $\mathrm{h}_\mu(T^{-1}) = \mathrm{h}_\mu(T)$.

Proof. (a) Let $k \in \mathbb{N}$. Using Theorem 9.4.11(h), we get

$$\begin{aligned} k\,\mathrm{h}_\mu(T) &= \sup\{k\,\mathrm{h}_\mu(T,\alpha) : \alpha \in \mathrm{Part}_{\mathrm{Fin}}(X,\mathcal{A})\} \\ &= \sup\{\mathrm{h}_\mu(T^k,\alpha^k) : \alpha \in \mathrm{Part}_{\mathrm{Fin}}(X,\mathcal{A})\} \\ &\le \sup\{\mathrm{h}_\mu(T^k,\beta) : \beta \in \mathrm{Part}_{\mathrm{Fin}}(X,\mathcal{A})\} = \mathrm{h}_\mu(T^k).\end{aligned}$$

On the other hand, since $\alpha \le \alpha^k$ for all $k \in \mathbb{N}$, Theorem 9.4.11(d) and (h) give

$$\mathrm{h}_\mu(T^k, \alpha) \le \mathrm{h}_\mu(T^k, \alpha^k) = k\,\mathrm{h}_\mu(T, \alpha).$$

Passing to the supremum over all finite partitions α of X on both sides, we obtain the desired inequality, namely, $\mathrm{h}_\mu(T^k) \le k\,\mathrm{h}_\mu(T)$.

(b) To distinguish the action of T from the action of T^{-1} on a partition, we shall use the respective notation α_T^n and $\alpha_{T^{-1}}^n$. Using Lemmas 9.4.7(b) and 9.4.4(b) in turn, we deduce that

$$\begin{aligned}
\mathrm{H}_\mu(\alpha_{T^{-1}}^n) = \mathrm{H}_\mu\left(\bigvee_{i=0}^{n-1} (T^{-1})^{-i}\alpha\right) &= \mathrm{H}_\mu\left(\bigvee_{i=0}^{n-1} T^{i}\alpha\right) \\
&= \mathrm{H}_\mu\left(T^{-(n-1)}\left(\bigvee_{i=0}^{n-1} T^{i}\alpha\right)\right) \\
&= \mathrm{H}_\mu\left(\bigvee_{i=0}^{n-1} T^{-(n-1-i)}\alpha\right) \\
&= \mathrm{H}_\mu\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) = \mathrm{H}_\mu(\alpha_T^n).
\end{aligned}$$

It follows that $\mathrm{h}_\mu(T^{-1}, \alpha) = \mathrm{h}_\mu(T, \alpha)$ for every partition α, and thus, passing to the supremum on both sides, we conclude that $\mathrm{h}_\mu(T^{-1}) = \mathrm{h}_\mu(T)$. □

In Theorem 7.2.24, we observed that the topological entropy of an expansive dynamical system can be determined by simply calculating the entropy of that system with respect to any cover of sufficiently small diameter. We intend to prove the corresponding result for measure-theoretic entropy by the end of this section. We begin the journey to that destination with a purely measure-theoretical lemma. It says that given a finite Borel partition α of a compact metric space X and given any Borel partition β of X of sufficiently small diameter, we can group the atoms of β together in such a way that we nearly reconstruct the partition α. Notice that β may be countably infinite.

To simplify notation, we shall write $\mathrm{Part}(X) := \mathrm{Part}(X, \mathcal{B}(X))$.

Lemma 9.4.14. *Let X be a compact metric space and $\mu \in M(X)$. Let also $\alpha = \{A_1, A_2, \dots, A_n\} \in \mathrm{Part}_{\mathrm{Fin}}(X)$. Then for all $\varepsilon > 0$ there exists $\delta > 0$ so that for every $\beta \in \mathrm{Part}(X)$ with $\mathrm{diam}(\beta) < \delta$ there is $\beta' = \{B_1', B_2', \dots, B_n'\} \in \mathrm{Part}_{\mathrm{Fin}}(X)$ such that*

$$\beta' \le \beta \quad \textit{and} \quad \mu(B_i' \triangle A_i) < \varepsilon, \quad \forall 1 \le i \le n.$$

Proof. Fix $\varepsilon > 0$. Since μ is regular, for each $1 \le i \le n$ there exists a compact set $K_i \subseteq A_i$ such that

$$\mu(A_i \backslash K_i) < \frac{\varepsilon}{n}.$$

As usual, let d denote the metric on X and set

$$\theta = \min\{d(K_i, K_j) : i \neq j\}.$$

Then $\theta > 0$, as the sets K_i are compact and mutually disjoint. Let $\delta = \theta/2$ and let β be a partition with $\operatorname{diam}(\beta) < \delta$. For each $1 \leq i \leq n$, let

$$B'_i = \bigcup_{B \in \beta : B \cap K_i \neq \emptyset} B.$$

Then each B'_i is a Borel set such that $B'_i \supseteq K_i$. Furthermore, due to the choice of δ,

$$B'_i \cap B'_j = \emptyset, \quad \forall i \neq j.$$

However, the family of pairwise disjoint Borel sets $\{B'_i\}_{i=1}^n$ may not cover X completely. Indeed, there may be some sets $C \in \beta$ such that $C \cap \bigcup_{i=1}^n K_i = \emptyset$. Take all those sets and put them into B'_1. Then $\beta' = \{B'_i\}_{i=1}^n$ is a Borel partition of X such that $\beta' \leq \beta$. Moreover, since $B'_j \supseteq K_j$ for all $1 \leq j \leq n$, we get

$$\begin{aligned}
\mu(B'_i \triangle A_i) &= \mu(B'_i \backslash A_i) + \mu(A_i \backslash B'_i) \\
&= \mu\big((X \backslash \cup_{j \neq i} B'_j) \backslash A_i\big) + \mu(A_i \backslash B'_i) \\
&\leq \mu\big((X \backslash \cup_{j \neq i} K_j) \backslash A_i\big) + \mu(A_i \backslash K_i) \\
&= \mu\big((\cup_{k=1}^n A_k \backslash \cup_{j \neq i} K_j) \backslash A_i\big) + \mu(A_i \backslash K_i) \\
&= \mu(\cup_{k \neq i} A_k \backslash \cup_{j \neq i} K_j) + \mu(A_i \backslash K_i) \\
&\leq \mu(\cup_{j \neq i} A_j \backslash K_j) + \mu(A_i \backslash K_i) \\
&= \sum_{j=1}^n \mu(A_j \backslash K_j) < n \cdot \frac{\varepsilon}{n} = \varepsilon.
\end{aligned}$$

□

From the above result, we will show that the conditional entropy of a partition α given a partition β can be made as small as desired provided that β has a small enough diameter. Indeed, from Theorem 9.4.3(e), given partitions α, β and β' as in the above lemma, we have that $\mathrm{H}_\mu(\alpha|\beta) \leq \mathrm{H}_\mu(\alpha|\beta')$, where the partition β' is designed to resemble the partition α. In order to estimate the conditional entropy $\mathrm{H}_\mu(\alpha|\beta')$, we must estimate the contribution of all atoms of the partition $\alpha \vee \beta'$. There are essentially two kinds of atoms to be taken into account, namely, atoms of the form $A_i \cap B'_i$ and atoms of the form $A_i \cap B'_j$ with $i \neq j$. Intuitively, because A_i is more or less equal to B'_i (after all, $\mu(A_i \triangle B'_i)$ is small), the information provided by A_i assuming that measurement β' resulted in B'_i is small. On the other hand, since A_i is nearly disjoint from B'_j when $i \neq j$ (given that A_i is close to B'_i and $B'_i \cap B'_j = \emptyset$), the information obtained from getting A_i given that observation B'_j occurred is also small. This is what we now prove rigorously. The proof will make use of the function k from Definition 9.3.4.

Lemma 9.4.15. *Let X be a compact metric space and $\mu \in M(X)$. Let also $\alpha \in \mathrm{Part}_{\mathrm{Fin}}(X)$. For every $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\beta \in \mathrm{Part}(X),\ \mathrm{diam}(\beta) < \delta \Longrightarrow \mathrm{H}_\mu(\alpha|\beta) < \varepsilon.$$

Proof. Let $\alpha = \{A_1, A_2, \ldots, A_n\}$ be a finite Borel partition of X. As atoms of measure zero do not affect conditional entropy, we may assume that $\mu(A_i) > 0$ for all $1 \le i \le n$. Fix $\varepsilon > 0$ and let $0 < \overline{\varepsilon} < e^{-1}$ be so small that

$$\max\{k(\overline{\varepsilon}), k(1-\overline{\varepsilon})\} < \frac{\varepsilon}{2n}.$$

Then there exists $\widehat{\varepsilon} > 0$ such that

$$0 < \frac{\widehat{\varepsilon}}{\mu(A_i) - \widehat{\varepsilon}} < \overline{\varepsilon} \quad \text{and} \quad \frac{\mu(A_i) - \widehat{\varepsilon}}{\mu(A_i) + \widehat{\varepsilon}} > 1 - \overline{\varepsilon} \tag{9.3}$$

for all $1 \le i \le n$. In particular, the left relation in (9.3) imposes that $\mu(A_i) > \widehat{\varepsilon}$ for all i. Let $\delta > 0$ be the number ascribed to $\widehat{\varepsilon}$ in Lemma 9.4.14. Let β be a partition with $\mathrm{diam}(\beta) < \delta$, and let $\beta' = \{B_1', B_2', \ldots, B_n'\} \le \beta$ be such that $\mu(A_i \triangle B_i') \le \widehat{\varepsilon}$ for all $1 \le i \le n$, also as prescribed in Lemma 9.4.14. Since $\mu(A_i) > \widehat{\varepsilon}$ for all i, this implies that $\mu(B_i') > 0$ for all i. Moreover,

$$\left|\mu(A_i) - \mu(B_i')\right| \le \mu(A_i \triangle B_i') \le \widehat{\varepsilon}. \tag{9.4}$$

Therefore,

$$0 < \mu(A_i) - \widehat{\varepsilon} \le \mu(A_i) - \mu(A_i \triangle B_i') \le \mu(B_i') \le \mu(A_i) + \mu(A_i \triangle B_i') \le \mu(A_i) + \widehat{\varepsilon}. \tag{9.5}$$

Hence,

$$\mu(A_i \cap B_i') = \mu(A_i) - \mu(A_i \backslash B_i') \ge \mu(A_i) - \mu(A_i \triangle B_i') \ge \mu(A_i) - \widehat{\varepsilon} > 0.$$

Using this, (9.5) and (9.3), we get

$$\frac{\mu(A_i \cap B_i')}{\mu(B_i')} \ge \frac{\mu(A_i) - \widehat{\varepsilon}}{\mu(A_i) + \widehat{\varepsilon}} > 1 - \overline{\varepsilon}.$$

By choice of $\overline{\varepsilon}$, the function k is decreasing on the interval $[1-\overline{\varepsilon}, 1] \subseteq [1-e^{-1}, 1] \subseteq [e^{-1}, 1]$, and thus

$$k\left(\frac{\mu(A_i \cap B_i')}{\mu(B_i')}\right) \le k(1-\overline{\varepsilon}) < \frac{\varepsilon}{2n} \tag{9.6}$$

for all i.

Now, suppose that $i \neq j$. Since $\alpha = \{A_k\}_{k=1}^n$ is a partition of X, we know that $A_i \cap B'_j \subseteq B'_j \setminus A_j \subseteq A_j \triangle B'_j$. Using this, (9.5), (9.4) and (9.3), we infer that

$$\frac{\mu(A_i \cap B'_j)}{\mu(B'_j)} \leq \frac{\mu(A_j \triangle B'_j)}{\mu(A_j) - \mu(A_j \triangle B'_j)} \leq \frac{\widehat{\varepsilon}}{\mu(A_j) - \widehat{\varepsilon}} < \overline{\varepsilon}.$$

By choice of $\overline{\varepsilon}$, the function k is increasing on the interval $[0, \overline{\varepsilon}] \subseteq [0, e^{-1}]$, and hence

$$k\left(\frac{\mu(A_i \cap B'_j)}{\mu(B'_j)}\right) \leq k(\overline{\varepsilon}) < \frac{\varepsilon}{2n} \tag{9.7}$$

for all $i \neq j$. Then, by Theorem 9.4.3(e) and (9.6)–(9.7), we have

$$\begin{aligned}
\mathrm{H}_\mu(\alpha|\beta) \leq \mathrm{H}_\mu(\alpha|\beta') &= \sum_{A \in \alpha} \sum_{B' \in \beta'} -\mu(A \cap B') \log \frac{\mu(A \cap B')}{\mu(B')} \\
&= \sum_{i,j=1}^n \mu(B'_j) k\left(\frac{\mu(A_i \cap B'_j)}{\mu(B'_j)}\right) \\
&= \sum_{i=1}^n \mu(B'_i) k\left(\frac{\mu(A_i \cap B'_i)}{\mu(B'_i)}\right) + \sum_{\substack{i,j=1 \\ i \neq j}}^n \mu(B'_j) k\left(\frac{\mu(A_i \cap B'_j)}{\mu(B'_j)}\right) \\
&< \sum_{i=1}^n \mu(B'_i) \frac{\varepsilon}{2n} + \sum_{i=1}^n \sum_{j=1}^n \mu(B'_j) \frac{\varepsilon}{2n} = \frac{\varepsilon}{2n} + \sum_{i=1}^n \frac{\varepsilon}{2n} = \frac{\varepsilon}{2n} + n \cdot \frac{\varepsilon}{2n} \\
&\leq \varepsilon.
\end{aligned}$$

□

From the above lemma, we can infer that any sequence of partitions whose diameters tend to 0 provide asymptotically as much information as any given finite partition can.

Corollary 9.4.16. *Let X be a compact metric space and $\mu \in M(X)$. Let also $(\alpha_n)_{n=1}^\infty$ be a sequence in* $\mathrm{Part}(X)$ *such that* $\lim_{n\to\infty} \mathrm{diam}(\alpha_n) = 0$. *Then*

$$\lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\alpha_n) = 0$$

for every $\alpha \in \mathrm{Part}_{\mathrm{Fin}}(X)$.

Proof. Let α be a finite Borel partition of X. By Lemma 9.4.15, for every $\varepsilon > 0$ there exists a $\delta > 0$ such that if $\mathrm{diam}(\beta) < \delta$ then $\mathrm{H}_\mu(\alpha|\beta) < \varepsilon$. Since $\mathrm{diam}(\alpha_n) \to 0$, it follows that $\mathrm{H}_\mu(\alpha|\alpha_n) \to 0$ as $n \to \infty$. □

The above corollary about conditional entropy of partitions allows us to deduce the following result on the measure-theoretic entropy of a system. This is the counterpart of Lemma 7.2.20.

Theorem 9.4.17. *Let X be a compact metric space and $\mu \in M(X)$. Let also $T : X \to X$ be a measure-preserving dynamical system on $(X, \mathcal{B}(X), \mu)$ and $(\alpha_n)_{n=1}^{\infty}$ be a sequence in* $\mathrm{Part}_{\mathrm{Fin}}(X)$ *such that* $\lim_{n\to\infty} \mathrm{diam}(\alpha_n) = 0$. *Then*

$$\mathrm{h}_\mu(T) = \lim_{n\to\infty} \mathrm{h}_\mu(T, \alpha_n).$$

Proof. Let α be a finite partition of X consisting of Borel sets. By Corollary 9.4.16, we know that $\lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\alpha_n) = 0$. By Theorem 9.4.11(j), it follows that

$$\mathrm{h}_\mu(T, \alpha) \le \liminf_{n\to\infty} \mathrm{h}_\mu(T, \alpha_n) \le \limsup_{n\to\infty} \mathrm{h}_\mu(T, \alpha_n) \le \mathrm{h}_\mu(T).$$

Since this is true for any finite Borel partition α, we deduce from a passage to the supremum that

$$\mathrm{h}_\mu(T) \le \liminf_{n\to\infty} \mathrm{h}_\mu(T, \alpha_n) \le \limsup_{n\to\infty} \mathrm{h}_\mu(T, \alpha_n) \le \mathrm{h}_\mu(T).$$

Hence, $\mathrm{h}_\mu(T) = \lim_{n\to\infty} \mathrm{h}_\mu(T, \alpha_n)$. □

We can easily deduce a counterpart to Lemma 7.2.22.

Corollary 9.4.18. *Let X be a compact metric space and $\mu \in M(X)$. Let also $T : X \to X$ be a measure-preserving dynamical system on $(X, \mathcal{B}(X), \mu)$ and $\alpha \in$* $\mathrm{Part}_{\mathrm{Fin}}(X)$ *be such that* $\lim_{n\to\infty} \mathrm{diam}(\alpha^n) = 0$. *Then*

$$\mathrm{h}_\mu(T) = \mathrm{h}_\mu(T, \alpha).$$

Proof. By Theorems 9.4.17 and 9.4.11(g), we have that

$$\mathrm{h}_\mu(T) = \lim_{n\to\infty} \mathrm{h}_\mu(T, \alpha^n) = \lim_{n\to\infty} \mathrm{h}_\mu(T, \alpha) = \mathrm{h}_\mu(T, \alpha).$$ □

Partitions α such that $\lim_{n\to\infty} \mathrm{diam}(\alpha^n) = 0$ allow us (when they exist) to find the entropy of a transformation by simply computing the entropy of the transformation with respect to one such partition. As in Definition 7.2.21, we give them a special name.

Definition 9.4.19. Let X be a compact metric space and $\mu \in M(X)$. Let also $T : X \to X$ be a measure-preserving dynamical system on $(X, \mathcal{B}(X), \mu)$. Any $\alpha \in \mathrm{Part}_{\mathrm{Fin}}(X)$ such that $\lim_{n\to\infty} \mathrm{diam}(\alpha^n) = 0$ is called a *generator* for T.

We have already seen in Lemma 7.2.23 that expansive dynamical systems admit generators.

Theorem 9.4.20. *Let $T : X \to X$ be an expansive dynamical system preserving a Borel probability measure μ. If $\alpha \in$* $\mathrm{Part}_{\mathrm{Fin}}(X)$ *satisfies* $\mathrm{diam}(\alpha) < \delta(T)$, *where $\delta(T)$ is an expansive constant for T, then α is a generator for T and* $\mathrm{h}_\mu(T) = \mathrm{h}_\mu(T, \alpha)$.

Proof. Let $\alpha = \{A_k\}_{k=1}^m$ be a finite Borel partition with $\mathrm{diam}(\alpha) < \delta(T)$. Define $\delta = (\delta(T) - \mathrm{diam}(\alpha))/2$. The finite open cover $\widetilde{\alpha} = \{B(A_k, \delta)\}_{k=1}^m$ has diameter $\mathrm{diam}(\widetilde{\alpha}) \leq \delta(T)$. Lemma 7.2.23 asserts that $\lim_{n\to\infty} \mathrm{diam}(\widetilde{\alpha}^n) = 0$. As $\mathrm{diam}(\alpha^n) \leq \mathrm{diam}(\widetilde{\alpha}^n)$ for all $n \in \mathbb{N}$, it ensues that $\lim_{n\to\infty} \mathrm{diam}(\alpha^n) = 0$ and the result thus follows from Corollary 9.4.18. □

Let us now give some examples, all but the first one of which are applications of Theorem 9.4.17, Corollary 9.4.18 and/or Theorem 9.4.20.

Example 9.4.21. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. If there exists a finite measurable set $Y \subseteq X$ of full measure, then $\mathrm{h}_\mu(T) = 0$. Indeed, for any $\beta \in \mathrm{Part}(X, \mathcal{A})$ we have

$$\mathrm{H}_\mu(\beta) = -\sum_{B\in\beta} \mu(B) \log \mu(B) = -\sum_{B\in\beta} \mu(B \cap Y) \log \mu(B \cap Y).$$

This means that the entropy of a partition of X is equal to the entropy of the projection of that partition onto Y. In other words, the entropy of a partition of X coincides with the entropy of a partition of Y. Since Y is finite, there are only finitely many such partitions. Therefore, $\mathrm{H}_\mu(\beta)$ can only take finitely many values. Consequently, the entropies $\mathrm{H}_\mu(\beta^n)$, $n \in \mathbb{N}$, can also only take finitely many values. Thus

$$\mathrm{h}_\mu(T, \beta) = \lim_{n\to\infty} \frac{1}{n} \mathrm{H}_\mu(\beta^n) = 0.$$

Since β was arbitrary, we conclude that

$$\mathrm{h}_\mu(T) = \sup\{\mathrm{h}_\mu(T, \beta) : \beta \in \mathrm{Part}_{\mathrm{Fin}}(X, \mathcal{A})\} = 0.$$

Example 9.4.22. The entropy of any homeomorphism $T : \mathbb{S}^1 \to \mathbb{S}^1$ of the unit circle $\mathbb{S}^1$ is equal to 0 with respect to any T-invariant Borel probability measure.

Indeed, let μ be any T-invariant Borel probability measure. Let α and β be finite partitions of $\mathbb{S}^1$ into intervals. Then $\alpha \vee \beta$ is a partition of $\mathbb{S}^1$ into at most $(\#\alpha + \#\beta)$ intervals since $\#(\alpha \vee \beta)$ is equal to the number of endpoints of the intervals in $\alpha \vee \beta$, which is bounded above by the sum of the number of endpoints of the intervals in α and the number of endpoints of the intervals in β. Moreover, since T is a homeomorphism, we know that $T^{-k}\alpha$ is a partition of $\mathbb{S}^1$ into $\#\alpha$ intervals for every $k \in \mathbb{N}$. Therefore, α^n is a partition of $\mathbb{S}^1$ into at most $\#(\alpha^n) \leq n \cdot \#\alpha$ intervals. Consequently,

$$0 \leq \mathrm{H}_\mu(\alpha^n) \leq \log \#(\alpha^n) \leq \log n + \log \#\alpha.$$

We deduce that

$$0 \leq \mathrm{h}_\mu(T, \alpha) = \lim_{n\to\infty} \frac{1}{n} \mathrm{H}_\mu(\alpha^n) \leq \lim_{n\to\infty} \frac{1}{n}(\log n + \log \#\alpha) = 0.$$

Now, for every $m \in \mathbb{N}$, let α_m be a partition of $\mathbb{S}^1$ into m intervals of equal length. Then $\lim_{m\to\infty} \operatorname{diam}(\alpha_m) = 0$ and, by the above result, $\mathrm{h}_\mu(T, \alpha_m) = 0$ for all $m \in \mathbb{N}$. It follows from Theorem 9.4.17 that

$$\mathrm{h}_\mu(T) = \lim_{m\to\infty} \mathrm{h}_\mu(T, \alpha_m) = 0.$$

Example 9.4.23. Let E be a finite set and $A : E \times E \to \{0, 1\}$ be an incidence matrix. We proved in Example 4.1.3 that the shift map $\sigma : E_A^\infty \to E_A^\infty$ is expanding, and hence is expansive. More precisely, any $0 < \delta < 1$ is an expansive constant when E_A^∞ is endowed with the metric $d_s(\omega, \tau) = s^{|\omega\wedge\tau|}$, for any $0 < s < 1$. Let

$$\alpha = \{[e] : e \in E\}$$

be the partition of E_A^∞ into its initial 1-cylinders. Then

$$\operatorname{diam}(\alpha) = s < 1.$$

If μ is any σ-invariant Borel probability measure, then

$$\mathrm{h}_\mu(\sigma) = \mathrm{h}_\mu(\sigma, \alpha)$$

according to Theorem 9.4.20.

In particular, let us consider the full E-shift. Let μ be the product measure determined by its value on the cylinder sets; in other words,

$$\mu\big([\omega_1\omega_2 \ldots \omega_n]\big) = \prod_{k=1}^{n} P(\omega_k),$$

where P is a probability measure on the σ-algebra of all subsets of E and $P(e) := P(\{e\})$. It was shown in Example 8.1.14 that μ is σ-invariant (and σ-ergodic per Example 8.2.32). Furthermore, it is possible to show by induction that

$$\mathrm{H}_\mu(\alpha^n) = -n \sum_{e\in E} P(e) \log P(e).$$

Thus

$$\mathrm{h}_\mu(\sigma) = \mathrm{h}_\mu(\sigma, \alpha) = \lim_{n\to\infty} \frac{1}{n} \mathrm{H}_\mu(\alpha^n) = -\sum_{e\in E} P(e) \log P(e).$$

Example 9.4.24. Let $T_n : \mathbb{S}^1 \to \mathbb{S}^1$ be the n-fold map defined by $T_n(x) = nx \pmod 1$, where $\mathbb{S}^1$ is equipped with the σ-algebra of Borel sets and with the Lebesgue measure λ. We have already seen in Example 8.1.10 that T_n preserves λ (we also showed in Example 8.2.30 that T_n is ergodic with respect to λ). Consider the partition

$$\alpha = \left\{\left[\frac{j}{n}, \frac{j+1}{n}\right) : 0 \le j < n\right\}.$$

The map T_n is expanding, and thus expansive, with any $0 < \delta < 1/n$ as expansive constant. As $\operatorname{diam}(\alpha) = 1/n$, Theorem 9.4.20 does not apply to α directly. Nevertheless, observe that

$$\alpha^k = \left\{\left[\frac{j}{n^k}, \frac{j+1}{n^k}\right) : 0 \le j < n^k\right\}$$

for all $k \in \mathbb{N}$. Thus $\operatorname{diam}(\alpha^k) = 1/n^k < 1/n$ for all $k \geq 2$. Using Theorems 9.4.20 and 9.4.11(g), we deduce that

$$\mathrm{h}_\lambda(T_n) = \mathrm{h}_\lambda(T_n, \alpha^k) = \mathrm{h}_\lambda(T_n, \alpha).$$

Moreover,

$$\begin{aligned}
\mathrm{H}_\lambda(\alpha^k) &= \sum_{I\in\alpha^k} -\lambda(I)\log\lambda(I) \\
&= \sum_{j=0}^{n^k-1} -\lambda\left(\left[\frac{j}{n^k}, \frac{j+1}{n^k}\right)\right)\log\lambda\left(\left[\frac{j}{n^k}, \frac{j+1}{n^k}\right)\right) \\
&= \sum_{j=0}^{n^k-1} -\frac{1}{n^k}\log\frac{1}{n^k} = -\log\frac{1}{n^k} = k\log n.
\end{aligned}$$

Consequently,

$$\mathrm{h}_\lambda(T_n) = \mathrm{h}_\lambda(T_n, \alpha) = \lim_{k\to\infty}\frac{1}{k}\mathrm{H}_\lambda(\alpha^k) = \log n.$$

9.5 Shannon–McMillan–Breiman theorem

The Shannon–McMillan–Breiman theorem is a central result in information theory and can be thought of as a sort of ergodic theorem for measure-theoretic entropy. Indeed, the proof relies heavily on Birkhoff's ergodic theorem (Theorem 8.2.11). It also uses the following result.

Lemma 9.5.1. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. Let $\alpha \in \operatorname{Part}(X, \mathcal{A})$. Let also*

$$f_n := I_\mu(\alpha|\alpha_1^n) \quad \textit{and} \quad f^* := \sup_{n\in\mathbb{N}} f_n.$$

Then, for all $r \in \mathbb{R}$ and all $A \in \alpha$, we have

$$\mu(\{x \in A : f^*(x) > r\}) \le \min\{\mu(A), e^{-r}\}.$$

Proof. Let $A \in \alpha$ and fix $n \in \mathbb{N}$. To shorten notation, let

$$f_n^A = -\log E(\mathbb{1}_A|\sigma(\alpha_1^n)),$$

where $\sigma(\alpha_1^n)$ is the sub-σ-algebra generated by the countable partition α_1^n. Fix $x \in A$. Then, using Example A.1.62 in Appendix A, we get

$$\begin{aligned} f_n^A(x) &= -\log E(\mathbb{1}_A | \sigma(\alpha_1^n))(x) = -\log\Bigg[\frac{1}{\mu(\alpha_1^n(x))} \int\limits_{\alpha_1^n(x)} \mathbb{1}_A \, d\mu\Bigg] \\ &= -\log \frac{\mu(A \cap \alpha_1^n(x))}{\mu(\alpha_1^n(x))} = -\log \frac{\mu(\alpha(x) \cap \alpha_1^n(x))}{\mu(\alpha_1^n(x))} = I_\mu(\alpha|\alpha_1^n)(x) \\ &= f_n(x) \end{aligned}$$

for all $x \in A$. Hence,

$$f_n = \sum_{A \in \alpha} \mathbb{1}_A \cdot f_n^A.$$

Now, for $n \in \mathbb{N}$ and $r \in \mathbb{R}$ consider the set

$$B_n^{A,r} = \Big\{x \in X : \max_{1 \le i < n} f_i^A(x) \le r \text{ while } f_n^A(x) > r\Big\}.$$

The family $\{B_n^{A,r}\}_{n=1}^\infty$ consists of mutually disjoint sets. Also, recall that $\alpha_1^n \le \alpha_1^{n+1}$, and thus $\sigma(\alpha_1^n) \subseteq \sigma(\alpha_1^{n+1})$ for each $n \in \mathbb{N}$. By definition, each f_n^A is measurable with respect to $\sigma(\alpha_1^n)$. Consequently, $B_n^{A,r} \in \sigma(\alpha_1^n)$. Then

$$\begin{aligned} \mu(A \cap B_n^{A,r}) &= \int\limits_{B_n^{A,r}} \mathbb{1}_A \, d\mu = \int\limits_{B_n^{A,r}} E(\mathbb{1}_A | \sigma(\alpha_1^n)) \, d\mu \\ &= \int\limits_{B_n^{A,r}} \exp(-f_n^A) \, d\mu \le \int\limits_{B_n^{A,r}} e^{-r} \, d\mu = e^{-r} \mu(B_n^{A,r}). \end{aligned}$$

But

$$\begin{aligned} \{x \in A : f^*(x) > r\} &= \{x \in A : \exists\, n \in \mathbb{N} \text{ such that } f_n(x) > r\} \\ &= \{x \in A : \exists\, n \in \mathbb{N} \text{ such that } f_n^A(x) > r\} = A \cap \bigcup_{n=1}^\infty B_n^{A,r}. \end{aligned}$$

Using the disjointness of the $B_n^{A,r}$'s, it ensues that

$$\mu(\{x \in A : f^*(x) > r\}) = \sum_{n=1}^\infty \mu(A \cap B_n^{A,r}) \le \sum_{n=1}^\infty e^{-r} \mu(B_n^{A,r}) = e^{-r} \mu\Bigg(\bigcup_{n=1}^\infty B_n^{A,r}\Bigg) \le e^{-r}. \qquad \square$$

Corollary 9.5.2. *In addition to the hypotheses of Lemma* 9.5.1, *assume that* $\mathrm{H}_\mu(\alpha) < \infty$. *Then* $f^* \in L^1(X, \mathcal{A}, \mu)$ *and* $\|f^*\|_1 \le \mathrm{H}_\mu(\alpha) + 1$.

Proof. Since $f^* \ge 0$, we have $\int_X |f^*| \, d\mu = \int_X f^* \, d\mu$. Using Lemmas 9.5.1 and A.1.37, we obtain

$$\begin{aligned} \|f^*\|_1 &= \int\limits_X f^* \, d\mu = \sum_{A \in \alpha} \int\limits_A f^* \, d\mu \\ &= \sum_{A \in \alpha} \int\limits_0^\infty \mu(\{x \in A : f^*(x) > r\}) \, dr \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{A\in\alpha} \int_0^\infty \min\{\mu(A), e^{-r}\}\,dr \\
&= \sum_{A\in\alpha} \left(\int_0^{-\log\mu(A)} \mu(A)\,dr + \int_{-\log\mu(A)}^{\infty} e^{-r}\,dr \right) \\
&= \sum_{A\in\alpha} \left(-\mu(A)\log\mu(A) + \left[-e^{-r}\right]_{-\log\mu(A)}^{\infty} \right) \\
&= \sum_{A\in\alpha} -\mu(A)\log\mu(A) + \sum_{A\in\alpha} \mu(A) \\
&= \mathrm{H}_\mu(\alpha) + 1 < \infty. \qquad \square
\end{aligned}$$

Corollary 9.5.3. *Under the same hypotheses as Corollary* 9.5.2*, the sequence* $(f_n)_{n=1}^\infty$ *converges* μ*-a. e. and in* $L^1(X, \mathcal{A}, \mu)$*.*

Proof. Recall that $\alpha_1^n \leq \alpha_1^{n+1}$, and thus $\sigma(\alpha_1^n) \subseteq \sigma(\alpha_1^{n+1})$ for each $n \in \mathbb{N}$. For any $x \in A \in \alpha$, we have $f_n(x) = f_n^A(x) = -\log E(\mathbb{1}_A|\sigma(\alpha_1^n))(x)$ and Doob's martingale convergence theorem for conditional expectations (Theorem A.1.67) guarantees that $\lim_{n\to\infty} E(\mathbb{1}_A|\sigma(\alpha_1^n))$ exists μ-almost everywhere. Hence, the sequence of nonnegative functions $(f_n)_{n=1}^\infty$ converges μ-a. e. to some limit function $g \geq 0$. Since $|f_n| = f_n \leq f^*$ for all n, we have $|g| = g \leq f^*$, and thus $|f_n - g| \leq 2f^*$ μ-almost everywhere. Applying Lebesgue's dominated convergence theorem (Theorem A.1.38) to the sequence $(|f_n - g|)_{n=1}^\infty$, we obtain

$$\lim_{n\to\infty} \|f_n - g\|_1 = \lim_{n\to\infty} \int_X |f_n - g|\,d\mu = \int_X \lim_{n\to\infty} |f_n - g|\,d\mu = 0.$$

In other words, $f_n \to g$ in $L^1(X, \mathcal{A}, \mu)$. $\square$

We are finally in a position to prove the main result of this section and chapter.

Theorem 9.5.4 (Shannon–McMillan–Breiman theorem). *Let* $T : X \to X$ *be a measure-preserving dynamical system on a probability space* $(X, \mathcal{A}, \mu)$ *and let* $\alpha \in \mathrm{Part}(X, \mathcal{A})$ *be such that* $\mathrm{H}_\mu(\alpha) < \infty$*. Then the following limits exist:*

$$f := \lim_{n\to\infty} I_\mu(\alpha|\alpha_1^n) \quad \textit{and} \quad \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j = E(f|\mathcal{I}_\mu) \quad \mu\textit{-a. e.},$$

where $\mathcal{I}_\mu$ *is the sub-*σ*-algebra of all* μ*-almost* T*-invariant sets (see Definition* 8.2.5*).*

Moreover, the following statements hold:

(a) $\lim_{n\to\infty} \frac{1}{n} I_\mu(\alpha^n) = E(f|\mathcal{I}_\mu)$ *μ-a. e. and in* $L^1(\mu)$*.*

(b) $\mathrm{h}_\mu(T, \alpha) = \lim_{n\to\infty} \frac{1}{n}\mathrm{H}_\mu(\alpha^n) = \int_X E(f|\mathcal{I}_\mu)\,d\mu = \int_X f\,d\mu$.

Proof. According to Corollary 9.5.3, the first sequence of functions $(f_n)_{n=1}^{\infty} = (I_\mu(\alpha|\alpha_1^n))_{n=1}^{\infty}$ converges μ-a. e. to an integrable function f. The second limit exists by virtue of Birkhoff's ergodic theorem (Theorem 8.2.11). Note also that all functions $(f_n)_{n=1}^{\infty}$ are nonnegative, and thus so are f and $E(f|\mathcal{I}_\mu)$. To prove the remaining two statements, let us first assume that (a) holds and derive (b) from it. Then we will prove (a).

Let us assume that (a) holds. Using Scheffé's lemma (Lemma A.1.39) and the fact that $(I_\mu(\alpha^n))_{n=1}^{\infty}$ and $E(f|\mathcal{I}_\mu)$ are nonnegative, we obtain that

$$\mathrm{h}_\mu(T,\alpha) = \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\alpha^n) = \lim_{n\to\infty}\frac{1}{n}\int_X I_\mu(\alpha^n)\,d\mu = \int_X E(f|\mathcal{I}_\mu)\,d\mu = \int_X f\,d\mu.$$

This establishes (b).

In order to prove (a), first notice that by Lemma 9.4.6 we have

$$I_\mu(\alpha^n) = \sum_{k=1}^{n} I_\mu(\alpha|\alpha_1^k)\circ T^{n-k} = \sum_{j=0}^{n-1} I_\mu(\alpha|\alpha_1^{n-j})\circ T^j = \sum_{j=0}^{n-1} f_{n-j}\circ T^j.$$

Then, by the triangle inequality,

$$\begin{aligned}\left|\frac{1}{n}I_\mu(\alpha^n) - E(f|\mathcal{I}_\mu)\right| &= \left|\frac{1}{n}\sum_{j=0}^{n-1}(f_{n-j}\circ T^j - f\circ T^j) + \frac{1}{n}\sum_{j=0}^{n-1} f\circ T^j - E(f|\mathcal{I}_\mu)\right| \\ &\le \left|\frac{1}{n}\sum_{j=0}^{n-1}(f_{n-j}-f)\circ T^j\right| + \left|\frac{1}{n}S_n f - E(f|\mathcal{I}_\mu)\right| \\ &\le \frac{1}{n}\sum_{j=0}^{n-1}\left|f_{n-j}-f\right|\circ T^j + \left|\frac{1}{n}S_n f - E(f|\mathcal{I}_\mu)\right|. \end{aligned} \tag{9.8}$$

Birkhoff's ergodic theorem (Theorem 8.2.11) asserts that the second term on the right-hand side tends to 0 in $L^1(\mu)$. Let us now investigate the first term on that right-hand side. Set $g_n = |f_n - f|$. Since $(f_n)_{n=1}^{\infty}$ converges to f in $L^1(\mu)$ according to Corollary 9.5.3, the sequence $(g_n)_{n=1}^{\infty}$ converges to 0 in $L^1(\mu)$. So do its Cesàro averages $(\frac{1}{n}\sum_{i=1}^{n} g_i)_{n=1}^{\infty}$. Then

$$\begin{aligned}\lim_{n\to\infty}\left\|\frac{1}{n}\sum_{j=0}^{n-1} g_{n-j}\circ T^j\right\|_1 &= \lim_{n\to\infty}\int_X\left|\frac{1}{n}\sum_{j=0}^{n-1} g_{n-j}\circ T^j\right| d\mu \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\int_X g_{n-j}\circ T^j\,d\mu \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\int_X g_{n-j}\,d\mu\end{aligned}$$

$$= \lim_{n\to\infty} \int_X \frac{1}{n} \sum_{i=1}^{n} g_i \, d\mu$$
$$= 0.$$

That is, the functions $\frac{1}{n}\sum_{j=0}^{n-1} g_{n-j} \circ T^j$ converge to 0 in $L^1(\mu)$. Thus the first term on the right-hand side of (9.8), like the second term, converges to 0 in $L^1(\mu)$. This implies that the sequence $(\frac{1}{n} I_\mu(\alpha^n))_{n=1}^{\infty}$ converges to $E(f|\mathcal{I}_\mu)$ in $L^1(\mu)$.

It only remains to show convergence μ-a. e. of that same sequence. To this end, for each $N \in \mathbb{N}$ let $G_N = \sup_{n\geq N} g_n$. The sequence of functions $(G_N)_{N=1}^{\infty}$ is decreasing and bounded below by 0, so it converges to some function. As $f_n \to f$ μ-a. e., we know that $g_n = |f_n - f| \to 0$ μ-almost everywhere. It follows that $G_N \searrow 0$ μ-almost everywhere. Also, the functions G_N are uniformly bounded above by an integrable function since

$$0 \leq G_N \leq G_1 = \sup_{n\in\mathbb{N}} g_n \leq \sup_{n\in\mathbb{N}}\big(|f_n| + |f|\big) \leq f^* + f \in L^1(\mu),$$

where $f^*, f \in L^1(\mu)$ according to Corollaries 9.5.2–9.5.3. So $G_N \in L^1(\mu)$ for all $N \in \mathbb{N}$ and Lebesgue's dominated convergence theorem affirms that

$$\lim_{N\to\infty} \int_X E(G_N|\mathcal{I}_\mu)\, d\mu = \lim_{N\to\infty} \int_X G_N \, d\mu = \int_X \lim_{N\to\infty} G_N \, d\mu = 0.$$

Moreover, according to Proposition A.1.60, since $(G_N)_{N=1}^{\infty}$ is decreasing and bounded below by 0, so is the sequence of conditional expectations $(E(G_N|\mathcal{I}_\mu))_{N=1}^{\infty}$ μ-almost everywhere. Summarizing, we have established that $E(G_N|\mathcal{I}_\mu) \searrow$ μ-a. e., $E(G_N|\mathcal{I}_\mu) \geq 0$ μ-a. e. and $\int_X E(G_N|\mathcal{I}_\mu)\, d\mu \searrow 0$ as $N \to \infty$. It ensues that $E(G_N|\mathcal{I}_\mu) \searrow 0$ μ-a. e. as $N \to \infty$.

Fix temporarily $N \in \mathbb{N}$. Then for any $n > N$, we have

$$\begin{aligned}
\frac{1}{n}\sum_{j=0}^{n-1} g_{n-j} \circ T^j &= \frac{1}{n}\sum_{j=0}^{n-N} g_{n-j} \circ T^j + \frac{1}{n}\sum_{j=n-N+1}^{n-1} g_{n-j} \circ T^j \\
&\leq \frac{n-N}{n} \cdot \frac{1}{n-N}\sum_{j=0}^{n-N} G_N \circ T^j + \frac{1}{n}\sum_{j=n-N+1}^{n-1} G_1 \circ T^j.
\end{aligned}$$

Let $F_N = \sum_{j=0}^{N-2} G_1 \circ T^j$. Using Birkhoff's ergodic theorem (Theorem 8.2.11), we deduce that

$$\begin{aligned}
0 \leq \limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} g_{n-j} \circ T^j &\leq \lim_{n\to\infty} \frac{1}{n-N}\sum_{j=0}^{n-N} G_N \circ T^j + \limsup_{n\to\infty} \frac{1}{n} F_N \circ T^{n-N+1} \\
&= E(G_N|\mathcal{I}_\mu) + \limsup_{n\to\infty} \frac{1}{n} F_N \circ T^{n-N+1} \quad \mu\text{-a. e.} \\
&= E(G_N|\mathcal{I}_\mu) \quad \mu\text{-a. e.}.
\end{aligned}$$

Since N was chosen arbitrarily and we have showed earlier that $E(G_N|\mathcal{I}_\mu) \to 0$ μ-a. e. as $N \to \infty$, we conclude that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} g_{n-j} \circ T^j = 0 \quad \mu\text{-a. e.}.$$

This establishes the μ-a. e. convergence of the first term on the right-hand side of (9.8). The μ-a. e. convergence of the second term on that right-hand side follows from Birkhoff's ergodic theorem (Theorem 8.2.11). Therefore, the sequence $(\frac{1}{n}I_\mu(\alpha^n))_{n=1}^\infty$ converges to $E(f|\mathcal{I}_\mu)$ μ-almost everywhere. □

Corollary 9.5.5 (Ergodic case of Shannon–McMillan–Breiman theorem). *Let $T : X \to X$ be an ergodic measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$ and let $\alpha \in \mathrm{Part}(X, \mathcal{A})$ be such that $\mathrm{H}_\mu(\alpha) < \infty$. Then*

$$\mathrm{h}_\mu(T, \alpha) = \lim_{n\to\infty} \frac{1}{n} I_\mu(\alpha^n)(x) \quad \textit{for } \mu\textit{-a. e. } x \in X.$$

Proof. This follows immediately from Shannon–McMillan–Breiman theorem (Theorem 9.5.4) and the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14). □

When $\mathrm{h}_\mu(T, \alpha) > 0$, this corollary reveals that $\mu(\alpha^n(x))$ tends to 0 with exponential rate $e^{-\mathrm{h}_\mu(T,\alpha)}$ for μ-a. e. $x \in X$ (see Exercise 9.7.11).

The right-hand side in the above equality can be viewed as a local entropy at x. The corollary then states that at almost every x the local entropy exists and is equal to the entropy of the transformation relative to the partition. Another approach to local entropy is discussed next.

9.6 Brin–Katok local entropy formula

We now derive the celebrated Brin–Katok local entropy formula.

In preparation for this, we show that given any Borel probability measure μ there exist finite Borel partitions of arbitrarily small diameters whose atoms have negligible boundaries with respect to μ.

Lemma 9.6.1. *Let (X, d) be a compact metric space and $\mu \in M(X)$. For every $\varepsilon > 0$, there exists a finite Borel partition α of X such that $\mathrm{diam}(\alpha) < \varepsilon$ and $\mu(\partial A) = 0$ for all $A \in \alpha$.*

Proof. Let $\varepsilon > 0$ and let $\{x_1, \ldots, x_n\}$ be an $(\varepsilon/4)$-spanning set of X. For each $1 \le i \le n$, the sets $\{x \in X : d(x, x_i) = r\}$, where $\varepsilon/4 < r < \varepsilon/2$, are mutually disjoint, and thus only countably many of them may have positive μ-measure. Hence, there exists $\varepsilon/4 < t < \varepsilon/2$ such that

$$\mu(\{x \in X : d(x, x_i) = t\}) = 0, \quad \forall 1 \le i \le n. \tag{9.9}$$

Define the sets A_i, $1 \le i \le n$, inductively by

$$A_i := \{x \in X : d(x, x_i) \le t\} \backslash (\cup_{j=1}^{i-1} A_j).$$

Since $t < \varepsilon/2$, the family $\alpha := \{A_1, \dots, A_n\}$ is a Borel partition of X with diameter smaller than ε. Noting that $\partial(A \backslash B) \subseteq \partial A \cup \partial B$ and $\partial(A \cup B) \subseteq \partial A \cup \partial B$, it follows from (9.9) that $\mu(\partial A_i) = 0$ for all $1 \le i \le n$. □

We now recall the concept, frequently used in coding theory, of *Hamming metric*. Let E be a nonempty finite set and $n \in \mathbb{N}$. The Hamming metric $\rho_{E,n}^{(H)}$ on E^n is defined by

$$\rho_{E,n}^{(H)}(\omega, \tau) = \frac{1}{n} \sum_{k=1}^{n} (1 - \delta_{\omega_k \tau_k}),$$

where δ_{ab} is the Kronecker delta symbol, that is,

$$\delta_{ab} = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b. \end{cases}$$

Equivalently,

$$\rho_{E,n}^{(H)}(\omega, \tau) = \frac{1}{n} \#\{1 \le k \le n : \omega_k \neq \tau_k\}. \tag{9.10}$$

It is well known and a straightforward exercise to check that $\rho_{E,n}^{(H)}$ is a metric on E^n. Given $\omega \in E^n$ and $r \ge 0$, we naturally denote by $B_{E,n}^{(H)}(\omega, r)$ the open ball, in the Hamming metric $\rho_{E,n}^{(H)}$, centered at ω and of radius r. Formally,

$$B_{E,n}^{(H)}(\omega, r) = \{\tau \in E^n : \rho_{E,n}^{(H)}(\omega, \tau) < r\}.$$

Standard combinatorial considerations show that the number of elements in the ball $B_{E,n}^{(H)}(\omega, r)$ depends only on $\#E$, n, and r, and is equal to

$$\#B_{E,n}^{(H)}(\omega, r) = \sum_{k=0}^{[rn]} (\#E - 1)^k \binom{n}{k}.$$

As Katok writes in [31], using this and Stirling's formula, it is easy to verify that for every $r \in (0, \frac{\#E-1}{\#E})$ we have

$$\lim_{n\to\infty} \frac{1}{n} \log \#B_{E,n}^{(H)}(\omega, r) = r \log(\#E - 1) - r \log r - (1 - r) \log(1 - r) =: g(r). \tag{9.11}$$

Observe that

$$\lim_{r\to 0} g(r) = 0, \tag{9.12}$$

and thus for every $r \in (0, \frac{\#E-1}{\#E})$ there is $N(r) \in \mathbb{N}$ such that

$$\#B^{(H)}_{E,n}(\omega, r) \le e^{g(r)n}, \quad \forall n \ge N(r). \tag{9.13}$$

Returning to dynamics, let (X, d) be a metric space, let $T : X \to X$ be a Borel measurable self-transformation and let μ be a Borel probability measure on X. Let also α be a Borel partition of X.

As the symbol α^n might be interpreted in two different ways in the proof of the forthcoming Theorem 9.6.2, we introduce further notation. As before, the nth refined partition of α with respect to the map T will be denoted by $\alpha^n := \bigvee_{i=0}^{n-1} T^{-i}\alpha$. The n-folded Cartesian product $\alpha \times \alpha \ldots \times \alpha$ will be denoted by $\widehat{\alpha}^n$.

In the proof of Theorem 9.6.2, we will work with the Hamming metrics $\rho^{(H)}_{E,n}$ on the sets $\widehat{\alpha}^n$, $n \in \mathbb{N}$. We introduce two mappings.

First, we define the map

$$\alpha^n \ni A \longmapsto \widehat{A} \in \widehat{\alpha}^n$$

as follows. Given that α is a partition, every $A \in \alpha^n$ is uniquely represented as

$$A = \bigcap_{i=0}^{n-1} T^{-i}(A_i),$$

where $A_i \in \alpha$ for all $0 \le i < n$. We naturally set

$$\widehat{A} := (A_0, A_1, \ldots, A_{n-1}) \in \widehat{\alpha}^n$$

and we note that the map $\alpha^n \ni A \longmapsto \widehat{A} \in \widehat{\alpha}^n$ is one-to-one.

Second, we define the map

$$\widehat{\alpha}^n \ni A = (A_1, A_2, \ldots, A_n) \longmapsto \check{A} \in \alpha^n \cup \{\emptyset\}$$

by the formula

$$\check{A} := \bigcap_{i=0}^{n-1} T^{-i}(A_{i+1})$$

and we note that the map $\widehat{\alpha}^n \ni A \longmapsto \check{A} \in \alpha^n \cup \{\emptyset\}$ is one-to-one on ${}^{\vee-1}(\alpha^n) = \{A \in \widehat{\alpha}^n : \check{A} \ne \emptyset\}$ and by restricting the first mapping to that set and the second mapping to the image of that set, the two restricted mappings are inverse of one another.

Introducing more notation, we denote

$$\widehat{G} := \{\widehat{g} \mid g \in G\} \quad \text{and} \quad \check{H} := \{\check{h} \mid h \in H\}$$

for all $G \subseteq \alpha^n$ and $H \subseteq \widehat{\alpha}^n$. We abbreviate

$$\check{B}^{(H)}_{\alpha,n}(A, r) := \widecheck{B^{(H)}_{\alpha,n}(A, r)} \quad \text{and} \quad \widehat{\alpha}^n(x) := \widehat{\alpha^n(x)}$$

for every $A \in \widehat{\alpha}^n$, $r > 0$ and $x \in X$. Finally, we denote

$$\cup\beta := \bigcup_{B\in\beta} B$$

for every $\beta \subseteq \alpha^n$.

We now present and prove the ergodic version of Brin and Katok's formula from [12]. The general (i. e. nonergodic) case is considerably more complicated and rarely needed in applications. However, unlike [12], we do not assume that the map $T : X \to X$ is continuous but merely that it is Borel measurable. The proof we provide is motivated by Pesin's relevant considerations in [56].

Theorem 9.6.2 (Brin–Katok local entropy formula). *Let (X, d) be a compact metric space and let $T : X \to X$ be a Borel measurable map. If μ is an ergodic T-invariant Borel probability measure on X, then for μ-almost every $x \in X$ we have*

$$\mathrm{h}_\mu(T) = \lim_{\delta\to 0} \liminf_{n\to\infty} \frac{-\log \mu(B_n(x,\delta))}{n} = \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{-\log \mu(B_n(x,\delta))}{n},$$

where $B_n(x, \delta)$ is the dynamical (n, δ)-ball at x (see Section 7.3).

Proof. It suffices to prove that for μ-almost every $x \in X$,

$$\mathrm{h}_\mu(T) \le \lim_{\delta\to 0} \liminf_{n\to\infty} \frac{-\log \mu(B_n(x,\delta))}{n} \le \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{-\log \mu(B_n(x,\delta))}{n} \le \mathrm{h}_\mu(T). \tag{9.14}$$

The middle inequality being obvious, we start with the right inequality, as it is simpler to establish.

Temporarily fix $\delta > 0$. Since X is a compact metric space, there is a finite Borel partition α_δ of X such that $\operatorname{diam}(\alpha_\delta) < \delta$. Then $\alpha_\delta^n(x) \subseteq B_n(x, \delta)$ for all $x \in X$ and all $n \in \mathbb{N}$. By the ergodic case of Shannon–McMillan–Breiman theorem (Corollary 9.5.5), we know that there exists a Borel set $X_1(\alpha_\delta)$ such that

$$\mu(X_1(\alpha_\delta)) = 1 \tag{9.15}$$

and

$$\lim_{n\to\infty} \frac{-\log \mu(\alpha_\delta^n(x))}{n} = \mathrm{h}_\mu(T, \alpha_\delta), \quad \forall x \in X_1(\alpha_\delta).$$

As $\alpha^n(x) \subseteq B_n(x, \delta)$, we deduce that

$$\limsup_{n\to\infty} \frac{-\log \mu(B_n(x,\delta))}{n} \le \mathrm{h}_\mu(T, \alpha_\delta) \le \mathrm{h}_\mu(T), \quad \forall x \in X_1(\alpha_\delta). \tag{9.16}$$

It follows from this and (9.15) that the set

$$X_1 := \bigcap_{k=1}^{\infty} X_1(\alpha_{1/k})$$

satisfies

$$\mu(X_1) = 1 \tag{9.17}$$

and

$$\limsup_{n\to\infty} \frac{-\log \mu(B_n(x, 1/k))}{n} \le \mathrm{h}_\mu(T), \quad \forall x \in X_1,\ \forall k \in \mathbb{N}. \tag{9.18}$$

Observing that the left-hand sides of (9.16) and (9.18) are a decreasing function of δ and an increasing function of k, respectively, we conclude that

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \frac{-\log \mu(B_n(x,\delta))}{n} = \lim_{k\to\infty} \limsup_{n\to\infty} \frac{-\log \mu(B_n(x, 1/k))}{n} \le \mathrm{h}_\mu(T),\ \forall x \in X_1.$$

This is the right inequality in (9.14).

To achieve the left inequality in (9.14), temporarily fix $\varepsilon \in \left(0, \frac{\#\alpha-1}{\#\alpha}\right)$ with α a finite Borel partition of X such that

$$\mu(\partial\alpha) = 0, \tag{9.19}$$

where $\partial\alpha$ denotes the boundary of the partition α. For any $\eta > 0$, set

$$U_\eta(\alpha) := \{x \in X \,:\, B(x,\eta) \not\subseteq \alpha(x)\}.$$

Since $\bigcap_{\eta>0} U_\eta(\alpha) = \partial\alpha$ and $U_{\eta_1}(\alpha) \subseteq U_{\eta_2}(\alpha)$ whenever $\eta_1 \le \eta_2$, it follows from (9.19) that

$$\lim_{\eta\to 0} \mu(U_\eta(\alpha)) = 0.$$

Consequently, there exists $\eta_\varepsilon > 0$ such that $\mu(U_\eta(\alpha)) < \varepsilon$ for every $0 < \eta \le \eta_\varepsilon$. By the ergodic case of Birkhoff's ergodic theorem for an indicator function (Corollary 8.2.15) and by Egorov's theorem (Theorem A.1.44), for every $\eta \in (0, \eta_\varepsilon]$ there exist a Borel set $X(\varepsilon,\eta) \subseteq X$ and an integer $M(\varepsilon,\eta) \in \mathbb{N}$ such that

$$\mu(X(\varepsilon,\eta)) \ge 1 - \varepsilon \tag{9.20}$$

and

$$\frac{1}{n}\#\{0 \le i < n : T^i(x) \in U_\eta(\alpha)\} < \varepsilon, \quad \forall x \in X(\varepsilon,\eta),\ \forall n \ge M(\varepsilon,\eta).$$

Now, observe that if $y \in B_n(x,\eta)$ then for each $0 \le i < n$,

$$\text{either} \quad \alpha(T^i(x)) = \alpha(T^i(y)) \quad \text{or} \quad T^i(x) \in U_\eta(\alpha).$$

So, if $x \in X(\varepsilon,\eta)$ and $y \in B_n(x,\eta)$ for some $n \ge M(\varepsilon,\eta)$, then

$$\rho^{(H)}_{\alpha,n}(\widehat{\alpha}^n(x), \widehat{\alpha}^n(y)) < \varepsilon.$$

(See (9.10) for the definition of $\rho^{(H)}_{\alpha,n}$.) Equivalently,

$$B_n(x,\eta) \subseteq \breve{B}^{(H)}_{\alpha,n}(\widehat{\alpha}^n(x),\varepsilon), \quad \forall x \in X(\varepsilon,\eta),\ \forall n \geq M(\varepsilon,\eta). \tag{9.21}$$

We thus need an upper estimate on $\mu(\breve{B}^{(H)}_{\alpha,n}(\widehat{\alpha}^n(x),\varepsilon))$. For every $n \in \mathbb{N}$ define

$$Z_n := \{A \in \alpha^n : \mu(A) \geq \exp((-\mathrm{h}_\mu(T,\alpha) + 3g(\varepsilon))n)\},$$

with $g(\cdot)$ from (9.11). As sets in Z_n are mutually disjoint and $\mu(X) = 1$, we deduce that

$$\#Z_n \leq \exp((\mathrm{h}_\mu(T,\alpha) - 3g(\varepsilon))n). \tag{9.22}$$

To get an appropriate upper estimate, there are "good" and "bad" atoms in α^n. Let

$$\mathrm{Bad}(\alpha^n,\varepsilon) := \{A \in \alpha^n : B^{(H)}_{\alpha,n}(\widehat{A},\varepsilon) \cap \widehat{Z_n} \neq \emptyset\} \subseteq \breve{B}^{(H)}_{\alpha,n}(\widehat{Z_n},\varepsilon). \tag{9.23}$$

Using (9.13), if $n \geq N(\varepsilon)$ and $A \in \alpha^n \setminus \mathrm{Bad}(\alpha^n,\varepsilon)$ then we obtain

$$\begin{aligned} \mu(\breve{B}^{(H)}_{\alpha,n}(\widehat{A},\varepsilon)) &\leq \#B^{(H)}_{\alpha,n}(\widehat{A},\varepsilon)\exp((-\mathrm{h}_\mu(T,\alpha) + 3g(\varepsilon))n) \\ &\leq \exp((-\mathrm{h}_\mu(T,\alpha) + 4g(\varepsilon))n). \end{aligned} \tag{9.24}$$

Along with (9.21), this implies that

$$\mu(B_n(x,\eta)) \leq \exp((-\mathrm{h}_\mu(T,\alpha) + 4g(\varepsilon))n) \tag{9.25}$$

if $x \in X(\varepsilon,\eta) \setminus \cup\mathrm{Bad}(\alpha^n,\varepsilon)$ for some $n \geq \max\{N(\varepsilon), M(\varepsilon,\eta)\}$. Hence,

$$\lim_{\delta\to 0}\liminf_{n\to\infty} \frac{-\log\mu(B_n(x,\delta))}{n} \geq \liminf_{n\to\infty}\frac{-\log\mu(B_n(x,\eta))}{n} \geq \mathrm{h}_\mu(T,\alpha) - 4g(\varepsilon) \tag{9.26}$$

for all $x \in \bigcup_{i=1}^\infty \bigcap_{j=i}^\infty (X(\varepsilon,\eta) \setminus \cup\mathrm{Bad}(\alpha^j,\varepsilon)) = X(\varepsilon,\eta) \cap \bigcup_{i=1}^\infty \bigcap_{j=i}^\infty (X \setminus \cup\mathrm{Bad}(\alpha^j,\varepsilon))$.

We need to show that this latter set is big measurewise, i. e. that its measure is ε-close to 1. For this, it suffices to show that one of its subsets is big. This subset has the similar form $X(\varepsilon,\eta) \cap \bigcup_{i=1}^\infty \bigcap_{j=i}^\infty (\cup\beta_j(\varepsilon) \setminus \cup\mathrm{Bad}(\alpha^j,\varepsilon))$. We define $\beta_j(\varepsilon)$ now and then estimate the measure of the said subset in several steps. This is the most arduous task.

By the ergodic case of the Shannon–McMillan–Breiman theorem (Corollary 9.5.5) and Egorov's theorem (Theorem A.1.44), there exists a Borel set $Y(\varepsilon) \subseteq X$ and an integer $N_1(\varepsilon) \geq N(\varepsilon)$ such that

$$\mu(Y(\varepsilon)) > 1 - \varepsilon \tag{9.27}$$

and

$$\frac{-\log\mu(\alpha^n(x))}{n} \geq \mathrm{h}_\mu(T,\alpha) - g(\varepsilon), \quad \forall x \in Y(\varepsilon),\ \forall n \geq N_1(\varepsilon).$$

Equivalently,

$$\mu(\alpha^n(x)) \le \exp(-(\mathrm{h}_\mu(T,\alpha) - g(\varepsilon))n), \quad \forall x \in Y(\varepsilon),\ \forall n \ge N_1(\varepsilon).$$

Let

$$\beta_n(\varepsilon) := \{\alpha^n(x) : x \in Y(\varepsilon)\}.$$

Fix temporarily $n \ge N_1(\varepsilon)$. Then

$$\mu(A) \le \exp(-(\mathrm{h}_\mu(T,\alpha) - g(\varepsilon))n), \quad \forall A \in \beta_n(\varepsilon). \tag{9.28}$$

Let

$$D_n(\varepsilon) := \{A \in \beta_n(\varepsilon) : \widehat{A} \in B^{(H)}_{\alpha,n}(\widehat{Z_n},\varepsilon)\} = \beta_n(\varepsilon) \cap \breve{B}^{(H)}_{\alpha,n}(\widehat{Z_n},\varepsilon).$$

Using (9.28) as well as (9.13) and (9.22), we get

$$\begin{aligned}\mu(\cup D_n(\varepsilon)) &\le \exp(-(\mathrm{h}_\mu(T,\alpha) - g(\varepsilon))n)\#(\beta_n(\varepsilon) \cap \breve{B}^{(H)}_{\alpha,n}(\widehat{Z_n},\varepsilon))\\ &\le \exp(-(\mathrm{h}_\mu(T,\alpha) - g(\varepsilon))n)\#B^{(H)}_{\alpha,n}(\widehat{Z_n},\varepsilon)\\ &\le \exp(-(\mathrm{h}_\mu(T,\alpha) - g(\varepsilon))n)\#Z_n \exp(g(\varepsilon)n)\\ &\le \exp(-g(\varepsilon)n).\end{aligned} \tag{9.29}$$

Using (9.23), we obtain that

$$\begin{aligned}\cup\beta_n(\varepsilon) \setminus \cup\mathrm{Bad}(\alpha^n,\varepsilon) &= \cup\beta_n(\varepsilon) \setminus (\cup\beta_n(\varepsilon) \cap \cup\mathrm{Bad}(\alpha^n,\varepsilon))\\ &\supseteq \cup\beta_n(\varepsilon) \setminus (\cup\beta_n(\varepsilon) \cap \cup\breve{B}^{(H)}_{\alpha,n}(\widehat{Z_n},\varepsilon))\\ &= \cup\beta_n(\varepsilon) \setminus \cup(\beta_n(\varepsilon) \cap \breve{B}^{(H)}_{\alpha,n}(\widehat{Z_n},\varepsilon))\\ &= \cup\beta_n(\varepsilon) \setminus \cup D_n(\varepsilon).\end{aligned}$$

Therefore, for every $k \ge N_1(\varepsilon)$ we obtain that

$$\begin{aligned}\bigcap_{n=k}^{\infty}(\cup\beta_n(\varepsilon) \setminus \cup\mathrm{Bad}(\alpha^n,\varepsilon)) &\supseteq \bigcap_{n=k}^{\infty}(\cup\beta_n(\varepsilon) \setminus \cup D_n(\varepsilon)) \supseteq \bigcap_{n=k}^{\infty}\cup\beta_n(\varepsilon) \Big\backslash \bigcup_{n=k}^{\infty}\cup D_n(\varepsilon)\\ &\supseteq \bigcap_{n=k}^{\infty} Y(\varepsilon) \Big\backslash \bigcup_{n=k}^{\infty}\cup D_n(\varepsilon) = Y(\varepsilon) \Big\backslash \bigcup_{n=k}^{\infty}\cup D_n(\varepsilon).\end{aligned}$$

From this, (9.27) and (9.29), we deduce that

$$\begin{aligned}\mu\left(\bigcap_{n=k}^{\infty}(\cup\beta_n(\varepsilon) \setminus \cup\mathrm{Bad}(\alpha^n,\varepsilon))\right) &\ge \mu(Y(\varepsilon)) - \mu\left(\bigcup_{n=k}^{\infty}\cup D_n(\varepsilon)\right)\\ &\ge \mu(Y(\varepsilon)) - \sum_{n=k}^{\infty}\mu(\cup D_n(\varepsilon))\end{aligned}$$

$$\begin{aligned} &\geq 1-\varepsilon-\sum_{n=k}^{\infty} e^{-g(\varepsilon)n} \\ &= 1-\varepsilon-\frac{e^{-g(\varepsilon)k}}{1-e^{-g(\varepsilon)}}. \end{aligned} \tag{9.30}$$

Consequently,

$$\mu\left(\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}(\cup\beta_n(\varepsilon)\setminus\cup\mathrm{Bad}(\alpha^n,\varepsilon))\right)=\lim_{k\to\infty}\mu\left(\bigcap_{n=k}^{\infty}(\cup\beta_n(\varepsilon)\setminus\cup\mathrm{Bad}(\alpha^n,\varepsilon))\right)\geq 1-\varepsilon. \tag{9.31}$$

It follows from (9.31) and (9.20) that

$$\mu\left(X(\varepsilon,\eta)\cap\left(\bigcup_{i=1}^{\infty}\bigcap_{j=i}^{\infty}(\cup\beta_j(\varepsilon)\setminus\cup\mathrm{Bad}(\alpha^j,\varepsilon))\right)\right)\geq 1-2\varepsilon.$$

Thus

$$\mu\left(\bigcup_{q=k}^{\infty}\left(X(1/q,\eta_{1/q})\cap\left(\bigcup_{i=1}^{\infty}\bigcap_{j=i}^{\infty}(\cup\beta_j(1/q)\setminus\cup\mathrm{Bad}(\alpha^j,1/q))\right)\right)\right)=1,\quad\forall k\in\mathbb{N}.$$

If

$$\widehat{X}_2(\alpha):=\bigcap_{k=1}^{\infty}\bigcup_{q=k}^{\infty}\left(X(1/q,\eta_{1/q})\cap\left(\bigcup_{i=1}^{\infty}\bigcap_{j=i}^{\infty}(\cup\beta_j(1/q)\setminus\cup\mathrm{Bad}(\alpha^j,1/q))\right)\right),$$

then

$$\mu(\widehat{X}_2(\alpha))=1 \tag{9.32}$$

and, by virtue of (9.26) and (9.12), we deduce that

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\frac{-\log\mu(B_n(x,\delta))}{n}\geq \mathrm{h}_\mu(T,\alpha),\quad\forall x\in\widehat{X}_2(\alpha). \tag{9.33}$$

As the metric space X is compact, it follows from Lemma 9.6.1 and Theorem 9.4.17 that there exists a sequence $(\alpha_k)_{k=1}^{\infty}$ of finite Borel partitions of X such that $\mu(\partial\alpha_k)=0$ for every $k\in\mathbb{N}$ and

$$\lim_{k\to\infty}\mathrm{h}_\mu(T,\alpha_k)=\mathrm{h}_\mu(T).$$

Setting

$$\widehat{X}_2:=\bigcap_{k=1}^{\infty}\widehat{X}_2(\alpha_k),$$

we have by (9.32) that

$$\mu(\widehat{X}_2) = 1 \tag{9.34}$$

and by (9.33) that

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\frac{-\log\mu(B_n(x,\delta))}{n} \ge \mathrm{h}_\mu(T), \quad \forall x \in \widehat{X}_2.$$

This is the left inequality in (9.14).

As $\mu(X_1 \cap \widehat{X}_2) = 1$ by (9.17) and (9.34), and as all three inequalities in (9.14) are valid on $X_1 \cap \widehat{X}_2$, the result ensues. □

Remark 9.6.3. As a matter of fact, only total boundedness of the metric d is needed for Theorem 9.6.2 to hold. More precisely, in Lemma 9.6.1 total boundedness is sufficient, and compactness has not been used anywhere in the proof of Theorem 9.6.2.

In the case of an expansive system T, we have the following stronger and simpler version of Theorem 9.6.2.

Theorem 9.6.4 (Brin–Katok local entropy formula for expansive maps). *Let $T : X \to X$ be an expansive topological dynamical system and let d be a metric compatible with the topology on X. If $\delta > 0$ is an expansive constant for T corresponding to this metric, then for every $\zeta \in (0,\delta]$, every ergodic T-invariant Borel probability measure μ on X and μ-almost every $x \in X$, we have*

$$\mathrm{h}_\mu(T) = \lim_{n\to\infty}\frac{-\log\mu(B_n(x,\zeta))}{n}.$$

Proof. For every $x \in X$, denote

$$\overline{h}_\mu(T,\zeta,x) := \limsup_{n\to\infty}\frac{-\log\mu(B_n(x,\zeta))}{n} \quad \text{and} \quad \underline{h}_\mu(T,\zeta,x) := \liminf_{n\to\infty}\frac{-\log\mu(B_n(x,\zeta))}{n}.$$

Since $B_n(x,\zeta) \subseteq B_n(x,\delta)$ for every $n \in \mathbb{N}$, it is clear that

$$\overline{h}_\mu(T,\delta,x) \le \overline{h}_\mu(T,\zeta,x). \tag{9.35}$$

On the other hand, using Observation 5.2.4 we obtain that

$$\begin{aligned}
\underline{h}_\mu(T,\delta,x) &= \liminf_{n\to\infty}\frac{-\log\mu(B_n(x,\delta))}{n}\\
&\ge \liminf_{n\to\infty}\frac{-\log\mu(B_{n-N(\zeta/2)}(x,\zeta))}{n}\\
&= \liminf_{n\to\infty}\frac{n-N(\zeta/2)}{n}\cdot\frac{-\log\mu(B_{n-N(\zeta/2)}(x,\zeta))}{n-N(\zeta/2)}
\end{aligned}$$

$$= \liminf_{n\to\infty} \frac{-\log \mu(B_{n-N(\zeta/2)}(x,\zeta))}{n - N(\zeta/2)}$$
$$= \underline{h}_\mu(T,\zeta,x). \tag{9.36}$$

Since, by Theorem 9.6.2,

$$\lim_{\zeta\to 0} \underline{h}_\mu(T,\zeta,x) = \mathrm{h}_\mu(T) = \lim_{\zeta\to 0} \overline{h}_\mu(T,\zeta,x) \quad \text{for } \mu\text{-a. e. } x \in X,$$

we infer from (9.36) and (9.35) that $\underline{h}_\mu(T,\delta,x) \geq \mathrm{h}_\mu(T)$ and $\overline{h}_\mu(T,\delta,x) \leq \mathrm{h}_\mu(T)$ for μ-almost every $x \in X$. As it is always true that $\underline{h}_\mu(T,\delta,x) \leq \overline{h}_\mu(T,\delta,x)$, we deduce that $\underline{h}_\mu(T,\delta,x) = \mathrm{h}_\mu(T) = \overline{h}_\mu(T,\delta,x)$ and the result holds when $\zeta = \delta$. Since any $\zeta \in (0,\delta]$ is also an expansive constant for T, the theorem is validated. □

9.7 Exercises

Exercise 9.7.1. The objective of this exercise is to prove Theorem 9.1.1. Using axioms (A1)–(A4), proceed as follows.

(a) Given $n \in \mathbb{N}$, prove by induction that $f(n^k) = kf(n)$ for all $k \geq 0$. (Think about the meaning of this relationship.)

(b) Given $n \geq 2$, for every $r \in \mathbb{N}$ there exists a unique $k \geq 0$ such that $n^k \leq 2^r < n^{k+1}$. Show that

$$kf(n) \leq rf(2) < (k+1)f(n).$$

(c) Prove that (b) holds with f replaced by log.

(d) Given $n \geq 2$, deduce that

$$\left|\frac{f(2)}{f(n)} - \frac{\log 2}{\log n}\right| < \frac{1}{r}, \quad \forall r \in \mathbb{N}.$$

(e) Conclude that $f(n) = (f(2)/\log 2)\log n$ for all $n \in \mathbb{N}$.

(f) Let $C = f(2)/\log 2$. Observe that in order to establish that

$$\mathrm{H}(p, 1-p) = -Cp\log p - C(1-p)\log(1-p), \tag{9.37}$$

it suffices to prove that this relation holds for all rational $p \in (0,1)$.

(g) Accordingly, let $p = r/s \in \mathbb{Q} \cap (0,1)$. By partitioning some experiment appropriately, show that

$$f(s) = \mathrm{H}\left(\frac{r}{s}, \frac{s-r}{s}\right) + \frac{r}{s}f(r) + \frac{s-r}{s}f(s-r).$$

(h) Deduce (9.37) and observe that the function H extends continuously to $[0,1]$.

(i) Hence, the formula in Theorem 9.1.1 holds when $n = 2$. Prove by induction that it holds for any $n \in \mathbb{N}$.

Exercise 9.7.2. Let $(X, \mathcal{A})$ be a measurable space. Show that $\le$ is a partial order relation on the set $\mathrm{Part}(X, \mathcal{A})$.

Exercise 9.7.3. Let $(X, \mathcal{A})$ be a measurable space and $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. Prove that $\alpha \le \beta$ if and only if $\beta(x) \subseteq \alpha(x)$ for all $x \in X$.

Exercise 9.7.4. Show that $\alpha \le \beta$ if and only if $A = \bigcup\{B \in \beta : B \cap A \neq \emptyset\}$ for all $A \in \alpha$.

Exercise 9.7.5. Prove that $\alpha \le \beta$ if and only if $A = \bigcup\{B \in \beta : B \subseteq A\}$ for all $A \in \alpha$.

Exercise 9.7.6. Find a probability space $(X, \mathcal{A}, \mu)$ and $\alpha, \beta, \gamma \in \mathrm{Part}(X, \mathcal{A})$ such that $\beta \le \gamma$ but $I_\mu(\alpha|\beta)(x) \ngeq I_\mu(\alpha|\gamma)(x)$ for some $x \in X$.

Exercise 9.7.7. Find a probability space $(X, \mathcal{A}, \mu)$ and $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$ such that $I_\mu(\alpha \vee \beta)(x) \nleq I_\mu(\alpha)(x) + I_\mu(\beta)(x)$ for some $x \in X$.

Exercise 9.7.8. Find a probability space $(X, \mathcal{A}, \mu)$ and $\alpha, \beta, \gamma \in \mathrm{Part}(X, \mathcal{A})$ such that $I_\mu(\alpha|\gamma)(x) \nleq I_\mu(\alpha|\beta)(x) + I_\mu(\beta|\gamma)(x)$ for some $x \in X$.

Exercise 9.7.9. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. Given $\alpha \in \mathrm{Part}(X, \mathcal{A})$, show that the sequence $(\mathrm{H}_\mu(\alpha^n))_{n=1}^\infty$ is subadditive. Then deduce from Lemma 3.2.17 that the limit $\lim_{n\to\infty} \frac{1}{n}\mathrm{H}_\mu(\alpha^n) = \mathrm{h}_\mu(T, \alpha)$ exists and is nonnegative.

Exercise 9.7.10. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. Show that the mapping $\mu \mapsto \mathrm{h}_\mu(T)$ is affine on the set $M(T, \mathcal{A})$ of all T-invariant probability measures on $(X, \mathcal{A})$. In other words, show that if $T : X \to X$ is a dynamical system preserving two probability measures μ and ν on the measurable space $(X, \mathcal{A})$, then

$$\mathrm{h}_{s\mu+(1-s)\nu}(T) = s\mathrm{h}_\mu(T) + (1-s)\mathrm{h}_\nu(T)$$

for all $0 \le s \le 1$.

Exercise 9.7.11. Let $T : X \to X$ be an ergodic measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha \in \mathrm{Part}(X, \mathcal{A})$ be such that $\mathrm{H}_\mu(\alpha) < \infty$. According to Corollary 9.5.5,

$$\mathrm{h}_\mu(T, \alpha) = \lim_{n\to\infty} \frac{1}{n} I_\mu(\alpha^n)(x) \quad \text{for } \mu\text{-a. e. } x \in X.$$

Let $0 < \varepsilon < 1$ and for each $n \in \mathbb{N}$ let $N_n(\varepsilon)$ be the minimum number of atoms of α^n needed to construct a set of measure at least $1 - \varepsilon$. Show that

$$\mathrm{h}_\mu(T, \alpha) = \lim_{n\to\infty} \frac{1}{n} \log N_n(\varepsilon).$$

Exercise 9.7.12. Prove that for the full E-shift equipped with the product measure μ as in Example 9.4.23, and the partition $\alpha := \{[e] : e \in E\}$,

$$\mathrm{H}_\mu(\alpha^n) = -n \sum_{e \in E} P(e) \log P(e), \quad \forall n \in \mathbb{N}.$$

Exercise 9.7.13. Let $T : [0,1] \to [0,1]$ be the tent map (see Example 1.1.3). Show that T preserves the Lebesgue measure on $[0,1]$ and that its entropy with respect to Lebesgue measure is equal to $\log 2$.

Exercise 9.7.14. Let $(X, \mathcal{A})$ be a measurable space and $(\alpha_n)_{n=1}^\infty$ be a sequence of increasingly finer countable measurable partitions of $(X, \mathcal{A})$ which generates the σ-algebra $\mathcal{A}$, that is, such that

$$\alpha_n \le \alpha_{n+1}, \quad \forall n \in \mathbb{N} \quad \text{and} \quad \sigma\left(\bigcup_{n=1}^\infty \mathcal{A}_n\right) = \mathcal{A}.$$

Suppose that μ and ν are probability measures on $(X, \mathcal{A})$ such that $\mu \ll \nu$. Let $\rho = d\mu/d\nu$ be the Radon–Nikodym derivative of μ with respect to ν (cf. Theorem A.1.50). Using Example A.1.62 and the martingale convergence theorem for conditional expectations (Theorem A.1.67), show that

$$\rho(x) = \lim_{n\to\infty} \frac{\mu(\alpha_n(x))}{\nu(\alpha_n(x))} \text{ for } \nu\text{-a. e. } x \in X.$$

Exercise 9.7.15. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces and $T : X \to X$ and $S : Y \to Y$ be measurable transformations. A measurable transformation $\pi : X \to Y$ is called a *factor map* between T and S if $\pi \circ T = S \circ \pi$.

If $(X, \mathcal{A}, \mu)$ is a measure space, then recall that π induces the push down measure $\mu \circ \pi^{-1}$ on the measurable space $(Y, \mathcal{B})$.

Let $\alpha, \beta \in \mathrm{Part}(Y, \mathcal{B})$. Prove the following statements:

(a) $\pi^{-1}(\alpha \vee \beta) = (\pi^{-1}\alpha) \vee (\pi^{-1}\beta)$.
(b) $\pi^{-1}(\alpha_m^n) = (\pi^{-1}\alpha)_m^n$ for all $m, n \ge 0$.
(c) π^{-1} preserves the partial order $\le$, that is, if $\alpha \le \beta$ then $\pi^{-1}\alpha \le \pi^{-1}\beta$.
(d) $(\pi^{-1}\alpha)(x) = \pi^{-1}(\alpha(\pi(x)))$ for all $x \in X$.
(e) $I_\mu(\pi^{-1}\alpha|\pi^{-1}\beta) = I_{\mu\circ\pi^{-1}}(\alpha|\beta) \circ \pi$.
(f) $I_\mu(\pi^{-1}\alpha) = I_{\mu\circ\pi^{-1}}(\alpha) \circ \pi$.
(g) $\mathrm{H}_\mu(\pi^{-1}\alpha|\pi^{-1}\beta) = \mathrm{H}_{\mu\circ\pi^{-1}}(\alpha|\beta)$.
(h) $\mathrm{H}_\mu(\pi^{-1}\alpha) = \mathrm{H}_{\mu\circ\pi^{-1}}(\alpha)$.
(i) $\mathrm{h}_\mu(T, \pi^{-1}\alpha) = \mathrm{h}_{\mu\circ\pi^{-1}}(S, \alpha)$.
(j) $\mathrm{h}_\mu(T) \ge \mathrm{h}_{\mu\circ\pi^{-1}}(S)$.
(k) If π is bimeasurable (i. e., measurable, bijective and its inverse is measurable), then $\mathrm{h}_\mu(T) = \mathrm{h}_{\mu\circ\pi^{-1}}(S)$.

10 Infinite invariant measures

In this chapter, we deal with measurable transformations preserving measures that are no longer assumed to be finite. The outlook is then substantially different than in the case of finite measures. As far as we know, only J. Aaronson's book [1] is entirely dedicated to infinite ergodic theory.

In Section 10.1, we introduce and investigate in detail the notions of quasi-invariant measures, ergodicity, and conservativity. We also prove Halmos' recurrence theorem, which is a generalization of Poincaré's recurrence theorem for quasi-invariant measures that are not necessarily finite.

In Section 10.2, we discuss first return times, first return maps, and induced systems. We further establish relations between invariant measures for the original transformation and the induced transformation.

In Section 10.3, we study implications of Birkhoff's ergodic theorem for finite and infinite measure spaces. Among others, we demonstrate Hopf's ergodic theorem, which applies to measure-preserving transformations of σ-finite spaces.

Finally, in Section 10.4, we seek a condition under which, given a quasi-invariant probability measure, one can construct a σ-finite invariant measure which is absolutely continuous with respect to the original measure. To this end, we introduce a class of transformations, called Martens maps, that have this feature and even more. In fact, these maps have the property that any quasi-invariant probability measure admits an equivalent σ-finite invariant one.

Applications of these concepts and results can be found in Chapters 13–14 of the second volume and Chapters 29–32 of the third volume.

10.1 Quasi-invariant measures, ergodicity and conservativity

By definition, quasi-invariant measures preserve sets of measure zero.

Definition 10.1.1. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. A measure μ on $(X, \mathcal{A})$ is called *quasi-T-invariant* if $\mu \circ T^{-1} \ll \mu$.

Obviously, invariant measures are quasi-invariant but the converse statement does not hold in general.

The concept of ergodicity defined in Chapter 8 for transformations of probability spaces readily generalizes to transformations of arbitrary measure spaces.

Definition 10.1.2. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. Then T is *ergodic* with respect to μ if

$$T^{-1}(A) = A \quad \Longrightarrow \quad \mu(A) = 0 \;\text{ or }\; \mu(X \backslash A) = 0.$$

Alternatively, μ is said to be ergodic with respect to T.

https://doi.org/10.1515/9783110702682-010

The next result states that, like a T-invariant measure, a quasi-T-invariant measure μ is ergodic if and only if every μ-a. e. T-invariant set is trivial in a measure-theoretic sense, that is, has measure zero or its complement is of measure zero. This is a generalization of Proposition 8.2.4.

Proposition 10.1.3. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. Then T is ergodic with respect to μ if and only if*

$$\mu\big(T^{-1}(A) \triangle A\big) = 0 \quad \Longrightarrow \quad \mu(A) = 0 \ \textit{or} \ \mu(X \backslash A) = 0.$$

Proof. The proof goes along similar lines to that of Proposition 8.2.4 and is left to the reader. □

In Section 1.4, we studied the concept of wandering points. We revisit this dynamical behavior from a measure-theoretic standpoint.

Definition 10.1.4. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. A set $W \in \mathcal{A}$ is a *wandering set for T* if its preimages $(T^{-n}(W))_{n=0}^{\infty}$ are mutually disjoint.

One way of constructing wandering sets is now described.

Lemma 10.1.5. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. For every $A \in \mathcal{A}$, the set $W_A := A \backslash \bigcup_{n=1}^{\infty} T^{-n}(A)$ is a wandering set for T.*

To lighten notation, define

$$A^{-} := \bigcup_{n=1}^{\infty} T^{-n}(A) \quad \text{and thus} \quad W_A := A \backslash A^{-}.$$

Proof. Suppose for a contradiction that W_A is not wandering for T, that is, $T^{-k}(W_A) \cap T^{-l}(W_A) \neq \emptyset$ for some $0 \le k < l$. This means that $T^{-k}\big(W_A \cap T^{-(l-k)}(W_A)\big) \neq \emptyset$, and thus $W_A \cap T^{-(l-k)}(W_A) \neq \emptyset$. Set $j = l - k \in \mathbb{N}$. Fix $x \in W_A \cap T^{-j}(W_A)$. On one hand, $x \in W_A$ implies that $x \notin A^{-}$. So $x \notin T^{-j}(A)$. On the other hand, $W_A \subseteq A$ and $x \in T^{-j}(W_A)$ imply that $x \in T^{-j}(A)$. This is a contradiction and W_A must therefore be a wandering set. □

Next, we introduce the notion of conservativity.

Definition 10.1.6. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. Then T is *conservative* with respect to μ if $\mu(W) = 0$ for every wandering set W for T.

Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. For every $B \in \mathcal{A}$, define the set $B_\infty \in \mathcal{A}$ to be

$$B_\infty := \{x \in X : T^n(x) \in B \text{ for infinitely many } n \in \mathbb{N}\} = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} T^{-n}(B).$$

Clearly, $T^{-1}(B_\infty) = B_\infty$ and so $T^{-1}(X\backslash B_\infty) = X\backslash B_\infty$. Notice also that if W is a wandering set for T, then $W \cap \bigcup_{n=1}^\infty T^{-n}(W) = \emptyset$. In particular, $W \cap W_\infty = \emptyset$.

Poincaré's recurrence theorem (Theorem 8.1.16) asserts that if μ is T-invariant and finite, then $\mu(B\backslash B_\infty) = 0$. We shall now prove its generalization in two respects: namely, by assuming only (1) that μ is quasi-T-invariant and (2) that μ may be infinite.

Theorem 10.1.7 (Halmos' recurrence theorem). *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. For every $A \in \mathcal{A}$, the following equivalence holds: $\mu(B\backslash B_\infty) = 0$ for all measurable sets $B \subseteq A$ if and only if $\mu(W) = 0$ for all wandering sets $W \subseteq A$.*

Proof. Fix $A \in \mathcal{A}$. If $\mu(A) = 0$, then $\mu(B\backslash B_\infty) = 0$ for all measurable sets $B \subseteq A$ and $\mu(W) = 0$ for all wandering sets $W \subseteq A$. Thus the equivalence is trivially satisfied when $\mu(A) = 0$, and we may assume in the sequel that $\mu(A) > 0$.

[$\Rightarrow$] We will prove the contrapositive statement. Suppose that $\mu(W) > 0$ for some wandering set $W \subseteq A$. Then $W \cap W_\infty = \emptyset$. Therefore, $\mu(W\backslash W_\infty) > 0$. Thus W is a measurable set $B \subseteq A$ such that $\mu(B\backslash B_\infty) > 0$.

[$\Leftarrow$] Let us now prove the converse implication. Assume that $\mu(W) = 0$ for all wandering sets $W \subseteq A$. Fix a measurable set $B \subseteq A$ and for all $n \geq 0$ let

$$B_n := B \cap T^{-n}(B) \Big\backslash \bigcup_{\ell=n+1}^{\infty} T^{-\ell}(B),$$

that is, B_n is the set of points in B that return to B at time n but never again thereafter. So

$$B\backslash B_\infty = \bigcup_{n\geq 0} B_n.$$

Suppose for a contradiction that $\mu(B\backslash B_\infty) > 0$. That is, suppose there is $n \geq 0$ such that $\mu(B_n) > 0$. Lemma 10.1.5 asserts that W_{B_n} is a wandering set for T. Since $B_n \subseteq A$, the hypothesis implies that $\mu(W_{B_n}) = 0$. This means that $\mu(B_n\backslash B_n^-) = 0$. Since $\mu(B_n) > 0$, there thus exists $x \in B_n \cap B_n^- = B_n \cap \bigcup_{k=1}^\infty T^{-k}(B_n)$. So $x \in B_n$. There is also $k \in \mathbb{N}$ such that $x \in T^{-k}(B_n)$. As $T^{-k}(B_n) \subseteq T^{-(n+k)}(B)$, it turns out that $x \in T^{-\ell}(B)$, where $\ell = n + k$. So $x \notin B_n$. This is a contradiction. Consequently, $\mu(B\backslash B_\infty) = 0$ for any measurable set $B \subseteq A$. $\square$

Taking $A = X$ in Theorem 10.1.7, we get the following special case.

Corollary 10.1.8. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. Then T is conservative if and only if $\mu(B\backslash B_\infty) = 0$ for all $B \in \mathcal{A}$.*

In particular, Poincaré's recurrence theorem (Theorem 8.1.16) confirms that every measure-preserving transformation of a finite measure space is conservative.

Corollary 10.1.9. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a finite T-invariant measure. Then T is conservative.*

Conservativity also has the following consequence.

Corollary 10.1.10. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. If T is conservative, then*

$$\sum_{n=0}^{\infty} \mu(T^{-n}(A)) = \infty$$

for all sets $A \in \mathcal{A}$ such that $\mu(A) > 0$.

Proof. Let $A \in \mathcal{A}$. If $\sum_{n=0}^{\infty} \mu(T^{-n}(A)) < \infty$, then $\mu(A_\infty) = 0$ by Borel–Cantelli Lemma (Lemma A.1.20). Moreover, $\mu(A \backslash A_\infty) = 0$ according to Corollary 10.1.8. Therefore, we conclude that $\mu(A) = \mu(A \cap A_\infty) + \mu(A \backslash A_\infty) = 0$. □

We now prove a characterization of ergodicity + conservativity.

Theorem 10.1.11. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. Then T is conservative and ergodic if and only if $\mu(X \backslash A_\infty) = 0$ for every $A \in \mathcal{A}$ such that $\mu(A) > 0$.*

Proof. Assume first that $T : X \to X$ is conservative and ergodic. Fix $A \in \mathcal{A}$ with $\mu(A) > 0$. We have earlier observed that $T^{-1}(A_\infty) = A_\infty$. Then, due to the ergodicity of μ, either $\mu(A_\infty) = 0$ or $\mu(X \backslash A_\infty) = 0$. In the former case, $\mu(A) = \mu(A \backslash A_\infty)$ and Corollary 10.1.8 implies that $\mu(A) = 0$. As $\mu(A) > 0$ by assumption, this means that this case never happens. Only the latter case $\mu(X \backslash A_\infty) = 0$ occurs, and this proves the implication ⇒.

We now prove the converse implication.

Let us first show conservativity. Let $A \in \mathcal{A}$. If $\mu(A) = 0$ then obviously $\mu(A \backslash A_\infty) = 0$. If $\mu(A) > 0$, then by assumption $\mu(X \backslash A_\infty) = 0$ and again $\mu(A \backslash A_\infty) = 0$. Corollary 10.1.8 then confirms the conservativity of T.

Now we establish ergodicity. Let $A \in \mathcal{A}$ be such that $T^{-1}(A) = A$. Then $A_\infty = A$. We must show that either $\mu(A) = 0$ or $\mu(X \backslash A) = 0$. Suppose that $\mu(A) > 0$. By assumption, we then have $\mu(X \backslash A_\infty) = 0$. Since $A_\infty = A$, this means that $\mu(X \backslash A) = 0$. □

For finite invariant measures, Theorem 10.1.11, in conjunction with Corollary 10.1.9, provides yet another characterization of ergodicity.

Corollary 10.1.12. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a finite T-invariant measure. Then T is ergodic if and only if $\mu(X \backslash A_\infty) = 0$ for every $A \in \mathcal{A}$ such that $\mu(A) > 0$.*

Sets that are visited by almost every point in the space will also play a crucial role. Accordingly, we make the following definition.

Definition 10.1.13. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. A set $A \in \mathcal{A}$ is said to be *absorbing* with respect to μ if $0 < \mu(A) < \infty$ and $\mu(X \setminus \bigcup_{k=0}^{\infty} T^{-k}(A)) = 0$.

Notice that any invariant measure which admits an absorbing set is σ-finite.

Obviously, a set $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$ and $\mu(X \setminus A_\infty) = 0$, is absorbing with respect to μ. Therefore, we have the following corollary to Theorem 10.1.11.

Corollary 10.1.14. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. If T is conservative and ergodic, then every $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$ is absorbing.*

To end this section, we briefly return to transformations of probability spaces and introduce a concept of weak metric exactness. Recall that the notion of metric exactness was described in Definition 8.4.4.

Definition 10.1.15. Let $(X, \mathcal{F}, \mu)$ be a Lebesgue probability space and $T : X \to X$ a transformation for which μ is a quasi-T-invariant measure. Then T is said to be *weakly metrically exact* if for each $A \in \mathcal{F}$ such that $\mu(A) > 0$ we have

$$\limsup_{n\to\infty} \mu\big(T^n(A)\big) = 1.$$

Note that each set $T^n(A)$ is measurable since T is a measurable transformation of a Lebesgue space.

Theorem 10.1.16. *Every weakly metrically exact transformation $T : (X, \mathcal{F}, \mu) \to (X, \mathcal{F}, \mu)$ is conservative and ergodic.*

Proof. We first prove ergodicity. Let $A \in \mathcal{A}$ be such that $T^{-1}(A) = A$. Then $T^n(A) \subseteq A$ for all $n \in \mathbb{N}$. Since μ is a probability measure, we must show that $\mu(A) \in \{0, 1\}$. Suppose that $\mu(A) > 0$. The weak metric exactness of T implies that

$$1 = \limsup_{n\to\infty} \mu\big(T^n(A)\big) \le \mu(A) \le 1.$$

So $\mu(A) = 1$ and T is ergodic.

To prove the conservativity of T, suppose for a contradiction that there exists a wandering set W such that $\mu(W) > 0$. Then $W \cap \bigcup_{n=1}^{\infty} T^{-n}(W) = \emptyset$. Therefore $\bigcup_{n=1}^{\infty} T^n(W) \subseteq X \setminus W$, and the weak metric exactness of T yields that

$$1 = \limsup_{n\to\infty} \mu\big(T^n(W)\big) \le \mu(X \setminus W) = 1 - \mu(W).$$

So $\mu(W) = 0$. This is a contradiction. Consequently, T is conservative. □

10.2 Invariant measures and inducing

Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a T-invariant measure. Fix $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$ and $\mu(A \setminus A_\infty) = 0$. Then $\mu(A \cap A_\infty) = \mu(A)$. Let

$A'_\infty = A \cap A_\infty$. The function $\tau_{A'_\infty} : A'_\infty \to \mathbb{N}$ given by the formula

$$\tau_{A'_\infty}(x) = \min\{n \in \mathbb{N} : T^n(x) \in A'_\infty\} \tag{10.1}$$

is well-defined and measurable when A'_∞ is endowed with the σ-algebra $\mathcal{A}|_{A'_\infty} := \{B \subseteq A'_\infty : B \in \mathcal{A}\}$ and $\mathbb{N}$ is equipped with the σ-algebra $\mathcal{P}(\mathbb{N})$ of all subsets of $\mathbb{N}$. Consequently, the map $T_{A'_\infty} : A'_\infty \to A'_\infty$ defined by

$$T_{A'_\infty}(x) = T^{\tau_{A'_\infty}(x)}(x) \tag{10.2}$$

is well-defined and measurable. The number $\tau_{A'_\infty}(x) \in \mathbb{N}$ is called the *first return time* of x to the set A'_∞ and, accordingly, the map $T_{A'_\infty}$ is called the *first return map* or the *induced map*. Given that $A'_\infty \subseteq A$ and $\mu(A'_\infty) = \mu(A)$, without loss of generality we will assume that $A'_\infty = A$ from the outset, hence alleging notation to τ_A and T_A.

Finally, in a similar way to Definition A.1.70, let μ_A be the conditional probability measure on $(A, \mathcal{A}|_A)$ defined by

$$\mu_A(B) = \frac{\mu(B)}{\mu(A)}, \quad \forall B \in \mathcal{A}|_A.$$

This measure has the following properties.

Theorem 10.2.1. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$.*

If μ is T-invariant and $\mu(A\backslash A_\infty) = 0$, then μ_A is T_A-invariant and T_A is conservative.

Conversely, if ν is a T_A-invariant probability measure on $(A, \mathcal{A}|_A)$, then there exists a T-invariant measure $\tilde{\mu}$ on $(X, \mathcal{A})$ such that $\tilde{\mu}_A = \nu$ on $(A, \mathcal{A}|_A)$. In fact, $\tilde{\mu}$ may be constructed as follows: for any $B \in \mathcal{A}$, let

$$\tilde{\mu}(B) = \sum_{n=0}^{\infty} \nu\left(A \cap T^{-n}(B) \setminus \bigcup_{k=1}^{n} T^{-k}(A) \right) = \sum_{n=0}^{\infty} \nu\big(\{x \in A \cap T^{-n}(B) : \tau_A(x) > n\}\big). \tag{10.3}$$

In particular, the set A is absorbing with respect to $\tilde{\mu}$ and $\tilde{\mu}$ is σ-finite. The measure $\tilde{\mu}$ is said to be induced *by ν.*

Proof. First, suppose that μ is T-invariant and that $0 < \mu(A) < \infty$ and $\mu(A\backslash A_\infty) = 0$. Let $B \in \mathcal{A}|_A$. Then

$$\begin{aligned}
\mu\big(T_A^{-1}(B)\big) &= \sum_{n=1}^{\infty} \mu\big(T_A^{-1}(B) \cap \tau_A^{-1}(n)\big) \\
&= \sum_{n=1}^{\infty} \mu\big(A \cap T^{-n}(B) \cap \tau_A^{-1}(n)\big) \\
&= \sum_{n=1}^{\infty} \mu\left(A \cap T^{-n}(B) \setminus \bigcup_{k=1}^{n-1} T^{-k}(A) \right) \\
&= \sum_{n=1}^{\infty} \mu\big(A \cap T^{-1}(B_{n-1})\big),
\end{aligned} \tag{10.4}$$

where

$$B_0 := B \quad \text{and} \quad B_n := T^{-n}(B) \Big\backslash \bigcup_{k=0}^{n-1} T^{-k}(A), \quad \forall n \in \mathbb{N}.$$

Observe that $\mu(B_n) \leq \mu(T^{-n}(B)) = \mu(B) \leq \mu(A) < \infty$ for every $n \geq 0$. Since $T^{-1}(B_{n-1}) = (A \cap T^{-1}(B_{n-1})) \cup B_n$ and $(A \cap T^{-1}(B_{n-1})) \cap B_n = \emptyset$ for all $n \in \mathbb{N}$, we obtain that

$$\mu(A \cap T^{-1}(B_{n-1})) = \mu(T^{-1}(B_{n-1})) - \mu(B_n) = \mu(B_{n-1}) - \mu(B_n)$$

for all $n \in \mathbb{N}$. Therefore, by (10.4),

$$\mu(T_A^{-1}(B)) = \lim_{n\to\infty}(\mu(B) - \mu(B_n)) \leq \mu(B). \tag{10.5}$$

This relation also holds for $A \backslash B$, that is,

$$\mu(T_A^{-1}(A \backslash B)) \leq \mu(A \backslash B).$$

Since $A = T_A^{-1}(A) = T_A^{-1}(B) \cup T_A^{-1}(A\backslash B)$ and $T_A^{-1}(B) \cap T_A^{-1}(A\backslash B) = \emptyset$, it follows that

$$\mu(T_A^{-1}(B)) = \mu(A) - \mu(T_A^{-1}(A\backslash B)) \geq \mu(A) - \mu(A\backslash B) = \mu(B). \tag{10.6}$$

It ensues from (10.5) and (10.6) that $\mu(T_A^{-1}(B)) = \mu(B)$. Thus $\mu_A \circ T_A^{-1} = \mu_A$. So μ_A is T_A-invariant. The conservativity of T_A with respect to μ_A is a direct consequence of Corollary 10.1.9.

To prove the converse implication, assume that ν is a probability measure on $(A, \mathcal{A}|_A)$ such that $\nu \circ T_A^{-1} = \nu$. We shall first show that the measure $\widetilde{\mu}$ given by (10.3) is T-invariant. Indeed, let $B \in \mathcal{A}$. Then

$$\begin{aligned}
\widetilde{\mu}(T^{-1}(B)) &= \sum_{n=0}^{\infty} \nu(\{x \in A \cap T^{-n}(T^{-1}(B)) : \tau_A(x) > n\}) \\
&= \sum_{n=0}^{\infty} \nu(\{x \in A \cap T^{-(n+1)}(B) : \tau_A(x) > n+1\}) \\
&\quad + \sum_{n=0}^{\infty} \nu(\{x \in A \cap T^{-(n+1)}(B) : \tau_A(x) = n+1\}) \\
&= \widetilde{\mu}(B) - \nu(A \cap B) + \sum_{n=1}^{\infty} \nu(\{x \in A \cap T^{-n}(B) : \tau_A(x) = n\}) \\
&= \widetilde{\mu}(B) - \nu(A \cap B) + \sum_{n=1}^{\infty} \nu(\{x \in A \cap T^{-n}(A \cap B) : \tau_A(x) = n\}) \\
&= \widetilde{\mu}(B) - \nu(A \cap B) + \sum_{n=1}^{\infty} \nu(T_A^{-1}(A \cap B) \cap \tau_A^{-1}(n)) \\
&= \widetilde{\mu}(B) - \nu(A \cap B) + \nu(T_A^{-1}(A \cap B)) \\
&= \widetilde{\mu}(B).
\end{aligned}$$

Thus $\widetilde{\mu}$ is T-invariant when ν is T_A-invariant.

Moreover, if $B \subseteq A$ then (10.3) reduces to $\tilde{\mu}(B) = \nu(A \cap B) = \nu(B)$. In particular, $\tilde{\mu}(A) = \nu(A) = 1$, and hence $\tilde{\mu}_A(B) = \frac{\tilde{\mu}(B)}{\tilde{\mu}(A)} = \tilde{\mu}(B) = \nu(B)$. That is, $\tilde{\mu} = \tilde{\mu}_A = \nu$ on $(A, \mathcal{A}|_A)$.

Furthermore, (10.3) gives

$$\tilde{\mu}\left(X \Big\backslash \bigcup_{k=1}^{\infty} T^{-k}(A)\right) = \sum_{n=0}^{\infty} \nu\left(\left\{x \in A \Big\backslash \bigcup_{k=1}^{\infty} T^{-(n+k)}(A) : \tau_A(x) > n\right\}\right) = \sum_{n=0}^{\infty} \nu(\emptyset) = 0.$$

Thus A is absorbing with respect to $\tilde{\mu}$. In addition, this shows that $\tilde{\mu}$ is σ-finite when $\tilde{\mu}(T^{-k}(A)) < \infty$ for all $k \in \mathbb{N}$. Since $\tilde{\mu}$ is T-invariant, this condition is equivalent to $\tilde{\mu}(A) < \infty$. And we saw above that $\tilde{\mu}(A) = 1$. Thus $\tilde{\mu}$ is σ-finite. □

Remark 10.2.2. It follows from (10.5) and (10.6) that $\lim_{n\to\infty} \mu(B_n) = 0$. In particular, taking $B = A$, we get that

$$\lim_{n\to\infty} \mu\left(T^{-n}(A) \Big\backslash \bigcup_{k=0}^{n-1} T^{-k}(A)\right) = 0.$$

Theorem 10.2.1 raises some interesting questions. Among others, in the second part of the theorem, is the induced measure $\tilde{\mu}$ unique? In Exercise 10.5.1, you will learn that this is generally not the case. However, we now prove that uniqueness prevails when the backward orbit of the set A covers the space almost everywhere.

Proposition 10.2.3. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. If μ is a T-invariant measure that admits an absorbing set A, then $\mu = \tilde{\mu}$ if ν is replaced by μ in formula* (10.3).
(*Formally, $\mu = \mu(A) \cdot \tilde{\mu}$, where $\tilde{\mu}$ is given by* (10.3) *with $\nu = \mu_A$.*)

Proof. Let $B \in \mathcal{A}$ be such that $\mu(B) < \infty$. For every $j \geq 0$, let

$$B_j := T^{-j}(B) \Big\backslash \bigcup_{k=0}^{j} T^{-k}(A).$$

Observe that $\mu(B_j) \leq \mu(T^{-j}(B)) = \mu(B) < \infty$ for every $j \geq 0$. As μ is T-invariant and $T^{-1}(B_{j-1}) \supseteq B_j$ for all $j \in \mathbb{N}$, we get for every $n \in \mathbb{N}$ that

$$\begin{aligned}
\mu(B \backslash A) - \mu(B_n) &= \sum_{j=1}^{n} [\mu(B_{j-1}) - \mu(B_j)] = \sum_{j=1}^{n} [\mu(T^{-1}(B_{j-1})) - \mu(B_j)] \\
&= \sum_{j=1}^{n} \mu(T^{-1}(B_{j-1}) \backslash B_j) \\
&= \sum_{j=1}^{n} \mu\left(\left(T^{-j}(B) \Big\backslash \bigcup_{k=1}^{j} T^{-k}(A)\right) \Big\backslash \left(T^{-j}(B) \Big\backslash \bigcup_{k=0}^{j} T^{-k}(A)\right)\right) \\
&= \sum_{j=1}^{n} \mu\left(A \cap T^{-j}(B) \Big\backslash \bigcup_{k=1}^{j} T^{-k}(A)\right).
\end{aligned}$$

Then

$$\mu(B) - \mu(B_n) = \mu(B \cap A) + \mu(B\backslash A) - \mu(B_n) = \sum_{j=0}^{n} \mu\left(A \cap T^{-j}(B) \backslash \bigcup_{k=1}^{j} T^{-k}(A) \right).$$

Replacing ν by μ in formula (10.3), we deduce that

$$\widetilde{\mu}(B) = \sum_{j=0}^{\infty} \mu\left(A \cap T^{-j}(B) \backslash \bigcup_{k=1}^{j} T^{-k}(A) \right) = \mu(B) - \lim_{n\to\infty} \mu(B_n).$$

Therefore, in order to complete the proof, we need to show that

$$\lim_{n\to\infty} \mu(B_n) = 0.$$

Fix $\varepsilon > 0$. Since A is absorbing, we have that $\mu(X \backslash \bigcup_{k=0}^{\infty} T^{-k}(A)) = 0$. Equivalently,

$$\mu\left(X \Big\backslash \bigcup_{k=0}^{\infty} A^{(k)} \right) = 0, \tag{10.7}$$

where

$$A^{(k)} := T^{-k}(A) \Big\backslash \bigcup_{j=0}^{k-1} T^{-j}(A).$$

The usefulness of the $A^{(k)}$s lies in their mutual disjointness. Indeed, relation (10.7) implies that $\mu(B \backslash \bigcup_{k=0}^{\infty} A^{(k)}) = 0$. Then

$$\mu(B) = \mu\left(B \cap \bigcup_{k=0}^{\infty} A^{(k)} \right) = \mu\left(\bigcup_{k=0}^{\infty} (B \cap A^{(k)}) \right) = \sum_{k=0}^{\infty} \mu(B \cap A^{(k)}). \tag{10.8}$$

Since $\mu(B) < \infty$, there exists $\ell_\varepsilon \in \mathbb{N}$ so large that

$$\sum_{\ell=\ell_\varepsilon+1}^{\infty} \mu(B \cap A^{(\ell)}) < \frac{\varepsilon}{2}. \tag{10.9}$$

Relation (10.7) also ensures that $\mu(B_n \backslash \bigcup_{k=0}^{\infty} A^{(k)}) = 0$ for every $n \geq 0$. So, like for B,

$$\mu(B_n) = \sum_{k=0}^{\infty} \mu(B_n \cap A^{(k)}).$$

But

$$\begin{aligned} B_n \cap A^{(k)} &= \left[T^{-n}(B) \Big\backslash \bigcup_{i=0}^{n} T^{-i}(A) \right] \cap \left[T^{-k}(A) \Big\backslash \bigcup_{j=0}^{k-1} T^{-j}(A) \right] \\ &= [T^{-n}(B) \cap T^{-k}(A)] \Big\backslash \bigcup_{i=0}^{\max\{n,k-1\}} T^{-i}(A) \end{aligned}$$

$$= \begin{cases} \emptyset & \text{if } k \le n \\ T^{-n}(B \cap T^{-(k-n)}(A)) \backslash \bigcup_{i=0}^{k-1} T^{-i}(A) & \text{if } k > n \end{cases}$$
$$= \begin{cases} \emptyset & \text{if } k \le n \\ [T^{-n}(B \cap T^{-(k-n)}(A)) \backslash \bigcup_{i=n}^{k-1} T^{-i}(A)] \backslash \bigcup_{i=0}^{n-1} T^{-i}(A) & \text{if } k > n \end{cases}$$
$$= \begin{cases} \emptyset & \text{if } k \le n \\ T^{-n}(B \cap A^{(k-n)}) \backslash \bigcup_{i=0}^{n-1} T^{-i}(A) & \text{if } k > n. \end{cases}$$

Consequently,

$$\mu(B_n) = \sum_{k>n} \mu(B_n \cap A^{(k)}) = \sum_{k=n+1}^{\infty} \mu\left(T^{-n}(B \cap A^{(k-n)}) \Big\backslash \bigcup_{i=0}^{n-1} T^{-i}(A) \right) = \sum_{\ell=1}^{\infty} \mu(B_n^{(\ell)}),$$

where

$$B_n^{(\ell)} := T^{-n}(B \cap A^{(\ell)}) \Big\backslash \bigcup_{i=0}^{n-1} T^{-i}(A) \subseteq T^{-(n+\ell)}(A) \Big\backslash \bigcup_{i=0}^{n+\ell-1} T^{-i}(A) = A^{(n+\ell)}.$$

Remark 10.2.2 asserts that $\lim_{N\to\infty} \mu(A^{(N)}) = 0$. Then

$$\lim_{n\to\infty} \mu(B_n^{(\ell)}) = 0, \quad \forall \ell \in \mathbb{N}. \tag{10.10}$$

As $\mu(B_n^{(\ell)}) \le \mu(T^{-n}(B \cap A^{(\ell)})) = \mu(B \cap A^{(\ell)})$ for each $\ell, n \in \mathbb{N}$, it follows from (10.9) that

$$\sum_{\ell=\ell_\varepsilon+1}^{\infty} \mu(B_n^{(\ell)}) < \frac{\varepsilon}{2}, \quad \forall n \in \mathbb{N}.$$

By (10.10), there exists $n_\varepsilon \in \mathbb{N}$ so large that for all $1 \le \ell \le \ell_\varepsilon$ and all $n \ge n_\varepsilon$,

$$\mu(B_n^{(\ell)}) \le \frac{\varepsilon}{2\ell_\varepsilon}.$$

So, for all $n \ge n_\varepsilon$, we have

$$\mu(B_n) = \sum_{\ell=1}^{\infty} \mu(B_n^{(\ell)}) = \sum_{\ell=1}^{\ell_\varepsilon} \mu(B_n^{(\ell)}) + \sum_{\ell=\ell_\varepsilon+1}^{\infty} \mu(B_n^{(\ell)}) \le \sum_{\ell=1}^{\ell_\varepsilon} \frac{\varepsilon}{2\ell_\varepsilon} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus $\lim_{n\to\infty} \mu(B_n) = 0$ and the proof is complete for sets B of finite measure.

Now, let $B \in \mathcal{A}$ be any set. Since $\mu(A^{(k)}) \le \mu(T^{-k}(A)) = \mu(A) < \infty$ for all $k \ge 0$, the sets $(B \cap A^{(k)})_{k=0}^{\infty}$ are of finite measure. Then the first part of this proof shows that $\mu(B \cap A^{(k)}) = \tilde{\mu}(B \cap A^{(k)})$. By (10.8) and the mutual disjointness of the $A^{(k)}$s, we conclude that

$$\mu(B) = \sum_{k=0}^{\infty} \mu(B \cap A^{(k)}) = \sum_{k=0}^{\infty} \tilde{\mu}(B \cap A^{(k)}) = \tilde{\mu}\left(B \cap \bigcup_{k=0}^{\infty} A^{(k)} \right) = \tilde{\mu}(B).$$

The last equality follows from the fact that $\tilde{\mu}(X \backslash \bigcup_{k=0}^{\infty} A^{(k)}) = 0$ according to Theorem 10.2.1. So $\mu = \tilde{\mu}$. □

Corollary 10.2.4. *Let $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ be an ergodic conservative measure-preserving transformation and any $A \in \mathcal{A}$ with $0 < \mu(A) < \infty$. Then $\mu = \widetilde{\mu}$ if ν is replaced by μ in formula* (10.3).

Proof. According to Corollary 10.1.14, any $A \in \mathcal{A}$ with $0 < \mu(A) < \infty$ is absorbing with respect to μ and Proposition 10.2.3 applies to any such A. □

Let $\varphi : X \to \mathbb{R}$ be a measurable function and $n \in \mathbb{N}$. Recall from Definition 8.2.10 that the nth *Birkhoff sum* of φ at a point $x \in X$ is

$$S_n\varphi(x) = \sum_{j=0}^{n-1} \varphi(T^j(x)).$$

Let $A \in \mathcal{A}$. Define the function $\varphi_A : A \to \mathbb{R}$ by the formula

$$\varphi_A(x) = S_{\tau_A(x)}\varphi(x). \tag{10.11}$$

In the next proposition, we describe properties that φ_A inherits from φ.

Proposition 10.2.5. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. Let also $\varphi : X \to \mathbb{R}$ be a measurable function. If μ is a T-invariant measure and A is an absorbing set with respect to μ, then:*

(a) *If $\varphi \in L^1(\mu)$, then $\varphi_A \in L^1(\mu_A)$.*

(b) *If $\varphi \geq 0$ or $\varphi \in L^1(\mu)$, then*

$$\int_A \varphi_A \, d\mu_A = \frac{1}{\mu(A)} \int_X \varphi \, d\mu.$$

If, in addition, T is conservative and ergodic, then the above two statements apply to all sets $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$.

Proof. Suppose first that $\varphi = \mathbb{1}_B$ for some $B \in \mathcal{A}$ such that $0 < \mu(B) < \infty$. In view of Proposition 10.2.3, we have

$$\begin{aligned}
\int_X \mathbb{1}_B \, d\mu = \mu(B) &= \sum_{k=0}^{\infty} \mu(\{x \in A \cap T^{-k}(B) : \tau_A(x) > k\}) \\
&= \sum_{n=1}^{\infty} \sum_{j=0}^{n-1} \mu(\{x \in A \cap T^{-j}(B) : \tau_A(x) = n\}) \\
&= \sum_{n=1}^{\infty} \sum_{j=0}^{n-1} \int_{\tau_A^{-1}(n)} \mathbb{1}_{T^{-j}(B)} \, d\mu \\
&= \sum_{n=1}^{\infty} \sum_{j=0}^{n-1} \int_{\tau_A^{-1}(n)} \mathbb{1}_B \circ T^j \, d\mu
\end{aligned}$$

$$= \sum_{n=1}^{\infty} \int_{\tau_A^{-1}(n)} S_n \mathbb{1}_B \, d\mu$$
$$= \sum_{n=1}^{\infty} \int_{\tau_A^{-1}(n)} \varphi_A \, d\mu$$
$$= \int_A \varphi_A \, d\mu = \mu(A) \int_A \varphi_A \, d\mu_A,$$

and we are done with this case. If $\varphi : X \to \mathbb{R}$ is a simple measurable function, that is, $\varphi = \sum_{i=1}^n a_i \varphi^{(i)}$, where all $a_i \in \mathbb{R}$, $1 \le i \le n$, and all $\varphi^{(i)}$, $1 \le i \le n$, are characteristic functions of some measurable sets with positive and finite measures, then

$$\int_X \varphi \, d\mu = \sum_{i=1}^n a_i \int_X \varphi^{(i)} \, d\mu$$
$$= \mu(A) \sum_{i=1}^n a_i \int_A \varphi_A^{(i)} \, d\mu_A$$
$$= \mu(A) \int_A \sum_{i=1}^n a_i \varphi_A^{(i)} \, d\mu_A$$
$$= \mu(A) \int_A \left(\sum_{i=1}^n a_i \varphi^{(i)} \right)_A \, d\mu_A$$
$$= \mu(A) \int_A \varphi_A \, d\mu_A,$$

and we are done in this case as well. The next case is to consider an arbitrary nonnegative measurable function $\varphi : X \to [0, \infty)$. Then φ is the pointwise limit of an increasing sequence of nonnegative simple measurable functions $(\varphi^{(n)})_{n=1}^{\infty}$. It is easy to see that φ_A is the pointwise limit of the increasing sequence of nonnegative measurable functions $(\varphi_A^{(n)})_{n=1}^{\infty}$. Applying twice the monotone convergence theorem (Theorem A.1.35), we then get that

$$\int_X \varphi \, d\mu = \lim_{n\to\infty} \int_X \varphi^{(n)} \, d\mu = \lim_{n\to\infty} \mu(A) \int_A \varphi_A^{(n)} \, d\mu_A = \mu(A) \int_A \lim_{n\to\infty} \varphi_A^{(n)} \, d\mu_A = \mu(A) \int_A \varphi_A \, d\mu_A.$$

We are also done in this case. Since $|\varphi_A| \le |\varphi|_A$, this in particular shows that if $\varphi \in L^1(\mu)$, then $\varphi_A \in L^1(\mu_A)$. Moreover, writing $\varphi = \varphi^+ - \varphi^-$, where $\varphi^+ = \max\{\varphi, 0\}$ and $\varphi^- = \max\{-\varphi, 0\}$, both functions φ^+ and φ^- are in $L^1(\mu)$ when φ is and

$$\int_X \varphi \, d\mu = \int_X (\varphi^+ - \varphi^-) \, d\mu = \mu(A) \int_A (\varphi_A^+ - \varphi_A^-) \, d\mu_A = \mu(A) \int_A \varphi_A \, d\mu_A. \qquad \square$$

Observe that if $\varphi \equiv 1$, then $\varphi_A \equiv \tau_A$. As an immediate consequence of the previous proposition, we obtain that the average of the first return time to a set is inversely proportional to the relative measure of that set in the space.

Lemma 10.2.6 (Kac's lemma). *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. If μ is a T-invariant measure and A is an absorbing set with respect to μ, then*

$$\int_A \tau_A \, d\mu_A = \frac{\mu(X)}{\mu(A)}.$$

In particular:

(a) *The measure μ is finite ($\mu(X) < \infty$) if and only if*

$$\int_A \tau_A \, d\mu_A < \infty.$$

(b) *If μ is a probability measure, then*

$$\int_A \tau_A \, d\mu_A = \frac{1}{\mu(A)}.$$

If, in addition, T is conservative and ergodic, then the above statements apply to every set $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$.

We now study the transmission of ergodicity between T and T_A.

Proposition 10.2.7. *Let $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ be a measure-preserving transformation and $A \in \mathcal{A}$ with $0 < \mu(A) < \infty$.*

If $T : X \to X$ is ergodic and conservative with respect to μ, then $T_A : A \to A$ is ergodic with respect to μ_A.

Conversely, if A is absorbing with respect to μ and $T_A : A \to A$ is ergodic with respect to μ_A, then $T : X \to X$ is ergodic with respect to μ.

Proof. First, suppose that $T : X \to X$ is ergodic and conservative. Let $C \subseteq A$ be completely T_A-invariant and assume that $\mu_A(C) > 0$. This latter assumption implies that $\mu(C) > 0$. By Theorem 10.1.11, we know that $\mu(X \backslash C_\infty) = 0$. But since $T_A^{-1}(C) = C$, we also have that $T_A^{-1}(A \backslash C) = A \backslash C$. Therefore, $A \backslash C \subseteq X \backslash C_\infty$, and hence $\mu(A \backslash C) = 0$. So $\mu_A(A \backslash C) = 0$ and T_A is ergodic.

In order to prove the converse, suppose that $T_A : A \to A$ is ergodic with respect to μ_A and let $B \in \mathcal{A}$ be such that $T^{-1}(B) = B$ and $\mu(B) > 0$. Suppose also that A is absorbing with respect to μ. Then $\mu(X \backslash \bigcup_{n=0}^\infty T^{-n}(A)) = 0$ and there exists $k \geq 0$ such that $\mu(B \cap T^{-k}(A)) > 0$. Therefore,

$$\mu(B \cap A) = \mu\big(T^{-k}(B \cap A)\big) = \mu\big(T^{-k}(B) \cap T^{-k}(A)\big) = \mu\big(B \cap T^{-k}(A)\big) > 0.$$

Recall that $(B\cap A)_\infty = \bigcap_{k=1}^\infty \bigcup_{n=k}^\infty T^{-n}(B\cap A)$ and write $(B\cap A)_\infty^{T_A} := \bigcap_{k=1}^\infty \bigcup_{n=k}^\infty T_A^{-n}(B\cap A)$. Since μ_A is a probability measure and $T_A : A \to A$ is ergodic, Corollary 10.1.12 states that $\mu_A(A\backslash(B\cap A)_\infty^{T_A}) = 0$ and this implies that

$$1 = \mu_A\big((B\cap A)_\infty^{T_A}\big) \le \mu_A\left(\bigcup_{n=0}^\infty T_A^{-n}(B\cap A)\right) \le \mu_A\left(A\cap\bigcup_{n=0}^\infty T^{-n}(B\cap A)\right) \le 1.$$

This means that $\mu(A\backslash\bigcup_{n=0}^\infty T^{-n}(B\cap A)) = 0$. Using the T-invariance of μ, it follows that $\mu(\bigcup_{k=0}^\infty T^{-k}(A)\backslash\bigcup_{n=0}^\infty T^{-n}(B\cap A)) = 0$. By hypothesis, $\mu(X\backslash\bigcup_{k=0}^\infty T^{-k}(A)) = 0$. Then $\mu(X\backslash\bigcup_{n=0}^\infty T^{-n}(B\cap A)) = 0$. Consequently,

$$\mu(X\backslash B) = \mu\left(X\backslash\bigcup_{n=0}^\infty T^{-n}(B)\right) \le \mu\left(X\backslash\bigcup_{n=0}^\infty T^{-n}(B\cap A)\right) = 0.$$

Hence, $\mu(X\backslash B) = 0$, and thus T is ergodic. □

10.3 Ergodic theorems

Birkhoff's ergodic theorem (Theorem 8.2.11 and Corollaries 8.2.14–8.2.15) concerns measure-preserving dynamical systems acting on probability spaces. Its ramifications are manifold. We studied some of them in the last two chapters. In this section, we use it with the inducing procedure described in the previous section to prove Hopf's ergodic theorem, which holds for measure-preserving transformations of σ-finite measure spaces.

But first, as a straightforward consequence of Birkhoff's ergodic theorem, we have the following two useful facts.

Proposition 10.3.1. *Let $T : X \to X$ be an ergodic measure-preserving transformation of a probability space $(X, \mathcal{A}, \mu)$. Fix $A \in \mathcal{A}$ such that $\mu(A) > 0$. For every $x \in X$, let $(k_n(x))_{n=1}^\infty$ be the sequence of successive times at which the iterates of x visit the set A. Then*

$$\lim_{n\to\infty}\frac{k_{n+1}(x)}{k_n(x)} = 1 \quad \textit{for } \mu\textit{-a. e. } x \in X.$$

Proof. Note that $S_{k_n(x)}\mathbb{1}_A(x) = n$. It follows from the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14) that for μ-a. e. $x \in X$,

$$\begin{aligned}\lim_{n\to\infty}\frac{k_{n+1}(x)}{k_n(x)} &= \lim_{n\to\infty}\left(\frac{n}{k_n(x)}\cdot\frac{k_{n+1}(x)}{n+1}\right)\\ &= \lim_{n\to\infty}\frac{1}{k_n(x)}S_{k_n(x)}\mathbb{1}_A(x)\cdot\frac{1}{\lim_{n\to\infty}\frac{1}{k_{n+1}(x)}S_{k_{n+1}(x)}\mathbb{1}_A(x)}\\ &= \mu(A)\cdot\frac{1}{\mu(A)} = 1.\end{aligned}$$

□

Proposition 10.3.2. *Let $T : X \to X$ be an ergodic measure-preserving transformation of a probability space $(X, \mathcal{A}, \mu)$. If $f \in L^1(\mu)$, then*

$$\lim_{n\to\infty} \frac{1}{n} f(T^n(x)) = 0 \quad \text{for } \mu\text{-a.e. } x \in X.$$

Proof. It follows from the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14) that for μ-a. e. $x \in X$,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} f(T^n(x)) &= \lim_{n\to\infty} \frac{1}{n+1} f(T^n(x)) = \lim_{n\to\infty} \frac{1}{n+1}\big(S_{n+1}f(x) - S_n f(x)\big)\\ &= \lim_{n\to\infty} \frac{1}{n+1} S_{n+1}f(x) - \lim_{n\to\infty} \frac{1}{n+1} S_n f(x)\\ &= \lim_{n\to\infty} \frac{1}{n+1} S_{n+1}f(x) - \lim_{n\to\infty} \frac{1}{n} S_n f(x) = 0.\end{aligned}$$ □

The following result is an application of Birkhoff's ergodic theorem to the ergodic theory of transformations preserving σ-finite measures.

Theorem 10.3.3 (Hopf's ergodic theorem). *Let $T : X \to X$ be an ergodic and conservative measure-preserving transformation of a σ-finite measure space $(X, \mathcal{A}, \mu)$. If $f, g \in L^1(\mu)$ and $\int_X g\, d\mu \neq 0$, then*

$$\lim_{n\to\infty} \frac{S_n f(x)}{S_n g(x)} = \frac{\int_X f\, d\mu}{\int_X g\, d\mu} \quad \text{for } \mu\text{-a.e. } x \in X.$$

Proof. (Note: We strongly recommend that the reader work on Exercise 10.5.2 before examining this proof.) Since μ is σ-finite, there are mutually disjoint sets $\{X_j\}_{j=1}^\infty$ such that $0 < \mu(X_j) < \infty$ for all $j \in \mathbb{N}$ and $\bigcup_{j=1}^\infty X_j = X$.

Fix $j \in \mathbb{N}$. Since T is measure-preserving and conservative, Corollary 10.1.8 affirms that $\mu(X_j \backslash (X_j)_\infty) = 0$. Thus the first return time to X_j and the first return map to X_j are well-defined by (10.1) and (10.2), respectively. Let $\tau_j := \tau_{X_j}$ and $T_j := T_{X_j}$. Let also $\varphi : X \to \mathbb{R}$ and set $\varphi_j := \varphi_{X_j}$ per (10.11). Given $x \in X_j$, let $S_n^j \varphi_j(x) := \sum_{i=0}^{n-1} \varphi_j(T_j^i(x))$. For every $k \in \mathbb{N}$, let $j_k(x)$ be the largest integer $n \geq 0$ such that $\sum_{i=0}^{n-1} \tau_j(T_j^i(x)) \leq k$. Then

$$S_k\varphi(x) = S_{j_k(x)}^j \varphi_j(x) + S_{\Delta k(x)}\varphi\big(T_j^{j_k(x)}(x)\big),$$

where $\Delta k(x) := k - \sum_{i=0}^{j_k(x)-1} \tau_j(T_j^i(x)) \geq 0$. Consequently,

$$\frac{1}{j_k(x)} S_k\varphi(x) = \frac{1}{j_k(x)} S_{j_k(x)}^j \varphi_j(x) + \frac{1}{j_k(x)} S_{\Delta k(x)}\varphi\big(T_j^{j_k(x)}(x)\big). \tag{10.12}$$

But

$$\left|\frac{1}{j_k(x)} S_{\Delta k(x)}\varphi\big(T_j^{j_k(x)}(x)\big)\right| \leq \frac{1}{j_k(x)} S_{\Delta k(x)}|\varphi|\big(T_j^{j_k(x)}(x)\big) \leq \frac{1}{j_k(x)} |\varphi|_j\big(T_j^{j_k(x)}(x)\big). \tag{10.13}$$

Let $\mu_j := \mu_{X_j}$. Since T is measure-preserving, ergodic and conservative, Theorem 10.2.1 and Proposition 10.2.7 assert that T_j is measure-preserving, conservative and ergodic with respect to μ_j. Moreover, Proposition 10.2.5 states that if $\varphi \in L^1(\mu)$, then $|\varphi|_j \in L^1(\mu_j)$. It follows from Proposition 10.3.2 (with $T = T_j$, $\mu = \mu_j$, and $f = |\varphi|_j$) that the right-hand side of (10.13) approaches 0 for μ-a. e. $x \in X_j$, and so does the left-hand side:

$$\lim_{k\to\infty}\left|\frac{1}{j_k(x)}S_{\Delta k(x)}\varphi(T_j^{j_k(x)}(x))\right| = 0 \quad \text{for } \mu\text{-a. e. } x \in X_j.$$

It ensues from this, (10.12), the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14) (with $T = T_j$, $\mu = \mu_j$ and $\varphi = f_j, g_j$) and Proposition 10.2.5 that for μ-a. e. $x \in X_j$,

$$\frac{S_kf(x)}{S_kg(x)} = \frac{\frac{1}{j_k(x)}S_kf(x)}{\frac{1}{j_k(x)}S_kg(x)} = \frac{\frac{1}{j_k(x)}S^j_{j_k(x)}f_j(x) + \frac{1}{j_k(x)}S_{\Delta k(x)}f(T_j^{j_k(x)}(x))}{\frac{1}{j_k(x)}S^j_{j_k(x)}g_j(x) + \frac{1}{j_k(x)}S_{\Delta k(x)}g(T_j^{j_k(x)}(x))} \underset{k\to\infty}{\longrightarrow} \frac{\int_{X_j} f_j\,d\mu_j}{\int_{X_j} g_j\,d\mu_j} = \frac{\int_X f\,d\mu}{\int_X g\,d\mu}.$$

Since $\bigcup_{j=1}^{\infty} X_j = X$, the conclusion holds for μ-a. e. $x \in X$. □

The following result rules out any hope for an ergodic theorem closer to Birkhoff's ergodic theorem in the case of infinite measures.

Corollary 10.3.4. *Let $T : X \to X$ be an ergodic and conservative measure-preserving transformation of a σ-finite and infinite measure space $(X, \mathcal{A}, \mu)$. If $f \in L^1(\mu)$, then*

$$\lim_{n\to\infty}\frac{1}{n}S_nf(x) = 0 \quad \textit{for } \mu\textit{-a. e. } x \in X.$$

Proof. Since μ is σ-finite and $\mu(X) = \infty$, there exists a sequence $(A_k)_{k=1}^{\infty}$ of measurable sets such that

$$0 < \mu(A_k) < \infty,\ \forall k \in \mathbb{N} \quad \text{and} \quad \lim_{k\to\infty}\mu(A_k) = \infty. \tag{10.14}$$

As $|f| \in L^1(\mu)$, we deduce from Hopf's ergodic theorem (Theorem 10.3.3) that for every $k \in \mathbb{N}$,

$$\limsup_{n\to\infty}\frac{1}{n}S_n|f| \le \limsup_{n\to\infty}\frac{S_n|f|}{S_n\mathbb{1}_{A_k}} = \frac{\int_X |f|\,d\mu}{\int_X \mathbb{1}_{A_k}\,d\mu} = \frac{\|f\|_1}{\mu(A_k)} \quad \mu\text{-a. e.}$$

So, by (10.14), $\limsup_{n\to\infty}\frac{1}{n}S_n|f| = 0$ μ-a. e. Since $|\frac{1}{n}S_nf| \le \frac{1}{n}S_n|f|$, we are done. □

As a matter of fact, as the next two results show, Hopf's ergodic theorem precludes the existence of even weaker forms of Birkhoff's ergodic theorem in the case of infinite invariant measures. Indeed, for all $f \in L^1(\mu)$ Corollary 10.3.4 of Hopf's ergodic theorem implies that

$$\lim_{n\to\infty}\frac{1}{a_n}S_nf(x) = 0 \quad \text{for } \mu\text{-a. e. } x \in X$$

if there exists $C > 0$ such that $a_n \geq Cn$ for all n. We will now show that there are no constants $a_n > 0$ such that

$$\lim_{n\to\infty} \frac{1}{a_n} S_n f(x) = \int_X f\, d\mu \quad \text{for } \mu\text{-a. e. } x \in X, \quad \forall f \in L^1(\mu).$$

We will accomplish this in two steps. The first step will require the following proposition.

Proposition 10.3.5. *Let $T : X \to X$ be an ergodic and conservative measure-preserving transformation of a probability space $(X, \mathcal{A}, \mu)$. Let $a : [0,\infty) \to [0,\infty)$ be continuous, strictly increasing, and satisfying $\frac{a(x)}{x} \searrow 0$ as $x \nearrow \infty$. If $\int_X a(|f|)\, d\mu < \infty$, then*

$$\lim_{n\to\infty} \frac{a(|S_n f(x)|)}{n} = 0 \quad \textit{for } \mu\textit{-a. e. } x \in X.$$

Proof. An outline of a proof can be found in Exercise 10.5.3. □

In the first step, a sequence $(a(n))_{n=1}^{\infty}$ will be imposed properties that mimic those of the function a in Proposition 10.3.5. We will show that the outcome takes the form of a dichotomy: $\liminf_{n\to\infty} \frac{1}{a(n)} S_n f(x)$ is either 0 or ∞, for μ-a. e. $x \in X$ for all $f \in L^1_+(\mu) := \{f \in L^1(\mu) : f \geq 0 \text{ and } \int_X f\, d\mu > 0\}$.

Theorem 10.3.6. *Let $T : X \to X$ be an ergodic and conservative measure-preserving transformation of a σ-finite and infinite measure space $(X, \mathcal{A}, \mu)$. Let $(a(n))_{n=1}^{\infty}$ be a sequence such that*

$$a(n) \nearrow \infty \quad \textit{and} \quad \frac{a(n)}{n} \searrow 0 \qquad \textit{as } n \nearrow \infty.$$

(a) *If there exists $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$ and $\int_A a(\tau_A(x))\, d\mu(x) < \infty$, then*

$$\lim_{n\to\infty} \frac{1}{a(n)} S_n f(x) = \infty \quad \textit{for } \mu\textit{-a. e. } x \in X, \quad \forall f \in L^1_+(\mu).$$

(b) *Otherwise,*

$$\liminf_{n\to\infty} \frac{1}{a(n)} S_n f(x) = 0 \quad \textit{for } \mu\textit{-a. e. } x \in X, \quad \forall f \in L^1_+(\mu).$$

Proof. (a) Suppose that $A \in \mathcal{A}$ satisfies $0 < \mu(A) < \infty$ and $\int_A a(\tau_A(x))\, d\mu(x) < \infty$. Clearly, the set

$$I := \left\{ x \in X : \lim_{n\to\infty} \frac{S_n \mathbb{1}_A(x)}{a(n)} = \infty \right\}$$

is T-invariant. As T is ergodic, either $\mu(I) = 0$ or $\mu(X \backslash I) = 0$. We will show that I contains μ-a. e. $x \in A$. This will allow us to conclude that $\mu(I) \geq \mu(A) > 0$, and hence $\mu(X \backslash I) = 0$.

Since T is measure-preserving and conservative, Corollary 10.1.8 affirms that $\mu(A\backslash A_\infty) = 0$. Thus the nth return time to A and the nth return map to A are well-defined for all $n \in \mathbb{N}$. See Exercise 10.5.2. Since T is measure-preserving, ergodic, and conservative, Theorem 10.2.1 and Proposition 10.2.7 assert that T_A is measure-preserving, conservative and ergodic with respect to μ_A. As $\int_A a(\tau_A(x))\,d\mu_A(x) < \infty$, it follows from Proposition 10.3.5 (with $T = T_A$, $\mu = \mu_A$, and $f = \tau_A$) that

$$\lim_{n\to\infty} \frac{a(\tau_A^n(x))}{n} = \lim_{n\to\infty} \frac{a(S_n^{T_A}\tau_A(x))}{n} = 0 \quad \text{for } \mu\text{-a. e. } x \in A. \tag{10.15}$$

Since T is measure-preserving, ergodic and conservative, Theorem 10.1.11 states that $\mu(X\backslash A_\infty) = 0$. For every $x \in A_\infty$ and $n \in \mathbb{N}$, let $k_n(x)$ be the largest integer $k \geq 0$ such that $\tau_A^k(x) \leq n$. Since $S_{\tau_A^n}\mathbb{1}_A \equiv n$, we then have

$$\frac{S_n\mathbb{1}_A(x)}{a(n)} \geq \frac{S_{\tau_A^{k_n(x)}}\mathbb{1}_A(x)}{a(\tau_A^{k_n(x)+1}(x))} = \frac{k_n(x)}{a(\tau_A^{k_n(x)+1}(x))} = \frac{k_n(x)+1}{a(\tau_A^{k_n(x)+1}(x))}\frac{k_n(x)}{k_n(x)+1}. \tag{10.16}$$

Using (10.15), we deduce from (10.16) that the set I contains μ-a. e. $x \in A$. As pointed out at the beginning of the proof, this implies that $\mu(X\backslash I) = 0$. Let $f \in L^1_+(\mu)$. It ensues from Hopf's ergodic theorem (Theorem 10.3.3) that

$$\lim_{n\to\infty} \frac{S_nf(x)}{a(n)} = \lim_{n\to\infty} \frac{S_nf(x)}{S_n\mathbb{1}_A(x)} \lim_{n\to\infty} \frac{S_n\mathbb{1}_A(x)}{a(n)} = \frac{\int_X f\,d\mu}{\mu(A)} \lim_{n\to\infty} \frac{S_n\mathbb{1}_A(x)}{a(n)} = \infty$$

for μ-a. e. $x \in I$, that is, for μ-a. e. $x \in X$ since $\mu(X\backslash I) = 0$.

(b) Assume that there does not exist a set $B \in \mathcal{A}$ such that $0 < \mu(B) < \infty$ and $\int_B a(\tau_B(x))\,d\mu(x) < \infty$. Suppose for a contradiction that the conclusion of (b) is not satisfied. That is, suppose that there exist a set $A \in \mathcal{A}$ with $0 < \mu(A) < \infty$ and a function $f \in L^1_+(\mu)$ such that

$$F(x) := \liminf_{n\to\infty} \frac{1}{a(n)} S_nf(x) > 0 \quad \text{for } \mu\text{-a. e. } x \in A. \tag{10.17}$$

By Hopf's ergodic theorem, this actually holds for every $f \in L^1_+(\mu)$, with the same set A.

Moreover, since $\frac{a(n)}{n} \searrow$ as $n \nearrow \infty$ and $f \geq 0$, we have

$$\frac{S_nf}{a(n)} \circ T \leq \frac{1}{n} \cdot \frac{n}{a(n)} S_{n+1}f \leq \frac{1}{n} \cdot \frac{n+1}{a(n+1)} S_{n+1}f = \left(1 + \frac{1}{n}\right)\frac{S_{n+1}f}{a(n+1)}.$$

Using this inequality and Corollary 10.3.4, the function F satisfies

$$F \circ T = \sup_{n\in\mathbb{N}} \inf_{k\geq n} \frac{S_kf \circ T}{a(k)} \leq \sup_{n\in\mathbb{N}} \inf_{k\geq n} \left(1 + \frac{1}{k}\right)\frac{S_{k+1}f}{a(k+1)} = \sup_{n\in\mathbb{N}} \inf_{k\geq n} \frac{S_{k+1}f}{a(k+1)} = F$$

μ-a. e. on X. That is, F is μ-a. e. T-subinvariant, and hence F is constant μ-a. e. on X by the ergodicity of T (Exercise 8.5.47 generalizes to any measure space). By (10.17), we conclude that for every $f \in L^1_+(\mu)$ there is a constant $c = c(f) > 0$ such that

$$\liminf_{n\to\infty} \frac{1}{a(n)} S_n f(x) = c \quad \text{for } \mu\text{-a. e. } x \in A.$$

In particular, this applies to the function $\mathbb{1}_A$ on the set A, that is,

$$\liminf_{n\to\infty} \frac{1}{a(n)} S_n \mathbb{1}_A(x) = c > 0 \quad \text{for } \mu\text{-a. e. } x \in A.$$

According to Egorov's theorem (Theorem A.1.44), there is $B \in \mathcal{A}$ such that $B \subseteq A$, $\mu(B) > 0$, and over which the sequence $\left(\inf_{k\ge n} \frac{S_k \mathbb{1}_A}{a(k)}\right)_{k=1}^{\infty}$ converges uniformly to c. Consequently, there is $K \in \mathbb{N}$ such that

$$S_k \mathbb{1}_A(x) \ge \frac{c}{2} a(k), \quad \forall x \in B, \ \forall k \ge K.$$

Since $S_1 \mathbb{1}_A(x) = 1$ for all $x \in B$ and $f \ge 0$, we know that $S_k \mathbb{1}_A(x) \ge 1$ for all $k \in \mathbb{N}$ and all $x \in B$. Then there is $0 < \tilde{c} \le c/2$ such that

$$S_k \mathbb{1}_A(x) \ge \tilde{c}\, a(k), \quad \forall x \in B, \ \forall k \in \mathbb{N}.$$

In particular,

$$S_{\tau_B(x)} \mathbb{1}_A(x) \ge \tilde{c}\, a(\tau_B(x)), \quad \forall x \in B.$$

It ensues from the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14) (with $T = T_B$, $\mu = \mu_B$, and $\varphi = (\mathbb{1}_A)_B$) and Proposition 10.2.5 that for μ-a. e. $x \in B$,

$$\begin{aligned}
\frac{1}{n} \sum_{k=0}^{n-1} a(\tau_B(T_B^k(x))) &\le \frac{1}{n\tilde{c}} \sum_{k=0}^{n-1} S_{\tau_B(T_B^k(x))} \mathbb{1}_A(T_B^k(x)) \\
&= \frac{1}{\tilde{c}\, n} S_{\tau_B^n(x)} \mathbb{1}_A(x) = \frac{1}{\tilde{c}} \cdot \frac{1}{n} S_n^{T_B} (\mathbb{1}_A)_B(x) \\
&\underset{n\to\infty}{\longrightarrow} \frac{1}{\tilde{c}} \int_B (\mathbb{1}_A)_B \, d\mu_B = \frac{1}{\tilde{c}} \frac{1}{\mu(B)} \int_X \mathbb{1}_A \, d\mu = \frac{1}{\tilde{c}} \frac{\mu(A)}{\mu(B)}.
\end{aligned}$$

Since μ_B is T_B-invariant, it follows that

$$\int_B a(\tau_B)\, d\mu = \frac{1}{n} \sum_{k=0}^{n-1} \int_B a(\tau_B) \circ T_B^k \, d\mu$$

for every $n \in \mathbb{N}$. Passing to the limit $n \to \infty$, we conclude that $\int_B a(\tau_B)\, d\mu \le \frac{\mu(A)}{\tilde{c}\, \mu(B)} < \infty$, hence contradicting the hypothesis in (b). $\square$

In the second step, no restriction other than strict positivity will be put on the sequence $(a_n)_{n=1}^{\infty}$. We will construct a related sequence $(a(n))_{n=1}^{\infty}$ that satisfies the first step. We will again show that there are only two possibilities: either $\liminf_{n\to\infty} \frac{1}{a_n} S_n f(x) = 0$ or $\limsup_{n\to\infty} \frac{1}{a_n} S_n f(x) = \infty$, for μ-a. e. $x \in X$ for all $f \in L^1_+(\mu)$.

Theorem 10.3.7. *Let $T : X \to X$ be an ergodic and conservative measure-preserving transformation of a σ-finite and infinite measure space $(X, \mathcal{A}, \mu)$. Let $(a_n)_{n=1}^{\infty}$ be a sequence such that $a_n > 0$ for all $n \in \mathbb{N}$. Then*

(a) *either* $\liminf\limits_{n\to\infty} \dfrac{S_n f(x)}{a_n} = 0$ *for μ-a. e. $x \in X$, $\forall f \in L^1_+(\mu)$;*

(b) *or there is $n_k \nearrow \infty$ such that* $\lim\limits_{k\to\infty} \dfrac{S_{n_k} f(x)}{a_{n_k}} = \infty$ *for μ-a. e. $x \in X$, $\forall f \in L^1_+(\mu)$.*

Proof. If $(a_n)_{n=1}^{\infty}$ is bounded, then (b) holds. Indeed, let $f \in L^1_+(\mu)$. Since $\int_X f\, d\mu > 0$, there exist $\varepsilon > 0$ and $B \in \mathcal{A}$ such that $\mu(B) > 0$ and $f \geq \varepsilon$ on B. As T is conservative and ergodic, Theorem 10.1.11 affirms that $\mu(X \backslash B_\infty) = 0$. Therefore, for μ-a. e. $x \in X$ there exists a sequence $(n_k(x))_{k=1}^{\infty}$ such that $n_k(x) \nearrow \infty$ and $T^{n_k(x)}(x) \in B$. Hence, $S_{n_k(x)} f(x) \geq k\varepsilon$. In fact, $S_n f(x) \geq k\varepsilon$ for any $n \geq n_k(x)$ since $f \in L^1_+(\mu)$. Therefore, $\lim_{n\to\infty} S_n f(x) = \infty$. As $(a_n)_{n=1}^{\infty}$ is bounded, it follows that $\lim_{n\to\infty} \frac{S_n f(x)}{a_n} = \infty$ for μ-a. e. $x \in X$. So (b) holds with $(n_k)_{k=1}^{\infty} = (n)_{n=1}^{\infty}$ when $(a_n)_{n=1}^{\infty}$ is bounded.

We can thereby restrict our attention to the case $\limsup_{n\to\infty} a_n = \infty$. Suppose that (a) does not hold. That is, there exist a set $A \in \mathcal{A}$ with $\mu(A) > 0$ and a function $f \in L^1_+(\mu)$ such that

$$F(x) := \liminf_{n\to\infty} \frac{S_n f(x)}{a_n} > 0 \quad \text{for } \mu\text{-a. e. } x \in A. \tag{10.18}$$

By Hopf's ergodic theorem (Theorem 10.3.3), this actually holds for every $f \in L^1_+(\mu)$, with the same set A. Then for μ-a. e. $x \in A$,

$$0 \leq \limsup_{n\to\infty} \frac{a_n}{n} \leq \limsup_{n\to\infty} \frac{a_n}{S_n f(x)} \limsup_{n\to\infty} \frac{S_n f(x)}{n} = \left[\liminf_{n\to\infty} \frac{S_n f(x)}{a_n}\right]^{-1} \lim_{n\to\infty} \frac{S_n f(x)}{n} = 0$$

by (10.18) and Corollary 10.3.4. Thus $a_n = o(n)$ as $n \to \infty$. For every $n \in \mathbb{N}$, set

$$\overline{a}_n = \max_{1\leq k\leq n} a_k.$$

Clearly, $a_n \leq \overline{a}_n$ for all $n \in \mathbb{N}$ and $\overline{a}_n \nearrow \infty$ as $n \nearrow \infty$. Moreover, for each $n \in \mathbb{N}$ there is $1 \leq k(n) \leq n$ such that $\overline{a}_n = a_{k(n)}$. Note that $k(n) \to \infty$ as $n \to \infty$. Then

$$\liminf_{n\to\infty} \frac{S_n f}{a_n} \geq \liminf_{n\to\infty} \frac{S_n f}{\overline{a}_n} = \liminf_{n\to\infty} \frac{S_n f}{a_{k(n)}} \geq \liminf_{n\to\infty} \frac{S_{k(n)} f}{a_{k(n)}} \geq \liminf_{n\to\infty} \frac{S_n f}{a_n},$$

where the last inequality follows from the fact that the lim inf of a subsequence of a sequence is greater than or equal to the lim inf of the full sequence. By this and (10.18),

$$\liminf_{n\to\infty} \frac{S_n f}{\overline{a}_n} = \liminf_{n\to\infty} \frac{S_n f}{a_n} > 0 \quad \mu\text{-a. e. on } A, \quad \forall f \in L^1_+(\mu). \tag{10.19}$$

Next, set $b_n = \frac{\overline{a}_n}{n}$, and let $1 = n_0 < n_1 < \cdots$ be defined by

$$\{n_k\}_{k\in\mathbb{N}} = \{j \geq 2 : b_i > b_j,\ \forall 1 \leq i \leq j-1\}.$$

For every $k \geq 0$,

$$b_{n_k} > b_{n_{k+1}} \quad \text{and} \quad n_k b_{n_k} \leq n_{k+1} b_{n_{k+1}},$$

whence

$$0 < \frac{n_k}{n_{k+1}} \leq \frac{b_{n_{k+1}}}{b_{n_k}} < 1.$$

Thus there exists $\alpha_k \in (0,1]$ such that

$$\left(\frac{n_k}{n_{k+1}}\right)^{\alpha_k} = \frac{b_{n_{k+1}}}{b_{n_k}}.$$

Define

$$b(x) = \frac{b_{n_k} n_k^{\alpha_k}}{x^{\alpha_k}}, \quad x \in [n_k, n_{k+1}],\ k \in \mathbb{N}, \quad \textit{and} \quad a(x) = x b(x).$$

Evidently,

$$a(n_k) = \overline{a}_{n_k}, \quad \forall k \in \mathbb{N}.$$

By definition of the n_k's, we have that for $k \in \mathbb{N}$, $n \in [n_k, n_{k+1})$,

$$b_{n_k} \leq b_n \quad \text{and hence} \quad b(n) \leq b_n,$$

whereby

$$a(n) \leq \overline{a}_n, \quad \forall n \in \mathbb{N}.$$

Hence, following (10.19),

$$\liminf_{n\to\infty} \frac{S_n f}{a(n)} > 0 \quad \mu\text{-a. e. on } A, \quad \forall f \in L^1_+(\mu).$$

It is evident that

$$a(n) \nearrow \infty \quad \text{and} \quad \frac{a(n)}{n} \searrow 0 \quad \text{as } n \nearrow \infty.$$

So by Theorem 10.3.6,

$$\lim_{n\to\infty} \frac{S_n f}{a(n)} = \infty \quad \mu\text{-a. e. on } X, \quad \forall f \in L^1_+(\mu).$$

Then (b) follows since $a_{n_k} \leq \overline{a}_{n_k} = a(n_k)$. □

10.4 Absolutely continuous σ-finite invariant measures

In this section, we establish a very useful, relatively easy to verify, sufficient condition for a quasi-invariant probability measure to admit an absolutely continuous σ-finite invariant measure. This condition actually provides a σ-finite invariant measure equivalent to the original quasi-invariant probability measure. It goes back to the work of Marco Martens [44] and has been used many times, notably in [40]. It obtained its nearly final form in [70]. In contrast to Martens, where σ-compact metric spaces form the setting, the sufficient condition in [70] is stated for abstract measure spaces, and the proof uses the concept of Banach limit rather than weak convergence. In this section, we somewhat strengthen the assertions in [70].

We first identify the possible relations between σ-finite invariant measures with respect to which a system is ergodic and conservative. The following result is an analogue of Theorem 8.2.21.

Theorem 10.4.1. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. If μ_1 and μ_2 are σ-finite T-invariant measures and if T is ergodic and conservative with respect to both measures, then either μ_1 and μ_2 are mutually singular or else they coincide up to a positive multiplicative constant.*

Proof. We may assume that neither $\mu_1 \equiv 0$ nor $\mu_2 \equiv 0$. Since both measures are σ-finite, there is a sequence $(Y_n)_{n=1}^{\infty}$ of mutually disjoint measurable sets such that $\max\{\mu_1(Y_n), \mu_2(Y_n)\} < \infty$ for all $n \in \mathbb{N}$ and $X = \bigcup_{n=1}^{\infty} Y_n$.

First, suppose that $\mu_1(Y_n) > 0$ for some $n \in \mathbb{N}$ and that μ_1 and μ_2 coincide on Y_n up to a positive multiplicative constant. Without loss of generality, assume that $n = 1$ and that $\mu_1|_{Y_1} = \mu_2|_{Y_1}$. It immediately follows from Corollary 10.2.4 that $\mu_1 = \mu_2$, and we are done in this case.

Now, assume that μ_1 and μ_2 do not coincide on X up to any positive multiplicative constant. For each $n \in \mathbb{N}$, select a set Z_n in the following way:

(1) If $\mu_1(Y_n) \cdot \mu_2(Y_n) > 0$, it ensues from the previous case that $\mu_1|_{Y_n}$ and $\mu_2|_{Y_n}$ cannot be equal up to any positive multiplicative constant. Hence, $\mu_1|_{Y_n} \neq \mu_2|_{Y_n}$. Combining Proposition 10.2.7 and Theorem 8.2.21, we deduce that the measures $\mu_1|_{Y_n}$ and $\mu_2|_{Y_n}$ are mutually singular, that is, there is a measurable set $Z_n \subseteq Y_n$ such that $\mu_1(Z_n) = 0$ and $\mu_2(Y_n \backslash Z_n) = 0$.
(2) If $\mu_1(Y_n) = 0$, set $Z_n = Y_n$.
(3) Otherwise, set $Z_n = \emptyset$. (In this case, $\mu_2(Y_n) = 0$.)

Observe that $Z_n \subseteq Y_n$ for all $n \in \mathbb{N}$. Therefore, the sets $(Z_n)_{n=1}^{\infty}$ are mutually disjoint. Moreover, setting $Z := \bigcup_{n=1}^{\infty} Z_n \in \mathcal{A}$, it turns out that

$$\mu_1(Z) = \sum_{n=1}^{\infty} \mu_1(Z_n) = 0$$

while

$$\mu_2(X\backslash Z) = \sum_{n=1}^{\infty} \mu_2(Y_n\backslash Z) = \sum_{n=1}^{\infty} \mu_2(Y_n\backslash Z_n) = 0.$$

So the measures μ_1 and μ_2 are mutually singular. □

The preceding theorem allows us to derive the uniqueness of a σ-finite invariant measure which is absolutely continuous with respect to a given quasi-invariant measure, assuming that the transformation is ergodic and conservative with respect to the quasi-invariant measure.

Theorem 10.4.2. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and m a σ-finite quasi-T-invariant measure. If T is ergodic and conservative with respect to m then, up to a positive multiplicative constant, there exists at most one nonzero σ-finite T-invariant measure μ which is absolutely continuous with respect to m.*

Proof. Suppose that μ_1 and μ_2 are nonzero σ-finite T-invariant measures absolutely continuous with respect to m. Since m is ergodic and conservative, so are the measures μ_1 and μ_2. It follows from Theorem 10.4.1 that if μ_1 and μ_2 do not coincide up to a positive multiplicative constant, then they are mutually singular. But this means that there exists a measurable set $Y \subseteq X$ such that $\mu_1(Y) = 0$ and $\mu_2(X\backslash Y) = 0$. So

$$0 \le \mu_1\left(\bigcup_{n=0}^{\infty} T^{-n}(Y)\right) \le \sum_{n=0}^{\infty} \mu_1(T^{-n}(Y)) = \sum_{n=0}^{\infty} \mu_1(Y) = 0. \tag{10.20}$$

On the other hand, $\mu_2(Y) > 0$. Since $\mu_2 \ll m$, this implies that $m(Y) > 0$. Thus $m(X\backslash \bigcup_{n=0}^{\infty} T^{-n}(Y)) = 0$ by virtue of Theorem 10.1.11. Since $\mu_1 \ll m$, this forces $\mu_1(X\backslash \bigcup_{n=0}^{\infty} T^{-n}(Y)) = 0$. Along with (10.20), this gives that $\mu_1(X) = 0$. This contradicts the assumption that $\mu_1(X) \neq 0$. □

We now introduce the concept of Martens map.

Definition 10.4.3. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. Let also m be a quasi-T-invariant probability measure. The transformation T is called a *Martens map* if it admits a countable family $\{X_n\}_{n=0}^{\infty}$ of subsets of X with the following properties:

(a) $X_n \in \mathcal{A}$, $\forall n \ge 0$.
(b) $m(X\backslash \bigcup_{n=0}^{\infty} X_n) = 0$.
(c) For all $m, n \ge 0$, there exists $j \ge 0$ such that $m(X_m \cap T^{-j}(X_n)) > 0$.
(d) For all $j \ge 0$, there exists $K_j \ge 1$ such that for all $A, B \in \mathcal{A}$ with $A \cup B \subseteq X_j$ and for all $n \ge 0$,

$$m(T^{-n}(A))m(B) \le K_j m(A) m(T^{-n}(B)).$$

(e) $\sum_{n=0}^{\infty} m(T^{-n}(X_0)) = \infty$.

(f) $T(\bigcup_{j=l}^{\infty} Y_j) \in \mathcal{A}$ for all $l \geq 0$, where $Y_j := X_j \setminus \bigcup_{i<j} X_i$.
(g) $\lim_{l\to\infty} m(T(\bigcup_{j=l}^{\infty} Y_j)) = 0$.

The family $\{X_n\}_{n=0}^{\infty}$ is called a *Martens cover*.

Remark 10.4.4.
(1) Without loss of generality, condition (b) can be replaced by $\bigcup_{n=0}^{\infty} X_n = X$.
(2) Condition (c) imposes that $m(X_n) > 0$ for all $n \geq 0$.
(3) In light of Corollary 10.1.10, if T is conservative with respect to μ then condition (e) is fulfilled.
(4) In conditions (f-g), note that $\bigcup_{j=l}^{\infty} Y_j = \bigcup_{j=0}^{\infty} X_j \setminus \bigcup_{i<l} X_i \subseteq X \setminus \bigcup_{i<l} X_i$.
(5) If the map $T : X \to X$ is finite-to-one, then condition (g) is satisfied. For then, $\bigcap_{l=1}^{\infty} T(\bigcup_{j=l}^{\infty} Y_j) = \emptyset$.

Let l^∞ denote the Banach space of all bounded real-valued sequences $x = (x_n)_{n=1}^{\infty}$ with norm $\|x\|_\infty := \sup_{n\in\mathbb{N}} |x_n|$. Recall that a *Banach limit* is a shift-invariant positive continuous linear functional $l_B : l^\infty \to \mathbb{R}$ which extends the usual limits. More precisely, for all sequences $x = (x_n)_{n=1}^{\infty}$, $y = (y_n)_{n=1}^{\infty} \in l^\infty$ and $\alpha, \beta \in \mathbb{R}$, the following properties hold:
(a) $l_B(\alpha x + \beta y) = \alpha\, l_B(x) + \beta\, l_B(y)$ (linearity).
(b) $\|l_B\| := \sup\{|l_B(x)| : \|x\|_\infty \leq 1\} < \infty$ (continuity/boundedness).
(c) If $x \geq 0$, that is, if $x_n \geq 0$ for all $n \in \mathbb{N}$, then $l_B(x) \geq 0$ (positivity).
(d) $l_B(\sigma(x)) = l_B(x)$, where $\sigma : l^\infty \to l^\infty$ is the (left) shift map defined by $(\sigma(x))_n = x_{n+1}$ for all $n \in \mathbb{N}$ (shift-invariance).
(e) If x is a convergent sequence, then $l_B(x) = \lim_{n\to\infty} x_n$.

It follows from properties (a), (c) and (e) that a Banach limit also satisfies:
(f) $\liminf_{n\to\infty} x_n \leq l_B(x) \leq \limsup_{n\to\infty} x_n$.
(g) If $x \leq y$, that is, if $x_n \leq y_n$ for all $n \in \mathbb{N}$, then $l_B(x) \leq l_B(y)$.

As already announced, the main result of this section is the following.

Theorem 10.4.5. *Let $(X, \mathcal{A}, m)$ be a probability space and $T : X \to X$ a Martens map with Martens cover $\{X_j\}_{j=0}^{\infty}$ and for which m is quasi-T-invariant. Then there exists a σ-finite T-invariant measure μ equivalent to m on X. In addition, $0 < \mu(X_j) < \infty$, $\forall j \geq 0$.*

A measure μ with the above properties can be constructed as follows. Let $l_B : l^\infty \to \mathbb{R}$ be a Banach limit and let $Y_j := X_j \setminus \bigcup_{i<j} X_i$ for every $j \geq 0$. For each $A \in \mathcal{A}$, set

$$m_n(A) := \frac{\sum_{k=0}^{n} m(T^{-k}(A))}{\sum_{k=0}^{n} m(T^{-k}(X_0))}. \tag{10.21}$$

If $A \in \mathcal{A}$ and $A \subseteq Y_j$ for some $j \geq 0$, then $(m_n(A))_{n=1}^{\infty} \in l^\infty$ and set

$$\mu(A) := l_B((m_n(A))_{n=1}^{\infty}). \tag{10.22}$$

For a general $A \in \mathcal{A}$, set

$$\mu(A) := \sum_{j=0}^{\infty} \mu(A \cap Y_j).$$

If $(m_n(A))_{n=1}^{\infty} \in l^{\infty}$ for some $A \in \mathcal{A}$, then

$$\mu(A) = l_B\big((m_n(A))_{n=1}^{\infty}\big) - \lim_{l\to\infty} l_B\Bigg(\bigg(m_n\Big(A \cap \bigcup_{j=l}^{\infty} Y_j\Big)\bigg)_{n=0}^{\infty}\Bigg). \tag{10.23}$$

In particular, if $A \in \mathcal{A}$ is contained in a finite union of sets X_j, $j \geq 0$, then

$$\mu(A) = l_B\big((m_n(A))_{n=1}^{\infty}\big).$$

Finally, if T is ergodic and conservative with respect to m, then μ is unique up to a positive multiplicative constant and T is ergodic and conservative with respect to μ.

In order to prove Theorem 10.4.5, we need several lemmas.

Lemma 10.4.6. *Let $(Z, \mathcal{F})$ be a measurable space such that:*
(a) $Z = \bigcup_{j=0}^{\infty} Z_j$ *for some mutually disjoint sets $Z_j \in \mathcal{F}$; and*
(b) ν_j *is a finite measure on Z_j for each $j \geq 0$.*

Then the set function $\nu : \mathcal{F} \to [0, \infty]$ defined by

$$\nu(F) := \sum_{j=0}^{\infty} \nu_j(F \cap Z_j)$$

is a σ-finite measure on Z.

Proof. Clearly, $\nu(\emptyset) = 0$. Let $F \in \mathcal{F}$ and $\{F_n\}_{n=1}^{\infty}$ a partition of F into sets in $\mathcal{F}$. Then

$$\nu(F) = \sum_{j=0}^{\infty} \nu_j(F \cap Z_j) = \sum_{j=0}^{\infty} \nu_j\left(\bigcup_{n=1}^{\infty}(F_n \cap Z_j)\right) = \sum_{j=0}^{\infty}\sum_{n=1}^{\infty} \nu_j(F_n \cap Z_j) = \sum_{n=1}^{\infty}\sum_{j=0}^{\infty} \nu_j(F_n \cap Z_j) = \sum_{n=1}^{\infty} \nu(F_n),$$

where the order of summation could be changed since all terms involved are non-negative. Thus ν is a measure. Moreover, by definition, $Z = \bigcup_{j=0}^{\infty} Z_j$ and $\nu(Z_j) = \nu_j(Z_j) < \infty$ for all $j \geq 0$. Therefore, ν is σ-finite. □

From this point on, all lemmas rely on the same main hypotheses as Theorem 10.4.5.

Lemma 10.4.7. *For all $n, j \geq 0$ and all $A, B \in \mathcal{A}$ with $A \cup B \subseteq X_j$, we have*

$$m_n(A) m(B) \leq K_j\, m(A)\, m_n(B).$$

Proof. This follows directly from the definition of m_n and condition (d) of Definition 10.4.3. □

Lemma 10.4.8. *For every $j \geq 0$, we have $(m_n(X_j))_{n=1}^{\infty} \in l^{\infty}$ and $\mu(Y_j) \leq \mu(X_j) < \infty$.*

Proof. Fix $j \geq 0$. In virtue of condition (c) of Definition 10.4.3, there exists $q \geq 0$ such that $m(X_j \cap T^{-q}(X_0)) > 0$. By Lemma 10.4.7 and the definition of m_n, for all $n \geq 0$ we have that

$$\begin{aligned} m_n(Y_j) \leq m_n(X_j) &\leq K_j \frac{m(X_j)}{m(X_j \cap T^{-q}(X_0))} m_n(X_j \cap T^{-q}(X_0)) \\ &\leq K_j \frac{m(X_j)}{m(X_j \cap T^{-q}(X_0))} m_n(T^{-q}(X_0)) \\ &\leq K_j \frac{m(X_j)}{m(X_j \cap T^{-q}(X_0))} \frac{\sum_{k=0}^{n+q} m(T^{-k}(X_0))}{\sum_{k=0}^{n} m(T^{-k}(X_0))} \\ &= K_j \frac{m(X_j)}{m(X_j \cap T^{-q}(X_0))} \left[1 + \frac{\sum_{k=n+1}^{n+q} m(T^{-k}(X_0))}{\sum_{k=0}^{n} m(T^{-k}(X_0))}\right] \\ &\leq K_j \frac{m(X_j)}{m(X_j \cap T^{-q}(X_0))} \left[1 + \frac{q}{m(X_0)}\right]. \end{aligned} \tag{10.24}$$

Consequently, $(m_n(X_j))_{n=1}^{\infty} \in l^{\infty}$ and properties (g) and (e) of a Banach limit yield that

$$\mu(Y_j) \leq K_j \frac{m(X_j)}{m(X_j \cap T^{-q}(X_0))} \left[1 + \frac{q}{m(X_0)}\right] < \infty.$$

Since $X_j = \bigcup_{i=0}^{j} Y_i$ and the Y_i's are mutually disjoint, we deduce that

$$\mu(Y_j) \leq \sum_{i=0}^{j} \mu(X_j \cap Y_i) = \sum_{i=0}^{\infty} \mu(X_j \cap Y_i) =: \mu(X_j) \leq \sum_{i=0}^{j} \mu(Y_i) < \infty.$$ □

For every $j \geq 0$, set $\mu_j := \mu|_{Y_j}$.

Lemma 10.4.9. *For every $j \geq 0$ such that $\mu(Y_j) > 0$ and for every measurable set $A \subseteq Y_j$, we have*

$$K_j^{-1} \frac{\mu(Y_j)}{m(Y_j)} m(A) \leq \mu_j(A) \leq K_j \frac{\mu(Y_j)}{m(Y_j)} m(A).$$

Proof. This follows from the definition of μ, and by setting $B = Y_j$ in Lemma 10.4.7 and using properties (a) and (g) of a Banach limit. □

Lemma 10.4.10. *For each $j \geq 0$, μ_j is a finite measure on Y_j.*

Proof. Let $j \geq 0$. If $\mu_j(Y_j) = 0$, then the result is trivial. So assume that $\mu_j(Y_j) > 0$. Let $A \subseteq Y_j$ be a measurable set and $(A_k)_{k=1}^{\infty}$ a countable measurable partition of A. Using termwise operations on sequences, for every $l \in \mathbb{N}$ we have

$$\left(\sum_{k=1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty} - \sum_{k=1}^{l}(m_n(A_k))_{n=1}^{\infty} = \left(\sum_{k=1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty} - \left(\sum_{k=1}^{l} m_n(A_k)\right)_{n=1}^{\infty}$$
$$= \left(\sum_{k=l+1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty}.$$

It therefore follows from Lemma 10.4.7 (with $A = A_k$ and $B = Y_j$) that

$$\begin{aligned}\left\|\left(\sum_{k=1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty} - \sum_{k=1}^{l}(m_n(A_k))_{n=1}^{\infty}\right\|_{\infty} &= \left\|\left(\sum_{k=l+1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty}\right\|_{\infty} \\ &\leq \left\|\frac{K_j}{m(Y_j)}\left(m_n(Y_j)\sum_{k=l+1}^{\infty} m(A_k)\right)_{n=1}^{\infty}\right\|_{\infty} \\ &= \frac{K_j}{m(Y_j)}\left\|\left(m_n(Y_j)\sum_{k=l+1}^{\infty} m(A_k)\right)_{n=1}^{\infty}\right\|_{\infty}.\end{aligned}$$

Since $(m_n(Y_j))_{n=1}^{\infty} \in l^{\infty}$ by Lemma 10.4.8 and since $\lim_{l\to\infty}\sum_{k=l+1}^{\infty} m(A_k) = 0$, we conclude that

$$\lim_{l\to\infty}\left\|\left(\sum_{k=1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty} - \sum_{k=1}^{l}(m_n(A_k))_{n=1}^{\infty}\right\|_{\infty} = 0.$$

This means that

$$\left(\sum_{k=1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty} = \sum_{k=1}^{\infty}(m_n(A_k))_{n=1}^{\infty} \quad \text{in } l^{\infty}.$$

Hence, using the continuity of the Banach limit $l_B : l^{\infty} \to \mathbb{R}$, we get

$$\begin{aligned}\mu(A) &= l_B\big((m_n(A))_{n=1}^{\infty}\big) = l_B\left(\left(m_n\Big(\bigcup_{k=1}^{\infty} A_k\Big)\right)_{n=1}^{\infty}\right) = l_B\left(\left(\sum_{k=1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty}\right) \\ &= \sum_{k=1}^{\infty} l_B\big((m_n(A_k))_{n=1}^{\infty}\big) = \sum_{k=1}^{\infty}\mu(A_k).\end{aligned}$$

So μ_j is countably additive. Also, $\mu_j(\emptyset) = 0$. Thus μ_j is a measure. By Lemma 10.4.8, this measure μ_j is finite. □

Combining Lemmas 10.4.6, 10.4.8, 10.4.9, and 10.4.10, and condition (b) of Definition 10.4.3, we obtain the following.

Lemma 10.4.11. *The set function μ is a σ-finite measure on X equivalent to m. Moreover, $\mu(Y_j) \le \mu(X_j) < \infty$ and $\mu(X_j) > 0$ for all $j \ge 0$.*

Lemma 10.4.12. *Formula* (10.23) *holds.*

Proof. Fix $A \in \mathcal{A}$ such that $(m_n(A))_{n=1}^\infty \in l^\infty$. For every $l \in \mathbb{N}$, we then have

$$\begin{aligned} l_B\big((m_n(A))_{n=1}^\infty\big) &= l_B\left(\sum_{j=0}^{l}(m_n(A\cap Y_j))_{n=1}^\infty\right) + l_B\left(\left(m_n\Big(\bigcup_{j=l+1}^{\infty} A\cap Y_j\Big)\right)_{n=1}^\infty\right) \\ &= \sum_{j=0}^{l} l_B\big((m_n(A\cap Y_j))_{n=1}^\infty\big) + l_B\left(\left(m_n\Big(A\cap \bigcup_{j=l+1}^{\infty} Y_j\Big)\right)_{n=1}^\infty\right). \end{aligned}$$

Letting $l \to \infty$, we deduce that

$$\begin{aligned} l_B\big((m_n(A))_{n=1}^\infty\big) &= \sum_{j=0}^{\infty} l_B\big((m_n(A\cap Y_j))_{n=1}^\infty\big) + \lim_{l\to\infty} l_B\left(\left(m_n\Big(A\cap \bigcup_{j=l+1}^{\infty} Y_j\Big)\right)_{n=1}^\infty\right) \\ &= \sum_{j=0}^{\infty} \mu(A\cap Y_j) + \lim_{l\to\infty} l_B\left(\left(m_n\Big(A\cap \bigcup_{j=l}^{\infty} Y_j\Big)\right)_{n=1}^\infty\right) \\ &= \mu(A) + \lim_{l\to\infty} l_B\left(\left(m_n\Big(A\cap \bigcup_{j=l}^{\infty} Y_j\Big)\right)_{n=1}^\infty\right). \end{aligned}$$

This establishes formula (10.23). In particular, if $A \subseteq \bigcup_{j=0}^{k} X_j$ for some $k \in \mathbb{N}$, then $A \cap \bigcup_{j=l}^{\infty} Y_j \subseteq (\bigcup_{j=0}^{k} X_j) \cap (X\setminus \bigcup_{i<l} X_i) = \emptyset$ for all $l > k$. In that case, the equation above reduces to

$$l_B\big((m_n(A))_{n=1}^\infty\big) = \mu(A). \qquad \square$$

Lemma 10.4.13. *The σ-finite measure μ is T-invariant.*

Proof. Let $i \ge 0$ be such that $m(Y_i) > 0$. Fix a measurable set $A \subset Y_i$. By definition, $\mu(A) = l_B((m_n(A))_{n=1}^\infty)$. Furthermore, for all $n \ge 0$ notice that

$$\big|m_n(T^{-1}(A)) - m_n(A)\big| = \frac{\big|m(T^{-(n+1)}(A)) - m(A)\big|}{\sum_{k=0}^{n} m(T^{-k}(X_0))} \le \frac{1}{\sum_{k=0}^{n} m(T^{-k}(X_0))}.$$

Thus $\big(m_n(T^{-1}(A))\big)_{n=1}^\infty \in l^\infty$ because $(m_n(A))_{n=1}^\infty \in l^\infty$. Moreover, by condition (e) of Definition 10.4.3, it follows from the above and properties (a), (e), and (g) of a Banach limit that $l_B\big((m_n(T^{-1}(A)))_{n=1}^\infty\big) = l_B\big((m_n(A))_{n=1}^\infty\big) = \mu(A)$.

Keep A a measurable subset of Y_i. Fix $l \in \mathbb{N}$. We then have

$$\begin{aligned} m_n\left(T^{-1}(A)\cap \bigcup_{j=l}^{\infty} Y_j\right) &= \frac{\sum_{k=0}^{n} m\big(T^{-k}(T^{-1}(A)\cap \bigcup_{j=l}^{\infty} Y_j)\big)}{\sum_{k=0}^{n} m(T^{-k}(X_0))} \\ &\le \frac{\sum_{k=0}^{n} m\big(T^{-(k+1)}(A\cap T(\bigcup_{j=l}^{\infty} Y_j))\big)}{\sum_{k=0}^{n} m(T^{-k}(X_0))} \end{aligned}$$

$$\le m_{n+1}\Big(A \cap T\Big(\bigcup_{j=l}^{\infty} Y_j\Big)\Big) \cdot \frac{\sum_{k=0}^{n+1} m(T^{-k}(X_0))}{\sum_{k=0}^{n} m(T^{-k}(X_0))}$$
$$\le K_i \frac{m_{n+1}(Y_i)}{m(Y_i)} \cdot m\Big(A \cap T\Big(\bigcup_{j=l}^{\infty} Y_j\Big)\Big) \cdot \frac{\sum_{k=0}^{n+1} m(T^{-k}(X_0))}{\sum_{k=0}^{n} m(T^{-k}(X_0))},$$

where the last inequality sign holds by Lemma 10.4.7 since $A \subseteq Y_i$. When $n \to \infty$, the last quotient on the right-hand side approaches 1. Therefore,

$$0 \le l_B\Big(\Big(m_n\Big(T^{-1}(A) \cap \bigcup_{j=l}^{\infty} Y_j\Big)\Big)_{n=1}^{\infty}\Big) \le K_i \frac{\mu(Y_i)}{m(Y_i)} m\Big(T\Big(\bigcup_{j=l}^{\infty} Y_j\Big)\Big).$$

Hence, by virtue of condition (g) of Definition 10.4.3,

$$0 \le \lim_{l\to\infty} l_B\Big(\Big(m_n\Big(T^{-1}(A) \cap \bigcup_{j=l}^{\infty} Y_j\Big)\Big)_{n=1}^{\infty}\Big) \le K_i \frac{\mu(Y_i)}{m(Y_i)} \lim_{l\to\infty} m\Big(T\Big(\bigcup_{j=l}^{\infty} Y_j\Big)\Big) = 0.$$

So

$$\lim_{l\to\infty} l_B\Big(\Big(m_n\Big(T^{-1}(A) \cap \bigcup_{j=l}^{\infty} Y_j\Big)\Big)_{n=1}^{\infty}\Big) = 0.$$

It thus follows from Lemma 10.4.12 that

$$\mu(T^{-1}(A)) = l_B((m_n(T^{-1}(A)))_{n=1}^{\infty}) = l_B((m_n(A))_{n=1}^{\infty}) = \mu(A).$$

For an arbitrary $A \in \mathcal{A}$, write $A = \bigcup_{j=0}^{\infty}(A \cap Y_j)$ and observe that

$$\mu(T^{-1}(A)) = \mu\Big(\bigcup_{j=0}^{\infty} T^{-1}(A \cap Y_j)\Big) = \sum_{j=0}^{\infty} \mu(T^{-1}(A \cap Y_j)) = \sum_{j=0}^{\infty} \mu(A \cap Y_j) = \mu(A). \qquad \square$$

Proof of Theorem 10.4.5. Combining Lemmas 10.4.8, 10.4.11, 10.4.12, and 10.4.13, with Theorems 10.4.2 and 10.1.11, we obtain Theorem 10.4.5. $\square$

Remark 10.4.14. In the course of the proof of Theorem 10.4.5, we have shown that

$$0 < \inf\{m_n(A) : n \in \mathbb{N}\} \le \sup\{m_n(A) : n \in \mathbb{N}\} < \infty$$

for all $j \ge 0$ and all measurable sets $A \subseteq X_j$ such that $m(A) > 0$.

10.5 Exercises

Exercise 10.5.1. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation for which there is a completely T-invariant set $A \in \mathcal{A}\backslash\{\emptyset, X\}$. Assume also that there exist a $T|_A$-invariant probability measure ν on $(A, \mathcal{A}|_A)$ and a $T|_{X\backslash A}$-invariant probability measure κ on $(X\backslash A, \mathcal{A}|_{X\backslash A})$. Set $\mu(B) = \nu(B \cap A) + \kappa(B\backslash A)$ for all $B \in \mathcal{A}$.

(a) Prove that μ is a T-invariant measure on $(X, \mathcal{A})$.
(b) Show that $0 < \mu(A) < \infty$ and $\mu(A \backslash A_\infty) = 0$.
(c) Deduce that $\mu_A = \nu$ and is T_A-invariant.
(d) Show that $\mu(X \backslash \bigcup_{k=0}^{\infty} T^{-k}(A)) \neq 0$.
(e) Let $\widetilde{\mu}$ be the measure induced by $\nu = \mu_A$.
  (1) Prove that $\mu(A) = 1 = \widetilde{\mu}(A)$.
  (2) Prove that $\mu(X \backslash A) = 1$ whereas $\widetilde{\mu}(X \backslash A) = 0$.
  (3) Conclude that there is no $c \in \mathbb{R}$ such that $\widetilde{\mu} = c\mu$. In particular, $\widetilde{\mu} \neq \mu$.

Exercise 10.5.2. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a T-invariant measure. Fix $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$ and $\mu(A \backslash A_\infty) = 0$. So, from this point on, A will be identified with $A \cap A_\infty$.

The first return time function $\tau_A : A \to \mathbb{N}$ was defined in (10.1) while the first return map $T_A : A \to A$ was introduced in (10.2).

Similarly, for every $n \in \mathbb{N}$ the nth return time function $\tau_A^n : A \to \mathbb{N}$ is defined by

$$\tau_A^n(x) := \min\{k \in \mathbb{N} : \#\{1 \leq j \leq k : T^j(x) \in A\} = n\}.$$

The nth return map $T_A^n : A \to A$ is subsequently defined as

$$T_A^n(x) = T^{\tau_A^n(x)}(x).$$

Finally, let $\varphi : X \to \mathbb{R}$ be a measurable function. The function $\varphi_A : A \to \mathbb{R}$ was defined in (10.11). The nth *Birkhoff sum* of φ_A under T_A at a point $x \in A$ is denoted by $S_n^{T_A}\varphi_A : A \to \mathbb{R}$ and is naturally given by

$$S_n^{T_A}\varphi_A(x) = \sum_{i=0}^{n-1} \varphi_A(T_A^i(x)).$$

(a) Show that $T_A^n = T_A \circ T_A \ldots \circ T_A$, with n copies of T_A in the composition. In other words, show that the nth return map T_A^n is the usual n-time composition of the first return map T_A.
(b) Prove that $\tau_A^n(x) = \sum_{i=0}^{n-1} \tau_A(T_A^i(x)) = S_n^{T_A}\tau_A(x)$. In other terms, the nth return time is the sum of the first return times of the first n iterates of x that fall into A.
(c) Deduce that $\tau_A^{n+1}(x) - \tau_A^n(x) = \tau_A(T_A^n(x))$.
(d) Show that $S_{\tau_A^n(x)}\varphi(x) = \sum_{k=0}^{n-1} S_{\tau_A(T_A^k(x))}\varphi(T_A^k(x)) = S_n^{T_A}\varphi_A(x)$.
(e) Given $x \in A$ and $k \in \mathbb{N}$, let $n(x)$ be the largest integer $n \geq 0$ such that $\tau_A^n(x) \leq k$. In other words, $n(x)$ is the number of times that the iterates of x visit A by time k, and $\tau_A^{n(x)}(x)$ is the last time at which an iterate of x falls into A prior to or at time k. Demonstrate that

$$S_k\varphi(x) = S_{n(x)}^{T_A}\varphi_A(x) + S_{\Delta k(x)}\varphi(T_A^{n(x)}(x)),$$

where $\Delta k(x) := k - \tau_A^{n(x)}(x) \geq 0$.

(f) Show that $|\varphi_A| \le |\varphi|_A$.
(g) Prove that $S_{\Delta k(x)}\varphi(T_A^{n(x)}(x)) \le |\varphi|_A(T_A^{n(x)}(x))$.

Exercise 10.5.3. In this exercise, you will give a proof of Proposition 10.3.5. You will first establish the proposition under the additional assumption that $a(0) = 0$.
(a) Prove that $a(x+y) \le a(x) + a(y)$ for all $x, y \in [0, \infty)$.
(b) Show that

$$\lim_{M\to\infty} a\big(|f|\mathbb{1}_{\{|f|\ge M\}}(x)\big) = 0 \quad \text{for } \mu\text{-a.e. } x \in X.$$

(c) Deduce that

$$\lim_{M\to\infty} \int_X a\big(|f|\mathbb{1}_{\{|f|\ge M\}}\big)\, d\mu = 0.$$

(d) Fix $\varepsilon > 0$. From (c), identify two nonnegative functions $g, h \in L^1(\mu)$ such that

$$|f| = g + h, \quad \sup_{x\in X} g(x) < \infty, \quad \text{and} \quad \int_X a(h)\, d\mu < \varepsilon.$$

(e) Using (a) and (d), show that

$$\limsup_{n\to\infty} \frac{a\big(|S_n f(x)|\big)}{n} \le \int_X a(h)\, d\mu.$$

(f) Conclude that

$$\lim_{n\to\infty} \frac{a\big(|S_n f(x)|\big)}{n} = 0 \quad \text{for } \mu\text{-a.e. } x \in X.$$

You will now establish the proposition without the assumption that $a(0) = 0$. Choose $m > 0$ such that $a(m) = \alpha m$ for some $0 < \alpha < 1$, and define

$$\tilde{a}(x) = \begin{cases} \alpha x & \text{if } 0 \le x \le m \\ a(x) & \text{if } x \ge m. \end{cases}$$

(g) Show that $\tilde{a}$ is continuous, strictly increasing, $\tilde{a} \equiv a$ on $[m, \infty)$, $\frac{\tilde{a}(x)}{x} \searrow 0$ as $x \nearrow \infty$, and $\tilde{a}(0) = 0$.
(h) Deduce that

$$\lim_{n\to\infty} \frac{\tilde{a}\big(|S_n f(x)|\big)}{n} = 0 \quad \text{for } \mu\text{-a.e. } x \in X.$$

(i) Conclude that

$$\lim_{n\to\infty} \frac{a\big(|S_n f(x)|\big)}{n} = 0 \quad \text{for } \mu\text{-a.e. } x \in X.$$

Exercise 10.5.4. Let $X = \overline{\mathbb{R}}$, λ be the Lebesgue measure on $\overline{\mathbb{R}}$, and $T : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ be Boole's transformation defined by $T(x) = x - \frac{1}{x}$.

(a) Prove that λ is T-invariant.

(b) Show that T is conservative.

Exercise 10.5.5. For each $i = 1, 2$, let $T_i : (X_i, \mathcal{A}_i, \mu_i) \to (X_i, \mathcal{A}_i, \mu_i)$ be a measurable transformation of a probability space $(X_i, \mathcal{A}_i, \mu_i)$ such that μ_i is quasi-T_i-invariant. The Cartesian product transformation $T = T_1 \times T_2$ is defined on the product space $(X, \mathcal{A}, \mu) := (X_1 \times X_2, \mathcal{A}_1 \times \mathcal{A}_2, m_1 \times m_2)$ by $T(x_1, x_2) = (T_1(x_1), T_2(x_2))$ (cf. Example 8.1.13). Prove that if T_1 is measure-preserving and T_2 is conservative, then $T_1 \times T_2$ is conservative.

Exercise 10.5.6. Generalize Exercise 8.5.47 to every measure space.

Exercise 10.5.7. Let $S : (X_S, \mathcal{A}_S, \mu_S) \to (X_S, \mathcal{A}_S, \mu_S)$ be a measurable transformation of a σ-finite measure space $(X_S, \mathcal{A}_S, \mu_S)$. Let $\varphi : X_S \to \mathbb{N}$ be a measurable function. The *Kakutani tower* over T with height function φ is the transformation T of the σ-finite measure space $(X_T, \mathcal{A}_T, \mu_T)$ defined as follows:

$$X_T = \{(x, n) : x \in X_S \text{ and } n \leq \varphi(x)\},$$
$$\mathcal{A}_T = \sigma\big(\{A \times \{n\} : n \in \mathbb{N} \text{ and } A = A_S \cap \varphi^{-1}([n, \infty)) \text{ for some } A_S \in \mathcal{A}_S\}\big),$$
$$\mu_T(A \times \{n\}) = \mu_S(A),$$

and

$$T(x, n) = \begin{cases} (S(x), \varphi(S(x))) & \text{if } n = 1, \\ (x, n-1) & \text{if } n \geq 2. \end{cases}$$

Prove the following statements:

(a) If μ_S is quasi-S-invariant, then μ_T is quasi-T-invariant.

(b) If S is conservative, then so is T.

(c) $T_{X_S \times \{1\}}(x, 1) \equiv (S(x), 1)$ and $\varphi_{X_S \times \{1\}}(x, 1) \equiv \varphi(x)$.

(d) If μ_S is S-invariant, then μ_T is T-invariant.

(e) If S is ergodic, then so is T.

11 Topological pressure

In the forthcoming three chapters (the third one being part of the second volume), we introduce and extensively deal with the fundamental concepts and results of thermodynamic formalism, including topological pressure, the so-called variational principle, equilibrium states, and Gibbs states.

Thermodynamic formalism originated from the works of David Ruelle in the late 1960s. Ruelle's motivation came from statistical mechanics, particularly glass lattices. The foundations, classical concepts, and theorems of thermodynamic formalism were developed throughout the 1970s in the early works of Ruelle [59, 60], Rufus Bowen [10], Peter Walters [75], and Yakov Sinai [68]. More recent and modern expositions can be found in [57, 61, 76], among others. Also worthy of mention is Michal Misiurewicz's paper [49], where an elegant, short, and simple proof of the variational principle was provided. This is the proof we shall reproduce in Chapter 12.

In Chapter 11, we define and investigate the properties of topological pressure. Like topological entropy, this is a topological concept and a topological conjugacy invariant. We further give Bowen's characterization of pressure in terms of separated and spanning sets, which however requires a metric.

In Chapter 12, we relate topological pressure with Kolmogorov–Sinai metric entropy by proving the variational principle, the very cornerstone of thermodynamic formalism. This principle naturally leads to the concepts of equilibrium states and measures of maximal entropy. We deal with those at length in that chapter, particularly through the problem of existence of equilibrium states. Among others, we will show that under a continuous potential every expansive system admits an equilibrium state.

Whereas in Chapters 11 and 12 we consider general topological dynamical systems, in Chapter 13 we will restrict our attention to transitive open distance expanding maps. We will introduce therein the concept of Gibbs measures for such maps. We will also prove their existence and uniqueness for Hölder continuous potentials. We will further demonstrate that they coincide with equilibrium states for these potentials, concomitantly establishing the uniqueness of equilibrium states.

Gibbs states are measures with particularly fine and transparent stochastic properties, such as the central limit theorem, the law of the iterated logarithm, and exponential decay of correlations. They are used to describe long-term unstable behaviors of typical orbits of a given dynamical system having Gibbs states. Apart from such direct application to dynamical systems, thermodynamic formalism was used to provided a full account of SRB (Sinai–Ruelle–Bowen) measures for Axiom A diffeomorphisms and flows, and many other dynamical systems. Via Bowen's formula, thermodynamic formalism is also an indispensable tool for studying the fractal geometry of nonlinear smooth dynamical systems, particularly conformal and holomorphic ones. This is at the heart of Chapter 16 and a guiding theme in all subsequent chapters.

The main tool for dealing with Gibbs states is the transfer, also frequently called Perron–Frobenius, operator. We prove many of its functional analytic properties;

https://doi.org/10.1515/9783110702682-011

among them its almost periodicity if acting on the Banach space of continuous functions and its quasi–compactness if acting on the Banach space of Hölder continuous functions. In addition to quasi-compactness, we show that this operator has only one (and real) eigenvalue of maximal modulus and this eigenvalue is simple. The corresponding eigenfunction turns out to be the Radon–Nikodym derivative of the invariant Gibbs state with respect to the eigenmeasure of the dual transfer operator.

11.1 Definition of topological pressure via open covers

Recall that a topological dynamical system $T : X \to X$ is a self-transformation T of a compact metrizable space X. Let $\varphi : X \to \mathbb{R}$ be a real-valued continuous function. In the context of topological pressure (for historical, physical reasons), such a function is usually referred to as a *potential*.

The topological pressure of a potential is defined in two stages. This may seem surprising since topological entropy was defined in three stages in Chapter 7. However, the main reason for defining topological entropy in three stages was so as to later mirror it in the definition of measure-theoretic entropy in Chapter 9. Indeed, the first stage might just as well have been omitted and we would then have proceeded immediately to the second stage by defining $Z_n(\mathcal{U})$ directly and deriving its properties without relying upon the fact that $Z_n(\mathcal{U}) = Z_1(\mathcal{U}^n)$. The first stage for topological entropy proves to be useless when defining topological pressure since, as we will shortly discover, $Z_n(\varphi,\mathcal{U}) \neq Z_1(\varphi,\mathcal{U}^n)$ in general.

11.1.1 First stage: pressure of a potential relative to an open cover

Let us first recall the notion of Birkhoff (or ergodic) sum (cf. Definition 8.2.10). The nth *Birkhoff sum* of a potential φ at a point $x \in X$ is given by

$$S_n\varphi(x) := \sum_{j=0}^{n-1} \varphi(T^j(x)).$$

This is the sum of the values of the potential φ at the first n iterates of x under T.

Definition 11.1.1. For every $Y \subseteq X$ and $n \in \mathbb{N}$, define

$$\overline{S}_n\varphi(Y) := \sup_{y\in Y} S_n\varphi(y) \quad \text{and} \quad \underline{S}_n\varphi(Y) := \inf_{y\in Y} S_n\varphi(y).$$

Now, let $\mathcal{U}$ be an open cover of X. The minimum number $Z_n(\mathcal{U})$ of elements of $\mathcal{U}^n$ required to cover X (cf. Definition 7.2.6) generalises to the real numbers $Z_n(\varphi,\mathcal{U})$ and $z_n(\varphi,\mathcal{U})$ as follows.

Definition 11.1.2. Let $T : X \to X$ be a topological dynamical system and let $\varphi : X \to \mathbb{R}$ be a potential. Let $\mathcal{U}$ be an open cover of X. For each $n \in \mathbb{N}$, define the nth level functions (sometimes called *partition functions*) of $\mathcal{U}$ with respect to the potential φ by

$$Z_n(\varphi,\mathcal{U}) := \inf\left\{\sum_{V\in\mathcal{V}} e^{\overline{S}_n\varphi(V)} : \mathcal{V} \text{ is a subcover of } \mathcal{U}^n\right\}$$

and

$$z_n(\varphi,\mathcal{U}) := \inf\left\{\sum_{V\in\mathcal{V}} e^{\underline{S}_n\varphi(V)} : \mathcal{V} \text{ is a subcover of } \mathcal{U}^n\right\}.$$

Remark 11.1.3.
(a) It is sufficient to take the infimum over all **finite** subcovers since the exponential function takes only positive values and every subcover has itself a finite subcover. However, this infimum may not be achieved if $\mathcal{U}$ is infinite.
(b) In general, $Z_n(\varphi,\mathcal{U}) \neq Z_1(\varphi,\mathcal{U}^n)$ and $z_n(\varphi,\mathcal{U}) \neq z_1(\varphi,\mathcal{U}^n)$.
(c) If $\varphi \equiv 0$, then $Z_n(0,\mathcal{U}) = z_n(0,\mathcal{U}) = Z_n(\mathcal{U})$ for all $n \in \mathbb{N}$ and any open cover $\mathcal{U}$ of X.
(d) If $\varphi \equiv c$ for some $c \in \mathbb{R}$, then $Z_n(c,\mathcal{U}) = z_n(c,\mathcal{U}) = e^{nc}Z_n(\mathcal{U})$ for all $n \in \mathbb{N}$ and every open cover $\mathcal{U}$ of X.
(e) For all open covers $\mathcal{U}$ of X and all $n \in \mathbb{N}$, we have

$$e^{n\inf\varphi}Z_n(\mathcal{U}) \le Z_n(\varphi,\mathcal{U}) \le e^{n\sup\varphi}Z_n(\mathcal{U})$$

and

$$e^{n\inf\varphi}Z_n(\mathcal{U}) \le z_n(\varphi,\mathcal{U}) \le e^{n\sup\varphi}Z_n(\mathcal{U}).$$

We have seen in Chapter 7 that the functions $Z_n(\cdot)$, $n \in \mathbb{N}$, behave well with respect to all cover operations. In particular, it was observed in Lemma 7.2.8 that they respect the refinement relation, that is, if $\mathcal{U} \prec \mathcal{V}$ then $Z_n(\mathcal{U}) \le Z_n(\mathcal{V})$ for every $n \in \mathbb{N}$. This is not necessarily true for the partition functions $Z_n(\varphi,\cdot)$, $n \in \mathbb{N}$. The corresponding inequality is more intricate. It involves the concept of oscillation.

Definition 11.1.4. The *oscillation* of φ with respect to an open cover $\mathcal{U}$ is defined to be

$$\operatorname{osc}(\varphi,\mathcal{U}) := \sup_{U\in\mathcal{U}} \sup_{x,y\in U} \left|\varphi(y) - \varphi(x)\right|.$$

Note that $\operatorname{osc}(\varphi,\cdot) \le 2\|\varphi\|_\infty$. Also, $\operatorname{osc}(c,\cdot) = 0$ for all $c \in \mathbb{R}$.

Lemma 11.1.5. *For every $n \in \mathbb{N}$ and every open cover $\mathcal{U}$ of X,*

$$\operatorname{osc}(S_n\varphi,\mathcal{U}^n) \le n\operatorname{osc}(\varphi,\mathcal{U}).$$

Proof. Let $V := U_0 \cap \cdots \cap T^{-(n-1)}(U_{n-1}) \in \mathcal{U}^n$ and $x, y \in V$. For each $0 \leq j < n$, we have that $T^j(x), T^j(y) \in U_j \in \mathcal{U}$. Hence, for all $0 \leq j < n$,

$$\left|\varphi(T^j(x)) - \varphi(T^j(y))\right| \leq \operatorname{osc}(\varphi, \mathcal{U}).$$

Therefore,

$$\left|S_n\varphi(x) - S_n\varphi(y)\right| \leq \sum_{j=0}^{n-1}\left|\varphi(T^j(x)) - \varphi(T^j(y))\right| \leq n\operatorname{osc}(\varphi, \mathcal{U}).$$

Since this is true for all $x, y \in V$ and all $V \in \mathcal{U}^n$, the result follows. □

We now look at the relationship between the Z_n's and the z_n's.

Lemma 11.1.6. *For all $n \in \mathbb{N}$ and all open covers U of X, the following inequalities hold:*

$$z_n(\varphi, \mathcal{U}) \leq Z_n(\varphi, \mathcal{U}) \leq e^{n\operatorname{osc}(\varphi, \mathcal{U})} z_n(\varphi, \mathcal{U}).$$

Proof. The left inequality is obvious. To ascertain the right one, let $\mathcal{W}$ be a subcover of $\mathcal{U}^n$. Then

$$\begin{aligned}\sum_{W\in\mathcal{W}} e^{\overline{S}_n\varphi(W)} &\leq \exp\Big(\sup_{W\in\mathcal{W}}\big[\,\overline{S}_n\varphi(W) - \underline{S}_n\varphi(W)\,\big]\Big)\sum_{W\in\mathcal{W}} e^{\underline{S}_n\varphi(W)} \\ &\leq e^{\operatorname{osc}(S_n\varphi, \mathcal{U}^n)}\sum_{W\in\mathcal{W}} e^{\underline{S}_n\varphi(W)} \leq e^{n\operatorname{osc}(\varphi, \mathcal{U})}\sum_{W\in\mathcal{W}} e^{\underline{S}_n\varphi(W)}.\end{aligned}$$

Taking the infimum over all subcovers of $\mathcal{U}^n$ on both sides results in the right inequality. □

In the next few results, we will see that the Z_n's and the z_n's have distinct properties.

Lemma 11.1.7. *If $\mathcal{U} \prec \mathcal{V}$, then for all $n \in \mathbb{N}$ we have that*

$$Z_n(\varphi, \mathcal{U})e^{-n\operatorname{osc}(\varphi, \mathcal{U})} \leq Z_n(\varphi, \mathcal{V}) \quad \textit{while} \quad z_n(\varphi, \mathcal{U}) \leq z_n(\varphi, \mathcal{V}).$$

Proof. Fix $n \in \mathbb{N}$. Let $i : \mathcal{V} \to \mathcal{U}$ be a map such that $V \subseteq i(V)$ for all $V \in \mathcal{V}$. The map i induces a map $i_n : \mathcal{V}^n \to \mathcal{U}^n$ in the following way. For every $W := V_0 \cap \cdots \cap T^{-(n-1)}(V_{n-1}) \in \mathcal{V}^n$, define

$$i_n(W) := i(V_0) \cap \cdots \cap T^{-(n-1)}\big(i(V_{n-1})\big).$$

Observe that $W \subseteq i_n(W) \in \mathcal{U}^n$ for all $W \in \mathcal{V}^n$. Moreover, if $x \in W$ and $y \in i_n(W)$, then for each $0 \leq j < n$ we have that $T^j(x) \in V_j \subseteq i(V_j) \ni T^j(y)$. So $T^j(x), T^j(y) \in i(V_j)$ for all $0 \leq j < n$. Hence, $x, y \in i_n(W) \in \mathcal{U}^n$, and thus

$$S_n\varphi(x) \geq S_n\varphi(y) - \operatorname{osc}\big(S_n\varphi, \mathcal{U}^n\big).$$

Taking the supremum over all $x \in W$ on the left-hand side and over all $y \in i_n(W)$ on the right-hand side yields

$$\overline{S}_n\varphi(W) \geq \overline{S}_n\varphi(i_n(W)) - \operatorname{osc}(S_n\varphi, \mathcal{U}^n).$$

Now, let $\mathcal{W}$ be a subcover of $\mathcal{V}^n$. Then $i_n(\mathcal{W}) := \{i_n(W) : W \in \mathcal{W}\}$ is a subcover of $\mathcal{U}^n$. Therefore,

$$\begin{aligned}\sum_{W\in\mathcal{W}} e^{\overline{S}_n\varphi(W)} &\geq e^{-\operatorname{osc}(S_n\varphi,\mathcal{U}^n)} \sum_{W\in\mathcal{W}} e^{\overline{S}_n\varphi(i_n(W))} \\ &\geq e^{-\operatorname{osc}(S_n\varphi,\mathcal{U}^n)} \sum_{Y\in i_n(\mathcal{W})} e^{\overline{S}_n\varphi(Y)} \\ &\geq e^{-\operatorname{osc}(S_n\varphi,\mathcal{U}^n)} Z_n(\varphi,\mathcal{U}).\end{aligned}$$

Taking the infimum over all subcovers $\mathcal{W}$ of $\mathcal{V}^n$ on the left-hand side and using Lemma 11.1.5, we conclude that

$$Z_n(\varphi,\mathcal{V}) \geq e^{-\operatorname{osc}(S_n\varphi,\mathcal{U}^n)} Z_n(\varphi,\mathcal{U}) \geq e^{-n\operatorname{osc}(\varphi,\mathcal{U})} Z_n(\varphi,\mathcal{U}).$$

The proof of the inequality for the z_n's is left to the reader. □

In Lemma 7.2.9, we saw that the functions $Z_n(\cdot)$ are submultiplicative with respect to the join operation; in other words, $Z_n(\mathcal{U} \vee \mathcal{V}) \leq Z_n(\mathcal{U})Z_n(\mathcal{V})$ for all $n \in \mathbb{N}$. The corresponding property for the functions $Z_n(\varphi,\cdot)$ and $z_n(\varphi,\cdot)$ is the following.

Lemma 11.1.8. *Let $\mathcal{U}$ and $\mathcal{V}$ be open covers of X and let $n \in \mathbb{N}$. Then*

$$Z_n(\varphi,\mathcal{U} \vee \mathcal{V}) \leq \min\{Z_n(\varphi,\mathcal{U}) \cdot Z_n(\mathcal{V}), Z_n(\mathcal{U}) \cdot Z_n(\varphi,\mathcal{V})\}$$

and

$$z_n(\varphi,\mathcal{U} \vee \mathcal{V}) \leq \min\{e^{n\operatorname{osc}(\varphi,\mathcal{U})} z_n(\varphi,\mathcal{U}) \cdot Z_n(\mathcal{V}), Z_n(\mathcal{U}) \cdot e^{n\operatorname{osc}(\varphi,\mathcal{V})} z_n(\varphi,\mathcal{V})\}.$$

Proof. The proof is left to the reader as an exercise. □

We have also seen in Lemma 7.2.10 that the sequence $(Z_n(\mathcal{U}))_{n=1}^{\infty}$ is submultiplicative. This property is retained by the $Z_n(\varphi,\mathcal{U})$'s but generally not by the $z_n(\varphi,\mathcal{U})$'s (see Exercise 11.5.2).

Lemma 11.1.9. *Given an open cover $\mathcal{U}$ of X, the sequence $(Z_n(\varphi,\mathcal{U}))_{n=1}^{\infty}$ is submultiplicative.*

Proof. Fix $m, n \in \mathbb{N}$, let $\mathcal{V}$ be a subcover of $\mathcal{U}^m$ and $\mathcal{W}$ a subcover of $\mathcal{U}^n$. Note that $\mathcal{V} \vee T^{-m}(\mathcal{W})$ is a subcover of $\mathcal{U}^{m+n}$ since it is a cover and

$$\mathcal{V} \vee T^{-m}(\mathcal{W}) \subseteq \mathcal{U}^m \vee T^{-m}(\mathcal{U}^n) = \mathcal{U}^{m+n}.$$

Take arbitrary $V \in \mathcal{V}$ and $W \in \mathcal{W}$. Then for every $x \in V \cap T^{-m}(W)$, we have $x \in V$ and $T^m(x) \in W$, and hence

$$S_{m+n}\varphi(x) = S_m\varphi(x) + S_n\varphi\big(T^m(x)\big) \le \overline{S}_m\varphi(V) + \overline{S}_n\varphi(W).$$

Taking the supremum over all $x \in V \cap T^{-m}(W)$, we deduce that

$$\overline{S}_{m+n}\varphi\big(V \cap T^{-m}(W)\big) \le \overline{S}_m\varphi(V) + \overline{S}_n\varphi(W).$$

Therefore,

$$\begin{aligned} Z_{m+n}(\varphi,\mathcal{U}) &\le \sum_{E\in\mathcal{V}\vee T^{-m}(\mathcal{W})} e^{\overline{S}_{m+n}\varphi(E)} \\ &\le \sum_{V\in\mathcal{V}}\sum_{W\in\mathcal{W}} e^{\overline{S}_{m+n}\varphi(V\cap T^{-m}(W))} \\ &\le \sum_{V\in\mathcal{V}}\sum_{W\in\mathcal{W}} e^{\overline{S}_m\varphi(V)}e^{\overline{S}_n\varphi(W)} \\ &= \sum_{V\in\mathcal{V}} e^{\overline{S}_m\varphi(V)} \sum_{W\in\mathcal{W}} e^{\overline{S}_n\varphi(W)}. \end{aligned}$$

Taking the infimum of the right-hand side over all subcovers $\mathcal{V}$ of $\mathcal{U}^m$ and over all subcovers $\mathcal{W}$ of $\mathcal{U}^n$ gives

$$Z_{m+n}(\varphi,\mathcal{U}) \le Z_m(\varphi,\mathcal{U})Z_n(\varphi,\mathcal{U}). \qquad \square$$

We immediately deduce the following fact.

Corollary 11.1.10. *The sequence* $(\log Z_n(\varphi,\mathcal{U}))_{n=1}^{\infty}$ *is subadditive for every open cover* $\mathcal{U}$ *of* X.

Thanks to this fact, we can define the topological pressure of a potential with respect to an open cover. This constitutes the first step in the definition of the topological pressure of a potential.

Definition 11.1.11. The *topological pressure* of a potential $\varphi : X \to \mathbb{R}$ with respect to an open cover $\mathcal{U}$ of X, denoted by $\mathrm{P}(T,\varphi,\mathcal{U})$, is defined to be

$$\mathrm{P}(T,\varphi,\mathcal{U}) := \lim_{n\to\infty}\frac{1}{n}\log Z_n(\varphi,\mathcal{U}) = \inf_{n\in\mathbb{N}}\frac{1}{n}\log Z_n(\varphi,\mathcal{U}).$$

The existence of the limit and its equality with the infimum follow from Lemma 3.2.17 and Corollary 11.1.10, just as in the corresponding Definition 7.2.12 for topological entropy.

It is also possible to define similar quantities using the $z_n(\varphi,\mathcal{U})$'s rather than the $Z_n(\varphi,\mathcal{U})$'s.

Definition 11.1.12. Given a potential $\varphi : X \to \mathbb{R}$ and an open cover $\mathcal{U}$ of X, let

$$\underline{p}(T,\varphi,\mathcal{U}) := \liminf_{n\to\infty} \frac{1}{n} \log z_n(\varphi,\mathcal{U}) \quad \text{and} \quad \overline{p}(T,\varphi,\mathcal{U}) := \limsup_{n\to\infty} \frac{1}{n} \log z_n(\varphi,\mathcal{U}).$$

Remark 11.1.13. Let $\mathcal{U}$ be an open cover of X.
(a) $\mathrm{P}(T,0,\mathcal{U}) = \underline{p}(T,0,\mathcal{U}) = \overline{p}(T,0,\mathcal{U}) = \mathrm{h_{top}}(T,\mathcal{U})$ by Remark 11.1.3(c).
(b) By Remark 11.1.3(e),

$$-\infty < \mathrm{h_{top}}(T,\mathcal{U}) + \inf\varphi \le \mathrm{P}(T,\varphi,\mathcal{U}) \le \mathrm{h_{top}}(T,\mathcal{U}) + \sup\varphi < \infty.$$

These inequalities also hold with $\mathrm{P}(T,\varphi,\mathcal{U})$ replaced by $\underline{p}(T,\varphi,\mathcal{U})$ and $\overline{p}(T,\varphi,\mathcal{U})$, respectively.
(c) Using Lemma 11.1.6,

$$\underline{p}(T,\varphi,\mathcal{U}) \le \overline{p}(T,\varphi,\mathcal{U}) \le \mathrm{P}(T,\varphi,\mathcal{U}) \le \underline{p}(T,\varphi,\mathcal{U}) + \mathrm{osc}(\varphi,\mathcal{U}).$$

We have seen in Proposition 7.2.14 that the topological entropy relative to covers respects the refinement relation and is subadditive with respect to the join operation. The topological pressure satisfies the following similar properties.

Proposition 11.1.14. *Let $\mathcal{U}$ and $\mathcal{V}$ be open covers of X.*
(a) *If $\mathcal{U} \prec \mathcal{V}$, then* $\mathrm{P}(T,\varphi,\mathcal{U}) - \mathrm{osc}(\varphi,\mathcal{U}) \le \mathrm{P}(T,\varphi,\mathcal{V})$ *while*

$$\underline{p}(T,\varphi,\mathcal{U}) \le \underline{p}(T,\varphi,\mathcal{V}) \quad \textit{and} \quad \overline{p}(T,\varphi,\mathcal{U}) \le \overline{p}(T,\varphi,\mathcal{V}).$$

(b) $\mathrm{P}(T,\varphi,\mathcal{U}\vee\mathcal{V}) \le \min\{\mathrm{P}(T,\varphi,\mathcal{U}) + \mathrm{h_{top}}(T,\mathcal{V}), \mathrm{P}(T,\varphi,\mathcal{V}) + \mathrm{h_{top}}(T,\mathcal{U})\}$ *whereas*

$$\begin{aligned}\overline{p}(T,\varphi,\mathcal{U}\vee\mathcal{V}) \le \min\{&\overline{p}(T,\varphi,\mathcal{U}) + \mathrm{osc}(\varphi,\mathcal{U}) + \mathrm{h_{top}}(T,\mathcal{V}),\\ &\overline{p}(T,\varphi,\mathcal{V}) + \mathrm{osc}(\varphi,\mathcal{V}) + \mathrm{h_{top}}(T,\mathcal{U})\}\end{aligned}$$

and a similar inequality with $\overline{p}$ replaced by $\underline{p}$.

Proof. Part (a) is an immediate consequence of Lemma 11.1.7 while (b) follows directly from Lemma 11.1.8. □

We have proved in Lemma 7.2.15 that the entropy of a system relative to covers remains the same for all dynamical covers generated by a given cover. The topological pressure of a potential has a similar property.

Lemma 11.1.15. *If $\mathcal{U}$ is an open cover of X, then*

$$\underline{p}(T,\varphi,\mathcal{U}^n) = \underline{p}(T,\varphi,\mathcal{U}) \quad \textit{and} \quad \overline{p}(T,\varphi,\mathcal{U}^n) = \overline{p}(T,\varphi,\mathcal{U})$$

whereas $\mathrm{P}(T,\varphi,\mathcal{U}^n) \le \mathrm{P}(T,\varphi,\mathcal{U})$ *for all $n \in \mathbb{N}$. In addition, if $\mathcal{U}$ is an open partition of X, then* $\mathrm{P}(T,\varphi,\mathcal{U}^n) = \mathrm{P}(T,\varphi,\mathcal{U})$ *for all $n \in \mathbb{N}$.*

Proof. Fix $n \in \mathbb{N}$. For all $k \in \mathbb{N}$ and all $x \in X$, we already know that

$$S_{k+n-1}\varphi(x) = S_k\varphi(x) + S_{n-1}\varphi(T^k(x)).$$

Therefore,

$$S_k\varphi(x) - \|S_{n-1}\varphi\|_\infty \le S_{k+n-1}\varphi(x) \le S_k\varphi(x) + \|S_{n-1}\varphi\|_\infty.$$

Hence, for any subset Y of X,

$$\overline{S}_k\varphi(Y) - \|S_{n-1}\varphi\|_\infty \le \overline{S}_{k+n-1}\varphi(Y) \le \overline{S}_k\varphi(Y) + \|S_{n-1}\varphi\|_\infty \tag{11.1}$$

and

$$\underline{S}_k\varphi(Y) - \|S_{n-1}\varphi\|_\infty \le \underline{S}_{k+n-1}\varphi(Y) \le \underline{S}_k\varphi(Y) + \|S_{n-1}\varphi\|_\infty. \tag{11.2}$$

We claim that

$$e^{-\|S_{n-1}\varphi\|_\infty} Z_k(\varphi, \mathcal{U}^n) \le Z_{k+n-1}(\varphi, \mathcal{U}) \tag{11.3}$$

and

$$e^{-\|S_{n-1}\varphi\|_\infty} z_k(\varphi, \mathcal{U}^n) \le z_{k+n-1}(\varphi, \mathcal{U}) \le e^{\|S_{n-1}\varphi\|_\infty} z_k(\varphi, \mathcal{U}^n). \tag{11.4}$$

Let us first prove (11.3). Recall that $(\mathcal{U}^n)^k \prec \mathcal{U}^{k+n-1} \prec (\mathcal{U}^n)^k$ for all $k \in \mathbb{N}$ (cf. Lemma 7.1.12(d)). However, this is insufficient to declare that a subcover of $\mathcal{U}^{k+n-1}$ is also a subcover of $(\mathcal{U}^n)^k$, or vice versa. We need to remember that $\mathcal{U} \vee \mathcal{U} \supseteq \mathcal{U}$, and thus $(\mathcal{U}^n)^k \supseteq \mathcal{U}^{k+n-1}$, that is, $\mathcal{U}^{k+n-1}$ is a subcover of $(\mathcal{U}^n)^k$. Let $\mathcal{V}$ be a subcover of $\mathcal{U}^{k+n-1}$. Then $\mathcal{V}$ is a subcover of $(\mathcal{U}^n)^k$. Using the left inequality in (11.1) with Y replaced by each $V \in \mathcal{V}$ successively, we obtain

$$e^{-\|S_{n-1}\varphi\|_\infty} Z_k(\varphi, \mathcal{U}^n) \le e^{-\|S_{n-1}\varphi\|_\infty} \sum_{V\in\mathcal{V}} e^{\overline{S}_k\varphi(V)} \le \sum_{V\in\mathcal{V}} e^{\overline{S}_{k+n-1}\varphi(V)}.$$

Taking the infimum over all subcovers $\mathcal{V}$ of $\mathcal{U}^{k+n-1}$ yields (11.3). Similarly, using the left inequality in (11.2), we get that

$$e^{-\|S_{n-1}\varphi\|_\infty} z_k(\varphi, \mathcal{U}^n) \le e^{-\|S_{n-1}\varphi\|_\infty} \sum_{V\in\mathcal{V}} e^{\underline{S}_k\varphi(V)} \le \sum_{V\in\mathcal{V}} e^{\underline{S}_{k+n-1}\varphi(V)}.$$

Taking the infimum over all subcovers $\mathcal{V}$ of $\mathcal{U}^{k+n-1}$ yields the left inequality in (11.4). Regarding the right inequality, since $\mathcal{U}^{k+n-1} \prec (\mathcal{U}^n)^k$, there exists a map $i : (\mathcal{U}^n)^k \to \mathcal{U}^{k+n-1}$ such that $W \subseteq i(W)$ for all $W \in (\mathcal{U}^n)^k$. Let $\mathcal{W}$ be a subcover of $(\mathcal{U}^n)^k$. Then $i(\mathcal{W})$

is a subcover of $\mathcal{U}^{k+n-1}$ and, using the right inequality in (11.2), we deduce that

$$\begin{aligned}\sum_{W\in\mathcal{W}} e^{\underline{S}_k\varphi(W)} &\geq \sum_{W\in\mathcal{W}} e^{\underline{S}_k\varphi(i(W))} \geq \sum_{Z\in i(\mathcal{W})} e^{\underline{S}_k\varphi(Z)} \\ &\geq \sum_{Z\in i(\mathcal{W})} e^{\underline{S}_{k+n-1}\varphi(Z)-\|S_{n-1}\varphi\|_\infty} \\ &\geq e^{-\|S_{n-1}\varphi\|_\infty} z_{k+n-1}(\varphi,\mathcal{U}).\end{aligned}$$

Taking the infimum over all subcovers of $(\mathcal{U}^n)^k$ on the left-hand side gives the right inequality in (11.4). So (11.3) and (11.4) always hold.

Moreover, if $\mathcal{U}$ is a partition then $\mathcal{U}\vee\mathcal{U}=\mathcal{U}$, and thus $(\mathcal{U}^n)^k=\mathcal{U}^{k+n-1}$ for all $k\in\mathbb{N}$. Let $\mathcal{W}$ be a subcover of $(\mathcal{U}^n)^k$. Using the right inequality in (11.1), we conclude that

$$\sum_{W\in\mathcal{W}} e^{\overline{S}_k\varphi(W)} \geq \sum_{W\in\mathcal{W}} e^{\overline{S}_{k+n-1}\varphi(W)-\|S_{n-1}\varphi\|_\infty} \geq e^{-\|S_{n-1}\varphi\|_\infty} Z_{k+n-1}(\varphi,\mathcal{U}).$$

Taking the infimum over all subcovers of $(\mathcal{U}^n)^k$ on the left-hand side gives

$$Z_k(\varphi,\mathcal{U}^n) \geq e^{-\|S_{n-1}\varphi\|_\infty} Z_{k+n-1}(\varphi,\mathcal{U}). \tag{11.5}$$

Finally, for the passage from the z_n's to $\overline{p}$, it follows from (11.4) that

$$\frac{k}{k+n-1}\cdot\frac{1}{k}\log z_k(\varphi,\mathcal{U}^n) - \frac{\|S_{n-1}\varphi\|_\infty}{k+n-1} \leq \frac{1}{k+n-1}\log z_{k+n-1}(\varphi,\mathcal{U})$$

and

$$\frac{1}{k+n-1}\log z_{k+n-1}(\varphi,\mathcal{U}) \leq \frac{k}{k+n-1}\cdot\frac{1}{k}\log z_k(\varphi,\mathcal{U}^n) + \frac{\|S_{n-1}\varphi\|_\infty}{k+n-1}.$$

Taking the lim sup as $k\to\infty$ in these two relations yields

$$\overline{p}(T,\varphi,\mathcal{U}^n) \leq \overline{p}(T,\varphi,\mathcal{U}) \leq \overline{p}(T,\varphi,\mathcal{U}^n).$$

Taking the lim inf instead, results in a corresponding conclusion for $\underline{p}$. Similarly, one deduces from (11.3) that $\mathrm{P}(T,\varphi,\mathcal{U}^n) \leq \mathrm{P}(T,\varphi,\mathcal{U})$ and, when $\mathcal{U}$ is a partition, it ensues from (11.5) that $\mathrm{P}(T,\varphi,\mathcal{U}^n) \geq \mathrm{P}(T,\varphi,\mathcal{U})$. □

11.1.2 Second stage: the pressure of a potential

Recall that the topological entropy of a system is defined to be the supremum over all open covers of the entropy of the system with respect to an open cover (cf. Definition 7.2.16). However, due to Proposition 11.1.14(a), taking the supremum of the pressure relative to all covers does not always lead to a quantity that has natural properties. Instead, we take the supremum of the difference between the pressure relative to a cover minus the oscillation of the potential with respect to that cover. This definition is purely topological.

Definition 11.1.16. Let $T : X \to X$ be a topological dynamical system and $\varphi : X \to \mathbb{R}$ a potential. The *topological pressure* of the potential φ, denoted $\mathrm{P}(T, \varphi)$, is defined by

$$\mathrm{P}(T, \varphi) := \sup\{\mathrm{P}(T, \varphi, \mathcal{U}) - \operatorname{osc}(\varphi, \mathcal{U}) : \mathcal{U} \text{ is an open cover of } X\}.$$

In light of Proposition 11.1.14(a), we may define the counterparts $\underline{p}(T, \varphi)$ and $\overline{p}(T, \varphi)$ of $\mathrm{P}(T, \varphi)$ by simply taking the supremum over all covers.

Definition 11.1.17. Let $T : X \to X$ be a topological dynamical system and $\varphi : X \to \mathbb{R}$ a potential. Define

$$\underline{p}(T, \varphi) := \sup\{\underline{p}(T, \varphi, \mathcal{U}) : \mathcal{U} \text{ is an open cover of } X\},$$

and

$$\overline{p}(T, \varphi) := \sup\{\overline{p}(T, \varphi, \mathcal{U}) : \mathcal{U} \text{ is an open cover of } X\}.$$

Clearly, $\underline{p}(T, \varphi) \le \overline{p}(T, \varphi)$. In fact, $\underline{p}(T, \varphi)$ and $\overline{p}(T, \varphi)$ are just other expressions of the topological pressure.

Theorem 11.1.18. *For any topological dynamical system $T : X \to X$ and potential $\varphi : X \to \mathbb{R}$, it turns out that $\underline{p}(T, \varphi) = \overline{p}(T, \varphi) = \mathrm{P}(T, \varphi)$.*

Proof. From a rearrangement of the right inequality in Remark 11.1.13(c), it follows that $\mathrm{P}(T, \varphi) \le \underline{p}(T, \varphi) \le \overline{p}(T, \varphi)$.

To prove that $\overline{p}(T, \varphi) \le \mathrm{P}(T, \varphi)$, let $(\mathcal{U}_n)_{n=1}^{\infty}$ be a sequence of open covers such that $\lim_{n\to\infty} \overline{p}(T, \varphi, \mathcal{U}_n) = \overline{p}(T, \varphi)$. Each open cover $\mathcal{U}_n$ has a Lebesgue number $\delta_n > 0$. The compactness of X guarantees that there are finitely many open balls of radius $\min\{\delta_n, 1/(2n)\}$ that cover X. These balls thereby constitute a refinement of $\mathcal{U}_n$ of diameter at most $1/n$. Thanks to Proposition 11.1.14(a), this means that we may assume without loss of generality that the sequence $(\mathcal{U}_n)_{n=1}^{\infty}$ is such that $\lim_{n\to\infty} \operatorname{diam}(\mathcal{U}_n) = 0$. Since φ is uniformly continuous, it ensues that $\lim_{n\to\infty} \operatorname{osc}(\varphi, \mathcal{U}_n) = 0$. Consequently, using the left inequality in Remark 11.1.13(c), we conclude that

$$\begin{aligned}
\mathrm{P}(T, \varphi) &\ge \sup_{n\in\mathbb{N}}\big[\mathrm{P}(T, \varphi, \mathcal{U}_n) - \operatorname{osc}(\varphi, \mathcal{U}_n)\big] \\
&\ge \sup_{n\in\mathbb{N}}\big[\overline{p}(T, \varphi, \mathcal{U}_n) - \operatorname{osc}(\varphi, \mathcal{U}_n)\big] \\
&\ge \lim_{n\in\mathbb{N}}\big[\overline{p}(T, \varphi, \mathcal{U}_n) - \operatorname{osc}(\varphi, \mathcal{U}_n)\big] = \overline{p}(T, \varphi).
\end{aligned}$$

□

Remark 11.1.19.

(a) $\mathrm{P}(T, 0) = \mathrm{h_{top}}(T)$. Thus topological pressure generalizes topological entropy. This is a consequence of Remark 11.1.13(a) and the fact that $\operatorname{osc}(0, \mathcal{U}) = 0$ for all $\mathcal{U}$.

(b) By Remark 11.1.13(b),

$$\mathrm{h_{top}}(T) + \inf \varphi - \operatorname{osc}(\varphi, X) \le \mathrm{P}(T, \varphi) \le \mathrm{h_{top}}(T) + \sup \varphi.$$

(c) $\mathrm{P}(T, \varphi) = \infty$ if and only if $\mathrm{h_{top}}(T) = \infty$, according to part (b).

We saw in Proposition 7.2.18 that topological entropy cannot be greater for a factor of a given map than for the original map. We expect the same for its generalization, topological pressure, with a small twist, namely that the potentials to which the dynamical systems are subject must correspond.

Proposition 11.1.20. *Suppose that $S : Y \to Y$ is a factor of $T : X \to X$ via the factor map $h : X \to Y$. Then for every potential $\varphi : Y \to \mathbb{R}$, we have that $\mathrm{P}(S, \varphi) \le \mathrm{P}(T, \varphi \circ h)$.*

Proof. Let $\mathcal{V}$ be an open cover of Y. Recall (cf. proof of Proposition 7.2.18) that $h^{-1}(\mathcal{V}_S^n) = (h^{-1}(\mathcal{V}))_T^n$ for all $n \in \mathbb{N}$. Without loss of generality, we may restrict our attention to nondegenerate subcovers, that is, subcovers whose members are all different from one another. Letting C be the collection of all nondegenerate subcovers of $\mathcal{V}_S^n$, the map $\mathcal{C} \mapsto h^{-1}(\mathcal{C}), \mathcal{C} \in C$, defines a bijection between the nondegenerate subcovers of $\mathcal{V}_S^n$ and the nondegenerate subcovers of $h^{-1}(\mathcal{V}_S^n) = (h^{-1}(\mathcal{V}))_T^n$, since h is a surjection. We leave it to the reader to show that

$$\overline{S}_n^T(\varphi \circ h)(h^{-1}(Z)) = \overline{S}_n^S\varphi(Z), \quad \forall Z \subseteq Y.$$

It then follows that (again this is left to the reader)

$$Z_n(T, \varphi \circ h, h^{-1}(\mathcal{V})) = Z_n(S, \varphi, \mathcal{V}).$$

Therefore,

$$\mathrm{P}(T, \varphi \circ h, h^{-1}(\mathcal{V})) = \mathrm{P}(S, \varphi, \mathcal{V}).$$

Observe further that $\mathrm{osc}(\varphi \circ h, h^{-1}(\mathcal{V})) = \mathrm{osc}(\varphi, \mathcal{V})$. Then

$$\begin{aligned}\mathrm{P}(T, \varphi \circ h) &\ge \mathrm{P}(T, \varphi \circ h, h^{-1}(\mathcal{V})) - \mathrm{osc}(\varphi \circ h, h^{-1}(\mathcal{V}))\\ &= \mathrm{P}(S, \varphi, \mathcal{V}) - \mathrm{osc}(\varphi, \mathcal{V}).\end{aligned}$$

Taking the supremum over all open covers $\mathcal{V}$ of Y yields $\mathrm{P}(T, \varphi \circ h) \ge \mathrm{P}(S, \varphi)$. □

An immediate but important consequence of this lemma is the following.

Corollary 11.1.21. *If $T : X \to X$ and $S : Y \to Y$ are topologically conjugate dynamical systems via a conjugacy $h : X \to Y$, then $\mathrm{P}(S, \varphi) = \mathrm{P}(T, \varphi \circ h)$ for all potentials $\varphi : Y \to \mathbb{R}$.*

We now study the behavior of topological pressure with respect to the iterates of the system. This is a generalization of Theorem 7.2.19.

Theorem 11.1.22. *For every $n \in \mathbb{N}$, we have that $\mathrm{P}(T^n, S_n\varphi) = n\,\mathrm{P}(T, \varphi)$.*

Proof. Fix $n \in \mathbb{N}$. Let $\mathcal{U}$ be an open cover of X. The action of the map T^n on $\mathcal{U}$ until time $j - 1$ will be denoted by $\mathcal{U}_{T^n}^j$. Recall (cf. proof of Theorem 7.2.19) that $\mathcal{U}^{mn} = (\mathcal{U}^n)_{T^n}^m$ for all $m \in \mathbb{N}$. Furthermore, for all $x \in X$,

$$S_{mn}\varphi(x) = \sum_{k=0}^{mn-1} \varphi \circ T^k(x) = \sum_{j=0}^{m-1} (S_n\varphi) \circ T^{jn}(x) = \sum_{j=0}^{m-1} (S_n\varphi) \circ (T^n)^j(x) = S_m^{T^n}(S_n\varphi)(x),$$

where $S_m^{T^n}\psi(x) = \sum_{j=0}^{m-1}\psi((T^n)^j(x))$. Hence, $\underline{S}_{mn}\varphi(Y) = \underline{S}_m^{T^n}(S_n\varphi)(Y)$ for all subsets Y of X, and in particular for all $Y \in \mathcal{U}^{mn} = (\mathcal{U}^n)_{T^n}^m$. Thus,

$$z_{mn}(T,\varphi,\mathcal{U}) = z_m(T^n, S_n\varphi, \mathcal{U}^n), \quad \forall m \in \mathbb{N}.$$

Using this and Lemma 11.1.14(a), we get

$$\begin{aligned}\overline{p}(T,\varphi,\mathcal{U}) = \limsup_{m\to\infty}\frac{1}{m}\log z_m(T,\varphi,\mathcal{U}) &\geq \limsup_{m\to\infty}\frac{1}{mn}\log z_{mn}(T,\varphi,\mathcal{U}) \\ &= \frac{1}{n}\limsup_{m\to\infty}\frac{1}{m}\log z_m(T^n,S_n\varphi,\mathcal{U}^n) \\ &= \frac{1}{n}\overline{p}(T^n,S_n\varphi,\mathcal{U}^n) \geq \frac{1}{n}\overline{p}(T^n,S_n\varphi,\mathcal{U}).\end{aligned}$$

Taking the supremum over all open covers $\mathcal{U}$ of X yields

$$\overline{p}(T,\varphi) \geq \frac{1}{n}\overline{p}(T^n,S_n\varphi).$$

Similarly,

$$\underline{p}(T,\varphi) \leq \frac{1}{n}\underline{p}(T^n,S_n\varphi).$$

The result ensues from the previous two relations and Theorem 11.1.18. □

As a generalization of topological entropy, in a metrizable space topological pressure is determined by any sequence of covers whose diameters tend to zero. The next result is an extension of Lemma 7.2.20.

Lemma 11.1.23. *The following quantities are all equal:*
(a) $\mathrm{P}(T,\varphi)$.
(b) $\overline{p}(T,\varphi)$.
(c) $\lim_{\varepsilon\to 0}[\sup\{\mathrm{P}(T,\varphi,\mathcal{U}) : \mathcal{U}$ *open cover with* $\mathrm{diam}(\mathcal{U}) \leq \varepsilon\}]$.
(d) $\sup\{\overline{p}(T,\varphi,\mathcal{U}) : \mathcal{U}$ *open cover with* $\mathrm{diam}(\mathcal{U}) \leq \delta\}$ *for any* $\delta > 0$.
(e) $\lim_{\varepsilon\to 0}\mathrm{P}(T,\varphi,\mathcal{U}_\varepsilon)$ *for any open covers* $(\mathcal{U}_\varepsilon)_{\varepsilon\in(0,\infty)}$ *such that* $\lim_{\varepsilon\to 0}\mathrm{diam}(\mathcal{U}_\varepsilon) = 0$.
(f) $\lim_{\varepsilon\to 0}\overline{p}(T,\varphi,\mathcal{U}_\varepsilon)$ *for any open covers* $(\mathcal{U}_\varepsilon)_{\varepsilon\in(0,\infty)}$ *such that* $\lim_{\varepsilon\to 0}\mathrm{diam}(\mathcal{U}_\varepsilon) = 0$.
(g) $\lim_{n\to\infty}\mathrm{P}(T,\varphi,\mathcal{U}_n)$ *for any open covers* $(\mathcal{U}_n)_{n=1}^\infty$ *such that* $\lim_{n\to\infty}\mathrm{diam}(\mathcal{U}_n) = 0$.
(h) $\lim_{n\to\infty}\overline{p}(T,\varphi,\mathcal{U}_n)$ *for any open covers* $(\mathcal{U}_n)_{n=1}^\infty$ *such that* $\lim_{n\to\infty}\mathrm{diam}(\mathcal{U}_n) = 0$.

Note that $\overline{p}$ *can be replaced by* $\underline{p}$ *in the statements above.*

Proof. We already know that (a)=(b) by Lemma 11.1.18. It is clear that (b)≥(d). It is also obvious that (d)≥(f) and (c)≥(e) for any family $(\mathcal{U}_\varepsilon)_{\varepsilon\in(0,\infty)}$ as described, and that (d)≥(h) and (c)≥(g) for any sequence $(\mathcal{U}_n)_{n=1}^\infty$ as specified. It thus suffices to prove that (f)≥(b), that (h)≥(b), that (e)≥(a), that (g)≥(a), and that (b)≥(c).

We will prove that (g)≥(a). The proofs of the other inequalities are similar. Let $\mathcal{V}$ be any open cover of X. Since $\lim_{n\to\infty}\mathrm{diam}(\mathcal{U}_n) = 0$, there exists $N \in \mathbb{N}$ such that $\mathcal{V} \prec \mathcal{U}_n$

for all $n \geq N$ (cf. proof of Lemma 7.2.20). By Proposition 11.1.14(a), we obtain that for all sufficiently large n,

$$\mathrm{P}(T,\varphi,\mathcal{U}_n) \geq \mathrm{P}T,\varphi,\mathcal{V}) - \mathrm{osc}(\varphi,\mathcal{V}).$$

We immediately deduce that

$$\liminf_{n\to\infty} \mathrm{P}(T,\varphi,\mathcal{U}_n) \geq \mathrm{P}(T,\varphi,\mathcal{V}) - \mathrm{osc}(\varphi,\mathcal{V}).$$

As the open cover $\mathcal{V}$ was chosen arbitrarily, passing to the supremum over all open covers allows us to conclude that

$$\liminf_{n\to\infty} \mathrm{P}(T,\varphi,\mathcal{U}_n) \geq \mathrm{P}(T,\varphi).$$

But $\lim_{n\to\infty} \mathrm{osc}(\varphi,\mathcal{U}_n) = 0$ since $\lim_{n\to\infty} \mathrm{diam}(\mathcal{U}_n) = 0$ and φ is uniformly continuous. Therefore,

$$\begin{aligned}
\mathrm{P}(T,\varphi) &= \sup_{\mathcal{V}}\big[\mathrm{P}(T,\varphi,\mathcal{V}) - \mathrm{osc}(\varphi,\mathcal{V})\big] \\
&\geq \limsup_{n\to\infty}\big[\mathrm{P}(T,\varphi,\mathcal{U}_n) - \mathrm{osc}(\varphi,\mathcal{U}_n)\big] \\
&= \limsup_{n\to\infty} \mathrm{P}(T,\varphi,\mathcal{U}_n) - \lim_{n\to\infty} \mathrm{osc}(\varphi,\mathcal{U}_n) \\
&= \limsup_{n\to\infty} \mathrm{P}(T,\varphi,\mathcal{U}_n) \geq \liminf_{n\to\infty} \mathrm{P}(T,\varphi,\mathcal{U}_n) \geq \mathrm{P}(T,\varphi).
\end{aligned}$$

Hence, $\mathrm{P}(T,\varphi) = \lim_{n\to\infty} \mathrm{P}(T,\varphi,\mathcal{U}_n)$. □

We can now obtain a slightly stronger estimate than Remark 11.1.19(b) for the difference between topological entropy and topological pressure when the underlying space is metrizable.

Corollary 11.1.24. $\mathrm{h}_{\mathrm{top}}(T) + \inf\varphi \leq \mathrm{P}(T,\varphi) \leq \mathrm{h}_{\mathrm{top}}(T) + \sup\varphi$.

Proof. The upper bound was already mentioned in Remark 11.1.19(b). In order to derive the lower bound, we return to Remark 11.1.13(b). Let $(\mathcal{U}_n)_{n=1}^{\infty}$ be a sequence of open covers of X such that $\lim_{n\to\infty} \mathrm{diam}(\mathcal{U}_n) = 0$. According to Remark 11.1.13(b), for each $n \in \mathbb{N}$ we have

$$\mathrm{h}_{\mathrm{top}}(T,\mathcal{U}_n) + \inf\varphi \leq \mathrm{P}(T,\varphi,\mathcal{U}_n).$$

Passing to the limit $n \to \infty$ and using Lemmas 7.2.20 and 11.1.23, we conclude that

$$\mathrm{h}_{\mathrm{top}}(T) + \inf\varphi \leq \mathrm{P}(T,\varphi).$$ □

The preceding lemma characterized the topological pressure of a potential as the limit of the topological pressure of the potential relative to a sequence of covers. An even better result would be the characterization of the topological pressure of a potential as the topological pressure of that potential with respect to a single cover. As might by now be expected, such a characterization exists when the system has a generator. This is a generalization of Lemma 7.2.22 (see also Definition 7.2.21).

Lemma 11.1.25. *If a system $T : X \to X$ has a generator $\mathcal{U}$, then*

$$\mathrm{P}(T, \varphi) = \underline{p}(T, \varphi, \mathcal{U}) = \overline{p}(T, \varphi, \mathcal{U}).$$

Moreover, if the generator $\mathcal{U}$ is a partition, then

$$\mathrm{P}(T, \varphi) = \mathrm{P}(T, \varphi, \mathcal{U}).$$

Proof. It follows from Lemmas 11.1.23 and 11.1.15 that

$$\mathrm{P}(T, \varphi) = \lim_{n\to\infty} \overline{p}(T, \varphi, \mathcal{U}^n) = \lim_{n\to\infty} \overline{p}(T, \varphi, \mathcal{U}) = \overline{p}(T, \varphi, \mathcal{U}).$$

A similar argument leads to the statements for $\underline{p}$ and for a generating partition. □

We then have the following generalization of Theorem 7.2.24.

Theorem 11.1.26. *If $T : X \to X$ is a δ-expansive dynamical system on a compact metric space (X, d), then*

$$\mathrm{P}(T, \varphi) = \underline{p}(T, \varphi, \mathcal{U}) = \overline{p}(T, \varphi, \mathcal{U})$$

for any open cover $\mathcal{U}$ of X with $\mathrm{diam}(\mathcal{U}) \leq \delta$. *Moreover,*

$$\mathrm{P}(T, \varphi) = \mathrm{P}(T, \varphi, \mathcal{U})$$

for any open partition $\mathcal{U}$ of X with $\mathrm{diam}(\mathcal{U}) \leq \delta$.

Proof. This is an immediate consequence of Lemmas 11.1.25 and 7.2.23. □

11.2 Bowen's definition of topological pressure

We have seen in Theorem 7.3.8 and Corollary 7.3.12 that topological entropy can also be defined using separated or spanning sets. This definition may be generalized to yield a definition of topological pressure, which coincides with the one from the previous section. To lighten notation, for any $n \in \mathbb{N}$ and $Y \subseteq X$, let

$$\Sigma_n(Y) = \sum_{x\in Y} e^{S_n\varphi(x)}.$$

Theorem 11.2.1. *For all $n \in \mathbb{N}$ and all $\varepsilon > 0$, let $E_n(\varepsilon)$ be a maximal (n,ε)-separated set and $F_n(\varepsilon)$ be a minimal (n,ε)-spanning set. Then*

$$\begin{aligned}\mathrm{P}(T,\varphi) = \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\Sigma_n(E_n(\varepsilon)) &= \lim_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\Sigma_n(E_n(\varepsilon))\\ &\le \liminf_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\Sigma_n(F_n(\varepsilon)).\end{aligned}$$

Proof. Fix $\varepsilon > 0$. Let $\mathcal{U}_\varepsilon$ be an open cover of X consisting of balls of radius $\varepsilon/2$. Fix $n \in \mathbb{N}$. Let $\mathcal{U}$ be a subcover of $\mathcal{U}_\varepsilon^n$ such that $Z_n(\varphi,\mathcal{U}_\varepsilon) \ge e^{-1}\sum_{U\in\mathcal{U}}\exp(\overline{S}_n\varphi(U))$. For each $x \in E_n(\varepsilon)$, let $U(x)$ be an element of the cover $\mathcal{U}$ which contains x and define the function $i : E_n(\varepsilon) \to \mathcal{U}$ by setting $i(x) = U(x)$. We have already shown in the proof of Theorem 7.3.8 that this function is an injection. Therefore,

$$Z_n(\varphi,\mathcal{U}_\varepsilon) \ge e^{-1}\sum_{U\in\mathcal{U}} e^{\overline{S}_n\varphi(U)} \ge e^{-1}\sum_{x\in E_n(\varepsilon)} e^{\overline{S}_n\varphi(U(x))} \ge e^{-1}\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)}.$$

Since this is true for all $n \in \mathbb{N}$, we deduce that

$$\mathrm{P}(T,\varphi,\mathcal{U}_\varepsilon) = \lim_{n\to\infty}\frac{1}{n}\log Z_n(\varphi,\mathcal{U}_\varepsilon) \ge \limsup_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)}.$$

Letting $\varepsilon \to 0$ and using Lemma 11.1.23 yields that

$$\mathrm{P}(T,\varphi) \ge \limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)}. \tag{11.6}$$

On the other hand, if $\mathcal{V}$ is an arbitrary open cover of X, if $\delta(\mathcal{V})$ is a Lebesgue number for $\mathcal{V}$, if $0 < \varepsilon < \delta(\mathcal{V})/2$ and if $n \in \mathbb{N}$, then for all $0 \le k < n$ and all $x \in E_n(\varepsilon)$ we have

$$T^k(B_n(x,\varepsilon)) \subseteq B(T^k(x),\varepsilon) \implies \mathrm{diam}(T^k(B_n(x,\varepsilon))) \le 2\varepsilon < \delta(\mathcal{V}).$$

Hence, for all $0 \le k < n$, the set $T^k(B_n(x,\varepsilon))$ is contained in at least one element of $\mathcal{V}$. Denote one such element by $V_k(x)$. Then $B_n(x,\varepsilon) \subseteq \bigcap_{k=0}^{n-1} T^{-k}(V_k(x))$. But this latter intersection is simply an element of $\mathcal{V}^n$. Let us denote it by $V(x)$.

Since $E_n(\varepsilon)$ is a maximal (n,ε)-separated set, by Lemma 7.3.7 it is also (n,ε)-spanning, so the family $\{B_n(x,\varepsilon)\}_{x\in E_n(\varepsilon)}$ is an open cover of X. Each of these balls is contained in the corresponding set $V(x)$. Hence, the family $\{V(x)\}_{x\in E_n(\varepsilon)}$ is also an open cover of X. Therefore, it is a subcover of $\mathcal{V}^n$. Consequently,

$$Z_n(\varphi,\mathcal{V}) \le \sum_{x\in E_n(\varepsilon)} e^{\overline{S}_n\varphi(V(x))} \le e^{n\,\mathrm{osc}(\varphi,\mathcal{V})}\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)},$$

where the last inequality is due to Lemma 11.1.5. It follows that

$$\mathrm{P}(T,\varphi,\mathcal{V}) \le \mathrm{osc}(\varphi,\mathcal{V}) + \liminf_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)}.$$

Since $\mathcal{V}$ is independent of $\varepsilon > 0$, we deduce that

$$\mathrm{P}(T,\varphi,\mathcal{V}) - \operatorname{osc}(\varphi,\mathcal{V}) \le \liminf_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)}.$$

Then, as $\mathcal{V}$ was chosen to be an arbitrary open cover of X, we conclude that

$$\mathrm{P}(T,\varphi) \le \liminf_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)}. \tag{11.7}$$

Inequalities (11.6)–(11.7) establish the result for the separated sets. We can deduce the result for the spanning sets as in the proof of Theorem 7.3.8. □

In Theorem 11.2.1, the topological pressure of the system is expressed in terms of a specific family of maximal separated (resp. minimal spanning) sets. However, to derive theoretical results, it is sometimes simpler to use the following quantities.

Definition 11.2.2. For all $n \in \mathbb{N}$ and $\varepsilon > 0$, let

$$P_n(T,\varphi,\varepsilon) = \sup\{\Sigma_n(E_n(\varepsilon)) \,:\, E_n(\varepsilon) \text{ maximal } (n,\varepsilon)\text{-separated set}\}$$
$$Q_n(T,\varphi,\varepsilon) = \inf\{\Sigma_n(F_n(\varepsilon)) \,:\, F_n(\varepsilon) \text{ minimal } (n,\varepsilon)\text{-spanning set}\}.$$

Thereafter, let

$$\underline{P}(T,\varphi,\varepsilon) = \liminf_{n\to\infty}\frac{1}{n}\log P_n(T,\varphi,\varepsilon), \qquad \overline{P}(T,\varphi,\varepsilon) = \limsup_{n\to\infty}\frac{1}{n}\log P_n(T,\varphi,\varepsilon)$$
$$\underline{Q}(T,\varphi,\varepsilon) = \liminf_{n\to\infty}\frac{1}{n}\log Q_n(T,\varphi,\varepsilon), \qquad \overline{Q}(T,\varphi,\varepsilon) = \limsup_{n\to\infty}\frac{1}{n}\log Q_n(T,\varphi,\varepsilon).$$

The following are key observations constitute a generalization of Remark 7.3.10.

Remark 11.2.3. Let $m \le n \in \mathbb{N}$ and $0 < \varepsilon < \varepsilon'$. The following relations hold:
(a) $P_m(T,\varphi,\varepsilon) \le P_n(T,\varphi,\varepsilon)e^{(n-m)\|\varphi\|_\infty}$ by Remark 7.3.2(a).
(b) $e^{-n\|\varphi\|_\infty} \le P_n(T,\varphi,\varepsilon) \le r_n(\varepsilon)e^{n\|\varphi\|_\infty}$ and $P_n(T,0,\varepsilon) = r_n(\varepsilon)$.
(c) $Q_m(T,\varphi,\varepsilon) \le Q_n(T,\varphi,\varepsilon)e^{(n-m)\|\varphi\|_\infty}$ by Remark 7.3.6(a).
(d) $e^{-n\|\varphi\|_\infty} \le Q_n(T,\varphi,\varepsilon) \le s_n(\varepsilon)e^{n\|\varphi\|_\infty}$ and $Q_n(T,0,\varepsilon) = s_n(\varepsilon)$.
(e) $P_n(T,\varphi,\varepsilon) \ge P_n(T,\varphi,\varepsilon')$ and $Q_n(T,\varphi,\varepsilon) \ge Q_n(T,\varphi,\varepsilon')$ by Remarks 7.3.2 and 7.3.6(b).
(f) $0 < Q_n(T,\varphi,\varepsilon) \le P_n(T,\varphi,\varepsilon) < \infty$ by Lemma 7.3.7.
(g) $\underline{P}(T,\varphi,\varepsilon) \le \overline{P}(T,\varphi,\varepsilon)$ and $\underline{Q}(T,\varphi,\varepsilon) \le \overline{Q}(T,\varphi,\varepsilon)$.
(h) $-\|\varphi\|_\infty \le \underline{P}(T,\varphi,\varepsilon) \le \underline{r}(\varepsilon) + \|\varphi\|_\infty$ and $-\|\varphi\|_\infty \le \overline{P}(T,\varphi,\varepsilon) \le \overline{r}(\varepsilon) + \|\varphi\|_\infty$ by (b).
(i) $-\|\varphi\|_\infty \le \underline{Q}(T,\varphi,\varepsilon) \le \underline{s}(\varepsilon) + \|\varphi\|_\infty$ and $-\|\varphi\|_\infty \le \overline{Q}(T,\varphi,\varepsilon) \le \overline{s}(\varepsilon) + \|\varphi\|_\infty$ by (d).
(j) $\underline{P}(T,\varphi,\varepsilon) \ge \underline{P}(T,\varphi,\varepsilon')$ and $\overline{P}(T,\varphi,\varepsilon) \ge \overline{P}(T,\varphi,\varepsilon')$ by (e).
(k) $\underline{Q}(T,\varphi,\varepsilon) \ge \underline{Q}(T,\varphi,\varepsilon')$ and $\overline{Q}(T,\varphi,\varepsilon) \ge \overline{Q}(T,\varphi,\varepsilon')$ by (e).
(l) $-\|\varphi\|_\infty \le \overline{Q}(T,\varphi,\varepsilon) \le \overline{P}(T,\varphi,\varepsilon) \le \infty$ and $-\|\varphi\|_\infty \le \underline{Q}(T,\varphi,\varepsilon) \le \underline{P}(T,\varphi,\varepsilon) \le \infty$ by (f).

We now describe a relationship between the P_n's, the Q_n's and the cover-related quantities Z_n's and z_n's. This is the counterpart of Lemma 7.3.11.

Lemma 11.2.4. *The following relations hold:*
(a) *If $\mathcal{U}$ is an open cover of X with Lebesgue number 2δ, then*

$$z_n(T,\varphi,\mathcal{U}) \le Q_n(T,\varphi,\delta) \le P_n(T,\varphi,\delta).$$

(b) *If $\varepsilon > 0$ and $\mathcal{V}$ is an open cover of X with* $\operatorname{diam}(\mathcal{V}) \le \varepsilon$*, then*

$$Q_n(T,\varphi,\varepsilon) \le P_n(T,\varphi,\varepsilon) \le Z_n(T,\varphi,\mathcal{V}).$$

Proof. We already know that $Q_n(T,\varphi,\delta) \le P_n(T,\varphi,\delta)$.

(a) Let $\mathcal{U}$ be an open cover with Lebesgue number 2δ and let F be an (n,δ)-spanning set. Then the dynamic balls $\{B_n(x,\delta) : x \in F\}$ form a cover of X. For every $0 \le i < n$, the ball $B(T^i(x),\delta)$, which has diameter at most 2δ, is contained in an element of $\mathcal{U}$. Therefore $B_n(x,\delta) = \bigcap_{i=0}^{n-1} T^{-i}(B(T^i(x),\delta))$ is contained in an element of $\mathcal{U}^n = \bigvee_{i=0}^{n-1} T^{-i}(\mathcal{U})$. That is, $\mathcal{U}^n \prec \{B_n(x,\delta) : x \in F\}$. Then there exists a map $i : \{B_n(x,\delta) : x \in F\} \to \mathcal{U}^n$ such that $B_n(x,\delta) \subseteq i(B_n(x,\delta))$ for every $x \in F$. Let $\mathcal{W}$ be a subcover of $\{B_n(x,\delta) : x \in F\}$. Then $i(\mathcal{W})$ is a subcover of $\mathcal{U}^n$ and thus

$$\begin{aligned}\Sigma_n(F) = \sum_{x\in F} e^{S_n\varphi(x)} \ge \sum_{x\in F} e^{\underline{S}_n\varphi(B_n(x,\delta))} &\ge \sum_{W\in\mathcal{W}} e^{\underline{S}_n\varphi(W)} \ge \sum_{W\in\mathcal{W}} e^{\underline{S}_n\varphi(i(W))}\\ &\ge \sum_{Z\in i(\mathcal{W})} e^{\underline{S}_n\varphi(Z)} \ge z_n(T,\varphi,\mathcal{U}).\end{aligned}$$

Since F is an arbitrary (n,δ)-spanning set, it ensues that $Q_n(T,\varphi,\delta) \ge z_n(T,\varphi,\mathcal{U})$.

(b) Let $\mathcal{V}$ be an open cover with $\operatorname{diam}(\mathcal{V}) \le \varepsilon$ and let E be an (n,ε)-separated set. Let $\mathcal{W}$ be a subcover of $\mathcal{V}^n$. Let $i : E \to \mathcal{W}$ be such that $x \in i(x)$ for all $x \in E$. This map is injective as no element of the cover $\mathcal{V}^n$ can contain more than one element of E. Then

$$\Sigma_n(E) = \sum_{x\in E} e^{S_n\varphi(x)} \le \sum_{x\in E} e^{\overline{S}_n\varphi(i(x))} = \sum_{W\in i(E)} e^{\overline{S}_n\varphi(W)} \le \sum_{W\in\mathcal{W}} e^{\overline{S}_n\varphi(W)}.$$

As $\mathcal{W}$ is an arbitrary subcover of $\mathcal{V}^n$, it follows that $\Sigma_n(E) \le Z_n(T,\varphi,\mathcal{V})$. Since E is an arbitrary (n,ε)-separated set, we deduce that $P_n(T,\varphi,\varepsilon) \le Z_n(T,\varphi,\mathcal{V})$. □

These inequalities have the following immediate consequences.

Corollary 11.2.5. *The following relations hold:*
(a) *If $\mathcal{U}$ is an open cover of X with Lebesgue number 2δ, then*

$$\underline{p}(T,\varphi,\mathcal{U}) \le \underline{Q}(T,\varphi,\delta) \le \underline{P}(T,\varphi,\delta).$$

(b) *If $\varepsilon > 0$ and $\mathcal{V}$ is an open cover of X with* $\operatorname{diam}(\mathcal{V}) \le \varepsilon$*, then*

$$\overline{Q}(T,\varphi,\varepsilon) \le \overline{P}(T,\varphi,\varepsilon) \le \mathrm{P}(T,\varphi,\mathcal{V}).$$

We can then surmise new expressions for the topological pressure (cf. Corollary 7.3.12).

Corollary 11.2.6. *The following equalities hold:*

$$\mathrm{P}(T,\varphi) = \lim_{\varepsilon\to 0} \underline{P}(T,\varphi,\varepsilon) = \lim_{\varepsilon\to 0} \overline{P}(T,\varphi,\varepsilon) = \lim_{\varepsilon\to 0} \underline{Q}(T,\varphi,\varepsilon) = \lim_{\varepsilon\to 0} \overline{Q}(T,\varphi,\varepsilon).$$

Proof. Let $(\mathcal{U}_\varepsilon)_{\varepsilon\in(0,\infty)}$ be a family of open covers such that $\lim_{\varepsilon\to\infty} \operatorname{diam}(\mathcal{U}_\varepsilon) = 0$. Let δ_ε be a Lebesgue number for $\mathcal{U}_\varepsilon$. Then $\lim_{\varepsilon\to\infty} \delta_\varepsilon = 0$, as $\delta_\varepsilon \le \operatorname{diam}(\mathcal{U}_\varepsilon)$. Using Lemma 11.1.23 and Corollary 11.2.5(a), we deduce that

$$\mathrm{P}(T,\varphi) = \lim_{\varepsilon\to\infty} \underline{p}(T,\varphi,\mathcal{U}_\varepsilon) \le \lim_{\varepsilon\to 0} \underline{Q}(T,\varphi,\varepsilon) \le \lim_{\varepsilon\to 0} \underline{P}(T,\varphi,\varepsilon). \tag{11.8}$$

On the other hand, using Lemma 11.1.23 and Corollary 11.2.5(b), we obtain

$$\lim_{\varepsilon\to 0} \overline{Q}(T,\varphi,\varepsilon) \le \lim_{\varepsilon\to 0} \overline{P}(T,\varphi,\varepsilon) \le \lim_{\varepsilon\to 0} \sup_{\operatorname{diam}(\mathcal{V})\le\varepsilon} \mathrm{P}(T,\varphi,\mathcal{V}) = \mathrm{P}(T,\varphi). \tag{11.9}$$

Combining (11.8) and (11.9) allows us to conclude. □

Corollary 11.2.6 is useful to derive theoretical results. Nevertheless, in practice, Theorem 11.2.1 is simpler to use, as only one family of sets is needed. Sometimes a single sequence of sets is enough (cf. Theorem 7.3.13).

Theorem 11.2.7. *If a topological dynamical system $T : X \to X$ admits a generator with Lebesgue number 2δ, then the following statements hold for all $0 < \varepsilon \le \delta$:*

(a) *If $(E_n(\varepsilon))_{n=1}^{\infty}$ is a sequence of maximal (n,ε)-separated sets in X, then*

$$\mathrm{P}(T,\varphi) = \lim_{n\to\infty} \frac{1}{n} \log \Sigma_n(E_n(\varepsilon)).$$

(b) *If $(F_n(\varepsilon))_{n=1}^{\infty}$ is a sequence of minimal (n,ε)-spanning sets in X, then*

$$\mathrm{P}(T,\varphi) \le \liminf_{n\to\infty} \frac{1}{n} \log \Sigma_n(F_n(\varepsilon)).$$

(c) $\mathrm{P}(T,\varphi) = \lim_{n\to\infty} \frac{1}{n} \log P_n(T,\varphi,\varepsilon)$.

(d) $\mathrm{P}(T,\varphi) = \lim_{n\to\infty} \frac{1}{n} \log Q_n(T,\varphi,\varepsilon)$.

Proof. We will prove (a) and leave it to the reader to show the other parts using similar arguments.

Let $\mathcal{U}$ be a generator with Lebesgue number 2δ. Then $\mathrm{P}(T,\varphi) = \underline{p}(T,\varphi,\mathcal{U})$ by Lemma 11.1.25. Set $0 < \varepsilon \le \delta$. Observe that 2ε is also a Lebesgue number for $\mathcal{U}$. Choose any sequence $(E_n(\varepsilon))_{n=1}^{\infty}$ of maximal (n,ε)-separated sets. Since maximal (n,ε)-separated sets are (n,ε)-spanning sets, it follows from Lemma 11.2.4(a) that $z_n(T,\varphi,\mathcal{U}) \le Q_n(T,\varphi,\varepsilon) \le \Sigma_n(E_n(\varepsilon))$. Therefore,

$$\mathrm{P}(T,\varphi) = \underline{p}(T,\varphi,\mathcal{U}) = \liminf_{n\to\infty} \frac{1}{n} \log z_n(T,\varphi,\mathcal{U}) \le \liminf_{n\to\infty} \frac{1}{n} \log \Sigma_n(E_n(\varepsilon)). \tag{11.10}$$

On the other hand, since $\mathcal{U}$ is a generator, there exists $K \in \mathbb{N}$ such that $\operatorname{diam}(\mathcal{U}^k) \leq \varepsilon$ for all $k \geq K$. It ensues from Lemma 11.2.4(b) that $\Sigma_n(E_n(\varepsilon)) \leq P_n(T, \varphi, \varepsilon) \leq Z_n(T, \varphi, \mathcal{U}^k)$ for all $k \geq K$. Consequently,

$$\limsup_{n\to\infty} \frac{1}{n} \log \Sigma_n(E_n(\varepsilon)) \leq \lim_{n\to\infty} \frac{1}{n} \log Z_n(T, \varphi, \mathcal{U}^k) = \mathrm{P}(T, \varphi, \mathcal{U}^k)$$

for all $k \geq K$. It follows from Lemma 11.1.23(g) that

$$\limsup_{n\to\infty} \frac{1}{n} \log \Sigma_n(E_n(\varepsilon)) \leq \lim_{k\to\infty} \mathrm{P}(T, \varphi, \mathcal{U}^k) = \mathrm{P}(T, \varphi). \tag{11.11}$$

Combining (11.10) and (11.11) gives (a). □

Recall that for expansive systems, the Lebesgue number can be expressed in terms of the expansive constant.

Theorem 11.2.8. *If $T : X \to X$ is a δ_0-expansive dynamical system on a compact metric space (X, d), then Theorem* 11.2.7 *applies with any* $0 < \delta < \delta_0/4$.

Proof. See the proof of Theorem 7.3.14. □

11.3 Basic properties of topological pressure

In this section, we give some of the most basic properties of topological pressure. First, we show that the addition or subtraction of a constant to the potential increases or decreases the pressure of the potential by that same constant.

Proposition 11.3.1. *Let $T : X \to X$ be a topological dynamical system and $\varphi : X \to \mathbb{R}$ a potential. For any constant $c \in \mathbb{R}$, we have* $\mathrm{P}(T, \varphi + c) = \mathrm{P}(T, \varphi) + c$.

Proof. For each $n \in \mathbb{N}$ and $\varepsilon > 0$, let $E_n(\varepsilon)$ be a maximal (n, ε)-separated set. By Theorem 11.2.1,

$$\begin{aligned}
\mathrm{P}(T, \varphi + c) &= \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \sum_{x\in E_n(\varepsilon)} e^{S_n(\varphi+c)(x)} \\
&= \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)} e^{nc} \\
&= \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \left[\log\left(\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)} \right) + nc \right] \\
&= \mathrm{P}(T, \varphi) + c.
\end{aligned}$$
□

Next, we show that the pressure, as a function of the potential, is increasing.

Proposition 11.3.2. *Let $T : X \to X$ be a topological dynamical system and $\varphi, \psi : X \to \mathbb{R}$ be potentials. If $\varphi \leq \psi$, then* $\mathrm{P}(T, \varphi) \leq \mathrm{P}(T, \psi)$. *In particular,*

$$\mathrm{h}_{\mathrm{top}}(T) + \inf \varphi \leq \mathrm{P}(T, \varphi) \leq h_{top}(T) + \sup \varphi.$$

Proof. That $\mathrm{P}(T,\varphi) \le \mathrm{P}(T,\psi)$ whenever $\varphi \le \psi$ is obvious from Theorem 11.2.1. The second statement was proved in Corollary 11.1.24 but also follows from the first statement, Proposition 11.3.1, and the fact that $0+\inf\varphi \le \varphi \le 0+\sup\varphi$ and $\mathrm{P}(T,0) = \mathrm{h}_{\mathrm{top}}(T)$. □

In general, it is not the case that $\mathrm{P}(T,c\,\varphi) = c\mathrm{P}(T,\varphi)$. For example, suppose that $\mathrm{P}(T,0) \neq 0$. Then the equation $\mathrm{P}(T,c\,0) = c\,\mathrm{P}(T,0)$ only holds when $c = 1$.

11.4 Examples

Example 11.4.1. Let E be a finite alphabet and let $\sigma : E^\infty \to E^\infty$ be the full E-shift map. Let $\widetilde{\varphi} : E \to \mathbb{R}$ be a function. Then the function $\varphi : E^\infty \to \mathbb{R}$ defined by $\varphi(\omega) := \widetilde{\varphi}(\omega_1)$ is a continuous function on E^∞ which depends only upon the first coordinate ω_1 of the word $\omega \in E^\infty$. We will show that

$$\mathrm{P}(\sigma,\varphi) = \log\sum_{e\in E}\exp(\widetilde{\varphi}(e)).$$

According to Example 5.1.4, the shift map σ is δ-expansive for any $0 < \delta < 1$ when E^∞ is endowed with the metric $d_s(\omega,\tau) = s^{|\omega\wedge\tau|}$, where $0 < s < 1$. Choose $\mathcal{U} = \{[e] : e \in E\}$ as (finite) open cover of E^∞. So $\mathcal{U}$ is the partition of E^∞ into its initial 1-cylinders. Since $\mathrm{diam}(\mathcal{U}) = s < 1$, Theorem 11.1.26 states that $\mathrm{P}(\sigma,\varphi) = \mathrm{P}(\sigma,\varphi,\mathcal{U})$.

In order to compute $\mathrm{P}(\sigma,\varphi,\mathcal{U})$, observe that $\mathcal{U}^n = \{[\omega] : \omega \in E^n\}$ is the partition of E^∞ into its initial n-cylinders. Then

$$\begin{aligned}
\mathrm{P}(\sigma,\varphi) = \mathrm{P}(\sigma,\varphi,\mathcal{U}) &= \lim_{n\to\infty}\frac{1}{n}\log Z_n(\varphi,\mathcal{U}) = \lim_{n\to\infty}\frac{1}{n}\log\sum_{U\in\mathcal{U}^n} e^{\overline{S}_n\varphi(U)} \\
&= \lim_{n\to\infty}\frac{1}{n}\log\sum_{\omega\in E^n} e^{\overline{S}_n\varphi([\omega])} \\
&= \lim_{n\to\infty}\frac{1}{n}\log\sum_{\omega_1\ldots\omega_n\in E^n}\exp(\widetilde{\varphi}(\omega_1)+\cdots+\widetilde{\varphi}(\omega_n)) \\
&= \lim_{n\to\infty}\frac{1}{n}\log\Big(\sum_{\omega_1\in E}\exp(\widetilde{\varphi}(\omega_1))\cdots\sum_{\omega_n\in E}\exp(\widetilde{\varphi}(\omega_n))\Big) \\
&= \lim_{n\to\infty}\frac{1}{n}\log\Big(\sum_{e\in E}\exp(\widetilde{\varphi}(e))\Big)^n \\
&= \log\sum_{e\in E}\exp(\widetilde{\varphi}(e)).
\end{aligned}$$

Example 11.4.2. Let E be a finite alphabet and let $\sigma : E^\infty \to E^\infty$ be the full E-shift map. Let $\widetilde{\varphi} : E^2 \to \mathbb{R}$ be a function. Then the function $\varphi : E^\infty \to \mathbb{R}$ defined by $\varphi(\omega) = \widetilde{\varphi}(\omega_1,\omega_2)$ is a continuous function on E^∞ which depends only upon the first two coordinates of the word $\omega \in E^\infty$.

As in the previous example, $\mathrm{P}(\sigma,\varphi) = \mathrm{P}(\sigma,\varphi,\mathcal{U})$, where $\mathcal{U} = \{[e] : e \in E\}$ is the (finite) open partition of E^∞ into its initial 1-cylinders and

$$\mathrm{P}(\sigma,\varphi) = \mathrm{P}(\sigma,\varphi,\mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{\omega\in E^n} e^{\overline{S}_n\varphi([\omega])}.$$

But in this case

$$\begin{aligned}\sum_{\omega\in E^n} e^{\overline{S}_n\varphi([\omega])} &= \sum_{\omega\in E^n} \exp\begin{pmatrix} \widetilde{\varphi}(\omega_1,\omega_2) + \widetilde{\varphi}(\omega_2,\omega_3) + \cdots \\ + \widetilde{\varphi}(\omega_{n-1},\omega_n) + \max_{e\in E} \widetilde{\varphi}(\omega_n,e) \end{pmatrix} \\ &= \sum_{\omega_1\in E}\sum_{\omega_2\in E} e^{\widetilde{\varphi}(\omega_1,\omega_2)} \sum_{\omega_3\in E} e^{\widetilde{\varphi}(\omega_2,\omega_3)} \cdots \sum_{\omega_n\in E} e^{\widetilde{\varphi}(\omega_{n-1},\omega_n)} \cdot \max_{e\in E} \exp(\widetilde{\varphi}(\omega_n,e)).\end{aligned}$$

Since

$$m := \min_{e,f\in E} \exp(\widetilde{\varphi}(f,e)) \le \max_{e\in E} \exp(\widetilde{\varphi}(\omega_n,e)) \le \max_{e,f\in E} \exp(\widetilde{\varphi}(f,e)) =: M$$

for all $n \in \mathbb{N}$ and all $\omega_n \in E$, we have that

$$\sum_{\omega\in E^n} e^{\overline{S}_n\varphi([\omega])} \asymp \sum_{\omega_1\in E}\sum_{\omega_2\in E} e^{\widetilde{\varphi}(\omega_1,\omega_2)} \sum_{\omega_3\in E} e^{\widetilde{\varphi}(\omega_2,\omega_3)} \cdots \sum_{\omega_n\in E} e^{\widetilde{\varphi}(\omega_{n-1},\omega_n)}$$

for all n, with uniform constant of comparability $C = \max\{m^{-1}, M\}$.

Let $A : E^2 \to \mathbb{R}_+$ be the positive matrix whose entries are $A_{ef} = \exp(\widetilde{\varphi}(e,f))$. Equip this matrix with the norm $\|A\| = \sum_{e\in E}\sum_{f\in E} A_{ef}$. It is easy to prove by induction that

$$\|A^{n-1}\| = \sum_{\omega_1\in E}\sum_{\omega_2\in E} e^{\widetilde{\varphi}(\omega_1,\omega_2)} \sum_{\omega_3\in E} e^{\widetilde{\varphi}(\omega_2,\omega_3)} \cdots \sum_{\omega_n\in E} e^{\widetilde{\varphi}(\omega_{n-1},\omega_n)}$$

for all $n \ge 2$, and hence

$$\sum_{\omega\in E^n} e^{\overline{S}_n\varphi([\omega])} \asymp \|A^{n-1}\|.$$

Therefore,

$$\begin{aligned}\mathrm{P}(T,\varphi) &= \lim_{n\to\infty} \frac{1}{n} \log \sum_{\omega\in E^n} e^{\overline{S}_n\varphi([\omega])} \\ &= \lim_{n\to\infty} \frac{1}{n} \log\|A^{n-1}\| = \log \lim_{n\to\infty} \|A^n\|^{1/n} = \log r(A),\end{aligned}$$

where $r(A)$ is the spectral radius of A, that is, the largest eigenvalue of A (in absolute value).

11.5 Exercises

Exercise 11.5.1. Let $T : X \to X$ be a dynamical system and $\mathcal{U}$ be an open cover of X. Show that $Z_n(\varphi,\mathcal{U}) \neq Z_1(\varphi,\mathcal{U}^n)$ in general. That is, find a potential φ such that $Z_n(\varphi,\mathcal{U}) \neq Z_1(\varphi,\mathcal{U}^n)$ for some $n \in \mathbb{N}$.

Note: It is possible to find a potential for which the above nonequality holds for any $n > 1$.

Exercise 11.5.2. Using a symbolic dynamical system, give an example of a sequence $(z_n(\mathcal{U}))_{n=1}^{\infty}$ which is not submultiplicative.

Exercise 11.5.3. Prove Lemma 11.1.8.

Exercise 11.5.4. Show that for every $t \geq 0$ there exists a dynamical system $T : X \to X$ whose topological entropy is equal to t.

Exercise 11.5.5. Consider the full shift $\sigma : \{0,1\}^{\infty} \to \{0,1\}^{\infty}$. Let $\varphi : \{0,1\}^{\infty} \to \mathbb{R}$ be given by the formula

$$\varphi(\omega_1\omega_2\ldots) := \begin{cases} -\log 4 & \text{if } \omega_1 = 0 \\ \log 3 - \log 4 & \text{if } \omega_1 = 1. \end{cases}$$

Show that $\mathrm{P}(\sigma,\varphi) = 0$.

Exercise 11.5.6. Let $T : X \to X$ be a dynamical system. Show that the following are equivalent:
(a) $\mathrm{h}_{\mathrm{top}}(T)$ is finite.
(b) There exists a continuous function $\varphi : X \to \mathbb{R}$ such that $\mathrm{P}(T,\varphi)$ is finite.
(c) $\mathrm{P}(T,\varphi)$ is finite for every continuous function $\varphi : X \to \mathbb{R}$.

Exercise 11.5.7. Let $T : X \to X$ be a dynamical system such that $\mathrm{h}_{\mathrm{top}}(T) < \infty$. Prove that the topological pressure function $\mathrm{P}(T,\bullet) : C(X) \to \mathbb{R}$ is Lipschitz continuous with Lipschitz constant 1, and convex.

Exercise 11.5.8. Generalize Examples 11.4.1 and 11.4.2 to the case where $k \in \mathbb{N}$ and φ depends on k coordinates.

Exercise 11.5.9. Show that the pressure function is not linear; more precisely, in general $\mathrm{P}(T,t\varphi) \neq t\mathrm{P}(T,\varphi)$.

Exercise 11.5.10. Show that if $\inf\varphi \leq 0$, then the pressure function $\mathbb{R} \ni t \mapsto \mathrm{P}(T,t\varphi)$ is convex, that is,

$$\mathrm{P}\big(T,(st_1 + (1-s)t_2)\varphi\big) \leq s\,\mathrm{P}(T,t_1\varphi) + (1-s)\,\mathrm{P}(T,t_2\varphi), \quad \forall s \in [0,1],\ \forall t_1,t_2 \in \mathbb{R}.$$

Conclude that the function $\mathbb{R} \ni t \mapsto \mathrm{P}(T,t\varphi)$ is differentiable at all but at most countably many t's.

12 The variational principle and equilibrium states

In Section 12.1, we state and prove a fundamental result of thermodynamic formalism known as the *variational principle*. This deep result establishes a crucial relationship between topological dynamics and ergodic theory, by way of a formula linking topological pressure and measure-theoretic entropy. The variational principle in its classical form and full generality was proved in [75] and [10]. The proof we present follows that of Michal Misiurewicz [49], which is particularly elegant, short, and simple.

In Section 12.2, we introduce the concept of equilibrium states, give sufficient conditions for their existence, such as the upper semicontinuity of the metric entropy function (which prevails under any expansive system). We single out a special class of equilibrium states, those corresponding to a potential identically equal to zero, and following tradition, call them measures of maximal entropy. We do not deal in this chapter with the issue of the uniqueness of equilibrium states. Nevertheless, we provide an example of a topological dynamical system with positive and finite topological entropy which does not have any measure of maximal entropy.

12.1 The variational principle

For any topological dynamical system $T : X \to X$, subject to a potential $\varphi : X \to \mathbb{R}$ and equipped with a T-invariant measure μ, the quantity $\mathrm{h}_\mu(T) + \int \varphi \, d\mu$ is called the *free energy* of the system T with respect to μ under the potential φ. The variational principle states that the topological pressure of a system is the supremum of the free energy generated by that system.

Recall that $M(T)$ is the set of all T-invariant Borel probability measures on X and that by definition any potential φ is continuous.

Theorem 12.1.1 (Variational principle). *Let $T : X \to X$ be a topological dynamical system and $\varphi : X \to \mathbb{R}$ a potential. Then*

$$\mathrm{P}(T, \varphi) = \sup\left\{\mathrm{h}_\mu(T) + \int_X \varphi \, d\mu : \mu \in M(T)\right\}.$$

Remark 12.1.2. In fact, as we shall see in Corollary 12.1.10, the supremum can be restricted to the subset $E(T)$ of ergodic measures in $M(T)$.

The proof of the variational principle will be given in two parts. In Part I, we will show that $\mathrm{P}(T, \varphi) \geq \mathrm{h}_\mu(T) + \int \varphi \, d\mu$ for every measure $\mu \in M(T)$. Part II consists in the proof of the inequality $\sup\{\mathrm{h}_\mu(T) + \int \varphi \, d\mu : \mu \in M(T)\} \geq \mathrm{P}(T, \varphi)$.

The first part is relatively easier to prove than the second one. In the proof of Part I, we will need Jensen's inequality. Recall that a function $k : I \to \mathbb{R}$, where $I \subseteq \mathbb{R}$ is an

https://doi.org/10.1515/9783110702682-012

interval, is convex on I if

$$\psi(tx + (1-t)y) \le t\psi(x) + (1-t)\psi(y), \quad \forall t \in [0,1], \forall x, y \in I.$$

Theorem 12.1.3 (Jensen's inequality). *Let $(X, \mathcal{A}, \mu)$ be a probability space. Let $-\infty \le a < b \le \infty$ and $\psi : (a,b) \to \mathbb{R}$ be a convex function. If $f \in L^1(\mu)$ and $f(X) \subseteq (a,b)$, then*

$$\psi\Big(\int_X f\, d\mu\Big) \le \int_X \psi \circ f\, d\mu.$$

Proof. See, for instance, [58]. □

We shall also need the following lemma, which states that any finite Borel partition α of X can be, from a measure-theoretic entropy viewpoint, approximated as closely as desired by a finite Borel partition β whose elements are compact and are, with one exception, contained in those of α.

Lemma 12.1.4. *Let $\mu \in M(X)$, let $\alpha := \{A_1, \ldots, A_n\}$ be a finite Borel partition of X, and let $\varepsilon > 0$. Then there exist compact sets $B_i \subseteq A_i$, $1 \le i \le n$, such that the partition $\beta := \{B_1, \ldots, B_n, X \backslash (B_1 \cup \cdots \cup B_n)\}$ satisfies*

$$\mathrm{H}_\mu(\alpha|\beta) \le \varepsilon.$$

Proof. Let the measure μ and the partition α be as stated and let $\varepsilon > 0$. Recall from Definition 9.3.4 the nonnegative continuous function $k : [0,1] \to [0,1]$ defined by

$$k(t) = -t \log t,$$

where it is understood that $0 \cdot (-\infty) = 0$. The continuity of k at 0 implies that there exists $\delta > 0$ such that $k(t) < \varepsilon/n$ when $0 \le t < \delta$. Since μ is regular and X is compact, for each $1 \le i \le n$ there exists a compact set $B_i \subseteq A_i$ such that $\mu(A_i \backslash B_i) < \delta$. Then $k(\mu(A_i \backslash B_i)) < \varepsilon/n$ for all $1 \le i \le n$. Observe further that $X \backslash \bigcup_{j=1}^n B_j = \bigcup_{j=1}^n A_j \backslash B_j$. By Definition 9.4.2 of conditional entropy, it follows that

$$\begin{aligned}
\mathrm{H}_\mu(\alpha|\beta) &= \sum_{j=1}^n \sum_{i=1}^n -\mu(A_i \cap B_j) \log \frac{\mu(A_i \cap B_j)}{\mu(B_j)} \\
&\quad + \sum_{i=1}^n -\mu(A_i \cap (X \backslash \cup_{j=1}^n B_j)) \log \frac{\mu(A_i \cap (X \backslash \cup_{j=1}^n B_j))}{\mu(X \backslash \cup_{j=1}^n B_j)} \\
&= \sum_{j=1}^n -\mu(B_j) \log \frac{\mu(B_j)}{\mu(B_j)} + \sum_{i=1}^n -\mu(A_i \cap (\cup_{j=1}^n A_j \backslash B_j)) \log \frac{\mu(A_i \cap (\cup_{j=1}^n A_j \backslash B_j))}{\mu(\cup_{j=1}^n A_j \backslash B_j)} \\
&= 0 + \sum_{i=1}^n -\mu(A_i \backslash B_i) \log \frac{\mu(A_i \backslash B_i)}{\mu(\cup_{j=1}^n A_j \backslash B_j)}
\end{aligned}$$

$$= \sum_{i=1}^{n} -\mu(A_i \backslash B_i)\left[\log \mu(A_i \backslash B_i) - \log \mu\left(\cup_{j=1}^{n} A_j \backslash B_j\right)\right]$$
$$= \sum_{i=1}^{n} k\left(\mu(A_i \backslash B_i)\right) + \sum_{i=1}^{n} \mu(A_i \backslash B_i) \log \mu\left(\cup_{j=1}^{n} A_j \backslash B_j\right)$$
$$\leq \sum_{i=1}^{n} k\left(\mu(A_i \backslash B_i)\right) \leq n \cdot \frac{\varepsilon}{n} = \varepsilon. \qquad \square$$

We are now in a position to begin the proof of the first part of the variational principle.

Proof of Part I. Recall that our aim is to establish the inequality

$$\mathrm{P}(T, \varphi) \geq \mathrm{h}_\mu(T) + \int_X \varphi \, d\mu, \quad \forall \mu \in M(T). \tag{12.1}$$

We claim that it is sufficient to prove that there exists a constant $C \in \mathbb{R}$, independent of T, φ and μ, such that

$$\mathrm{P}(T, \varphi) \geq \mathrm{h}_\mu(T) + \int_X \varphi \, d\mu + C. \tag{12.2}$$

Indeed, suppose that such a constant exists. In particular, this means that this constant works not only for the system (X, T) under the potential φ and a measure $\mu \in M(T)$ but also for any higher-iterate system (X, T^n) under the potential $S_n\varphi = \sum_{k=0}^{n-1} \varphi \circ T^k$ and the same measure μ, since any T-invariant measure is T^n-invariant. Fix temporarily $n \in \mathbb{N}$. Using successively Theorem 11.1.22, inequality (12.2) with the quadruple $(X, T^n, S_n\varphi, \mu)$ instead of (X, T, φ, μ), and Theorems 9.4.13 and 8.1.18, we then obtain that

$$n\,\mathrm{P}(T, \varphi) = \mathrm{P}(T^n, S_n\varphi) \geq \mathrm{h}_\mu(T^n) + \int_X S_n\varphi \, d\mu + C = n\,\mathrm{h}_\mu(T) + n \int_X \varphi \, d\mu + C.$$

Dividing by n and letting n tend to infinity yields inequality (12.1).

Of course, to obtain (12.2) it suffices to show that

$$\mathrm{P}(T, \varphi) \geq \mathrm{h}_\mu(T, \alpha) + \int_X \varphi \, d\mu + C \tag{12.3}$$

for all finite Borel partitions α of X (see Definition 9.4.12). So let α be any such partition and let $\varepsilon > 0$. To obtain (12.3), it is enough to prove that

$$\mathrm{P}(T, \varphi) \geq \mathrm{h}_\mu(T, \alpha) + \int_X \varphi \, d\mu + C - 2\varepsilon. \tag{12.4}$$

By Theorem 11.2.1, it suffices to demonstrate that

$$\limsup_{n\to\infty}\frac{1}{n}\log\sum_{y\in E_n(\delta)}e^{S_n\varphi(y)}\geq \mathrm{h}_\mu(T,\alpha)+\int_X\varphi\,d\mu+C-2\varepsilon \tag{12.5}$$

for all sufficiently small $\delta>0$ and some family $\{E_n(\delta):n\in\mathbb{N},\delta>0\}$ of (n,δ)-separated sets. In light of Definition 9.4.10 and of Theorem 8.1.18, it is sufficient to prove that

$$\frac{1}{n}\log\sum_{y\in E_n(\delta)}e^{S_n\varphi(y)}\geq\frac{1}{n}\mathrm{H}_\mu(\alpha^n)+\frac{1}{n}\int_X S_n\varphi\,d\mu+C-2\varepsilon \tag{12.6}$$

for all sufficiently small $\delta>0$, all large enough $n\in\mathbb{N}$ and all (n,δ)-separated sets $E_n(\delta)$.

To this end, let β be the finite Borel partition given by Lemma 12.1.4. Then $\mathrm{H}_\mu(\alpha|\beta)\leq\varepsilon$. Momentarily fix $n\in\mathbb{N}$. By Theorem 9.4.3(i) and Lemma 9.4.7(c), we know that

$$\mathrm{H}_\mu(\alpha^n)\leq\mathrm{H}_\mu(\beta^n)+\mathrm{H}_\mu(\alpha^n|\beta^n)\leq\mathrm{H}_\mu(\beta^n)+n\,\mathrm{H}_\mu(\alpha|\beta)\leq\mathrm{H}_\mu(\beta^n)+n\varepsilon. \tag{12.7}$$

From (12.6) and (12.7), it thus suffices to establish that

$$\log\sum_{y\in E_n(\delta)}e^{S_n\varphi(y)}\geq\mathrm{H}_\mu(\beta^n)+\int_X S_n\varphi\,d\mu+(C-\varepsilon)n \tag{12.8}$$

for all sufficiently small $\delta>0$, all large enough $n\in\mathbb{N}$ and all (n,δ)-separated sets $E_n(\delta)$. To prove this inequality, we will estimate the term $\mathrm{H}_\mu(\beta^n)+\int S_n\varphi\,d\mu$ from above. Since the logarithm function is concave (so its negative is convex), Jensen's inequality (Theorem 12.1.3) implies that

$$\begin{aligned}\mathrm{H}_\mu(\beta^n)+\int_X S_n\varphi\,d\mu &\leq\sum_{B\in\beta^n}\mu(B)\big[-\log\mu(B)+S_n\varphi(B)\big]\\ &=\sum_{B\in\beta^n}\mu(B)\log\frac{\exp(S_n\varphi(B))}{\mu(B)}\\ &=\int_X\log\frac{\exp(S_n\varphi(\beta^n(x)))}{\mu(\beta^n(x))}\,d\mu(x)\\ &\leq\log\int_X\frac{\exp(S_n\varphi(\beta^n(x)))}{\mu(\beta^n(x))}\,d\mu(x)\\ &=\log\sum_{B\in\beta^n}e^{S_n\varphi(B)}.\end{aligned} \tag{12.9}$$

Since each set $B_i\in\beta$ is compact, it follows that $d(B_i,B_j)>0$ for all $i\neq j$. As φ is uniformly continuous, let $0<\delta<\frac{1}{2}\min\{d(B_i,B_j):i\neq j\}$ be such that

$$d(x,y)<\delta\Longrightarrow|\varphi(x)-\varphi(y)|<\varepsilon. \tag{12.10}$$

Now, consider an arbitrary maximal (n, δ)-separated set $E_n(\delta)$ and fix temporarily $B \in \beta$. According to Lemma 7.3.7, each maximal (n, δ)-separated set is an (n, δ)-spanning set. So for every $x \in B$, there exists $y \in E_n(\delta)$ such that $x \in B_n(y, \delta)$ and, therefore, $|S_n\varphi(x) - S_n\varphi(y)| < n\varepsilon$ by (12.10). As the set $E_n(\delta)$ is finite, there is $y_B \in E_n(\delta)$ such that

$$S_n\varphi(B) \le S_n\varphi(y_B) + n\varepsilon \quad \text{and} \quad B \cap B_n(y_B, \delta) \ne \emptyset. \tag{12.11}$$

Moreover, since $d(B_i, B_j) > 2\delta$ for each $i \ne j$, any ball $B(z, \delta)$, $z \in X$, intersects at most one B_i and perhaps $X \setminus \bigcup_j B_j$. Hence,

$$\#\{B \in \beta : B \cap B(z, \delta) \ne \emptyset\} \le 2 \tag{12.12}$$

for all $z \in X$. Thus,

$$\#\{B \in \beta^n : B \cap B_n(z, \delta) \ne \emptyset\} \le 2^n \tag{12.13}$$

for all $z \in X$. So the function $f : \beta^n \to E_n(\delta)$ defined by $f(B) = y_B$ is at most 2^n-to-one. Consequently, by (12.11) we obtain that

$$2^n \sum_{y \in E_n(\delta)} e^{S_n\varphi(y)} \ge \sum_{B \in \beta^n} e^{S_n\varphi(y_B)} \ge \sum_{B \in \beta^n} e^{S_n\varphi(B)} \cdot e^{-n\varepsilon}.$$

Multiplying both sides by 2^{-n}, then taking the logarithm of both sides and applying (12.9) yields

$$\begin{aligned} \log \sum_{y \in E_n(\delta)} e^{S_n\varphi(y)} &\ge \log \sum_{B \in \beta^n} e^{S_n\varphi(B)} - n\varepsilon - n\log 2 \\ &\ge \mathrm{H}_\mu(\beta^n) + \int_X S_n\varphi \, d\mu + n(-\log 2 - \varepsilon). \end{aligned}$$

This inequality, which is nothing other than the sought inequality (12.8) with $C = -\log 2$, holds for all $0 < \delta < \frac{1}{2}\min\{d(B_i, B_j) : i \ne j\}$, all $n \in \mathbb{N}$ and all maximal (n, δ)-separated sets $E_n(\delta)$. This concludes the proof of Part I. □

Remark 12.1.5. Observe that the constant $C = -\log 2$ originates from relation (12.12), and thus depends solely on the **existence** of the Borel partition β, which is ensured by Lemma 12.1.4.

Let us move on to the proof of Part II of the variational principle. In addition to Lemma 9.6.1, we shall need the following three lemmas.

The first of those states that given any finite Borel partition α whose atoms have boundaries with zero μ-measure, the entropy of α, as a function of the underlying Borel probability measure, is continuous at μ.

Lemma 12.1.6. *Let $\mu \in M(X)$. If α is a finite Borel partition of X such that $\mu(\partial A) = 0$ for all $A \in \alpha$, then the function*

$$\begin{array}{rccc} \mathrm{H}_{\bullet}(\alpha) & : & M(X) & \longrightarrow & [0,\infty] \\ & & \nu & \longmapsto & \mathrm{H}_{\nu}(\alpha) \end{array}$$

is continuous at μ.

Proof. This follows directly from the fact that according to the Portmanteau theorem (Theorem A.1.56), a sequence of Borel probability measures $(\mu_n)_{n=1}^{\infty}$ converges weakly* to a measure μ if and only if $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for every Borel set A with $\mu(\partial A) = 0$. Recall that $\mathrm{H}_{\nu}(\alpha) = -\sum_{A\in\alpha} \nu(A)\log\nu(A)$. □

In the second lemma, we show that the entropy of α, as a function of the underlying Borel probability measure, is concave.

Lemma 12.1.7. *For any finite Borel partition α of X, the function $\mathrm{H}_{\bullet}(\alpha)$ is concave.*

Proof. Let α be a finite Borel partition of X, and μ and ν Borel probability measures on X. Let also $t \in (0,1)$. Since the function $k(x) = -x\log x$ is concave, for each $A \in \alpha$ we have

$$k\big(t\mu(A) + (1-t)\nu(A)\big) \geq t\,k\big(\mu(A)\big) + (1-t)k\big(\nu(A)\big).$$

Therefore,

$$\begin{aligned} \mathrm{H}_{t\mu+(1-t)\nu}(\alpha) &= \sum_{A\in\alpha} k\big(t\mu(A) + (1-t)\nu(A)\big) \\ &\geq t\sum_{A\in\alpha} k\big(\mu(A)\big) + (1-t)\sum_{A\in\alpha} k\big(\nu(A)\big) \\ &= t\mathrm{H}_{\mu}(\alpha) + (1-t)\mathrm{H}_{\nu}(\alpha). \end{aligned}$$ □

Finally, the third lemma is a generalization of the Krylov–Bogolyubov theorem (Theorem 8.1.22).

Lemma 12.1.8. *Let $T : X \to X$ be a dynamical system. If $(\mu_n)_{n=1}^{\infty}$ is a sequence of measures in $M(X)$, then every* weak* *limit point of the sequence $(m_n)_{n=1}^{\infty}$, where*

$$m_n := \frac{1}{n}\sum_{i=0}^{n-1} \mu_n \circ T^{-i},$$

is a T-invariant measure.

Proof. By the compactness of $M(X)$, the sequence $(m_n)_{n=1}^{\infty}$ has accumulation points. Let $(m_{n_j})_{j=1}^{\infty}$ be a subsequence which converges weakly* to, say, $m \in M(X)$. Let $f \in C(X)$.

Using Lemma 8.1.2, we obtain that

$$\begin{aligned}\left|\int_X f\circ T\,dm-\int_X f\,dm\right| &= \lim_{j\to\infty}\left|\int_X f\circ T\,dm_{n_j}-\int_X f\,dm_{n_j}\right| \\ &= \lim_{j\to\infty}\left|\frac{1}{n_j}\int_X\sum_{i=0}^{n_j-1}(f\circ T^{i+1}-f\circ T^i)\,d\mu_{n_j}\right| \\ &= \lim_{j\to\infty}\frac{1}{n_j}\left|\int_X(f\circ T^{n_j}-f)\,d\mu_{n_j}\right| \\ &\le \lim_{j\to\infty}\frac{2\|f\|_\infty}{n_j}=0.\end{aligned}$$

Thus, by Theorem 8.1.18 the measure m is T-invariant. □

We are now ready to prove Part II of the variational principle.

Proof of Part II. We aim to show that

$$\sup\left\{\mathrm{h}_\mu(T)+\int_X\varphi\,d\mu:\mu\in M(T)\right\}\ge \mathrm{P}(T,\varphi).$$

Fix $\varepsilon>0$. Let $(E_n(\varepsilon))_{n=1}^\infty$ be a sequence of maximal (n,ε)-separated sets in X. For every $n\in\mathbb{N}$, define the measures μ_n and m_n by

$$\mu_n:=\frac{\sum_{x\in E_n(\varepsilon)}e^{S_n\varphi(x)}\delta_x}{\sum_{x\in E_n(\varepsilon)}e^{S_n\varphi(x)}}\quad\text{and}\quad m_n:=\frac{1}{n}\sum_{k=0}^{n-1}\mu_n\circ T^{-k},$$

where δ_x denotes the Dirac measure concentrated at the point x. Let $(n_i)_{i=1}^\infty$ be a strictly increasing sequence in $\mathbb{N}$ such that $(m_{n_i})_{i=1}^\infty$ converges weakly* to, say, m, and such that

$$\lim_{i\to\infty}\frac{1}{n_i}\log\sum_{x\in E_{n_i}(\varepsilon)}e^{S_n\varphi(x)}=\limsup_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)}e^{S_n\varphi(x)}. \tag{12.14}$$

For ease of exposition, define

$$s_n:=\sum_{x\in E_n(\varepsilon)}e^{S_n\varphi(x)}\quad\text{and}\quad \mu(x):=\mu(\{x\}).$$

From Lemma 12.1.8, the limit measure m belongs to $M(T)$. Also, in view of Lemma 9.6.1, there exists a finite Borel partition α such that $\mathrm{diam}(\alpha)<\varepsilon$ and $m(\partial A)=0$ for all $A\in\alpha$. Since $\#(A\cap E_n(\varepsilon))\le 1$ for all $A\in\alpha^n$, we obtain that

$$\begin{aligned}\mathrm{H}_{\mu_n}(\alpha^n)+\int_X S_n\varphi\,d\mu_n &= \sum_{x\in E_n(\varepsilon)}\mu_n(x)[-\log\mu_n(x)+S_n\varphi(x)] \\ &= \sum_{x\in E_n(\varepsilon)}\frac{e^{S_n\varphi(x)}}{s_n}\left[-\log\frac{e^{S_n\varphi(x)}}{s_n}+S_n\varphi(x)\right]\end{aligned}$$

$$= \frac{1}{s_n} \sum_{x \in E_n(\varepsilon)} e^{S_n\varphi(x)} [-S_n\varphi(x) + \log s_n + S_n\varphi(x)]$$
$$= \log s_n = \log \sum_{x \in E_n(\varepsilon)} e^{S_n\varphi(x)}. \tag{12.15}$$

Now, fix $M \in \mathbb{N}$ and let $n \geq 2M$. For $j = 0, 1, \ldots, M-1$, define $s(j) := \lfloor \frac{n-j}{M} \rfloor - 1$, where $\lfloor r \rfloor$ denotes the integer part of r. Note that

$$\bigvee_{k=0}^{s(j)} T^{-(kM+j)}(\alpha^M) = \bigvee_{\ell=j}^{(s(j)+1)M+j-1} T^{-\ell}(\alpha)$$

and

$$(s(j)+1)M + j - 1 = \left\lfloor \frac{n-j}{M} \right\rfloor M + j - 1 \leq n - j + j - 1 = n - 1.$$

Observe also that

$$\begin{aligned}(n-1) - ((s(j)+1)M + j) &= n - 1 - \left(\left\lfloor \frac{n-j}{M} \right\rfloor M + j \right) \\ &\leq n - 1 - \left(\frac{n-j}{M} - 1 \right) M - j = M - 1.\end{aligned}$$

Setting $R_j := \{0, 1, \ldots, j-1\} \cup \{(s(j)+1)M + j, \ldots, n-1\}$, we have $\#R_j \leq 2M$ and

$$\alpha^n = \bigvee_{k=0}^{s(j)} T^{-(kM+j)}(\alpha^M) \vee \bigvee_{i \in R_j} T^{-i}(\alpha).$$

Hence, using Theorem 9.4.3(g) and (9.2), we get that

$$\begin{aligned}\mathrm{H}_{\mu_n}(\alpha^n) &\leq \sum_{k=0}^{s(j)} \mathrm{H}_{\mu_n}(T^{-(kM+j)}(\alpha^M)) + \mathrm{H}_{\mu_n}\Big(\bigvee_{i \in R_j} T^{-i}(\alpha) \Big) \\ &\leq \sum_{k=0}^{s(j)} \mathrm{H}_{\mu_n \circ T^{-(kM+j)}}(\alpha^M) + \log \# \Big(\bigvee_{i \in R_j} T^{-i}(\alpha) \Big) \\ &\leq \sum_{k=0}^{s(j)} \mathrm{H}_{\mu_n \circ T^{-(kM+j)}}(\alpha^M) + \log(\#\alpha)^{\#R_j} \\ &\leq \sum_{k=0}^{s(j)} \mathrm{H}_{\mu_n \circ T^{-(kM+j)}}(\alpha^M) + 2M \log \#\alpha.\end{aligned}$$

Summing over all $j = 0, 1, \ldots, M-1$ and using Lemma 12.1.7, we obtain

$$\begin{aligned}M\mathrm{H}_{\mu_n}(\alpha^n) &\leq \sum_{j=0}^{M-1} \sum_{k=0}^{s(j)} \mathrm{H}_{\mu_n \circ T^{-(kM+j)}}(\alpha^M) + 2M^2 \log \#\alpha \\ &\leq \sum_{l=0}^{n-1} \mathrm{H}_{\mu_n \circ T^{-l}}(\alpha^M) + 2M^2 \log \#\alpha\end{aligned}$$

$$\le n\mathrm{H}_{\frac{1}{n}\sum_{l=0}^{n-1}\mu_n\circ T^{-l}}(\alpha^M) + 2M^2\log\#\alpha$$
$$= n\mathrm{H}_{m_n}(\alpha^M) + 2M^2\log\#\alpha.$$

Adding $M\int_X S_n\varphi\,d\mu_n$ to both sides and applying (12.15) yields

$$M\log\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)} \le n\mathrm{H}_{m_n}(\alpha^M) + M\int_X S_n\varphi\,d\mu_n + 2M^2\log\#\alpha.$$

As $\frac{1}{n}\int_X S_n\varphi\,d\mu_n = \int_X \varphi\,dm_n$ by Lemma 8.1.2, dividing both sides of the above inequality by Mn gives us that

$$\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)} \le \frac{1}{M}\mathrm{H}_{m_n}(\alpha^M) + \int_X \varphi\,dm_n + \frac{2M}{n}\log\#\alpha.$$

Since $\partial T^{-1}(A) \subseteq T^{-1}(\partial A)$ for every set $A\subseteq X$ and $\partial(A\cap B)\subset\partial A\cup\partial B$ for all sets $A,B\in X$, the m-measure of the boundary of each atom of the partition α^M is, as for α, equal to zero. Therefore, upon letting n tend to infinity along the subsequence $(n_i)_{i=1}^\infty$, we know that $(m_{n_i})_{i=1}^\infty$ converges weakly* to m and that (12.14) holds, so we infer from the above inequality and from Lemma 12.1.6 that

$$\limsup_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)} \le \frac{1}{M}\mathrm{H}_m(\alpha^M) + \int_X \varphi\,dm.$$

Letting $M\to\infty$, we obtain by Definition 9.4.10 that

$$\limsup_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)} \le \mathrm{h}_m(T,\alpha) + \int_X \varphi\,dm \le \sup\left\{\mathrm{h}_\mu(T) + \int_X \varphi\,d\mu : \mu\in M(T)\right\}.$$

As $\varepsilon > 0$ is arbitrary, Theorem 11.2.1 yields the desired inequality. This completes the proof of Part II. □

12.1.1 Consequences of the variational principle

Let us now state some immediate consequences of the variational principle. A first consequence concerns the topological entropy of the system. The topological entropy of a system is the supremum of all measure-theoretic entropies of the system.

Corollary 12.1.9. $\mathrm{h}_{\mathrm{top}}(T) = \sup\{\mathrm{h}_\mu(T) : \mu\in M(T)\}$.

Proof. This follows directly upon letting $\varphi\equiv 0$. □

Furthermore, the pressure of the system is determined by the supremum of the free energy of the system with respect to its ergodic measures. Recall that $E(T)$ denotes the subset of ergodic measures in $M(T)$.

Corollary 12.1.10. *For every $\mu \in M(T)$, there exists $\nu \in E(T)$ such that* $\mathrm{h}_\nu(T) + \int_X \varphi\, d\nu \geq \mathrm{h}_\mu(T) + \int_X \varphi\, d\mu$. *Consequently,*

$$\mathrm{P}(T, \varphi) = \sup\Big\{\mathrm{h}_\nu(T) + \int_X \varphi\, d\nu : \nu \in E(T)\Big\}.$$

Proof. Let $\mu \in M(T)$. According to Theorem 8.2.26, the measure μ has a decomposition into ergodic measures. More precisely, there exists a Borel probability space $(Y, \mathcal{B}(Y), \tau)$ and a measurable map $Y \ni y \mapsto \mu_y \in M(X)$ such that $\mu_y \in E(T)$ for τ-almost every $y \in Y$ and $\mu = \int_Y \mu_y\, d\tau(y)$. Then

$$\int_X \varphi\, d\mu = \int_Y \Big(\int_X \varphi\, d\mu_y\Big)\, d\tau(y).$$

Moreover, using a generalization of Exercise 9.7.10, we have that

$$\mathrm{h}_\mu(T) = \mathrm{h}_{\int_Y \mu_y\, d\tau(y)}(T) = \int_Y \mathrm{h}_{\mu_y}(T)\, d\tau(y).$$

It follows that

$$\mathrm{h}_\mu(T) + \int_X \varphi\, d\mu = \int_Y \Big[\mathrm{h}_{\mu_y}(T) + \int_X \varphi\, d\mu_y\Big]\, d\tau(y).$$

It is a well-known fact from measure theory (a simple consequence of Lemma A.1.34(a)) that there is $Z \in \mathcal{B}(Y)$ such that $\tau(Z) > 0$ and $\mathrm{h}_{\mu_z}(T) + \int_X \varphi\, d\mu_z \geq \mathrm{h}_\mu(T) + \int_X \varphi\, d\mu$ for every $z \in Z$. Given that $\tau(E(T)) = 1$, it follows that $\tau(Z \cap E(T)) > 0$. So there exists $\nu \in E(T)$ such that $\mathrm{h}_\nu(T) + \int_X \varphi\, d\nu \geq \mathrm{h}_\mu(T) + \int_X \varphi\, d\mu$. □

We now show that the pressure function is Lipschitz continuous.

Corollary 12.1.11. *If $T : X \to X$ is a dynamical system such that* $\mathrm{h}_{\mathrm{top}}(T) < \infty$, *then the pressure function* $\mathrm{P}(T, \bullet) : C(X) \to \mathbb{R}$ *is Lipschitz continuous with Lipschitz constant* 1.

Proof. Let $\psi, \varphi \in C(X)$. Let also $\varepsilon > 0$. By the variational principle, there exists $\mu \in M(T)$ such that

$$\mathrm{P}(T, \psi) \leq \mathrm{h}_\mu(T) + \int_X \psi\, d\mu + \varepsilon.$$

Then, using the variational principle once again, we get

$$\mathrm{P}(T, \psi) \leq \mathrm{h}_\mu(T) + \int_X \varphi\, d\mu + \int_X (\psi - \varphi)\, d\mu + \varepsilon \leq \mathrm{P}(T, \varphi) + \|\psi - \varphi\|_\infty + \varepsilon.$$

Since this is true for all $\varepsilon > 0$, we conclude that

$$\mathrm{P}(T, \psi) - \mathrm{P}(T, \varphi) \leq \|\psi - \varphi\|_\infty. \qquad \square$$

Finally, we show that the pressure of any subsystem is at most the pressure of the entire system.

Corollary 12.1.12. *If $T : X \to X$ is a topological dynamical system, $\varphi : X \to \mathbb{R}$ a potential and Y a closed T-invariant subset of X, then* $\mathrm{P}(T|_Y, \varphi|_Y) \le \mathrm{P}(T, \varphi)$.

Proof. Each $T|_Y$-invariant measure μ on Y generates the T-invariant measure $\overline{\mu}(B) = \mu(B \cap Y)$ on X and $\overline{\mu}$ is such that $\mathrm{h}(T, \overline{\mu}) = \mathrm{h}(T|_Y, \mu)$ and $\int_X \varphi \, d\overline{\mu} = \int_Y \varphi \, d\mu$. □

12.2 Equilibrium states

In light of the variational principle, the measures that maximize the free energy of the system, that is, the measures which respect to which the free energy of the system coincides with its pressure, are given a special name.

Definition 12.2.1. Let $T : X \to X$ be a topological dynamical system and $\varphi : X \to \mathbb{R}$ a potential. A measure $\mu \in M(T)$ is called an *equilibrium state* for φ provided that

$$\mathrm{P}(T, \varphi) = \mathrm{h}_\mu(T) + \int_X \varphi \, d\mu.$$

Notice that if a given potential φ has an equilibrium state, then φ has an ergodic equilibrium state according to Corollary 12.1.10. When $\varphi \equiv 0$, the equilibrium states are also called *measures of maximal entropy*, that is, measures for which $\mathrm{h}_\mu(T) = \mathrm{h}_{\mathrm{top}}(T)$. In particular, if $\mathrm{h}_{\mathrm{top}}(T) = 0$, then every invariant measure is a measure of maximal entropy for T. Recall that this is the case for homeomorphisms of the unit circle (see Exercise 7.6.10), among other examples.

A simple consequence of the variational principle is the following.

Theorem 12.2.2. *If $T : X \to X$ is a topological dynamical system and $\varphi : X \to \mathbb{R}$ is a Hölder continuous potential such that* $\mathrm{P}(T, \varphi) > \sup \varphi$, *then*

$$\mathrm{h}_\mu(T) > 0$$

for every equilibrium state μ of φ.

Proof. Since μ is an equilibrium state for φ, we have that

$$\mathrm{P}(T, \varphi) = \mathrm{h}_\mu(T) + \int_X \varphi \, d\mu \le \mathrm{h}_\mu(T) + \sup \varphi.$$

Rearranging the terms,

$$\mathrm{h}_\mu(T) \ge \mathrm{P}(T, \varphi) - \sup \varphi > 0.$$ □

It is natural to wonder whether equilibrium states exist for all topological dynamical systems. As the following example demonstrates, the answer is negative.

Example 12.2.3. We construct a system with positive, finite topological entropy but without any measure of maximal entropy. Let $(T_n : X_n \to X_n)_{n=1}^{\infty}$ be a sequence of topological dynamical systems with the property that

$$\mathrm{h_{top}}(T_n) < \mathrm{h_{top}}(T_{n+1}), \ \forall n \in \mathbb{N} \quad \text{and} \quad \sup_{n\in\mathbb{N}} \mathrm{h_{top}}(T_n) < \infty.$$

Let $\bigsqcup_{n=1}^{\infty} X_n$ denote the disjoint union of the spaces X_n, and let $X = \{\omega\} \cup \bigsqcup_{n=1}^{\infty} X_n$ be the one-point compactification of $\bigsqcup_{n=1}^{\infty} X_n$. Define the map $T : X \to X$ by

$$T(x) := \begin{cases} T_n(x) & \text{if } x \in X_n \\ \omega & \text{if } x = \omega. \end{cases}$$

Then T is continuous. Suppose that μ is an ergodic measure of maximal entropy for T. Then $\mu(\{\omega\}) \in \{0,1\}$ since $T^{-1}(\{\omega\}) = \{\omega\}$. But if $\mu(\{\omega\}) = 1$, then we would have $\mu(\bigsqcup_{n=1}^{\infty} X_n) = 0$. Hence, on one hand, we would have $\mathrm{h}_\mu(T) = 0$, while, on the other hand, $\mathrm{h}_\mu(T) = \mathrm{h_{top}}(T) \geq \sup_{n\in\mathbb{N}} \mathrm{h_{top}}(T_n) > 0$. This contradiction imposes that $\mu(\{\omega\}) = 0$. Similarly, $\mu(X_n) \in \{0,1\}$ for all $n \in \mathbb{N}$ since $T^{-1}(X_n) = X_n$. Therefore, there exists a unique $N \in \mathbb{N}$ such that $\mu(X_N) = 1$. It follows that

$$\mathrm{h_{top}}(T) = \mathrm{h}_\mu(T) = \mathrm{h}_\mu(T_N) \leq \mathrm{h_{top}}(T_N) < \sup_{n\in\mathbb{N}} \mathrm{h_{top}}(T_n) \leq \mathrm{h_{top}}(T).$$

This contradiction implies that there is no measure of maximal entropy for the system T.

Given that equilibrium states do not always exist, we would like to find conditions under which they do exist. But since the function $\mu \mapsto \int_X \varphi \, d\mu$ is continuous in the weak* topology on the compact space $M(T)$, the function $\mu \mapsto \mathrm{h}_\mu(T)$ cannot be continuous in general. Otherwise, the sum of these last two functions would be continuous and would hence attain a maximum on the compact space $M(T)$, that is, equilibrium states would always exist. Nevertheless, the function $\mu \mapsto \mathrm{h}_\mu(T)$ is sometimes upper semicontinuous and this is sufficient to ensure the existence of an equilibrium state. Let us first recall the notion of upper (and lower) semicontinuity.

Definition 12.2.4. Let X be a topological space. A function $f : X \to [-\infty, \infty]$ is *upper semicontinuous* if for all $x \in X$,

$$\limsup_{y\to x} f(y) \leq f(x).$$

Equivalently, f is upper semicontinuous if the set $\{x \in X : f(x) < r\}$ is open in X for all $r \in \mathbb{R}$. A function $f : X \to [-\infty, \infty]$ is *lower semicontinuous* if $-f$ is upper semicontinuous.

Evidently, a function $f : X \to [-\infty, \infty]$ is continuous if and only if it is both upper and lower semicontinuous. Like continuous functions, upper semicontinuous functions attain their upper bound (while lower semicontinuous functions reach their lower bound) on every compact set.

One class of dynamical systems for which the function $\mu \mapsto \mathrm{h}_\mu(T)$ is upper semicontinuous are the expansive maps T.

Theorem 12.2.5. *If $T : X \to X$ is expansive, then the function*

$$\begin{array}{rcccl} \mathrm{h}_\bullet(T) & : & M(T) & \longrightarrow & [0, \infty] \\ & & \mu & \longmapsto & \mathrm{h}_\mu(T) \end{array}$$

is upper semicontinuous. Hence, each potential $\varphi : X \to \mathbb{R}$ has an equilibrium state.

Proof. Fix $\delta > 0$ an expansive constant for T and let $\mu \in M(T)$. According to Lemma 9.6.1, there exists a finite Borel partition α of X such that $\operatorname{diam}(\alpha) < \delta$ and $\mu(\partial A) = 0$ for each $A \in \alpha$. Let $\varepsilon > 0$. As $\mathrm{h}_\mu(T) \geq \mathrm{h}_\mu(T, \alpha) = \inf_{n \in \mathbb{N}} \frac{1}{n} \mathrm{H}_\mu(\alpha^n)$ by Definitions 9.4.12 and 9.4.10, there exists $m \in \mathbb{N}$ such that

$$\frac{1}{m} \mathrm{H}_\mu(\alpha^m) \leq \mathrm{h}_\mu(T) + \frac{\varepsilon}{2}.$$

Let $(\mu_n)_{n=1}^\infty$ be a sequence of measures in $M(T)$ converging weakly* to μ. Since $\operatorname{diam}(\alpha) < \delta$, it follows from Theorem 9.4.20 that

$$\mathrm{h}_{\mu_n}(T) = \mathrm{h}_{\mu_n}(T, \alpha)$$

for all $n \in \mathbb{N}$. Moreover, by Lemma 12.1.6 (with α replaced by α^m), we have

$$\lim_{n \to \infty} \mathrm{H}_{\mu_n}(\alpha^m) = \mathrm{H}_\mu(\alpha^m).$$

Therefore, there exists $N \in \mathbb{N}$ such that

$$\frac{1}{m} \left| \mathrm{H}_{\mu_n}(\alpha^m) - \mathrm{H}_\mu(\alpha^m) \right| \leq \frac{\varepsilon}{2}$$

for all $n \geq N$. Hence, for all $n \geq N$, we deduce that

$$\mathrm{h}_{\mu_n}(T) = \mathrm{h}_{\mu_n}(T, \alpha) \leq \frac{1}{m} \mathrm{H}_{\mu_n}(\alpha^m) \leq \frac{1}{m} \mathrm{H}_\mu(\alpha^m) + \frac{\varepsilon}{2} \leq \mathrm{h}_\mu(T) + \varepsilon.$$

Consequently, $\limsup_{n\to\infty} \mathrm{h}_{\mu_n}(T) \leq \mathrm{h}_\mu(T)$ for any sequence $(\mu_n)_{n=1}^\infty$ in $M(T)$ converging weakly* to μ. Thus $\limsup_{\nu \to \mu} \mathrm{h}_\nu(T) \leq \mathrm{h}_\mu(T)$, or, in other words, $\mu \mapsto \mathrm{h}_\mu(T)$ is upper semicontinuous.

Since the function $\mu \mapsto \int_X \varphi \, d\mu$ is continuous in the weak* topology on the compact space $M(T)$, it follows that the function $\mu \mapsto \mathrm{h}_\mu(T) + \int_X \varphi \, d\mu$ is upper semicontinuous. Since upper semicontinuous functions attain their upper bound on any compact set, we conclude from the variational principle that each potential φ admits an equilibrium state. □

Recall the class of piecewise monotone continuous maps of the interval. These are not necessarily expansive maps (e. g., the tent map is not expansive). Nonetheless, the function $\mu \mapsto \mathrm{h}_\mu(T)$ is upper semicontinuous for any such map.

Theorem 12.2.6. *If $T : X \to X$ is a piecewise monotone continuous map of the interval, then the function $\mu \mapsto \mathrm{h}_\mu(T)$ is upper semicontinuous. Hence, each potential $\varphi : X \to \mathbb{R}$ has an equilibrium state.*

Proof. The avid reader is referred to [50]. □

12.3 Examples of equilibrium states

Example 12.3.1. By Corollary 12.1.10, any uniquely ergodic system has a unique equilibrium state for every continuous potential. This unique equilibrium state is obviously the unique ergodic invariant measure of the system.

For instance, recall that (cf. Proposition 8.2.43) a translation of the torus $L_\gamma : \mathbb{T}^n \to \mathbb{T}^n$, where $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n) \in \mathbb{T}^n$, is uniquely ergodic if and only if the numbers $1, \gamma_1, \gamma_2, \ldots, \gamma_n$ are linearly independent over $\mathbb{Q}$. Such a translation has a unique equilibrium state.

Let us now look at a symbolic example.

Example 12.3.2. We revisit Example 11.4.1, where E is a finite alphabet and $\sigma : E^\infty \to E^\infty$ is the one-sided full E-shift map. Recall that any function $\widetilde{\varphi} : E \to \mathbb{R}$ generates a continuous potential $\varphi : E^\infty \to \mathbb{R}$ defined by $\varphi(\omega) := \widetilde{\varphi}(\omega_1)$ on E^∞. This potential depends only on the first coordinate ω_1 of the word $\omega \in E^\infty$.

Let $\mathcal{F}$ be the σ-algebra $\mathcal{P}(E)$ of all subsets of E and let P be a probability measure/vector on E, that is, $\sum_{e\in E} P(\{e\}) = 1$. Recall from Examples 8.1.14 and 8.2.32 that the one-sided Bernoulli shift $(\sigma : E^\infty \to E^\infty, \mu_P)$ is an ergodic measure-preserving system.

Let $S = \sum_{e\in E} \exp(\widetilde{\varphi}(e))$. Note that $0 < S < \infty$. We will show that μ_P is an equilibrium state for $\sigma : E^\infty \to E^\infty$ when

$$P(\{e\}) = \frac{1}{S}\exp(\widetilde{\varphi}(e)), \quad \forall e \in E. \tag{12.16}$$

First, let us consider $\mathrm{h}_{\mu_P}(\sigma)$. Let $\alpha := \{[e]\}_{e\in E}$ be the partition of E^∞ into its initial 1-cylinders. It is easy to see that $\alpha^n = \{[\omega]\}_{\omega\in E^n}$, that is, α^n is the partition of E^∞ into its initial n-cylinders. Recall that α is a generator for σ (see Definition 9.4.19 and Example 9.4.23). By Theorem 9.4.20 and Definition 9.4.10, we know that

$$\mathrm{h}_{\mu_P}(\sigma) = \mathrm{h}_{\mu_P}(\sigma, \alpha) = \inf_{n\in\mathbb{N}} \frac{1}{n}\mathrm{H}_{\mu_P}(\alpha^n).$$

By induction on n, it is not difficult to establish that $\mathrm{H}_{\mu_P}(\alpha^n) = n\mathrm{H}_{\mu_P}(\alpha)$. Therefore,

$$\begin{aligned}
\mathrm{h}_{\mu_P}(\sigma) &= \mathrm{H}_{\mu_P}(\alpha) \\
&= -\sum_{e\in E} \mu_P([e]) \log \mu_P([e]) = -\sum_{e\in E} P(\{e\}) \log P(\{e\}) \\
&= -\sum_{e\in E} \frac{1}{S} \exp(\widetilde{\varphi}(e)) \log\left[\frac{1}{S}\exp(\widetilde{\varphi}(e))\right] \\
&= -\frac{1}{S}\sum_{e\in E} \exp(\widetilde{\varphi}(e))[\widetilde{\varphi}(e) - \log S] \\
&= -\frac{1}{S}\sum_{e\in E} \widetilde{\varphi}(e)\exp(\widetilde{\varphi}(e)) + \log S.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\int_{E^\infty} \varphi\, d\mu_P &= \sum_{e\in E}\int_{[e]} \varphi(\omega)\, d\mu_P(\omega) = \sum_{e\in E}\int_{[e]} \widetilde{\varphi}(\omega_1)\, d\mu_P(\omega) \\
&= \sum_{e\in E} \widetilde{\varphi}(e)\mu_P([e]) = \sum_{e\in E} \widetilde{\varphi}(e) P(\{e\}) \\
&= \frac{1}{S}\sum_{e\in E} \widetilde{\varphi}(e)\exp(\widetilde{\varphi}(e)).
\end{aligned}$$

It ensues that

$$\mathrm{h}_{\mu_P}(\sigma) + \int_{E^\infty} \varphi\, d\mu_P = \log S = \mathrm{P}(\sigma, \varphi)$$

where the last equality was derived in Example 11.4.1. Hence, μ_P is an equilibrium state for the potential φ when P satisfies (12.16).

Further examples will be given in Subsection 13.7.3 and in Chapters 17 and 27 onward.

12.4 Exercises

Exercise 12.4.1. Generalize Example 12.2.3 to show that if in addition $\varphi : X \to \mathbb{R}$ is a potential such that

$$\mathrm{P}(T_n, \varphi|_{X_n}) < \mathrm{P}(T_{n+1}, \varphi|_{X_{n+1}})$$

for all $n \in \mathbb{N}$, then φ has no equilibrium state.

Exercise 12.4.2. In the context of Exercise 11.5.8, give an explicit description of the equilibrium states of φ when $k = 1$ and $k = 2$.

Exercise 12.4.3. Let $T : X \to X$ be a topological dynamical system and $\varphi : X \to \mathbb{R}$ a potential. For any $n \in \mathbb{N}$, prove that if μ is an equilibrium state for the couple (T, φ), then μ is an equilibrium state for the couple $(T^n, S_n\varphi)$, too.

Exercise 12.4.4. Let $T : X \to X$ be a topological dynamical system and $\varphi : X \to \mathbb{R}$ a continuous potential. Show that the set of all equilibrium states for φ is a convex subset of $M(T)$. Deduce that if φ has a unique ergodic equilibrium state, then it has unique equilibrium state. Conclude also that if φ has two different equilibrium states, then it has uncountably many (in fact, a continuum of) equilibrium states.

Exercise 12.4.5. Let $T : X \to X$ be a topological dynamical system. Two continuous functions $\varphi, \psi : X \to \mathbb{R}$ are said to be cohomologous modulo a constant (or, equivalently, $\varphi - \psi$ is cohomologous to a constant) in the additive group $C(X)$ if there exist a continuous function $u : X \to \mathbb{R}$ and a constant $c \in \mathbb{R}$ such that

$$\varphi - \psi = u \circ T - u + c.$$

Show that such potentials φ and ψ share the same equilibrium states.

Exercise 12.4.6. Going beyond Example 12.2.3, give an example of a transitive topological dynamical system which does not have any measure of maximal entropy.

Exercise 12.4.7. Using Example 12.2.3, give an example of a topological dynamical system which admits infinitely many equilibrium states under a certain potential.

Exercise 12.4.8. Give an example of a transitive topological dynamical system which has infinitely many measures of maximal entropy.

Exercise 12.4.9. If $T : X \to X$ is a dynamical system such that $\mathrm{h}_{\mathrm{top}}(T) < \infty$, then deduce from the variational principle that the pressure function $\mathrm{P}(T, \bullet) : C(X) \to \mathbb{R}$ is convex.

Appendix A – A selection of classical results

This appendix lists classical definitions and results that will be used in this volume. Several of these results are stated without proofs. We sometimes complemented them with classical examples.

A.1 Measure theory

Let us begin by gathering together some of the standard results from measure theory that will be needed in this book. Measure theory is one of the main tools in ergodic theory, so it is important to be familiar with it. Proofs and further explanations of the results can be found in many books on measure theory, for instance, Billingsley [7, 8] and Rudin [58].

A.1.1 Collections of sets and measurable spaces

Given a set X, we shall denote the set of all subsets of X by $\mathcal{P}(X)$. Let us recall the definitions of some important collections of subsets of a set. The most basic collection is called a π-system.

Definition A.1.1. Let X be a set. A nonempty family $\mathcal{P} \subseteq \mathcal{P}(X)$ is a *π-system* on X if $P_1 \cap P_2 \in \mathcal{P}$ for all $P_1, P_2 \in \mathcal{P}$.

In other words, a π-system is a collection that is closed under finite intersections. For example, the family of open intervals $\{(a, \infty) : a \in \mathbb{R}\}$ constitutes a π-system on $\mathbb{R}$. So does the family of closed intervals $\{[a, \infty) : a \in \mathbb{R}\}$. Other examples are the families $\{(-\infty, b) : b \in \mathbb{R}\}$ and $\{(-\infty, b] : b \in \mathbb{R}\}$.

A “slightly” more complex collection is a semialgebra.

Definition A.1.2. Let X be a set. A family $\mathcal{S} \subseteq \mathcal{P}(X)$ is called a *semialgebra* on X if it satisfies the following three conditions:

(a) $\emptyset \in \mathcal{S}$.
(b) $\mathcal{S}$ is a π-system.
(c) If $S \in \mathcal{S}$, then $X \setminus S$ can be written as a finite union of mutually disjoint sets in $\mathcal{S}$. That is, $X \setminus S = \bigcup_{i=1}^{n} S_i$ for some $n \in \mathbb{N}$ and $S_1, S_2, \ldots, S_n \in \mathcal{S}$ with $S_i \cap S_j = \emptyset$ whenever $i \neq j$.

Every semialgebra is a π-system but the converse is not true in general. For instance, none of the π-systems described above is a semialgebra. However, the collection of all intervals forms a semialgebra on $\mathbb{R}$.

An even more intricate collection is an algebra.

https://doi.org/10.1515/9783110702682-013

Definition A.1.3. Let X be a set. A family $\mathcal{A} \subseteq \mathcal{P}(X)$ is said to be an *algebra* on X if it satisfies the following three conditions:
(a) $\emptyset \in \mathcal{A}$.
(b) $\mathcal{A}$ is a π-system.
(c) If $A \in \mathcal{A}$, then $X \setminus A \in \mathcal{A}$.

Every algebra is a semialgebra, although the converse is not true in general. For instance, the semialgebra outlined earlier is not an algebra. Nevertheless, as we will observe in the next lemma, the collection of all subsets of $\mathbb{R}$ that can be expressed as a finite union of intervals is an algebra on $\mathbb{R}$.

The fact that an algebra is stable under finite intersections and complementation implies that an algebra is stable under finitely many set operations (e. g., unions, intersections, differences, symmetric differences, complementation, and combinations thereof).

Note that $\{\emptyset, X\}$ and $\mathcal{P}(X)$ are trivial algebras on X. Since the intersection of any family of algebras is an algebra, the algebra $\mathcal{A}(\mathcal{C})$ generated by any collection $\mathcal{C}$ of subsets of X is well-defined as the smallest, in the sense of set inclusion, of all algebras on X that contain $\mathcal{C}$. If the collection $\mathcal{C}$ is a semialgebra, then it is easy to describe the algebra it generates.

Lemma A.1.4. *Let $\mathcal{S}$ be a semialgebra on X. The algebra $\mathcal{A}(\mathcal{S})$ generated by $\mathcal{S}$ consists of those subsets A of X, which can be written as a finite union of mutually disjoint sets in $\mathcal{S}$, that is, all sets $A \subseteq X$ such that $A = \bigcup_{i=1}^{n} S_i$ for some $S_1, S_2, \ldots, S_n \in \mathcal{S}$ with $S_i \cap S_j = \emptyset$ whenever $i \neq j$.*

Proof. Let

$$\mathcal{A} := \left\{ A \subseteq X \mid \exists S_1, \ldots, S_n \in \mathcal{S},\ S_i \cap S_j = \emptyset, \forall i \neq j \text{ such that } A = \bigcup_{i=1}^{n} S_i \right\}.$$

It is easy to see that $\mathcal{A}$ is an algebra containing $\mathcal{S}$. Therefore, $\mathcal{A} \supseteq \mathcal{A}(\mathcal{S})$. On the other hand, since any algebra is closed under finite unions, any algebra containing $\mathcal{S}$ must contain $\mathcal{A}$. Thus $\mathcal{A}(\mathcal{S}) \supseteq \mathcal{A}$. Hence, $\mathcal{A} = \mathcal{A}(\mathcal{S})$. □

In measure theory, the most important type of collection of subsets of a given set is a σ-algebra.

Definition A.1.5. Let X be a set. A family $\mathcal{B} \subseteq \mathcal{P}(X)$ is called a *σ-algebra* on X if it satisfies the following three conditions:
(a) $\emptyset \in \mathcal{B}$.
(b) $\bigcap_{n=1}^{\infty} B_n \in \mathcal{B}$ for every sequence $(B_n)_{n=1}^{\infty}$ of sets in $\mathcal{B}$.
(c) If $B \in \mathcal{B}$ then $X \setminus B \in \mathcal{B}$.

Note that condition (b) can be replaced by:

(b′) $\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}$ for every sequence $(B_n)_{n=1}^{\infty}$ of sets in $\mathcal{B}$.

A σ-algebra on X is thus a family of subsets of X which is closed under countably many set operations. Clearly, any σ-algebra is an algebra, though the converse is not true in general.

Note that $\{\emptyset, X\}$ and $\mathcal{P}(X)$ are trivial σ-algebras on X. They are respectively called the *indiscrete* and *discrete* σ-algebras. As the intersection of any family of σ-algebras is itself a σ-algebra, the σ-algebra $\sigma(\mathcal{C})$ generated by any collection $\mathcal{C}$ of subsets of X is well-defined as the smallest σ-algebra that contains $\mathcal{C}$. In particular, if the collection $\mathcal{C}$ is finite then the algebra $\mathcal{A}(\mathcal{C})$ it generates is also finite, and thus $\sigma(\mathcal{C}) = \mathcal{A}(\mathcal{C})$ (see Exercise 8.5.1).

A set X equipped with a σ-algebra $\mathcal{B}$ is called a *measurable space* and the elements of $\mathcal{B}$ are accordingly called *measurable sets*.

Example A.1.6. Let X be a topological space and let $\mathcal{T}$ be the topology of X, that is, the collection of all open subsets of X. Then $\sigma(\mathcal{T})$ is a σ-algebra on X called the *Borel σ-algebra* of X. Henceforth, we will denote this latter by $\mathcal{B}(X)$. In particular, $\mathcal{B}(X)$ contains all open sets and closed sets, as well as all countable unions of closed sets and all countable intersections of open sets, that is, all F_σ- and G_δ-sets, respectively. Note that $\mathcal{T}$ is a π-system but not a semialgebra in general.

In the Euclidean space $\mathbb{R}$, the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is generated by the even simpler π-system of open intervals $\{(a, \infty) : a \in \mathbb{R}\}$. Similarly, it is generated by the semialgebra comprising all intervals.

Sometimes the functions considered take values in the extended real numbers $\overline{\mathbb{R}} := [-\infty, \infty]$. A base for the order topology of $\overline{\mathbb{R}}$ is the π-system of all open intervals, namely $\{[-\infty, b) : b \in \overline{\mathbb{R}}\} \cup \{(a, b) : a, b \in \overline{\mathbb{R}}\} \cup \{(a, \infty] : a \in \overline{\mathbb{R}}\} \cup \{[-\infty, \infty]\}$. The Borel σ-algebra $\mathcal{B}(\overline{\mathbb{R}})$ is generated by the even simpler π-system of open intervals $\{(a, \infty] : a \in \mathbb{R}\}$. Note that

$$\mathcal{B}(\overline{\mathbb{R}}) = \big\{B, \{-\infty\} \cup B, B \cup \{\infty\}, \{-\infty\} \cup B \cup \{\infty\} \mid B \in \mathcal{B}(\mathbb{R})\big\}.$$

More examples of algebras and σ-algebras are presented in Exercises 8.5.2–8.5.4.

We now introduce λ-systems, also called Dynkin systems. These collections of sets are closed under complementation and countable disjoint unions.

Definition A.1.7. Let X be a set. A family $\mathcal{L} \subseteq \mathcal{P}(X)$ is called a *λ-system* on X if it satisfies the following three conditions:

(a) $X \in \mathcal{L}$.

(b) $\bigcup_{n=1}^{\infty} L_n \in \mathcal{L}$ for every sequence $(L_n)_{n=1}^{\infty}$ of sets in $\mathcal{L}$ such that $L_n \cap L_m = \emptyset$ for all $m \neq n$.

(c) If $L \in \mathcal{L}$, then $X \setminus L \in \mathcal{L}$.

Note that condition (c) can be replaced by:

(c′) If $K, L \in \mathcal{L}$ and $K \subseteq L$, then $L \setminus K \in \mathcal{L}$.

Every σ-algebra is a λ-system but the converse is not true in general. Nevertheless, it is not difficult to see how these two concepts are related.

Lemma A.1.8. *A collection of sets forms a σ-algebra if and only if it is both a λ-system and a π-system.*

Proof. We have already observed that every σ-algebra is a λ-system and a π-system. So suppose that $\mathcal{B}$ is both a λ-system and a π-system on a set X. Since $\mathcal{B}$ is a π-system that enjoys properties (a) and (c) of a λ-system, it is clear that $\mathcal{B}$ is an algebra. Therefore, it just remains to prove that $\mathcal{B}$ satisfies condition (b') of Definition A.1.5. Let $(B_n)_{n=1}^{\infty}$ be a sequence of sets in $\mathcal{B}$. For every $n \in \mathbb{N}$, let $B_n' = \bigcup_{k=1}^{n} B_k$. As $\mathcal{B}$ is an algebra, $B_n' \in \mathcal{B}$ for all $n \in \mathbb{N}$. The sequence $(B_n')_{n=1}^{\infty}$ is ascending and is such that $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} B_n'$. Thus it suffices to prove condition (b') for ascending sequences in $\mathcal{B}$. Moreover, observe that $\bigcup_{n=1}^{\infty} B_n' = B_1' \cup \bigcup_{n=1}^{\infty} (B_{n+1}' \setminus B_n')$. By condition (c') of a λ-system, we know that $B_{n+1}' \setminus B_n' \in \mathcal{B}$ for each $n \in \mathbb{N}$. Furthermore, the sets B_1' and $B_{n+1}' \setminus B_n'$, $n \in \mathbb{N}$, are mutually disjoint. By condition (b) of a λ-system, it follows that

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} B_n' = B_1' \cup \bigcup_{n=1}^{\infty} (B_{n+1}' \setminus B_n') \in \mathcal{B}.$$

Hence, $\mathcal{B}$ is an algebra satisfying condition (b') of Definition A.1.5. So $\mathcal{B}$ is a σ-algebra. □

The importance and usefulness of λ-systems mostly lie in the following theorem.

Theorem A.1.9 (Dynkin's π-λ theorem). *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system such that $\mathcal{P} \subseteq \mathcal{L}$, then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.*

Proof. See Theorem 3.2 in Billingsley [7]. □

Furthermore, as the intersection of any family of λ-systems is a λ-system, the λ-system $\mathcal{L}(\mathcal{C})$ generated by any collection $\mathcal{C} \subseteq \mathcal{P}(X)$ is well-defined as the intersection of all λ-systems that comprise $\mathcal{C}$. When $\mathcal{C}$ is a π-system, the λ-system and the σ-algebra that are generated by $\mathcal{C}$ are one and the same.

Corollary A.1.10. *If $\mathcal{P}$ is a π-system, then $\sigma(\mathcal{P}) = \mathcal{L}(\mathcal{P})$.*

Proof. This immediately follows from Lemma A.1.8 and Theorem A.1.9. □

Finally, let us recall yet another type of collection of sets named, for obvious reasons, a monotone class.

Definition A.1.11. Let X be a set. A family $\mathcal{M} \subseteq \mathcal{P}(X)$ is called a *monotone class* on X if it is stable under countable monotone unions and countable monotone intersections.

In other words,

$$\text{if } (M_n)_{n=1}^{\infty} \subseteq \mathcal{M} \text{ is such that } M_1 \subseteq M_2 \subseteq M_3 \subseteq \cdots, \quad \text{then } \bigcup_{n=1}^{\infty} M_n \in \mathcal{M}$$

and

$$\text{if } (M_n)_{n=1}^{\infty} \subseteq \mathcal{M} \text{ is such that } M_1 \supseteq M_2 \supseteq M_3 \supseteq \cdots, \quad \text{then } \bigcap_{n=1}^{\infty} M_n \in \mathcal{M}.$$

Every σ-algebra is a monotone class but the converse is not true in general. Nevertheless, a monotone class which is an algebra is a σ-algebra. There is an analogue of Dynkin's theorem for monotone classes.

Theorem A.1.12 (Halmos' monotone class theorem). *If $\mathcal{A}$ is an algebra and $\mathcal{M}$ is a monotone class such that $\mathcal{A} \subseteq \mathcal{M}$, then $\sigma(\mathcal{A}) \subseteq \mathcal{M}$.*

Proof. See Theorem 3.4 in Billingsley [7]. □

Because the intersection of any family of monotone classes is a monotone class, the monotone class $\mathcal{M}(\mathcal{C})$ generated by any collection $\mathcal{C} \subseteq \mathcal{P}(X)$ is well-defined as the intersection of all monotone classes that comprise $\mathcal{C}$. When $\mathcal{C}$ is a semialgebra, the σ-algebra and the monotone class generated by $\mathcal{C}$ coincide.

Theorem A.1.13. *If $\mathcal{S}$ is a semialgebra, then $\sigma(\mathcal{S}) = \sigma(\mathcal{A}(\mathcal{S})) = \mathcal{M}(\mathcal{A}(\mathcal{S})) = \mathcal{M}(\mathcal{S})$.*

Proof. The equalities $\sigma(\mathcal{S}) = \sigma(\mathcal{A}(\mathcal{S}))$ and $\mathcal{M}(\mathcal{A}(\mathcal{S})) = \mathcal{M}(\mathcal{S})$ are obvious. Since a σ-algebra is a monotone class and $\sigma(\mathcal{A}(\mathcal{S})) \supseteq \mathcal{A}(\mathcal{S})$, it is evident that $\sigma(\mathcal{A}(\mathcal{S})) \supseteq \mathcal{M}(\mathcal{A}(\mathcal{S}))$. The opposite inclusion is the object of Halmos' monotone class theorem. □

A.1.2 Measurable transformations

We now look at maps between measurable spaces. Recall that a measurable space is a set X equipped with a σ-algebra $\mathcal{A}$. The elements of $\mathcal{A}$ are the measurable sets in that space.

Definition A.1.14. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. A transformation $T : X \to Y$ is said to be *measurable* provided that $T^{-1}(B) \in \mathcal{A}$ for every $B \in \mathcal{B}$.

We have earlier mentioned that σ-algebras are the most important collections of sets in measure theory. However, it is generally impossible to describe, in a simple form, the sets in a σ-algebra. Luckily, σ-algebras that are generated by smaller and simpler structures like π-systems, semialgebras, or algebras, are much easier to cope with. In this situation, proving that some interesting property is satisfied for the sets in these smaller and simpler structures is often sufficient to guarantee that that property holds for all sets in the σ-algebra. This is the case for the measurability of transformations.

Theorem A.1.15. *Let $T : (X, \mathcal{A}) \to (Y, \mathcal{B})$ be a transformation. If $\mathcal{B} = \sigma(\mathcal{C})$ is a σ-algebra generated by a collection $\mathcal{C} \subseteq \mathcal{P}(Y)$, then T is measurable if and only if $T^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{C}$.*

Proof. It is clear that if T is measurable, then $T^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{C}$ since $\mathcal{C} \subseteq \sigma(\mathcal{C}) = \mathcal{B}$. Conversely, suppose that $T^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{C}$. Consider the collection of sets $\mathcal{B}' = \{B \subseteq Y : T^{-1}(B) \in \mathcal{A}\}$. By assumption, $\mathcal{B}' \supseteq \mathcal{C}$. It is also easy to see that $\mathcal{B}'$ is a σ-algebra. Thus $\mathcal{B}' \supseteq \sigma(\mathcal{C}) = \mathcal{B}$, and hence T is measurable. □

If the range Y of a transformation is a Borel subset of $\overline{\mathbb{R}}$, then unless otherwise stated Y will be assumed to be endowed with its Borel σ-algebra, which is just the projection of $\mathcal{B}(\overline{\mathbb{R}})$ onto Y (see Exercise 8.5.5). In this context, we will use the term function instead of transformation.

Example A.1.16. Let $(X, \mathcal{A})$ be a measurable space.
(a) Let $A \subseteq X$. The *indicator function* $\mathbb{1}_A : X \to \{0, 1\}$ (also called *characteristic function*) defined by

$$\mathbb{1}_A(x) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

is measurable if and only if A is measurable, that is, $A \in \mathcal{A}$.
(b) A function $s : X \to \mathbb{R}$ which takes only finitely many values is called a *simple function*. Such a function can be expressed in the form

$$s = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i},$$

where $A_i = \{x \in X : s(x) = \alpha_i\}$ and the α_i's are the values of the function s. Such a function is measurable if and only if each set A_i is measurable.

The following theorem shows the utility of simple functions. It states that any nonnegative measurable function is the pointwise limit of a nondecreasing sequence of nonnegative measurable simple functions.

Theorem A.1.17. *Let $(X, \mathcal{A})$ be a measurable space and $f : X \to [0, \infty]$ be a measurable function. Then there exists a sequence $(s_n)_{n=1}^{\infty}$ of measurable simple functions on X such that*
(a) $0 \le s_1 \le s_2 \le \cdots \le f$.
(b) $\lim_{n\to\infty} s_n(x) = f(x)$, $\forall x \in X$.

Proof. See Theorem 1.17 in Rudin [58]. □

A.1.3 Measure spaces

The concept of measure is obviously central to measure theory.

Definition A.1.18. Let $(X, \mathcal{A})$ be a measurable space. A set function $\mu : \mathcal{A} \to [0, \infty]$ is said to be a *measure* on X provided that
(a) $\mu(\emptyset) = 0$.
(b) μ is *countably additive*, that is, for each sequence $(A_n)_{n=1}^{\infty}$ of pairwise disjoint sets belonging to $\mathcal{A}$, the function μ is such that

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

The triple $(X, \mathcal{A}, \mu)$ is called a *measure space*.

If $\mu(X) < \infty$, then μ is said to be a *finite measure*. If $\mu(X) = 1$, then μ is a *probability measure*. Finally, μ is said to be *σ-finite* if there exists a sequence $(A_n)_{n=1}^{\infty}$ of sets in $\mathcal{A}$ such that $\mu(A_n) < \infty$ for all $n \in \mathbb{N}$ and $\bigcup_{n=1}^{\infty} A_n = X$.

Here are a few basic properties of measures.

Lemma A.1.19. *Let $(X, \mathcal{A}, \mu)$ be a measure space and $A, B \in \mathcal{A}$.*
(a) *If $A \subseteq B$, then $\mu(A) \le \mu(B)$.*
(b) *If $A \subseteq B$ and $\mu(B \setminus A) < \infty$, then $\mu(A) = \mu(B) - \mu(B \setminus A)$.*
(c) $\mu(A \cup B) \le \mu(A) + \mu(B)$.
(d) *If $(A_n)_{n=1}^{\infty}$ is an ascending sequence in $\mathcal{A}$ (i. e., $A_n \subseteq A_{n+1}, \forall n \in \mathbb{N}$), then*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n) = \sup_{n\in\mathbb{N}} \mu(A_n).$$

(e) *If $(B_n)_{n=1}^{\infty}$ is a descending sequence in $\mathcal{A}$ (i. e., $A_n \supseteq A_{n+1}, \forall n \in \mathbb{N}$) and if $\mu(B_1) < \infty$, then*

$$\mu\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mu(B_n) = \inf_{n\in\mathbb{N}} \mu(B_n).$$

(f) *If $(C_n)_{n=1}^{\infty}$ is any sequence in $\mathcal{A}$, then*

$$\mu\left(\bigcup_{n=1}^{\infty} C_n\right) \le \sum_{n=1}^{\infty} \mu(C_n).$$

(g) *If $(D_n)_{n=1}^{\infty}$ is a sequence in $\mathcal{A}$ such that $\mu(D_m \cap D_n) = 0$ for all $m \ne n$, then*

$$\mu\left(\bigcup_{n=1}^{\infty} D_n\right) = \sum_{n=1}^{\infty} \mu(D_n).$$

Proof. The proof is left to the reader as an exercise. □

A more intricate property of measures deserves a special name.

Lemma A.1.20 (Borel–Cantelli lemma). *Let $(X, \mathcal{A}, \mu)$ be a measure space and $(A_n)_{n=1}^{\infty}$ a sequence in $\mathcal{A}$. If $\sum_{n=1}^{\infty} \mu(A_n) < \infty$, then $\mu(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n) = 0$.*

Proof. The proof is left to the reader as an exercise. □

We now provide two simple examples of measures. The first of these may seem insignificant at first glance but turns out to be very useful in practice.

Example A.1.21. Let X be a nonempty set and $\mathcal{P}(X)$ be the discrete σ-algebra on X.

(a) Choose a point $x \in X$. Define the set function $\delta_x : \mathcal{P}(X) \to \{0, 1\}$ by setting

$$\delta_x(A) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

One readily verifies that δ_x is a probability measure. It is referred to as the *Dirac point mass* or *Dirac measure* concentrated at the point x.

(b) For any $A \subseteq X$ define $m(A)$ to be the number of elements in the set A if A is finite and set $m(A) = \infty$ if the set A is infinite. Then $m : \mathcal{P}(X) \to [0, \infty]$ is called the *counting measure* on X.

There is a notion of completeness for measure spaces.

Definition A.1.22. A measure space $(X, \mathcal{A}, \mu)$ is said to be *complete* if every subset of a set of measure zero is measurable. That is, if $A \subseteq X$ and there is $B \in \mathcal{A}$ such that $A \subseteq B$ and $\mu(B) = 0$, then $A \in \mathcal{A}$.

Note that any measure space can be extended to a complete one (see Exercises 8.5.7–8.5.8).

In Example A.1.6, we introduced the concept of Borel σ-algebra. We now introduce Borel measures and describe different forms of regularity for these measures.

Definition A.1.23. Let X be a topological space and $\mathcal{B}(X)$ be the Borel σ-algebra on X.

(a) A *Borel measure* μ on X is a measure defined on the Borel σ-algebra $\mathcal{B}(X)$ of X. The resulting measure space $(X, \mathcal{B}(X), \mu)$ is called a Borel measure space. In particular, if μ is a probability measure then $(X, \mathcal{B}(X), \mu)$ is called a Borel probability space.

(b) A Borel measure μ is said to be *inner regular* if

$$\mu(B) = \sup\{\mu(K) : K \subseteq B,\ K \text{ compact}\}, \quad \forall B \in \mathcal{B}(X).$$

(c) A Borel measure μ is said to be *outer regular* if

$$\mu(B) = \inf\{\mu(G) : B \subseteq G,\ G \text{ open}\}, \quad \forall B \in \mathcal{B}(X).$$

(d) A Borel measure is called *regular* if it is both inner and outer regular.

Theorem A.1.24. *Every Borel probability measure on a separable, completely metrizable space is regular.*

Proof. See Theorem 17.11 in Kechris [35]. □

Example A.1.25. There exists a complete, regular measure λ defined on a σ-algebra $\mathcal{L}$ on $\mathbb{R}^k$ with the following properties:

(a) $\lambda(R) = \mathrm{Vol}(R)$ for every k-rectangle $R \subseteq \mathbb{R}^k$, where Vol denotes the usual k-dimensional volume in $\mathbb{R}^k$.

(b) $\mathcal{L}$ is the completion of the Borel σ-algebra $\mathcal{B}(\mathbb{R}^k)$; more precisely, $E \in \mathcal{L}$ if and only if there exist an F_σ-set F and a G_δ-set G such that $F \subseteq E \subseteq G$ and $\lambda(G \setminus F) = 0$.

(c) λ is *translation invariant*, that is, $\lambda(E + x) = \lambda(E)$ for all $E \in \mathcal{L}$ and all $x \in \mathbb{R}^k$.

Moreover, up to a multiplicative constant, λ is the only translation-invariant Borel measure on $\mathbb{R}^k$ that gives finite measure to all compact sets.

The measure λ is called the *Lebesgue measure on* $\mathbb{R}^k$ and, accordingly, the sets in $\mathcal{L}$ are said to be *Lebesgue measurable*. Unlike in Example A.1.21, not all subsets of $\mathbb{R}^k$ are Lebesgue measurable. Indeed, Vitali (see 5.7, The Vitali Monsters on p. 120 of [29]) showed that it is impossible to construct a measure having properties (a)–(c) on the set of all subsets of $\mathbb{R}^k$.

The following lemma is one more eloquent manifestation of the relevance of structures simpler than σ-algebras.

Lemma A.1.26. *Let X be a set and $\mathcal{P}$ be a π-system on X. Let μ and ν be probability measures on $(X, \sigma(\mathcal{P}))$. Then*

$$\mu = \nu \iff \mu(P) = \nu(P), \quad \forall P \in \mathcal{P}.$$

Proof. The implication $\Rightarrow$ is trivial. For the opposite one $\Leftarrow$, suppose that $\mu(P) = \nu(P)$ for all $P \in \mathcal{P}$. Consider the collection of sets $\mathcal{A} := \{A \subseteq X : \mu(A) = \nu(A)\}$. By assumption, $\mathcal{P} \subseteq \mathcal{A}$. It is also easy to see that $\mathcal{A}$ is a λ-system. Per Corollary A.1.10, it follows that $\sigma(\mathcal{P}) = \mathcal{L}(\mathcal{P}) \subseteq \mathcal{A}$. □

In summary, two probability measures that agree on a π-system are equal on the σ-algebra generated by that π-system. However, this result does not generally hold for infinite measures (see Exercise 8.5.9).

A.1.4 Extension of set functions to measures

The main shortcoming of the preceding lemma lies in the assumption that the set functions μ and ν are measures defined on a σ-algebra. In particular, this means that they are countably additive on that entire σ-algebra. However, we frequently define a set function $\mu : \mathcal{C} \to [0, \infty]$ on a collection $\mathcal{C}$ of subsets of a set X on which the values of μ are naturally determined but it may be unclear whether μ may be extended to a measure on $\sigma(\mathcal{C})$. The forthcoming results are extremely useful in that regard. The first one concerns the extension of a finitely/countably additive set function from a semialgebra to an algebra.

Theorem A.1.27. *Let $\mathcal{S}$ be a semialgebra on a set X and let $\mu : \mathcal{S} \to [0,\infty]$ be a finitely additive set function, that is, a function such that*

$$\mu\left(\bigcup_{i=1}^{n} S_i\right) = \sum_{i=1}^{n} \mu(S_i)$$

for every finite family $(S_i)_{i=1}^n$ of mutually disjoint sets in $\mathcal{S}$ such that $\bigcup_{i=1}^n S_i \in \mathcal{S}$. Then there exists a unique finitely additive set function $\overline{\mu} : \mathcal{A}(\mathcal{S}) \to [0,\infty]$ which is an extension of μ to $\mathcal{A}(\mathcal{S})$, the algebra generated by $\mathcal{S}$. Moreover, the extension $\overline{\mu}$ is countably additive whenever the original set function μ is.

Proof. This directly follows from Lemma A.1.4. For more detail, see Theorems 3.4 and 3.5 in Kingman and Taylor [38]. □

The second result concerns the extension of a countably additive set function from an algebra to a σ-algebra.

Theorem A.1.28 (Carathéodory's extension theorem). *Let $\mathcal{A}$ be an algebra on a set X and let $\mu : \mathcal{A} \to [0,1]$ be a countably additive set function, that is, a function such that*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

for every sequence $(A_i)_{i=1}^\infty$ of mutually disjoint sets in $\mathcal{A}$ such that $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$. Suppose also that $\mu(X) = 1$. Then there exists a unique probability measure $\overline{\mu} : \sigma(\mathcal{A}) \to [0,1]$, which is an extension of $\mu : \mathcal{A} \to [0,1]$.

Proof. See Theorem 3.1 in Billingsley [7] or Theorem 4.2 in Kingman and Taylor [38]. □

Carathéodory's extension theorem thus reduces the problem to demonstrating that a set function is countably additive on an algebra. This can be hard to prove, so we sometimes rely upon the following result.

Lemma A.1.29. *Let $\mathcal{A}$ be an algebra on a set X and $\mu : \mathcal{A} \to [0,\infty)$ be a finitely additive function. Then μ is countably additive if and only if*

$$\lim_{i\to\infty} \mu(A_i) = 0 \tag{A.1}$$

for every descending sequence $(A_i)_{i=1}^\infty$ of sets in $\mathcal{A}$ such that $\bigcap_{i=1}^\infty A_i = \emptyset$.

Furthermore, if (A.1) *holds, then μ has a unique extension to a σ-additive function (a measure) from* $\sigma(\mathcal{A})$*, the σ-algebra generated by* $\mathcal{A}$*, to* $[0, \infty)$.

Proof. Recall that measures enjoy property (d) of Lemma A.1.19. However, the set function μ considered here is not a measure since it is defined on an algebra rather than on a σ-algebra. Nevertheless, we will show that μ satisfies property (d) on the algebra $\mathcal{A}$ because of hypothesis (A.1). Let $(A_i)_{i=1}^{\infty}$ be an ascending sequence of sets in $\mathcal{A}$ such that $A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. Since $\mathcal{A}$ is an algebra, we know that $A \setminus A_i \in \mathcal{A}$ for all $i \in \mathbb{N}$. Therefore, the sequence $(A \setminus A_i)_{i=1}^{\infty}$ is a descending sequence of sets in $\mathcal{A}$ such that

$$\bigcap_{i=1}^{\infty}(A \setminus A_i) = A \setminus \left(\bigcup_{i=1}^{\infty} A_i\right) = \emptyset.$$

By hypothesis (A.1), we infer that

$$\lim_{i\to\infty} \mu(A \setminus A_i) = 0.$$

Moreover, since $A = (A \setminus A_i) \cup A_i$ and μ is finitely additive on $\mathcal{A}$ and finite, we deduce that

$$\mu(A \setminus A_i) = \mu(A) - \mu(A_i), \quad \forall i \in \mathbb{N}.$$

Hence,

$$\lim_{i\to\infty} \big(\mu(A) - \mu(A_i)\big) = 0.$$

Thus

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mu(A_i). \tag{A.2}$$

Now, let $(B_i)_{i=1}^{\infty}$ be any sequence of mutually disjoint sets in $\mathcal{A}$ such that $\bigcup_{i=1}^{\infty} B_i \in \mathcal{A}$. For every $i \in \mathbb{N}$, let $B_i' = \bigcup_{j=1}^{i} B_j$. As $\mathcal{A}$ is an algebra, $B_i' \in \mathcal{A}$ for all $i \in \mathbb{N}$. The sequence $(B_i')_{i=1}^{\infty}$ is ascending and is such that $\bigcup_{i=1}^{\infty} B_i' = \bigcup_{i=1}^{\infty} B_i \in \mathcal{A}$. By (A.2), we know that

$$\mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \mu\left(\bigcup_{i=1}^{\infty} B_i'\right) = \lim_{i\to\infty} \mu(B_i').$$

Using this, the fact that the B_i's are mutually disjoint and that μ is finitely additive, we conclude that

$$\mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \lim_{i\to\infty} \mu(B_i') = \lim_{i\to\infty} \mu\left(\bigcup_{j=1}^{i} B_j\right) = \lim_{i\to\infty} \sum_{j=1}^{i} \mu(B_j) = \sum_{i=1}^{\infty} \mu(B_i).$$

That is, μ is countably additive on $\mathcal{A}$.

The second part of the statement directly follows from the first one and Theorem A.1.28. □

As an immediate consequence of Lemma A.1.29 we get the following.

Lemma A.1.30. *Let $\mathcal{A}$ be an algebra on a set X, let $\nu : \sigma(\mathcal{A}) \to [0, \infty)$ be a finite measure, and let $\mu : \mathcal{A} \to [0, \infty)$ be a finitely additive function which is absolutely continuous with respect to ν, meaning that $\mu(A) = 0$ whenever $A \in \mathcal{A}$ and $\nu(A) = 0$. Then μ is countably additive and has a unique extension to a σ-additive function (a measure) from $\sigma(\mathcal{A})$ to $[0, \infty)$.*

Finally, as a straightforward consequence of Lemma A.1.29 we have the following.

Lemma A.1.31. *Let $\mathcal{A}$ be a σ-algebra on a set X, let $\nu : \mathcal{A} \to [0, \infty)$ be a finite measure and let $\mu : \mathcal{A} \to [0, \infty)$ be a finitely additive function which is absolutely continuous with respect to ν on some algebra generating $\mathcal{A}$, meaning that $\mu(A) = 0$ whenever A belongs to this algebra and $\nu(A) = 0$. Then μ is a measure.*

Another feature of an algebra is that any element of a σ-algebra generated by an algebra can be approximated as closely as desired by an element of the algebra. Before stating the precise result, recall that the *symmetric difference* of two sets A and B is denoted by $A \triangle B$ and is the set of all points that belong to exactly one of those two sets. That is,

$$A \triangle B := (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B).$$

Properties of the symmetric difference are examined in Exercises 8.5.10–8.5.11.

Lemma A.1.32. *Let $\mathcal{A}$ be an algebra on a set X and μ be a probability measure on $(X, \sigma(\mathcal{A}))$. Then for every $\varepsilon > 0$ and $B \in \sigma(\mathcal{A})$ there exists some $A \in \mathcal{A}$ such that $\mu(A \triangle B) < \varepsilon$.*

Proof. See Theorem 4.4 in Kingman and Taylor [38]. □

A.1.5 Integration

Let us now briefly recollect some facts about integration. First, the definition of the integral of a measurable function with respect to a measure.

Definition A.1.33. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $A \in \mathcal{A}$.

(a) If $s : X \to [0, \infty)$ is a measurable simple function of the form,

$$s = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i},$$

then the integral of the function s over the set A with respect to the measure μ is defined as

$$\int_A s \, d\mu := \sum_{i=1}^{n} \alpha_i \mu(A_i \cap A).$$

We use the convention that $0 \cdot \infty = 0$ in case it happens that $\alpha_i = 0$ and $\mu(A_i \cap A) = \infty$ for some $1 \leq i \leq n$.

(b) If $f : X \to [0, \infty]$ is a measurable function, then the integral of the function f over the set A with respect to the measure μ is defined as

$$\int_A f \, d\mu := \sup \int_A s \, d\mu,$$

where the supremum is taken over all measurable simple functions $0 \leq s \leq f$. Note that if f is simple, then definitions (a) and (b) coincide.

(c) If $f : X \to \overline{\mathbb{R}}$ is a measurable function, then the integral of the function f over the set A with respect to the measure μ is defined as

$$\int_A f \, d\mu := \int_A f_+ \, d\mu - \int_A f_- \, d\mu,$$

as long as $\min\{\int_A f_+ \, d\mu, \int_A f_- \, d\mu\} < \infty$, where f_+ and f_- respectively denote the positive and negative parts of f. That is, $f_+(x) := \max\{f(x), 0\}$ whereas $f_-(x) := \max\{-f(x), 0\}$.

(d) A measurable function $f : X \to \overline{\mathbb{R}}$ is said to be *integrable* if $\int_X |f| \, d\mu < \infty$. We denote this by $f \in L^1(X, \mathcal{A}, \mu)$. If there can be no confusion, we simply write $f \in L^1(\mu)$.

(e) A property is said to hold *μ-almost everywhere* (sometimes abbreviated μ-a. e.) if the property holds on the entire space except possibly on a set of μ-measure zero.

The following properties follow from this definition.

Lemma A.1.34. *Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f, g \in L^1(X, \mathcal{A}, \mu)$, $A, B \in \mathcal{A}$, and $a, b \in \mathbb{R}$.*

(a) *If $f \leq g$ μ-a. e., then $\int_A f \, d\mu \leq \int_A g \, d\mu$. Also, if $f < g$ μ-a. e., then $\int_A f \, d\mu < \int_A g \, d\mu$.*

(b) *If $A \subseteq B$ and $0 \leq f$ μ-a. e., then $0 \leq \int_A f \, d\mu \leq \int_B f \, d\mu$.*

(c)

$$\left| \int_A f \, d\mu \right| \leq \int_A |f| \, d\mu.$$

(d) *Linearity:*

$$\int_A (af + bg) \, d\mu = a \int_A f \, d\mu + b \int_A g \, d\mu.$$

(e) *If $(A_n)_{n=1}^{\infty}$ is a sequence of mutually disjoint measurable sets, then*

$$\int_{\bigcup_{n=1}^{\infty} A_n} f \, d\mu = \sum_{n=1}^{\infty} \int_{A_n} f \, d\mu.$$

(f) *$f = g$ μ-a. e. $\Longleftrightarrow \int_A f\, d\mu = \int_A g\, d\mu$, $\forall A \in \mathcal{A}$.*
(g) *The relation $f = g$ μ-a. e. is an equivalence relation on the set $L^1(X, \mathcal{A}, \mu)$. The equivalence classes generated by this relation form a Banach space also denoted by $L^1(X, \mathcal{A}, \mu)$ (or $L^1(\mu)$, for short) with norm*

$$\|f\|_1 := \int_X |f|\, d\mu < \infty.$$

A sequence $(f_n)_{n=1}^{\infty}$ in $L^1(\mu)$ is said to converge to f in $L^1(\mu)$ if $\lim_{n\to\infty} \|f_n - f\|_1 = 0$.

A.1.6 Convergence theorems

In measure theory, there are fundamental theorems that are especially helpful for finding the integral of functions that are the pointwise limits of sequences of functions.

The first of these results applies to monotone sequences of functions. A sequence of functions $(f_n)_{n=1}^{\infty}$ is monotone if it is increasing pointwise ($f_{n+1}(x) \geq f_n(x)$ for all $x \in X$ and all $n \in \mathbb{N}$) or decreasing pointwise ($f_{n+1}(x) \leq f_n(x)$ for all $x \in X$ and all $n \in \mathbb{N}$). We state the theorem for increasing sequences, but its counterpart for decreasing sequences can be easily deduced from it.

Theorem A.1.35 (Monotone convergence theorem). *Let $(X, \mathcal{A}, \mu)$ be a measure space. If $(f_n)_{n=1}^{\infty}$ is an increasing sequence of nonnegative measurable functions, then the integral of their pointwise limit is equal to the limit of their integrals, that is,*

$$\int_X \lim_{n\to\infty} f_n\, d\mu = \lim_{n\to\infty} \int_X f_n\, d\mu.$$

Proof. See Theorem 1.26 in Rudin [58]. □

Note that this theorem holds for almost everywhere increasing sequences of almost everywhere nonnegative measurable functions with an almost everywhere pointwise limit.

For general sequences of nonnegative functions, we have the following immediate consequence.

Lemma A.1.36 (Fatou's lemma). *Let $(X, \mathcal{A}, \mu)$ be a measure space. For any sequence $(f_n)_{n=1}^{\infty}$ of nonnegative measurable functions,*

$$\int_X \liminf_{n\to\infty} f_n\, d\mu \leq \liminf_{n\to\infty} \int_X f_n\, d\mu.$$

Proof. For every $x \in X$ and $n \in \mathbb{N}$, define $g_n(x) = \inf\{f_i(x) : 1 \leq i \leq n\}$ and apply the monotone convergence theorem to the sequence $(g_n)_{n=1}^{\infty}$. For more detail, see Theorem 1.28 in Rudin [58]. □

The following lemma is another application of the monotone convergence theorem. It offers another way of integrating a nonnegative function.

Lemma A.1.37. *Let $(X, \mathcal{A}, \mu)$ be a measure space. Let f be a nonnegative measurable function and $A \in \mathcal{A}$. Then*

$$\int_A f\, d\mu = \int_0^\infty \mu(\{x \in A : f(x) > r\})\, dr.$$

Proof. Suppose that $f = \mathbb{1}_B$ for some $B \in \mathcal{A}$. Then

$$\begin{aligned}
\int_0^\infty \mu(\{x \in A : f(x) > r\})\, dr &= \int_0^1 \mu(\{x \in A : \mathbb{1}_B(x) > r\})\, dr \\
&= \int_0^1 \mu(A \cap B)\, dr \\
&= \mu(A \cap B) \\
&= \int_X \mathbb{1}_{A\cap B}\, d\mu = \int_X \mathbb{1}_A \cdot \mathbb{1}_B\, d\mu = \int_A f\, d\mu.
\end{aligned}$$

So the equality holds for characteristic functions. We leave it to the reader to show that the equality prevails for all nonnegative measurable simple functions. If f is a general nonnegative measurable function, then by Theorem A.1.17 there exists an increasing sequence $(s_n)_{n=1}^\infty$ of nonnegative measurable simple functions such that $\lim_{n\to\infty} s_n(x) = f(x)$ for every $x \in X$. For every $r \geq 0$, let

$$\tilde{f}(r) = \mu(\{x \in A : f(x) > r\}).$$

This function is obviously nonnegative and decreasing. Hence, by Theorem A.1.15 it is Borel measurable since $\tilde{f}^{-1}((t, \infty])$ is an interval for all $t \in [0, \infty]$ and the sets $\{(t, \infty]\}_{t\in[0,\infty]}$ generate the Borel σ-algebra of $[0, \infty]$. Fix momentarily $r \geq 0$. Since $s_n \nearrow f$, the sets $(\{x \in A : s_n(x) > r\})_{n=1}^\infty$ form an ascending sequence such that

$$\bigcup_{n=1}^\infty \{x \in A : s_n(x) > r\} = \{x \in A : f(x) > r\}.$$

Then by Lemma A.1.19(d),

$$\tilde{f}(r) = \mu\left(\bigcup_{n=1}^\infty \{x \in A : s_n(x) > r\}\right) = \lim_{n\to\infty} \mu(\{x \in A : s_n(x) > r\}) = \lim_{n\to\infty} \tilde{s}_n(r).$$

Observe that $(\tilde{s}_n)_{n=1}^\infty$ is an increasing sequence of nonnegative decreasing functions. So $(\tilde{s}_n)_{n=1}^\infty$ is a sequence of nonnegative Borel measurable functions, which increases pointwise to $\tilde{f}$. It follows from the monotone convergence theorem (Theorem A.1.35) that

$$
\begin{aligned}
\int_0^\infty \mu(\{x \in A : f(x) > r\})\, dr &= \int_0^\infty \widetilde{f}(r)\, dr = \int_0^\infty \lim_{n\to\infty} \widetilde{s}_n(r)\, dr \\
&= \lim_{n\to\infty} \int_0^\infty \widetilde{s}_n(r)\, dr \\
&= \lim_{n\to\infty} \int_0^\infty \mu(\{x \in A : s_n(x) > r\})\, dr \\
&= \lim_{n\to\infty} \int_A s_n\, d\mu = \int_A \lim_{n\to\infty} s_n\, d\mu \\
&= \int_A f\, d\mu. \qquad \square
\end{aligned}
$$

Pointwise convergence of a sequence of integrable functions does not guarantee convergence in L^1 (see Exercise 8.5.14). However, under one relatively weak additional assumption, this becomes true. The second fundamental theorem of convergence applies to sequences of functions which have an almost everywhere pointwise limit and are dominated (i. e., uniformly bounded) almost everywhere by an integrable function.

Theorem A.1.38 (Lebesgue's dominated convergence theorem). *If a sequence of measurable functions $(f_n)_{n=1}^\infty$ on a measure space $(X, \mathcal{A}, \mu)$ converges pointwise μ-a. e. to a function f and if there exists $g \in L^1(\mu)$ such that $|f_n(x)| \le g(x)$ for all $n \in \mathbb{N}$ and μ-a. e. $x \in X$, then $f \in L^1(\mu)$ and*

$$\lim_{n\to\infty} \|f_n - f\|_1 = 0 \quad \textit{and} \quad \lim_{n\to\infty} \int_X f_n\, d\mu = \int_X f\, d\mu.$$

Proof. Apply Fatou's lemma to $2g - |f_n - f| \ge 0$. See Theorem 1.34 in Rudin [58] for more detail. $\square$

Note that

$$\lim_{n\to\infty} \|f_n - f\|_1 = 0 \implies \lim_{n\to\infty} \|f_n\|_1 = \|f\|_1$$

since $\big|\|f_n\|_1 - \|f\|_1\big| \le \|f_n - f\|_1$ and

$$\lim_{n\to\infty} \|f_n - f\|_1 = 0 \implies \lim_{n\to\infty} \int_X f_n\, d\mu = \int_X f\, d\mu$$

by applying Lemma A.1.34(c) to $f_n - f$. The opposite implications do not hold in general. Nevertheless, the following lemma states that any sequence of integrable functions

$(f_n)_{n=1}^\infty$ that converges pointwise almost everywhere to an integrable function f will also converge to that function in L^1 if and only if their L^1 norms converge to the L^1 norm of f.

Lemma A.1.39 (Scheffé's lemma). *Let $(X, \mathcal{A}, \mu)$ be a measure space. If a sequence $(f_n)_{n=1}^\infty$ of functions in $L^1(\mu)$ converges pointwise μ-a. e. to a function $f \in L^1(\mu)$, then*

$$\lim_{n\to\infty} \|f_n - f\|_1 = 0 \iff \lim_{n\to\infty} \|f_n\|_1 = \|f\|_1.$$

In particular, if $f_n \geq 0$ μ-a. e. for all $n \in \mathbb{N}$, then $f \geq 0$ μ-a. e. and

$$\lim_{n\to\infty} \|f_n - f\|_1 = 0 \iff \lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu.$$

Proof. The direct implication $\Rightarrow$ is trivial. For the converse implication, assume that $\lim_{n\to\infty} \|f_n\|_1 = \|f\|_1$. Suppose first that $f_n \geq 0$ for all $n \in \mathbb{N}$. Then $f \geq 0$ and hence our assumption reduces to $\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu$. Let $\ell_n = \min\{f, f_n\}$ and $u_n = \max\{f, f_n\}$. Then both $(\ell_n)_{n=1}^\infty$ and $(u_n)_{n=1}^\infty$ converge pointwise μ-a. e. to f. Also, $|\ell_n| = \ell_n \leq f$ for all n, so Lebesgue's dominated convergence theorem asserts that

$$\lim_{n\to\infty} \int_X \ell_n \, d\mu = \int_X f \, d\mu.$$

Observing that $u_n = f + f_n - \ell_n$, we also get that

$$\lim_{n\to\infty} \int_X u_n \, d\mu = \int_X f \, d\mu + \lim_{n\to\infty} \int_X f_n \, d\mu - \lim_{n\to\infty} \int_X \ell_n \, d\mu = \int_X f \, d\mu.$$

Thus

$$\lim_{n\to\infty} \|f_n - f\|_1 = \lim_{n\to\infty} \int_X |f_n - f| \, d\mu = \lim_{n\to\infty} \left(\int_X u_n \, d\mu - \int_X \ell_n \, d\mu \right) = 0.$$

So the implication $\Leftarrow$ holds for nonnegative functions. As $g = g_+ - g_-$ and $((f_n)_+)_{n=1}^\infty$ and $((f_n)_-)_{n=1}^\infty$ are sequences of functions in $L^1(\mu)$ converging pointwise μ-a. e. to $f_+ \in L^1(\mu)$ and $f_- \in L^1(\mu)$, respectively, it is easy to see that the general case follows from the case for nonnegative functions. $\square$

If g is a L^1 function on a general measure space $(X, \mathcal{A}, \mu)$, then the sequence of nonnegative measurable functions $(g_M)_{M=1}^\infty$, where $g_M = |g| \cdot \mathbb{1}_{\{|g|\geq M\}}$, decreases to 0 pointwise and is dominated by $|g|$. Therefore, the monotone convergence theorem (or, alternatively, Lebesgue's dominated convergence theorem) affirms that

$$\lim_{M\to\infty} \int_{\{|g|\geq M\}} |g| \, d\mu = 0.$$

This suggests introducing the following concept.

Definition A.1.40. Let $(X, \mathcal{A}, \mu)$ be a measure space. A sequence of measurable functions $(f_n)_{n=1}^{\infty}$ is *uniformly integrable* if

$$\lim_{M\to\infty} \sup_{n\in\mathbb{N}} \int_{\{|f_n|\geq M\}} |f_n|\, d\mu = 0.$$

On finite measure spaces, there exists a generalization of Lebesgue's dominated convergence theorem (Theorem A.1.38).

Theorem A.1.41. *Let $(X, \mathcal{A}, \mu)$ be a finite measure space and $(f_n)_{n=1}^{\infty}$ a sequence of measurable functions that converges pointwise μ-a. e. to a function f.*

(a) *If $(f_n)_{n=1}^{\infty}$ is uniformly integrable, then $f_n \in L^1(\mu)$ for all $n \in \mathbb{N}$ and $f \in L^1(\mu)$. Moreover,*

$$\lim_{n\to\infty} \|f_n - f\|_1 = 0 \quad \textit{and} \quad \lim_{n\to\infty} \int_X f_n\, d\mu = \int_X f\, d\mu.$$

(b) *If $f, f_n \in L^1(\mu)$ and $f_n \geq 0$ μ-a. e. for all $n \in \mathbb{N}$, then $\lim_{n\to\infty} \int_X f_n\, d\mu = \int_X f\, d\mu$ implies that $(f_n)_{n=1}^{\infty}$ is uniformly integrable.*

Proof. See Theorem 16.14 in Billingsley [7]. □

Corollary A.1.42. *Let $(X, \mathcal{A}, \mu)$ be a finite measure space and $(f_n)_{n=1}^{\infty}$ a sequence of integrable functions that converges pointwise μ-a. e. to an integrable function f. Then the following conditions are equivalent:*

(a) *The sequence $(f_n)_{n=1}^{\infty}$ is uniformly integrable.*
(b) $\lim_{n\to\infty} \|f_n - f\|_1 = 0$.
(c) $\lim_{n\to\infty} \|f_n\|_1 = \|f\|_1$.

Proof. Part (a) of Theorem A.1.41 yields (a)⇒(b). That (b)⇒(c) follows from $\big|\|f_n\|_1 - \|f\|_1\big| \leq \|f_n - f\|_1$. Finally, replacing f_n by $|f_n - f|$ and f by 0 in part (b) of Theorem A.1.41 gives (c)⇒(a). □

Obviously, any sequence of measurable functions that converges uniformly on an entire space does converge pointwise. It is well known that the converse is not true in general, and it is thus natural to ask whether, in some way, a pointwise convergent sequence converges "almost" uniformly.

Definition A.1.43. Let $(X, \mathcal{A}, \mu)$ be a measure space. A sequence $(f_n)_{n=1}^{\infty}$ of measurable functions on X is said to *converge μ-almost uniformly* to a function f if for every $\varepsilon > 0$ there exists $Y \in \mathcal{A}$ such that $\mu(Y) < \varepsilon$ and $(f_n)_{n=1}^{\infty}$ converges uniformly to f on $X \setminus Y$.

It is clear that almost uniform convergence implies almost everywhere pointwise convergence. The converse is not true in general but these two types of convergence are one and the same on any finite measure space.

Theorem A.1.44 (Egorov's theorem). *Let $(X, \mathcal{A}, \mu)$ be a finite measure space. A sequence $(f_n)_{n=1}^{\infty}$ of measurable functions on X converges pointwise μ-almost everywhere to a limit function f if and only if that sequence converges μ-almost uniformly to f.*

Proof. See Chapter 3, Exercise 16 in Rudin [58]. □

The reader ought to convince themself that this result does not generally hold on infinite spaces.

Convergence in measure is another interesting type of convergence.

Definition A.1.45. Let $(X, \mathcal{A}, \mu)$ be a measure space. A sequence $(f_n)_{n=1}^{\infty}$ of measurable functions *converges in measure* to a measurable function f provided that for each $\varepsilon > 0$,

$$\lim_{n\to\infty} \mu(\{x \in X : |f_n(x) - f(x)| > \varepsilon\}) = 0.$$

Lemma A.1.46. *Let $(X, \mathcal{A}, \mu)$ be a measure space. If a sequence $(f_n)_{n=1}^{\infty}$ of measurable functions converges in $L^1(\mu)$ to a measurable function f, then $(f_n)_{n=1}^{\infty}$ converges in measure to f.*

Proof. Let $\varepsilon > 0$. Then

$$\mu(\{x \in X : |f_n(x) - f(x)| > \varepsilon\}) \leq \int_{\{x\in X : |f_n(x)-f(x)|>\varepsilon\}} \frac{|f_n - f|}{\varepsilon}\, d\mu \leq \frac{1}{\varepsilon}\|f_n - f\|_1.$$

Taking the limit of both sides as $n \to \infty$ completes the proof. □

When the measure is finite, there is a close relationship between pointwise convergence and convergence in measure.

Theorem A.1.47. *Let $(X, \mathcal{A}, \mu)$ be a finite measure space and $(f_n)_{n=1}^{\infty}$ a sequence of measurable functions.*

(a) *If $(f_n)_{n=1}^{\infty}$ converges pointwise μ-a. e. to a function f, then $(f_n)_{n=1}^{\infty}$ converges in measure to f.*
(b) *If $(f_n)_{n=1}^{\infty}$ converges in measure to a function f, then there exists a subsequence $(f_{n_k})_{k=1}^{\infty}$ which converges pointwise μ-a. e. to f.*
(c) *$(f_n)_{n=1}^{\infty}$ converges in measure to a function f if and only if each subsequence $(f_{n_k})_{k=1}^{\infty}$ admits a further subsequence $(f_{n_{k_l}})_{l=1}^{\infty}$ that converges pointwise μ-a. e. to f.*

Proof. See Theorem 20.5 in Billingsley [7]. □

The previous two results reveal that, on a finite measure space, a sequence of integrable functions that converges in L^1 to an integrable function admits a subsequence which converges pointwise almost everywhere to that function. In general, the sequence itself might not converge pointwise almost everywhere (see Exercise 8.5.17).

In some sense, the following result is a form of convergence theorem. It asserts that Borel measurable functions can be approximated by continuous functions on "arbitrarily large" portions of their domain.

Theorem A.1.48. *Let $(X, \mathcal{B}(X), \mu)$ be a finite Borel measure space and let $f : X \to \overline{\mathbb{R}}$ be a Borel measurable function. Given any $\varepsilon > 0$, for every $B \in \mathcal{B}(X)$ there is a closed set E with $\mu(B \setminus E) < \varepsilon$ such that $f|_E$ is continuous. If B is locally compact, then the set E can be chosen to be compact and then there is a continuous function $f_\varepsilon : X \to \overline{\mathbb{R}}$ with compact support that coincides with f on E and such that $\sup_{x \in X} |f_\varepsilon(x)| \le \sup_{x \in X} |f(x)|$.*

A.1.7 Mutual singularity, absolute continuity and equivalence of measures

We now leave aside convergence of sequences of functions and recall the definitions of mutually singular, absolutely continuous, and equivalent measures.

Definition A.1.49. Let $(X, \mathcal{A})$ be a measurable space, and μ and ν be two measures on $(X, \mathcal{A})$.

(a) The measures μ and ν are said to be *mutually singular*, denoted by $\mu \perp \nu$, if there exist disjoint sets $X_\mu, X_\nu \in \mathcal{A}$ such that $\mu(X \setminus X_\mu) = 0 = \nu(X \setminus X_\nu)$.

(b) The measure μ is said to be *absolutely continuous with respect to* ν, denoted $\mu \ll \nu$, if $\nu(A) = 0 \Longrightarrow \mu(A) = 0$.

(c) The measures μ and ν are said to be *equivalent* if $\mu \ll \nu$ and $\nu \ll \mu$.

The Radon–Nikodym theorem provides a characterization of absolute continuity. Though it is valid for σ-finite measures, the following version for finite measures is sufficient for our purposes.

Theorem A.1.50 (Radon–Nikodym theorem). *Let $(X, \mathcal{A})$ be a measurable space and let μ and ν be two finite measures on $(X, \mathcal{A})$. Then the following statements are equivalent:*

(a) $\mu \ll \nu$.

(b) *For every $\varepsilon > 0$ there exists $\delta > 0$ such that $\nu(A) < \delta \Longrightarrow \mu(A) < \varepsilon$.*

(c) *There exists a ν-a. e. unique function $f \in L^1(\nu)$ such that $f \ge 0$ and*

$$\mu(A) = \int_A f \, d\nu, \quad \forall A \in \mathcal{A}.$$

Proof. See relation (32.4) and Theorem 32.2 in Billingsley [7]. □

Remark A.1.51. The function f is often denoted by $\frac{d\mu}{d\nu}$ and called the *Radon–Nikodym derivative* of μ with respect to ν.

Per Lemma A.1.34(f), two integrable functions are equal almost everywhere if and only if their integrals are equal over every measurable set. When the measure is finite, we can restrict our attention to any generating π-system.

Corollary A.1.52. *Let $(X, \mathcal{A}, \nu)$ be a finite measure space and suppose that $\mathcal{A} = \sigma(\mathcal{P})$ for some π-system $\mathcal{P}$. Let $f, g \in L^1(\nu)$. Then*

$$f = g \ \ \nu\text{-a.e.} \quad \Longleftrightarrow \quad \int_P f\, d\nu = \int_P g\, d\nu, \quad \forall P \in \mathcal{P}.$$

Proof. The direct implication $\Rightarrow$ is obvious. So let us assume that $\int_P f\, d\nu = \int_P g\, d\nu$ for all $P \in \mathcal{P}$. The measures $\mu_f(A) := \int_A f\, d\nu$ and $\mu_g(A) := \int_A g\, d\nu$ are equal on the π-system $\mathcal{P}$. According to Lemma A.1.26, this implies that $\mu_f = \mu_g$. It follows from the uniqueness part of the Radon–Nikodym theorem that $f = g$ ν-almost everywhere. □

A.1.8 The space $C(X)$, its dual $C(X)^*$ and the subspace $M(X)$

Another important result is Riesz representation theorem. Before stating it, we first establish some notation. Let X be a compact metrizable space. Let $C(X)$ be the set of all continuous real-valued functions on X. This set becomes a normed vector space when endowed with the supremum norm

$$\|f\|_\infty := \sup\{|f(x)| : x \in X\}. \tag{A.3}$$

This norm defines a metric on $C(X)$ in the usual way:

$$d_\infty(f, g) := \|f - g\|_\infty = \sup\{|f(x) - g(x)| : x \in X\}.$$

The topology induced by the metric d_∞ on $C(X)$ is called the *topology of uniform convergence* on X. Indeed, $\lim_{n\to\infty} d_\infty(f_n, f) = 0$ if and only if the sequence $(f_n)_{n=1}^\infty$ converges to f uniformly on X. It is not hard to see that $C(X)$ is a separable Banach space (i. e., a separable and complete normed vector space).

Let $C(X)^*$ denote the *dual space* of $C(X)$, that is,

$$C(X)^* := \{F : C(X) \to \mathbb{R} \mid F \text{ is continuous and linear}\}.$$

Recall that a real-valued function F defined on $C(X)$ is called a *functional* on $C(X)$. It is well known that a linear functional F is continuous if and only if it is *bounded*, that is, if and only if its operator norm $\|F\|$ is finite, where

$$\|F\| := \sup\{|F(f)| : f \in C(X) \text{ and } \|f\|_\infty \le 1\}. \tag{A.4}$$

So $C(X)^*$ can also be described as the normed vector space of all bounded linear functionals on $C(X)$. The operator norm defines a metric on $C(X)^*$ in the usual manner:

$$d(F, G) := \|F - G\|.$$

The topology induced by the metric d on $C(X)^*$ is called the *operator norm topology*, or strong topology, on $C(X)^*$.

It is not difficult to see that $C(X)^*$ is a separable Banach space. Furthermore, a linear functional F is said to be *normalized* if $F(1) = 1$ and is called *positive* if $F(f) \geq 0$ whenever $f \geq 0$.

Finally, we denote the set of all Borel probability measures on X by $M(X)$. This set is clearly convex and can be characterized as follows.

Theorem A.1.53 (Riesz representation theorem). *Let X be a compact metrizable space, and let F be a normalized and positive linear functional on $C(X)$. Then there exists a unique $\mu \in M(X)$ such that*

$$F(f) = \int_X f\, d\mu, \quad \forall f \in C(X). \tag{A.5}$$

Conversely, any $\mu \in M(X)$ defines a normalized positive linear functional on $C(X)$ via formula (A.5). This linear functional is bounded.

Proof. The converse statement is straightforward to check. For the other direction, see Theorem 2.14 in Rudin [58]. □

It immediately follows from Riesz representation theorem that every Borel probability measure on a compact metrizable space is uniquely determined by the way it integrates continuous functions on that space.

Corollary A.1.54. *If μ and ν are two Borel probability measures on a compact metrizable space X, then*

$$\mu = \nu \quad \Longleftrightarrow \quad \int_X f\, d\mu = \int_X f\, d\nu, \quad \forall f \in C(X).$$

Let us now discuss the weak* *topology* on the set $M(X)$. Recall that if Z is a set and $(Z_\alpha)_{\alpha \in A}$ is a family of topological spaces, then the weak topology induced on Z by a collection of maps $\{\psi_\alpha : Z \to Z_\alpha \mid \alpha \in A\}$ is the smallest topology on Z that makes each ψ_α continuous. Evidently, the sets $\psi_\alpha^{-1}(U_\alpha)$, for U_α open in Z_α, constitute a subbase for the weak topology. The weak* topology on $M(X)$ is the weak topology induced by $C(X)$ on its dual space $C(X)^*$, where measures in $M(X)$ and normalized positive linear functionals in $C(X)^*$ are identified via the Riesz representation theorem. Note that $M(X)$ is metrizable, although $C(X)^*$ with the weak* topology usually is not. Indeed, both $C(X)$ and its subspace $C(X, [0, 1])$ of continuous functions on X taking values in $[0, 1]$, are separable since X is a compact metrizable space. Then for any dense subset $\{f_n\}_{n=1}^{\infty}$ of $C(X, [0, 1])$, a metric on $M(X)$ is

$$d(\mu, \nu) = \sum_{n=1}^{\infty} \frac{1}{2^n} \left| \int_X f_n\, d\mu - \int_X f_n\, d\nu \right|.$$

In this book, we will denote the convergence of a sequence of measures $(\mu_n)_{n=1}^\infty$ to a measure μ in the weak* topology of $M(X)$ by $\mu_n \xrightarrow{*} \mu$.

Remark A.1.55. Note that this notion is often presented as "weak convergence" of measures. This can be slightly confusing at first sight, but it helps to bear in mind that, as we have seen above, the weak* topology is just one instance of a weak topology.

The following theorem gives several equivalent characterizations of weak* convergence of Borel probability measures.

Theorem A.1.56 (Portmanteau theorem). *Let $(\mu_n)_{n=1}^\infty$ and μ be Borel probability measures on a compact metrizable space X. The following statements are equivalent:*
(a) $\mu_n \xrightarrow{*} \mu$.
(b) *For all continuous functions $f : X \to \mathbb{R}$,*

$$\lim_{n\to\infty} \int_X f\, d\mu_n = \int_X f\, d\mu.$$

(c) *For all closed sets $F \subseteq X$,*

$$\limsup_{n\to\infty} \mu_n(F) \le \mu(F).$$

(d) *For all open sets $G \subseteq X$,*

$$\liminf_{n\to\infty} \mu_n(G) \le \mu(G).$$

(e) *For all sets $A \in \mathcal{B}(X)$ such that $\mu(\partial A) = 0$,*

$$\lim_{n\to\infty} \mu_n(A) = \mu(A).$$

Proof. See Theorem 2.1 in Billingsley [8]. □

For us, the most important result concerning weak* convergence of measures is that the set $M(X)$ of all Borel probability measures on a compact metrizable space X is a compact and convex set in the weak* topology.

In order to establish this, we need to remember Banach–Alaoglu's theorem. In this theorem, note that the boundedness and closedness are with respect to the operator norm on the dual space while the compactness is with respect to the weak* topology on the dual space.

Theorem A.1.57 (Banach–Alaoglu's theorem). *The closed unit ball in the dual space B^* of a Banach space B is compact in the* weak* *topology on B^*. Furthermore, every closed, bounded subset of B^* is compact in the* weak* *topology on B^*.*

Proof. See Theorem V.4.2 and Corollary V.4.3 in Dunford and Schwartz [20]. □

Theorem A.1.58. *Let X be a compact metrizable space. The set $M(X)$ is compact and convex in the* weak* *topology of $C(X)^*$.*

Proof. The set $M(X)$ is closed with respect to the operator norm topology on $C(X)^*$. Indeed, suppose that $(\mu_n)_{n=1}^{\infty}$ is a sequence in $M(X)$ which converges to a $F \in C(X)^*$ in the operator norm topology of $C(X)^*$. In other words, suppose that $\lim_{n\to\infty} \|\mu_n - F\| = 0$. By definition of the operator norm (see (A.4)) and thanks to the linearity of F, this implies that

$$F(f) = \lim_{n\to\infty} \int_X f \, d\mu_n, \quad \forall f \in C(X).$$

In particular, F is normalized (since $F(1) = 1$) and positive (as $F(f) \geq 0$ for all $f \geq 0$). By Riesz representation theorem (Theorem A.1.53), there is $\mu \in M(X)$ that represents F. So $F \in M(X)$, and thus $M(X)$ is closed in the operator norm topology on $C(X)^*$.

The set $M(X)$ is also bounded in that topology. Indeed, if $\mu \in M(X)$ then

$$\|\mu\| \leq \sup\left\{ \int_X |f| \, d\mu : f \in C(X), \|f\|_\infty \leq 1 \right\} = 1.$$

Since X is a compact metrizable space, the space $C(X)$ is a Banach space, as earlier mentioned. We can then infer from Banach–Alaoglu's theorem that the set $M(X)$ is compact in the weak* topology.

The convexity of $M(X)$ is obvious. Indeed, if $\mu, \nu \in M(X)$ so is any convex combination $m = \alpha\mu + (1-\alpha)\nu$, where $\alpha \in [0,1]$. □

A.1.9 Expected values and conditional expectation functions

The mean or expected value of a function over a set is a straightforward generalisation of the mean value of a real-valued function defined on an interval of the real line.

Definition A.1.59. Let $(X, \mathcal{A}, \mu)$ be a probability space and let $\varphi \in L^1(\mu)$. The *mean* or *expected value* $E(\varphi|A)$ of the function φ over the set $A \in \mathcal{A}$ is defined to be

$$E(\varphi|A) := \begin{cases} \frac{1}{\mu(A)} \int_A \varphi \, d\mu & \text{if } \mu(A) > 0 \\ 0 & \text{if } \mu(A) = 0. \end{cases}$$

Given that μ is a probability measure, the expected value of φ over the entire space is simply given by

$$E(\varphi) := E(\varphi|X) = \int_X \varphi \, d\mu =: \mu(\varphi).$$

Our next goal is to give the definition of the conditional expectation of a function with respect to a σ-algebra. Let $(X, \mathcal{A}, \mu)$ be a probability space and $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Let also $\varphi \in L^1(X, \mathcal{A}, \mu)$. Notice that $\varphi : X \to \mathbb{R}$ is not necessarily measurable if X is endowed with the sub-σ-algebra $\mathcal{B}$ instead of the σ-algebra $\mathcal{A}$. In short, we say that φ is $\mathcal{A}$-measurable but not necessarily $\mathcal{B}$-measurable. We aim to find a function $E(\varphi|\mathcal{B}) \in L^1(X, \mathcal{B}, \mu)$ such that

$$\int_B E(\varphi|\mathcal{B})\, d\mu = \int_B \varphi\, d\mu, \quad \forall B \in \mathcal{B}. \tag{A.6}$$

This condition means that the function $E(\varphi|\mathcal{B})$ has the same expected value as φ on every measurable set belonging to the sub-σ-algebra $\mathcal{B}$. Accordingly, $E(\varphi|\mathcal{B})$ is called the *conditional expectation of φ with respect to $\mathcal{B}$*.

We now demonstrate the existence and μ-a. e. uniqueness of the conditional expectation. Let us begin with the existence of that function. Suppose first that the function φ is nonnegative. If $\varphi = 0$ μ-a. e., then simply set $E(\varphi|\mathcal{B}) = 0$. If $\varphi \neq 0$ μ-a. e. then the set function $\nu(A) := \int_A \varphi\, d\mu$ defines a finite measure on $(X, \mathcal{A})$ which is absolutely continuous with respect to μ. The restriction of ν to $\mathcal{B}$ also determines a finite measure on $(X, \mathcal{B})$ which is absolutely continuous with respect to the restriction of μ to $\mathcal{B}$. So by the Radon–Nikodym theorem (Theorem A.1.50), there exists a μ-a. e. unique nonnegative function $\widehat{\varphi} \in L^1(X, \mathcal{B}, \mu)$ such that $\nu(B) = \int_B \widehat{\varphi}\, d\mu$ for every $B \in \mathcal{B}$. Then

$$\int_B \widehat{\varphi}\, d\mu = \nu(B) = \int_B \varphi\, d\mu, \quad \forall B \in \mathcal{B}.$$

The point here is that although it may look as if we have not really achieved anything, we have actually gained that $\widehat{\varphi}$ is $\mathcal{B}$-measurable, whereas φ may not be. Therefore, $\widehat{\varphi}$ is the sought-after conditional expectation $E(\varphi|\mathcal{B})$ of φ with respect to $\mathcal{B}$.

If φ takes both negative and positive values, write $\varphi = \varphi_+ - \varphi_-$, where $\varphi_+(x) := \max\{\varphi(x), 0\}$ is the positive part of φ and $\varphi_-(x) := \max\{-\varphi(x), 0\}$ is the negative part of φ. Then define the conditional expectation linearly, that is, set

$$E(\varphi|\mathcal{B}) := E(\varphi_+|\mathcal{B}) - E(\varphi_-|\mathcal{B}).$$

This proves the existence of the conditional expectation function. Its μ-a. e. uniqueness follows from its defining property (A.6) and Lemma A.1.34(f).

The conditional expectation exhibits several natural properties. We mention a few of them in the next proposition.

Proposition A.1.60. *Let $(X, \mathcal{A}, \mu)$ be a probability space, let $\mathcal{B}$ and $\mathcal{C}$ denote sub-σ-algebras of $\mathcal{A}$ and let $\varphi \in L^1(X, \mathcal{A}, \mu)$.*
(a) *If $\varphi \geq 0$ μ-a. e., then $E(\varphi|\mathcal{B}) \geq 0$ μ-a. e.*
(b) *If $\varphi_1 \geq \varphi_2$ μ-a. e., then $E(\varphi_1|\mathcal{B}) \geq E(\varphi_2|\mathcal{B})$ μ-a. e.*
(c) $|E(\varphi|\mathcal{B})| \leq E(|\varphi|\,|\,\mathcal{B})$.

(d) *The functional* $E(\cdot|\mathcal{B})$ *is linear, i. e. for any* $c_1, c_2 \in \mathbb{R}$ *and* $\varphi_1, \varphi_2 \in L^1(X, \mathcal{A}, \mu)$,

$$E(c_1\varphi_1 + c_2\varphi_2\big|\mathcal{B}) = c_1E(\varphi_1|\mathcal{B}) + c_2E(\varphi_2|\mathcal{B}).$$

(e) *If* φ *is already* $\mathcal{B}$*-measurable, then* $E(\varphi|\mathcal{B}) = \varphi$. *In particular, we have that* $E\big(E(\varphi|\mathcal{B})\big|\mathcal{B}\big) = E(\varphi|\mathcal{B})$. *Also, if* $\varphi = c \in \mathbb{R}$ *is a constant function, then* $E(\varphi|\mathcal{B}) = \varphi = c$.

(f) *If* $\mathcal{C} \subseteq \mathcal{B}$, *then* $E(\varphi|\mathcal{C}) = E\big(\,E(\varphi|\mathcal{B})\,\big|\,\mathcal{C}\big)$.

Proof. This is left as an exercise to the reader. □

We will now determine the conditional expectation of an arbitrary integrable function φ with respect to various sub-σ-algebras of particular interest.

Example A.1.61. Let $(X, \mathcal{A}, \mu)$ be a probability space. The family $\mathcal{N}$ of all measurable sets that are either of null or of full measure constitutes a sub-σ-algebra of $\mathcal{A}$. Let $\varphi \in L^1(X, \mathcal{A}, \mu)$. Then the function $E(\varphi|\mathcal{N})$ has to belong to $L^1(X, \mathcal{N}, \mu)$ and must satisfy condition (A.6). In particular, $E(\varphi|\mathcal{N})$ must be $\mathcal{N}$-measurable. This means that for each Borel subset R of $\mathbb{R}$, the function $E(\varphi|\mathcal{N})$ must be such that $E(\varphi|\mathcal{N})^{-1}(R) \in \mathcal{N}$. Among others, for every $t \in \mathbb{R}$ we must have $E(\varphi|\mathcal{N})^{-1}(\{t\}) \in \mathcal{N}$; in other words, for each $t \in \mathbb{R}$ the set $E(\varphi|\mathcal{N})^{-1}(\{t\})$ must be of measure zero or of measure one. Also bear in mind that

$$X = E(\varphi|\mathcal{N})^{-1}(\mathbb{R}) = \bigcup_{t\in\mathbb{R}} E(\varphi|\mathcal{N})^{-1}\big(\{t\}\big).$$

Since the above union consists of mutually disjoint sets of measure zero and one, it follows that only one of these sets can be of measure one. In other words, there exists a unique $t \in \mathbb{R}$ such that $E(\varphi|\mathcal{N})^{-1}(\{t\}) = A$ for some $A \in \mathcal{A}$ with $\mu(A) = 1$. Because the function $E(\varphi|\mathcal{N})$ is unique up to a set of measure zero, we may assume without loss of generality that $A = X$. Hence, $E(\varphi|\mathcal{N})$ is a constant function. More specifically, its value is

$$E(\varphi|\mathcal{N}) = \int_X E(\varphi|\mathcal{N})\,d\mu = \int_X \varphi\,d\mu.$$

Example A.1.62. Let $(X, \mathcal{A})$ be a measurable space and let $\alpha = \{A_n\}_{n=1}^{\infty}$ be a countable measurable partition of X. That is, each $A_n \in \mathcal{A}$, $A_i \cap A_j = \emptyset$ for all $i \neq j$ and $X = \bigcup_{n=1}^{\infty} A_n$. The sub-$\sigma$-algebra of $\mathcal{A}$ generated by α is the family of all sets which can be written as a union of elements of α, that is,

$$\sigma(\alpha) = \Big\{A \subseteq X : A = \bigcup_{j\in J} A_j \text{ for some } J \subseteq \mathbb{N}\Big\}.$$

When α is finite, so is $\sigma(\alpha)$. When α is countably infinite, $\sigma(\alpha)$ is uncountable. Let μ be a probability measure on $(X, \mathcal{A})$. Let $\varphi \in L^1(X, \mathcal{A}, \mu)$ and set $\mathcal{B} = \sigma(\alpha)$. Then the conditional expectation $E(\varphi|\mathcal{B}) : X \to \mathbb{R}$ has to be a $L^1(X, \mathcal{B}, \mu)$ function that satisfies condition (A.6). In particular, $E(\varphi|\mathcal{B})$ must be $\mathcal{B}$-measurable. Thus, for any $t \in \mathbb{R}$ we must have $E(\varphi|\mathcal{B})^{-1}(\{t\}) \in \mathcal{B}$, that is, the set $E(\varphi|\mathcal{B})^{-1}(\{t\})$ must be a union of elements of α. This means that the conditional expectation function $E(\varphi|\mathcal{B})$ is constant on each element of α. Let $A_n \in \alpha$. If $\mu(A_n) = 0$ then $E(\varphi|\mathcal{B})|_{A_n} = 0$. Otherwise,

$$E(\varphi|\mathcal{B})|_{A_n} = \frac{1}{\mu(A_n)} \int_{A_n} E(\varphi|\mathcal{B})\, d\mu = \frac{1}{\mu(A_n)} \int_{A_n} \varphi\, d\mu = E(\varphi|A_n).$$

In summary, the conditional expectation $E(\varphi|\mathcal{B})$ of a function φ with respect to a sub-σ-algebra generated by a countable measurable partition is constant on each element of that partition. More precisely, on any given element of the partition, $E(\varphi|\mathcal{B})$ is equal to the mean value of φ on that element.

The next result is a special case of a theorem originally due to Doob and called the martingale convergence theorem. But, first, let us define the martingale itself.

Definition A.1.63. Let $(X, \mathcal{A}, \mu)$ be a probability space. Let $(\mathcal{A}_n)_{n=1}^{\infty}$ be a sequence of sub-σ-algebras of $\mathcal{A}$. Let also $(\varphi_n : X \to \mathbb{R})_{n=1}^{\infty}$ be a sequence of $\mathcal{A}$-measurable functions. The sequence $((\varphi_n, \mathcal{A}_n))_{n=1}^{\infty}$ is called a *martingale* if the following conditions are satisfied:
(a) $(\mathcal{A}_n)_{n=1}^{\infty}$ is an ascending sequence, that is, $\mathcal{A}_n \subseteq \mathcal{A}_{n+1}$ for all $n \in \mathbb{N}$.
(b) φ_n is $\mathcal{A}_n$-measurable for all $n \in \mathbb{N}$.
(c) $\varphi_n \in L^1(\mu)$ for all $n \in \mathbb{N}$.
(d) $E(\varphi_{n+1}|\mathcal{A}_n) = \varphi_n$ μ-a. e. for all $n \in \mathbb{N}$.

Theorem A.1.64 (Martingale convergence theorem). *Let $(X, \mathcal{A}, \mu)$ be a probability space. If $((\varphi_n, \mathcal{A}_n))_{n=1}^{\infty}$ is a martingale such that*

$$\sup_{n\in\mathbb{N}} \|\varphi_n\|_1 < \infty,$$

then there exists $\widehat{\varphi} \in L^1(X, \mathcal{A}, \mu)$ such that

$$\lim_{n\to\infty} \varphi_n(x) = \widehat{\varphi}(x) \quad \text{for } \mu\text{-a. e. } x \in X \quad \text{and} \quad \|\widehat{\varphi}\|_1 \le \sup_{n\in\mathbb{N}} \|\varphi_n\|_1.$$

Proof. See Theorem 35.5 in Billingsley [7]. □

One natural martingale is formed by the conditional expectations of a function with respect to an ascending sequence of sub-σ-algebras.

Example A.1.65. Let $(X, \mathcal{A}, \mu)$ be a probability space and let $(\mathcal{A}_n)_{n=1}^{\infty}$ be an ascending sequence of sub-σ-algebras of $\mathcal{A}$. For any $\varphi \in L^1(X, \mathcal{A}, \mu)$, the sequence $\{(E(\varphi|\mathcal{A}_n), \mathcal{A}_n)\}_{n=1}^{\infty}$

is a martingale. Indeed, set $\varphi_n = E(\varphi|\mathcal{A}_n)$ for all $n \in \mathbb{N}$. Condition (a) in Definition A.1.63 is automatically fulfilled. Conditions (b) and (c) follow from the very definition of the conditional expectation function. Regarding condition (d), a straightforward application of Proposition A.1.60(f) gives

$$E(\varphi_{n+1}|\mathcal{A}_n) = E\big(E(\varphi|\mathcal{A}_{n+1}) \mid \mathcal{A}_n\big) = E(\varphi|\mathcal{A}_n) = \varphi_n \ \ \mu\text{-a. e.}, \quad \forall n \in \mathbb{N}.$$

So $\{(E(\varphi|\mathcal{A}_n), \mathcal{A}_n)\}_{n=1}^{\infty}$ is a martingale. Using Proposition A.1.60(c), note that

$$\sup_{n\in\mathbb{N}} \|\varphi_n\|_1 = \sup_{n\in\mathbb{N}} \int_X \big|E(\varphi|\mathcal{A}_n)\big| \, d\mu \le \sup_{n\in\mathbb{N}} \int_X E\big(|\varphi| \mid \mathcal{A}_n\big) \, d\mu = \int_X |\varphi| \, d\mu = \|\varphi\|_1 < \infty.$$

According to Theorem A.1.64, there thus exists $\widehat{\varphi} \in L^1(X, \mathcal{A}, \mu)$ such that

$$\lim_{n\to\infty} E(\varphi|\mathcal{A}_n)(x) = \widehat{\varphi}(x) \quad \text{for } \mu\text{-a. e. } x \in X \quad \text{and} \quad \|\widehat{\varphi}\|_1 \le \|\varphi\|_1.$$

What is $\widehat{\varphi}$? This is the question we will answer in Theorem A.1.67.

Beforehand, we establish the uniform integrability of this martingale (see Definition A.1.40).

Lemma A.1.66. *Let $(X, \mathcal{A}, \mu)$ be a probability space and let $(\mathcal{A}_n)_{n=1}^{\infty}$ be a sequence of sub-σ-algebras of $\mathcal{A}$. For any $\varphi \in L^1(X, \mathcal{A}, \mu)$, the sequence $(E(\varphi|\mathcal{A}_n))_{n=1}^{\infty}$ is uniformly integrable.*

Proof. Without loss of generality, we may assume that $\varphi \ge 0$. Let $\varepsilon > 0$. Since $\nu(A) = \int_A \varphi \, d\mu$ is absolutely continuous with respect to μ, it follows from the Radon–Nikodym theorem (Theorem A.1.50) that there exists $\delta > 0$ such that

$$A \in \mathcal{A}, \ \mu(A) < \delta \quad \Longrightarrow \quad \int_A \varphi \, d\mu < \varepsilon. \tag{A.7}$$

Set $M > \int_X \varphi \, d\mu / \delta$. For each $n \in \mathbb{N}$, let

$$X_n(M) = \{x \in X : E(\varphi|\mathcal{A}_n)(x) \ge M\}.$$

Observe that $X_n(M) \in \mathcal{A}_n$ since $E(\varphi|\mathcal{A}_n)$ is $\mathcal{A}_n$-measurable. Therefore,

$$\mu\big(X_n(M)\big) \le \frac{1}{M} \int_{X_n(M)} E(\varphi|\mathcal{A}_n) \, d\mu = \frac{1}{M} \int_{X_n(M)} \varphi \, d\mu \le \frac{1}{M} \int_X \varphi \, d\mu < \delta$$

for all $n \in \mathbb{N}$. Consequently, by (A.7),

$$\int_{X_n(M)} E(\varphi|\mathcal{A}_n) \, d\mu = \int_{X_n(M)} \varphi \, d\mu < \varepsilon$$

for all $n \in \mathbb{N}$. Thus

$$\sup_{n\in\mathbb{N}} \int_{\{E(\varphi|\mathcal{A}_n)\geq M\}} E(\varphi|\mathcal{A}_n)\, d\mu \leq \varepsilon.$$

Since this holds for all large enough M's and since $\varepsilon > 0$ is arbitrary, we have

$$\lim_{M\to\infty} \sup_{n\in\mathbb{N}} \int_{\{E(\varphi|\mathcal{A}_n)\geq M\}} E(\varphi|\mathcal{A}_n)\, d\mu = 0,$$

that is, the sequence $(E(\varphi|\mathcal{A}_n))_{n=1}^{\infty}$ is uniformly integrable. □

Theorem A.1.67 (Martingale convergence theorem for conditional expectations). *Let $(X, \mathcal{A}, \mu)$ be a probability space and $\varphi \in L^1(X, \mathcal{A}, \mu)$. Let $(\mathcal{A}_n)_{n=1}^{\infty}$ be an ascending sequence of sub-σ-algebras of $\mathcal{A}$ and*

$$\mathcal{A}_\infty := \sigma\left(\bigcup_{n=1}^{\infty} \mathcal{A}_n\right).$$

Then

$$\lim_{n\to\infty} \left\| E(\varphi|\mathcal{A}_n) - E(\varphi|\mathcal{A}_\infty) \right\|_1 = 0 \quad \text{and} \quad \lim_{n\to\infty} E(\varphi|\mathcal{A}_n) = E(\varphi|\mathcal{A}_\infty) \;\; \mu\text{-a. e. on } X.$$

Proof. Let $\varphi_n = E(\varphi|\mathcal{A}_n)$. In Example A.1.65 and Lemma A.1.66, we have seen that $((\varphi_n, \mathcal{A}_n))_{n=1}^{\infty}$ is a uniformly integrable martingale such that

$$\lim_{n\to\infty} \varphi_n = \widehat{\varphi} \quad \mu\text{-a. e. on } X$$

for some $\widehat{\varphi} \in L^1(X, \mathcal{A}, \mu)$. For all $n \in \mathbb{N}$ the function φ_n is $\mathcal{A}_\infty$-measurable since it is $\mathcal{A}_n$-measurable and $\mathcal{A}_n \subseteq \mathcal{A}_\infty$. Thus $\widehat{\varphi}$ is $\mathcal{A}_\infty$-measurable, too. Moreover, it follows from Theorem A.1.41 that

$$\lim_{n\to\infty} \|\varphi_n - \widehat{\varphi}\|_1 = 0 \quad \text{and} \quad \lim_{n\to\infty} \int_A \varphi_n\, d\mu = \int_A \widehat{\varphi}\, d\mu, \quad \forall A \in \mathcal{A}.$$

Therefore, it just remains to show that $\widehat{\varphi} = E(\varphi|\mathcal{A}_\infty)$.

Let $k \in \mathbb{N}$ and $A \in \mathcal{A}_k$. If $n \geq k$, then $A \in \mathcal{A}_n \subseteq \mathcal{A}_\infty$, and thus

$$\int_A \varphi_n\, d\mu = \int_A E(\varphi|\mathcal{A}_n)\, d\mu = \int_A \varphi\, d\mu = \int_A E(\varphi|\mathcal{A}_\infty)\, d\mu.$$

Letting $n \to \infty$ yields

$$\int_A \widehat{\varphi}\, d\mu = \int_A E(\varphi|\mathcal{A}_\infty)\, d\mu, \quad \forall A \in \mathcal{A}_k.$$

Since k is arbitrary,

$$\int_B \widehat{\varphi}\, d\mu = \int_B E(\varphi|\mathcal{A}_\infty)\, d\mu, \quad \forall B \in \bigcup_{k=1}^{\infty} \mathcal{A}_k.$$

Since $\bigcup_{k=1}^{\infty} \mathcal{A}_k$ is a π-system generating $\mathcal{A}_\infty$ and since both $\widehat{\varphi}$ and $E(\varphi|\mathcal{A}_\infty)$ are $\mathcal{A}_\infty$-measurable, Corollary A.1.52 affirms that $\widehat{\varphi} = E(\varphi|\mathcal{A}_\infty)$ μ-a. e. □

There is also a counterpart of this theorem for descending sequences of σ-algebras.

Theorem A.1.68 (Reversed martingale convergence theorem for conditional expectations). *Let $(X, \mathcal{A}, \mu)$ be a probability space and $\varphi \in L^1(X, \mathcal{A}, \mu)$. If $(\mathcal{A}_n)_{n=1}^{\infty}$ is a descending sequence of sub-σ-algebras of $\mathcal{A}$, then*

$$\lim_{n\to\infty} \left\| E(\varphi|\mathcal{A}_n) - E\left(\varphi \;\middle|\; \bigcap_{n=1}^{\infty} \mathcal{A}_n\right) \right\|_1 = 0 \quad \textit{and} \quad \lim_{n\to\infty} E(\varphi|\mathcal{A}_n) = E\left(\varphi \;\middle|\; \bigcap_{n=1}^{\infty} \mathcal{A}_n\right) \mu\text{-}a.\,e.$$

Proof. See Theorem 35.9 in Billingsley [7]. □

Theorems A.1.67/A.1.68 are especially useful for the calculation of the conditional expectation of a function with respect to a sub-σ-algebra generated by an uncountable measurable partition which can be approached by an ascending/descending sequence of sub-σ-algebras generated by countable measurable partitions. See Exercise 8.5.22.

They can also be used to approximate a measurable set by one from a generating sequence of sub-σ-algebras.

Corollary A.1.69. *Let $(X, \mathcal{A}, \mu)$ be a probability space. Let $(\mathcal{A}_n)_{n=1}^{\infty}$ be an ascending sequence of sub-σ-algebras of $\mathcal{A}$ and set $\mathcal{A}_\infty = \sigma(\bigcup_{n=1}^{\infty} \mathcal{A}_n)$. Let $B \in \mathcal{A}_\infty$. For every $\varepsilon > 0$, there exists $A \in \bigcup_{n=1}^{\infty} \mathcal{A}_n$ such that $\mu(A \triangle B) < \varepsilon$.*

Proof. Let $B \in \mathcal{A}_\infty$. It ensues from Theorem A.1.67 that

$$\lim_{n\to\infty} E(\mathbb{1}_B|\mathcal{A}_n) = E(\mathbb{1}_B|\mathcal{A}_\infty) = \mathbb{1}_B \quad \mu\text{-a. e. on } X.$$

By Theorem A.1.47, we deduce that

$$\lim_{n\to\infty} \mu\left(\left\{x \in X : \left|E(\mathbb{1}_B|\mathcal{A}_n)(x) - \mathbb{1}_B(x)\right| \geq \frac{1}{8}\right\}\right) = 0.$$

For every $n \in \mathbb{N}$, let

$$B_n := \left\{x \in X : \left|E(\mathbb{1}_B|\mathcal{A}_n)(x) - \mathbb{1}_B(x)\right| \geq \frac{1}{8}\right\}.$$

Then there exists $N = N(\varepsilon) \in \mathbb{N}$ such that

$$\mu(B_n) \leq \varepsilon/2, \quad \forall\, n \geq N.$$

For every $n \in \mathbb{N}$, let

$$A_n := \left\{x \in X : \left|E(\mathbb{1}_B|\mathcal{A}_n)(x) - 1\right| \le \frac{1}{4}\right\} \in \mathcal{A}_n.$$

On one hand,

$$x \in B \setminus A_n \implies \left|E(\mathbb{1}_B|\mathcal{A}_n)(x) - \mathbb{1}_B(x)\right| > \frac{1}{4} \implies x \in B_n.$$

This means that

$$B \setminus A_n \subseteq B_n.$$

On the other hand,

$$x \in A_n \setminus B \implies \left|E(\mathbb{1}_B|\mathcal{A}_n)(x) - \mathbb{1}_B(x)\right| = \left|E(\mathbb{1}_B|\mathcal{A}_n)(x)\right| \ge \frac{3}{4} \implies x \in B_n.$$

This means that

$$A_n \setminus B \subseteq B_n.$$

Therefore,

$$\mu(A_n \triangle B) = \mu(A_n \setminus B) + \mu(B \setminus A_n) \le 2\mu(B_n) \le \varepsilon, \quad \forall n \ge N.$$

Since $A_n \in \mathcal{A}_n$, we have found some $A \in \bigcup_{n=1}^{\infty} \mathcal{A}_n$ with $\mu(A \triangle B) < \varepsilon$. □

We will now give a proof of Lemma A.1.32 in the case where the algebra is countable.

Proof. Let $\mathcal{A} = \{A_n\}_{n=1}^{\infty}$ be a countable algebra on a set X and let μ be a probability measure on $(X, \sigma(\mathcal{A}))$. Set $\mathcal{A}'_n = \{A_k\}_{k=1}^{n}$. Since $\mathcal{A}'_n$ is a finite set, the algebra $\mathcal{A}(\mathcal{A}'_n)$ it generates is also finite, and thus $\sigma(\mathcal{A}'_n) = \mathcal{A}(\mathcal{A}'_n)$. Moreover, since $\mathcal{A}$ is an algebra, $\sigma(\mathcal{A}'_n) = \mathcal{A}(\mathcal{A}'_n) \subseteq \mathcal{A}$. It follows that

$$\sigma(\mathcal{A}) = \sigma\left(\bigcup_{n=1}^{\infty} \sigma(\mathcal{A}'_n)\right).$$

Let $B \in \sigma(\mathcal{A})$ and $\varepsilon > 0$. Since $(\sigma(\mathcal{A}'_n))_{n=1}^{\infty}$ is an ascending sequence of sub-σ-algebras of $\sigma(\mathcal{A})$ such that $\sigma(\mathcal{A}) = \sigma(\bigcup_{n=1}^{\infty} \sigma(\mathcal{A}'_n))$, it ensues from Corollary A.1.69 that there exists $A \in \bigcup_{n=1}^{\infty} \sigma(\mathcal{A}'_n)$ with $\mu(A \triangle B) < \varepsilon$. Since $\bigcup_{n=1}^{\infty} \sigma(\mathcal{A}'_n) \subseteq \mathcal{A}$, we have found some $A \in \mathcal{A}$ such that $\mu(A \triangle B) < \varepsilon$. □

Finally, we introduce the concept of conditional measure and relate it to the concept of expected value.

Definition A.1.70. Let $(X, \mathcal{A}, \mu)$ be a probability space and let $B \in \mathcal{A}$ be such that $\mu(B) > 0$. The set function $\mu_B : \mathcal{A} \to [0, 1]$ defined by setting

$$\mu_B(A) := \frac{\mu(A \cap B)}{\mu(B)}, \quad \forall A \in \mathcal{A}$$

is a probability measure on $(X, \mathcal{A})$ called the *conditional measure of* μ *on* B.

Note that for every $\varphi \in L^1(X, \mathcal{A}, \mu)$,

$$\int_X \varphi \, d\mu_B = \int_B \varphi \, d\mu_B + \int_{X \setminus B} \varphi \, d\mu_B = \frac{1}{\mu(B)} \int_B \varphi \, d\mu + 0 = E(\varphi | B).$$

A.2 Analysis

Theorem A.2.1 (Inverse function theorem). *Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and $x_0 \in \mathcal{X}$. If F is a C^1 (i. e., continuously differentiable) function on some neighborhood of x_0 such that $F'(x_0)$ is invertible, then there exists an open neighborhood U of x_0 such that $F(U)$ is open in $\mathcal{Y}$ and F bijectively maps U to $F(U)$. Furthermore, the inverse function of $F|_U$, mapping $F(U)$ to (U), is continuously differentiable.*

Bibliography

[1] Aaronson, J., An Introduction to Infinite Ergodic Theory. Mathematical Surveys and Monographs, vol. **50**. American Mathematical Society (1997).

[2] Adler, R. L., Konheim, A. G., and McAndrew, M. H., Topological entropy. Trans. Am. Math. Soc. **114**, 309–319 (1965).

[3] Alsedá, L., Libre, J., and Misiurewicz, M., Combinatorial Dynamics and Entropy in Dimension One, 2nd edn. World Scientific (2000).

[4] Anosov, D., Geodesic flows on Riemannian manifolds with negative curvature. Proceedings of the Steklov Institute of Mathematics, vol. **90**. AMS, Providence, RI (1967).

[5] Anosov, D., Sinai, Y., Some smooth ergodic systems. Russ. Math. Surv. **22**(5), 103–167 (1967).

[6] Avila, A., Forni, G., Weak mixing for interval exchange transformations and translation flows. Ann. Math. **165**, 637–664 (2007).

[7] Billingsley, P., Probability and Measure, 3rd edn. John Wiley and Sons, New York (1995).

[8] Billingsley, P., Convergence of Probability Measures, 2nd edn. John Wiley and Sons, New York (1999).

[9] Birkhoff, G. D., Proof of the ergodic theorem. Proc. Natl. Acad. Sci. USA **17**, 656–660 (1931).

[10] Bowen, R., Entropy for group endomorphisms and homogeneous maps. Trans. Am. Math. Soc. **153**, 401–414 (1971).

[11] Bowen, R., Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms. Lect. Notes in Math., vol. **470**. Springer, (1975).

[12] Brin, M., Katok, A., On Local Entropy. In: Geometric Dynamics. Lecture Notes in Mathematics, vol. **1007**, pp. 30–38. Springer, Berlin-New York (1983).

[13] Brin, M., Stuck, G., Introduction to Dynamical Systems. Cambridge University Press, Cambridge (2002).

[14] Bryant, B. F., On expansive homeomorphisms. Pac. J. Math. **10**, 1163–1167 (1960).

[15] Conway, J. B., A Course in Functional Analysis, 2nd edn. Springer, New York (1990).

[16] Cornfeld, I. P., Fomin, S. V., and Sinai, Y. G., Ergodic Theory. Springer, New York (1982).

[17] Coven, E. M., Reddy, W. L., Positively expansive maps of compact manifolds. Lect. Notes in Math., vol. **819** (1980).

[18] Denker, M., Urbański, M., Absolutely continuous invariant measures for expansive rational maps with rationally indifferent periodic points. Forum Math. **3**, 561–579 (1991).

[19] Devaney, R. L., An Introduction to Chaotic Dynamical Systems, 2nd edn. Westview Press, Boulder, Colorado (2003).

[20] Dunford, N., Schwartz, J. T., Linear Operators, Part I: General Theory. John Wiley and Sons, New York (1964).

[21] Einsiedler, M., Ward, T., Ergodic Theory with a view towards Number Theory. Graduate Texts in Mathematics, vol. **259**. Springer, London (2011).

[22] Epstein, D., Shub, M., Expanding endomorphisms of flat manifolds. Topology **7**(2), 139–141 (1968).

[23] Frink, A. H., Distance functions and the metrization problem. Bull. Am. Math. Soc. **43**, 133–142 (1937).

[24] Goetz, A., Piecewise isometries – an emerging area of dynamical systems. http://math.sfsu.edu/goetz/Research/graz/graz.pdf.

[25] Gottschalk, W. H., Hedlund, G. A., Topological Dynamics. Amer. Math. Soc. Colloquium Publications, vol. **36**. Providence (1955).

[26] Hadamard, J., Les surfaces à courbures opposées et leur lignes géodésiques. J. Math. Pures Appl. **4**, 27–73 (1898).

[27] Halmos, P. R., Measure Theory. Van Nostrand, New York (1950).

[28] Hatcher, A., Algebraic Topology. Cambridge University Press, Cambridge (2002).

https://doi.org/10.1515/9783110702682-014

[29] Herrlich, H., Axiom of Choice. Lect. Notes in Math., vol. **1876**. Springer, Berlin Heidelberg (2006).

[30] Hirsch, M. W., Differential Topology. Graduate Texts in Mathematics, vol. **33**. Springer, New York (1976).

[31] Katok, A., Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. Publ. Math. IHÉS **51**, 137–173 (1980).

[32] Katok, A., Interval exchange transformations and some special flows are not mixing. Isr. J. Math. **35**(4), 301–310 (1980).

[33] Katok, A., Hasselblatt, B., Introduction to the Modern Theory of Dynamical Systems. Cambridge University Press, Cambridge (1996).

[34] Keane, M., Interval exchange transformations. Math. Z. **141**, 25–31 (1975).

[35] Kechris, A., Classical Descriptive Set Theory. Graduate Texts in Mathematics. Springer, New York (1995).

[36] Kesseböhmer, M., Stratmann, B. O., Fractal analysis for sets of non-differentiability of Minkowski's question mark function. J. Number Theory **128**, 2663–2686 (2008).

[37] Keynes, H. B., Robertson, J. B., Generators for topological entropy and expansiveness. Math. Syst. Theory **3**, 51–59 (1969).

[38] Kingman, J. F. C., Taylor, S. J., Introduction to Measure and Probability. Cambridge University Press, London (1966).

[39] Kitchens, B., Symbolic Dynamics: One-Sided, Two-Sided and Countable State Markov Shifts. Universitext. Springer, Berlin Heidelberg (1998).

[40] Kotus, J., Urbański, M., Existence of invariant measures for transcendental subexpanding functions. Math. Z. **243**, 25–36 (2003).

[41] Kotus, J., Urbański, M., Ergodic Theory, Geometric Measure Theory, Conformal Measures and the Dynamics of Elliptic Functions, preprint. http://arxiv.org/pdf/2007.13235.pdf.

[42] Krzyżewski, K., Szlenk, W., On invariant measures for expanding differentiable mappings. Stud. Math. **33**, 83–92 (1969).

[43] Lind, D. A., Marcus, B., An Introduction to Symbolic Dynamics and Coding. Cambridge University Press, Cambridge (1995).

[44] Martens, M., The existence of σ-finite invariant measures, Applications to real one-dimensional dynamics. Math. ArXiv, http://front.math.ucdavis.edu/math.DS/9201300.

[45] Masur, H., Interval exchange transformations and measured foliations. Ann. Math. **115**(1), 169–200 (1982).

[46] Mattila, P., Geometry of Sets and Measures in Euclidean Spaces. Cambridge University Press, Cambridge (1995).

[47] Mauldin, R. D., Urbański, M., Graph Directed Markov Systems: Geometry and Dynamics of Limit Sets. Cambridge Tracts in Mathematics, vol. **148**. Cambridge University Press, Cambridge (2003).

[48] Minkowski, H., Geometrie der Zahlen. Gesammelte Abhandlungen, vol. **2** (1911). Reprinted by Chelsea, New York, 43–52 (1967).

[49] Misiurewicz, M., A short proof of the variational principle for a $\mathbb{Z}_+^N$ action on a compact space. Astérisque **40**, 147–187 (1976).

[50] Misiurewicz, M., Szlenk, W., Entropy of piecewise monotone mappings. Stud. Math. **67**(1), 45–63 (1980).

[51] Morse, M., Hedlund, G. A., Symbolic Dynamics. Am. J. Math. **60**, 815–816 (1938).

[52] von Neumann, J., Proof of the quasi-ergodic hypothesis. Proc. Natl. Acad. Sci. USA **18**(1), 70–82 (1932).

[53] Nicol, M., Petersen, K., Ergodic Theory: Basic Examples and Constructions. Encyclopedia of Complexity and Systems Science. Springer (2009).

[54] Oxtoby, J. C., Measure and Category. Springer, New York (1980).

[55] Parthasarathy, K. R., Probability Measures on Metric Spaces. Academic Press, New York (1967).
[56] Pesin, Y. B., Dimension Theory in Dynamical Systems. Chicago Lectures in Mathematics. University of Chicago Press, Chicago (1997).
[57] Przytycki, F., Urbański, M., Conformal Fractals: Ergodic Theory Methods. London Mathematical Society Lecture Note Series, vol. **371**. Cambridge University Press (2010).
[58] Rudin, W., Real and Complex Analysis, 3rd edn. McGraw-Hill, Singapore (1987).
[59] Ruelle, D., Statistical Mechanics. Rigorous Results. Benjamin, New York (1969).
[60] Ruelle, D., Statistical mechanics on a compact set with $\mathbb{Z}^\nu$ action satisfying expansiveness and specification. Bull. Am. Math. Soc. **78**, 988–991 (1972). Transactions AMS **185**, 237–251 (1973).
[61] Ruelle, D., Thermodynamic formalism. Encyclopedia of Math. and Appl., vol. **5**. Addison-Wesley, Reading Mass (1976).
[62] Salem, R., On some singular monotonic functions which are strictly increasing. Trans. Am. Math. Soc. **53**(3), 427–439 (1943).
[63] Sears, M., Expansiveness on locally compact spaces. Math. Syst. Theory **7**(4), 377–382 (1973).
[64] Shannon, C., A Mathematical Theory of Communication. Bell Syst. Tech. J. **27**, 379–423 & 623–656 (1948).
[65] Shannon, C., The Mathematical Theory of Communication. University of Illinois Press, Urbana, Illinois (1949).
[66] Shub, M., Endomorphisms of compact differentiable manifolds. Am. J. Math. **91**(1), 175–199 (1969).
[67] Sinai, Ya. G., On the Notion of Entropy of a Dynamical System. Dokl. Russ. Acad. Sci. **124**, 768–771 (1959).
[68] Sinai, Ya. G., Gibbs Measures in Ergodic Theory. Russ. Math. Surv. **27**, 21–70 (1972).
[69] Stroppel, M., Locally Compact Groups. European Mathematical Society (2006).
[70] Sumi, H., Urbański, M., Measures and dimensions of Julia sets of semi-hyperbolic rational semigroups. Discrete Contin. Dyn. Syst., Ser. A **30**, 313–363 (2011).
[71] Szlenk, W., An Introduction to the Theory of Smooth Dynamical Systems. PWN/Wiley, Warszawa/New York (1984).
[72] Taylor, J. C., An Introduction to Measure and Probability. Springer, New York (1997).
[73] Utz, W. R., Unstable homeomorphisms. Proc. Am. Math. Soc. **1**, 769–774 (1950).
[74] Veech, W. A., Gauss measures for transformations on the space of interval exchange maps. Ann. Math. **115**(1), 201–242 (1982).
[75] Walters, P., A Variational Principle for the Pressure of Continuous Transformations. Am. J. Math. **97**, 937–971 (1975).
[76] Walters, P., An Introduction to Ergodic Theory. Graduate Text in Mathematics. Springer (2000).
[77] Willard, S., General Topology. Addison-Wesley, Mass. (1970).
[78] Williams, R. F., A note on unstable homeomorphisms. Proc. Am. Math. Soc. **6**, 308–309 (1955).
[79] Yomdin, Y., Volume growth and entropy. Isr. J. Math. **57**, 285–300 (1987).
[80] Zinsmeister, M., Thermodynamic Formalism and Holomorphic Dynamical Systems. SMF/AMS Texts and Monographs, vol. **2**. American Mathematical Society (1999).

Index

De Gruyter Expositions in Mathematics

Volume 68
Yuri A. Bahturin
Identical Relations in Lie Algebras, 2021
ISBN 978-3-11-056557-7, e-ISBN 978-3-11-056665-9

Volume 34
Marek Jarnicki, Peter Pflug
Extension of Holomorphic Functions, 2nd Edition, 2020
ISBN 978-3-11-062766-4, e-ISBN 978-3-11-063027-5

Volume 67
Alcides Lins Neto, Bruno Scárdua
Complex Algebraic Foliations, 2020
ISBN 978-3-11-060107-7, e-ISBN 978-3-11-060205-0

Volume 66
Anke Kalauch, Onno van Gaans
Pre-Riesz Spaces, 2018
ISBN 978-3-11-047539-5, e-ISBN 978-3-11-047629-3

Volume 65
Yakov Berkovich, Zvonimir Janko
Groups of Prime Power Order. Volume 6, 2018
ISBN 978-3-11-053097-1, e-ISBN 978-3-11-053314-9

Volume 64
Yakov Berkovich, Lev S. Kazarin, Emmanuel M. Zhmud'
Characters of Finite Groups. Volume 2, 2019
ISBN 978-3-11-022408-5, e-ISBN 978-3-11-022409-2

Volume 63
Yakov Berkovich, Lev S. Kazarin, Emmanuel M. Zhmud'
Characters of Finite Groups. Volume 1, 2017
ISBN 978-3-11-022406-1, e-ISBN 978-3-11-022407-8

Volume 62
Yakov Berkovich, Zvonimir Janko
Groups of Prime Power Order. Volume 5, 2016
ISBN 978-3-11-029534-4, e-ISBN 978-3-11-029535-1

Mariusz Urbański, Mario Roy, Sara Munday
Non-Invertible Dynamical Systems

De Gruyter Expositions in Mathematics

Volume 69/1

Mariusz Urbański, Mario Roy, Sara Munday

Non-Invertible Dynamical Systems

Volume 1: Ergodic Theory – Finite and Infinite, Thermodynamic Formalism, Symbolic Dynamics and Distance Expanding Maps

DE GRUYTER

Mathematics Subject Classification 2010
37A05, 37A25, 37A30, 37A35, 37A40, 37B10, 37B25, 37B40, 37B65, 37C05, 37C20, 37C40, 37D20, 37D35, 37E05, 37E10

Authors
Prof. Dr. Mariusz Urbański
University of North Texas
Department of Mathematics
1155 Union Circle #311430
Denton, TX 76203-5017
USA
mariusz@urbanskimath.com

Dr. Sara Munday
John Cabot University
Via della Lungara 233
00165 Rome
Italy
smunday@johncabot.edu

Prof. Dr. Mario Roy
York University
Glendon College
2275 Bayview Avenue
Toronto, M4N 3M6
Canada
mroy@glendon.yorku.ca

ISBN 978-3-11-070264-4
e-ISBN (PDF) 978-3-11-070268-2
e-ISBN (EPUB) 978-3-11-070275-0
ISSN 0938-6572

Library of Congress Control Number: 2021940887

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Typesetting: VTeX UAB, Lithuania
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Mariusz Urbański dedicates this book to his wife, Irena.
À mes parents Thérèse et Jean-Guy, à ma famille et à mes amis, sans qui ce livre n'aurait pu voir la vie… du fond du coeur, merci! Mario

Preface

Dynamical systems and ergodic theory is a rapidly evolving field of mathematics with a large variety of subfields, which use advanced methods from virtually all areas of mathematics. These subfields comprise but are by no means limited to: abstract ergodic theory, topological dynamical systems, symbolic dynamical systems, smooth dynamical systems, holomorphic/complex dynamical systems, conformal dynamical systems, one-dimensional dynamical systems, hyperbolic dynamical systems, expanding dynamical systems, thermodynamic formalism, geodesic flows, Hamiltonian systems, KAM theory, billiards, algebraic dynamical systems, iterated function systems, group actions, and random dynamical systems.

All of these branches of dynamical systems are mutually intertwined in many involved ways. Each of these branches nonetheless also has its own unique methods and techniques, in particular embracing methods which arise from the fields of mathematics the branch is closely related to. For example, complex dynamics borrows advanced methods from complex analysis, both of one and several variables; geodesic flows utilize methods from differential geometry; and abstract ergodic theory and thermodynamic formalism rely heavily on measure theory and functional analysis.

Indeed, it is truly fascinating how large the field of dynamical systems is and how many branches of mathematics it overlaps with. In this book, we focus on some selected subfields of dynamical systems, primarily noninvertible ones.

In the first volume, we give introductory accounts of topological dynamical systems acting on compact metrizable spaces, of finite-state symbolic dynamical systems, and of abstract ergodic theory of measure-theoretic dynamical systems acting on probability measure spaces, the latter including the metric entropy theory of Kolmogorov and Sinai. More advanced topics include infinite ergodic theory, general thermodynamic formalism, and topological entropy and pressure. This volume also includes a treatment of several classes of dynamical systems, which are interesting on their own and will be studied at greater length in the second volume: we provide a fairly detailed account of distance expanding maps and discuss Shub expanding endomorphisms, expansive maps, and homeomorphisms and diffeomorphisms of the circle.

The second volume is somewhat more advanced and specialized. It opens with a systematic account of thermodynamic formalism of Hölder continuous potentials for open transitive distance expanding systems. One chapter comprises no dynamics but rather is a concise account of fractal geometry, treated from the point of view of dynamical systems. Both of these accounts are later used to study conformal expanding repellers. Another topic exposed at length is that of thermodynamic formalism of countable-state subshifts of finite type. Relying on this latter, the theory of conformal graph directed Markov systems, with their special subclass of conformal iterated function systems, is described. Here, in a similar way to the treatment of conformal expanding repellers, the main focus is on Bowen's formula for the Hausdorff dimension

https://doi.org/10.1515/9783110702682-201

of the limit set and multifractal analysis. A rather short examination of Lasota–Yorke maps of an interval is also included in this second volume.

The third volume is entirely devoted to the study of the dynamics, ergodic theory, thermodynamic formalism, and fractal geometry of rational functions of the Riemann sphere. We present a fairly complete account of classical as well as more advanced topological theory of Fatou and Julia sets. Nevertheless, primary emphasis is placed on measurable dynamics generated by rational functions and fractal geometry of their Julia sets. These include the thermodynamic formalism of Hölder continuous potentials with pressure gaps, the theory of Sullivan's conformal measures, invariant measures and their dimensions, entropy, and Lyapunov exponents. We further examine in detail the classes of expanding, subexpanding, and parabolic rational functions. We also provide, with proofs, several of the fundamental tools from complex analysis that are used in complex dynamics. These comprise Montel's Theorem, Koebe's Distortion Theorems and Riemann–Hurwitz formulas, with their ramifications.

In virtually each chapter of this book, we describe a large number of concrete selected examples illustrating the theory and serving as examples in other chapters. Also, each chapter of the book is supplied with a number of exercises. These vary in difficulty, from very easy ones asking to verify fairly straightforward logical steps to more advanced ones enhancing largely the theory developed in the chapter.

This book originated from the graduate lectures Mariusz Urbański delivered at the University of North Texas in the years 2005–2010 and that Sara Munday took notes of. With the involvement of Mario Roy, the book evolved and grew over many years. The last 2 years (2020 and 2021) of its writing were most dramatic and challenging because of the COVID-19 pandemic. Our book borrows widely from many sources including the books [41, 47, 57]. We nevertheless tried to keep it as self-contained as possible, avoiding to refer the reader too often to specific results from special papers or books. Toward this end, an appendix comprising classical results, mostly from measure theory, functional analysis and complex analysis, is included. The book covers quite a many topics treated with various degrees of completeness, none of which are fully exhausted because of their sheer largeness and their continuous dynamical growth.

List of Figures

Introduction to Volume 1

In the first volume of this book, we give introductory accounts of topological dynamical systems acting on compact metrizable spaces, of finite-alphabet symbolic systems, and of ergodic theory of measure-theoretic dynamical systems acting on probability spaces, the latter including the metric entropy theory of Kolmogorov and Sinai. More advanced topics include infinite ergodic theory, general thermodynamic formalism, and topological entropy and pressure. This volume also includes a treatment of several classes of dynamical systems, which are interesting on their own and will be studied at greater length in the second volume: we provide a fairly detailed account of distance expanding maps and discuss Shub expanding endomorphisms, positively expansive maps, and homeomorphisms and diffeomorphisms of the circle.

We now describe the content of each chapter of this first volume in more detail, including their mutual dependence and interrelations.

Chapter 1 – Dynamical systems

In the first few sections of Chapter 1, we introduce the basic concepts in the theory of topological dynamical systems: orbits, periodic points, preperiodic points, ω-limit sets, factors, and subsystems. In particular, we introduce the concept of topological conjugacy and identify the number of periodic points of any given period as a simple (topological conjugacy) invariant. We further examine the following invariants: minimality, transitivity, topological mixing, strong transitivity, and topological exactness. Finally, we provide the first two classes of examples, namely rotations on compact topological groups and some continuous maps on compact intervals.

Chapter 2 – Homeomorphisms of the circle

In Chapter 2, we temporarily step away from the general theory of dynamical systems to consider more specific examples. We investigate homeomorphisms of the unit circle and examine the notions of lift and rotation number for homeomorphisms. Then we study in more detail the subclass of diffeomorphisms of the unit circle. The main result of this chapter is Denjoy's theorem, which states that if a C^2 diffeomorphism has an irrational rotation number, then this diffeomorphism is a minimal system which is topologically conjugate to an irrational rotation.

Chapter 3 – Symbolic dynamics

In Chapter 3, we discuss symbolic dynamical systems. We treat them as objects in their own right, but later (in Chapter 4, among others) we apply the ideas developed here to more general systems. We restrict ourselves to the case of finitely many letters, as symbolic systems born out of finite alphabets give rise to systems acting on compact metrizable spaces. Nevertheless, note that in Chapter 17 of the second volume,

https://doi.org/10.1515/9783110702682-202

we will consider countable-alphabet symbolic dynamics. In Section 3.1, we introduce full shifts. In Section 3.2, we study subshifts of finite type and in particular the characterizations of topological transitivity and exactness in terms of the underlying matrix associated with such systems. Finally, in Section 3.3 we examine general subshifts of finite type.

Chapter 4 – Distance expanding maps

In Chapter 4, we define and give some examples of distance expanding maps. In Section 4.2, we study the properties of their local inverse branches. This is a way of dealing with the noninvertibility of these maps. In Section 4.3, we examine the all important concepts of pseudo-orbit and shadowing. In Section 4.4, we introduce the powerful concept of Markov partitions and establish their existence for open, distance expanding systems. We then show in Section 4.5 how to use Markov partitions to represent symbolically the dynamics of open, distance expanding systems. This is a beautiful application of the symbolic dynamics studied in Chapter 3. The final theorem of the chapter describes the properties of the coding map between the underlying compact metric space (the phase space) and some subshift of finite type (a symbolic space).

Chapter 5 – Expansive maps

In Chapter 5, we introduce the concept of expansiveness. Amidst the large variety of dynamical behaviors, which can be thought of as expansionary in some sense, expansiveness has turned out to be a rather weak but useful notion. Indeed, all distance expanding maps are expansive and so, more particularly, all subshifts over a finite alphabet are expansive. But expansiveness is not so far from expandingness, as we demonstrate in this chapter that every expansive system is in fact expanding with respect to some metric compatible with the topology. This means that many of the results proved in Chapter 4, such as the existence of Markov partitions and of a nice symbolic representation, the density of periodic points, the closing lemma, and the shadowing property, hold for all positively expansive maps. Nevertheless, expansiveness is weaker than expandingness, and we provide at the end of the chapter a class of expansive maps that are not distance expanding. Expansive maps are important for other reasons as well. One of them is that expansiveness is a topological conjugacy invariant. More crucially, the measure-theoretic entropy function is upper semicontinuous within that class of maps. In particular, all expansive maps admit a measure of maximal entropy and, more generally, equilibrium states under all continuous potentials (see Chapter 12).

Chapter 6 – Shub expanding endomorphisms

In Section 6.2, we give a systematic account of Shub's expanding endomorphisms. These maps constitute a large, beautiful subclass of distance expanding maps and are far-reaching generalizations of the expanding endomorphisms of the circle, which will be first introduced in Section 6.1. After a digression into albegraic topology, we establish in Section 6.4 that Shub expanding endomorphisms are structurally stable, form an open set in an appropriate topology of smooth maps, are topologically exact, have at least one fixed point as well as a dense set of periodic points, and their universal covering space is diffeomorphic to $\mathbb{R}^n$.

Chapter 7 – Topological entropy

In Chapter 7, we study the central notion of topological entropy, one of the most useful and widely-applicable topological invariant thus far discovered. It was introduced to dynamical systems by Adler, Konheim, and McAndrew in 1965. Their definition was motivated by Kolmogorov and Sinai's definition of metric/measure-theoretic entropy introduced less than a decade earlier. The topological entropy of a dynamical system, which we introduce in Section 7.2, is a nonnegative extended real number that measures the complexity of the system. Topological entropy is a topological conjugacy invariant but by no means a complete invariant. In Section 7.3, we treat at length Bowen's characterization of topological entropy in terms of separated and spanning sets. In Chapter 11, we will introduce and deal with topological pressure, which is a substantial generalization of topological entropy. Our approach to topological pressure will stem from and extend that for topological entropy. In this sense, this chapter can be viewed as a preparation to Chapter 11.

Chapter 8 – Ergodic theory

In Chapter 8, we move away from the study of purely topological dynamical systems to consider instead dynamical systems that come equipped with a measure. That is, instead of self-maps acting on compact metrizable spaces, we now ask that the self-maps act upon measure spaces. We introduce in Section 8.1 the basic object of study in ergodic theory, namely, invariant measures. We also prove Poincaré's recurrence theorem. Section 8.2 presents the notion of ergodicity and comprises a demonstration of Birkhoff's ergodic theorem. This theorem is one of the most fundamental results in ergodic theory. It is extremely useful in numerous applications. The class of ergodic measures for a given transformation is then studied in more detail. The penultimate Section 8.3 contains an introduction to various measure-theoretic mixing properties that a system may satisfy, and shows that ergodicity is a very weak form of mixing. In the final Section 8.4, Rokhlin's natural extension of any given dynamical system is described and the mixing properties of this extension are investigated.

Chapter 9 –Measure-theoretic entropy

In Chapter 9, we study the measure-theoretic entropy of a (probability) measure-preserving dynamical system, also known as metric entropy or Kolmogorov–Sinai metric entropy. It was introduced by A. Kolmogorov and Ya. Sinai in the late 1950s. Since then, its account has been presented in virtually every textbook on ergodic theory. Its introduction to dynamical systems was motivated by Ludwig Boltzmann's concept of entropy in statistical mechanics and Claude Shannon's work on information theory. We first study measurable partitions in Section 9.2. Then we examine the concepts of information and conditional information in Section 9.3. In Section 9.4, we finally define the metric entropy of a measure-preserving dynamical system. And in Section 9.5, we formulate and prove the full version of Shannon–McMillan–Breiman's characterization of metric entropy. Finally, in Section 9.6 we shed further light on the nature of entropy, by proving the Brin–Katok local entropy formula. Like the Shannon–McMillan–Breiman theorem, the Brin–Katok local entropy formula is very useful in applications.

Chapter 10 – Infinite invariant measures

In Chapter 10, we deal with measurable transformations preserving measures that are no longer assumed to be finite. The outlook is then substantially different than in the case of finite measures. In Section 10.1, we investigate in detail the notions of quasi-invariant measures, ergodicity, and conservativity. We also prove Halmos' recurrence theorem, which is a generalization of Poincaré's recurrence theorem for quasi-invariant measures that are not necessarily finite. In Section 10.2, we discuss first return times, first return maps, and induced systems. We further establish relations between invariant measures for the original transformation and the induced one. In Section 10.3, we study implications of Birkhoff's ergodic theorem for finite and infinite measure spaces. Among others, we demonstrate Hopf's ergodic theorem, which applies to measure-preserving transformations of σ-finite spaces. Finally, in Section 10.4, we seek a condition under which, given a quasi-invariant probability measure, one can construct a σ-finite invariant measure which is absolutely continuous with respect to the original measure. To this end, we introduce a class of transformations, called Martens maps, that have this feature and even more. In fact, these maps have the property that any quasi-invariant probability measure admits an equivalent σ-finite invariant one. Applications of these concepts and results can be found in Chapters 13–14 of the second volume and Chapters 29–32 of the third volume.

Chapter 11 – Topological pressure

Chapter 12 – The variational principle and equilibrium states

In the last two chapters of this first volume, we introduce and extensively deal with the fundamental concepts and results of thermodynamic formalism, including topological pressure, the variational principle, and equilibrium states. This topic has a continuation throughout the whole second volume, first and perhaps most notably, in the first chapter of that volume, which is devoted to the thermodynamic formalism of distance expanding maps and Hölder continuous potentials. It will be enriched by the seminal concepts of Gibbs states and transfer (Perron–Frobenius, Ruelle, Araki) operators.

Thermodynamic formalism originated in the late 1960s with the works of David Ruelle. The motivation for Ruelle came from statistical mechanics, particularly glass lattices. The foundations, classical concepts and theorems of thermodynamic formalism were developed throughout the 1970s by Ruelle, Rufus Bowen, Peter Walters, and Yakov Sinai.

In Chapter 11, we define and investigate the properties of topological pressure. Like topological entropy, this is a topological concept and a topological conjugacy invariant. We further give Bowen's characterization of pressure in terms of separated and spanning sets.

In Chapter 12, we relate topological pressure with metric entropy by proving the variational principle, the very cornerstone of thermodynamic formalism. This principle naturally leads to the concepts of equilibrium states and measures of maximal entropy. Among others, we show that under a continuous potential every expansive dynamical system admits an equilibrium state.

Contents

Volume 2

Volume 3

1 Dynamical systems

In the first few sections of this chapter, we introduce the basic concepts in the theory of topological dynamical systems: orbits, periodic points, preperiodic points, ω-limit sets, factors, and subsystems. In particular, we introduce in Section 1.2 the concept of topological conjugacy and identify the number of periodic points of any given period as a simple (topological conjugacy) invariant. In Section 1.5, we examine the following invariants: minimality, transitivity, topological mixing, strong transitivity, and topological exactness. Finally, in Section 1.6 we provide the first two classes of examples, namely rotations on compact topological groups and some continuous maps of compact intervals.

1.1 Basic definitions

Throughout this book, a (discrete) *topological dynamical system* is a continuous map $T : X \to X$ of a nonempty compact metrizable space X. When emphasis on a metric is desirable, we write (X, d). The study of a dynamical system consists of determining its long-term behaviors, also referred to as asymptotic behaviors. That is, if we denote by T^n the nth *iterate* of T, which is defined to be

$$T^n := \underbrace{T \circ \cdots \circ T}_{n \text{ times}},$$

in order to study a dynamical system T, we investigate the sequence of iterates $(T^n)_{n=0}^{\infty}$. The long-term behavior of a point $x \in X$ can be determined by looking at this sequence of iterates evaluated at the point x.

Definition 1.1.1. Let $x \in X$. The *forward orbit* of x under T is the set

$$\mathcal{O}_+(x) := \{T^n(x) : n \geq 0\}.$$

Moreover, the *backward orbit* of x is the set

$$\mathcal{O}_-(x) := \{T^{-n}(x) : n \geq 0\} = \{T^n(x) : n \leq 0\},$$

while the *full orbit* of x is the set

$$\mathcal{O}(x) := \{T^n(x) : n \in \mathbb{Z}\} = \mathcal{O}_-(x) \cup \mathcal{O}_+(x).$$

The simplest (forward) orbits that may be observed in a dynamical system are those that consist of only finitely many points. Among these are the orbits that are cyclic.

https://doi.org/10.1515/9783110702682-001

Definition 1.1.2. A point $x \in X$ is said to be *periodic* for a system T if

$$T^n(x) = x$$

for some $n \in \mathbb{N}$. Then n is called a *period* of x. The smallest period of a periodic point x is called the *prime period* of x. The set of all periodic points of period n for T shall be denoted by $\mathrm{Per}_n(T)$. In particular, if

$$T(x) = x$$

then x is called a *fixed point* for T. The set of all fixed points will be denoted by $\mathrm{Fix}(T)$. Hence, $\mathrm{Fix}(T) = \mathrm{Per}_1(T)$. Finally, we let $\mathrm{Per}(T) = \bigcup_{n=1}^{\infty} \mathrm{Per}_n(T)$ denote the set of all periodic points for T.

Example 1.1.3.

(a) Define the map $T : [0,1] \to [0,1]$ by setting

$$T(x) := \begin{cases} 2x & \text{if } x \in [0,1/2) \\ 2-2x & \text{if } x \in [1/2,1]. \end{cases}$$

This map is known in the literature as the *tent map*. Its graph, which makes clear the reasoning behind the name, is shown in Figure 1.1. The tent map has for fixed points $\mathrm{Fix}(T) = \{0, 2/3\}$ and has 2^n periodic points of period n for each $n \in \mathbb{N}$. These are given by

$$\mathrm{Per}_n(T) = \left\{0, \frac{k}{2^n-1}, \frac{k}{2^n+1}, \frac{2^n}{2^n+1} : k \in \{2,4,6,\ldots,2^n-2\}\right\}.$$

These periodic points are the points of intersection of the graph of T^n with the diagonal line $y = x$.

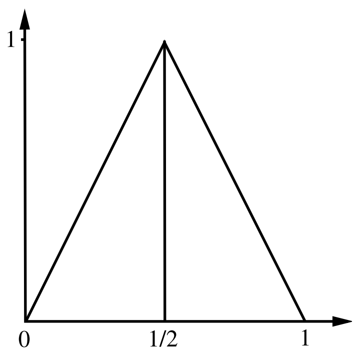

Figure 1.1: The tent map $T : [0,1] \to [0,1]$.

(b) Let $\mathbb{S}^1$ denote the unit circle, where $\mathbb{S}^1 := \mathbb{R}/\mathbb{Z}$, or, equivalently, $\mathbb{S}^1 := [0,1] \pmod 1$. Fix $m \in \mathbb{N}$ and define the map $T_m : \mathbb{S}^1 \to \mathbb{S}^1$ by setting $T_m(x) := mx \pmod 1$. One example of such a map is shown in Figure 1.2. The map T_m is simply a piecewise linear map that sends each interval $[i/m, (i+1)/m]$, for $0 \le i \le m-1$, onto $\mathbb{S}^1$. It can be expressed by the formula

$$T_m(x) = mx - i, \quad \forall x \in \left[\frac{i}{m}, \frac{i+1}{m}\right], \forall 0 \le i \le m-1.$$

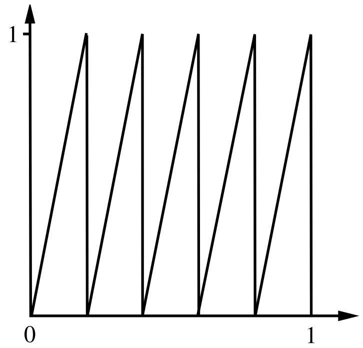

Figure 1.2: The map $T_m : [0,1] \to [0,1]$, where $m = 5$.

The map T_m has $m-1$ fixed points. They are the points of intersection of the graph of T_m with the diagonal line $y = x$. More precisely,

$$\mathrm{Fix}(T_m) = \left\{\frac{i}{m-1} : 0 \le i < m-1\right\}.$$

Similarly, it can be shown that the nth iterate T_m^n has $m^n - 1$ fixed points (see Exercise 1.7.1). We will return to this example later in the book, specifically in Chapters 4 and 9.

We now observe a general fact about convergent sequences of iterates of a point.

Lemma 1.1.4. *Let $T : X \to X$ be a dynamical system. Suppose that there exists $x \in X$ such that*

$$\lim_{n\to\infty} T^n(x) = y.$$

Then y is a fixed point for T.

Proof. Using the continuity of T, we obtain that

$$T(y) = T\Big(\lim_{n\to\infty} T^n(x)\Big) = \lim_{n\to\infty} T^{n+1}(x) = y. \qquad \square$$

Note that this fact applies only when the entire sequence of iterates converges. It does not generally hold for convergent subsequences.

Definition 1.1.5. A point $x \in X$ is said to be *preperiodic* for a system T if one of its (forward) iterates is a periodic point. That is, if there exists $k \in \mathbb{N}$ such that $T^k(x)$ is a periodic point. In other words, this means that there exists $n \in \mathbb{N}$ such that $T^{k+n}(x) = T^k(x)$.

The forward orbit $\mathcal{O}_+(x)$ is finite if and only if x is periodic or preperiodic. Equivalently, the sequence of forward iterates $(T^n(x))_{n=0}^{\infty}$ consist of mutually distinct points if and only if x is neither periodic nor preperiodic. Indeed, $\mathcal{O}_+(x)$ is infinite if and only if the sequence $(T^n(x))_{n=0}^{\infty}$ consist of mutually distinct points.

1.2 Topological conjugacy and structural stability

Suppose that we have two topological dynamical systems, $T : X \to X$ and $S : Y \to Y$. In this section, we describe a particular condition under which these two systems should be considered dynamically equivalent, that is, as dynamically "the same" in some sense. More precisely, we will establish when the orbits of two systems behave in the same way. Establishing an equivalence relation between dynamical systems can be extremely helpful, since it gives us the opportunity to apply our knowledge of systems we understand well to systems we have less information about.

Definition 1.2.1. Two dynamical systems $T : X \to X$ and $S : Y \to Y$ are said to be *topologically conjugate* if there exists a homeomorphism $h : X \to Y$, called a *conjugacy map*, such that

$$h \circ T = S \circ h.$$

In other words, T and S are topologically conjugate if there exists a homeomorphism h such that the following diagram commutes:

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[d, "h"'] & X \arrow[d, "h"] \\ Y \arrow[r, "S"'] & Y \end{tikzcd}$$

Remark 1.2.2.
(a) Topological conjugacy defines an equivalence relation on the space of all dynamical systems (see Exercise 1.7.6).
(b) If two dynamical systems T and S are topologically conjugate via a conjugacy map h, then all of their corresponding iterates are topologically conjugate by means of h. That is,

$$h \circ T^n = S^n \circ h, \quad \forall n \in \mathbb{N}.$$

Therefore, there exists a one-to-one correspondence between the orbits of T and those of S. This is why two topologically conjugate systems are considered dynamically equivalent.

Example 1.2.3. Recall the definition of the tent map from Example 1.1.3. We shall now give an example of another system that is topologically conjugate to the tent map. Define the map $F : [0,1] \to [0,1]$ by setting

$$F(x) := \begin{cases} \frac{x}{1-x} & \text{if } x \in [0, \frac{1}{2}] \\ \frac{1-x}{x} & \text{if } x \in [\frac{1}{2}, 1]. \end{cases}$$

The map F is called the *Farey map* and its graph is shown in Figure 1.3.

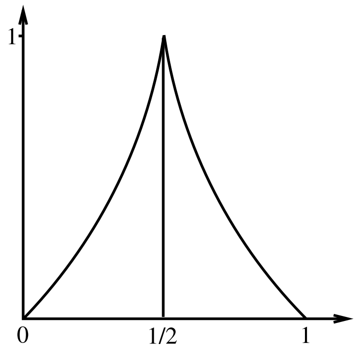

Figure 1.3: The Farey map $F : [0,1] \to [0,1]$.

The Farey map may be familiar to any reader who has studied the continued fraction expansion of real numbers, as it is related to the *Gauss map*, also known as the continued fraction map. Let us just briefly recall that a *continued fraction* is an expression of the form

$$\cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots}}},$$

where $a_i \in \mathbb{N}$ for all $i \in \mathbb{N}$. We write $[a_1, a_2, \ldots]$ for the above expression. It turns out that every continued fraction represents an irrational number in $[0,1]$ and, conversely, every irrational number in $[0,1]$ can be written as a continued fraction. This relation is a bijection.

For $x \in (1/2, 1] \setminus \mathbb{Q}$, the continued fraction representation of x is given by $[1, a_2(x), a_3(x), \ldots]$, where $a_i(x) \in \mathbb{N}$ for all $i \geq 2$. In this case, we deduce that

$$F(x) = \frac{1}{x} - 1 = 1 + [a_2(x), a_3(x), \ldots] - 1 = [a_2(x), a_3(x), \ldots].$$

For $x \in [0, 1/2] \setminus \mathbb{Q}$, the first entry of the continued fraction representation of x is strictly greater than 1 and so it follows that

$$F(x) = \frac{x}{1-x} = \frac{1}{\frac{1}{x} - 1} = [a_1(x) - 1, a_2(x), a_3(x), \ldots].$$

It is known that *Minkowski's question-mark function* is a conjugacy map between the tent map T and the Farey map F. Minkowski's question-mark function is the map $Q : [0,1] \to [0,1]$ defined by

$$Q(x) := -2\sum_{k=1}^{\infty}(-1)^k 2^{-\sum_{i=1}^{k} a_i(x)},$$

whenever $x \in [0,1] \setminus \mathbb{Q}$, where $a_i(x)$ is the ith entry of the continued fraction expansion of x. The map Q is an increasing bijection and is continuous on $[0,1] \setminus \mathbb{Q}$. Recall that a map $f : (Y, d_Y) \to (Z, d_Z)$ between two metric spaces is said to be *Hölder continuous with exponent* α if there exists a constant $C \geq 0$ such that

$$d_Z(f(x), f(y)) \leq C(d_Y(x,y))^{\alpha}, \quad \forall x, y \in Y.$$

It was shown by Salem in [62] that the map Q is Hölder continuous with exponent $\log 2/(2\log\gamma)$, where $\gamma := (1+\sqrt{5})/2$ is the golden mean. Furthermore, since $[0,1] \setminus \mathbb{Q}$ is dense in $[0,1]$, the map Q can be uniquely extended to an increasing homeomorphism of $[0,1]$ (this follows from a topological result whose proof is left to Exercise 1.7.5).

Historically, this map was designed by the German mathematician, Hermann Minkowski (1864–1909), to map the rational numbers in $[0,1]$ to the set of dyadic rational numbers $\bigcup_{n=1}^{\infty}\{i/2^n : i = 0, 1, \ldots, 2^n\}$ and the quadratic surds onto the non-dyadic rationals in an order preserving way. The graph of Q is shown in Figure 1.4. For further information on Minkowski's question-mark function, the reader is referred to [48] and [36].

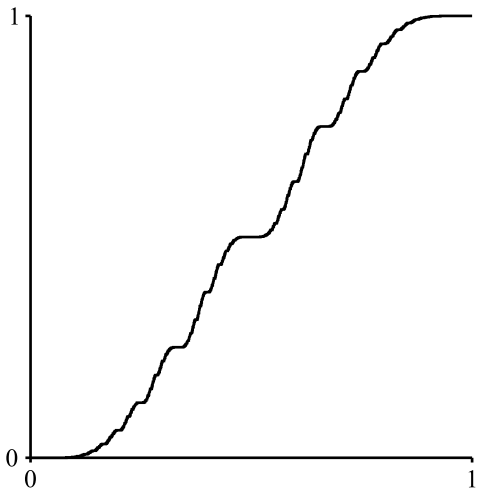

Figure 1.4: Minkowski's question-mark function $Q : [0,1] \to [0,1]$.

Let us now demonstrate that Q really does conjugate the tent and Farey systems. For this, suppose first that $x \in [0,1/2] \setminus \mathbb{Q}$. Then $Q(x) \in [0,1/2]$ and

$$\begin{aligned} T(Q(x)) &= 2\left(-2\sum_{k=1}^{\infty}(-1)^k 2^{-\sum_{i=1}^k a_i(x)}\right) \\ &= -2\left(\sum_{k=1}^{\infty}(-1)^k 2^{-(a_1(x)-1)-\sum_{i=2}^k a_i(x)}\right) \\ &= Q([a_1(x)-1, a_2(x), a_3(x), \ldots]) = Q(F(x)). \end{aligned}$$

Now, suppose that $x \in (1/2,1] \setminus \mathbb{Q}$, that is, $x = [1, a_2(x), a_3(x), \ldots]$. Then $Q(x) \in (1/2,1]$ and

$$\begin{aligned} T(Q(x)) &= 2 - 2\left(2\cdot 2^{-1} - 2\sum_{k=2}^{\infty}(-1)^k 2^{-1-\sum_{i=2}^k a_i(x)}\right) \\ &= -2\left(\sum_{k=2}^{\infty}(-1)^{k-1} 2^{-\sum_{i=2}^k a_i(x)}\right) \\ &= Q([a_2(x), a_3(x), \ldots]) = Q(F(x)). \end{aligned}$$

Thus, $T(Q(x)) = Q(F(x))$ for all $x \in [0,1] \setminus \mathbb{Q}$. Since this latter set is dense in $[0,1]$, the continuity of T, F, and Q guarantees that $T(Q(x)) = Q(F(x))$ for all $x \in [0,1]$.

Directly from the notion of topological conjugacy, we can derive the following notion of an invariant for a dynamical system.

Definition 1.2.4. A *(topological conjugacy) invariant* is a property of dynamical systems that is preserved under a topological conjugacy map.

Remark 1.2.5.
(a) By definition, topologically conjugate dynamical systems share the same set of topological conjugacy invariants. Thus, if a property is a topological conjugacy invariant and if a given dynamical system has this property while another one does not, then we can immediately deduce that these two dynamical systems are not topologically conjugate.
(b) Among the collection of invariants, there are those which are called *complete invariants*. An invariant is *complete* if two systems that share this invariant are automatically topologically conjugate. Note that this is not true of all invariants, as we will very shortly see. In fact, there are no known complete invariants that exist for arbitrary dynamical systems. Later in this chapter, we shall give examples of topological conjugacy invariants that are complete for a subfamily of dynamical systems.
(c) If $T : X \to X$ is topologically conjugate to $S : Y \to Y$ via a conjugacy map $h : X \to Y$ and if $x \in \mathrm{Per}_n(T)$, then by Remark 1.2.2(b) we deduce that

$$S^n(h(x)) = h(T^n(x)) = h(x),$$

that is, $h(x) \in \mathrm{Per}_n(S)$. Thus, h induces a one-to-one correspondence between periodic points. This correspondence preserves the prime period of a periodic point. Therefore, the number of periodic points of any given period is a topological conjugacy invariant. However, the number of periodic points of any given period is not a complete invariant. Below we give an example of two dynamical systems that have the same number of fixed points despite not being topologically conjugate. Another example will be given in Chapter 3.

(d) The cardinality of X is also an invariant, but again it is not a complete invariant, as we show in the examples below.

Example 1.2.6.

(a) Recall the maps $T_m : \mathbb{S}^1 \to \mathbb{S}^1$ defined in Example 1.1.3. It turns out that the dynamical systems $(\mathbb{S}^1, T_n)$ and $(\mathbb{S}^1, T_m)$ are not topologically conjugate whenever $n \neq m$. Indeed, by Remark 1.2.5(c), we know that if they were topologically conjugate they would have the same number of fixed points. However, T_n has n fixed points, whereas T_m has m.

(b) Let $f : [0,1] \to [0,1]$ be defined by

$$f(x) = \sqrt{x}$$

and let $g : [0,1] \to [0,1]$ be defined by

$$g(x) = 3x(1-x).$$

Then $\mathrm{Fix}(f) = \{0,1\}$ and $\mathrm{Fix}(g) = \{0, 2/3\}$. However, these two systems are not topologically conjugate. This can be seen by supposing $h : [0,1] \to [0,1]$ to be a conjugacy map between f and g. Then we would have to have either $h(0) = 0$ and $h(1) = 2/3$, or $h(0) = 2/3$ and $h(1) = 0$. In either case, it is impossible to construct a homeomorphism of the unit interval into itself satisfying these properties, as such a homeomorphism has to be either strictly increasing or strictly decreasing.

Let us now define the related concept of *structural stability*. Let (X, d) be a compact metric space and let $C(X,X)$ be the space of all continuous maps from X to X. Define the metric d_∞ on $C(X,X)$ by setting

$$d_\infty(T,S) := \sup_{x \in X} d\big(T(x), S(x)\big).$$

The topology on $C(X,X)$ induced by the metric d_∞ is called the *topology of uniform convergence* on X. This terminology is appropriate since $\lim_{n\to\infty} d_\infty(T_n, T) = 0$ if and only if the sequence $(T_n)_{n=1}^\infty$ converges to $T \in C(X,X)$ uniformly. It is not hard to see (and we leave it as an exercise for the reader) that the metric space $(C(X,X), d_\infty)$ is complete and separable.

Let $\mathcal{C}$ be an arbitrary subset of $C(X,X)$. Let τ be a topology on $\mathcal{C}$ which is finer than or coincides with the topology of uniform convergence inherited from $C(X,X)$. We say that an element T of $\mathcal{C}$ is *structurally stable* relative to $\mathcal{C}$ if there exists a neighborhood U of T in the topology τ on $\mathcal{C}$ such that for every $S \in U$ there is a homeomorphism $h = h(S) \in C(X,X)$ for which

$$h \circ T = S \circ h.$$

In other words, T is structurally stable relative to $\mathcal{C}$ if it is topologically conjugate to all systems S in one of its neighborhoods U in the topology τ on $\mathcal{C}$. The system T is *strongly structurally stable* (relative to $\mathcal{C}$) if for every $\varepsilon > 0$ there exists a neighborhood U_ε of T in the topology τ on $\mathcal{C}$ such that for every $S \in U_\varepsilon$ there is a homeomorphism $h \in B_{d_\infty}(\mathrm{Id}_X, \varepsilon)$ for which $T \circ h = h \circ S$. Here, the notation $B_{d_\infty}(g,\varepsilon)$ denotes the ε-ball around the map g:

$$B_{d_\infty}(g,\varepsilon) = \{f \in C(X,X) : d_\infty(f,g) < \varepsilon\}.$$

Later we will provide classes of structurally stable dynamical systems, most notably Shub's expanding endomorphisms (see Chapter 6).

1.3 Factors

A weaker relationship than that of topological conjugacy between two dynamical systems is that of a factor.

Definition 1.3.1. Let $T : X \to X$ and $S : Y \to Y$ be two dynamical systems. If there exists a continuous surjection $h : X \to Y$ such that $h \circ T = S \circ h$, then S is called a *factor* of T. The map h is hereafter called a *factor map*.

In general, the existence of a factor map between two systems is not sufficient to make them topologically conjugate. Nonetheless, if S is a factor of T, then every orbit of T is projected to an orbit of S. As every factor map is by definition surjective, this means that all of the orbits of S have an analogue in T. However, as a factor map needs not be injective, more than one orbit of T may be projected to the same orbit of S. In other words, some orbits of S may have more than one analogue in T. Therefore, the dynamical system T can usually be thought of as more "complicated" than the factor S. In particular, periodic points of period n for T are projected to periodic points for S whose periods are factors of n.

Example 1.3.2. Let $T : X \to X$ be a dynamical system and let $S : Y \to Y$ be given by $Y := \{y\}$ and S equal to the identity map. Then the map $h : X \to Y$ defined by $h(x) := y$ for all $x \in X$ is a factor map. This is, of course, a trivial example. In Chapter 3, we will encounter a class of nontrivial examples.

1.4 Subsystems

Our next aim is to introduce the concept of a subsystem of a dynamical system. In order to do this, we first define the notion of invariance for sets.

Definition 1.4.1. Let $T : X \to X$ be a dynamical system. A subset F of X is said to be
(a) *forward T-invariant* if $T^{-1}(F) \supseteq F$.
(b) *backward T-invariant* if $T^{-1}(F) \subseteq F$.
(c) *completely T-invariant* if $T^{-1}(F) = F$.

If the identity of the map T is clear, then we will sometimes omit it. We also often refer to forward invariant sets simply as "invariant." Note that the condition of being forward invariant is equivalent to $T(F) \subseteq F$.

Remark 1.4.2.
(a) A set is completely invariant if and only if it is both forward and backward invariant.
(b) The closure of an invariant set is invariant.
(c) A set F is invariant if and only if it is equal to the union of the forward orbits of all of its points, that is, $F = \bigcup_{x\in F} \mathcal{O}_+(x)$.
(d) A closed set F is invariant if and only if it is equal to the union of the closure of the forward orbit of all of its points, that is, $F = \bigcup_{x\in F} \overline{\mathcal{O}_+(x)}$. By (c), this means that

$$\overline{\bigcup_{x\in F} \mathcal{O}_+(x)} = F = \bigcup_{x\in F} \overline{\mathcal{O}_+(x)}.$$

We are now in a position to define the concept of subsystem.

Definition 1.4.3. Let $T : X \to X$ be a dynamical system. If $F \subseteq X$ is a closed T-invariant set, then the dynamical system induced by the restriction of T to F, that is, $T|_F : F \to F$ is called a *subsystem* of $T : X \to X$.

Note that as X is a compact metrizable space and, therefore, a compact Hausdorff space, the word "closed" can be replaced by "compact" in the above definition.

Remark 1.4.4. If a dynamical system $S : Y \to Y$ is a factor of a dynamical system $T : X \to X$ via a factor map $h : X \to Y$ and if $Z \subseteq Y$ is forward (resp., backward/completely) S-invariant, then $h^{-1}(Z)$ is forward (resp., backward/completely) T-invariant. Indeed, if $Z \subseteq Y$ is forward S-invariant, that is, $S^{-1}(Z) \supseteq Z$, then

$$T^{-1}(h^{-1}(Z)) = (h \circ T)^{-1}(Z) = (S \circ h)^{-1}(Z) = h^{-1}(S^{-1}(Z)) \supseteq h^{-1}(Z),$$

that is, $h^{-1}(Z)$ is forward T-invariant.

In particular, if $S|_Z$ is a subsystem of $S : Y \to Y$ then $T|_{h^{-1}(Z)}$ is a subsystem of $T : X \to X$. This uses the fact that a compact subset of a Hausdorff space is closed, that the preimage of a closed set under a continuous map is closed, and that a closed subset of a compact space is compact.

If the two systems are topologically conjugate, then any conjugacy map h induces a one-to-one correspondence between the T-invariant sets and the S-invariant sets. In particular, h induces a one-to-one correspondence between the subsystems of T and those of S.

Observe that every orbit is T-invariant, since for every $x \in X$ we have

$$T(\mathcal{O}_+(x)) = \{T(T^n(x)) : n \geq 0\} = \{T^{n+1}(x) : n \geq 0\} \subseteq \mathcal{O}_+(x).$$

By Remark 1.4.2(b), we deduce that the closure of every orbit is T-invariant and, therefore, the restriction of a system to the closure of any of its orbits constitutes a subsystem of that system.

Above and beyond the orbits of a system, the limit points, sometimes called accumulation points, of these orbits are also of interest.

Definition 1.4.5. Let $x \in X$. The set of limit points of the sequence of forward iterates $(T^n(x))_{n=0}^{\infty}$ of x is called the *ω-limit set* of x. It is denoted by $\omega(x)$.

In other words, $y \in \omega(x)$ if and only if there exists a strictly increasing sequence $(n_j)_{j=1}^{\infty}$ of nonnegative integers such that $\lim_{j\to\infty} T^{n_j}(x) = y$.

Remark 1.4.6.
(a) In general, the ω-limit set of a point x is not the set of limit points of the forward orbit $\mathcal{O}_+(x)$ of x. See Exercises 1.7.12, 1.7.13, and 1.7.14.
(b) By the very definition of an ω-limit set, it is easy to see that $\omega(x) \subseteq \overline{\mathcal{O}_+(x)}$. In fact, $\mathcal{O}_+(x) \cup \omega(x) = \overline{\mathcal{O}_+(x)}$. See Exercises 1.7.13 and 1.7.15.

Proposition 1.4.7. *Every ω-limit set is nonempty, closed, and T-invariant. Furthermore, for every $x \in X$ we have that $T(\omega(x)) = \omega(x)$.*

Proof. Let $x \in X$. Since X is compact, the set $\omega(x)$ is nonempty. Moreover, the set $\omega(x)$ is closed as the limit points of any sequence form a closed set (we leave the proof of this fact to Exercise 1.7.16). It only remains to show that $T(\omega(x)) = \omega(x)$. Let $y \in \omega(x)$. Then there exists a strictly increasing sequence $(n_j)_{j=1}^{\infty}$ of nonnegative integers such that $\lim_{j\to\infty} T^{n_j}(x) = y$. The continuity of T then ensures that

$$T(y) = T\Big(\lim_{j\to\infty} T^{n_j}(x)\Big) = \lim_{j\to\infty} T(T^{n_j}(x)) = \lim_{j\to\infty} T^{n_j+1}(x).$$

This shows that $T(y) \in \omega(x)$ and, in turn, proves that $T(\omega(x)) \subseteq \omega(x)$. To establish the reverse inclusion, again fix $y \in \omega(x)$. Then there exists a strictly increasing sequence $(n_j)_{j=1}^{\infty}$ of positive integers such that $\lim_{j\to\infty} T^{n_j}(x) = y$. Consider the sequence $(T^{n_j-1}(x))_{j=1}^{\infty}$. Since X is compact, this sequence admits a convergent subsequence

$(T^{n_{j_k}-1}(x))_{k=1}^{\infty}$, where $(n_{j_k})_{k=1}^{\infty}$ is some subsequence of $(n_j)_{j=1}^{\infty}$. Let $z := \lim_{k\to\infty} T^{n_{j_k}-1}(x)$. Then $z \in \omega(x)$ and

$$T(z) = T\Big(\lim_{k\to\infty} T^{n_{j_k}-1}(x)\Big) = \lim_{k\to\infty} T^{n_{j_k}}(x) = \lim_{j\to\infty} T^{n_j}(x) = y.$$

Consequently, $y \in T(\omega(x))$. This proves that $\omega(x) \subseteq T(\omega(x))$. □

The above proposition shows in particular that the restriction of a dynamical system to any of its ω-limit sets is a subsystem of that system.

If T is a homeomorphism, then we can define the counterpart of an ω-limit set by looking at the backward iterates of a point. For $x \in X$, we define the *α-limit set* of x as the set of accumulation points of the sequence of backward iterates $(T^{-n}(x))_{n=0}^{\infty}$ of x. It is denoted by $\alpha(x)$. In this case, we have that $y \in \alpha(x)$ if and only if there exists a strictly increasing sequence $(n_j)_{j=1}^{\infty}$ of nonnegative integers such that $\lim_{j\to\infty} T^{-n_j}(x) = y$. By the definition of the α-limit set, we have that $\alpha(x)$ is contained in the closure of the backward orbit $\mathcal{O}_-(x)$. The α-limit sets satisfy the same properties under T^{-1} as those of the ω-limit sets under T.

Definition 1.4.8. Let $T : X \to X$ be a topological dynamical system. A point x is said to be *wandering* for T if there exists an open neighborhood U of x such that the preimages of U are mutually disjoint, that is,

$$T^{-m}(U) \cap T^{-n}(U) = \emptyset, \quad \forall m \neq n \geq 0.$$

Accordingly, a point x is called *nonwandering* for T if each of its open neighborhoods U revisits itself under iteration by T, that is, for each neighborhood U of x there is $n \in \mathbb{N}$ such that $T^{-n}(U) \cap U \neq \emptyset$. The *nonwandering set* for T, which consists of all nonwandering points, is denoted by $\Omega(T)$.

Theorem 1.4.9. *The nonwandering set $\Omega(T)$ of a system $T : X \to X$ enjoys the following properties:*
(a) $\Omega(T)$ *is closed.*
(b) $\emptyset \neq \bigcup_{x\in X} \omega(x) \subseteq \Omega(T)$.
(c) $\mathrm{Per}(T) \subseteq \Omega(T)$.
(d) $\Omega(T)$ *is forward T-invariant.*
(e) *If T is a homeomorphism, then $\Omega(T) = \Omega(T^{-1})$ and is completely T-invariant.*

Proof.
(a) The nonwandering set $\Omega(T)$ is closed since its complement, the set of wandering points $X \setminus \Omega(T)$, is open. Indeed, if a point x is wandering, then there exists an open neighborhood U of x such that the preimages of U are mutually disjoint. Therefore, all points of U are wandering as well. So $X \setminus \Omega(T)$ is open.

(b) Let $x \in X$ and $y \in \omega(x)$. Then there is a strictly increasing sequence $(n_k)_{k=1}^{\infty}$ of nonnegative integers such that $\lim_{k\to\infty} T^{n_k}(x) = y$. Thus, given any open neighborhood U of y, there are numbers $n_k < n_l$ such that $T^{n_k}(x) \in U$ and $T^{n_l}(x) \in U$. Then, letting $n = n_l - n_k$ and $z = T^{n_k}(x)$, we have $z \in U$ and $T^n(z) \in U$, that is, $T^{-n}(U) \cap U \neq \emptyset$. As this is true for every open neighborhood U of y, we deduce that $y \in \Omega(T)$. Hence, $\bigcup_{x\in X} \omega(x) \subseteq \Omega(T)$. In particular, $\Omega(T) \neq \emptyset$ since $\omega(x) \neq \emptyset$ for every x.

(c) Since $\omega(x) = \mathcal{O}_+(x) \ni x$ for every periodic point x, all the periodic points of T belong to $\Omega(T)$. More simply, every periodic point is nonwandering as it eventually returns to itself under iteration.

(d) Let $x \in \Omega(T)$ and U an open neighborhood of $T(x)$. Then $T^{-1}(U)$ is an open neighborhood of x. As $x \in \Omega(T)$, there exists $n \in \mathbb{N}$ such that $T^{-n}(T^{-1}(U)) \cap T^{-1}(U) \neq \emptyset$. That is, $T^{-1}(T^{-n}(U) \cap U) \neq \emptyset$, which implies that $T^{-n}(U) \cap U \neq \emptyset$. Since this is true for every open neighborhood U of $T(x)$, we conclude that $T(x) \in \Omega(T)$, and hence $T(\Omega(T)) \subseteq \Omega(T)$.

(e) Suppose T is a homeomorphism. It is easy to show that $\Omega(T^{-1}) = \Omega(T)$. By (d), we then have $T(\Omega(T)) \subseteq \Omega(T)$ and $T^{-1}(\Omega(T)) \subseteq \Omega(T)$. This implies $T^{-1}(\Omega(T)) = \Omega(T)$.

Parts (a) and (d) tell us that the restriction of a system to its nonwandering set forms a subsystem of that system. Part (b) reveals that this subsystem comprises all ω-limit subsystems. □

Finally, we introduce the notion of invariance for a function.

Definition 1.4.10. A continuous function $g : X \to \mathbb{R}$ is said to be T-*invariant* if $g \circ T = g$.

Remark 1.4.11. A function g is T-invariant if and only if $g \circ T^n = g$ for every $n \in \mathbb{N}$. In other words, g is T-invariant if and only if g is constant along each orbit of T, and thus if and only if g is constant on the closure of each orbit of the system T.

1.5 Mixing and irreducibility

In this section, we investigate various forms of topological mixing and irreducibility that can be observed in some dynamical systems. For a dynamical system, topological mixing can intuitively be conceived as witnessing some parts of the underlying space becoming mixed under iteration with other parts of the space. Irreducibility of a dynamical system means that the system does not admit any “nontrivial” subsystem, which takes a different meaning depending on the stronger or weaker form of irreducibility. In any case, the absence of nontrivial subsystems forces irreducible systems to exhibit some form of mixing.

1.5.1 Minimality

We will now define one way in which a dynamical system $T : X \to X$ can be said to be irreducible. As mentioned above, by irreducibility we mean that T admits no nontrivial subsystem, for some sense of nontriviality. One natural form of irreducibility would be that the only subsystems of T are the empty system and T itself. Another way of saying this is that the only closed T-invariant subsets of X are the empty set and the whole of X. The concept we need for this is minimality.

Definition 1.5.1. Let $T : X \to X$ be a dynamical system. A set $F \subseteq X$ is said to be a *minimal set* for T if the following three conditions are satisfied:
(a) The set F is nonempty and closed.
(b) The set F is T-invariant.
(c) If $G \subseteq F$ is nonempty, closed and T-invariant, then $G = F$.

A minimal set F induces the *minimal subsystem* $T|_F : F \to F$. We now address the question of the existence of minimal sets.

Theorem 1.5.2. *Every dynamical system admits a minimal set (that induces a minimal subsystem).*

Proof. Let $\mathcal{F}$ be the family of all nonempty, closed, T-invariant subsets of X. This family is nonempty, as it at least contains X. It is also partially ordered under the relation of backward set inclusion $\supseteq$. We shall use the Kuratowski–Zorn lemma (often referred to simply as Zorn's lemma) to establish the existence of a minimal set. Accordingly, let $\{F_\lambda\}_{\lambda\in\Lambda}$ be a chain in $\mathcal{F}$, that is, a totally ordered subset of $\mathcal{F}$, and let $F = \bigcap_{\lambda\in\Lambda} F_\lambda$. Then F is nonempty and closed (cf. Exercise 1.7.17). Moreover,

$$T(F) = T\Big(\bigcap_{\lambda\in\Lambda} F_\lambda\Big) \subseteq \bigcap_{\lambda\in\Lambda} T(F_\lambda) \subseteq \bigcap_{\lambda\in\Lambda} F_\lambda = F.$$

Thus, F is T-invariant and constitutes the maximal element of the chain $\{F_\lambda\}_{\lambda\in\Lambda}$. Hence, every chain in $\mathcal{F}$ has a maximal element, and by Zorn's lemma, we infer that the family $\mathcal{F}$ has a maximal element under the relation of backward set inclusion $\supseteq$, that is, the family $\mathcal{F}$ has a minimal element under the relation of set inclusion $\subseteq$. This element is a minimal set for T. □

We can now define the concept of minimality for dynamical systems.

Definition 1.5.3. A dynamical system $T : X \to X$ is said to be *minimal* if X is a minimal set for T (and thus is the only minimal set for T).

Minimality is a strong form of irreducibility since minimal systems admit no nonempty subsystems other than themselves.

Let us now give a characterization of minimal sets. In particular, the next result shows that the strong form of irreducibility that we call minimality is also a strong form of mixing, since minimal systems are characterized by having only dense orbits.

Theorem 1.5.4. *Let F be a nonempty closed T-invariant subset of X. Then the following three statements are equivalent:*
(a) *F is minimal.*
(b) *$\omega(x) = F$ for every $x \in F$.*
(c) *$\overline{\mathcal{O}_+(x)} = F$ for every $x \in F$.*

Proof. We shall prove this theorem by establishing the sequence of implications (a)⇒(b)⇒(c)⇒(a).

To begin, suppose that F is a minimal set for T, and let $x \in F$. Then, since F is T-invariant and closed, we obtain that $\omega(x) \subseteq \overline{\mathcal{O}_+(x)} \subseteq F$. Moreover, in light of Proposition 1.4.7, $\omega(x)$ is nonempty, closed and T-invariant. So, by the definition of a minimal set, we must have that $\omega(x) = F$. This proves that (a) implies (b).

Toward the proof of the second implication, recall that as F is T-invariant and closed, we have that $\omega(x) \subseteq \overline{\mathcal{O}_+(x)} \subseteq F$. Thus, if $\omega(x) = F$ then $\overline{\mathcal{O}_+(x)} = F$. This proves that (b) implies (c).

Finally, assume that $\overline{\mathcal{O}_+(x)} = F$ for every $x \in F$. Let $E \subseteq F$ be a nonempty closed T-invariant set. It suffices to show that $E = F$. To that end, let $x \in E$. As E is T-invariant and closed, we have $\overline{\mathcal{O}_+(x)} \subseteq E$. Moreover, $x \in E \subseteq F$ implies that $\overline{\mathcal{O}_+(x)} = F$. Therefore, $F = \overline{\mathcal{O}_+(x)} \subseteq E \subseteq F$, and hence $E = F$. Thus, F is minimal. This proves the remaining implication, namely, that (c) implies (a). □

Remark 1.5.5.
(a) Theorem 1.5.4(c) characterizes a minimal set F by the requirement that the orbit of each point of F stays in F and is dense in F. Another way to think of this is that irreducibility in the sense that a system admits no nonempty subsystem other than itself, is equivalent to mixing in the sense that every orbit is dense. In particular, a consequence of this property is that a minimal system must be surjective. It also implies that an infinite minimal system does not admit any periodic point, as $\overline{\mathcal{O}_+(x)} = \mathcal{O}_+(x)$ for any periodic point x.
(b) Minimality is a topological conjugacy invariant. However, it is not a complete invariant (see Exercise 1.7.19).

1.5.2 Transitivity and topological mixing

In this section, we introduce a weaker form of mixing called transitivity. We have shown in the previous section that minimal systems have only dense orbits; transitive systems are only required to exhibit one dense orbit. Nonetheless, as we shall soon see, the existence of one dense orbit forces the existence of a dense G_δ-set of points with dense orbits.

Definition 1.5.6. Let $T : X \to X$ be a dynamical system.
(a) A point $x \in X$ is said to be *transitive* for T if $\omega(x) = X$.
(b) The system T is called *transitive* if it admits at least one transitive point.
(c) A point $x \in X$ is said to be *weakly transitive* for T if $\overline{\mathcal{O}_+(x)} = X$.
(d) The system T is said to be *weakly transitive* if it admits at least one weakly transitive point.

Remark 1.5.7.
(a) In light of Theorem 1.5.4(b), every minimal system is transitive. There are, of course, transitive systems which are not minimal. For instance, we shall see in Chapter 3 that full shifts are transitive, but not minimal since they admit periodic points.
(b) Transitivity is a topological conjugacy invariant. However, it is not a complete invariant. Indeed, as minimality is not a complete invariant and transitivity is weaker than minimality, transitivity cannot be a complete invariant.
(c) A transitive system is surjective since, given any transitive point x, we have that $T(X) = T(\omega(x)) = \omega(x) = X$.
(d) As $\overline{\mathcal{O}_+(x)} = \mathcal{O}_+(x) \cup \omega(x)$, every transitive system is weakly transitive. Note that there are weakly transitive systems which are not transitive, as Example 1.5.8 below demonstrates.
(e) If $T : X \to X$ is weakly transitive, then every continuous T-invariant function is constant. To see this, let $g : X \to \mathbb{R}$ be a T-invariant continuous function. Also let $x \in X$ be such that the forward orbit of x is dense in X. Then by Remark 1.4.11, we have that $g|_{\mathcal{O}_+(x)} = g(x)$. This means that the continuous function g is constant on a dense set of points, so it must be constant everywhere.

Example 1.5.8. Let $X = \{0\} \cup \{1/n : n \in \mathbb{N}\} \subseteq \mathbb{R}$, and let $T : X \to X$ be defined by $T(0) = 0$ and $T(1/n) = 1/(n+1)$. Then T is continuous. Moreover, as it is not surjective (its range does not include 1), T cannot be transitive. Alternatively, we might argue that $\omega(x) = \{0\}$ for every $x \in X$. However, observe that $\mathcal{O}_+(1) = \{1, 1/2, 1/3, \ldots\} = \{1/n : n \in \mathbb{N}\}$. So, $\overline{\mathcal{O}_+(1)} = X$ and, therefore, T is weakly transitive.

In fact, it turns out that surjectivity is the only difference between weakly transitive and transitive systems, as we now show.

Theorem 1.5.9. *A dynamical system is transitive if and only if it is weakly transitive and surjective.*

Proof. We have already observed in Remark 1.5.7 that transitive systems are weakly transitive and surjective. Suppose now that a system $T : X \to X$ is weakly transitive and surjective. Let $x \in X$ be such that $\overline{\mathcal{O}_+(x)} = X$. Since $X = \overline{\mathcal{O}_+(x)} = \mathcal{O}_+(x) \cup \omega(x)$, we deduce that $X \setminus \mathcal{O}_+(x) \subseteq \omega(x)$.

On one hand, if $T^{-1}(x) \cap \mathcal{O}_+(x) = \emptyset$ then by the surjectivity of T we have that $\emptyset \neq T^{-1}(x) \subseteq X \setminus \mathcal{O}_+(x) \subseteq \omega(x)$. As $\omega(x)$ is T-invariant, we obtain that $\{x\} = T(T^{-1}(x)) \subseteq T(\omega(x)) = \omega(x)$. Using the T-invariance of $\omega(x)$ once again, we deduce that $\mathcal{O}_+(x) \subseteq \omega(x)$. Therefore $X = \overline{\mathcal{O}_+(x)} \subseteq \omega(x) \subseteq X$, that is, $\omega(x) = X$.

On the other hand, if $T^{-1}(x) \cap \mathcal{O}_+(x) \neq \emptyset$, then x is a periodic point, and hence $\omega(x) = \mathcal{O}_+(x)$. It follows that $X = \overline{\mathcal{O}_+(x)} = \mathcal{O}_+(x) \cup \omega(x) = \mathcal{O}_+(x) = \omega(x)$. (That is, every point in X is a periodic point in the orbit of x and X is finite.)

In either case, we have demonstrated that x is a transitive point. □

Let us now introduce the concept of topological mixing. We shall very shortly see the connection between this notion and transitivity.

Definition 1.5.10. A dynamical system $T : X \to X$ is said to be *topologically mixing* if any of the following equivalent statements hold:

(a) For all nonempty open subsets U and V of X, there exists $n \in \mathbb{N}$ such that $T^n(U) \cap V \neq \emptyset$.

(b) For all nonempty open subsets U and V of X, there exists $n \in \mathbb{N}$ such that $U \cap T^{-n}(V) \neq \emptyset$.

(c) For all nonempty open subsets U and V of X and for all $N \in \mathbb{N}$, there exists $n \geq N$ such that $T^n(U) \cap V \neq \emptyset$.

(d) For all nonempty open subsets U and V of X and for all $N \in \mathbb{N}$, there exists $n \geq N$ such that $U \cap T^{-n}(V) \neq \emptyset$.

(e) For all nonempty open subsets U and V of X, there exist infinitely many $n \in \mathbb{N}$ such that $T^n(U) \cap V \neq \emptyset$.

(f) For all nonempty open subsets U and V of X, there exist infinitely many $n \in \mathbb{N}$ such that $U \cap T^{-n}(V) \neq \emptyset$.

(g) $\overline{\bigcup_{n\in\mathbb{N}} T^n(U)} = X$ for every nonempty open subset U of X.

(h) $\overline{\bigcup_{n\in\mathbb{N}} T^{-n}(U)} = X$ for every nonempty open subset U of X.

We leave it to the reader to provide a proof that (a) and (b) are equivalent, (c) and (d) are equivalent and (e) and (f) are equivalent. It is also straightforward to show the chain of equivalences (b)⇔(d) ⇔(f) by using the fact that $W := T^{-(N-1)}(V)$ is an open set. Statement (g) is just a rewriting of (a) while (h) is a reformulation of (b).

Furthermore, note that each of these statements implies that T is surjective. Indeed, since $T(X) \subseteq X$, we have by induction that $T^{n+1}(X) \subseteq T^n(X)$ for all $n \in \mathbb{N}$. So the sequence of compact sets $(T^n(X))_{n=0}^{\infty}$ is descending. By (g), $T(X) = \overline{T(X)} = \overline{\bigcup_{n\in\mathbb{N}} T^n(X)} = X$.

The next theorem gives the promised connection between transitivity and topological mixing, in addition to another characterization of transitive systems in terms of nowhere-dense sets. For more information on nowhere-dense sets, the reader is referred to [77] and [54]. The most relevant fact for us is that a closed set is nowhere-dense if and only if it has empty interior.

Theorem 1.5.11. *If $T : X \to X$ is a surjective dynamical system, then the following statements are equivalent:*
(a) *T is transitive.*
(b) *Whenever F is a closed T-invariant subset of X, either $F = X$ or F is nowhere-dense. In other words, T admits no subsystem with nonempty interior other than itself.*
(c) *T is topologically mixing.*
(d) *$\{x \in X : \overline{\mathcal{O}_+(x)} = X\}$ is a dense G_δ-subset of X.*
(e) *$\{x \in X : \overline{\mathcal{O}_+(x)} = X\}$ contains a dense G_δ-subset of X.*

Proof. We shall prove this theorem by establishing the implications (a)⇒(b)⇒(c)⇒(d)⇒(e)⇒(a).

To prove the first implication, let x be a transitive point and F a nonempty closed T-invariant set. Since x is transitive, we have that $\omega(x) = X$. Suppose that F has nonempty interior. Then there is a nonempty open set $U \subseteq F$. Therefore, there exists $p \geq 0$ with $T^p(x) \in U \subseteq F$. As F is T-invariant, all higher iterates of x lie in F as well. Since F is closed, this implies that $\omega(x) \subseteq F$. Therefore $X = \omega(x) \subseteq F \subseteq X$, that is, $F = X$. Thus, either F has empty interior or $F = X$. That is, either F is nowhere-dense or $F = X$. This proves that (a) implies (b).

For the second implication, suppose that (b) holds and that U and V are nonempty open subsets of X. By the surjectivity of T, the union $\bigcup_{n=1}^{\infty} T^{-n}(V)$ is a nonempty open subset of X. Therefore, the closed set $F := X \setminus \bigcup_{n=1}^{\infty} T^{-n}(V) \neq X$ satisfies

$$\begin{aligned} T^{-1}(F) &= X \setminus T^{-1}\left(\bigcup_{n=1}^{\infty} T^{-n}(V)\right) \\ &= X \setminus \bigcup_{n=2}^{\infty} T^{-n}(V) \supseteq X \setminus \bigcup_{n=1}^{\infty} T^{-n}(V) = F. \end{aligned}$$

This means that $T(F) \subseteq F$, that is, the set F is T-invariant. Thus, either $F = X$ or F is nowhere-dense. As $F \neq X$, the set F is nowhere-dense and its complement $\bigcup_{n=1}^{\infty} T^{-n}(V)$ is dense in X. Consequently, there exists some $n \in \mathbb{N}$ such that $U \cap T^{-n}(V) \neq \emptyset$. So T is topologically mixing. This establishes that (b) implies (c).

Now, suppose that T is topologically mixing and let $\{U_n : n \in \mathbb{N}\}$ be a countable base for the topology of X. Such a base exists since X is a compact metrizable space. Then

$$\{x \in X : \overline{\mathcal{O}_+(x)} = X\} = \bigcap_{n=1}^{\infty} \bigcup_{m=0}^{\infty} T^{-m}(U_n).$$

Since T is topologically mixing, for each $n \in \mathbb{N}$ the open set $\bigcup_{m=0}^{\infty} T^{-m}(U_n)$ intersects every nonempty open set, that is, this set is dense in X. By the Baire category theorem, it follows that $\bigcap_{n=1}^{\infty} \bigcup_{m=0}^{\infty} T^{-m}(U_n)$ is a dense G_δ-set. This proves that (c) implies (d).

The implication (d)⇒(e) is obvious.

Finally, suppose that $\{x \in X : \overline{\mathcal{O}_+(x)} = X\}$ contains a dense G_δ-set. This implies immediately that T is weakly transitive. As T is surjective, we deduce from Theorem 1.5.9 that T is in fact transitive. This demonstrates that (e) implies (a). □

In particular, Theorem 1.5.11 shows that transitivity corresponds to a weaker form of irreducibility than minimality. Indeed, as opposed to minimal systems which admit only the empty set and the whole of X as subsystems, nonminimal transitive systems admit nontrivial subsystems. Each of these subsystems is nevertheless nowhere-dense.

Rotations of the unit circle

We shall now discuss rotations of the unit circle $\mathbb{S}^1$. The unit circle may be defined in many different homeomorphic ways. It may be embedded in the complex plane by defining $\mathbb{S}^1 := \{z \in \mathbb{C} : |z| = 1\}$. It may also be defined to be the set of all angles $\theta \in [0, 2\pi]$ (mod 2π). Alternatively, as we have already seen in Example 1.1.3(b), it may be defined to be the quotient space $\mathbb{S}^1 := \mathbb{R}/\mathbb{Z}$ or as $\mathbb{S}^1 := [0, 1]$ (mod 1). We will use the form most appropriate to each specific situation.

Let $\mathbb{S}^1 = [0, 2\pi]$ (mod 2π). Let $\alpha \in \mathbb{R}$ and define the map $T_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ by

$$T_\alpha(x) = x + 2\pi\alpha \pmod{2\pi}.$$

Thus, T_α is the rotation of the unit circle by the angle $2\pi\alpha$. The dynamics of T_α are radically different depending on whether the number α is rational or irrational. We prove the following classical result.

Theorem 1.5.12. *Let $T_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ be defined as above. The following are equivalent:*
(a) $\alpha \in \mathbb{R}\backslash\mathbb{Q}$.
(b) T_α *is minimal.*
(c) T_α *is transitive.*

Moreover, when $\alpha \in \mathbb{Q}$ every point in $\mathbb{S}^1$ is a periodic point with the same prime period.

Proof. We shall prove that (b)⇒(c)⇒(a)⇒(b). Remark 1.5.7 already pointed out that (b)⇒(c).

[(c)⇒(a)] If $\alpha \in \mathbb{Q}$, say $\alpha = p/q$ for some $p, q \in \mathbb{Z}$ with p and q relatively prime and $q > 0$, then $T_\alpha^n(x) = x + 2\pi pn/q$ (mod 2π) for all $x \in \mathbb{S}^1$. In particular, $T_\alpha^q(x) = x + 2\pi p$ (mod 2π) $= x$ (mod 2π) for all $x \in \mathbb{S}^1$. Hence, T_α^q is the identity map, that is, every point in $\mathbb{S}^1$ is a periodic point of (prime) period q. In particular, T_α is not transitive. Therefore, if T_α is transitive then $\alpha \notin \mathbb{Q}$.

[(a)⇒(b)] Suppose now that $\alpha \notin \mathbb{Q}$ and that T_α is not minimal. Let F be a minimal set for T_α. Such a set exists by Theorem 1.5.2 and $F = \omega(x)$ for each $x \in F$ by Theorem 1.5.4. From this and Proposition 1.4.7, we deduce that $T_\alpha(F) = F$. Since T_α

is a bijection (in fact, a homeomorphism), we obtain that $T_\alpha^{-1}(F) = F = T_\alpha(F)$. Consequently, $T_\alpha^{-1}(\mathbb{S}^1 \setminus F) = \mathbb{S}^1 \setminus F = T_\alpha(\mathbb{S}^1 \setminus F)$. As by definition $F \neq \mathbb{S}^1$ and F is closed, the set $\mathbb{S}^1 \setminus F$ is nonempty and open. So it can be written as a countable union of open intervals

$$\mathbb{S}^1 \setminus F = \bigcup_{k=1}^{\infty} (a_k, b_k),$$

where the (a_k, b_k)'s are the connected components of $\mathbb{S}^1 \setminus F$. For each k, we have that $T_\alpha((a_k, b_k)) \subseteq \mathbb{S}^1 \setminus F$. Since (a_k, b_k) is connected and T_α is continuous, the image $T_\alpha((a_k, b_k))$ is also connected. This implies that none of the endpoints a_j, b_j lies in $T_\alpha((a_k, b_k))$. Therefore, there exists a unique $\ell = \ell(k) \in \mathbb{N}$ such that $T_\alpha((a_k, b_k)) \subseteq (a_\ell, b_\ell)$. Since $T_\alpha(a_k)$ and $T_\alpha(b_k)$ are in F, the continuity of T_α implies that $T_\alpha((a_k, b_k)) = (a_\ell, b_\ell)$. By induction on n, there exist unique a_{k_n} and b_{k_n} such that $T_\alpha^n((a_k, b_k)) = (a_{k_n}, b_{k_n})$. We claim that the family

$$\{T_\alpha^n((a_k, b_k)) : n \geq 0\} = \{(a_{k_n}, b_{k_n}) : n \geq 0\}$$

consists of mutually disjoint open arcs. If not, there would exist some $0 \leq p < q$ such that

$$T_\alpha^p((a_k, b_k)) \cap T_\alpha^q((a_k, b_k)) = (a_{k_p}, b_{k_p}) \cap (a_{k_q}, b_{k_q}) \neq \emptyset.$$

As $\{(a_j, b_j) : j \in \mathbb{N}\}$ is a pairwise disjoint family of open arcs, we deduce that

$$T_\alpha^p((a_k, b_k)) = T_\alpha^q((a_k, b_k)).$$

Consequently,

$$(a_k, b_k) = T_\alpha^{-p} \circ T_\alpha^q((a_k, b_k)) = T_\alpha^{q-p}((a_k, b_k)).$$

Writing $r := q - p \in \mathbb{N}$, this means that

$$T_\alpha^r((a_k, b_k)) = (a_k, b_k).$$

Thus, either $T_\alpha^r(a_k) = a_k$ or $T_\alpha^r(a_k) = b_k$. In either case, we would have that $T_\alpha^{2r}(a_k) = a_k$ and so a_k would be a periodic point of period $2r$ for T_α. That is, $T_\alpha^{2r}(a_k) = a_k + 4\pi r\alpha \pmod{2\pi} = a_k \pmod{2\pi}$. This means that $4\pi r\alpha = 0 \pmod{2\pi}$ or, in other words, α would be a rational number, which would contradict our original assumption.

Hence, we have shown that $\{T_\alpha^n((a_k, b_k))\}_{n=0}^\infty$ forms a family of mutually disjoint arcs. Moreover, $\mathrm{Leb}(T_\alpha^n((a_k, b_k))) = \mathrm{Leb}((a_k, b_k)) > 0$ for all $n \geq 0$, where Leb denotes the Lebesgue measure on $\mathbb{S}^1$. So the circle, which has finite Lebesgue measure, apparently contains an infinite family of disjoint arcs of equal positive Lebesgue measure. This is obviously impossible. This contradiction leads us to conclude that T_α is minimal whenever $\alpha \notin \mathbb{Q}$. □

Remark 1.5.13. An immediate consequence of the above theorem is that the orbit $\mathcal{O}_+(x)$ of every $x \in \mathbb{S}^1$ is dense in $\mathbb{S}^1$ when α is irrational. This is sometimes called Jacobi's theorem.

Finally, among the many other forms of transitivity, let us mention two. First, strongly transitive systems are those for which all points have a dense backward orbit.

Definition 1.5.14. A dynamical system $T : X \to X$ is *strongly transitive* if any of the following equivalent statements holds:
(a) $\overline{\bigcup_{n\in\mathbb{N}} T^{-n}(x)} = X$ for every $x \in X$.
(b) $\bigcup_{n\in\mathbb{N}} T^n(U) = X$ for every nonempty open subset U of X.

The proof of the equivalency is left to the reader. It is obvious that a strongly transitive system is topologically mixing, and thus transitive when the underlying space X is metrizable. But the converse does not hold in general. In Lemma 4.2.10, we will find conditions under which a transitive system is strongly transitive.

Definition 1.5.15. A dynamical system $T : X \to X$ is *very strongly transitive* if for every nonempty open subset U of X, there is $N = N(U) \in \mathbb{N}$ such that $\bigcup_{n=1}^{N} T^n(U) = X$.

It is clear that a very strongly transitive system is strongly transitive, as the terminology suggests. The converse is not true in general but, given the compactness of X, every open strongly transitive system is very strongly transitive.

1.5.3 Topological exactness

Definition 1.5.16. A dynamical system $T : X \to X$ is called *topologically exact* if for each nonempty open set $U \subseteq X$, there exists some $n \in \mathbb{N}$ such that $T^n(U) = X$.

Note that topologically exact systems are very strongly transitive. However, they may not be minimal. Full shifts, which we shall study in Chapter 3, are topologically exact systems, which are not minimal. On the other hand, there are very strongly transitive systems, which are not topologically exact (see Exercise 1.7.25).

Remark 1.5.17. Topological exactness is a topological conjugacy invariant. However, it is not a complete invariant. We shall see in Chapter 3 examples of topologically exact systems that are not topologically conjugate. Among others, two full shifts are topologically conjugate if and only if they are built upon alphabets with the same number of letters (see Theorem 3.1.14).

1.6 Examples

1.6.1 Rotations of compact topological groups

For our first example, we consider *topological groups*. For a detailed introduction to these objects, the reader is referred to [69] and, for a dynamical viewpoint, to [21].

A topological group is simply a group G together with a topology on G that satisfies the following two properties:

(a) The product map $\pi : G \times G \to G$ defined by setting

$$\pi(g, h) := gh$$

is continuous when $G \times G$ is endowed with the product topology.

(b) The inverse map $i : G \to G$ defined by setting

$$i(g) := g^{-1}$$

is continuous.

Given $a \in G$, we define the map $L_a : G \to G$ by

$$L_a(g) := ag.$$

So L_a acts on the group G by left multiplication by a. The map L_a is often referred to as the left rotation of G by a. The map L_a is continuous since $L_a(g) = \pi(a, g)$. Moreover, observe that $L_a^n = L_{a^n}$ for every $n \in \mathbb{Z}$. In particular, $L_a^{-1} = L_{a^{-1}}$. The rotation L_a is thus a homeomorphism of G. In a similar way, we define the right rotation of G by a to be the continuous map $R_a : G \to G$, where

$$R_a(g) := ga.$$

For rotations of topological groups, transitivity and minimality are one and the same property, as the following theorem shows.

Theorem 1.6.1. *Let $L_a : G \to G$ be the left rotation of a topological group G by $a \in G$. Then L_a is minimal if and only if L_a is transitive. Similarly, the right rotation R_a is minimal if and only if it is transitive.*

Proof. If L_a is minimal, then L_a is transitive by Remark 1.5.7(a). For the converse, let x be a transitive point and let $y \in G$ be arbitrary. According to Theorem 1.5.4, it suffices to show that $\omega(y) = G$. Let $z \in \omega(x)$. Then there exists a strictly increasing sequence $(n_k)_{k=1}^{\infty}$ of nonnegative integers such that $\lim_{k\to\infty} L_a^{n_k}(x) = z$. Observe that

$$\begin{aligned}\lim_{k\to\infty} L_a^{n_k}(y) &= \lim_{k\to\infty} L_a^{n_k}(xx^{-1}y) = \lim_{k\to\infty} a^{n_k}xx^{-1}y = \lim_{k\to\infty} (a^{n_k}x)(x^{-1}y)\\ &= \lim_{k\to\infty} (L_a^{n_k}(x))(x^{-1}y) = \lim_{k\to\infty} R_{x^{-1}y}(L_a^{n_k}(x))\\ &= R_{x^{-1}y}\Big(\lim_{k\to\infty} L_a^{n_k}(x)\Big) = R_{x^{-1}y}(z).\end{aligned}$$

So $R_{x^{-1}y}(z) \in \omega(y)$. Since $z \in \omega(x)$ was chosen arbitrarily, we deduce that $R_{x^{-1}y}(\omega(x)) \subseteq \omega(y)$. As $\omega(x) = G$, we conclude that

$$G = R_{x^{-1}y}(G) = R_{x^{-1}y}(\omega(x)) \subseteq \omega(y) \subseteq G.$$

Hence, $\omega(y) = G$ for any arbitrary $y \in G$ and so G, and thus L_a, is minimal. The proof that R_α is minimal proceeds analogously and is thus left to the reader. □

We now give a characterization of all minimal rotations of a topological group.

Proposition 1.6.2. *The rotation $L_a : G \to G$ is minimal if and only if*

$$G = \overline{\mathcal{O}_+(e)} = \overline{\{a^n : n \geq 0\}},$$

where e denotes the identity element of G.

Proof. First, observe that $\overline{\mathcal{O}_+(e)} = \overline{\{a^n : n \geq 0\}}$ since $L_a^n(e) = a^n e = a^n$ for each $n \geq 0$. Now suppose that L_a is minimal. By Theorem 1.5.4, we have that $\overline{\mathcal{O}_+(g)} = G$ for every $g \in G$. In particular, $\overline{\mathcal{O}_+(e)} = G$.

For the converse, suppose that $G = \overline{\mathcal{O}_+(e)}$. According to Theorem 1.6.1, it is sufficient to prove that L_a is transitive. By Theorem 1.5.9, since L_a is surjective it is sufficient to show that L_a is weakly transitive. This is certainly the case since $\overline{\mathcal{O}_+(e)} = G$. □

This characterization shows that minimal rotations can only occur in abelian groups. Indeed, if L_a is minimal and $x, y \in G$, then there exist strictly increasing sequences $(n_j)_{j=1}^\infty$ and $(m_k)_{k=1}^\infty$ of nonnegative integers such that $x = \lim_{j\to\infty} a^{n_j}$ and $y = \lim_{k\to\infty} a^{m_k}$. Using the left and right continuity of the product map, we obtain

$$\begin{aligned}
xy &= \Big(\lim_{j\to\infty} a^{n_j}\Big)y = \lim_{j\to\infty}(a^{n_j}y) = \lim_{j\to\infty}\Big(a^{n_j} \lim_{k\to\infty} a^{m_k}\Big) \\
&= \lim_{j\to\infty}\Big(\lim_{k\to\infty}(a^{n_j}a^{m_k})\Big) = \lim_{j\to\infty}\Big(\lim_{k\to\infty}(a^{m_k}a^{n_j})\Big) \\
&= \lim_{j\to\infty}\Big(\Big(\lim_{k\to\infty} a^{m_k}\Big)a^{n_j}\Big) = \lim_{j\to\infty}(ya^{n_j}) \\
&= y \lim_{j\to\infty} a^{n_j} = yx.
\end{aligned}$$

Rotations of the n-dimensional torus $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$

We shall now study rotations (also sometimes called translations) of the n-dimensional torus

$$\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1,$$

that is, the n-times direct product of $\mathbb{S}^1 := [0, 1]$ (mod 1). Let $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n) \in \mathbb{T}^n$ and let $L_\gamma : \mathbb{T}^n \to \mathbb{T}^n$ be the rotation of $\mathbb{T}^n$ by γ, which is defined to be

$$L_\gamma(x_1, x_2, \ldots, x_n) := (x_1 + \gamma_1, x_2 + \gamma_2, \ldots, x_n + \gamma_n) \text{ (mod 1)}.$$

The proof of the following theorem uses Fourier coefficients and the Hilbert space $L^2(\lambda_n)$ of complex-valued functions whose squared modulus is integrable with respect to the Lebesgue measure λ_n on the n-dimensional torus. A good reference for those unfamiliar with Fourier coefficients or the Hilbert space $L^2(\lambda_n)$ is Rudin [58]. Those unfamiliar with measure theory may wish to consult Appendix A first.

Recall that the numbers $1, \gamma_1, \gamma_2, \ldots, \gamma_n$ are said to be *linearly independent over the field of rational numbers* $\mathbb{Q}$ if the equation

$$\alpha_0 + \alpha_1\gamma_1 + \alpha_2\gamma_2 + \cdots + \alpha_n\gamma_n = 0,$$

where the α_j are rational numbers, has for unique solution $\alpha_0 = \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$. Equivalently, the numbers $1, \gamma_1, \gamma_2, \ldots, \gamma_n$ are linearly independent over $\mathbb{Q}$ if

$$k_1\gamma_1 + k_2\gamma_2 + \cdots + k_n\gamma_n \in \mathbb{Z},$$

where each $k_j \in \mathbb{Z}$, only when $k_1 = k_2 = \cdots = k_n = 0$.

We now prove the following classical result, which is a significant generalization of Theorem 1.5.12 with a consequently more intricate proof.

Theorem 1.6.3. *Let $L_\gamma : \mathbb{T}^n \to \mathbb{T}^n$ be a translation of the torus. The following statements are equivalent:*
(a) *The numbers $1, \gamma_1, \gamma_2, \ldots, \gamma_n$ are linearly independent over $\mathbb{Q}$.*
(b) *L_γ is minimal.*
(c) *L_γ is transitive.*

Proof. According to Theorem 1.6.1, a rotation of a topological group is minimal if and only if it is transitive. Therefore, (b)⇔(c). We shall now prove that (a)⇔(c).

Suppose first that L_γ is transitive. Assume by way of contradiction that $\sum_{j=1}^n k_j\gamma_j \in \mathbb{Z}$, where each $k_j \in \mathbb{Z}$ and at least one of these numbers, say k_i, differs from zero. Let $\varphi : \mathbb{T}^n \to \mathbb{R}$ be the function defined by

$$\varphi(x) = \varphi(x_1, x_2, \ldots, x_n) = \sin\left(2\pi \sum_{j=1}^n k_j x_j\right).$$

Since x is in reality the equivalence class $[x] = \{x + \ell : \ell \in \mathbb{Z}^n\}$, we must check that φ is well-defined, that is, is constant on the entire equivalence class. This straightforward calculation is left to the reader as an exercise (see Exercise 1.7.30). As φ is a composition of continuous maps, it is continuous. Moreover, since $\varphi(0, 0, \ldots, 0) = 0$ but $\varphi(0, \ldots, 0, 1/(4k_i), 0, \ldots, 0) = \sin(2\pi k_i \frac{1}{4k_i}) = \sin(\frac{\pi}{2}) = 1$, the function φ is not constant. But given that

$$\varphi(L_\gamma(x)) = \varphi(x_1 + \gamma_1, x_2 + \gamma_2, \ldots, x_n + \gamma_n) = \sin\left(2\pi \sum_{j=1}^n k_j(x_j + \gamma_j)\right),$$

and as we have assumed that $\sum_{j=1}^n k_j\gamma_j \in \mathbb{Z}$, we deduce that

$$\varphi(L_\gamma(x)) = \sin\left(2\pi \sum_{j=1}^n k_j x_j + 2\pi \sum_{j=1}^n k_j\gamma_j\right) = \sin\left(2\pi \sum_{j=1}^n k_j x_j\right) = \varphi(x).$$

Hence, φ is L_γ-invariant. To summarize, φ is a nonconstant function, which is invariant under a transitive system. According to Remark 1.5.7(d+e) this is impossible, and thus all k_j must equal 0. Hence, $1, \gamma_1, \gamma_2, \ldots, \gamma_n$ are linearly independent over $\mathbb{Q}$.

For the converse implication, suppose that $1, \gamma_1, \gamma_2, \ldots, \gamma_n$ are linearly independent over $\mathbb{Q}$ and, again for a contradiction, suppose that L_γ is not transitive. By Theorem 1.5.11, we have that L_γ is not topologically mixing. So there exist nonempty open sets U and V contained in $\mathbb{T}^n$ such that

$$\bigcup_{n=1}^{\infty} L_\gamma^n(U) \cap V = \emptyset.$$

This means that $W := \bigcup_{n=1}^{\infty} L_\gamma^n(U) \subseteq X \setminus V$. Then $\overline{W} \subseteq X \setminus V$ since $X \setminus V$ is a closed set. Moreover,

$$L_\gamma(W) = L_\gamma\left(\bigcup_{n=1}^{\infty} L_\gamma^n(U)\right) = \bigcup_{n=1}^{\infty} L_\gamma^{n+1}(U) \subseteq W,$$

and we thus obtain that $\overline{L_\gamma(W)} \subseteq \overline{W}$. Also, the continuity of L_γ ensures that $L_\gamma(\overline{W}) \subseteq \overline{L_\gamma(W)}$. Therefore $L_\gamma(\overline{W}) \subseteq \overline{W}$. We aim to show that this is in fact an equality. To that end, let λ_n denote the Lebesgue measure on $\mathbb{T}^n$, and note that λ_n is translation invariant, which means that

$$\lambda_n(E + v) = \lambda_n(E), \quad \forall E \subseteq \mathbb{T}^n,\ \forall v \in \mathbb{T}^n.$$

So

$$\lambda_n(L_\gamma(\overline{W})) = \lambda_n(\overline{W}). \tag{1.1}$$

If it turned out that $L_\gamma(\overline{W}) \subsetneq \overline{W}$, then there would exist $x \in \overline{W} \setminus L_\gamma(\overline{W})$ and $\varepsilon > 0$ such that $\emptyset \neq B(x, \varepsilon) \cap W \subseteq \overline{W} \setminus L_\gamma(\overline{W})$. But $B(x, \varepsilon) \cap W$ is a nonempty open set, and hence has positive Lebesgue measure. Thus, we would have $\lambda_n(\overline{W} \setminus L_\gamma(\overline{W})) > 0$, which would contradict (1.1). Hence, $L_\gamma(\overline{W}) = \overline{W}$ and, since L_γ is invertible, $L_\gamma^{-1}(\overline{W}) = \overline{W}$.

Now, denote by $\mathbb{1}_A$ the characteristic function of a subset A of $\mathbb{T}^n$, that is,

$$\mathbb{1}_A := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

For any map T and any set A, we have that

$$\mathbb{1}_A \circ T = \mathbb{1}_{T^{-1}(A)}.$$

Since the set $\overline{W}$ is completely L_γ-invariant, we deduce that

$$\mathbb{1}_{\overline{W}} \circ L_\gamma = \mathbb{1}_{L_\gamma^{-1}(\overline{W})} = \mathbb{1}_{\overline{W}}.$$

That is, the function $\mathbb{1}_{\overline{W}}$ is L_γ-invariant.

For every $k \in \mathbb{R}^n$, let $\psi_k : \mathbb{T}^n \to \mathbb{C}$ be defined by

$$\psi_k(x) = e^{2\pi i \langle k,x\rangle} = \cos(2\pi\langle k,x\rangle) + i\sin(2\pi\langle k,x\rangle),$$

where $\langle k,x\rangle = \sum_{j=1}^n k_j x_j$ is the scalar product of the vectors k and x. Then the family $\{\psi_k\}_{k\in\mathbb{Z}^n}$ is an orthonormal basis for the Hilbert space $L^2(\lambda_n)$. Since $\mathbb{1}_{\overline{W}} \in L^2(\lambda_n)$, we can write

$$\mathbb{1}_{\overline{W}}(x) = \sum_{k\in\mathbb{Z}^n} a_k\psi_k(x) = \sum_{k\in\mathbb{Z}^n} a_k e^{2\pi i\langle k,x\rangle} \quad \text{for } \lambda_n\text{-a. e. } x \in \mathbb{T}^n, \tag{1.2}$$

where

$$a_k := \int \mathbb{1}_{\overline{W}}(y)\overline{\psi_k(y)}\, d\lambda_n(y)$$

are the Fourier coefficients of $\mathbb{1}_{\overline{W}}$. Then, for λ_n-a. e. $x \in \mathbb{T}^n$, we have

$$\begin{aligned}\mathbb{1}_{\overline{W}}(x) = \mathbb{1}_{\overline{W}}(L_\gamma(x)) &= \sum_{k\in\mathbb{Z}^n} a_k e^{2\pi i\langle k,x+\gamma\rangle}\\ &= \sum_{k\in\mathbb{Z}^n} a_k e^{2\pi i\langle k,\gamma\rangle}\psi_k(x).\end{aligned} \tag{1.3}$$

Since $\{\psi_k\}_{k\in\mathbb{Z}^n}$ is an orthonormal basis, we deduce from (1.2) and (1.3) that

$$a_k = a_k e^{2\pi i\langle k,\gamma\rangle}$$

for every $k \in \mathbb{Z}^n$. For each such k, this implies that

$$\text{either } a_k = 0 \text{ or } e^{2\pi i\langle k,\gamma\rangle} = 1.$$

In the latter case,

$$\langle k,\gamma\rangle = \sum_{j=1}^n k_j\gamma_j \in \mathbb{Z}.$$

Since the γ_i were assumed to be linearly independent over $\mathbb{Q}$, this implies that $k = (0,0,\ldots,0)$. So, for all $k \neq (0,0,\ldots,0)$, we must be in the former case, that is, we must have that $a_k = 0$. Hence,

$$\mathbb{1}_{\overline{W}}(x) = a_{(0,\ldots,0)} = \lambda_n(\overline{W}) \quad \text{for } \lambda_n\text{-a. e. } x \in \mathbb{T}^n.$$

As W is a nonempty open set, it follows that $\lambda_n(W) > 0$, and thus $\lambda_n(\overline{W}) > 0$. So $\mathbb{1}_{\overline{W}}(x) > 0$ for λ_n-a. e. $x \in \mathbb{T}^n$. However, recall that $\overline{W} \cap V = \emptyset$, and thus $\mathbb{1}_{\overline{W}}(V) = 0$. As V is a nonempty open set, $\lambda_n(V) > 0$ and, therefore, $\mathbb{1}_{\overline{W}}(x) = 0$ on a set of positive measure. We have reached a contradiction. This means that L_γ is transitive. □

1.6.2 Maps of the interval

Although there is no known topological conjugacy invariant which is a complete invariant for the family of all dynamical systems, some conjugacy invariants turn out to be complete invariants for subfamilies of systems. By a complete invariant for a subfamily, we mean that if two systems from the subfamily share this invariant, then they are automatically topologically conjugate.

For instance, we have seen that the number of periodic points of any given period is a topological conjugacy invariant, though not a complete invariant. However, the number of periodic points of a given period sometimes turns out to be a complete invariant if we restrict our attention to an appropriate subfamily of dynamical systems.

In this section, we show that the number of fixed points is a complete invariant for the family of all self-homeomorphisms of compact intervals which fix the endpoints of their domain and whose sets of fixed points are finite. These self-homeomorphisms can be characterized as the strictly increasing continuous self-maps of compact intervals that fix the endpoints of their domain and have only finitely many fixed points.

The proof of the complete invariance of the number of fixed points will be given in several stages.

Theorem 1.6.4. *Any two strictly increasing continuous maps of the unit interval that fix the endpoints of the interval and that have no other fixed points are topologically conjugate.*

Proof. Let $I := [0,1]$, and let $T_1, T_2 : I \to I$ be two strictly increasing continuous maps such that $\mathrm{Fix}(T_1) = \mathrm{Fix}(T_2) = \{0,1\}$. By the intermediate value theorem, each T_i is a surjection. As T_i is a strictly increasing map, it is also injective. Thus, T_i is a bijection and T_i^{-1} exists. Furthermore, T_i^{-1} is a strictly increasing continuous bijection. It follows immediately that T_i^n is a strictly increasing homeomorphism of I for all $n \in \mathbb{Z}$.

Let $x \in (0,1)$. Since $T_1(x) \neq x$, either $T_1(x) < x$ or $T_1(x) > x$. Similarly, either $T_2(x) < x$ or $T_2(x) > x$. Let us consider the case in which $T_1(x) > x$ and $T_2(x) > x$. The proofs of the other three cases are similar and are left to the reader. Then $T_1(y) > y$ for every $y \in (0,1)$; otherwise, the continuous map T_1 would have a fixed point between 0 and 1 by the intermediate value theorem (because of the change of sign in the values of the continuous map $T_1 - \mathrm{Id}_I$). Similarly, $T_2(y) > y$ for each $y \in (0,1)$. As T_i^n is strictly increasing for all $n \in \mathbb{Z}$, it follows easily that for all $m < n$ and all $y \in (0,1)$, we have

$$T_i^m(y) < T_i^n(y)$$

for each $i = 1, 2$.

Now, fix $0 < a < 1$. Note once again that $T_i(a) > a$ and let $\triangle_i = [a, T_i(a)]$. We now state and prove three claims that will allow us to complete the proof of the theorem.

Claim 1. *For all $m, n \in \mathbb{Z}$ such that $m < n$, we have*

$$T_i^m\big(\mathrm{Int}(\triangle_i)\big) \cap T_i^n\big(\mathrm{Int}(\triangle_i)\big) = \emptyset.$$

Proof. Let $j \in \mathbb{Z}$. Since T_i^j is strictly increasing and continuous, we have that

$$T_i^j(\mathrm{Int}(\triangle_i)) = (T_i^j(a), T_i^{j+1}(a)).$$

As $m+1 \le n$, it follows that $T_i^{m+1}(a) \le T_i^n(a)$ and, therefore,

$$T_i^m(\mathrm{Int}(\triangle_i)) \cap T_i^n(\mathrm{Int}(\triangle_i)) = (T_i^m(a), T_i^{m+1}(a)) \cap (T_i^n(a), T_i^{n+1}(a)) = \emptyset. \qquad \square$$

Claim 2. *For all* $x \in (0,1)$, *we have that*

$$\lim_{n\to\infty} T_i^{-n}(x) = 0 \quad \textit{and} \quad \lim_{n\to\infty} T_i^n(x) = 1.$$

Proof. Let $x \in (0,1)$. We shall establish the second limit; the first limit can be ascertained analogously. First, note that the limit must exist because the sequence $(T_i^n(x))_{n=0}^{\infty}$ is (strictly) increasing and bounded above by 1. So, let $y = \lim_{n\to\infty} T_i^n(x)$. According to Lemma 1.1.4, y is a fixed point of T_i. Moreover, y is clearly not equal to 0. Hence, $y = 1$. $\square$

Claim 3.

$$\bigcup_{n=-\infty}^{\infty} T_i^n(\triangle_i) = (0,1).$$

Proof. It is clear that $\bigcup_{n=-\infty}^{\infty} T_i^n(\triangle_i) \subseteq (0,1)$. To prove the opposite inclusion, let $x \in (0,1)$. If $x \in \triangle_i$, we are done. So, let $x \notin \triangle_i$. Then either $x < a$ or $x > T_i(a)$. In the former case, there exists by Claim 2 a largest $n \ge 0$ such that $T_i^n(x) < a$. Therefore, $T_i^{n+1}(x) \ge a$. Moreover, since T_i is strictly increasing, $T_i^n(x) < a$ implies $T_i^{n+1}(x) < T_i(a)$. Hence, $T_i^{n+1}(x) \in \triangle_i$, that is, $x \in T_i^{-(n+1)}(\triangle_i)$. In the latter case, by Claim 2 there exists a largest $n \ge 0$ such that $T_i^{-n}(x) > T_i(a)$. Then $T_i^{-(n+1)}(x) \le T_i(a)$. Moreover, since T_i^{-1} is strictly increasing, $T_i^{-(n+1)}(x) > T_i^{-1}(T_i(a)) = a$. Hence, $T_i^{-(n+1)}(x) \in \triangle_i$, that is, $x \in T_i^{(n+1)}(\triangle_i)$. In all cases, $x \in \bigcup_{n=-\infty}^{\infty} T_i^n(\triangle_i)$.

In order to make the idea behind the sequence of intervals $(T_i^n(\triangle_i))_{n\in\mathbb{Z}}$ clearer, see Figure 1.5.

We will now define a conjugacy map h between T_1 and T_2. First of all, suppose that $H : \triangle_1 \to \triangle_2$ is any homeomorphism satisfying

$$H(a) = a \quad \text{and} \quad H(T_1(a)) = T_2(a).$$

Let $x \in (0,1)$. By Claim 3, there exists $n(x) \in \mathbb{Z}$ such that $x \in T_1^{-n(x)}(\triangle_1)$. We shall shortly observe that $n(x)$ is uniquely defined for all $x \notin \{T_1^n(a) : n \in \mathbb{Z}\}$. When $x \in \{T_1^n(a) : n \in \mathbb{Z}\}$, then $n(x)$ takes the value of any one of two consecutive integers; so it is not uniquely defined and this will require prudence when defining h. Define the conjugacy map h by setting

$$h(x) = \begin{cases} T_2^{-n(x)} \circ H \circ T_1^{n(x)}(x) & \text{if } x \in (0,1) \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x = 1. \end{cases}$$

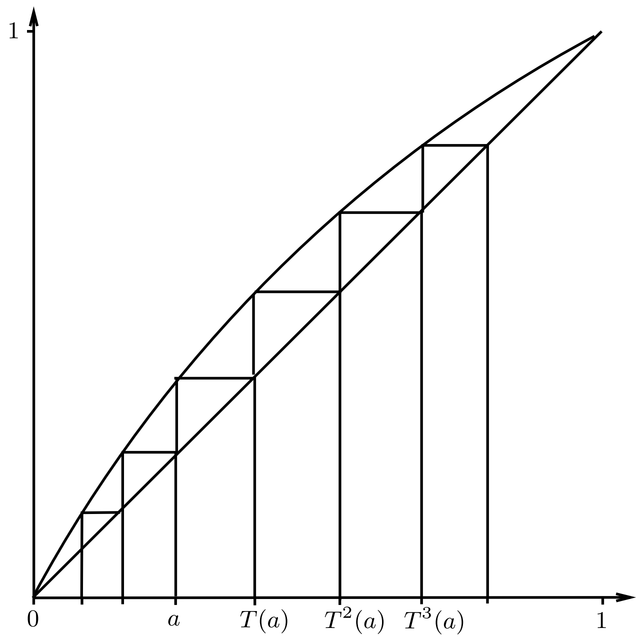

Figure 1.5: The map T is an orientation-preserving homeomorphism of the unit interval that fixes only the endpoints. The vertical dotted lines indicate the intervals $(T^n(\triangle))_{n\in\mathbb{Z}}$, where $\triangle := [a, T(a)]$.

We first check that this map is well-defined. Suppose that $T_1^k(x), T_1^\ell(x) \in \triangle_1$ for some $k < \ell$. Claim 1 above implies that $T_1^k(x), T_1^\ell(x) \in \partial\triangle_1$ and $k + 1 = \ell$. It follows that $T_1^k(x) = a$ and $T_1^\ell(x) = T_1^{k+1}(x) = T_1(a)$. Therefore,

$$\begin{aligned}
T_2^{-\ell} \circ H \circ T_1^\ell(x) &= T_2^{-(k+1)}\big(H\big(T_1(a)\big)\big) \\
&= T_2^{-(k+1)}\big(T_2(a)\big) \\
&= T_2^{-k}(a) \\
&= T_2^{-k}\big(H(a)\big) \\
&= T_2^{-k}\big(H\big(T_1^k(x)\big)\big) \\
&= T_2^{-k} \circ H \circ T_1^k(x).
\end{aligned}$$

Thus, the map h is well-defined.

We must also show that $T_2 \circ h = h \circ T_1$. Toward this end, first observe that we have $T_1^{n(x)-1}(T_1(x)) = T_1^{n(x)}(x) \in \triangle_1$, so we can choose $n(T_1(x))$ to be $n(x) - 1$, and then we obtain that

$$\begin{aligned}
h \circ T_1(x) &= T_2^{-n(T_1(x))} \circ H \circ T_1^{n(T_1(x))}\big(T_1(x)\big) \\
&= T_2^{-(n(x)-1)} \circ H \circ T_1^{n(x)-1}\big(T_1(x)\big)
\end{aligned}$$

$$= T_2 \circ T_2^{-n(x)} \circ H \circ T_1^{n(x)}(T_1^{-1}(T_1(x)))$$
$$= T_2 \circ h(x).$$

To complete the proof, it remains to show that h is a bijection and that it is continuous (recall that a continuous bijection between compact metrizable spaces is a homeomorphism). That h is continuous follows from the fact that

$$h|_{T_1^{-n}(\triangle_1)} = T_2^{-n} \circ H \circ T_1^{n}|_{T_1^{-n}(\triangle_1)}$$

for every $n \in \mathbb{Z}$. Indeed, as T_1^n, H, and T_2^{-n} are continuous, the restriction of h to each $T_1^{-n}(\triangle_1)$ is continuous. Using left and right continuity at the endpoints of the intervals $T_1^{-n}(\triangle_1)$, we conclude from Claim 3 that h is continuous on $(0,1)$. The continuity of h at 0 follows from the fact that $\lim_{x\to 0} n(x) = \infty$, that $h(x) \in T_2^{-n(x)}(\triangle_2) = [T_2^{-n(x)}(a), T_2^{-(n(x)-1)}(a)]$, and, by Claim 2, that $\lim_{n\to\infty} T_2^{-n}(a) = 0$. A similar argument establishes the continuity of h at 1.

For the injectivity of h, we shall show that h is strictly increasing. If $0 < x < y < 1$, then $n(x)$ and $n(y)$ can be chosen so that $n(x) \geq n(y)$. If $n(x) = n(y) =: n$, then $h(x) < h(y)$ since the restriction of h to $T_1^{-n}(\triangle_1)$ is strictly increasing (because T_1^n, H and T_2^{-n} are all strictly increasing). If $n(x) > n(y)$, then $T_2^{-n(x)}(\triangle_2)$ lies to the left of $T_2^{-n(y)}(\triangle_2)$ and, as $h(x) \in T_2^{-n(x)}(\triangle_2)$ while $h(y) \in T_2^{-n(y)}(\triangle_2)$, we deduce that $h(x) < h(y)$.

Finally, since h is continuous, $h(0) = 0$ and $h(1) = 1$, the map h is surjective by the intermediate value theorem. □

Corollary 1.6.5. *Any two strictly increasing continuous maps of compact intervals which have the same finite number of fixed points, including both endpoints of their respective domains, are topologically conjugate.*

Proof. We first prove that if $f : I \to I$ is a strictly increasing continuous map which only fixes the points 0 and 1 and if $g : [a,b] \to [a,b]$ is a strictly increasing continuous map which only fixes a and b, then f and g are topologically conjugate. Let $k : [a,b] \to I$ be defined by $k(x) = \frac{x-a}{b-a}$. Then k is a strictly increasing homeomorphism. Consequently, $k \circ g \circ k^{-1}$ is a strictly increasing continuous map of I which only fixes 0 and 1, just like f. By Theorem 1.6.4, there is a conjugacy map $h : I \to I$ between f and $k \circ g \circ k^{-1}$ (i. e., h is a homeomorphism such that $h \circ f = k \circ g \circ k^{-1} \circ h$). It follows that $k^{-1} \circ h : I \to [a,b]$ is a conjugacy map between f and g.

Recall that topological conjugacy is an equivalence relation. It follows from the argument above and the transitivity of topological conjugacy that any two strictly increasing continuous maps of compact intervals which only fix the endpoints of their respective domains, are topologically conjugate.

Now, let $T : I \to I$ and $S : [a,b] \to [a,b]$ be strictly increasing continuous maps with the same finite number of fixed points, which include the endpoints of their respective domains. Denote the sets of fixed points of T and S by $\{0, x_{T_1}, \ldots, x_{T_{k-1}}, 1\}$ and

$\{a, x_{S_1}, \dots, x_{S_{k-1}}, b\}$, respectively. These fixed points induce two finite sequences of maps

$$\begin{array}{rcccrccc} T_0: & [0, x_{T_1}] & \to & [0, x_{T_1}] & S_0: & [a, x_{S_1}] & \to & [a, x_{S_1}] \\ T_1: & [x_{T_1}, x_{T_2}] & \to & [x_{T_1}, x_{T_2}] & S_1: & [x_{S_1}, x_{S_2}] & \to & [x_{S_1}, x_{S_2}] \\ & & \vdots & & & & \vdots & \\ T_{k-1}: & [x_{T_{k-1}}, 1] & \to & [x_{T_{k-1}}, 1] & S_{k-1}: & [x_{S_{k-1}}, b] & \to & [x_{S_{k-1}}, b] \end{array}$$

maps which are the restrictions of T and S, and hence are strictly increasing continuous maps of compact intervals which only fix the endpoints of their domains. For each $0 \le i < k$, let h_i be a conjugacy map between T_i and S_i (such a map exists according to the discussion in the previous paragraph). Then set $h : I \to [a, b]$ to be the map defined by $h(x) = h_i(x)$ when $x \in [x_{T_i}, x_{T_{i+1}}]$. This map h is clearly a bijection and is continuous (the continuity of h at the points $\{x_{T_1}, \dots, x_{T_{k-1}}\}$ can be established by means of left and right continuity). Finally, $h \circ T = S \circ h$ since $h_i \circ T_i = S_i \circ h_i$ for all $0 \le i < k$.

Finally, it follows from this and the transitivity of topological conjugacy that any two strictly increasing continuous maps of compact intervals with the same finite number of fixed points, among whose are the endpoints of their respective domains, are topologically conjugate. □

However, note that any two strictly increasing continuous maps of compact intervals which have the same finite number of fixed points, one of which fixes the endpoints of its domain while the other does not, are not topologically conjugate (see Exercise 1.7.34). In particular, this establishes that the number of fixed points is not a complete invariant for the subfamily of all strictly increasing continuous maps of compact intervals which have the same given finite number of fixed points.

1.7 Exercises

Exercise 1.7.1. In this exercise, we revisit Example 1.1.3(b). Recall that, given $m \in \mathbb{N}$, we defined the map $T_m : \mathbb{S}^1 \to \mathbb{S}^1$ by setting $T_m(x) := mx \pmod 1$. The map T_m is simply a piecewise linear map that sends each interval $[i/m, (i+1)/m]$, for $0 \le i < m$, onto $\mathbb{S}^1$. It can be expressed by the formula

$$T_m(x) = mx - i \quad \text{for all } \frac{i}{m} \le x \le \frac{i+1}{m}.$$

Prove that for every $n \in \mathbb{N}$ the iterates of T can be expressed as

$$T_m^n(x) = m^n x - \sum_{k=1}^{n} m^{n-k} i_{n-k+1}$$

if

$$\frac{1}{m^n} \sum_{k=1}^{n} m^{n-k} i_{n-k+1} \le x \le \frac{1}{m^n} \left(\sum_{k=1}^{n} m^{n-k} i_{n-k+1} + 1 \right),$$

where $0 \le i_1, i_2, \dots, i_n < m$. Deduce that T_m has m^n periodic points of period n.

Exercise 1.7.2. Show that a point $x \in X$ is preperiodic for a system $T : X \to X$ if and only if its forward orbit $\mathcal{O}_+(x)$ is finite.

Exercise 1.7.3. Let $\mathrm{Per}(T)$ be the set of periodic points for a system $T : X \to X$. Prove that

$$\mathrm{PrePer}(T) = \bigcup_{x \in \mathrm{Per}(T)} \mathcal{O}_-(x),$$

where $\mathrm{PrePer}(T)$ denotes the set of all preperiodic points for the system T.

Exercise 1.7.4. Identify all the preperiodic points for the dynamical systems introduced in Example 1.1.3.

Exercise 1.7.5. Prove that if both X and Y are dense subsets of $\mathbb{R}$ and $g : X \to Y$ is an increasing bijection, then g extends uniquely to an increasing homeomorphism $\widetilde{g} : \mathbb{R} \to \mathbb{R}$.

Exercise 1.7.6. Prove that topological conjugacy defines an equivalence relation on the space of dynamical systems.

Exercise 1.7.7. Show that if two dynamical systems $T : X \to X$ and $S : Y \to Y$ are topologically conjugate via a conjugacy map $h : X \to Y$, then their corresponding iterates are topologically conjugate by means of the same conjugacy map h.

Exercise 1.7.8. Prove that for every $n \in \mathbb{N}$ there exists a one-to-one correspondence between the periodic points of period n of two topologically conjugate dynamical systems. Show that this implies for every $n \in \mathbb{N}$ the existence of a one-to-one correspondence between the periodic points of prime period n.

Exercise 1.7.9. Prove that if two dynamical systems $T : X \to X$ and $S : Y \to Y$ are topologically conjugate via a conjugacy map $h : X \to Y$, then h induces a one-to-one correspondence between preperiodic points. Deduce that the number of preperiodic points is a topological conjugacy invariant. By means of an example, show also that the number of preperiodic points is not a complete invariant.

Exercise 1.7.10. Suppose that a dynamical system $S : Y \to Y$ is a factor of a system $T : X \to X$. Show that every orbit of T is projected onto an orbit of S. Show also that for all $n \in \mathbb{N}$ every periodic point of period n for T is mapped to a periodic point for S whose period is a factor of n.

Exercise 1.7.11. Prove that the closure of every invariant set is invariant.

Exercise 1.7.12. Show that if $x \in X$ is a periodic point for a system $T : X \to X$, then $\omega(x) = \mathcal{O}_+(x) = \overline{\mathcal{O}_+(x)}$. Observe also that the set of limit points of $\mathcal{O}_+(x)$ is empty. Deduce that $\omega(x)$ does not coincide with the set of limit points of $\mathcal{O}_+(x)$.

Exercise 1.7.13. Prove that if $x \in X$ is a preperiodic point for a system $T : X \to X$, then $\omega(x) \neq \mathcal{O}_+(x) = \overline{\mathcal{O}_+(x)}$. So $\omega(x) \neq \overline{\mathcal{O}_+(x)}$. Moreover, as in Exercise 1.7.12, prove that $\omega(x)$ does not coincide with the set of limit points of the forward orbit $\mathcal{O}_+(x)$.

Exercise 1.7.14. Let $T : X \to X$ be a dynamical system. Prove that $\omega(x)$ is the set of limit points of $\mathcal{O}_+(x)$ if and only if x is not a periodic or preperiodic point.

Exercise 1.7.15. Let $T : X \to X$ be a dynamical system. Show that $\mathcal{O}_+(x) \cup \omega(x) = \overline{\mathcal{O}_+(x)}$ for any $x \in X$.

Exercise 1.7.16. Show that the set of limit points of any set is closed.

Hint: Prove that any accumulation point of accumulation points of a set S is an accumulation point of S.

Exercise 1.7.17. Prove that any intersection of a descending sequence of nonempty compact sets in a Hausdorff topological space is a nonempty compact set.

Exercise 1.7.18. Show that every minimal system is surjective.

Exercise 1.7.19. Prove that minimality is not a complete invariant.

Hint: Construct two finite minimal dynamical systems with different cardinalities.

Exercise 1.7.20. Prove that minimality is not a complete invariant for infinite systems.

Exercise 1.7.21. Let $+_2 : \{0,1\} \to \{0,1\}$ denote addition modulo 2 and endow the set $\{0,1\}$ with the discrete topology. Prove that the dynamical system $T : \mathbb{S}^1 \times \{0,1\} \to \mathbb{S}^1 \times \{0,1\}$ given by the formula

$$T(x,y) := \big(x + \sqrt{2} \,(\mathrm{mod}\ 1), y +_2 1\big)$$

is minimal.

Exercise 1.7.22. Prove or disprove (by providing a counterexample) that if $T : X \to X$ is minimal then $T^2 : X \to X$ is also minimal.

Exercise 1.7.23. Prove that the statements in Definition 1.5.10 are equivalent.

Exercise 1.7.24. Prove that topological transitivity, strong transitivity, very strong transitivity, and topological exactness are topological conjugacy invariants.

Exercise 1.7.25. Construct a very strongly transitive, open system which is not topologically exact. (Recall that a map is said to be open if the image of any open set under that map is open.)

Exercise 1.7.26. Build a strongly transitive system which is not very strongly transitive.

Exercise 1.7.27. Find a transitive system which is not strongly transitive.

Exercise 1.7.28. Let $T : X \to X$ be a dynamical system. Suppose that for every nonempty open subset U of X there exists $n \geq 0$ such that $T^n(U)$ is a dense subset of X. Prove that T is topologically exact.

Exercise 1.7.29. A continuous map $T : X \to X$ is said to be *locally eventually onto* provided that for every nonempty open subset U of X there exists $n \geq 0$ such that

$$\bigcup_{j=0}^{n} T^j(U) = X.$$

Each topologically exact map is locally eventually onto. Provide an example of a locally eventually onto map that is not topologically exact.

Exercise 1.7.30. Let $(k_1, k_2, \ldots, k_n) \in \mathbb{Z}^n$. Let $\varphi : \mathbb{T}^n \to \mathbb{R}$ be the function defined by

$$\varphi(x) = \varphi(x_1, x_2, \ldots, x_n) = \sin\left(2\pi \sum_{j=1}^{n} k_j x_j\right).$$

Show that φ is well-defined.

Exercise 1.7.31. Let $T : X \to X$ be a dynamical system. Prove that a subset Y of X is T-invariant if and only if $\mathbb{1}_Y$ is T-invariant, where $\mathbb{1}_Y$ denotes the characteristic function of the set Y, that is,

$$\mathbb{1}_Y(x) := \begin{cases} 1 & \text{if } x \in Y \\ 0 & \text{if } x \notin Y. \end{cases}$$

Exercise 1.7.32. Establish graphically that Claims 1, 2, and 3 in the proof of Theorem 1.6.4 hold.

Exercise 1.7.33. Prove Theorem 1.6.4 when $T_1(x) > x$ for all $x \in (0,1)$ and $T_2(x) < x$ for all $x \in (0,1)$.

Hint: Prove that Claim 1 still holds. Prove that Claim 2 holds for T_1. However, show that $\lim_{n\to\infty} T_2^n(x) = 0$ and $\lim_{n\to\infty} T_2^{-n}(x) = 1$. Then prove that Claim 3 holds. Finally, show that

$$h(x) = \begin{cases} T_2^{-n(x)} \circ H \circ T_1^{n(x)}(x) & \text{if } x \in (0,1) \\ 1 & \text{if } x = 0 \\ 0 & \text{if } x = 1. \end{cases}$$

is a conjugacy map between T_1 and T_2.

Exercise 1.7.34. Prove that any two strictly increasing continuous maps of compact intervals which have the same finite number of fixed points, one of which fixes the endpoints of its domain while the other does not, are not topologically conjugate.

Hint: Let $T : [a,b] \to [a,b]$ and $S : [c,d] \to [c,d]$ be strictly increasing continuous maps with T fixing both a and b. Suppose that $h : [a,b] \to [c,d]$ is a conjugacy map between T and S. Using the Intermediate Value Theorem (IVT), show that $h(a)$ is an extreme (in other words, the leftmost or rightmost) fixed point of S, while $h(b)$ is the other extreme fixed point. Using the IVT once more, show that $h([a,b]) = [h(a), h(b)]$. Deduce that S fixes both c and d.

2 Homeomorphisms of the circle

In this chapter, we temporarily step away from the general theory of dynamical systems to consider more specific examples. In the preparatory Section 2.1, we first study lifts of maps of the unit circle. Using lifts, we investigate homeomorphisms of the unit circle in Section 2.2. These homeomorphisms constitute the primary class of systems of interest in this chapter. After showing that rotations are homeomorphisms, we introduce Poincaré's notion of *rotation number* for homeomorphisms of the circle. Roughly speaking, this number is the average rotation that a homeomorphism induces on the points of the circle over the long term. In Section 2.3, we examine in more detail diffeomorphisms of the circle. The main result of this chapter is Denjoy's theorem (Theorem 2.3.4), which states that if a C^2 diffeomorphism has an irrational rotation number, then the diffeomorphism constitutes a minimal system which is topologically conjugate to an irrational rotation. Strictly speaking, it suffices that the modulus of the diffeomorphism's derivative be a function of bounded variation. Denjoy's theorem is a generalization of Theorem 1.5.12.

The concept of rotation number generalizes to all continuous degree-one self-maps of the circle. It is then called *rotation interval*. A systematic account of the theory of such maps and, in particular, an extended treatment of the rotation interval, can be found in [3].

2.1 Lifts of circle maps

In this section, we discuss general properties that are shared by all circle maps, be they one-to-one or not. As already mentioned in Section 1.5, the unit circle $\mathbb{S}^1$ can be defined in many homeomorphic ways. Here, we will regard $\mathbb{S}^1$ as the quotient space $\mathbb{R}/\mathbb{Z}$, that is, as the space of all equivalence classes

$$[x] = \{x + n : n \in \mathbb{Z}\},$$

where $x \in \mathbb{R}$, with metric

$$\begin{aligned}\rho([x],[y]) &= \inf\{\big|(x+n)-(y+m)\big| : n,m \in \mathbb{Z}\} = \inf\{|x-y+k| : k \in \mathbb{Z}\} \\ &= \min\{|x-y-1|, |x-y|, |x-y+1|\}.\end{aligned}$$

To study the dynamics of a map on the circle, it is helpful to *lift* that map from $\mathbb{S}^1 \cong \mathbb{R}/\mathbb{Z}$ to $\mathbb{R}$. This can be done via the continuous surjection

$$\begin{array}{cccc}\pi & : & \mathbb{R} & \longrightarrow & \mathbb{S}^1 \\ & & x & \longmapsto & \pi(x) = [x].\end{array}$$

Definition 2.1.1. Let $T : \mathbb{S}^1 \to \mathbb{S}^1$ be a continuous map of the circle. A continuous map $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ is called a *lift* of T to $\mathbb{R}$ if $\pi \circ \widetilde{T} = T \circ \pi$, that is, if the following diagram

https://doi.org/10.1515/9783110702682-002

commutes:

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{\widetilde{T}} & \mathbb{R} \\ \pi \downarrow & & \downarrow \pi \\ \mathbb{S}^1 & \xrightarrow[T]{} & \mathbb{S}^1 \end{array}$$

In other words, $T : \mathbb{S}^1 \to \mathbb{S}^1$ is a factor of $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ via the factor map π.

Note that π is countably infinite to one, and hence cannot be a conjugacy map. In fact, $\pi : \mathbb{R} \to \mathbb{S}^1$ is a *covering map*, that is, every $[x] \in \mathbb{S}^1$ has an open neighborhood $U_{[x]}$, which is *evenly covered* by π. Being evenly covered by π means that the preimage $\pi^{-1}(U_{[x]})$ is a union of disjoint open subsets of $\mathbb{R}$, called *sheets* of $\pi^{-1}(U_{[x]})$, each of which is mapped homeomorphically by π onto $U_{[x]}$. Observe further that π is a local isometry when $\mathbb{S}^1 \cong \mathbb{R}/\mathbb{Z}$ is equipped with the metric ρ. More precisely, it is an isometry on any interval of length at most 1/2.

Lemma 2.1.2. *Every continuous map $T : \mathbb{S}^1 \to \mathbb{S}^1$ admits a lift $\widetilde{T} : \mathbb{R} \to \mathbb{R}$.*

Proof. Fix $s_0 \in \mathbb{R}$ and $t_0 \in \pi^{-1}(T(\pi(s_0)))$. Define $\widetilde{T}(s_0) := t_0$. Then $\pi(\widetilde{T}(s_0)) = \pi(t_0) = T(\pi(s_0))$. In other words, $\widetilde{T}(s_0)$ is a lift of s_0. This is the starting point of our lift. By considering successive neighborhoods, we will gradually lift the points of $\mathbb{S}^1$ to $\mathbb{R}$. For each $t \in \mathbb{R}$, let $U_{\pi(t)} \subseteq \mathbb{S}^1$ be the largest open neighborhood centered on $\pi(t)$ which is evenly covered by π, that is,

$$U_{\pi(t)} = \{\pi(r) : t - 1/2 < r < t + 1/2\}.$$

We define the sought-after lift by successive steps, as follows. Let V_0 be the unique sheet of $\pi^{-1}(U_{\pi(\widetilde{T}(s_0))})$, which contains $\widetilde{T}(s_0)$, that is, let

$$V_0 := \{r \in \mathbb{R} : \widetilde{T}(s_0) - 1/2 < r < \widetilde{T}(s_0) + 1/2\}.$$

Since $T(\pi(s_0)) = \pi(\widetilde{T}(s_0)) \in U_{\pi(\widetilde{T}(s_0))}$, since T is continuous and since $U_{\pi(\widetilde{T}(s_0))}$ is open, there exists $s_1' > s_0$ such that $T(\pi(s)) \in U_{\pi(\widetilde{T}(s_0))}$ for all $s_0 \le s < s_1'$. Denote by s_1 the supremum of all such s_1'. For each $s_0 \le s < s_1$, define $\widetilde{T}(s)$ to be the unique point of V_0 such that $\pi(\widetilde{T}(s)) = T(\pi(s))$.

If $s_1 = \infty$, then the lift is defined for all $s \ge s_0$. If $s_1 < \infty$, define

$$\widetilde{T}(s_1) := \lim_{s \nearrow s_1} \widetilde{T}(s).$$

Then

$$\widetilde{T}(s_1) \in \{\widetilde{T}(s_0) \pm 1/2\}$$

and

$$\pi(\widetilde{T}(s_1)) = T(\pi(s_1)).$$

Just like we did from s_0, the map $\widetilde{T}$ can then be extended beyond s_1 as follows. Let V_1 be the unique sheet of $\pi^{-1}(U_{\pi(\widetilde{T}(s_1))})$ which contains $\widetilde{T}(s_1)$. Since $T(\pi(s_1)) = \pi(\widetilde{T}(s_1)) \in U_{\pi(\widetilde{T}(s_1))}$, since T is continuous and since $U_{\pi(\widetilde{T}(s_1))}$ is open, there exists $s_2' > s_1$ such that $T(\pi(s)) \in U_{\pi(\widetilde{T}(s_1))}$ for all $s_1 \le s < s_2'$. Denote by s_2 the supremum of all such s_2'. For each $s_1 \le s < s_2$, let $\widetilde{T}(s)$ be the unique point of V_1 such that $\pi(\widetilde{T}(s)) = T(\pi(s))$. If $s_2 < \infty$, define

$$\widetilde{T}(s_2) := \lim_{s \nearrow s_2} \widetilde{T}(s).$$

Then

$$\widetilde{T}(s_2) \in \{\widetilde{T}(s_1) \pm 1/2\} \subseteq \{\widetilde{T}(s_0) \pm k/2 : k = 0, 1, 2\}$$

and

$$\pi(\widetilde{T}(s_2)) = T(\pi(s_2)).$$

Continuing in this way, either the procedure ends with some $s_n = \infty$ or it does not, in which case a strictly increasing sequence $(s_n)_{n=1}^{\infty}$ is constructed recursively. We claim that $\lim_{n\to\infty} s_n = \infty$. Otherwise, let $s^* = \lim_{n\to\infty} s_n < \infty$. The continuity of T and π then ensures that $T(\pi(s^*)) = \lim_{n\to\infty} T(\pi(s_n)) = \lim_{n\to\infty} \pi(\widetilde{T}(s_n))$. Note also that $\pi(\widetilde{T}(s_n))$ coincides with $\pi(\widetilde{T}(s_0))$ or $\pi(\widetilde{T}(s_0) + 1/2)$ since $\widetilde{T}(s_n) \in \{\widetilde{T}(s_0) \pm k/2 : k \in \mathbb{Z}\}$. Therefore, $T(\pi(s^*))$ coincides with $\pi(\widetilde{T}(s_0))$ or $\pi(\widetilde{T}(s_0) + 1/2)$, and thus the sequence $\pi(\widetilde{T}(s_n))$, $n \ge 0$, is eventually constant. But this is impossible since $\pi(\widetilde{T}(s_{n+1})) \notin U_{\pi(\widetilde{T}(s_n))}$ for all $n \ge 0$ by definition. This contradiction means that the lift can be extended indefinitely to the right of s_0, and a similar argument shows that it can be indefinitely extended to the left as well. □

Reading the proof of the above lemma, the reader may have acquired the intuition that, given a starting point, the lift of a map is unique. Moreover, given that starting points can only differ by an integer, so should entire lifts. This intuition proves to be correct.

Lemma 2.1.3. *Let $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ be a lift of a continuous map $T : \mathbb{S}^1 \to \mathbb{S}^1$. Then $\widehat{T} : \mathbb{R} \to \mathbb{R}$ is a lift of T if and only if $\widehat{T} = \widetilde{T} + k$ for some $k \in \mathbb{Z}$. In particular, given $s \in \mathbb{R}$ and $t \in \pi^{-1}(T(\pi(s)))$, there is a unique lift $\widehat{T}$ so that $\widehat{T}(s) = t$.*

Proof. Suppose first that $\widehat{T} = \widetilde{T} + k$ for some $k \in \mathbb{Z}$. It follows that $\widehat{T}$ is continuous since $\widetilde{T}$ is continuous. Moreover, for every $x \in \mathbb{R}$ we have

$$\pi \circ \widehat{T}(x) = \pi(\widetilde{T}(x) + k) = \pi(\widetilde{T}(x)) = T \circ \pi(x).$$

Thus, $\widehat{T}$ is a lift of T. This proves one implication.

For the converse implication, suppose that $\widehat{T}$, like $\widetilde{T}$, is a lift of T. For every $x \in \mathbb{R}$, we then have

$$\pi \circ \widehat{T}(x) = T \circ \pi(x) = \pi \circ \widetilde{T}(x).$$

Therefore, $\widehat{T}(x) - \widetilde{T}(x) \in \mathbb{Z}$ for every $x \in \mathbb{R}$. Define the function $k : \mathbb{R} \to \mathbb{Z}$ by $k(x) = \widehat{T}(x) - \widetilde{T}(x)$. Since both $\widetilde{T}$ and $\widehat{T}$ are continuous on $\mathbb{R}$, so is the function k. Then $k(\mathbb{R})$, as the image of a connected set under a continuous function, is a connected set. But since $\mathbb{Z}$ is totally disconnected, the set $k(\mathbb{R})$ must be a singleton. In other words, the function k must be constant. Hence, $\widehat{T} = \widetilde{T} + k$ for some constant $k \in \mathbb{Z}$. □

Thus, once a lift is found, all the other lifts can be obtained by translating vertically the graph of the original lift by all the integers. In the following lemma, we shall describe an important property that all lifts have in common.

Lemma 2.1.4. *If $T : \mathbb{S}^1 \to \mathbb{S}^1$ is continuous, then the number $\widetilde{T}(x+1) - \widetilde{T}(x)$ is an integer independent of the point $x \in \mathbb{R}$ and of the choice of lift $\widetilde{T}$.*

Proof. For every $x \in \mathbb{R}$,

$$\pi(\widetilde{T}(x+1)) = T(\pi(x+1)) = T(\pi(x)) = \pi(\widetilde{T}(x)).$$

Thus $\widetilde{T}(x+1) - \widetilde{T}(x)$ is an integer. Since $\mathbb{R} \ni x \mapsto \widetilde{T}(x+1) - \widetilde{T}(x) \in \mathbb{Z}$ is a continuous function, it follows, as in the proof of Lemma 2.1.3, that it is constant. This implies the independence from the point $x \in \mathbb{R}$.

If $\widehat{T} : \mathbb{R} \to \mathbb{R}$ is another lift of T, then $\widehat{T} = \widetilde{T} + k$ for some $k \in \mathbb{Z}$ according to Lemma 2.1.3. Therefore,

$$\begin{aligned}(\widehat{T}(x+1) - \widehat{T}(x)) - (\widetilde{T}(x+1) - \widetilde{T}(x)) \\ &= (\widehat{T}(x+1) - \widetilde{T}(x+1)) - (\widehat{T}(x) - \widetilde{T}(x)) \\ &= k - k = 0.\end{aligned}$$

This establishes the independence from the choice of lift. □

It then makes sense to introduce the following notion.

Definition 2.1.5. If $T : \mathbb{S}^1 \to \mathbb{S}^1$ is continuous, then the integer number $\widetilde{T}(x+1) - \widetilde{T}(x)$, which is independent of the point $x \in \mathbb{R}$ and of the choice of lift $\widetilde{T}$, is called the *degree* of the map T and is denoted by $\deg(T)$.

We can now reformulate Lemma 2.1.4 as follows.

Lemma 2.1.6. *If $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ is a lift of the continuous map $T : \mathbb{S}^1 \to \mathbb{S}^1$, then*

$$\widetilde{T}(x+1) = \widetilde{T}(x) + \deg(T), \quad \forall x \in \mathbb{R}.$$

By way of an induction argument, this result yields the following corollary.

Corollary 2.1.7. *If $\tilde{T} : \mathbb{R} \to \mathbb{R}$ is a lift of the continuous map $T : \mathbb{S}^1 \to \mathbb{S}^1$, then*

$$\tilde{T}(x+k) = \tilde{T}(x) + k\deg(T), \quad \forall x \in \mathbb{R}, \forall k \in \mathbb{Z}.$$

Proof. Let $d = \deg(T)$. By Lemma 2.1.6, the statement holds for all $x \in \mathbb{R}$ when $k = 1$. Suppose that the statement holds for all $x \in \mathbb{R}$ for some $k \in \mathbb{N}$. Then

$$\tilde{T}(x+k+1) = \tilde{T}\big((x+k)+1\big) = \tilde{T}(x+k) + d = \big(\tilde{T}(x) + kd\big) + d = \tilde{T}(x) + (k+1)d$$

for all $x \in \mathbb{R}$. Thus, the statement holds for $k+1$ whenever it holds for $k \in \mathbb{N}$. By induction, the statement holds for all $x \in \mathbb{R}$ and all $k \in \mathbb{N}$. When $k \le 0$, we have that $-k \ge 0$ and, since the statement holds for $-k$, we obtain

$$\tilde{T}(x+k) = \tilde{T}(x+k) + (-k)d + kd = \tilde{T}\big(x+k+(-k)\big) + kd = \tilde{T}(x) + kd$$

for all $x \in \mathbb{R}$. □

The degree has the following property relative to the composition of maps.

Lemma 2.1.8. *If $S, T : \mathbb{S}^1 \to \mathbb{S}^1$ are continuous maps of the unit circle, then*

$$\deg(S \circ T) = \deg(S) \cdot \deg(T).$$

Proof. Let $\tilde{S}$, $\tilde{T}$ be lifts of S and T, respectively. Then $\tilde{S} \circ \tilde{T}$ is a lift of $S \circ T$ since

$$\pi \circ (\tilde{S} \circ \tilde{T}) = (\pi \circ \tilde{S}) \circ \tilde{T} = (S \circ \pi) \circ \tilde{T} = S \circ (\pi \circ \tilde{T}) = S \circ (T \circ \pi) = (S \circ T) \circ \pi.$$

Let $x \in \mathbb{R}$. Using Corollary 2.1.7 once for T and once for S, we obtain that

$$\begin{aligned}
\deg(S \circ T) &= \tilde{S} \circ \tilde{T}(x+1) - \tilde{S} \circ \tilde{T}(x) \\
&= \tilde{S}\big(\tilde{T}(x+1)\big) - \tilde{S}\big(\tilde{T}(x)\big) \\
&= \tilde{S}\big(\tilde{T}(x) + \deg(T)\big) - \tilde{S}\big(\tilde{T}(x)\big) \\
&= \tilde{S}\big(\tilde{T}(x)\big) + \deg(T) \cdot \deg(S) - \tilde{S}\big(\tilde{T}(x)\big) \\
&= \deg(S) \cdot \deg(T).
\end{aligned}$$
□

This has the following consequence for iterates of maps.

Corollary 2.1.9. *If $T : \mathbb{S}^1 \to \mathbb{S}^1$ is a continuous map of the unit circle, then*

$$\deg(T^n) = \big(\deg(T)\big)^n, \quad \forall n \in \mathbb{N}.$$

As a direct repercussion of Corollaries 2.1.7 and 2.1.9, we have the following fact.

Corollary 2.1.10. *If $T : \mathbb{S}^1 \to \mathbb{S}^1$ is continuous and $\tilde{T}$ is a lift of T, then*

$$\tilde{T}^n(x+k) = \tilde{T}^n(x) + k\big(\deg(T)\big)^n$$

for all $x \in \mathbb{R}$, all $k \in \mathbb{Z}$ and all $n \in \mathbb{N}$.

We can further describe the difference between the values of iterates of various lifts.

Corollary 2.1.11. *Let $T : \mathbb{S}^1 \to \mathbb{S}^1$ be a continuous map and $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ be a lift of T. If $\widehat{T}$ is another lift of T so that $\widehat{T} = \widetilde{T} + k$ for some $k \in \mathbb{Z}$, then*

$$\widehat{T}^n = \widetilde{T}^n + k \sum_{j=0}^{n-1} (\deg(T))^j$$

for all $n \in \mathbb{N}$.

Proof. By hypothesis, the statement holds when $n = 1$. Suppose now that it holds for some $n \in \mathbb{N}$. Let $x \in \mathbb{R}$. Then

$$\begin{aligned}
\widehat{T}^{n+1}(x) &= \widehat{T}^n(\widehat{T}(x)) \\
&= \widetilde{T}^n(\widehat{T}(x)) + k \sum_{j=0}^{n-1} (\deg(T))^j \\
&= \widetilde{T}^n(\widetilde{T}(x) + k) + k \sum_{j=0}^{n-1} (\deg(T))^j \\
&= \widetilde{T}^n(\widetilde{T}(x)) + k(\deg(T))^n + k \sum_{j=0}^{n-1} (\deg(T))^j \\
&= \widetilde{T}^{n+1}(x) + k \sum_{j=0}^{n} (\deg(T))^j.
\end{aligned}$$

The result follows by induction. □

We now observe that the degree, as a map, is locally constant.

Lemma 2.1.12. *If $C(\mathbb{S}^1, \mathbb{S}^1)$ is endowed with the topology of uniform convergence, then the degree map $C(\mathbb{S}^1, \mathbb{S}^1) \ni T \mapsto \deg(T) \in \mathbb{Z}$ is locally constant and hence continuous.*

Proof. We shall regard $\mathbb{S}^1$ as $(\mathbb{R}/\mathbb{Z}, \rho)$, with the metric ρ as defined at the beginning of this section. Let $T, S \in C(\mathbb{S}^1, \mathbb{S}^1)$ be such that $\rho_\infty(T, S) < 1/4$, where we recall that

$$\rho_\infty(T, S) = \sup\{\rho(T(x), S(x)) : x \in \mathbb{S}^1\}.$$

Let $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ be a lift of T. Since $\rho_\infty(T, S) < 1/4$, the restriction of π to the real interval $[\widetilde{T}(0) - \rho_\infty(T, S), \widetilde{T}(0) + \rho_\infty(T, S)]$ is an isometry. Thus, the connected set $\pi([\widetilde{T}(0) - \rho_\infty(T, S), \widetilde{T}(0) + \rho_\infty(T, S)])$, which is "centered" on $\pi(\widetilde{T}(0)) = T(\pi(0)) \in \mathbb{S}^1$, contains the point $S(\pi(0))$, since $\rho(T(\pi(0)), S(\pi(0))) \le \rho_\infty(T, S)$. Therefore, there exists some

$$t \in [\widetilde{T}(0) - \rho_\infty(T, S), \widetilde{T}(0) + \rho_\infty(T, S)]$$

such that $\pi(t) = S(\pi(0))$. There then exists a unique lift $\widetilde{S} : \mathbb{R} \to \mathbb{R}$ of S such that $\widetilde{S}(0) = t$. In particular,

$$\left|\widetilde{T}(0) - \widetilde{S}(0)\right| \le \rho_\infty(T, S).$$

We shall prove that

$$\left|\widetilde{T}(x) - \widetilde{S}(x)\right| \le \rho_\infty(T, S), \quad \forall x \in \mathbb{R}. \tag{2.1}$$

Indeed, suppose by way of contradiction that there exists $y \in \mathbb{R}$ such that $|\widetilde{T}(y) - \widetilde{S}(y)| > \rho_\infty(T, S)$. Since the function $x \mapsto |\widetilde{T}(x) - \widetilde{S}(x)|$ is continuous and $|\widetilde{T}(0) - \widetilde{S}(0)| \le \rho_\infty(T, S) < 1/4$, there must exist $w \in \mathbb{R}$ such that

$$\rho_\infty(T, S) < \left|\widetilde{T}(w) - \widetilde{S}(w)\right| < 1/4.$$

But since the restriction of π to $(\widetilde{T}(w) - 1/4, \widetilde{T}(w) + 1/4)$ is an isometry and $\widetilde{S}(w)$ belongs to that real interval, it follows that

$$\rho_\infty(T, S) \ge \rho\big(T(\pi(w)), S(\pi(w))\big) = \rho\big(\pi(\widetilde{T}(w)), \pi(\widetilde{S}(w))\big) = \left|\widetilde{T}(w) - \widetilde{S}(w)\right| > \rho_\infty(T, S),$$

which is impossible. This proves (2.1). Using that formula, we deduce that

$$\begin{aligned}
\left|\deg(T) - \deg(S)\right| &= \left|\big(\widetilde{T}(x+1) - \widetilde{T}(x)\big) - \big(\widetilde{S}(x+1) - \widetilde{S}(x)\big)\right| \\
&= \left|\big(\widetilde{T}(x+1) - \widetilde{S}(x+1)\big) - \big(\widetilde{T}(x) - \widetilde{S}(x)\big)\right| \\
&\le \left|\widetilde{T}(x+1) - \widetilde{S}(x+1)\right| + \left|\widetilde{T}(x) - \widetilde{S}(x)\right| \\
&\le 2\rho_\infty(T, S) < 1.
\end{aligned}$$

As $\deg(S)$ and $\deg(T)$ are integers, we conclude that $\deg(S) = \deg(T)$. □

Remark 2.1.13. Another way of stating (2.1) is to say that if $T, S \in C(\mathbb{S}^1, \mathbb{S}^1)$ are circle maps such that $\rho_\infty(T, S) < 1/4$, then T and S have lifts $\widetilde{T}$ and $\widetilde{S}$ such that $\widetilde{\rho}_\infty(\widetilde{T}, \widetilde{S}) \le \rho_\infty(T, S)$, where

$$\widetilde{\rho}_\infty(\widetilde{T}, \widetilde{S}) := \sup\left\{\left|\widetilde{T}(x) - \widetilde{S}(x)\right| : x \in [0, 1]\right\} = \sup\left\{\left|\widetilde{T}(x) - \widetilde{S}(x)\right| : x \in \mathbb{R}\right\}.$$

The following lemma gives us more information about lifts and their fixed points.

Lemma 2.1.14. *Every continuous map $T : \mathbb{S}^1 \to \mathbb{S}^1$ with $\deg(T) \ne 1$ has a lift $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ with a fixed point in $[-1/2, 1/2]$. Moreover,*

$$\operatorname{dist}(0, \operatorname{Fix}(\widetilde{T})) \to 0 \quad \textit{whenever } T \to E_k \textit{ uniformly},$$

where $E_k([x]) := [kx]$ and where $k = \deg(T)$.

Proof. Let $k = \deg(T)$ and $\widehat{T} : \mathbb{R} \to \mathbb{R}$ be a lift of T. Define a map $D : \mathbb{R} \to \mathbb{R}$ by setting

$$D(x) := \widehat{T}(x) - x.$$

By definition of k, we have that $\widehat{T}(1/2) = \widehat{T}(-1/2 + 1) = \widehat{T}(-1/2) + k$, and hence

$$D(1/2) = \widehat{T}(1/2) - 1/2 = \widehat{T}(-1/2) + k - 1/2 = D(-1/2) + k - 1.$$

Since $k \neq 1$, the interval $D([-1/2, 1/2])$ has length at least 1, and thus contains an integer, say m. In other words, there exists $x_0 \in [-1/2, 1/2]$ such that $D(x_0) = m$, that is, such that $\widehat{T}(x_0) - m = x_0$. Letting $\widetilde{T} = \widehat{T} - m$, which is also a lift of T by Lemma 2.1.3, we obtain that $\widetilde{T}(x_0) = x_0$ and the first assertion in our lemma is proved.

For the second part, fix $\delta > 0$. One immediately verifies that $\widetilde{E}_k(x) = kx$ is a lift of E_k to $\mathbb{R}$ and

$$\widetilde{E}_k(-\delta) - (-\delta) = -k\delta + \delta = -(k-1)\delta \quad \text{whereas} \quad \widetilde{E}_k(\delta) - \delta = (k-1)\delta.$$

Since $k \neq 1$, the numbers $\widetilde{E}_k(-\delta) - (-\delta)$ and $\widetilde{E}_k(\delta) - \delta$ have opposite signs. Therefore, in view of Remark 2.1.13, if T is sufficiently close to E_k, then there exists a lift $\widetilde{T}$ of T such that $\widetilde{T}(-\delta) - (-\delta)$ and $\widetilde{T}(\delta) - \delta$ have the same signs as $\widetilde{E}_k(-\delta) - (-\delta)$ and $\widetilde{E}_k(\delta) - \delta$, respectively. Consequently, there exists $s \in (-\delta, \delta)$ so that $\widetilde{T}(s) - s = 0$. □

Remark 2.1.15. Lemma 2.1.14 does not generally hold for circle maps of degree 1. Indeed, it clearly does not hold for any irrational rotation.

2.2 Orientation-preserving homeomorphisms of the circle

Every homeomorphism of the unit circle is either *orientation preserving* or *orientation reversing*, which means that either the homeomorphism preserves the order of points on the circle or it reverses their order. In this section, we shall study orientation-preserving homeomorphisms. The differences that occur when considering orientation-reversing homeomorphisms are covered in the exercises at the end of the chapter.

By convention, arcs will be traversed in the counterclockwise direction along the unit circle. That is, the closed arc $[a, b]$ is the arc that consists of all points that are met when moving in the counterclockwise direction from point a to point b, including these two points. The open arc (a, b) simply excludes the endpoints from the closed arc $[a, b]$. Hence, $[\pi(0), \pi(1/2)]$ is the upper-half circle while $[\pi(1/2), \pi(0)]$ corresponds to the lower-half circle. The left half circle is represented by $[\pi(1/4), \pi(3/4)]$ whereas the right-half circle is the arc $[\pi(3/4), \pi(1/4)]$.

Definition 2.2.1. A homeomorphism $f : \mathbb{S}^1 \to \mathbb{S}^1$ is said to be *orientation-preserving* if $f(c) \in (f(a), f(b))$ whenever $c \in (a, b)$.

We will show that this is equivalent to the fact that any lift $\widetilde{f}$ of f is an increasing homeomorphism of $\mathbb{R}$. Recall that lifts exist by Lemma 2.1.2, and are unique up to addition by an integer according to Lemma 2.1.3.

Lemma 2.2.2. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be a homeomorphism of the unit circle. Let $\widetilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of f. Then $\widetilde{f}$ is a homeomorphism of $\mathbb{R}$.*

Proof. Let $\widetilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of f. By definition, $\widetilde{f}$ is surjective and continuous. It remains to show that it is injective. Suppose that this is not the case, that is, there exist $x, y \in \mathbb{R}$ such that $\widetilde{f}(x) = \widetilde{f}(y)$. We claim that there then exist $\widetilde{x}, \widetilde{y} \in \mathbb{R}$ such that $|\widetilde{x}-\widetilde{y}| < 1$ and $\widetilde{f}(\widetilde{x}) = \widetilde{f}(\widetilde{y})$. Indeed, if $|x-y| < 1$ then simply let $\widetilde{x} = x$ and $\widetilde{y} = y$. Otherwise, that is, if $|x-y| \geq 1$ then there exists a unique $k \in \mathbb{N}$ such that $k \leq |x-y| < k+1$. Without loss of generality, we may assume that $x > y$. Then $y + k \leq x < y + k + 1$. Moreover,

$$\widetilde{f}(y+k) = \widetilde{f}(y) + k\deg(f) = \widetilde{f}(x) + k\deg(f)$$

while

$$\widetilde{f}(y+k+1) = \widetilde{f}(y) + (k+1)\deg(f) = \widetilde{f}(x) + k\deg(f) + \deg(f).$$

If $\deg(f) \neq 0$, applying the intermediate value theorem on the intervals $[y+k, x]$ and $[x, y+k+1]$ gives that there exist $x_1 \in (y+k, x)$ and $x_2 \in (x, y+k+1)$ such that

$$\widetilde{f}(x_1) = \widetilde{f}(x_2) = \widetilde{f}(x) + k\deg(f)/2.$$

In this case, let $\widetilde{x} = x_1$ and $\widetilde{y} = x_2$.

If $\deg(f) = 0$, then

$$\widetilde{f}(y+k) = \widetilde{f}(y+k+1) = \widetilde{f}(x).$$

If there exists $y+k < z < y+k+1$ such that $\widetilde{f}(z) \neq \widetilde{f}(x)$, then applying the intermediate value theorem on the intervals $[y+k, z]$ and $[z, y+k+1]$ yields points $z_1 \in (y+k, z)$ and $z_2 \in (z, y+k+1)$ such that

$$\widetilde{f}(z_1) = \widetilde{f}(z_2) = \big(\widetilde{f}(x) + \widetilde{f}(z)\big)/2.$$

In this case, let $\widetilde{x} = z_1$ and $\widetilde{y} = z_2$. Otherwise, $\widetilde{f}$ is equal to $\widetilde{f}(x)$ on the entire interval $[y+k, y+k+1]$ and we let $\widetilde{x} = x$ and $\widetilde{y} \in (y+k, y+k+1) \setminus \{x\}$.

In all cases, $|\widetilde{x} - \widetilde{y}| < 1$ and $\widetilde{f}(\widetilde{x}) = \widetilde{f}(\widetilde{y})$. It then follows that

$$f \circ \pi(\widetilde{x}) = \pi \circ \widetilde{f}(\widetilde{x}) = \pi \circ \widetilde{f}(\widetilde{y}) = f \circ \pi(\widetilde{y}).$$

Since f is injective, this means that $\pi(\widetilde{x}) = \pi(\widetilde{y})$. But $\pi(\widetilde{x}) \neq \pi(\widetilde{y})$ since $|\widetilde{x} - \widetilde{y}| < 1$. This contradiction shows that $\widetilde{f}$ is injective. In summary, $\widetilde{f}$ is a continuous bijection of $\mathbb{R}$. Then $\widetilde{f}$ is either strictly increasing or strictly decreasing. In particular, it is a homeomorphism and $\deg(f) \neq 0$. □

Corollary 2.2.3. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be a homeomorphism of the unit circle. Let $\widetilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of f. If f is orientation preserving, then $\widetilde{f}$ is an increasing homeomorphism of $\mathbb{R}$.*

Proof. By Lemma 2.2.2, we know that $\widetilde{f}$ is a homeomorphism of $\mathbb{R}$. Therefore, $\widetilde{f}$ is either strictly increasing or strictly decreasing. Suppose for a contradiction that $\widetilde{f}$ is strictly decreasing. Changing lift if necessary, we may assume that $\widetilde{f}(0) \in (0, 1]$. Since $\widetilde{f}$ is continuous and strictly decreasing, there exists $0 < \delta < 1$ such that $0 < \widetilde{f}(\delta) < \widetilde{f}(0)$. Consider the arc $(a, b) := (\pi(0), \pi(\delta)) = \pi((0, \delta)) \subseteq \mathbb{S}^1$. Then

$$(f(a), f(b)) = \big(f(\pi(0)), f(\pi(\delta))\big) = \big(\pi(\widetilde{f}(0)), \pi(\widetilde{f}(\delta))\big).$$

Let $c \in (a, b)$. Then there exists $0 < \widetilde{c} < \delta$ such that $\pi(\widetilde{c}) = c$. Therefore, $0 < \widetilde{f}(\delta) < \widetilde{f}(\widetilde{c}) < \widetilde{f}(0) \leq 1$, and hence

$$f(c) = f(\pi(\widetilde{c})) = \pi(\widetilde{f}(\widetilde{c})) \in \pi\big((\widetilde{f}(\delta), \widetilde{f}(0))\big) = \big(\pi(\widetilde{f}(\delta)), \pi(\widetilde{f}(0))\big).$$

Consequently, $f(c) \notin (f(a), f(b))$. This contradicts the hypothesis that f is orientation preserving. We thus conclude that $\widetilde{f}$ must be strictly increasing. □

Let us now show that the degree of every orientation-preserving homeomorphism of the circle is equal to 1, where we recall that the degree of a continuous map $T : \mathbb{S}^1 \to \mathbb{S}^1$ is defined to be the integer $\widetilde{T}(x + 1) - \widetilde{T}(x)$, which is independent of the choice of the point $x \in \mathbb{R}$ and of the choice of lift $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ according to Lemma 2.1.4.

Lemma 2.2.4. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be an orientation-preserving homeomorphism. Then* $\deg(f) = 1$.

Proof. Let $\widetilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of f. Since $\widetilde{f}$ is strictly increasing and $\widetilde{f}(1) = \widetilde{f}(0) + \deg(f)$, it is clear that $\deg(f) \geq 1$. Suppose that $\deg(f) \geq 2$. Then, as $\widetilde{f}$ is continuous and $\widetilde{f}(1) = \widetilde{f}(0) + \deg(f)$, the intermediate value theorem guarantees the existence of some real number $0 < y < 1$ such that $\widetilde{f}(y) = \widetilde{f}(0) + 1$. But then

$$f \circ \pi(y) = \pi \circ \widetilde{f}(y) = \pi\big(\widetilde{f}(0) + 1\big) = \pi \circ \widetilde{f}(0) = f \circ \pi(0).$$

Since f is injective (after all, it is a homeomorphism), this means that $\pi(y) = \pi(0)$. However, $\pi(y) \neq \pi(0)$ since $0 < y < 1$. This contradiction shows that the assumption $\deg(f) \geq 2$ cannot hold. Thus $\deg(f) = 1$. □

We leave it to the reader to prove that any homeomorphism F of $\mathbb{R}$ with the property that $F(x + 1) = F(x) + 1$ for all $x \in \mathbb{R}$ generates an orientation-preserving homeomorphism f of S^1 (see Exercise 2.4.3).

Observe that the inverse $f^{-1} : \mathbb{S}^1 \to \mathbb{S}^1$ of an orientation-preserving homeomorphism $f : \mathbb{S}^1 \to \mathbb{S}^1$ is also an orientation-preserving homeomorphism. Therefore, $\deg(f^{-1}) = \deg(f) = 1$. Moreover, note that if $\widetilde{f} : \mathbb{R} \to \mathbb{R}$ is a lift of $f : \mathbb{S}^1 \to \mathbb{S}^1$, then $\widetilde{f}^{-1} : \mathbb{R} \to \mathbb{R}$ is a lift of $f^{-1} : \mathbb{S}^1 \to \mathbb{S}^1$. We deduce the following for orientation-preserving homeomorphisms.

Corollary 2.2.5. *Let $\tilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of an orientation-preserving homeomorphism $f : \mathbb{S}^1 \to \mathbb{S}^1$. For all $x \in \mathbb{R}$, all $k \in \mathbb{Z}$ and all $n \in \mathbb{Z}$, we have that*

$$\tilde{f}^n(x + k) = \tilde{f}^n(x) + k.$$

Proof. The result is trivial when $n = 0$. In light of Lemma 2.2.4, the result follows directly from Corollary 2.1.10 for all $n \in \mathbb{N}$. Using f^{-1} instead of f in Lemma 2.2.4 and Corollary 2.1.10, the result follows for $n \le -1$. □

By induction one can also show the following (cf. Exercise 2.4.1).

Corollary 2.2.6. *Let $\tilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of an orientation-preserving homeomorphism $f : \mathbb{S}^1 \to \mathbb{S}^1$. Given any $k \in \mathbb{N}$, we have that*

$$|x - y| \le k \Longrightarrow \left|\tilde{f}^n(x) - \tilde{f}^n(y)\right| \le k, \quad \forall n \in \mathbb{Z}.$$

Moreover, if the left inequality is strict ($<$ instead of $\le$), then so is the right one.

Corollary 2.2.7. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be an orientation-preserving homeomorphism and $\tilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of f. If $\tilde{g}$ is another lift of f so that $\tilde{g} = \tilde{f} + k$ for some $k \in \mathbb{Z}$, then*

$$\tilde{g}^n = \tilde{f}^n + nk$$

for all $n \in \mathbb{Z}$.

Proof. Apply Corollary 2.1.11 with f and f^{-1} in lieu of T. Recall that $\deg(f) = \deg(f^{-1}) = 1$ since both f and f^{-1} are orientation-preserving homeomorphisms. □

Lemma 2.2.4 implies immediately that $\tilde{f} - \mathrm{Id}_{\mathbb{R}}$ is a periodic function with period 1, where $\mathrm{Id}_{\mathbb{R}} : \mathbb{R} \to \mathbb{R}$ is the identity map (cf. Exercise 2.4.2).

Corollary 2.2.8. *Let $\tilde{f}$ be a lift of f. Then $\tilde{f} - \mathrm{Id}_{\mathbb{R}}$ is a periodic function with period 1. More generally, an increasing homeomorphism $\tilde{g} : \mathbb{R} \to \mathbb{R}$ is a lift of an orientation-preserving homeomorphism of the circle if and only if $\tilde{g} - \mathrm{Id}_{\mathbb{R}}$ is a periodic function with period 1.*

Let us now give the simplest example of an orientation-preserving homeomorphism of the unit circle: a rotation.

Example 2.2.9. Let $\alpha \in \mathbb{R}$. If $f([x]) := [x + \alpha]$ is the rotation of the unit circle by the angle α, then $\tilde{f}(x) = x + \alpha$ is a lift of f and for all $x \in \mathbb{R}$ we have that

$$\lim_{n\to\infty} \frac{\tilde{f}^n(x)}{n} = \lim_{n\to\infty} \frac{x + n\alpha}{n} = \alpha.$$

2.2.1 Rotation numbers

In this section, we introduce a number that allows us to think of the dynamics of a given homeomorphism of the unit circle mimicking, in some sense, the dynamics of a rotation of the circle. Accordingly, this number will be called the *rotation number* of the said homeomorphism.

The first result generalizes to orientation-preserving homeomorphisms the observation about the ratio $\tilde{f}^n(x)/n$ made for rotations of the circle in Example 2.2.9.

Proposition 2.2.10. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be an orientation-preserving homeomorphism of the unit circle and $\tilde{f} : \mathbb{R} \to \mathbb{R}$ a lift of f. Then the following statements hold:*

(a) *The number*

$$\rho(\tilde{f}) := \lim_{n\to\infty} \frac{\tilde{f}^n(x)}{n}$$

exists for all $x \in \mathbb{R}$ and is independent of x.

(b) *If $\tilde{g} = \tilde{f} + k$ for some $k \in \mathbb{Z}$, then $\rho(\tilde{g}) = \rho(\tilde{f}) + k$. That is, the choice of lift only changes ρ by an integer.*

(c) *For every $m \in \mathbb{N}$, we have $\rho(\tilde{f}^m) = m \cdot \rho(\tilde{f})$.*

(d) *The number $\rho(\tilde{f})$ is an integer if and only if f has a fixed point.*

(e) *The number $\rho(\tilde{f})$ is rational if and only if f has a periodic point.*

(f) *Let $x \in \mathbb{R}$. If $q \geq 1$ and r are integers such that $\tilde{f}^q(x) \leq x + r$, then $q\rho(\tilde{f}) \leq r$.*

(g) *Let $x \in \mathbb{R}$. If $q \geq 1$ and r are integers such that $\tilde{f}^q(x) \geq x + r$, then $q\rho(\tilde{f}) \geq r$.*

Proof. (a) We prove this proposition in two steps. We first assume the existence of $\rho(\tilde{f})$ and prove its independence of the point x chosen. We then prove the existence of $\rho(\tilde{f})$.

Step 1: If $\rho(\tilde{f})$ exists for some $x \in \mathbb{R}$, then it exists for all $y \in \mathbb{R}$ and is the same for all y.

Suppose that $\rho(\tilde{f})$ exists for some $x \in \mathbb{R}$. Choose any $y \in \mathbb{R}$. Then there exists $k \in \mathbb{N}$ such that $|y - x| \leq k$. By Corollary 2.2.6, we have that for every $n \in \mathbb{N}$,

$$\left|\tilde{f}^n(x) - \tilde{f}^n(y)\right| \leq k.$$

Therefore, for all $n \in \mathbb{N}$ we obtain that

$$\frac{\tilde{f}^n(x)}{n} - \frac{k}{n} \leq \frac{\tilde{f}^n(y)}{n} \leq \frac{\tilde{f}^n(x)}{n} + \frac{k}{n}.$$

Passing to the limit as n tends to infinity, we conclude that

$$\lim_{n\to\infty} \frac{\tilde{f}^n(x)}{n} \leq \liminf_{n\to\infty} \frac{\tilde{f}^n(y)}{n} \leq \limsup_{n\to\infty} \frac{\tilde{f}^n(y)}{n} \leq \lim_{n\to\infty} \frac{\tilde{f}^n(x)}{n}.$$

Consequently,

$$\lim_{n\to\infty}\frac{\tilde{f}^n(y)}{n}=\lim_{n\to\infty}\frac{\tilde{f}^n(x)}{n}.$$

Since $y\in\mathbb{R}$ is arbitrary, the proof of Step 1 is complete.

Step 2. The rotation number $\rho(\tilde{f})$ always exists.

According to Step 1, if the limit $\rho(\tilde{f})$ exists for any particular x, it exists for all x. So, without loss of generality, set $x=0$. Fix momentarily $m\in\mathbb{N}$. Then there exists a unique $k\in\mathbb{Z}$ such that

$$k\le\tilde{f}^m(0)<k+1.$$

Using Corollary 2.2.5 and the fact that $\tilde{f}^m$ is strictly increasing, we deduce by induction that for any $n\in\mathbb{N}$,

$$nk\le\tilde{f}^{nm}(0)<n(k+1).$$

It follows that

$$\frac{k}{m}\le\frac{\tilde{f}^m(0)}{m}<\frac{k+1}{m}\quad\text{and}\quad\frac{k}{m}\le\frac{\tilde{f}^{nm}(0)}{nm}<\frac{k+1}{m}.$$

Consequently,

$$\left|\frac{\tilde{f}^m(0)}{m}-\frac{\tilde{f}^{nm}(0)}{nm}\right|<\frac{k+1}{m}-\frac{k}{m}=\frac{1}{m}.$$

Interchanging the roles of m and n yields the inequality

$$\left|\frac{\tilde{f}^n(0)}{n}-\frac{\tilde{f}^{nm}(0)}{nm}\right|<\frac{1}{n}.$$

By the triangle inequality, we get that

$$\left|\frac{\tilde{f}^n(0)}{n}-\frac{\tilde{f}^m(0)}{m}\right|<\frac{1}{n}+\frac{1}{m}.$$

This shows that the sequence $(\tilde{f}^n(0)/n)_{n=1}^{\infty}$ is a Cauchy sequence in $\mathbb{R}$ and is therefore convergent. Hence, $\rho(\tilde{f})$ exists.

(b) Suppose that $\tilde{g}=\tilde{f}+k$. By Corollary 2.2.7, we know that $\tilde{g}^n=\tilde{f}^n+nk$ for all $n\in\mathbb{N}$. It follows that

$$\rho(\tilde{g})=\lim_{n\to\infty}\frac{\tilde{g}^n(x)}{n}=\lim_{n\to\infty}\frac{\tilde{f}^n(x)}{n}+k=\rho(\tilde{f})+k.$$

(c) Given that $\tilde{f}^m$ is a lift of f^m for every $m \in \mathbb{Z}$, the number $\rho(\tilde{f}^m)$ is well-defined for all m due to statement (a). Using any $x \in \mathbb{R}$, we obtain that

$$\rho(\tilde{f}^m) = \lim_{n\to\infty} \frac{\tilde{f}^{mn}(x)}{n} = \lim_{n\to\infty} \frac{\tilde{f}^{mn}(x)}{mn} \cdot m = \rho(\tilde{f}) \cdot m.$$

(d) Assume first that f has a fixed point. This means that there exists $z \in \mathbb{S}^1$ such that $f(z) = z$. Let $x \in \mathbb{R}$ be such that $\pi(x) = z$. Then

$$\pi(x) = z = f(z) = f(\pi(x)) = \pi(\tilde{f}(x)).$$

Therefore, $\tilde{f}(x) - x = k$ for some $k \in \mathbb{Z}$. Invoking Corollary 2.2.5 once again, we know that $\tilde{f}^n(x) = x + nk$ for each $n \in \mathbb{N}$. Therefore,

$$\rho(\tilde{f}) = \lim_{n\to\infty} \frac{\tilde{f}^n(x)}{n} = \lim_{n\to\infty} \frac{x + nk}{n} = k \in \mathbb{Z}.$$

This proves one implication. To derive the converse, assume that $\rho(\tilde{f}) \in \mathbb{Z}$. We aim to show that f has a fixed point. Replacing $\tilde{f}$ by $\tilde{f} - \rho(\tilde{f})$, which, by Lemma 2.1.3, is also a lift of f, we may assume that $\rho(\tilde{f}) = 0$. Assume by way of contradiction that $\tilde{f}$ has no fixed point. By the intermediate value theorem, this means that either $\tilde{f}(x) > x$ for all $x \in \mathbb{R}$ or $\tilde{f}(x) < x$ for all $x \in \mathbb{R}$. Suppose that $\tilde{f}(x) > x$ for all $x \in \mathbb{R}$ (a similar argument holds in the other case). This implies in particular that $\tilde{f}(0) > 0$, and thus the sequence $(\tilde{f}^n(0))_{n=1}^{\infty}$ is increasing. We further claim that $\tilde{f}^n(0) < 1$ for all $n \in \mathbb{N}$. Indeed, if this were not the case, then we would have $\tilde{f}^N(0) \geq 1$ for some $N \in \mathbb{N}$. We would deduce by induction that $\tilde{f}^{nN}(0) \geq n$ for all $n \in \mathbb{N}$. Hence, we would conclude that

$$\rho(\tilde{f}) = \lim_{n\to\infty} \frac{\tilde{f}^{nN}(0)}{nN} \geq \frac{1}{N} > 0.$$

But this would contradict the fact that $\rho(\tilde{f}) = 0$. Thus, $0 < \tilde{f}^n(0) < 1$ for all $n \in \mathbb{N}$. Summarizing, $(\tilde{f}^n(0))_{n=1}^{\infty}$ is a bounded monotonic sequence and as such is convergent. Let $L := \lim_{n\to\infty} \tilde{f}^n(0)$. Because of the continuity of $\tilde{f}$, Lemma 1.1.4 implies that L is a fixed point of $\tilde{f}$. This contradicts our assumption that $\tilde{f}$ has no fixed point. Thus, $\tilde{f}$ has a fixed point and f, as a factor of $\tilde{f}$, has a fixed point too.

(e) Let $\tilde{f}$ be a lift of f. Note that f has a periodic point if and only if there exists $m \in \mathbb{N}$ for which f^m has a fixed point. By statements (c) and (d), this is equivalent to stating that f has a periodic point if and only if $\rho(\tilde{f})$ is rational.

(f) Suppose that $q \geq 1$ and r are integers such that $\tilde{f}^q(x) \leq x + r$. Using Corollary 2.2.5 and the fact that $\tilde{f}^q$ is increasing, we deduce by induction that $\tilde{f}^{nq}(x) \leq x + nr$ for each $n \in \mathbb{N}$. Then

$$\rho(\tilde{f}) = \lim_{n\to\infty} \frac{\tilde{f}^{qn}(x)}{qn} \leq \lim_{n\to\infty} \frac{x + nr}{qn} = \frac{r}{q}.$$

(g) The proof proceeds analogously to (f) and is left as an exercise for the reader (see Exercise 2.4.4). □

Proposition 2.2.10, in conjunction with Example 2.2.9, suggests the following definition and terminology.

Definition 2.2.11. The *rotation number* $\rho(f)$ of an orientation-preserving homeomorphism $f : \mathbb{S}^1 \to \mathbb{S}^1$ of the unit circle is defined to be $\rho(\tilde{f}) \pmod 1$, where $\tilde{f} : \mathbb{R} \to \mathbb{R}$ is any lift of f.

As mentioned at the very beginning of the chapter, the rotation number should be thought of as the average rotation that the homeomorphism induces on the points of $\mathbb{S}^1$ over the long term. Statements (a) and (b) of Proposition 2.2.10 ensure that the rotation number exists and is well-defined. Statements (c), (d), and (e) translate into the assertions below. Note, though, that statements (f) and (g) have no counterparts for $\rho(f)$.

Proposition 2.2.12. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be an orientation-preserving homeomorphism of the unit circle. Then the following statements hold:*
(c′) *For every $m \in \mathbb{N}$, we have that $\rho(f^m) = m \cdot \rho(f) \pmod 1$.*
(d′) *The rotation number $\rho(f)$ is equal to zero if and only if f has a fixed point.*
(e′) *The rotation number $\rho(f)$ is rational if and only if f has a periodic point.*

Given that every orientation-preserving homeomorphism of the circle has an associated rotation number, it is natural to ask whether the rotation number is a topological conjugacy invariant. This is, in fact, very nearly the case, as we now show.

Theorem 2.2.13. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ and $g : \mathbb{S}^1 \to \mathbb{S}^1$ be topologically conjugate orientation-preserving homeomorphisms of the unit circle. If the conjugacy map preserves orientation, then $\rho(f) = \rho(g)$. If the conjugacy map reverses orientation, then $\rho(f) + \rho(g) = 0 \pmod 1$.*

Proof. Let $f, g : \mathbb{S}^1 \to \mathbb{S}^1$ be topologically conjugate orientation-preserving homeomorphisms. Let $h : \mathbb{S}^1 \to \mathbb{S}^1$ be a conjugacy map between them, so that $h \circ f = g \circ h$, and let $\tilde{h}$ be a lift of h. Then $\tilde{h}(x+n) = \tilde{h}(x) + n$ for all $x \in \mathbb{R}$ and $n \in \mathbb{Z}$ if h is orientation preserving. If h is orientation reversing, then $\tilde{h}(x+n) = \tilde{h}(x) - n$ for all $x \in \mathbb{R}$ and $n \in \mathbb{Z}$. The proof of this last statement is deferred to Exercise 2.4.5(c). Denoting the integer and fractional parts of x by $\lfloor x \rfloor$ and $\langle x \rangle$, respectively, and observing that $\tilde{h}(\langle x \rangle)$ lies between $\tilde{h}(0)$ and $\tilde{h}(1)$, it follows that for all $x \in \mathbb{R}$,

$$\begin{aligned}
\lim_{x\to\infty} \frac{\tilde{h}(x)}{x} &= \lim_{x\to\infty} \frac{\tilde{h}(\lfloor x \rfloor + \langle x \rangle)}{\lfloor x \rfloor + \langle x \rangle} \\
&= \lim_{x\to\infty} \frac{\tilde{h}(\langle x \rangle) \pm \lfloor x \rfloor}{\lfloor x \rfloor} \cdot \frac{\lfloor x \rfloor}{\lfloor x \rfloor + \langle x \rangle} \\
&= \lim_{x\to\infty} \frac{\tilde{h}(\langle x \rangle) \pm \lfloor x \rfloor}{\lfloor x \rfloor} \cdot \lim_{x\to\infty} \frac{\lfloor x \rfloor}{\lfloor x \rfloor + \langle x \rangle} \\
&= \pm 1,
\end{aligned}$$

depending on whether $\widetilde{h}$ is orientation preserving (+1) or orientation reversing (−1). The same relation holds for $\widetilde{h}^{-1}$ since it is a lift of h^{-1}. If $\rho(f) = \rho(g) = 0$, then we are done. So, suppose that at least one of $\rho(f)$ and $\rho(g)$ is positive. Without loss of generality, suppose that $\rho(g) > 0$. Let $\widetilde{g}$ be the lift of g such that $\rho(\widetilde{g}) = \rho(g)$. Then the map $\widetilde{f} := \widetilde{h}^{-1} \circ \widetilde{g} \circ \widetilde{h}$ is a lift of f. Indeed, $\widetilde{f}$ is an increasing homeomorphism of $\mathbb{R}$ since $\widetilde{g}$ is an increasing homeomorphism, while $\widetilde{h}$ is either an increasing homeomorphism or a decreasing homeomorphism, depending on the nature of h. Moreover, for any $x \in \mathbb{R}$ notice that

$$\begin{aligned}\pi \circ \widetilde{f}(x) &= \pi \circ \widetilde{h}^{-1} \circ \widetilde{g} \circ \widetilde{h}(x) = h^{-1} \circ \pi \circ \widetilde{g} \circ \widetilde{h}(x) \\ &= h^{-1} \circ g \circ \pi \circ \widetilde{h}(x) = h^{-1} \circ g \circ h \circ \pi(x) \\ &= f \circ \pi(x).\end{aligned}$$

Furthermore,

$$\begin{aligned}\frac{\rho(\widetilde{f})}{\rho(\widetilde{g})} &= \lim_{n\to\infty} \frac{\widetilde{f}^n(x)/n}{\widetilde{g}^n(\widetilde{h}(x))/n} = \lim_{n\to\infty} \frac{\widetilde{f}^n(x)}{\widetilde{g}^n(\widetilde{h}(x))} \\ &= \lim_{n\to\infty} \frac{\widetilde{h}^{-1}(\widetilde{g}^n \circ \widetilde{h}(x))}{\widetilde{g}^n \circ \widetilde{h}(x)} = \pm 1.\end{aligned}$$

Therefore, $\rho(\widetilde{f}) = \pm\rho(\widetilde{g}) = \pm\rho(g)$.

When h, and thus h^{-1}, $\widetilde{h}$ and $\widetilde{h}^{-1}$, is orientation preserving, we have that $\rho(\widetilde{f}) = \rho(g)$ and, hence, $\rho(f) = \rho(\widetilde{f}) \pmod 1 = \rho(g)$.

On the other hand, when h, and thus h^{-1}, $\widetilde{h}$ and $\widetilde{h}^{-1}$, is orientation reversing, we have that $\rho(f) = \rho(\widetilde{f}) \pmod 1 = -\rho(g) \pmod 1$. It hence follows that $\rho(f) + \rho(g) = 0 \pmod 1$. □

The following lemma provides a partial converse to Theorem 2.2.13. It states that the rotation number is essentially a complete invariant for rotations of the circle.

Lemma 2.2.14. *Two rotations of the circle are topologically conjugate if and only if their rotation numbers are equal or sum to zero, modulo 1.*

Proof. By Theorem 2.2.13, two topologically conjugate rotations of the unit circle have rotation numbers that are equal or whose sum is 0 (mod 1). For the converse implication, suppose that $f([x]) = [x + \alpha]$ and $g([x]) = [x + \beta]$ for some $0 < \alpha, \beta < 1$. If $\alpha = \beta$, then f is trivially topologically conjugate to g. If $\alpha + \beta = 0 \pmod 1$, then the map $h([x]) = [-x]$ is a suitable conjugacy map. Indeed,

$$\begin{aligned}h \circ f([x]) = h([x + \alpha]) &= [-x - \alpha] \\ &= [-x + \beta] = g([-x]) = g \circ h([x]).\end{aligned}$$ □

2.3 Minimality for homeomorphisms and diffeomorphisms of the circle

Our first goal in this section is to give a classification of minimal orientation-preserving homeomorphisms of the circle. We will need the following lemma.

Lemma 2.3.1. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be an orientation-preserving homeomorphism and let $\widetilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift of f. Let A and B be the sets*

$$A := \{\widetilde{f}^n(0) + m : n, m \in \mathbb{Z}\} \subseteq \mathbb{R}$$
$$B := \{n\rho(\widetilde{f}) + m : n, m \in \mathbb{Z}\} \subseteq \mathbb{R}.$$

If $\rho(f)$ is irrational, then the map $H : A \to B$ defined by

$$H(\widetilde{f}^n(0) + m) = n\rho(\widetilde{f}) + m$$

is well-defined, bijective and increasing.

Proof. The map H is the composition $H_B \circ H_A$ of the two maps

$$H_A : A \to \mathbb{Z}^2, \quad \text{where} \quad H_A(\widetilde{f}^n(0) + m) = (n, m)$$

and

$$H_B : \mathbb{Z}^2 \to B, \quad \text{where} \quad H_B(n, m) = n\rho(\widetilde{f}) + m.$$

Thus, in order to show that H is a well-defined bijection, it suffices to show that H_A and H_B are well-defined bijections. It is clear that the map H_B is well-defined and surjective. To show that it is injective, suppose that $H_B(n_1, m_1) = H_B(n_2, m_2)$, that is, $n_1\rho(\widetilde{f}) + m_1 = n_2\rho(\widetilde{f}) + m_2$. If it were the case that $n_1 \neq n_2$, then we would have

$$\rho(\widetilde{f}) = (m_2 - m_1)/(n_1 - n_2) \in \mathbb{Q},$$

which would contradict the hypothesis that $\rho(f)$ is an irrational number. Thus, $n_1 = n_2$, which implies immediately that $m_1 = m_2$. Hence, H_B is injective. Let us now consider H_A. To prove that H_A is well-defined, assume that $\widetilde{f}^{n_1}(0) + m_1 = \widetilde{f}^{n_2}(0) + m_2$. If $n_1 \neq n_2$, then

$$f^{n_1}(\pi(0)) = \pi(\widetilde{f}^{n_1}(0)) = \pi(\widetilde{f}^{n_2}(0) + m_2 - m_1) = \pi(\widetilde{f}^{n_2}(0)) = f^{n_2}(\pi(0)).$$

Applying f^{-n_2} to both sides yields $f^{n_1 - n_2}(\pi(0)) = \pi(0)$, that is, $\pi(0)$ is a periodic point of f. But, according to Proposition 2.2.12, the rotation number of f would then be a rational number. Once again, this would contradict the hypothesis that $\rho(f)$ is irrational. So $n_1 = n_2$, which implies immediately that $m_1 = m_2$. Thus, H_A is well-defined. It is easy to see that H_A is bijective.

To show that the map H is increasing, suppose that $\tilde{f}^n(0) + m < \tilde{f}^k(0) + l$. If $k = n$ then $m < l$ and obviously $n\rho(\tilde{f}) + m < k\rho(\tilde{f}) + l$. If $k < n$ then applying $\tilde{f}^{-k}$ to each side of $\tilde{f}^n(0) + m < \tilde{f}^k(0) + l$ and using Corollary 2.2.5 along with the fact that $\tilde{f}^{-k}$ is increasing, we deduce that $\tilde{f}^{n-k}(0) < l - m$. From part (f) of Proposition 2.2.10, it follows that $(n-k)\rho(\tilde{f}) \le l-m$. Since $\rho(\tilde{f})$ is irrational, this last inequality must be strict: $(n-k)\rho(\tilde{f}) < l-m$. In other words, $n\rho(\tilde{f})+m < k\rho(\tilde{f})+l$. Similarly, if $k > n$ then applying $\tilde{f}^{-n}$ to each side of $\tilde{f}^n(0)+m < \tilde{f}^k(0)+l$ yields $m-l < \tilde{f}^{k-n}(0)$. From Proposition 2.2.10(g) and the fact that $\rho(\tilde{f})$ is irrational, we conclude that $m-l < (k-n)\rho(\tilde{f})$. In other words, $n\rho(\tilde{f})+m < k\rho(\tilde{f})+l$. In each case, we have shown that $n\rho(\tilde{f})+m < k\rho(\tilde{f})+l$, and hence H is a well-defined, increasing bijection. □

Our next result is the main one of this section. It states that every minimal orientation-preserving homeomorphism of the circle is topologically conjugate to a minimal rotation.

Theorem 2.3.2. *If $f : \mathbb{S}^1 \to \mathbb{S}^1$ is a minimal orientation-preserving homeomorphism of the circle, then f is topologically conjugate to the rotation $R_{\rho(f)} : \mathbb{S}^1 \to \mathbb{S}^1$ of the unit circle by the angle corresponding to the rotation number of f.*

Proof. If f is minimal, then by Remark 1.5.5 it admits no periodic points. In view of Proposition 2.2.12, this implies that $\rho(f)$ is an irrational number. By Lemma 2.3.1, the map H introduced in that lemma is then a well-defined increasing bijection from A to B.

We aim to extend H to a homeomorphism of $\mathbb{R}$ using Lemma 2.3.1 and the fact that an increasing bijection between dense subsets of $\mathbb{R}$ can be uniquely extended to an increasing homeomorphism of $\mathbb{R}$ (see Exercise 1.7.5). Toward that end, we shall now prove that A is dense in $\mathbb{R}$. To begin, choose an arbitrary $x \in \mathbb{R}$. Since f is minimal, we know that $\pi(x)$ belongs to $\omega(\pi(0)) = \mathbb{S}^1$ by Theorem 1.5.4. Therefore, there exists a strictly increasing sequence $(n_k)_{k=1}^{\infty}$ of nonnegative integers such that

$$\pi(x) = \lim_{k\to\infty} f^{n_k}(\pi(0)) = \lim_{k\to\infty} \pi(\tilde{f}^{n_k}(0)).$$

Since $\pi(y) = \pi(z)$ means that $y - z \in \mathbb{Z}$, we have that

$$\lim_{k\to\infty} \operatorname{dist}(\tilde{f}^{n_k}(0) - x, \mathbb{Z}) = 0.$$

Thus, for each $\varepsilon > 0$ there exists $k \in \mathbb{N}$ and $l \in \mathbb{Z}$ such that

$$|\tilde{f}^{n_k}(0) - x - l| < \varepsilon.$$

Therefore, $x \in \overline{A}$. As x was chosen arbitrarily in $\mathbb{R}$, we conclude that $\overline{A} = \mathbb{R}$.

Furthermore, B is also dense because, in light of Theorem 1.5.12, the rotation $R_{\rho(f)} : \mathbb{S}^1 \to \mathbb{S}^1$ is minimal when $\rho(f)$ is an irrational number. Indeed, observe that the set B consists of the orbit of 0 under $R_{\rho(f)}$ translated everywhere by adding each integer.

Since the rotation $R_{\rho(f)}$ is minimal, the orbit of 0 under $R_{\rho(f)}$ is dense in $\mathbb{S}^1$ or, equivalently, in $[0,1]$. As it comprises all integer translations of this orbit, the set B is dense in $\mathbb{R}$.

Because A and B are dense in $\mathbb{R}$, we infer that H extends uniquely to an increasing homeomorphism $\overline{H} : \mathbb{R} \to \mathbb{R}$.

Now, let $x = \tilde{f}^n(0) + m \in A$. Note that

$$H(x+1) = H\big(\tilde{f}^n(0) + m + 1\big) = n\rho(\tilde{f}) + m + 1 = H\big(\tilde{f}^n(0) + m\big) + 1 = H(x) + 1.$$

Thus, $H(x+1) = H(x) + 1$ for all $x \in A$. By continuity, the extension $\overline{H}$ must satisfy $\overline{H}(x+1) = \overline{H}(x) + 1$ for all $x \in \mathbb{R}$. Then $h(\pi(x)) := \pi \circ \overline{H}(x)$ is a well-defined orientation-preserving homeomorphism of the circle. Moreover, for every $x = \tilde{f}^n(0) + m \in A$ we have that

$$\begin{aligned} H \circ \tilde{f}(x) &= H \circ \tilde{f}\big(\tilde{f}^n(0) + m\big) = H\big(\tilde{f}^{n+1}(0) + m\big) \\ &= (n+1)\rho(\tilde{f}) + m = \big(n\rho(\tilde{f}) + m\big) + \rho(\tilde{f}) \\ &= H\big(\tilde{f}^n(0) + m\big) + \rho(\tilde{f}) = T_{\rho(\tilde{f})} \circ H\big(\tilde{f}^n(0) + m\big) \\ &= T_{\rho(\tilde{f})} \circ H(x), \end{aligned}$$

where the map $T_{\rho(\tilde{f})} : \mathbb{R} \to \mathbb{R}$ is the translation by $\rho(\tilde{f})$ on $\mathbb{R}$. This shows that $H \circ \tilde{f}(x) = T_{\rho(\tilde{f})} \circ H(x)$ for all $x \in A$. By continuity, $\overline{H} \circ \tilde{f}(x) = T_{\rho(\tilde{f})} \circ \overline{H}(x)$ for all $x \in \mathbb{R}$. Observe also that the real translation $T_{\rho(\tilde{f})}$ is a lift of the circle rotation $R_{\rho(f)}$. It then follows that

$$\begin{aligned} h \circ f\big(\pi(x)\big) &= h \circ \pi \circ \tilde{f}(x) = \pi \circ \overline{H} \circ \tilde{f}(x) \\ &= \pi \circ T_{\rho(\tilde{f})} \circ \overline{H}(x) = R_{\rho(f)} \circ \pi \circ \overline{H}(x) \\ &= R_{\rho(f)} \circ h\big(\pi(x)\big). \end{aligned}$$

So h is a conjugacy map between f and $R_{\rho(f)}$. □

2.3.1 Denjoy's theorem

The next result is the first for which we need the map f to be a diffeomorphism, rather than merely a homeomorphism. Recall that a diffeomorphism f is a homeomorphism with the property that both f' and $(f^{-1})'$ exist. We will also need the following definition.

Definition 2.3.3. The *(total) variation* $\operatorname{var}(\varphi)$ of a function $\varphi : \mathbb{S}^1 \to \mathbb{R}$ is defined to be

$$\operatorname{var}(\varphi) := \sup\left\{ \sum_{i=0}^{n-1} \big|\varphi(x_i) - \varphi(x_{i+1})\big| : x_0, x_1, \ldots, x_n = x_0 \text{ partition } \mathbb{S}^1, n \in \mathbb{N} \right\},$$

where the supremum is taken over all finite partitions of the circle. If $\operatorname{var}(\varphi)$ is finite, then φ is said to be of *bounded variation*.

The main result of this section is named after the French mathematician, Arnaud Denjoy (1884–1974). Denjoy made outstanding contributions to many areas of mathematics, in particular to the theory of functions of a real variable.

Theorem 2.3.4 (Denjoy's theorem). *Suppose that $f : \mathbb{S}^1 \to \mathbb{S}^1$ is an orientation-preserving C^1 diffeomorphism with derivative f' of bounded variation. If $\rho(f)$ is irrational, then $f : \mathbb{S}^1 \to \mathbb{S}^1$ is minimal.*

Before beginning the proof of Denjoy's theorem, we first establish three lemmas, which will be useful in the proof. In the remainder of this section, we adopt the usual convention that arcs of the unit circle shall be traversed in the counterclockwise direction. For instance, (x, y) is the open arc of the circle generated when moving from x to y along the circle in the counterclockwise direction. Note also that since f is orientation preserving, it holds that $f((x, y)) = (f(x), f(y))$.

Lemma 2.3.5. *Assume that $x_0 \in \mathbb{S}^1$ is such that for some $n \in \mathbb{N}$,*

$$(x_0, f^n(x_0)) \cap \{f^j(x_0) : |j| \le n\} = \emptyset. \tag{2.2}$$

Then, for all $0 \le k \le n$,

$$(f^{k-n}(x_0), f^k(x_0)) \cap \{f^j(x_0) : |j| \le n\} = \emptyset.$$

Proof. Assume that x_0 is as stated above and, by way of contradiction, suppose that there exist $0 \le k \le n$ and $|j| \le n$ such that

$$f^j(x_0) \in (f^{k-n}(x_0), f^k(x_0)). \tag{2.3}$$

Fix the largest k with this property. Of course, $j \neq k$. We shall examine two potential cases.

Case 1: $j \le 0$.

If it turned out that $j \le 0$, then (2.3) and the fact that f preserves orientation would result in

$$f^{j+1}(x_0) \in f((f^{k-n}(x_0), f^k(x_0))) = (f^{k+1-n}(x_0), f^{k+1}(x_0)).$$

Since $-n \le j + 1 \le n$, it follows that (2.3) would also be satisfied with $k + 1$ in place of k. Given that $k \le n$ was chosen to be the maximal number satisfying this property, the only way that this could be true is if $k + 1 > n$, that is, if $k = n$. Hence, (2.3) would reduce to $f^j(x_0) \in (x_0, f^n(x_0))$, which would contradict our original hypothesis (2.2). So this case never takes place.

Case 2: $j > 0$. This case is divided into two subcases, which are illustrated in Figure 2.1.

Note that for any given $x \in \mathbb{S}^1$, we have

$$f^r(x) \neq f^s(x), \quad \forall r, s \in \mathbb{Z}, \ r \neq s. \tag{2.4}$$

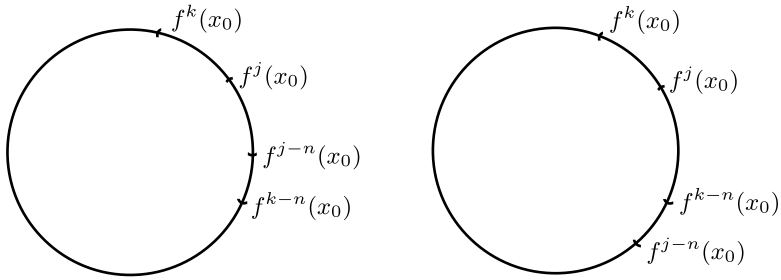

Figure 2.1: On the left, Subcase 2.1: $(f^{j-n}(x_0), f^j(x_0)) \subseteq (f^{k-n}(x_0), f^k(x_0))$. On the right, Subcase 2.2: $f^{k-n}(x_0) \in (f^{j-n}(x_0), f^j(x_0))$.

Otherwise, $f^{r-s}(x) = x$ and f would have a periodic point, that is, $\rho(f)$ would be rational according to Proposition 2.2.12. This would contradict our hypothesis that $\rho(f)$ is irrational.

Subcase 2.1: $f^{j-n}(x_0) \in (f^{k-n}(x_0), f^j(x_0))$.

This means that $(f^{j-n}(x_0), f^j(x_0)) \subseteq (f^{k-n}(x_0), f^j(x_0))$. In light of assumption (2.3), which we recall states that $f^j(x_0) \in (f^{k-n}(x_0), f^k(x_0))$, we actually have that

$$(f^{j-n}(x_0), f^j(x_0)) \subseteq (f^{k-n}(x_0), f^k(x_0)).$$

Then the continuity and orientation-preserving properties of f yield that

$$f^{j-k}([f^{k-n}(x_0), f^k(x_0)]) = [f^{j-n}(x_0), f^j(x_0)] \subseteq [f^{k-n}(x_0), f^k(x_0)].$$

So f^{j-k} maps a closed arc to a closed arc within itself. The intermediate value theorem then asserts that f^{j-k} has a fixed point (recall that $j \neq k$). Hence, f has a periodic point. According to Proposition 2.2.12, this means that $\rho(f)$ is a rational number. This contradicts our hypothesis that $\rho(f)$ is irrational, and thus this subcase cannot occur.

Subcase 2.2: $f^{j-n}(x_0) \notin (f^{k-n}(x_0), f^j(x_0))$.

In other words, $f^{j-n}(x_0) \in [f^j(x_0), f^{k-n}(x_0)]$. Since $f^r(x_0) \neq f^s(x_0)$ for all $r \neq s$, this actually means that $f^{j-n}(x_0) \in (f^j(x_0), f^{k-n}(x_0))$. Equivalently, this means that

$$f^{k-n}(x_0) \in (f^{j-n}(x_0), f^j(x_0)).$$

Then, as $-n \le k - n \le n$, we have a relation akin to (2.3) but with $k - n$ in place of j and j in place of k. But since k is maximal with this property, we deduce that $j \le k$. In fact, as $f^k(x_0) \notin (f^{k-n}(x_0), f^k(x_0))$, we know that $j < k$. Then, by (2.3), we obtain that

$$f^{n+j-k}(x_0) = f^{n-k}(f^j(x_0)) \in f^{n-k}((f^{k-n}(x_0), f^k(x_0))) = (x_0, f^n(x_0)).$$

Since $-n \le n + j - k \le n$, this contradicts our original hypothesis (2.2). This shows that this subcase does not happen either.

To summarize, neither Case 1 nor Case 2 can occur. So, whatever $0 \le k \le n$ might be, there is no $-n \le j \le n$ satisfying (2.3). This contradiction completes the proof. □

A rather straightforward consequence of Lemma 2.3.5 is the following.

Lemma 2.3.6. *If $x_0 \in \mathbb{S}^1$ is such that*

$$(x_0, f^n(x_0)) \cap \{f^j(x_0) : |j| \le n\} = \emptyset$$

for some $n \in \mathbb{N}$, then the arcs $\{(f^{k-n}(x_0), f^k(x_0)) : 0 \le k \le n\}$ are mutually disjoint.

Proof. If two such arcs intersected, then an endpoint of one of those arcs would belong to the other arc, which would contradict the previous lemma. □

Finally, we derive an estimate on the derivatives of the iterates of f.

Lemma 2.3.7. *There is a universal constant $v > 0$ with the property that for all $x_0 \in \mathbb{S}^1$ and $n \in \mathbb{N}$ such that*

$$(x_0, f^n(x_0)) \cap \{f^j(x_0) : |j| \le n\} = \emptyset,$$

we have

$$(f^n)'(x_0)(f^{-n})'(x_0) \ge e^{-v}.$$

Proof. Let $x_0 \in \mathbb{S}^1$ and $n \in \mathbb{N}$ be as stated above. Since f preserves orientation, both $(f^n)'(x)$ and $(f^{-n})'(x)$ are strictly positive for all $x \in \mathbb{S}^1$ and all $n \ge 0$. Let

$$a := \inf\{f'(x) \mid x \in \mathbb{S}^1\}.$$

As f is a C^1 function, its derivative f' is continuous, and hence its infimum on the compact set $\mathbb{S}^1$ is achieved. So $a > 0$.

By a simple application of the chain rule, we obtain that

$$\log(f^n)'(x_0) = \log\left(\prod_{k=0}^{n-1} f'(f^k(x_0))\right) = \sum_{k=0}^{n-1} \log f'(f^k(x_0))$$

and

$$\begin{aligned}
\log(f^{-n})'(x_0) &= \log\big((f^n)'(f^{-n}(x_0))\big)^{-1} \\
&= -\log(f^n)'(f^{-n}(x_0)) \\
&= -\log\left(\prod_{k=0}^{n-1} f'(f^k(f^{-n}(x_0)))\right) \\
&= -\sum_{k=0}^{n-1} \log f'(f^{k-n}(x_0)).
\end{aligned}$$

These two equalities yield the estimate

$$\begin{aligned}\log\Big((f^n)'(x_0)(f^{-n})'(x_0)\Big) &= \log(f^n)'(x_0) + \log(f^{-n})'(x_0)\\ &= \sum_{k=0}^{n-1}\Big(\log f'(f^k(x_0)) - \log f'(f^{k-n}(x_0))\Big)\\ &\geq -\sum_{k=0}^{n-1}\Big|\log f'(f^k(x_0)) - \log f'(f^{k-n}(x_0))\Big|.\end{aligned}$$

But, by the mean value theorem,

$$\Big|\log f'(f^k(x_0)) - \log f'(f^{k-n}(x_0))\Big| = \frac{1}{c_k}\Big|f'(f^k(x_0)) - f'(f^{k-n}(x_0))\Big|,$$

where c_k is between $f'(f^k(x_0))$ and $f'(f^{k-n}(x_0))$. In particular, we have that $c_k \geq a$. Moreover, according to Lemma 2.3.6, the family of arcs $\{(f^{k-n}(x_0), f^k(x_0)) : 0 \leq k \leq n\}$ are mutually disjoint. Consequently, the points $\{f^j(x_0) : |j| \leq n\}$ can be arranged in such a way that they form an ordered partition of $\mathbb{S}^1$ in which $f^k(x_0)$ immediately follows $f^{k-n}(x_0)$ for each $0 \leq k \leq n$. Hence,

$$\sum_{k=0}^{n-1}\Big|f'(f^k(x_0)) - f'(f^{k-n}(x_0))\Big| \leq \operatorname{var}(f').$$

Therefore,

$$\begin{aligned}\sum_{k=0}^{n-1}\Big|\log f'(f^k(x_0)) - \log f'(f^{k-n}(x_0))\Big| &= \sum_{k=0}^{n-1}\frac{1}{c_k}\Big|f'(f^k(x_0)) - f'(f^{k-n}(x_0))\Big|\\ &\leq \frac{1}{a}\sum_{k=0}^{n-1}\Big|f'(f^k(x_0)) - f'(f^{k-n}(x_0))\Big|\\ &\leq \frac{1}{a}\operatorname{var}(f') =: v < \infty.\end{aligned}$$

Then we obtain that

$$\log\Big((f^n)'(x_0)(f^{-n})'(x_0)\Big) \geq -\sum_{k=0}^{n-1}\Big|\log f'(f^k(x_0)) - \log f'(f^{k-n}(x_0))\Big| \geq -v.$$

Hence, $(f^n)'(x_0)(f^{-n})'(x_0) \geq e^{-v}$. Note that v depends only on f. □

We are now in a position to prove Denjoy's theorem.

Proof of Denjoy's theorem. Suppose, by way of contradiction, that the map $f : \mathbb{S}^1 \to \mathbb{S}^1$ is not minimal. According to Theorem 1.5.2, every dynamical system has a minimal set. Call this set $\triangle \subseteq \mathbb{S}^1$. Since we are assuming that f is not minimal, it follows that $\triangle$ is a proper subset of $\mathbb{S}^1$. Recall that $\triangle$ is a closed f-invariant set. Since f is a homeomorphism, not only is $f(\triangle) \subseteq \triangle$ but, in fact, $f(\triangle) = \triangle = f^{-1}(\triangle)$. That is, $\triangle$ is a completely invariant set. Therefore,

$$f(\mathbb{S}^1 \setminus \triangle) = \mathbb{S}^1 \setminus \triangle = f^{-1}(\mathbb{S}^1 \setminus \triangle).$$

Because $\mathbb{S}^1 \setminus \triangle$ is an open subset of the circle, we can write $\mathbb{S}^1 \setminus \triangle = \bigcup_{j=0}^{\infty} I_j$, where the I_j's form a countable union of maximal disjoint open arcs. This implies in particular that $\partial I_j \subseteq \triangle$ for each $j \geq 0$. We then have

$$\sum_{j=0}^{\infty} |I_j| = \mathrm{Leb}(\mathbb{S}^1 \setminus \triangle) \leq 1,$$

where Leb denotes the Lebesgue measure. Hence, $\lim_{j\to\infty} |I_j| = 0$.

We now establish two claims about the I_j's.

Claim 1. *For every $n \in \mathbb{Z}$, there is a unique $j_n \geq 0$ such that $f^n(I_0) \cap I_{j_n} \neq \emptyset$. In fact, $f^n(I_0) = I_{j_n}$.*

By definition, the I_j's are the connected components of $\mathbb{S}^1 \setminus \triangle$. Let n be an integer. Since $\mathbb{S}^1 \setminus \triangle$ is completely f-invariant and f is a homeomorphism, the set $f^n(I_0)$ is a connected component of $\mathbb{S}^1 \setminus \triangle$. Denote by I_{j_n} this unique component. This proves Claim 1.

Claim 2. *If $m \neq k$, then $I_{j_m} \cap I_{j_k} = \emptyset$. Moreover, $\lim_{|n|\to\infty} |I_{j_n}| = 0$.*

Suppose for a contradiction that $I_{j_m} \cap I_{j_k} \neq \emptyset$ for some $k < m$. Since these arcs are connected components of $\mathbb{S}^1 \setminus \triangle$, we have that $I_{j_m} = I_{j_k}$. Then

$$f^{m-k}(I_{j_k}) = f^{m-k}(f^k(I_0)) = f^m(I_0) = I_{j_m} = I_{j_k}.$$

Therefore, $f^{m-k}(\overline{I_{j_k}}) = \overline{I_{j_k}}$. In other words, f^{m-k} maps a closed arc within itself. Thus, f^{m-k} has a fixed point, and hence f has a periodic point. According to Proposition 2.2.12, this means that $\rho(f)$ is a rational number, which contradicts our original assumption that $\rho(f)$ is irrational. Consequently, the family $(I_{j_n})_{n=0}^{\infty}$ consists of mutually disjoint arcs, that is, all the j_n's differ and so there are infinitely many arcs I_{j_n}. It therefore follows that

$$\lim_{|n|\to\infty} |I_{j_n}| = 0.$$

This completes the proof of Claim 2.

Now, for each $n \geq 0$ consider the sets

$$\mathcal{J}_+^n := \{x \in I_0 : (f^n)'(x) \geq e^{-v/2}\} \quad \text{and} \quad \mathcal{J}_-^n := \{x \in I_0 : (f^{-n})'(x) \geq e^{-v/2}\},$$

where $v > 0$ is the universal constant arising from Lemma 2.3.7. We aim to show that there exists a strictly increasing sequence $(n_q)_{q=0}^{\infty}$ of nonnegative integers such that

$$I_0 = \mathcal{J}_+^{n_q} \cup \mathcal{J}_-^{n_q}, \quad \forall q \geq 0.$$

According to Lemma 2.3.7, it suffices to show that

$$\left(x_0, f^{n_q}(x_0)\right) \cap \left\{f^j(x_0) : |j| \leq n_q\right\} = \emptyset, \quad \forall x_0 \in I_0, \ \forall q \geq 0.$$

To prove this, write $I_0 := (a, b)$. We first construct a strictly increasing sequence of nonnegative integers $(p_q)_{q=0}^{\infty}$ and a sequence of integers $(m_{p_q})_{q=0}^{\infty}$, where $|m_{p_q}| \leq p_q$ and $\lim_{q\to\infty} |m_{p_q}| = \infty$, so that

$$\left(b, f^{m_{p_q}}(b)\right) \cap \left\{f^j(b) : |j| \leq p_q\right\} = \emptyset, \quad \forall q \geq 0. \tag{2.5}$$

Since $b \in \triangle$ and $\triangle$ is minimal, we know that $\omega(b) = \triangle$ by Theorem 1.5.4. Therefore, there is a strictly increasing sequence $(p_q)_{q=0}^{\infty}$ of nonnegative integers such that $\lim_{q\to\infty} f^{p_q}(b) = b$.

Note that

$$f^j(b) \neq f^k(b), \quad \forall j \neq k. \tag{2.6}$$

Otherwise, $f^{j-k}(b) = b$ for some $j > k$ and f would have a periodic point, that is, $\rho(f)$ would be rational according to Proposition 2.2.12. This would contradict the fact that $\rho(f)$ is irrational. In particular, this means that $f^{p_q}(b) \neq b$ for every q. By passing to a subsequence if necessary, we may thus assume that the points of the sequence $(f^{p_q}(b))_{q=0}^{\infty}$ successively edge closer to b.

For every q, denote by $f^{m_{p_q}}(b)$ the point among the iterates

$$f^{-p_q}(b), f^{-p_q+1}(b), \ldots, f^{-1}(b), f(b), \ldots, f^{p_q-1}(b), f^{p_q}(b)$$

(excluding b) which is closest to the point b. Clearly, $f^{m_{p_q}}(b) \to b$ since $f^{p_q}(b) \to b$ and the points of the sequence $(f^{m_{p_q}}(b))_{q=0}^{\infty}$ successively edge closer to b since the $f^{p_q}(b)$'s do. Thus, the sequence $(m_{p_q})_{q=0}^{\infty}$ accumulates to ∞ or $-\infty$, and hence admits a subsequence converging to ∞ or $-\infty$. By replacing f with f^{-1} if necessary, we may assume without loss of generality that the subsequence in question is positive and converges to ∞. Let us denote that subsequence by the same notation $(m_{p_q})_{q=0}^{\infty}$. Then, by construction of the $f^{m_{p_q}}(b)$'s, relation (2.5) holds. In other words,

$$\left\{f^j(b) : |j| \leq p_q\right\} \subseteq \left[f^{m_{p_q}}(b), b\right], \quad \forall q \geq 0. \tag{2.7}$$

In fact, by (2.6), we have

$$\left\{f^j(b) : |j| \leq p_q\right\} \subseteq \left(f^{m_{p_q}}(b), b\right), \quad \forall q \geq 0. \tag{2.8}$$

Moreover,

$$f^j(b) \in (b,a], \quad \forall j \neq 0. \tag{2.9}$$

Otherwise, $f^j(b) \in (a,b)$ would imply that $f^j((a,b)) \cap (a,b) \neq \emptyset$, which would force j to be equal to zero by Claims 1 and 2.

In particular, this implies that $f^{m_{p_q}}(b) \in (b,a]$. Hence,

$$\big(f^{m_{p_q}}(b), b\big) = (f^{m_{p_q}}(b), a] \cup (a,b), \quad \forall q \geq 0. \tag{2.10}$$

We deduce from (2.8), (2.9) and (2.10) that

$$\{f^j(b) : |j| \leq p_q\} \subseteq \big(f^{m_{p_q}}(b), b\big) \cap (b,a] = (f^{m_{p_q}}(b), a], \quad \forall q \geq 0. \tag{2.11}$$

Finally, by dropping the first few terms of the subsequence $(m_{p_q})_{q=0}^{\infty}$ if necessary, we can also guarantee that all $f^{m_{p_q}}(b)$'s are closer to b than the point a is.

Now, let $x_0 \in (a,b)$. First, note that $f^j(x_0) \notin (a,b)$ for all $j \neq 0$; otherwise, we would have that $f^j((a,b)) \cap (a,b) \neq \emptyset$, and Claims 1 and 2 would force j to be equal to zero. In particular, this implies that

$$(x_0, b) \cap \{f^j(x_0) : |j| \leq m_{p_q}\} \subseteq (x_0, b) \cap \{f^j(x_0) : j \in \mathbb{Z}\} = \emptyset. \tag{2.12}$$

Observe also that $f^j(x_0) \neq b$ for all j since $x_0 \in \mathbb{S}^1 \setminus \triangle$, the set $\mathbb{S}^1 \setminus \triangle$ is completely f-invariant and $b \in \triangle$. In particular, we obtain

$$\{b\} \cap \{f^j(x_0) : |j| \leq m_{p_q}\} \subseteq \{b\} \cap \{f^j(x_0) : j \in \mathbb{Z}\} = \emptyset. \tag{2.13}$$

Finally, we show that $f^{m_{p_q}}(x_0)$ is the point which is closest to b among the points $f^{-p_q}(x_0), f^{-p_q+1}(x_0), \ldots, f^{-1}(x_0), x_0, f(x_0), \ldots, f^{p_q-1}(x_0), f^{p_q}(x_0)$, except possibly x_0 itself. Let $q \geq 0$. Suppose that there exists j with $-p_q \leq j \leq p_q$, $j \neq m_{p_q}$, such that $f^j(x_0)$ is closer to b than $f^{m_{p_q}}(x_0)$. We already know that $f^j(x_0) \neq b$ and $f^j(x_0) \neq a$ by the complete f-invariance of $\mathbb{S}^1 \setminus \triangle$. On one hand, if $f^j(x_0) \in (a,b)$ then $f^j((a,b)) \cap (a,b) \neq \emptyset$. Claims 1 and 2 impose that j equal zero. Then $f^j(x_0) = x_0$. On the other hand, it might turn out that $f^j(x_0) \in (b,a)$. In this case, observe that

$$\begin{aligned} f^{m_{p_q}}(x_0) \in f^{m_{p_q}}\big((a,b)\big) = f^{m_{p_q}}(I_0) = I_{j_{m_{p_q}}} &= \big(f^{m_{p_q}}(a), f^{m_{p_q}}(b)\big) \\ &= (b,a) \cap \big(f^{m_{p_q}}(a), f^{m_{p_q}}(b)\big) \\ &\subseteq \big(b, f^{m_{p_q}}(b)\big). \end{aligned}$$

Since $f^{m_{p_q}}(b)$ is closer to b than the point a is and since $f^j(x_0)$ is assumed to be closer to b than $f^{m_{p_q}}(x_0)$, we would then have $f^j(x_0) \in (b, f^{m_{p_q}}(x_0))$. As $f^j(b) \in (f^{m_{p_q}}(b), a]$ according to (2.11), we would then deduce that $[f^{m_{p_q}}(x_0), f^{m_{p_q}}(b)] \subseteq [f^j(x_0), f^j(b)]$. As f is orientation preserving, this would imply that

$$f^{m_{p_q}-j}\big([x_0, b]\big) = f^{-j}\big([f^{m_{p_q}}(x_0), f^{m_{p_q}}(b)]\big) \subseteq f^{-j}\big([f^j(x_0), f^j(b)]\big) = [x_0, b].$$

Then f would have a periodic point, which is not the case. To summarize, $f^{m_{p_q}}(x_0)$ is indeed the point among $f^{-p_q}(x_0), \dots, f^{p_q}(x_0)$, which is closest to the point b, except possibly x_0 itself. As noted previously, $f^{m_{p_q}}(x_0) \in (b, a)$ for every q. This implies that

$$\big(b, f^{m_{p_q}}(x_0)\big) \cap \{f^j(x_0) : |j| \le m_{p_q}\} = \emptyset. \tag{2.14}$$

From (2.12), (2.13), and (2.14), we conclude that

$$\big(x_0, f^{m_{p_q}}(x_0)\big) \cap \{f^j(x_0) : |j| \le m_{p_q}\} = \emptyset.$$

Note that the sequence $(m_{p_q})_{q=0}^{\infty}$ is independent of $x_0 \in (a, b) = I_0$. Setting $n_q := m_{p_q}$, an application of Lemma 2.3.7 with $n = n_q$ allows us to deduce that $(f^{n_q})'(x_0)\,(f^{-n_q})'(x_0) \ge e^{-v}$ for all $q \ge 0$ and all $x_0 \in I_0$. This implies that $I_0 = \mathcal{J}_+^{n_q} \cup \mathcal{J}_-^{n_q}$, and hence that $\max\{\lambda(\mathcal{J}_+^{n_q}), \lambda(\mathcal{J}_-^{n_q})\} \ge |I_0|/2$ for all $q \ge 0$, where $\lambda = \mathrm{Leb}$ denotes the Lebesgue measure on $\mathbb{S}^1$. If $\lambda(\mathcal{J}_+^{n_q}) \ge |I_0|/2$, then

$$\begin{aligned}
|I_{j_{n_q}}| = \big|f^{n_q}(I_0)\big| &= \int_{I_0} (f^{n_q})'(x)\,dx \\
&\ge \int_{\mathcal{J}_+^{n_q}} (f^{n_q})'(x)\,d\lambda(x) \ge \int_{\mathcal{J}_+^{n_q}} e^{-v/2}\,d\lambda(x) \\
&= e^{-v/2}\lambda(\mathcal{J}_+^{n_q}) \ge \frac{e^{-v/2}}{2}|I_0|.
\end{aligned}$$

A similar argument yields the same conclusion if $\lambda(\mathcal{J}_-^{n_q}) \ge |I_0|/2$. Thus, $|I_{j_{n_q}}| \ge e^{-v/2}|I_0|/2$ for each q. This contradicts the fact that $\lim_{n\to\infty} |I_{j_n}| = 0$. Therefore, the minimal set $\triangle$ must be $\mathbb{S}^1$, which means that $f : \mathbb{S}^1 \to \mathbb{S}^1$ is minimal. □

As an immediate consequence of Denjoy's theorem, we obtain the following corollary.

Corollary 2.3.8. *Suppose that $f : \mathbb{S}^1 \to \mathbb{S}^1$ is an orientation-preserving C^1 diffeomorphism with derivative f' of bounded variation. If $\rho(f)$ is irrational, then f is topologically conjugate to the rotation around the circle by the angle $\rho(f)$.*

Proof. This follows directly from Theorems 2.3.2 and 2.3.4. □

Remark 2.3.9. Notice that if $\varphi : \mathbb{S}^1 \to \mathbb{R}$ is Lipschitz continuous, then φ is of bounded variation. Indeed, for any finite partition $x_0, x_1, \dots, x_n = x_0$ of the unit circle we have

$$\sum_{i=0}^{n-1} \big|\varphi(x_{i+1}) - \varphi(x_i)\big| \le \sum_{i=0}^{n-1} L|x_{i+1} - x_i| = L,$$

where L is any Lipschitz constant for φ.

As every C^1 function on $\mathbb{S}^1$ is Lipschitz continuous, every C^2 function f has a derivative f' which is C^1, and hence of bounded variation. Thus the previous corollary yields the following further result.

Corollary 2.3.10. *Suppose that $f : \mathbb{S}^1 \to \mathbb{S}^1$ is an orientation-preserving C^2 diffeomorphism. If $\rho(f)$ is irrational, then f is topologically conjugate to the rotation around the circle by the angle $\rho(f)$.*

2.3.2 Denjoy's counterexample

In light of Corollary 2.3.8, it is natural to ask whether all orientation-preserving homeomorphisms with irrational rotation numbers are topologically conjugate to a rotation around the circle by an irrational angle. This is not the case. In fact, there are even orientation-preserving C^1 diffeomorphisms with irrational rotation numbers which are not topologically conjugate to an irrational rotation.

We will now construct an orientation-preserving homeomorphism with irrational rotation number which is not topologically conjugate to an irrational rotation of the circle. This construction is also due to Denjoy. The idea is the following. We know that minimality is a topological invariant. Given that an irrational rotation of the circle is minimal, it suffices to construct an orientation-preserving homeomorphism with irrational rotation number which is not minimal. By Theorem 1.5.4, this reduces to devising an orientation-preserving homeomorphism with irrational rotation number which admits a nondense orbit. We will build such a map by performing a "surgery" on an irrational rotation of the circle. Let $R_\rho : \mathbb{S}^1 \to \mathbb{S}^1$ be an irrational rotation of the circle. Choose arbitrarily $\theta \in \mathbb{S}^1$. Cut the unit circle at the point θ, open it up and insert into the gap an arc I_0. Similarly, cut the circle at the point $R_\rho(\theta)$, open it up and insert into the gap an arc I_1. Perform a similar procedure at every point of the orbit of θ. That is, for every $n \in \mathbb{Z}$ cut the circle at the point $R_\rho^n(\theta)$, open it up and insert into the gap an arc I_n. Make sure to choose the arcs $\{I_n\}$ small enough that $\sum_{n=-\infty}^{\infty} |I_n| < \infty$. These surgeries result in a larger "circle" (more precisely, a simple closed curve). It only remains to extend the original irrational rotation to the larger curve by defining the extension on the union of the I_n's. Since the extension is required to preserve orientation, choose any orientation-preserving homeomorphism h_n mapping I_n to I_{n+1}. The extension is then an orientation-preserving homeomorphism of the larger circle. Note that the extension does not admit any periodic point, since the original map did not and the points in the inserted arcs visit all those arcs in succession without ever coming back to the same arc. Proposition 2.2.12 therefore allows us to infer that the extension must have an irrational rotation number. Moreover, the fact that the interior points of the inserted arcs never come back to their original arc under iteration shows that all these points have nondense orbits.

In fact, the above argument can be modified in such a way that the extension is a C^1 diffeomorphism whose derivative is not of bounded variation (see pp. 111–112 of [19]).

2.4 Exercises

Exercise 2.4.1. In this exercise, you shall prove Corollary 2.2.6. We suggest that you proceed as follows. Let $\widetilde{f}$ be a lift of f.

(a) Using Corollary 2.2.5 and the fact that $\widetilde{f}$ is increasing, show that if $|x - y| \le k$ for some $k \in \mathbb{Z}_+$, then $|\widetilde{f}(x) - \widetilde{f}(y)| \le k$.

(b) Deduce that if $|x - y| \le k$ for some $k \in \mathbb{Z}_+$, then $|\widetilde{f}^n(x) - \widetilde{f}^n(y)| \le k$ for any $n \in \mathbb{Z}$.

(c) Prove that we can replace $\le$ by $<$ above.

Exercise 2.4.2. Prove Corollary 2.2.8.

Exercise 2.4.3. Prove that any homeomorphism $F : \mathbb{R} \to \mathbb{R}$ with the property that $F(x + 1) = F(x) + 1$ for all $x \in \mathbb{R}$ generates an orientation-preserving homeomorphism $f : S^1 \to \mathbb{S}^1$.

Exercise 2.4.4. Taking inspiration from the proof of statement (f) in Proposition 2.2.10, prove statement (g) of the same proposition.

Exercise 2.4.5. This exercise is concerned with orientation-reversing homeomorphisms of the circle. You will be asked to prove several properties of lifts of such maps. In the end, you will discover that the concept of “rotation” number is useless for orientation-reversing homeomorphisms.

Let $\widetilde{f}$ be a lift of an orientation-reversing homeomorphism f. Prove the following statements:

(a) Show that any lift $\widetilde{f}$ is a decreasing homeomorphism of $\mathbb{R}$ (cf. Lemma 2.2.3).

(b) Prove that $\deg(f) = -1$ (cf. Lemma 2.2.4).

(c) Show that $\widetilde{f}^n(x + k) = \widetilde{f}^n(x) + (-1)^n k$ for all $x \in \mathbb{R}$, all $k \in \mathbb{Z}$, and all $n \in \mathbb{Z}$ (cf. Corollaries 2.1.10 and 2.2.5).

(d) Prove that if $|x - y| < k$ for some $k \in \mathbb{Z}_+$, then $|\widetilde{f}^n(x) - \widetilde{f}^n(y)| < k$ for any $n \in \mathbb{N}$. (This shows that Corollary 2.2.6 still holds.)

(e) If $\widetilde{g}$ is another lift of f so that $\widetilde{g} = \widetilde{f} + k$ for some $k \in \mathbb{Z}$, then $\widetilde{g}^n = \widetilde{f}^n + k \sin^2(n\pi/2)$ for all $n \in \mathbb{Z}$ (cf. Corollary 2.2.7).

(f) Show that $\widetilde{f} + \mathrm{Id}_{\mathbb{R}}$ is a periodic function with period 1. More generally, a decreasing homeomorphism $\widetilde{g} : \mathbb{R} \to \mathbb{R}$ is a lift of an orientation-reversing homeomorphism of the circle if and only if $\widetilde{g} + \mathrm{Id}_{\mathbb{R}}$ is a periodic function with period 1 (cf. Corollary 2.2.8).

(g) Show that the reflection of the unit circle in the x-axis is an orientation-reversing homeomorphism of $\mathbb{S}^1$. Then find its lifts.

(h) Prove that the number $\rho(\widetilde{f}) := \lim_{n\to\infty} \frac{\widetilde{f}^n(x)}{n}$ exists for all $x \in \mathbb{R}$ and is independent of x (cf. Proposition 2.2.10).

(i) Show that $\rho(\widetilde{f}) = 0$. This demonstrates that the concept of "rotation" number is useless for orientation-reversing homeomorphisms.

Exercise 2.4.6. Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be an orientation-preserving homeomorphism of the circle. Let $\varepsilon > 0$. Show that there exists $\delta > 0$ such that if $g : \mathbb{S}^1 \to \mathbb{S}^1$ is an orientation-preserving homeomorphism which is C^0-δ close to f, then

$$\left|\rho(g) - \rho(f)\right| < \varepsilon.$$

Hint: Reread the proof of statement (a) *in Proposition* 2.2.10.

Exercise 2.4.7. Suppose that $f, g : \mathbb{S}^1 \to \mathbb{S}^1$ are two orientation-preserving homeomorphisms of the circle. Show that if f and g commute, then $\rho(g \circ f) = \rho(g) + \rho(f)$.

Exercise 2.4.8. Suppose that $f : \mathbb{S}^1 \to \mathbb{S}^1$ is an orientation-preserving homeomorphism of the circle such that $\rho(f) \in \mathbb{R} \setminus \mathbb{Q}$. Show that $\{\rho(f^n) : n \in \mathbb{N}\}$ is dense in $\mathbb{S}^1$.

3 Symbolic dynamics

Symbolic dynamics is an extremely powerful tool for the analysis of general dynamical systems. The very rough idea is to break up a space into finitely or countably many parts, assign a symbol to each part and track the orbits of points by assigning sequences of symbols to them, representing the orbits visiting successive parts of the space. In the process, we glean information about the system by analyzing these symbolic orbits.

In this chapter, we exclusively deal with topological aspects of symbolic dynamics. Symbolic dynamics is however equally important, perhaps even more important, in the context of measure-preserving dynamical systems and ergodic theory, particularly thermodynamic formalism. We will see why in Chapters 4, 8 (especially Subsections 8.1.1 and 8.2.3) as well as in Chapters 13 and 17 onward in the second volume.

The first successful use of topological aspects of symbolic dynamics can be credited to Hadamard [26], who applied them to geodesic flows. However, it took another 40 years before the topic received its first systematic account and its name, in the foundational paper by Morse and Hedlund [51]. This paper is the first to treat symbolic dynamical systems as objects of study in and of themselves.

Since then, symbolic dynamics has found ever wider applications within dynamical systems as a whole, while still remaining an active area of research. For a deeper introduction to combinatorial and topological aspects of symbolic dynamics over finite alphabets, we refer the reader to Lind and Marcus [43]. There is also a nice chapter on symbolic dynamics over finite alphabets in the fairly recent book by Brin and Stuck [13]. For a treatment of topological symbolic dynamics over countable alphabets, we refer the reader to Kitchens [39].

In this chapter, we discuss symbolic dynamical systems as objects in their own right, but later on (in Chapter 4, among others mentioned above), we will apply the ideas developed here to more general systems. We restrict ourselves to finitely many letters, as symbolic systems over finite alphabets act on compact metrizable spaces. However, in Chapter 17, we will turn our attention to countable-alphabet symbolic dynamics.

In Section 3.1, we discuss full shifts. In Section 3.2, we talk about subshifts of finite type and in particular the characterizations of topological transitivity and exactness for such systems. Finally, in Section 3.3 we examine general subshifts of finite type.

3.1 Full shifts

Let us begin by introducing the simplest class of symbolic dynamical systems, namely the full shifts.

https://doi.org/10.1515/9783110702682-003

Definition 3.1.1. Let E be a set such that $2 \leq \#E < \infty$, where $\#E$ denotes the cardinality of E. This set will henceforth be referred to as an *alphabet*. The elements of E will be called *letters* or *symbols*.

(a) For each $n \in \mathbb{N}$, we shall denote by E^n the set of all *words* (also called *blocks*) comprising n letters from the alphabet E. For convenience, we also denote the *empty word*, that is, the word having no letters, by ϵ.
For instance, if $E = \{0, 1\}$ then $E^1 = E = \{0, 1\}$, $E^2 = \{00, 01, 10, 11\}$ and

$$E^3 = \{000, 100, 010, 001, 110, 101, 011, 111\}.$$

(b) We will denote by $E^* := \bigcup_{n=1}^{\infty} E^n$ the set of all finite nonempty words over the alphabet E.

(c) The set $E^\infty := E^{\mathbb{N}}$ of all one-sided infinite words over the alphabet E, that is, the set of all sequences or functions from $\mathbb{N}$ to E, is called the *one-sided full E-shift*, or, if no confusion about the underlying alphabet may arise, simply the full shift. In other words,

$$E^\infty := \{\omega = (\omega_j)_{j=1}^{\infty} : \omega_j \in E, \ \forall j \in \mathbb{N}\}.$$

When $E = \{0, 1, \ldots, n-1\}$ for some $n \in \mathbb{N}$, the full E-shift is usually referred to as the *full n-shift*.

(d) The *length* of a word $\omega \in E^* \cup E^\infty$ is defined in the natural way to be the number of letters that it consists of and is denoted by $|\omega|$. That is, $|\omega|$ is the unique $n \in \mathbb{N} \cup \{\infty\}$ such that $\omega \in E^n$. By convention, $|\epsilon| = 0$.
A word of length n is sometimes called an n-word or n-block. In our notation, the set of all n-words is simply E^n.

Note that for each $n \in \mathbb{N}$ the set E^n is finite, and hence the set E^* of finite words is countable. We can enumerate all the finite words by starting with the 1-words, followed by the 2-words, the 3-words, and so on. As $\#E \geq 2$, a classical Cantor diagonalization argument establishes that the full E-shift is uncountable. More precisely, the full E-shift has the cardinality $\mathfrak{c}$ of the continuum, that is, it is equinumerous to $\mathbb{R}$ and $[0, 1]$.

One of the most common examples of a full shift is the full 10-shift, which can serve to encode the decimal expansions of the real numbers between 0 and 1. For instance, the word $0^\infty = 0000\ldots$ corresponds to the number 0, while the word $1^\infty = 1111\ldots$ represents the number $0.1111\ldots = 1/9$. Furthermore, observe that the words $50^\infty = 5000\ldots$ and $49^\infty = 4999\ldots$ both encode $1/2$ since $0.5000\ldots = 0.4999\ldots = 1/2$. More generally, a word $\omega \in \{0, 1, \ldots, 9\}^\infty$ encodes the number

$$\sum_{j=1}^{\infty} \omega_j \left(\frac{1}{10}\right)^j.$$

As noted above, this coding is not one-to-one.

Another common example of a full shift is the full 2-shift. In computer science, the symbol 0 might correspond to having a device (e. g., a switch) turned off, while the symbol 1 would then correspond to the device being on. In physics, the Ising model describes particle spins that have only two possible states, up and down. So we can describe the successive states of a device or particle, at regular observation times, using a sequence of 0s and 1s. Words in the full 2-shift are also called binary sequences, as they correspond to the binary expansions of the numbers between 0 and 1. Indeed, a word $\omega \in \{0,1\}^\infty$ may be viewed as representing the real number

$$\sum_{j=1}^{\infty} \omega_j \left(\frac{1}{2}\right)^j.$$

For instance, the word 0^∞ corresponds to the number 0, as in the full 10-shift. However, the word 1^∞, which encoded the number 1/9 as a word in the full 10-shift, represents the number 1 in the full 2-shift; the number 1/9 is instead encoded by the word $(000111)^\infty = 000111000111\ldots$ in the full 2-shift. As for the number 1/2, it is encoded by the words $10^\infty = 1000\ldots$ and $01^\infty = 0111\ldots$. Hence, this binary coding is not one-to-one either.

The idea of approximating a real number by a rational number by cutting it after a certain number of decimals generalizes to the concept of an *initial block*. Initial blocks play an important role in symbolic dynamics.

Definition 3.1.2. If $\omega \in E^* \cup E^\infty$ and $n \in \mathbb{N}$ does not exceed the length of ω, we define the *initial block* $\omega|_n$ to be the initial n-word of ω, that is, the subword $\omega_1\omega_2\ldots\omega_n$.

In a similar vein, words which begin with the same strings of letters are intuitively close to one another and it is therefore useful to identify the initial subword that they share. To describe this, we introduce the wedge of two words.

Definition 3.1.3. Given two words $\omega, \tau \in E^* \cup E^\infty$, we define their *wedge* $\omega \wedge \tau \in \{\epsilon\} \cup E^* \cup E^\infty$ to be their longest common initial block.

The wedge of two words is better understood via examples. If $E = \{1,2,3\}$ and we have two words $\omega = 12321\ldots$ and $\tau = 12331\ldots$, then $\omega \wedge \tau = 123$. On the other hand, if $\gamma = 22331\ldots$ then $\omega \wedge \gamma = \epsilon$. Of course, if two (finite or infinite) words ω and τ are equal, then $\omega \wedge \tau = \omega = \tau$.

So far, we have talked about the set E^∞ and have given a natural sense to the closeness of any two of its words in terms of their common initial block. We will now endow E^∞ with a natural topology. First, the finite alphabet E is endowed with the discrete topology, that is, the topology in which every subset of E is both open and closed. Then, observing that $E^\infty = \prod_{n=1}^{\infty} E_n$, where $E_n = E$ for each $n \in \mathbb{N}$, that is, E^∞ is the product of countably many copies of E, we equip E^∞ with Tychonoff's product topology generated by the discrete topology on each copy of E. More precisely, that topology is determined by a countable base of open sets called *(initial) cylinder sets*.

Definition 3.1.4. Given a finite word $\omega \in E^*$, the *(initial) cylinder set* $[\omega]$ generated by ω is the set of all infinite words with initial block ω, that is,

$$[\omega] = \{\tau \in E^\infty : \tau|_{|\omega|} = \omega\} = \{\tau \in E^\infty : \tau_j = \omega_j, \ \forall 1 \le j \le |\omega|\}.$$

We take this opportunity to introduce more general cylinder sets.

Definition 3.1.5. Given a finite word $\omega \in E^*$ and $m, n \in \mathbb{N}$ such that $n - m + 1 = |\omega|$, the *(m, n)-cylinder set* $[\omega]_m^n$ generated by ω is the set of all infinite words whose subblock from coordinates m to n coincides with ω, that is,

$$[\omega]_m^n = \{\tau \in E^\infty : \tau_j = \omega_{j-m+1} \ \text{for all} \ m \le j \le n\}.$$

In particular, note that $[\omega]_1^{|\omega|} = [\omega]$.

Definition 3.1.6. When equipped with Tychonoff's product topology, that is, the topology generated by the (initial) cylinder sets, the set E^∞ is called the *full E-shift space* or, more simply, full shift space.

We now describe the most fundamental topological properties of full shift spaces. First, note that there are countably many (initial) cylinder sets since there are countably many finite words in E^*. Since these sets form a base for the topology, the space E^∞ is second countable.

Furthermore, since the alphabet E is finite, when endowed with the discrete topology it becomes a compact metrizable space (see Exercise 3.4.2). According to Tychonoff's theorem (see 17.8 in [77]), it then follows that E^∞, as a product of countably many copies of E, is also a compact metrizable space (see Exercise 3.4.2).

Moreover, the full shift space E^∞ is perfect, that is, it contains no isolated point. Indeed, notice that every point $\omega \in E^\infty$ is such that

$$\{\omega\} = \bigcap_{n=1}^{\infty} [\omega|_n].$$

Alternatively, note that every point $\omega \in E^\infty$ is the limit of the sequence of "periodic" points $((\omega|_n)^\infty)_{n=1}^\infty = ((\omega_1 \ldots \omega_n)^\infty)_{n=1}^\infty$.

Finally, since the complement of a cylinder set is a union of cylinder sets, the cylinder sets are both open and closed subsets of the full shift space. Therefore, E^∞ is totally disconnected (see Exercise 3.4.3). Summarizing these properties, we have obtained the following lemma.

Lemma 3.1.7. *The full shift space E^∞ is a totally disconnected, perfect, compact, metrizable space.*

To put this lemma in context, we recall the following definition.

Definition 3.1.8. A *Cantor space* (also frequently called a Cantor set) is a totally disconnected, perfect, compact, second-countable, Hausdorff topological space.

In light of this definition, we can restate Lemma 3.1.7 as follows.

Lemma 3.1.9. *The full shift space E^∞ is a Cantor space.*

Cantor spaces can be characterized as follows:
(a) They are totally disconnected, perfect, compact, metrizable topological spaces.
(b) They are homeomorphic to the middle-third Cantor set.
(c) They are homeomorphic to E^∞ for some finite set E having at least two elements.
(d) They are homeomorphic to E^∞ for every finite set E having at least two elements.

We now introduce a family of metrics on E^∞, each of which reflects the idea that two words are close if they share a long initial block. The longer their common initial subword, the closer two words are.

Definition 3.1.10. For each $s \in (0,1)$, let $d_s : E^\infty \times E^\infty \to [0,1]$ be defined by

$$d_s(\omega,\tau) = s^{|\omega\wedge\tau|}.$$

Remark 3.1.11. If $\omega,\tau \in E^\infty$ have no common initial block, then $\omega \wedge \tau = \epsilon$. Thus, $|\omega\wedge\tau| = 0$ and $d_s(\omega,\tau) = 1$. On the other hand, if $\omega = \tau$ then $|\omega\wedge\tau| = \infty$ and we adopt the convention that $s^\infty := 0$.

Proposition 3.1.12. *For every $s \in (0,1)$, the map $d_s : E^\infty \times E^\infty \to [0,1]$ defined above is an ultrametric, and thus a metric.*

Proof. First, note that $d_s(\omega,\omega) = s^\infty := 0$. Moreover, $d_s(\omega,\tau) = 0$ implies that $|\omega\wedge\tau| = \infty$, that is, $\omega = \tau$. Second, d_s is symmetric, as

$$d_s(\omega,\tau) = s^{|\omega\wedge\tau|} = s^{|\tau\wedge\omega|} = d_s(\tau,\omega).$$

It only remains to show that $d_s(\omega,\tau) \le \max\{d_s(\omega,\rho), d_s(\rho,\tau)\}$ for all $\omega,\rho,\tau \in E^\infty$. Fix $\omega,\rho,\tau \in E^\infty$. Observe that ω and ρ share the same initial block of length $|\omega\wedge\rho|$, while ρ and τ share the same initial block of length $|\rho\wedge\tau|$. This implies that ω, ρ, and τ all share the same initial block of length equal to $\min\{|\omega\wedge\rho|, |\rho\wedge\tau|\}$. Since $0 < s < 1$, we then have that

$$d_s(\omega,\tau) \le s^{\min\{|\omega\wedge\rho|,|\rho\wedge\tau|\}} = \max\{s^{|\omega\wedge\rho|}, s^{|\rho\wedge\tau|}\} = \max\{d_s(\omega,\rho), d_s(\rho,\tau)\}.$$

This shows that d_s is an ultrametric. In particular, it is a metric as the triangle inequality

$$d_s(\omega,\tau) \le d_s(\omega,\rho) + d_s(\rho,\tau)$$

is obviously satisfied. $\square$

We have now defined an uncountable family of metrics on E^∞, one for each $s \in (0,1)$. These metrics induce Tychonoff's topology on E^∞ (see Exercise 3.4.4). This implies that these metrics are topologically equivalent. In fact, they are Hölder equivalent but not Lipschitz equivalent (see Exercises 3.4.5 and 3.4.6).

Let us now describe what it means for a sequence to converge to a limit in the full shift space. Let $s \in (0,1)$. Let $(\omega^{(k)})_{k=1}^\infty$ be a sequence in E^∞. Observe that

$$\lim_{k\to\infty} d_s(\omega^{(k)},\omega) = \lim_{k\to\infty} s^{|\omega^{(k)}\wedge\omega|} = 0 \Leftrightarrow \lim_{k\to\infty} |\omega^{(k)} \wedge \omega| = \infty.$$

In other words, a sequence $(\omega^{(k)})_{k=1}^\infty$ converges to the infinite word ω if and only if for any $L \in \mathbb{N}$, the words in the sequence eventually all have $\omega|_L$ as initial L-block.

Now that we have explored the space E^∞, we would like to introduce some dynamics on it. To this end, we define the shift map, whose action consists in removing the first letter of each word and shifting all the remaining letters one space/coordinate to the left.

Definition 3.1.13. The *full left-shift map* $\sigma : E^\infty \to E^\infty$ is defined by $\sigma(\omega) = \sigma((\omega_j)_{j=1}^\infty) := (\omega_{j+1})_{j=1}^\infty$, that is,

$$\sigma(\omega_1\omega_2\omega_3\omega_4\ldots) := \omega_2\omega_3\omega_4\ldots$$

We will also often refer to this map simply as the shift map.

The shift map is $\#E$-to-one on E^∞. In other words, each word has $\#E$ preimages under the shift map. Indeed, given any letter $e \in E$ and any infinite word $\omega \in E^\infty$, the concatenation $e\omega = e\omega_1\omega_2\omega_3\ldots$ of e with ω is a preimage of ω under the shift map since $\sigma(e\omega) = \omega$.

The shift map is obviously continuous, since two words that are close share a long initial block and thus their images under the shift map, which result from dropping their first letters, will also share a long initial block. More precisely, for any $\omega, \tau \in E^\infty$ with $d_s(\omega,\tau) < 1$, that is, with $|\omega \wedge \tau| \geq 1$, we have that

$$d_s(\sigma(\omega),\sigma(\tau)) = s^{|\sigma(\omega)\wedge\sigma(\tau)|} = s^{|\omega\wedge\tau|-1} = s^{-1}s^{|\omega\wedge\tau|} = s^{-1}d_s(\omega,\tau).$$

So the shift map is Lipschitz continuous with Lipschitz constant s^{-1}. In particular, the shift map defines a dynamical system on E^∞. It is then natural to ask the following question: Given two finite sets E and F, under which conditions are the shift maps $\sigma_E : E^\infty \to E^\infty$ and $\sigma_F : F^\infty \to F^\infty$ topologically conjugate? Notice that the only fixed points of σ_E are the "constant" words e^∞, for each $e \in E$. Hence, the number of fixed points of σ_E is equal to $\#E$. Recall that the cardinality of the set of fixed points $\mathrm{Fix}(T)$ is a topological invariant. So, if $\#E \neq \#F$ then σ_E is not topologically conjugate to σ_F. In fact, as we will see in the following theorem, σ_E is topologically conjugate to σ_F precisely when $\#E = \#F$.

Theorem 3.1.14. *$\sigma_E : E^\infty \to E^\infty$ and $\sigma_F : F^\infty \to F^\infty$ are topologically conjugate if and only if $\#E = \#F$.*

Proof. If σ_E and σ_F are topologically conjugate, then it is clear from the discussion above that $\#E = \#F$. For the converse, assume that $\#E = \#F$. Therefore, there must exist some bijection $H : E \to F$. Now define the mapping $h : E^\infty \to F^\infty$ by concatenation, that is, by setting

$$h(\omega_1\omega_2\omega_3\dots) := H(\omega_1)H(\omega_2)H(\omega_3)\dots.$$

Then h is a homeomorphism: that h is a bijection follows from H being a bijection, while the continuity of both h and h^{-1} follows directly from the fact that h is an isometry, as $|h(\omega)\wedge h(\tau)| = |\omega\wedge\tau|$. It remains to show that the following diagram commutes:

$$\begin{tikzcd} E^\infty \arrow[r, "\sigma_E"] \arrow[d, "h"'] & E^\infty \arrow[d, "h"] \\ F^\infty \arrow[r, "\sigma_F"'] & F^\infty \end{tikzcd}$$

Indeed,

$$\begin{aligned} h\circ\sigma_E(\omega_1\omega_2\omega_3\dots) &= h(\omega_2\omega_3\omega_4\dots) \\ &= H(\omega_2)H(\omega_3)H(\omega_4)\dots \\ &= \sigma_F\big(H(\omega_1)H(\omega_2)H(\omega_3)\dots\big) \\ &= \sigma_F\circ h(\omega_1\omega_2\omega_3\dots). \end{aligned}$$ □

3.2 Subshifts of finite type

We now turn our attention to subsystems of full shift spaces. By definition, the subsystems of the full shift space E^∞ are all shift-invariant, compact subsets of E^∞. Recall that a set $F \subseteq E^\infty$ is shift-invariant (i. e., σ-invariant) if $\sigma(F) \subseteq F$.

The notion of forbidden word arises naturally in the study of subsets and subsystems of full shift spaces. Let $\mathcal{F} \subseteq E^*$ be a set of finite words, called *forbidden words* in the sequel. We shall denote by $E^\infty_{\mathcal{F}}$ the set of all those infinite words in E^∞ that do not contain any forbidden word as a subword. In other words,

$$E^\infty_{\mathcal{F}} := \{\omega \in E^\infty : \omega_m\omega_{m+1}\dots\omega_n \notin \mathcal{F},\ \forall m,n \in \mathbb{N},\ m \le n\}.$$

A subshift is a subset of a full shift that can be described by a set of forbidden words.

Definition 3.2.1. A subset F of the full shift E^∞ is called a *subshift* if there is a set $\mathcal{F} \subseteq E^*$ of forbidden words such that $F = E^\infty_{\mathcal{F}}$. It is sometimes said that $F = E^\infty_{\mathcal{F}}$ is the subshift generated by $\mathcal{F}$.

We briefly examine the relation between sets of forbidden words and the subshifts they generate.

Lemma 3.2.2. *Let $\mathcal{F} \subseteq E^*$ and $\mathcal{G} \subseteq E^*$.*
(a) *If $\mathcal{F} \subseteq \mathcal{G}$, then $E^\infty_{\mathcal{F}} \supseteq E^\infty_{\mathcal{G}}$.*
(b) *If $\mathcal{F} \subseteq \mathcal{G}$ and every word in $\mathcal{G}$ admits a subword which is in $\mathcal{F}$, then $E^\infty_{\mathcal{F}} = E^\infty_{\mathcal{G}}$.*

Proof.
(a) Suppose that $\mathcal{F} \subseteq \mathcal{G}$. If $\tau \notin \mathcal{G}$ then $\tau \notin \mathcal{F}$, and hence

$$\begin{aligned} E^\infty_{\mathcal{G}} &= \{\omega \in E^\infty : \omega_m \omega_{m+1} \ldots \omega_n \notin \mathcal{G},\ \forall m, n \in \mathbb{N},\ m \le n\} \\ &\subseteq \{\omega \in E^\infty : \omega_m \omega_{m+1} \ldots \omega_n \notin \mathcal{F},\ \forall m, n \in \mathbb{N},\ m \le n\} \\ &= E^\infty_{\mathcal{F}}. \end{aligned}$$

(b) Suppose that $\mathcal{F} \subseteq \mathcal{G}$ and that every word in $\mathcal{G}$ admits a subword which is in $\mathcal{F}$. By (a), we already know that $E^\infty_{\mathcal{G}} \subseteq E^\infty_{\mathcal{F}}$. Therefore, it only remains to establish that $E^\infty_{\mathcal{G}} \supseteq E^\infty_{\mathcal{F}}$. Let $\omega \in E^\infty \setminus E^\infty_{\mathcal{G}}$. Then there exist $m, n \in \mathbb{N}$, $m \le n$, such that $\omega_m \omega_{m+1} \ldots \omega_n \in \mathcal{G}$. Since every word in $\mathcal{G}$ contains a subword which is in $\mathcal{F}$, there exist $k, l \in \mathbb{N}$ such that $m \le k \le l \le n$ and $\omega_k \omega_{k+1} \ldots \omega_l \in \mathcal{F}$. Thus, $\omega \in E^\infty \setminus E^\infty_{\mathcal{F}}$. This means that $E^\infty \setminus E^\infty_{\mathcal{G}} \subseteq E^\infty \setminus E^\infty_{\mathcal{F}}$, and hence $E^\infty_{\mathcal{G}} \supseteq E^\infty_{\mathcal{F}}$. □

Taken together, the next two theorems demonstrate that the terms *subsystem of a full shift* and *subshift* can be used interchangeably.

Theorem 3.2.3. *Every subshift of E^∞ is a subsystem of the full shift E^∞.*

Proof. Let F be a subshift of E^∞. This means that there exists $\mathcal{F} \subseteq E^*$ such that $F = E^\infty_{\mathcal{F}}$. From its definition, it is clear that $E^\infty_{\mathcal{F}}$ is shift invariant. Moreover, observe that

$$\begin{aligned} E^\infty_{\mathcal{F}} &= \bigcap_{m=1}^{\infty} \bigcap_{n \ge m} \{\omega \in E^\infty : \omega_m \ldots \omega_n \notin \mathcal{F}\} \\ &= \bigcap_{m=1}^{\infty} \bigcap_{n \ge m} \bigcap_{\tau \in E^{n-m+1} \cap \mathcal{F}} \{\omega \in E^\infty : \omega_m \ldots \omega_n \ne \tau\} \\ &= \bigcap_{m=1}^{\infty} \bigcap_{n \ge m} \bigcap_{\tau \in E^{n-m+1} \cap \mathcal{F}} E^\infty \setminus [\tau]_m^n. \end{aligned}$$

Since every cylinder set is open in E^∞, the sets $E^\infty \setminus [\tau]_m^n$ are compact. Therefore, $E^\infty_{\mathcal{F}}$ is an intersection of compact sets and is thereby compact. In summary, $F = E^\infty_{\mathcal{F}}$ is a compact shift-invariant subset of E^∞. That is, it is a subsystem of the full shift E^∞. □

Theorem 3.2.4. *Let F be a subsystem of the full shift space E^∞. Let*

$$\mathcal{F}_F := \{\tau \in E^* : [\tau] \subseteq E^\infty \setminus F\}.$$

Then

$$F = E^{\infty}_{\mathcal{F}_F}.$$

In other words, the subsystem F of the full shift space E^{∞} coincides with the subshift $E^{\infty}_{\mathcal{F}_F}$ generated by the set of finite words $\mathcal{F}_F$.

Proof. By hypothesis, the set $F \subseteq E^{\infty}$ is σ-invariant and compact. In particular, F is closed. Therefore, $E^{\infty} \setminus F$ is open.

Let $\rho \in E^{\infty} \setminus F$. Since $E^{\infty} \setminus F$ is open, this is equivalent to the existence of $n \in \mathbb{N}$ such that $[\rho|_n] \subseteq E^{\infty} \setminus F$. In turn, this means that $\rho|_n \in \mathcal{F}_F$ and hence $\rho \notin E^{\infty}_{\mathcal{F}_F}$.

Conversely, assume that $\rho \notin E^{\infty}_{\mathcal{F}_F}$. There exist $m, n \in \mathbb{N}$, $m \le n$, such that $\rho_m\rho_{m+1}\rho_n \in \mathcal{F}_F$. In other terms, $\sigma^{m-1}(\rho)_1\sigma^{m-1}(\rho)_2 \dots \sigma^{m-1}(\rho)_{n-m+1} \in \mathcal{F}_F$. That is, $\sigma^{m-1}(\rho)|_{n-m+1} \in \mathcal{F}_F$. This is equivalent to $[\sigma^{m-1}(\rho)|_{n-m+1}] \subseteq E^{\infty} \setminus F$. In particular, $\sigma^{m-1}(\rho) \in E^{\infty} \setminus F$. Since F is σ-invariant, this implies that $\rho \in E^{\infty} \setminus F$. □

We shall now study a special class of subshifts. They are called *subshifts of finite type*.

Definition 3.2.5. A subshift F of the full shift E^{∞} is said to be of *finite type* if there is a *finite* set $\mathcal{F} \subseteq E^*$ of forbidden words such that $F = E^{\infty}_{\mathcal{F}}$.

In this case, it easily follows from Lemma 3.2.2(b) that the finite set $\mathcal{F}$ can be chosen so that $\mathcal{F} \subseteq E^q$ for some $q \in \mathbb{N}$. The set $\mathcal{F}$ then induces a function $A : E^q \to \{0, 1\}$ whose value is 0 on $\mathcal{F}$ (i. e., for all forbidden words of length q) and 1 on $E^q \setminus \mathcal{F}$ (i. e., for all other words of length q). We will revisit this general framework in the next section, where we will prove that all cases can be reduced to the case $q = 2$. For this reason, we will concentrate on this latter case in this section. Here, rather than using formally a function $A : E^2 \to \{0, 1\}$, subshifts of finite type are best understood by means of an incidence/transition matrix. An incidence/transition matrix is simply a square matrix consisting entirely of zeros and ones. To do this, we will work with the full shift on the alphabet $\{1, 2, \dots, \#E\}$ rather than the full E-shift itself. The incidence/transition matrix determines which letter/number(s) may follow a given letter/number.

Definition 3.2.6. Let A be an incidence matrix of size $\#E \times \#E$. The set of all infinite *A-admissible words* is the *subshift of finite type*

$$E^{\infty}_A := \{\omega \in E^{\infty} : A_{\omega_n\omega_{n+1}} = 1, \ \forall n \in \mathbb{N}\}.$$

E^{∞}_A is a subshift of finite type since $E^{\infty}_A = E^{\infty}_{\mathcal{F}}$, where the set of forbidden words $\mathcal{F}$ is the finite set of two-letter words

$$\mathcal{F} = \{ij \in E^2 : A_{ij} = 0\}.$$

A (finite) word $\omega_1\omega_2 \dots \omega_n$ is said to be *A-admissible* (or, if there can be no confusion about the matrix A, more simply, *admissible*) if

$$A_{\omega_k\omega_{k+1}} = 1, \quad \forall 1 \le k < n.$$

The set of all A-admissible n-words will naturally be denoted by E_A^n, while the set of all A-admissible finite words will naturally be denoted by E_A^*. An A-admissible path of length n from $i \in E$ to $j \in E$ is any A-admissible word ω of length n with $\omega_1 = i$ and $\omega_n = j$. Thus, the entry A_{ij} of the matrix A indicates the number of admissible words (or paths) of length 2 from i to j (which is necessarily either 0 or 1). By multiplying the matrix A with itself, we see that $(A^2)_{ij} = \sum_{k=1}^{\#E} A_{ik}A_{kj}$ specifies the number of admissible words of length 3 from i to j, since $A_{ik}A_{kj} = 1$ if and only if ikj is admissible. Similarly, $(A^n)_{ij}$ is the number of admissible words of length $n+1$ from i to j, and $(A^n)_{ij} > 0$ if and only if there is at least one such path.

Note that if a row of A does not contain any 1, then no infinite word can contain the letter corresponding to that row. This letter can then be thrown out of the alphabet because it is inessential. We will henceforth assume that this does not happen, that is, we will assume that all the letters are essential by imposing the condition that every row of A contains at least one 1. This is a standing assumption throughout this book.

Notice that if all the entries of the incidence matrix A are 1s, then $\mathcal{F} = \emptyset$ and $E_A^\infty = E^\infty$. However, if A has at least one 0 entry then E_A^∞ is a proper subshift of E^∞. In particular, if A is the identity matrix then $E_A^\infty = \{e^\infty : e \in E\}$, that is, E_A^∞ is the set of all constant words, which are the fixed points of σ in E^∞.

Alternatively, E_A^∞ can be represented by a directed graph. Imagine that each element e of E is a vertex of a directed graph. Then the directed graph has an edge directed from vertex e to vertex f if and only if $A_{ef} = 1$. The set of infinite A-admissible words E_A^∞ then corresponds to the set of all possible infinite walks along the directed graph. This is sometimes called a *vertex shift*.

Example 3.2.7. Let $E = \{1, 2, 3\}$ and let

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

What is E_A^∞? We strongly advise the reader to draw the corresponding directed graph. According to the incidence matrix, the letter 3 can only be followed by itself since $A_{33} = 1$ while $A_{31} = A_{32} = 0$. This means that vertex 3 of the directed graph has only one outgoing edge, and it is a self-loop. Thus, if $\omega \in E_A^\infty$ starts with a 3, then $\omega = 3^\infty$. According to the incidence matrix again, the letter 1 can only be followed by itself or by 3 since $A_{11} = A_{13} = 1$ whereas $A_{12} = 0$. This means that vertex 1 has two outgoing edges, one being a self-loop while the other terminates at vertex 3. Thus, if ω starts with a 1, then this can be followed by either a 3, in which case it is followed by infinitely many 3s, or by a 1, in which case we face the same choice again. Therefore, the admissible words starting with a 1 are 1^∞ and $1^n 3^\infty$, $n \in \mathbb{N}$. Finally, if ω starts with a 2 then, as $A_{21} = A_{22} = A_{23} = 1$, a 2 can be followed by any other letter. This means that vertex 2 has three outgoing edges, one terminating at each vertex. Hence the admissible words starting with a 2 are 2^∞, 2^m3^∞, 2^m1^∞ and $2^m1^n3^\infty$, where $m,n \in \mathbb{N}$. In summary,

$$E_A^\infty = \{1^\infty, 2^\infty, 3^\infty\} \cup \{2^m1^\infty : m \in \mathbb{N}\} \cup \{2^m1^n3^\infty : m,n \in \mathbb{Z}_+\}.$$

We now study the topological properties of E_A^∞. Since E_A^∞ is a subshift, Theorem 3.2.3 yields that this set is compact and σ-invariant. Nevertheless, we provide below a direct proof of this important fact.

Theorem 3.2.8. *E_A^∞ is a compact σ-invariant set.*

Proof. Let $\omega \in E_A^\infty$. Then $A_{\omega_n\omega_{n+1}} = 1$ for every $n \in \mathbb{N}$. In particular, this implies that $A_{\sigma(\omega)_n\sigma(\omega)_{n+1}} = A_{\omega_{n+1}\omega_{n+2}} = 1$ for all $n \in \mathbb{N}$. Thus, $\sigma(\omega) \in E_A^\infty$ and we therefore have that E_A^∞ is σ-invariant.

In order to show that E_A^∞ is compact, recall that a closed subset of a compact space is compact. As E^∞ is a compact space when endowed with the product topology, it is sufficient to prove that E_A^∞ is closed. Let $(\omega^{(k)})_{k=1}^\infty$ be a sequence in E_A^∞ and suppose that $\lim_{k\to\infty} \omega^{(k)} = \omega$. We must show that $\omega \in E_A^\infty$, or, in other words, we need to show that $A_{\omega_n\omega_{n+1}} = 1$ for all $n \in \mathbb{N}$. To that end, fix $n \in \mathbb{N}$. For each $k \in \mathbb{N}$, we have $A_{\omega_n^{(k)}\omega_{n+1}^{(k)}} = 1$ since $\omega^{(k)} \in E_A^\infty$. Moreover, $\lim_{k\to\infty} |\omega^{(k)} \wedge \omega| = \infty$ since $\omega^{(k)} \to \omega$. So, for sufficiently large k, we have $|\omega^{(k)} \wedge \omega| \geq n+1$. In particular, $\omega_n^{(k)} = \omega_n$ and $\omega_{n+1}^{(k)} = \omega_{n+1}$ for all k large enough. Hence, we deduce that $A_{\omega_n\omega_{n+1}} = A_{\omega_n^{(k)}\omega_{n+1}^{(k)}} = 1$ for all k large enough. Since n was chosen arbitrarily, we conclude that $\omega \in E_A^\infty$.

We now provide an alternative proof of the compactness of E_A^∞. Observe that

$$\begin{aligned} E_A^\infty &= \{\omega \in E^\infty : A_{\omega_n\omega_{n+1}} = 1, \forall n \in \mathbb{N}\} \\ &= \bigcap_{n=1}^\infty \{\omega \in E^\infty : A_{\omega_n\omega_{n+1}} = 1\} \\ &= \bigcap_{n=1}^\infty \Big[\bigcup_{\omega\in E_A^2} [\omega]_n^{n+1} \Big]. \end{aligned}$$

Recall that all cylinders are closed subsets of the compact space E^∞. Therefore, they are all compact. Since the set E_A^2 is finite, the union of cylinders $\bigcup_{\omega\in E_A^2}[\omega]_n^{n+1}$ is compact for all $n \in \mathbb{N}$. As an intersection of these latter sets, E_A^∞ is compact. □

The σ-invariance of E_A^∞ ensures that the map $\sigma : E_A^\infty \to E_A^\infty$ is well-defined. This restriction of the shift map is obviously continuous (in fact, Lipschitz continuous). The compactness of E_A^∞ makes the couple (E_A^∞, σ) a well-defined topological dynamical system. It is a subsystem of the full shift (E^∞, σ).

We shall now determine the condition under which $\sigma : E_A^\infty \to E_A^\infty$ is surjective.

Lemma 3.2.9. *The (restriction of) the shift map $\sigma : E_A^\infty \to E_A^\infty$ is surjective if and only if for every $j \in E$ there exists $i \in E$ such that $A_{ij} = 1$, that is, if and only if every column of A contains at least one* 1.

Proof. First, observe that if for every $j \in E$ there exists some $i \in E$ such that $A_{ij} = 1$, then $\sigma : E_A^\infty \to E_A^\infty$ is surjective. Indeed, given any $\omega \in E_A^\infty$, there exists $e \in E$ such that $A_{e\omega_1} = 1$. Then $e\omega \in E_A^\infty$ and $\sigma(e\omega) = \omega$.

To establish the converse, suppose A has a column consisting solely of 0s, that is, suppose there exists $j \in E$ such that $A_{ij} = 0$ for all $i \in E$. By our standing assumption, every row of A contains at least one 1. So there must be a word of the form $j\omega$ in E_A^∞. However, $j\omega \notin \sigma(E_A^\infty)$ since there is no word of the form $ij\omega$ in E_A^∞. □

After determining when the map $\sigma : E_A^\infty \to E_A^\infty$ is surjective, it is natural to next consider the condition under which the shift map is injective on a subshift of finite type.

Lemma 3.2.10. *The shift map $\sigma : E_A^\infty \to E_A^\infty$ is injective if and only if for every $j \in E$ there exists at most one $i \in E$ such that $A_{ij} = 1$, that is, if and only if A contains at most one* 1 *in each of its columns.*

The proof of this lemma is left to the reader as an exercise (see Exercise 3.4.11).

Corollary 3.2.11. *The shift map $\sigma : E_A^\infty \to E_A^\infty$ is bijective if and only if every column of A contains exactly one* 1. *Given our standing assumption that all letters are essential, we thus have that $\sigma : E_A^\infty \to E_A^\infty$ is bijective if and only if every row and every column of A contains exactly one* 1.

A matrix which has precisely one 1 in each of its rows and each of its columns is called a *permutation matrix*. Such a matrix has the property that A^n is the identity matrix for some $n \in \mathbb{N}$. This means that $\sigma : E_A^\infty \to E_A^\infty$ is bijective if and only if E_A^∞ consists solely of finitely many periodic points.

In addition to being continuous, the shift map is an open map. Recall that a map is said to be open if it sends open sets onto open sets. Moreover, note that as the cylinder sets form a base for the topology on E^∞, their restriction to E_A^∞, which we also call cylinders and which will be denoted by the same notation, constitute a base for the topology on E_A^∞. From this point on, a cylinder $[\omega]_m^n$ will be understood to be a cylinder in E_A^∞. That is to say,

$$[\omega]_m^n := \{\tau \in E_A^\infty : \tau_k = \omega_{k-m+1},\ \forall m \le k \le n\}.$$

Theorem 3.2.12. *The shift map $\sigma : E_A^\infty \to E_A^\infty$ is an open map.*

Proof. As the cylinder sets of length at least two form a base for the topology on E_A^∞, it suffices to prove that the image of an arbitrary cylinder of length at least two is a cylinder. Say $\omega = \omega_1\omega_2\ldots\omega_n$, where $n \ge 2$. Then $\sigma([\omega]) = [\omega_2\ldots\omega_n]$. Indeed, let $\tau \in [\omega]$. Then $\tau = \omega_1\omega_2\ldots\omega_n\tau_{n+1}\tau_{n+2}\ldots$, and thus $\sigma(\tau) = \omega_2\ldots\omega_n\tau_{n+1}\tau_{n+2}\ldots$. Hence $\sigma([\omega]) \subseteq [\omega_2\ldots\omega_n]$. Conversely, let $\gamma \in [\omega_2\ldots\omega_n]$. Since $A_{\omega_1\omega_2} = 1$, we have that $\omega_1\gamma = \omega_1\omega_2\ldots\omega_n\gamma_{n+1}\gamma_{n+2}\cdots \in E_A^\infty$. In fact, observe that $\omega_1\gamma = \omega\gamma_{n+1}\gamma_{n+2}\cdots \in [\omega]$.

Moreover, $\sigma(\omega_1\gamma) = \gamma$. Hence, $\sigma([\omega]) \supseteq [\omega_2 \dots \omega_n]$. We have thus established that

$$\sigma\big([\omega_1\omega_2 \dots \omega_n]\big) = [\omega_2 \dots \omega_n].$$

Hence, the image of a cylinder of length $n \geq 2$ is the cylinder of length $n - 1$ obtained by dropping the first symbol. □

3.2.1 Topological transitivity

We now describe the condition on the matrix A under which the subshift of finite type $\sigma : E_A^\infty \to E_A^\infty$ is topologically transitive. Recall that a dynamical system $T : X \to X$ is defined to be transitive if it admits at least one point x with the property that $\omega(x) = X$ (see Definition 1.5.6). We proved in Theorem 1.5.11 that this is equivalent to the system T being topologically mixing, that is, the orbit under T of every nonempty open set encounters every nonempty open set in X (cf. Definition 1.5.10).

Definition 3.2.13. An incidence matrix A is called *irreducible* if for each ordered pair $i, j \in E$ there exists $p := p(i, j) \in \mathbb{N}$ such that $(A^p)_{ij} > 0$.

Observe that an irreducible matrix cannot contain any row or column consisting solely of 0s. In light of Lemma 3.2.9, the irreducibility of a matrix A compels the surjectivity of the shift map $\sigma : E_A^\infty \to E_A^\infty$. We shall now prove that irreducibility of A is equivalent to the transitivity of σ.

Theorem 3.2.14. *The shift map $\sigma : E_A^\infty \to E_A^\infty$ is transitive if and only if the matrix A is irreducible.*

Proof. First, suppose that σ is transitive. By Remark 1.5.7(c), it is surjective, and thus topologically mixing according to Theorem 1.5.11. Fix $i, j \in E$. Then there exists $p := p(i, j) \in \mathbb{N}$ such that $\sigma^p([i]) \cap [j] \neq \emptyset$. So pick $\omega \in [i]$ so that $\sigma^p(\omega) \in [j]$. This means that $\omega_1\omega_2 \dots \omega_{p+1}$ is an admissible word of length $p + 1$ from $i = \omega_1$ to $j = \omega_{p+1}$. Thus $(A^p)_{ij} > 0$. Since $i, j \in E$ were chosen arbitrarily, we conclude that A is irreducible.

To prove the converse, suppose that A is irreducible. As the cylinder sets form a base for the topology, it is sufficient to restrict our attention to them. Let $\omega = \omega_1\omega_2 \dots \omega_k \in E_A^*$ and $\tau = \tau_1\tau_2 \dots \tau_l \in E_A^*$. We need to show that there exists some $n \in \mathbb{N}$ such that $\sigma^n([\omega]) \cap [\tau] \neq \emptyset$. Consider the pair of letters ω_k and τ_1. Since A is irreducible, there exists some $p := p(\omega_k, \tau_1) \in \mathbb{N}$ such that $(A^p)_{\omega_k\tau_1} > 0$. This means that there exists a finite word γ of length $p - 1$ such that $\omega_k\gamma\tau_1$ is an admissible word of length $p + 1$ from ω_k to τ_1. Consequently, $\omega\gamma\tau \in E_A^*$. Concatenate to this word a suffix $\overline{\tau}_{l+1}$ with the property that $A_{\tau_l\overline{\tau}_{l+1}} = 1$. This is possible thanks to our standing assumption that every row of A contains a 1. Continue concatenating in this way to build an infinite admissible word $\overline{\omega} = \omega\gamma\tau\overline{\tau}_{l+1}\overline{\tau}_{l+2} \dots \in E_A^\infty$. Then $\overline{\omega} \in [\omega]$ and

$\sigma^{k+p-1}(\overline{\omega}) = \tau\overline{\tau}_{l+1}\overline{\tau}_{l+2}\ldots \in [\tau]$. Hence $\sigma^{k+p-1}([\omega]) \cap [\tau] \neq \emptyset$. Since $\omega, \tau \in E_A^*$ were arbitrary, we deduce that σ is topologically mixing. The surjectivity of σ is also guaranteed by the irreducibility of A. Therefore, σ is transitive according to Theorem 1.5.11. $\square$

Note that the above theorem does not hold if our standing assumption that every row of A contains at least one 1 is dropped (see Exercise 3.4.12). Also, as mentioned before, it follows immediately from both the transitivity of σ and from the irreducibility of A that every column of A contains at least one 1, or equivalently, that σ is surjective (see Lemma 3.2.9).

3.2.2 Topological exactness

We now describe the condition on the matrix A under which the subshift of finite type $\sigma : E_A^\infty \to E_A^\infty$ is topologically exact. Recall that a dynamical system $T : X \to X$ is said to be topologically exact if every nonempty open set in X is eventually blown up onto the entire space X under iteration that is, for every nonempty open set $U \subseteq X$ there is $N \in \mathbb{N}$ such that $T^N(U) = X$ (cf. Definition 1.5.16). Note that this condition can only be fulfilled if the map T is surjective.

Definition 3.2.15. An incidence matrix A is called *primitive* if there exists some $p \in \mathbb{N}$ such that A^p has only positive entries, which is usually written as $A^p > 0$.

Note that every primitive matrix is irreducible, but that there exist irreducible matrices which are not primitive (see Exercise 3.4.13).

Theorem 3.2.16. *The shift map $\sigma : E_A^\infty \to E_A^\infty$ is topologically exact if and only if the matrix A is primitive.*

Proof. Suppose first that σ is topologically exact. Then for each $e \in E$ there exists some $p_e \in \mathbb{N}$ such that $\sigma^{p_e}([e]) = E_A^\infty$, since $[e]$ is an open set. Define $p := \max\{p_e : e \in E\}$ and note that p is finite. We claim that $A^p > 0$. To see this, fix an ordered pair $i, j \in E$. Since $\sigma^p([i]) = E_A^\infty$, we have $\sigma^p([i]) \cap [j] \neq \emptyset$. So if $\omega \in [i]$ and $\sigma^p(\omega) \in [j]$, then $\omega = i\omega_2\omega_3\ldots\omega_p j\omega_{p+2}\omega_{p+3}\ldots$. In other words, $i\omega_2\omega_3\ldots\omega_p j$ is an admissible word of length $p+1$ from i to j, and hence $(A^p)_{ij} > 0$. Since this is true for all $i, j \in E$, we conclude that $A^p > 0$.

To prove the converse, suppose that A is primitive, that is, suppose there exists $p \in \mathbb{N}$ such that $A^p > 0$. Then for each ordered pair $i, j \in E$ there is at least one admissible word of length $p+1$ from i to j. Since the cylinder sets $\{[\omega]\}_{\omega \in E_A^*}$ form a base for the topology on E_A^∞, it is sufficient to prove topological exactness for cylinder sets. So let $\omega = \omega_1\ldots\omega_n \in E_A^*$ and pick an arbitrary $\tau \in E_A^\infty$. There exists a finite word γ of length $p-1$ such that $\omega_n\gamma\tau_1$ is an admissible word of length $p+1$. Let $\overline{\omega} = \omega\gamma\tau \in E_A^\infty$. Then $\overline{\omega} \in [\omega]$ and $\sigma^{n+p-1}(\overline{\omega}) = \tau$. Therefore, as τ was arbitrarily chosen in E_A^∞, we deduce that $\sigma^{n+p-1}([\omega]) = E_A^\infty$ and σ is topologically exact. $\square$

The reader may be wondering under which condition on A the shift $\sigma : E_A^\infty \to E_A^\infty$ is minimal. The answer to this question can be found in Exercise 3.4.15.

3.2.3 Asymptotic behavior of periodic points

We now prove that for any matrix A, the maximal growth rate of the number of periodic points in E_A^∞ coincides with the logarithm of the spectral radius of A. Recall that the spectral radius of A is defined to be the largest eigenvalue of A (in absolute value). The *spectral radius* $r(A)$ can also be defined by

$$r(A) := \lim_{n\to\infty} \|A^n\|^{1/n}, \tag{3.1}$$

for any matrix norm $\|\cdot\|$ (for a proof of this fact, see Proposition 3.8 in Conway [15]). In what follows, it is convenient to choose the norm to be the sum of the absolute value of the entries of the matrix, that is, for a $k \times k$ matrix B, the norm is

$$\|B\| := \sum_{i,j=1}^{k} |B_{ij}|.$$

Before continuing with the growth rate of the number of periodic points, we first give a lemma which will turn out to be useful over and over again. Although it is a purely analytic result, we include its proof here for completeness. This result will be crucial not only here but also in Chapters 7 and 11. Recall that a sequence $(a_n)_{n=1}^\infty$ of real numbers is said to be *subadditive* if

$$a_{m+n} \le a_m + a_n, \quad \forall m, n \in \mathbb{N}.$$

Lemma 3.2.17. *If $(a_n)_{n=1}^\infty$ is a subadditive sequence of real numbers, then the sequence $(a_n/n)_{n=1}^\infty$ converges and*

$$\lim_{n\to\infty} \frac{1}{n} a_n = \inf_{n\in\mathbb{N}} \frac{1}{n} a_n.$$

If, moreover, $(a_n)_{n=1}^\infty$ is bounded from below, then $\inf_{n\in\mathbb{N}} \frac{1}{n} a_n \ge 0$.

Proof. Fix $m \in \mathbb{N}$. By the division algorithm, every $n \in \mathbb{N}$ can be uniquely written in the form $n = km + r$, where $0 \le r < m$. The subadditivity of the sequence implies that

$$\frac{a_n}{n} = \frac{a_{km+r}}{km+r} \le \frac{a_{km} + a_r}{km+r} \le \frac{ka_m + a_r}{km} = \frac{a_m}{m} + \frac{a_r}{km}.$$

Notice that for all $n \in \mathbb{N}$,

$$-\infty < \min_{0\le s<m} a_s \le a_r \le \max_{0\le s<m} a_s < \infty.$$

Therefore, as n tends to infinity, k also tends to infinity and thereby a_r/k approaches zero by the sandwich theorem. Hence,

$$\limsup_{n\to\infty}\frac{a_n}{n}\le\frac{a_m}{m}.$$

Since $m\in\mathbb{N}$ was chosen arbitrarily, taking the infimum over m yields that

$$\limsup_{n\to\infty}\frac{a_n}{n}\le\inf_{m\in\mathbb{N}}\frac{a_m}{m}.$$

Thus,

$$\limsup_{n\to\infty}\frac{a_n}{n}\le\inf_{m\in\mathbb{N}}\frac{a_m}{m}\le\liminf_{n\to\infty}\frac{a_n}{n}\le\limsup_{n\to\infty}\frac{a_n}{n}.$$

Consequently,

$$\lim_{n\to\infty}\frac{a_n}{n}=\inf_{m\in\mathbb{N}}\frac{a_m}{m}.$$

This proves the first assertion. The second one is obvious. □

Another purely analytic result that will be needed later is the following.

Lemma 3.2.18. *Let $(b_n)_{n=1}^\infty$ be a sequence of positive real numbers. Then*

$$\limsup_{n\to\infty}\frac{1}{n}\log b_n=\inf\left\{p\in\mathbb{R}:\sum_{n=1}^\infty b_ne^{-pn}<\infty\right\}.$$

Proof. First, assume that $-\infty<\limsup_{n\to\infty}\frac{1}{n}\log b_n$. Let $P\in\mathbb{R}$ be such that $P<\limsup_{n\to\infty}\frac{1}{n}\log b_n$. Then there exists a strictly increasing sequence $(n_m)_{m=1}^\infty$ of positive integers such that $P\le\frac{1}{n_m}\log b_{n_m}$ for all $m\in\mathbb{N}$. That is, $b_{n_m}\ge e^{Pn_m}$ for all $m\in\mathbb{N}$. Consequently,

$$\sum_{n=1}^\infty b_ne^{-Pn}\ge\sum_{m=1}^\infty b_{n_m}e^{-Pn_m}\ge\sum_{m=1}^\infty 1=\infty.$$

This implies that $P\le\inf\{p\in\mathbb{R}:\sum_{n=1}^\infty b_ne^{-pn}<\infty\}$. Since this is true for every $P<\limsup_{n\to\infty}\frac{1}{n}\log b_n$, we deduce that

$$\limsup_{n\to\infty}\frac{1}{n}\log b_n\le\inf\left\{p\in\mathbb{R}:\sum_{n=1}^\infty b_ne^{-pn}<\infty\right\}.\tag{3.2}$$

Obviously, this latter inequality holds as well when $-\infty=\limsup_{n\to\infty}\frac{1}{n}\log b_n$.

To prove the opposite inequality, suppose that $\limsup_{n\to\infty}\frac{1}{n}\log b_n<\infty$. Let $P\in\mathbb{R}$ be such that $\limsup_{n\to\infty}\frac{1}{n}\log b_n<P$. Let $Q\in\mathbb{R}$ be such that

$$\limsup_{n\to\infty}\frac{1}{n}\log b_n<Q<P.$$

Then there exists $N \in \mathbb{N}$ such that $\frac{1}{n}\log b_n \le Q$ for all $n \ge N$. That is, $b_n \le e^{Qn}$ for all $n \ge N$. It follows that

$$\begin{aligned}\sum_{n=1}^{\infty} b_n e^{-Pn} &= \sum_{n=1}^{N-1} b_n e^{-Pn} + \sum_{n=N}^{\infty} b_n e^{-Pn}\\ &\le \sum_{n=1}^{N-1} b_n e^{-Pn} + \sum_{n=1}^{\infty} e^{(Q-P)n}\\ &= \sum_{n=1}^{N-1} b_n e^{-Pn} + \sum_{n=1}^{\infty} (e^{Q-P})^n\\ &< \infty\end{aligned}$$

since the geometric series on the right has for ratio $0 < r := e^{Q-P} < 1$. This implies that $\inf\{p \in \mathbb{R} : \sum_{n=1}^{\infty} b_n e^{-pn} < \infty\} \le P$. Since this is true for every $P \in \mathbb{R}$ such that $\limsup_{n\to\infty} \frac{1}{n}\log b_n < P$, we deduce that

$$\inf\left\{p \in \mathbb{R} : \sum_{n=1}^{\infty} b_n e^{-pn} < \infty\right\} \le \limsup_{n\to\infty} \frac{1}{n}\log b_n. \tag{3.3}$$

Obviously, this latter inequality holds as well when $\limsup_{n\to\infty} \frac{1}{n}\log b_n = \infty$.

The result follows from (3.2) and (3.3). □

Let us now come back to the question of the number of periodic points that subshifts of finite type have. The following result holds for all subshifts of finite type. It relates the number of periodic points to the number of finite words, which in turn is related to the underlying matrix. Recall that a sequence $(a_n)_{n=1}^{\infty}$ of real numbers is said to be *submultiplicative* if

$$a_{m+n} \le a_m a_n, \quad \forall m, n \in \mathbb{N}.$$

Note that a sequence $(a_n)_{n=1}^{\infty}$ of positive real numbers is submultiplicative if and only if the sequence $(\log a_n)_{n=1}^{\infty}$ is subadditive.

Lemma 3.2.19. *Let A be an incidence matrix that generates a subshift of finite type $\sigma : E_A^{\infty} \to E_A^{\infty}$. The sequences $(\#E_A^n)_{n=1}^{\infty}$ and $(\|A^n\|)_{n=1}^{\infty}$ are nondecreasing and submultiplicative. Moreover,*

$$\|A^{n-1}\| = \#E_A^n \ge \#\mathrm{Per}_n(\sigma), \quad \forall n \in \mathbb{N}.$$

Proof. Since the matrix A has a 1 in every row by our standing assumption, every admissible word of length n can be extended to an admissible word of length $n+1$. Therefore,

$$\#E_A^n \le \#E_A^{n+1}, \quad \forall n \in \mathbb{N},$$

that is, the sequence $(\#E_A^n)_{n=1}^{\infty}$ is nondecreasing.

Regarding the submultiplicativity of that sequence, notice that each admissible word of length $m+n$ results from an admissible concatenation of an admissible word of length m with an admissible word of length n. This implies that

$$\#E_A^{m+n} \le \#E_A^m \cdot \#E_A^n, \quad \forall m, n \in \mathbb{N}.$$

(However, notice that the concatenation of an admissible word ω of length m with an admissible word τ of length n is an admissible word of length $m+n$ if and only if $\omega_m = \tau_1$. Thus, the equality does not hold in general.)

The nondecreasing and submultiplicative behaviors of the sequence $(\#E_A^n)_{n=1}^\infty$ are shared by the sequence $(\|A^n\|)_{n=1}^\infty$. Indeed, we have earlier observed that $(A^n)_{ij}$ is the number of admissible words of length $n+1$ starting with the letter i and ending with the letter j. Thus, $\sum_{i,j=1}^{\#E}(A^n)_{ij}$ is the number of words in E_A^{n+1}. This means that

$$\|A^n\| = \sum_{i,j=1}^{\#E} (A^n)_{ij} = \#E_A^{n+1}, \quad \forall n \in \mathbb{N}.$$

It is then obvious that

$$\|A^n\| = \#E_A^{n+1} \le \#E_A^{n+2} = \|A^{n+1}\|, \quad \forall n \in \mathbb{N},$$

that is, the sequence $(\|A^n\|)_{n=1}^\infty$ is nondecreasing. Submultiplicativity of that sequence follows from the fact that for all $m, n \in \mathbb{N}$,

$$\|A^{m+n}\| = \#E_A^{m+n+1} \le \#E_A^{m+n+2} \le \#E_A^{m+1} \cdot \#E_A^{n+1} = \|A^m\| \cdot \|A^n\|.$$

Moreover, note that every periodic point of period n is the infinitely repeated concatenation of its initial block of length n. This means that

$$\#\mathrm{Per}_n(\sigma) \le \#E_A^n = \|A^{n-1}\|, \quad \forall n \in \mathbb{N}. \qquad \square$$

Remark 3.2.20. In general, the sequence $(\#\mathrm{Per}_n(\sigma))_{n=1}^\infty$ is neither nondecreasing nor submultiplicative. See Exercise 3.4.20.

We can now obtain some information on the growth rate of the number of periodic points.

Theorem 3.2.21. *Let A be an incidence matrix that generates a subshift of finite type $\sigma : E_A^\infty \to E_A^\infty$. Then*

$$\begin{aligned}\log r(A) &= \lim_{n\to\infty} \frac{1}{n} \log \#E_A^n = \inf\Big\{p > 0 : \sum_{\omega \in E_A^*} e^{-p|\omega|} < \infty\Big\} \\ &\ge \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma) = \inf\Big\{p > 0 : \sum_{n=1}^{\infty} \#\mathrm{Per}_n(\sigma) \cdot e^{-pn} < \infty\Big\},\end{aligned}$$

where $r(A)$ is the spectral radius of A.

Proof. It follows from Lemma 3.2.19 that both of the sequences $(\log \#E_A^n)_{n=1}^\infty$ and $(\log \|A^n\|)_{n=1}^\infty$ are subadditive. By Lemma 3.2.17, we have that both limits $\lim_{n\to\infty} \frac{1}{n} \log \|A^n\|$ and $\lim_{n\to\infty} \frac{1}{n} \log \#E_A^n$ exist. Using Lemma 3.2.19, we deduce that

$$\begin{aligned}
\log r(A) &= \log \lim_{n\to\infty} \|A^n\|^{1/n} \\
&= \lim_{n\to\infty} \log \|A^n\|^{1/n} \\
&= \lim_{n\to\infty} \frac{1}{n} \log \|A^n\| \\
&= \lim_{n\to\infty} \frac{1}{n} \log \#E_A^{n+1} \\
&= \lim_{n\to\infty} \frac{n+1}{n} \cdot \frac{1}{n+1} \log \#E_A^{n+1} \\
&= \lim_{n\to\infty} \frac{n+1}{n} \cdot \lim_{n\to\infty} \frac{1}{n+1} \log \#E_A^{n+1} \\
&= \lim_{n\to\infty} \frac{1}{n} \log \#E_A^n \\
&\geq \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma).
\end{aligned}$$

From Lemma 3.2.18, it follows that

$$\limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma) = \inf\left\{ p \in \mathbb{R} : \sum_{n=1}^\infty \#\mathrm{Per}_n(\sigma) \cdot e^{-pn} < \infty \right\}$$

and

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log \#E_A^n &= \inf\left\{ p \in \mathbb{R} : \sum_{n=1}^\infty \#E_A^n e^{-pn} < \infty \right\} \\
&= \inf\left\{ p \in \mathbb{R} : \sum_{n=1}^\infty \sum_{\omega\in E_A^n} e^{-p|\omega|} < \infty \right\} \\
&= \inf\left\{ p \in \mathbb{R} : \sum_{\omega\in E_A^*} e^{-p|\omega|} < \infty \right\}.
\end{aligned}$$

Finally, note that the infima can be restricted to the positive real numbers p since the set E_A^* is infinite and the sets $\mathrm{Per}_n(\sigma)$ are nonempty for infinitely many n. □

The inequality in the statement of the previous theorem turns out to be an equality. We will demonstrate this in two steps. It is first simpler to prove it when the matrix A is irreducible.

Theorem 3.2.22. *Let A be an irreducible matrix. Then*

$$\begin{aligned}
\log r(A) = \lim_{n\to\infty} \frac{1}{n} \log \#E_A^n &= \inf\left\{ p > 0 : \sum_{\omega\in E_A^*} e^{-p|\omega|} < \infty \right\} \\
= \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma) &= \inf\left\{ p > 0 : \sum_{n=1}^\infty \#\mathrm{Per}_n(\sigma) \cdot e^{-pn} < \infty \right\}.
\end{aligned}$$

Proof. Since A is irreducible, there exists $p \geq 2$ such that for every $1 \leq i, j \leq \#E$ there is an admissible word of length at least 2 and at most p that begins with i and ends with j. Therefore, for any admissible word ω of length n there is an admissible word $\widetilde{\omega}$ of length at least 2 and at most p beginning with $\widetilde{\omega}_1 = \omega_n$ and ending with $\widetilde{\omega}_{|\widetilde{\omega}|} = \omega_1$. Consequently, the word $(\omega_1\omega_2\ldots\omega_n\widetilde{\omega}_2\widetilde{\omega}_3\ldots\widetilde{\omega}_{|\widetilde{\omega}|-1})^\infty$ is an admissible periodic point of period $n + |\widetilde{\omega}| - 2$, with $n \leq n + |\widetilde{\omega}| - 2 \leq n + p - 2$. This shows that every $\omega \in E_A^n$ generates at least one periodic point whose period is between n and $n + p - 2$, with different words $\omega \in E_A^n$ producing different periodic points $(\omega\widetilde{\omega}_2\widetilde{\omega}_3\ldots\widetilde{\omega}_{|\widetilde{\omega}|-1})^\infty$. Thus, if for every $n \in \mathbb{N}$ we choose $m(n)$ to be such that $n \leq m(n) \leq n + p - 2$ and

$$\#\mathrm{Per}_{m(n)}(\sigma) = \max_{n \leq m \leq n+p-2} \#\mathrm{Per}_m(\sigma),$$

we obtain that

$$\begin{aligned}\|A^{n-1}\| = \#E_A^n &\leq \#\mathrm{Per}_n(\sigma) + \#\mathrm{Per}_{n+1}(\sigma) + \cdots + \#\mathrm{Per}_{n+p-2}(\sigma)\\ &\leq (p-1)\#\mathrm{Per}_{m(n)}(\sigma).\end{aligned}$$

Using this estimate, we can make the following calculation:

$$\begin{aligned}\log r(A) &= \lim_{n\to\infty} \frac{1}{n-1}\log\|A^{n-1}\| \leq \limsup_{n\to\infty}\frac{1}{n-1}\log\big[(p-1)\cdot\#\mathrm{Per}_{m(n)}(\sigma)\big]\\ &= \limsup_{n\to\infty}\frac{1}{n-1}\big[\log(p-1) + \log\#\mathrm{Per}_{m(n)}(\sigma)\big]\\ &= \lim_{n\to\infty}\frac{1}{n-1}\log(p-1) + \limsup_{n\to\infty}\frac{1}{n-1}\log\#\mathrm{Per}_{m(n)}(\sigma)\\ &= 0 + \limsup_{n\to\infty}\left[\frac{m(n)}{n-1}\cdot\frac{1}{m(n)}\log\#\mathrm{Per}_{m(n)}(\sigma)\right]\\ &= \lim_{n\to\infty}\frac{m(n)}{n-1}\cdot\limsup_{n\to\infty}\frac{1}{m(n)}\log\#\mathrm{Per}_{m(n)}(\sigma)\\ &\leq 1\cdot\limsup_{n\to\infty}\frac{1}{n}\log\#\mathrm{Per}_n(\sigma)\\ &= \limsup_{n\to\infty}\frac{1}{n}\log\#\mathrm{Per}_n(\sigma).\end{aligned}$$

Combining this inequality with the opposite one proved in Theorem 3.2.21 completes the proof. □

In order to prove that the equality holds in full generality, we need to decompose reducible systems into maximal (in the sense of inclusion) irreducible subsystems. For this, we separate the letters of E into equivalence classes by means of a relation called *communication*.

Definition 3.2.23.

(a) A letter $i \in E$ *leads to* a letter $j \in E$ if there exists $p = p(i,j) \in \mathbb{N}$ such that $(A^p)_{ij} = 1$.
(b) A letter i is said to *communicate* with a letter j if i leads to j and j leads to i.

(c) A letter which communicates with itself or any other letter is called *communicating*.
(d) Otherwise, the letter is said to be *noncommunicating*.

The relation of communication defines an equivalence relation on the set of communicating letters. The corresponding equivalence classes are called *communication classes*.

Theorem 3.2.24. *Let A be an incidence matrix that generates a subshift of finite type $\sigma : E_A^\infty \to E_A^\infty$. Then*

$$\begin{aligned}
\log r(A) &= \lim_{n\to\infty} \frac{1}{n}\log \#E_A^n = \inf\Big\{p > 0 : \sum_{\omega\in E_A^*} e^{-p|\omega|} < \infty\Big\}\\
&= \max_{\substack{C\\ \textit{comm. class}}} \lim_{n\to\infty}\frac{1}{n}\log\#C_A^n\\
&= \max_{\substack{C\\ \textit{comm. class}}} \limsup_{n\to\infty}\frac{1}{n}\log\#\mathrm{Per}_n(\sigma_C)\\
&= \limsup_{n\to\infty}\frac{1}{n}\log\#\mathrm{Per}_n(\sigma) = \inf\Big\{p > 0 : \sum_{n=1}^{\infty}\#\mathrm{Per}_n(\sigma)\cdot e^{-pn} < \infty\Big\}\\
&= \max_{\substack{C\\ \textit{comm. class}}} \log r(A|_C).
\end{aligned}$$

Proof. Suppose that E admits k communication classes $C_1, C_2, \ldots, C_k$. Clearly, $(C_l)_A^\infty \subseteq E_A^\infty$ for each $1 \le l \le k$. Moreover, $(C_l)_A^\infty \cap (C_m)_A^\infty = \emptyset$ for all $l \ne m$ since $C_l \cap C_m = \emptyset$ for all $l \ne m$. Note further that the submatrix $A|_{C_l} : C_l \times C_l \to \{0,1\}$ is irreducible for each $1 \le l \le k$ by the very definition of communication classes. Therefore, Theorem 3.2.22 asserts that

$$\begin{aligned}
\log r(A|_{C_l}) &= \lim_{n\to\infty}\frac{1}{n}\log\#(C_l)_A^n = \inf\Big\{p > 0 : \sum_{\omega\in(C_l)_A^*} e^{-p|\omega|} < \infty\Big\}\\
&= \limsup_{n\to\infty}\frac{1}{n}\log\#\mathrm{Per}_n(\sigma_{C_l})
\end{aligned}\tag{3.4}$$

for every $1 \le l \le k$.

Let $I = E\backslash\bigcup_{l=1}^k C_l$. This set consists of all noncommunicating letters. If $\omega \in E_A^*$, then each noncommunicating letter can appear at most once in ω. Moreover, if ω contains at least one letter from the class C_l then ω can be uniquely written as $\beta_l\alpha_l\gamma_l$, where $\alpha_l \in (C_l)_A^*$ and $\beta_l, \gamma_l \in (E \setminus C_l)_A^*$, that is, α_l comprises only letters from the class C_l while none of the letters of β_l and γ_l are from C_l. Note that β_l and/or γ_l may be the empty word, while α_l is the longest subword of ω that has letters from C_l only. For each $\omega \in E_A^*$ and $1 \le l \le k$, let $\alpha_l(\omega)$ be the longest subword of ω in $(C_l)_A^*$. Note that $\alpha_l(\omega)$ may be the empty word for some l's. Then ω can be uniquely written as a concatenation of the subwords $\alpha_l(\omega)$, $1 \le l \le k$, and no more than $k+1$ subwords of noncommunicating letters, each of which consists of at most $\#I$ letters. Therefore, the map

$$\begin{array}{ccccccccccc} \alpha & : & E_A^* & \longrightarrow & (C_1)_A^* & \times & (C_2)_A^* & \times & \cdots & \times & (C_k)_A^* \\ & & \omega & \longmapsto & \big(\alpha_1(\omega) & , & \alpha_2(\omega) & , & \ldots & , & \alpha_k(\omega)\big) \end{array}$$

is such that each element of $(C_1)_A^* \times \cdots \times (C_k)_A^*$ has at most $(\#I \cdot \#I! + 1)^{k+1}$ preimages. Indeed, suppose that $\alpha(\tau) = \alpha(\omega)$ for some $\tau, \omega \in E_A^*$. Then $\alpha_l(\tau) = \alpha_l(\omega) =: \alpha_l$ for all $1 \le l \le k$. If $\alpha_l \ne \epsilon \ne \alpha_m$ for some $l \ne m$ and if τ contained the subword $\alpha_l\beta\alpha_m$ whereas ω contained the subword $\alpha_m\gamma\alpha_l$, then the word $\alpha_l\beta\alpha_m\gamma\alpha_l$ would be in E_A^*. This would imply that the classes C_l and C_m communicate, which would contradict their very definition. This reveals that the α_l, $1 \le l \le k$, must appear in the same order in both τ and ω. That is, τ and ω can only differ in the subwords of noncommunicating letters they contain. Now, there are at most $k+1$ subwords of noncommunicating letters in any word. And each of these subwords contains at most $\#I$ letters. Let $1 \le L \le \#I$. The number of words of length L with distinct letters drawn from I is at most $\#I \cdot (\#I - 1) \cdots (\#I - L + 1) \le \#I!$. Thus the number of nonempty words of length at most $\#I$ with distinct letters drawn from I is at most $\#I \cdot \#I!$. Add 1 for the empty word. This is an upper estimate of the number of possibilities for each instance of a subword of noncommunicating letters. Since there are at most $k+1$ such instances, a crude upper bound on the number of preimages for any point is $B := (\#I \cdot \#I! + 1)^{k+1}$.

For all $p > 0$, it then follows that

$$\begin{aligned} \sum_{\omega \in E_A^*} e^{-p|\omega|} &\le \sum_{\omega \in E_A^*} e^{-p\sum_{l=1}^k |\alpha_l(\omega)|} \\ &= \sum_{\omega \in E_A^*} \prod_{l=1}^k e^{-p|\alpha_l(\omega)|} \\ &\le B \prod_{l=1}^k \sum_{\omega_l \in (C_l)_A^*} e^{-p|\omega_l|}. \end{aligned} \tag{3.5}$$

Let

$$P > \max_{1 \le l \le k} \lim_{n \to \infty} \frac{1}{n} \log \#(C_l)_A^n.$$

We infer from (3.4) that

$$\sum_{\omega_l \in (C_l)_A^*} e^{-P|\omega_l|} < \infty, \quad \forall 1 \le l \le k$$

and thus by (3.5) we deduce that

$$\sum_{\omega \in E_A^*} e^{-P|\omega|} < \infty.$$

According to Theorem 3.2.21, this implies that

$$P > \lim_{n\to\infty} \frac{1}{n} \log \#E_A^n.$$

Since this is true for every $P > \max_{1\le l\le k} \lim_{n\to\infty} \frac{1}{n} \log \#(C_l)_A^n$, we obtain that

$$\lim_{n\to\infty} \frac{1}{n} \log \#E_A^n \le \max_{1\le l\le k} \lim_{n\to\infty} \frac{1}{n} \log \#(C_l)_A^n.$$

The opposite inequality is obvious. Hence,

$$\lim_{n\to\infty} \frac{1}{n} \log \#E_A^n = \max_{1\le l\le k} \lim_{n\to\infty} \frac{1}{n} \log \#(C_l)_A^n. \tag{3.6}$$

On the other hand, note that $\mathrm{Per}(\sigma_{C_l}) \cap \mathrm{Per}(\sigma_{C_m}) = \emptyset$ for all $l \ne m$ since $C_l \cap C_m = \emptyset$ for all $l \ne m$. Moreover, since a periodic point can comprise neither noncommunicating letters nor letters from two distinct communicating classes, we have that

$$\mathrm{Per}(\sigma_E) = \bigcup_{l=1}^{k} \mathrm{Per}(\sigma_{C_l}).$$

Therefore,

$$\max_{1\le l\le k} \#\mathrm{Per}_n(\sigma_{C_l}) \le \#\mathrm{Per}_n(\sigma_E) = \sum_{l=1}^{k} \#\mathrm{Per}_n(\sigma_{C_l}) \le k \max_{1\le l\le k} \#\mathrm{Per}_n(\sigma_{C_l}).$$

It follows immediately that

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log \max_{1\le l\le k} \#\mathrm{Per}_n(\sigma_{C_l}) &\le \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma_E)\\ &\le \limsup_{n\to\infty} \frac{1}{n} \log\Big[k \max_{1\le l\le k} \#\mathrm{Per}_n(\sigma_{C_l})\Big].\end{aligned}$$

Using Exercise 3.4.14, it follows that

$$\begin{aligned}\max_{1\le l\le k} \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma_{C_l}) &\le \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma_E)\\ &\le \lim_{n\to\infty} \frac{\log k}{n} + \max_{1\le l\le k} \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma_{C_l}).\end{aligned}$$

Hence,

$$\limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma_E) = \max_{1\le l\le k} \limsup_{n\to\infty} \frac{1}{n} \log \#\mathrm{Per}_n(\sigma_{C_l}). \tag{3.7}$$

Using Theorem 3.2.21 and relations (3.4), (3.6), and (3.7), the result follows. □

The use of the lim sup in Theorem 3.2.22 is indispensable. Indeed, there are transitive subshifts of finite type for which the limit does not exist (see Exercise 3.4.20). However, the limit does exist for all topologically exact subshifts of finite type.

Theorem 3.2.25. *Let A be a primitive matrix. Then*

$$\begin{aligned}\log r(A) &= \lim_{n\to\infty}\frac{1}{n}\log \#E_A^n = \inf\Big\{p>0 : \sum_{\omega\in E_A^*} e^{-p|\omega|}<\infty\Big\}\\ &= \lim_{n\to\infty}\frac{1}{n}\log \#\mathrm{Per}_n(\sigma) = \inf\Big\{p>0 : \sum_{n=1}^{\infty} \#\mathrm{Per}_n(\sigma)\cdot e^{-pn}<\infty\Big\}.\end{aligned}$$

Proof. In Lemma 3.2.19, we observed that $\#\mathrm{Per}_n(\sigma) \le \#E_A^n = \|A^{n-1}\|$ for all $n \in \mathbb{N}$. Now we use the primitivity of the matrix to establish a similar inequality in the other direction. Since A is primitive, there exists $P \in \mathbb{N}$ such that $A^P > 0$. Consequently, for every $1 \le i, j \le \#E$ there is an admissible word of length $P+1$ that begins with i and ends with j. Therefore, for any $\omega \in E_A^n$ there is a word $\widetilde{\omega} \in E_A^{P+1}$ which begins with $\widetilde{\omega}_1 = \omega_n$ and ends with $\widetilde{\omega}_{P+1} = \omega_1$. Then the word $(\omega_1\omega_2\ldots\omega_n\widetilde{\omega}_2\widetilde{\omega}_3\ldots\widetilde{\omega}_P)^\infty$ is an admissible periodic point of period $n+P-1$. This shows that every $\omega \in E_A^n$ generates at least one periodic point of period $n+P-1$, with different words $\omega \in E_A^n$ producing different periodic points $(\omega\widetilde{\omega}_2\widetilde{\omega}_3\ldots\widetilde{\omega}_P)^\infty \in \mathrm{Per}_{n+P-1}(\sigma)$. Hence, $\#E_A^n \le \#\mathrm{Per}_{n+P-1}(\sigma)$. Using this and Lemma 3.2.19, we get

$$\#E_A^n \le \#\mathrm{Per}_{n+P-1}(\sigma) \le \#E_A^{n+P-1} \le \#E_A^n\cdot\#E_A^{P-1}, \quad \forall n\in\mathbb{N}.$$

Hence,

$$\frac{1}{n}\log\#E_A^n \le \frac{1}{n}\log\#\mathrm{Per}_{n+P-1}(\sigma) \le \frac{1}{n}\log\#E_A^n + \frac{1}{n}\log\#E_A^{P-1}, \quad \forall n\in\mathbb{N}.$$

It follows from the squeeze theorem that

$$\lim_{n\to\infty}\frac{1}{n}\log\#\mathrm{Per}_{n+P-1}(\sigma) = \lim_{n\to\infty}\frac{1}{n}\log\#E_A^n.$$

Therefore,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\log\#\mathrm{Per}_n(\sigma) &= \lim_{n\to\infty}\frac{1}{n+P-1}\log\#\mathrm{Per}_{n+P-1}(\sigma)\\ &= \lim_{n\to\infty}\Big[\frac{n}{n+P-1}\cdot\frac{1}{n}\log\#\mathrm{Per}_{n+P-1}(\sigma)\Big]\\ &= \lim_{n\to\infty}\frac{n}{n+P-1}\cdot\lim_{n\to\infty}\frac{1}{n}\log\#\mathrm{Per}_{n+P-1}(\sigma)\\ &= \lim_{n\to\infty}\frac{1}{n}\log\#E_A^n.\end{aligned}$$

The rest follows from Theorem 3.2.22. ☐

3.3 General subshifts of finite type

Recall that a subshift of finite type is a subshift that can be described by a finite set $\mathcal{F}$ of forbidden words. In this case, the set $\mathcal{F}$ can be chosen so that $\mathcal{F} \subseteq E^q$ for some $q \in \mathbb{N}$. The set of forbidden words then induces a function A from E^q to $\{0,1\}$, where the function A takes the value 0 on the set $\mathcal{F}$ of forbidden words and takes the value 1 on the set $E^q \setminus \mathcal{F}$ of all admissible words. For the sake of simplicity, we concentrated on the case $q = 2$ in the previous section. We then pointed out that we would prove that all cases can be reduced to that case. We now do so.

Fix an integer $q \geq 2$ and a function $A : E^q \to \{0,1\}$. Let

$$E_A^\infty := \{\omega \in E^\infty : A(\omega_n, \omega_{n+1}, \dots, \omega_{n+q-1}) = 1, \forall n \in \mathbb{N}\}.$$

E_A^∞ is a subshift of finite type since it consists of all those infinite words that do not contain any word from the finite set of forbidden words

$$\mathcal{F} = \{\omega \in E^q : A(\omega_1, \omega_2, \dots, \omega_q) = 0\}.$$

Theorem 3.3.1. *The shift map $\sigma : E_A^\infty \to E_A^\infty$ is topologically conjugate to a shift map $\widetilde{\sigma} : \widetilde{E}_{\widetilde{A}}^\infty \to \widetilde{E}_{\widetilde{A}}^\infty$, where $\#\widetilde{E} = (\#E)^q$ and $\widetilde{A}$ is an incidence matrix on $\widetilde{E}$.*

Proof. Set $\widetilde{E} := E^q$ as an alphabet. In other words, a letter in the alphabet $\widetilde{E}$ is a word of length q over the alphabet E. Define the incidence matrix $\widetilde{A} : \widetilde{E} \times \widetilde{E} \to \{0,1\}$ by

$$\widetilde{A}_{\tau\rho} = \begin{cases} 1 & \text{if } A(\tau) = 1 = A(\rho) \text{ and } \tau_2\tau_3\dots\tau_q = \rho_1\rho_2\dots\rho_{q-1} \\ 0 & \text{otherwise} \end{cases}$$

for every $\tau, \rho \in \widetilde{E}$. Let

$$\widetilde{E}_{\widetilde{A}}^\infty := \{\widetilde{\omega} \in \widetilde{E}^\infty : \widetilde{A}_{\widetilde{\omega}_n\widetilde{\omega}_{n+1}} = 1,\ \forall n \in \mathbb{N}\}$$

be the subshift of finite type generated by the matrix $\widetilde{A}$. Define the map $H : E_A^\infty \to \widetilde{E}$ by $H(\omega) = \omega|_q$. That is, the map H associates to every A-admissible infinite word its initial subword of length q. Let $h : E_A^\infty \to \widetilde{E}_{\widetilde{A}}^\infty$ be defined by the concatenation

$$h(\omega) = H(\omega)H(\sigma(\omega))H(\sigma^2(\omega))\dots.$$

In other words, for any $\omega \in E_A^\infty$, we have that

$$h(\omega) = (\omega_1 \dots \omega_q)(\omega_2 \dots \omega_{q+1})(\omega_3 \dots \omega_{q+2}) \dots \in \widetilde{E}_{\widetilde{A}}^\infty.$$

We claim that h is a homeomorphism. Indeed, h is injective since if ω and τ are two distinct elements of E_A^∞, then $\omega_k \neq \tau_k$ for some $k \in \mathbb{N}$. It immediately follows that $\sigma^{k-1}(\omega)_1 = \omega_k \neq \tau_k = \sigma^{k-1}(\tau)_1$, and hence, by definition, $H(\sigma^{k-1}(\omega)) \neq H(\sigma^{k-1}(\tau))$. Thus, $h(\omega)_k \neq h(\tau)_k$ and so $h(\omega) \neq h(\tau)$. To show that the map h is surjective, let

$\widetilde{\tau} \in \widetilde{E}^\infty_{\widetilde{A}}$ be arbitrary. Recall that $\widetilde{\tau}_k \in \widetilde{E} = E^q$ for each $k \in \mathbb{N}$, that is, $\widetilde{\tau}_k$ is a word of length q from the alphabet E. Also bear in mind that by the definition of $\widetilde{E}^\infty_{\widetilde{A}}$, the word consisting of the last $q-1$ letters of $\widetilde{\tau}_k$ is equal to the initial $(q-1)$-word of $\widetilde{\tau}_{k+1}$. Construct the infinite word τ by concatenating the first letters of each word $\widetilde{\tau}_k$ in turn, that is,

$$\tau = (\widetilde{\tau}_1)_1(\widetilde{\tau}_2)_1(\widetilde{\tau}_3)_1 \ldots .$$

Then, for every $n \in \mathbb{N}$, we have that

$$\begin{aligned} A(\tau_n, \tau_{n+1}, \ldots, \tau_{n+q-1}) &= A\big((\widetilde{\tau}_n)_1, (\widetilde{\tau}_{n+1})_1, \ldots, (\widetilde{\tau}_{n+q-1})_1\big) \\ &= A\big((\widetilde{\tau}_n)_1, (\widetilde{\tau}_n)_2, \ldots, (\widetilde{\tau}_n)_q\big) \\ &= A(\widetilde{\tau}_n) \\ &= 1. \end{aligned}$$

Therefore, $\tau \in E^\infty_A$. Furthermore,

$$\begin{aligned} h(\tau) &= H(\tau)H(\sigma(\tau))H(\sigma^2(\tau))\ldots \\ &= \big((\widetilde{\tau}_1)_1(\widetilde{\tau}_2)_1\ldots(\widetilde{\tau}_q)_1\big)\big((\widetilde{\tau}_2)_1(\widetilde{\tau}_3)_1\ldots(\widetilde{\tau}_{q+1})_1\big)\big((\widetilde{\tau}_3)_1(\widetilde{\tau}_4)_1\ldots(\widetilde{\tau}_{q+2})_1\big)\ldots \\ &= \big((\widetilde{\tau}_1)_1(\widetilde{\tau}_1)_2\ldots(\widetilde{\tau}_1)_q\big)\big((\widetilde{\tau}_2)_1(\widetilde{\tau}_2)_2\ldots(\widetilde{\tau}_2)_q\big)\big((\widetilde{\tau}_3)_1(\widetilde{\tau}_3)_2\ldots(\widetilde{\tau}_3)_q\big)\ldots \\ &= (\widetilde{\tau}_1)(\widetilde{\tau}_2)(\widetilde{\tau}_3)\ldots \\ &= \widetilde{\tau}. \end{aligned}$$

Since $\widetilde{\tau}$ is arbitrary, this demonstrates that h is surjective.

Moreover, h is continuous. To see this, let $\omega, \tau \in E^\infty_A$. Denote the length of their wedge by $Q = |\omega \wedge \tau|$. If $Q \geq q$, then $|h(\omega) \wedge h(\tau)| = Q - q + 1$, for if ω and τ share the same first Q letters, then $h(\omega)$ and $h(\tau)$ share the same first $Q - q + 1$ letters. The fact that h is a homeomorphism follows from the fact that it is a continuous bijection between two Hausdorff compact topological spaces. It only remains to show that the following diagram commutes:

$$\begin{tikzcd} E^\infty_A \arrow[r, "\sigma"] \arrow[d, "h"'] & E^\infty_A \arrow[d, "h"] \\ \widetilde{E}^\infty_{\widetilde{A}} \arrow[r, "\widetilde{\sigma}"'] & \widetilde{E}^\infty_{\widetilde{A}} \end{tikzcd}$$

Indeed, for every $\omega \in E^\infty_A$ we have

$$\begin{aligned} h \circ \sigma(\omega) &= H(\sigma(\omega))H(\sigma(\sigma(\omega)))H(\sigma^2(\sigma(\omega)))\ldots \\ &= H(\sigma(\omega))H(\sigma^2(\omega))H(\sigma^3(\omega))\ldots \\ &= \widetilde{\sigma}\big(H(\omega)H(\sigma(\omega))H(\sigma^2(\omega))\ldots\big) \\ &= \widetilde{\sigma} \circ h(\omega). \end{aligned}$$

This completes the proof. □

3.4 Exercises

Exercise 3.4.1. Fix $n \in \mathbb{N}$. Show that the full n-shift can be used to encode all the numbers between 0 and 1. That is, show that to every number in $[0,1]$ can be associated an infinite word in $\{0,\dots,n-1\}^{\infty}$.

Exercise 3.4.2. Prove that the discrete topology on a set X, that is, the topology in which every subset of X is both open and closed, is metrizable by means of the distance function $d : X \times X \to \{0,1\}$ defined by

$$d(x_1,x_2) = \begin{cases} 1 & \text{if } x_1 \neq x_2 \\ 0 & \text{if } x_1 = x_2. \end{cases}$$

Then show that the product $\prod_{n=1}^{\infty} X$ is metrizable.

Exercise 3.4.3. Show that the family $\{[\omega] : \omega \in E^*\}$ of all initial cylinders forms a base of open sets for Tychonoff's product topology on E^{∞}. Deduce that all cylinders $\{[\omega]_m^n : \omega \in E^*, m,n \in \mathbb{N}\}$ are both open and closed sets. Deduce further that the space E^{∞} is totally disconnected.

Exercise 3.4.4. Prove that the metrics d_s, $s \in (0,1)$, introduced in Definition 3.1.10 induce Tychonoff's product topology on E^{∞}.

Exercise 3.4.5. Prove that the metrics d_s, for each $s \in (0,1)$, are Hölder equivalent. That is, show that for any pair $s, s' \in (0,1)$ there is an exponent $\alpha \geq 0$ and a constant $C \geq 1$ such that

$$C^{-1}\big(d_{s'}(\omega,\tau)\big)^{\alpha} \leq d_s(\omega,\tau) \leq C\big(d_{s'}(\omega,\tau)\big)^{\alpha}, \quad \forall \omega,\tau \in E^{\infty}.$$

Exercise 3.4.6. Show that the metrics d_s, $s \in (0,1)$, are not Lipschitz equivalent. That is, prove that for any pair $s, s' \in (0,1)$ there is no constant $C \geq 1$ such that

$$C^{-1} d_{s'}(\omega,\tau) \leq d_s(\omega,\tau) \leq C d_{s'}(\omega,\tau), \quad \forall \omega,\tau \in E^{\infty}.$$

Exercise 3.4.7. Prove directly that the space E^{∞} is separable. That is, find a countable dense set in E^{∞}.

Exercise 3.4.8. Prove directly that E^{∞} is compact.

Hint: Since E^{∞} is metrizable, it suffices to prove that E^{∞} is sequentially compact. That is, it is sufficient to prove that every sequence in E^{∞} admits a convergent subsequence.

Exercise 3.4.9. Describe the subshift of finite type E_A^{∞} generated by the incidence matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

This shift is called the *golden mean shift*. (We will explain why in Chapter 7.)

Exercise 3.4.10. Find a closed shift-invariant subset F of $\{0,1\}^\infty$ such that $\sigma|_F : F \to F$ is not open.

Exercise 3.4.11. Prove that the shift map $\sigma : E_A^\infty \to E_A^\infty$ is injective if and only if every column of the incidence matrix A contains at most one 1.

Exercise 3.4.12. Show that the shift map $\sigma : E_A^\infty \to E_A^\infty$ may be transitive even if the incidence matrix A contains a row of zeros. (Thus, Theorem 3.2.14 does not hold if one does not assume that every row of A contains at least one 1.)

Exercise 3.4.13. Construct an irreducible matrix which is not primitive.

Exercise 3.4.14. Let $(a_n)_{n=1}^\infty$ and $(b_n)_{n=1}^\infty$ be sequences of real numbers. Prove that

$$\limsup_{n\to\infty} \max\{a_n, b_n\} = \max\Big\{\limsup_{n\to\infty} a_n, \limsup_{n\to\infty} b_n\Big\}.$$

Show a similar result for lim inf. Also, show that max can be replaced by min and that the corresponding statements hold. Finally, show that the statements do not necessarily hold for lim.
Note: Though this exercise has been stated with two sequences only, these statements hold for any finite number of sequences.

Exercise 3.4.15. In this exercise, you will prove that the shift $\sigma : E_A^\infty \to E_A^\infty$ is minimal if and only if A is a permutation matrix with a unique class of communicating letters.

(a) Relying upon our standing assumption that A has a 1 in each of its rows, prove that E_A^∞ admits a periodic point. (In fact, it is possible to show that the set of eventually periodic points is dense in E_A^∞.)
(b) Suppose that $\sigma : E_A^\infty \to E_A^\infty$ is minimal. Deduce from the minimality of σ that E_A^∞ coincides with the orbit of a periodic point.
(c) Deduce that σ is a bijection.
(d) Deduce that A is a permutation matrix.
(e) Show that A has a unique communicating class.

To prove the converse, suppose that A is a permutation matrix with a unique class of communicating letters.

(f) Prove that E_A^∞ coincides with the orbit of a periodic point.
(g) Deduce that $\sigma : E_A^\infty \to E_A^\infty$ is minimal.

Exercise 3.4.16. Let C and C' be classes of communicating letters for a matrix A. Class C is said to *lead* to class C' if one of the letters in C leads to one of the letters in C'. Prove that the set of periodic points of E_A^∞ is dense if and only if A consists of classes of communicating letters, none of which leads to another (in particular, A does not have any noncommunicating letter).

In other words, the set of periodic points of E_A^∞ is dense if and only if E_A^∞ is a disjoint union of irreducible subsystems.

Exercise 3.4.17. Let C be the middle-third Cantor set. Show that the map $f : [0,1] \to [0,1]$ defined by

$$f(x) = 3x \pmod 1$$

restricted to C is continuous. Show that $f|_C : C \to C$ is topologically conjugate to the full shift map on two symbols.

Exercise 3.4.18. Show that if the shift map $\sigma : E_A^\infty \to E_A^\infty$ is such that $\sigma([e] \cap E_A^\infty) = E_A^\infty \neq \emptyset$ for every $e \in E$, then $E_A^\infty = E^\infty$.

Exercise 3.4.19. Show that the full shift map $\sigma : \{0,1\}^\infty \to \{0,1\}^\infty$ has uncountably many points that are not transitive (i. e., with a nondense orbit).

Exercise 3.4.20. Construct a transitive subshift of finite type whose periodic points have even periods.

4 Distance expanding maps

In this chapter, we first define and give some examples of distance expanding maps, which, as their name suggests, expand distances between points. On a compact metric space, this behavior may only be observed locally. Accordingly, the definition of an expanding map involves two constants: a constant describing the magnitude of the expansion of the system under scrutiny, and a constant delimiting the neighborhoods on which the expansion can be observed.

Distance expanding maps were introduced in [61]. A fairly complete account of them can be found in [57]. Our approach stems from that work, but in many instances is much more detailed. Moreover, the proof of the existence of Markov partitions in Section 4.4 is substantially simplified.

In Section 4.2, we introduce the notion and study the properties of inverse branches of a distance expanding map. This is a way of dealing with the noninvertibility of these maps.

In Section 4.3, we describe two new concepts: pseudo-orbit and shadowing. The latter makes precise the fact that, given a measuring device of some prescribed accuracy, sequences of points which remain sufficiently close to one another cannot be distinguished by the said device.

Sections 4.4 and 4.5 are crucial. In the former, we introduce the concept of Markov partitions and their existence for open, distance expanding systems, while in the latter we show exactly how to use them to represent the dynamics of such systems by means of the symbolic dynamics studied in Chapter 3. The final theorem of the chapter describes the properties of the coding map between the underlying compact metric space (the phase space) and some subshift of finite type (a symbolic space).

The concept of Markov partition was introduced to dynamical systems by Adler, Konheim, and McAndrew in the paper [2] in 1965. It achieved its full significance in Rufus Bowen's book [11]. It is in this book that the existence of Markov partitions was proved for Axiom A diffeomorphisms and the corresponding symbolic representation/dynamics along with thermodynamic formalism were developed. Our approach, via the book [57], traces back to Bowen's work. Markov partitions, in their various forms, play an enormous role in the modern (that is, after Bowen) theory of dynamical systems.

4.1 Definition and examples

Definition 4.1.1. A continuous map $T : X \to X$ of a compact metric space (X, d) is called *distance expanding* provided that there exist two constants $\lambda > 1$ and $\delta > 0$ such that

$$d(x,y) < 2\delta \Longrightarrow d\big(T(x),T(y)\big) \geq \lambda d(x,y).$$

https://doi.org/10.1515/9783110702682-004

The use of 2δ in the above definition, as opposed to simply δ, is only to make the forthcoming expressions and calculations simpler.

Remark 4.1.2.
(a) If $T : X \to X$ is a distance expanding map, then for every forward T-invariant closed set $F \subseteq X$ the map $T|_F : F \to F$ is also distance expanding.
(b) If $d(x,y) < 2\delta$, then $d(T(x),T(y)) \geq \lambda d(x,y)$. Therefore, if $x \neq y$, we have that $T(x) \neq T(y)$. Thus, if $T(x) = T(y)$ and $x \neq y$, then $d(x,y) \geq 2\delta$. In particular, this demonstrates that any distance expanding map is locally injective.

Example 4.1.3 (Subshifts of finite type). The full shift map $\sigma : E^\infty \to E^\infty$ and all of its subshifts of finite type $\sigma : E_A^\infty \to E_A^\infty$ are expanding whichever metric d_s is used. More precisely, σ is expanding with $\lambda = s^{-1}$ and any $0 < \delta \leq 1/2$. To see this, let $\omega, \tau \in E_A^\infty$ with $d_s(\omega,\tau) < 1$. This means that $|\omega \wedge \tau| \geq 1$ and, therefore, $|\sigma(\omega) \wedge \sigma(\tau)| = |\omega \wedge \tau| - 1$. So,

$$d_s\big(\sigma(\omega),\sigma(\tau)\big) = s^{|\sigma(\omega)\wedge\sigma(\tau)|} = s^{|\omega\wedge\tau|-1} = s^{-1}s^{|\omega\wedge\tau|} = s^{-1}d_s(\omega,\tau).$$

4.1.1 Expanding repellers

A large class of distance expanding maps are the expanding repellers, which we define and study in this subsection.

Definition 4.1.4. Let U be a nonempty open subset of $\mathbb{R}^d$ and $T : U \to \mathbb{R}^d$ a C^1 map (that is, T is continuously differentiable on U). Let X be a nonempty compact subset of U. The triple (X,U,T) is called an *expanding repeller* provided that the following conditions are satisfied:
(a) $T(X) = X$.
(b) There exists $\lambda > 1$ such that $\|T'(x)v\| \geq \lambda\|v\|$ for all $v \in \mathbb{R}^d$ and all $x \in X$.
(c) $\bigcap_{n=0}^{\infty} T^{-n}(U) = X$.

The set X is sometimes called the *limit set* of the repeller.

Recall that $T'(x) = D_xT : \mathbb{R}^d \to \mathbb{R}^d$ is, by definition, the unique bounded linear operator such that

$$\lim_{y\to x} \frac{\|T(y) - T(x) - D_xT(y-x)\|}{\|y-x\|} = 0,$$

where $\|v\| = \sum_{i=1}^d v_i^2$ is the standard Euclidean norm on $\mathbb{R}^d$. To shorten notation, we write $T'(x)v$ instead of $T'(x)(v)$ or $(T'(x))(v)$. Alternatively, we write $D_xT(v)$. In what follows, $|T'(x)|$ denotes the operator norm of $T'(x)$, that is,

$$\big|T'(x)\big| = \sup\{\|T'(x)v\| : \|v\| \leq 1\},$$

while $\|T'\|$ denotes the supremum of those norms on U, that is,

$$\|T'\| = \sup_{x\in U}|T'(x)|.$$

Condition (c) states that the points whose orbits are confined to U forever are exactly the points of X. Thus each point which is not in X eventually escapes from U under iteration by T. Usually, the closer such a point is to X the longer the escape from U takes. Such points are said to be *repelled* from X. We will now show that every expanding repeller, when restricted to the set X, is a distance expanding map. In fact, we will show that an expanding repeller is distance expanding on an open κ-neighborhood of X denoted by $B(X,\kappa) := \{y \in \mathbb{R}^d : d(y,X) < \kappa\}$. Note that the proof here uses many of the ideas we will present more generally in the next section. In particular, we use the following topological fact, which we state here without proof: For every open cover $\mathcal{U}$ of a compact metric space X, there exists a positive number ϵ, called a *Lebesgue number*, such that every subset of X of diameter less than ϵ is contained entirely in some element of the cover $\mathcal{U}$.

Theorem 4.1.5. *If (X,U,T) is an expanding repeller, then there exists $\kappa > 0$ such that the map $T|_{B(X,\kappa)} : B(X,\kappa) \to \mathbb{R}^d$ is distance expanding. In particular, the system $T|_X : X \to X$ is distance expanding.*

Proof. As this property depends solely on the first iterate of T, the proof relies solely on condition (b) of the definition of a repeller and on the compactness of X.

Since T is differentiable on U, its derivative $T'(x)$ exists at every point $x \in U$. Furthermore, as $T \in C^1(U)$, its derivative T' is continuous on U. Condition (b) guarantees that for each $x \in X$ there exists $r'_x > 0$ such that

$$\|T'(z)v\| \geq \frac{1+\lambda}{2}\|v\|, \quad \forall v \in \mathbb{R}^d,\ \forall z \in B(x,r'_x).$$

In particular, this implies that $T'(z)$ is one-to-one, and is therefore a linear isomorphism of $\mathbb{R}^d$, for every $z \in \bigcup_{x\in X} B(x,r'_x) \supseteq X$. Therefore, the inverse function theorem (Theorem A.2.1) asserts that for every $x \in X$ there exists $r''_x > 0$ such that $T : B(x,r''_x) \to T(B(x,r''_x))$ is a diffeomorphism, and $T : \bigcup_{x\in X} B(x,r''_x) \to \mathbb{R}^d$ is a local diffeomorphism. For every $x \in X$, let $r_x = \min\{r'_x, r''_x\}$. The family of open balls $\{B(x,r_x) : x \in X\}$ forms an open cover of X and hence admits a Lebesgue number $\delta > 0$. That is, for every $x \in X$ there is $\tilde{x} \in X$ such that $B(x,\delta/2) \subseteq B(\tilde{x}, r_{\tilde{x}})$. Setting $r = \delta/2$, it follows that

$$\|T'(z)v\| \geq \frac{1+\lambda}{2}\|v\|, \quad \forall v \in \mathbb{R}^d,\ \forall z \in B(X,r) := \bigcup_{x\in X} B(x,r) \tag{4.1}$$

and $T : B(x,r) \to T(B(x,r))$ is a diffeomorphism for all $x \in X$. Denote the inverse of the diffeomorphism $T : B(x,r) \to T(B(x,r))$ by

$$T_x^{-1} : T(B(x,r)) \to B(x,r). \tag{4.2}$$

As the set $T(B(x,r))$ is open for all $x \in X$, let q_x be the largest radius $Q > 0$ such that $B(T(x),Q) \subseteq T(B(x,r))$. As X is compact, we have $q := \inf_{x\in X} q_x > 0$. (Take the fact that $q > 0$ for granted for the moment; the proof of Lemma 4.2.2 below applies here.) Then

$$B(T(x),q) \subseteq T(B(x,r)), \quad \forall x \in X. \tag{4.3}$$

Furthermore, since it is the image of an open set under a diffeomorphism, the set $T_x^{-1}(B(T(x),q))$ is open for all $x \in X$. Let p_x be the largest $0 < P \le r$ such that $B(x,P) \subseteq T_x^{-1}(B(T(x),q))$. As X is compact, we have $p := \inf_{x\in X} p(x) > 0$. (Take the fact that $p > 0$ for granted for the moment; a variation of the proof of Lemma 4.2.2 below can also be applied here.) Note that $p \le r$ by definition. In addition,

$$B(x,p) \subseteq T_x^{-1}(B(T(x),q)), \quad \forall x \in X. \tag{4.4}$$

Let $y_1, y_2 \in B(X,p/2)$ be such that $\|y_1 - y_2\| < p/2$. Then there exists $x \in X$ such that $y_1, y_2 \in B(x,p)$. Therefore, $T(y_1), T(y_2) \in B(T(x),q)$ according to (4.4). Let $S = [T(y_1), T(y_2)]$ be the line segment joining $T(y_1)$ and $T(y_2)$. Due to the convexity of balls in $\mathbb{R}^d$, we know that $S \subseteq B(T(x),q)$. Moreover, $T_x^{-1}(T(y_1)) = y_1$ and $T_x^{-1}(T(y_2)) = y_2$. Therefore, the curve $T_x^{-1}(S)$ joins the points y_1 to y_2, and thus

$$\|y_1 - y_2\| \le \ell(T_x^{-1}(S)) = \int_S \|(T_x^{-1})'(w)u\|\, dw, \tag{4.5}$$

where $\ell(T_x^{-1}(S))$ stands for the length of the curve $T_x^{-1}(S)$ and u is the unit vector in the direction from $T(y_1)$ to $T(y_2)$. Since $T_x^{-1}(S) \subseteq T_x^{-1}(B(T(x),q)) \subseteq B(x,r)$, for any $w \in S$ inequality (4.1) can be applied with $z = T_x^{-1}(w)$ and $v = (T_x^{-1})'(w)u$ to yield

$$1 = \|u\| = \|T'(T_x^{-1}(w))((T_x^{-1})'(w)u)\| \ge \frac{1+\lambda}{2}\|(T_x^{-1})'(w)u\|.$$

Consequently,

$$\|(T_x^{-1})'(w)u\| \le \frac{2}{1+\lambda}, \quad \forall w \in S,$$

and hence (4.5) gives

$$\|y_1 - y_2\| \le \int_S \frac{2}{1+\lambda} dz = \frac{2}{1+\lambda}\ell(S) = \frac{2}{1+\lambda}\|T(y_1) - T(y_2)\|.$$

In other words,

$$d(T(y_1),T(y_2)) \ge \frac{1+\lambda}{2} d(y_1,y_2), \quad \forall y_1, y_2 \in B(X,p/2) \text{ with } d(y_1,y_2) < p/2.$$

Letting $\kappa = p/2$ completes the proof. □

4.1.2 Hyperbolic Cantor sets

In this section, we introduce and study in detail one special class of expanding repellers (and so of distance expanding maps), namely, hyperbolic Cantor sets. Their construction is a prototype of conformal iterated function systems and conformal graph directed Markov systems, whose systematic account will be given in Chapter 19.

Recall that a *similarity map* $S : \mathbb{R}^d \to \mathbb{R}^d$ is a bijection that multiplies all distances by the same positive real number r called *similarity ratio*, that is,

$$\|S(x) - S(y)\| = r\|x - y\|, \quad \forall x, y \in \mathbb{R}^d.$$

When $r = 1$, a similarity is called an isometry. Two sets are called *similar* if one is the image of the other under a similarity. A similarity $S : \mathbb{R}^d \to \mathbb{R}^d$ with ratio r takes the form

$$S(x) = rA(x) + b,$$

where $A : \mathbb{R}^d \to \mathbb{R}^d$ is an $d \times d$ orthogonal matrix and $b \in \mathbb{R}^d$ is a translation vector. Similarities preserve shapes, including line segments, lines, planes, parallelism, and perpendicularity. Similarities preserve angles but do not necessarily preserve orientation (in fact, S and A preserve orientation if and only if $\det(A) > 0$).

Note also that $S'(x) = rA$ for all $x \in \mathbb{R}^d$. Therefore, $\|S'(x)v\| = r\|v\|$ for all $v \in \mathbb{R}^d$. Consequently, $|S'(x)| = r$ for all $x \in \mathbb{R}^d$, and hence $\|S'\| = r$.

Let E be a finite set such that $\#E \geq 2$. Let $\varphi_e : \mathbb{R}^d \to \mathbb{R}^d$, $e \in E$, be similarities for which there exists a compact set $X_0 \subseteq \mathbb{R}^d$ with the following properties:

(i) $0 < \|\varphi_e'\| < 1$ for all $e \in E$.
(ii) $\varphi_e(X_0) \subseteq X_0$ for all $e \in E$.
(iii) $\varphi_e(X_0) \cap \varphi_f(X_0) = \emptyset$ for all $e, f \in E$ with $e \neq f$.

Construction of the limit set X.

We now define the *limit set X* by constructing a descending sequence of compact sets $(X_n)_{n=1}^\infty$, all of which are subsets of $\bigcup_{e\in E} \varphi_e(X_0)$. The set X_n is called the nth *level set* of the construction. The limit set X will be the intersection of the level sets.

We will use symbolic dynamics notation. For every $\omega \in E^*$, define

$$\varphi_\omega := \varphi_{\omega_1} \circ \varphi_{\omega_2} \circ \cdots \circ \varphi_{\omega_{|\omega|}}.$$

The maps φ_e, $e \in E$, are said to be the *generators* of the construction, and so we say that the map φ_ω is generated by the word ω. The nth level set X_n is the disjoint union of the images of X_0 under all maps generated by words of length n, namely

$$X_n := \bigcup_{\omega \in E^n} \varphi_\omega(X_0). \tag{4.6}$$

As a finite union of compact sets, each level set X_n is compact. Moreover, by condition (ii),

$$X_{n+1} = \bigcup_{\omega \in E^{n+1}} \varphi_{\omega|_n}(\varphi_{\omega_{n+1}}(X_0)) \subseteq \bigcup_{\tau \in E^n} \varphi_\tau(X_0) = X_n \tag{4.7}$$

for all $n \in \mathbb{N}$. As the intersection of a descending sequence of nonempty compact sets, the limit set

$$X := \bigcap_{n=1}^{\infty} X_n \tag{4.8}$$

is a nonempty compact set.

The limit set X is a topological Cantor set (the proof of this fact is postponed) with the property that

$$\bigcup_{e \in E} \varphi_e(X) = X.$$

Indeed, we first observe that the generators map the nth level set down to the $(n+1)$th level set since

$$\begin{aligned}
\bigcup_{e \in E} \varphi_e(X_n) &= \bigcup_{e \in E} \bigcup_{\omega \in E^n} \varphi_e(\varphi_\omega(X_0)) \\
&= \bigcup_{e \in E} \bigcup_{\omega \in E^n} \varphi_{e\omega}(X_0) \\
&= \bigcup_{\tau \in E^{n+1}} \varphi_\tau(X_0) \\
&= X_{n+1}
\end{aligned} \tag{4.9}$$

for all $n \in \mathbb{N}$.

Moreover, $x \in \bigcap_{n=1}^{\infty}[\bigcup_{e \in E} \varphi_e(X_n)]$ if and only if for every $n \in \mathbb{N}$ there exists $e_n \in E$ such that $x \in \varphi_{e_n}(X_n)$. Since $\varphi_e(X_n) \subseteq \varphi_e(X_0)$ for all $e \in E$, the e_n's are unique according to condition (iii). By this very same condition, since $x \in \varphi_{e_n}(X_n) \cap \varphi_{e_{n+1}}(X_{n+1}) \subseteq \varphi_{e_n}(X_0) \cap \varphi_{e_{n+1}}(X_0)$, it turns out that $e_n = e_{n+1}$ for all $n \in \mathbb{N}$. In summary, $x \in \bigcap_{n=1}^{\infty}[\bigcup_{e \in E} \varphi_e(X_n)]$ if and only if there is a unique $e \in E$ such that $x \in \varphi_e(X_n)$ for all $n \in \mathbb{N}$. In other words,

$$\bigcup_{e \in E}\left[\bigcap_{n=1}^{\infty} \varphi_e(X_n)\right] = \bigcap_{n=1}^{\infty}\left[\bigcup_{e \in E} \varphi_e(X_n)\right]. \tag{4.10}$$

It follows from (4.8), (4.9), (4.10), and the injectivity of the generators that

$$\begin{aligned}
\bigcup_{e \in E} \varphi_e(X) = \bigcup_{e \in E} \varphi_e\left(\bigcap_{n=1}^{\infty} X_n\right) &= \bigcup_{e \in E}\left[\bigcap_{n=1}^{\infty} \varphi_e(X_n)\right] \\
&= \bigcap_{n=1}^{\infty}\left[\bigcup_{e \in E} \varphi_e(X_n)\right] = \bigcap_{n=1}^{\infty} X_{n+1} = X.
\end{aligned}$$

By induction, we have that

$$\bigcup_{\omega\in E^n} \varphi_\omega(X) = X, \quad \forall n \in \mathbb{N}. \tag{4.11}$$

Construction of a neighborhood U of X.

Since $\varphi_e(X_0) \cap \varphi_f(X_0) = \emptyset$ for all $e \neq f$ and since there are finitely many compact sets $\varphi_e(X_0)$, $e \in E$, the continuity of the generators ensures the existence of an open ε-neighborhood $B(X_0, \varepsilon) := \{x \in \mathbb{R}^d : d(x, X_0) < \varepsilon\}$ of X_0 such that

$$\varphi_e\big(B(X_0,\varepsilon)\big) \cap \varphi_f\big(B(X_0,\varepsilon)\big) = \emptyset, \quad \forall e \neq f.$$

Let

$$U = \bigcup_{e\in E} \varphi_e\big(B(X_0,\varepsilon)\big). \tag{4.12}$$

Construction of a map $T : U \to \mathbb{R}^d$.

Finally, we define a map $T : U \to \mathbb{R}^d$ by

$$T|_{\varphi_e(B(X_0,\varepsilon))} = \varphi_e^{-1}. \tag{4.13}$$

This piecewise-similar map is well-defined since the sets $\varphi_e(B(X_0,\varepsilon))$, $e \in E$, are mutually disjoint.

Proof that the triple (X, U, T) is an expanding repeller.

Condition (a) for a repeller is rather easy to check. First, note that T maps the nth level set up to the $(n-1)$th level set, that is,

$$T(X_n) = X_{n-1}, \quad \forall n \in \mathbb{N}. \tag{4.14}$$

Indeed, let $n \in \mathbb{N}$. Since $\varphi_e(X_{n-1}) \subseteq \varphi_e(B(X_0,\varepsilon))$ and since $T|_{\varphi_e(B(X_0,\varepsilon))} = \varphi_e^{-1}$ for every $e \in E$, we have

$$\begin{aligned}
T(X_n) &= \bigcup_{\omega\in E^n} T\big(\varphi_\omega(X_0)\big)\\
&= \bigcup_{\tau\in E^{n-1}} \bigcup_{e\in E} T\big(\varphi_{e\tau}(X_0)\big)\\
&= \bigcup_{\tau\in E^{n-1}} \bigcup_{e\in E} T\circ\varphi_e\big(\varphi_\tau(X_0)\big)\\
&= \bigcup_{\tau\in E^{n-1}} \varphi_\tau(X_0)\\
&= X_{n-1}.
\end{aligned}$$

Then

$$T(X) = T\left(\bigcap_{n=1}^{\infty} X_n\right) \subseteq \bigcap_{n=1}^{\infty} T(X_n) = \bigcap_{n=1}^{\infty} X_{n-1} = X.$$

This establishes that $T(X) \subseteq X$. To prove the reverse inclusion, pick $x \in X$. Then $x \in X_n$ for every $n \in \mathbb{N}$. Fix an arbitrary $e \in E$. Then $\varphi_e(x) \in X_{n+1}$ for every $n \in \mathbb{N}$. Thus $\varphi_e(x) \in X$. It follows that $x = \varphi_e^{-1}(\varphi_e(x)) = T(\varphi_e(x)) \in T(X)$. Hence $X \subseteq T(X)$. Since both inclusions hold, we conclude that

$$T(X) = X. \tag{4.15}$$

Condition (b) for a repeller is also straightforward to verify. Indeed, let $x \in U$. There exists a unique $e_x \in E$ such that $x \in \varphi_{e_x}(B(X_0, \varepsilon))$. Then, for all $v \in \mathbb{R}^d$,

$$\begin{aligned} \|T'(x)v\| &= \|(T|_{\varphi_{e_x}(B(X_0,\varepsilon))})'(x)v\| = \|(\varphi_{e_x}^{-1})'(x)v\| = \|\varphi'_{e_x}\|^{-1}\|v\| \\ &\geq \min_{e\in E} \frac{1}{\|\varphi'_e\|}\|v\| = \frac{1}{M}\|v\|, \end{aligned} \tag{4.16}$$

where $M := \max_{e\in E} \|\varphi'_e\| < 1$ by condition (i).

Finally, we show that condition (c) for a repeller is fulfilled. First, we observe that

$$T^{-n}(U) = \bigcup_{\omega\in E^{n+1}} \varphi_\omega(B(X_0, \varepsilon)), \quad \forall n \in \mathbb{Z}_+. \tag{4.17}$$

Indeed, by definition of U, the relationship holds when $n = 0$. For the inductive step, let $x \in U$ and $n \in \mathbb{N}$. Then

$$\begin{aligned} x \in T^{-n}(U) &\iff T(x) \in T^{-(n-1)}(U) \\ &\iff T(x) \in \bigcup_{\omega\in E^n} \varphi_\omega(B(X_0, \varepsilon)) \\ &\iff \varphi_f^{-1}(x) \in \bigcup_{\omega\in E^n} \varphi_\omega(B(X_0, \varepsilon)), \text{ if } x \in \varphi_f(B(X_0, \varepsilon)) \\ &\iff x \in \bigcup_{\tau\in E^{n+1}} \varphi_\tau(B(X_0, \varepsilon)). \end{aligned}$$

By induction, (4.17) holds.

Now, observe that the similarity map φ_e enjoys the property that

$$\varphi_e(B(X_0, \varepsilon)) = B(\varphi_e(X_0), \|\varphi'_e\|\varepsilon) \subseteq B(\varphi_e(X_0), M\varepsilon).$$

By an induction argument, we deduce that

$$\varphi_\omega(B(X_0, \varepsilon)) = B(\varphi_\omega(X_0), \|\varphi'_\omega\|\varepsilon) \subseteq B(\varphi_\omega(X_0), M^{|\omega|}\varepsilon), \quad \forall \omega \in E^*. \tag{4.18}$$

Therefore,

$$\begin{aligned} \bigcup_{\omega\in E^n} \varphi_\omega(B(X_0, \varepsilon)) &\subseteq \bigcup_{\omega\in E^n} B(\varphi_\omega(X_0), M^n\varepsilon) \\ &= B\Big(\bigcup_{\omega\in E^n} \varphi_\omega(X_0), M^n\varepsilon\Big) \\ &= B(X_n, M^n\varepsilon), \quad \forall n \in \mathbb{N}. \end{aligned} \tag{4.19}$$

It follows that (cf. Exercise 4.6.4)

$$X = \bigcap_{n=1}^{\infty} X_n = \bigcap_{n=1}^{\infty} \bigcup_{\omega\in E^n} \varphi_\omega(X_0) \subseteq \bigcap_{n=1}^{\infty} \bigcup_{\omega\in E^n} \varphi_\omega(B(X_0,\varepsilon)) \subseteq \bigcap_{n=1}^{\infty} B(X_n, M^n\varepsilon) = X.$$

From this and (4.17), we conclude that

$$X = \bigcap_{n=1}^{\infty} \bigcup_{\omega\in E^n} \varphi_\omega(B(X_0,\varepsilon)) = \bigcap_{n=0}^{\infty} T^{-n}(U). \tag{4.20}$$

This establishes condition (c), and completes the proof that the triple (X,U,T) is an expanding repeller.

Alternative construction of the limit set X and proof that X is a topological Cantor set.

The limit set X can also be constructed in a slightly different way. Let $\omega \in E^\infty$. Since (by condition (ii))

$$\varphi_{\omega|_{n+1}}(X_0) = \varphi_{\omega|_n}(\varphi_{\omega_{n+1}}(X_0)) \subseteq \varphi_{\omega|_n}(X_0) \tag{4.21}$$

and since

$$\operatorname{diam}(\varphi_{\omega|_n}(X_0)) = \|\varphi'_{\omega|_n}\| \operatorname{diam}(X_0) \le M^n \operatorname{diam}(X_0) \tag{4.22}$$

for all $n \in \mathbb{N}$, the sets $(\varphi_{\omega|_n}(X_0))_{n=1}^{\infty}$ form a descending sequence of nonempty compact sets whose diameters tend to 0 (by condition (i)). Therefore, $\bigcap_{n=1}^{\infty} \varphi_{\omega|_n}(X_0)$ is a singleton. Define the *coding map* $\pi : E^\infty \to \mathbb{R}^d$ by

$$\{\pi(\omega)\} := \bigcap_{n=1}^{\infty} \varphi_{\omega|_n}(X_0).$$

This map is injective. Indeed, if $\omega \ne \tau \in E^\infty$, then there is a smallest $n \in \mathbb{N}$ such that $\omega_n \ne \tau_n$. It follows from the injectivity of the generators and condition (iii) that

$$\begin{aligned}
\{\pi(\omega)\} \cap \{\pi(\tau)\} &\subseteq \varphi_{\omega|_n}(X_0) \cap \varphi_{\tau|_n}(X_0) \\
&= [\varphi_{\omega|_{n-1}}(\varphi_{\omega_n}(X_0))] \cap [\varphi_{\omega|_{n-1}}(\varphi_{\tau_n}(X_0))] \\
&= \varphi_{\omega|_{n-1}}(\varphi_{\omega_n}(X_0) \cap \varphi_{\tau_n}(X_0)) \\
&= \emptyset.
\end{aligned}$$

We claim that

$$X = \pi(E^\infty). \tag{4.23}$$

Indeed, let $\omega \in E^\infty$. Since $\varphi_{\omega|_n}(X_0) \subseteq X_n$ for all $n \in \mathbb{N}$, we have $\{\pi(\omega)\} = \bigcap_{n=1}^\infty \varphi_{\omega|_n}(X_0) \subseteq \bigcap_{n=1}^\infty X_n = X$. Therefore, $\pi(E^\infty) \subseteq X$. Now, let $x \in X$. Then $x \in X_n$ for every $n \in \mathbb{N}$. This means that there exists a unique $\omega^{(n)} \in E^n$ such that $x \in \varphi_{\omega^{(n)}}(X_0)$. The uniqueness of the $\omega^{(n)}$'s implies that $\omega^{(n+1)}|_n = \omega^{(n)}$. Define $\omega \in E^\infty$ to be such that $\omega|_n = \omega^{(n)}$ for all $n \in \mathbb{N}$. Then $x \in \varphi_{\omega^{(n)}}(X_0) = \varphi_{\omega|_n}(X_0)$ for all $n \in \mathbb{N}$. Thus $x = \pi(\omega)$. Hence $X \subseteq \pi(E^\infty)$. Since both inclusions hold, the claim has been shown.

Furthermore, the map π is continuous. This ensues from the fact that, for every $\rho \in E^*$,

$$\operatorname{diam}(\pi([\rho])) \leq \operatorname{diam}(\varphi_\rho(X_0)) \leq M^{|\rho|} \operatorname{diam}(X_0). \tag{4.24}$$

In summary, the map $\pi : E^\infty \to X$ is a continuous bijection between two compact metrizable spaces. Thus $\pi : E^\infty \to X$ is a homeomorphism and X is a homeomorphic image of E^∞, that is, X is a topological Cantor set.

For later purposes, we further note that $\varphi_{\omega_1} \circ \pi \circ \sigma(\omega) = \pi(\omega)$ for all $\omega \in E^\infty$. In light of (4.13), this means that

$$\pi \circ \sigma = T \circ \pi, \tag{4.25}$$

that is, the symbolic system (E^∞, σ) is topologically conjugate to the dynamical system (X, T) via the coding map π.

One final word: Condition (i) is the reason for calling X an hyperbolic Cantor set. In differentiable dynamical systems, hyperbolicity takes the shape of a derivative that stays away from 1.

Example 4.1.6. Let $\varphi_e : \mathbb{R} \to \mathbb{R}$, $e \in E := \{0, 2\}$, be the two contracting similarities defined by

$$\varphi_e(x) = \frac{x+e}{3}.$$

Let $X_0 = I := [0,1]$ be the unit interval. The limit set X defined in Subsection 4.1.2 is called the *middle-third Cantor set* and is usually denoted by C. Let $\varepsilon = 1/3$. Then $B(X_0, \varepsilon) = (-1/3, 4/3)$ and

$$U = \varphi_0((-1/3, 4/3)) \cup \varphi_2((-1/3, 4/3)) = (-1/9, 4/9) \cup (5/9, 10/9).$$

Moreover, $T : U \to \mathbb{R}$ is defined by $T(x) = 3x - e$ if $x \in \varphi_e(B(X_0, \varepsilon))$, that is,

$$T(x) = \begin{cases} 3x & \text{if } x \in (-1/9, 4/9) \\ 3x - 2 & \text{if } x \in (5/9, 10/9). \end{cases}$$

In two dimensions, the most classic examples of hyperbolic Cantor sets are Sierpiński triangles (also called Sierpiński gaskets) and Sierpiński carpets. The following example is a natural generalization of the middle-third Cantor set.

Example 4.1.7. Let $\varphi_e : \mathbb{R}^2 \to \mathbb{R}^2$, $e \in E := \{0,1,2\}$, be the three contracting similarities defined by

$$\varphi_0(x,y) = \left(\frac{x}{3}, \frac{y}{3}\right)$$
$$\varphi_1(x,y) = \left(\frac{x+1}{3}, \frac{y+\sqrt{3}}{3}\right)$$
$$\varphi_2(x,y) = \left(\frac{x+2}{3}, \frac{y}{3}\right).$$

Let X_0 be the filled-in equilateral triangle with vertices $(0,0)$, $(1/2, \sqrt{3}/2)$ and $(1,0)$. The limit set X defined in Subsection 4.1.2 is a totally disconnected *Sierpinski triangle*. Let $\varepsilon = 1/3$. Then $U = \bigcup_{e\in E} \varphi_e(B(X_0,\varepsilon))$ and the map $T : U \to \mathbb{R}^2$ is defined by

$$T(x,y) := \varphi_e^{-1}(x,y), \quad \forall (x,y) \in \varphi_e\big(B(X_0,\varepsilon)\big), \quad \forall e \in E.$$

That is,

$$T(x,y) = \begin{cases} (3x, 3y) & \text{if } (x,y) \in \varphi_0(B(X_0,\varepsilon)) \\ (3x-1, 3y-\sqrt{3}) & \text{if } (x,y) \in \varphi_1(B(X_0,\varepsilon)) \\ (3x-2, 3y) & \text{if } (x,y) \in \varphi_2(B(X_0,\varepsilon)). \end{cases}$$

Example 4.1.8. Let $\varphi_e : \mathbb{R}^2 \to \mathbb{R}^2$, $e \in E := \{0,1,2,3\}$, be the following four contracting similarities:

$$\varphi_0(x,y) = \left(\frac{x}{4}, \frac{y}{4}\right), \quad \varphi_1(x,y) = \left(\frac{x+3}{4}, \frac{y}{4}\right),$$
$$\varphi_2(x,y) = \left(\frac{x+3}{4}, \frac{y+3}{4}\right), \quad \varphi_3(x,y) = \left(\frac{x}{4}, \frac{y+3}{4}\right).$$

Let $X_0 = I^2$ be the unit square. The limit set X defined in Subsection 4.1.2 is a totally disconnected *Sierpinski carpet* (see Figure 4.1). Let $\varepsilon = 1/2$. Then $U = \bigcup_{e\in E} \varphi_e(B(X_0,\varepsilon))$ and the map $T : U \to \mathbb{R}^2$ is defined by

$$T(x,y) := \varphi_e^{-1}(x,y), \quad \forall (x,y) \in \varphi_e\big(B(X_0,\varepsilon)\big), \quad \forall e \in E.$$

That is,

$$T(x,y) = \begin{cases} (4x, 4y) & \text{if } (x,y) \in \varphi_0(B(X_0,\varepsilon)) \\ (4x-3, 4y) & \text{if } (x,y) \in \varphi_1(B(X_0,\varepsilon)) \\ (4x-3, 4y-3) & \text{if } (x,y) \in \varphi_2(B(X_0,\varepsilon)) \\ (4x, 4y-3) & \text{if } (x,y) \in \varphi_3(B(X_0,\varepsilon)). \end{cases}$$

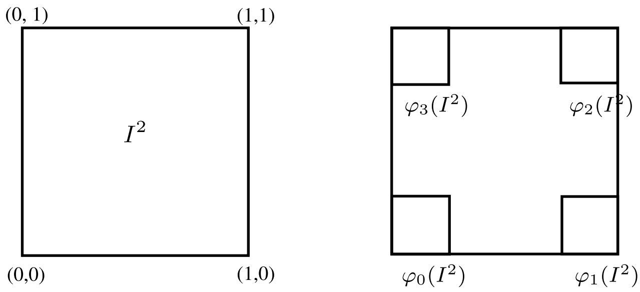

Figure 4.1: The action of the four contracting similarities $\varphi_0, \dots, \varphi_3$ on the closed unit square I^2.

4.2 Inverse branches

In order for a map to have a properly defined inverse, it is necessary that the map be injective. Nonetheless, we can get around the noninjectivity of a map by defining its *inverse branches* as long as the map is locally injective. The following proposition is the first step in the construction of the inverse branches of a distance expanding map.

Proposition 4.2.1. *Let $T : X \to X$ be a distance expanding map. For all $x \in X$, the restriction $T|_{B(x,\delta)}$ is injective.*

Proof. For each $x \in X$, apply Remark 4.1.2(b) to $B(x, \delta)$. □

We shall assume from this point on that $T : X \to X$ is an **open**, distance expanding map of a compact metric space X. Note that the restriction $T|_F$ of T to a closed forward T-invariant subset F of X need not be open (see Exercise 4.6.5). However, since T is open, for every $x \in X$ and $r > 0$ the set $T(B(x,r))$ is open and, therefore, contains a nonempty open ball centered at $T(x)$, say $B(T(x), s(r))$. Accordingly, we define

$$R(x,r) := \sup\{s > 0 : B(T(x), s) \subseteq T(B(x,r))\} > 0.$$

In fact, $R(x,r)$ is the radius of the largest ball centered at $T(x)$ which is contained in $T(B(x,r))$. In the following lemma, we investigate the greatest lower bound of the radii $R(x,r)$ for a fixed $r > 0$.

Lemma 4.2.2. *For every $r > 0$, we have $R(r) := \inf\{R(x,r) : x \in X\} > 0$.*

Proof. We shall prove this lemma by contradiction. Suppose that there exists some $r > 0$ for which $R(r) = 0$. This means that there exists a sequence $(x_n)_{n=1}^{\infty}$ in X such that $\lim_{n\to\infty} R(x_n, r) = 0$. Since X is compact, the sequence $(x_n)_{n=1}^{\infty}$ has a convergent subsequence, say $(x_{n_k})_{k=1}^{\infty}$. Define $x := \lim_{k\to\infty} x_{n_k}$. Then there exists $K \in \mathbb{N}$ such that $d(x_{n_k}, x) < r/2$ for all $k \geq K$. In particular, this implies that

$$B(x, r/2) \subseteq B(x_{n_k}, r), \quad \forall k \geq K. \tag{4.26}$$

Now, since T is an open map, there exists $\eta > 0$ such that

$$B\big(T(x),\eta\big) \subseteq T\big(B(x,r/2)\big). \tag{4.27}$$

Moreover, $T(x) = \lim_{k\to\infty} T(x_{n_k})$ since T is continuous. Thus there exists $K' \in \mathbb{N}$ such that $d(T(x_{n_k}),T(x)) < \eta/2$ for all $k \geq K'$. In particular, this implies that

$$B\big(T(x_{n_k}),\eta/2\big) \subseteq B\big(T(x),\eta\big), \quad \forall k \geq K'. \tag{4.28}$$

From (4.26), (4.27), and (4.28), it follows that for all $k \geq \max\{K,K'\}$,

$$B\big(T(x_{n_k}),\eta/2\big) \subseteq B\big(T(x),\eta\big) \subseteq T\big(B(x,r/2)\big) \subseteq T\big(B(x_{n_k},r)\big).$$

Therefore, $R(x_{n_k},r) \geq \eta/2$ for all $k \geq \max\{K,K'\}$, which contradicts our assumption that $\lim_{n\to\infty} R(x_n,r) = 0$. Thus $\inf\{R(x,r) : x \in X\} > 0$. □

Remark 4.2.3. Notice that we did not use the distance expanding property of T in the proof, so this result holds for all open, continuous maps of a compact metric space. The compactness of the space X ensures that for any fixed r, the radii $R(x,r)$ have a positive greatest lower bound $R(r)$. Thus the image of any ball of radius r contains a ball of radius $R(r)$.

Lemma 4.2.2 shows that for each $r > 0$ the image of the ball of radius r centered at the point x contains the ball of radius $R(r)$ centered at $T(x)$. For an open, distance expanding map T, the quantity

$$\xi := \min\{\delta, R(\delta)\} > 0, \tag{4.29}$$

where δ is a constant delimiting the neighborhoods of expansion of the map, is of particular interest. Indeed, given that, according to Proposition 4.2.1, the restricted map $T|_{B(x,\delta)}$ is injective for every $x \in X$, we can define its inverse

$$T|_{B(x,\delta)}^{-1} : T\big(B(x,\delta)\big) \to B(x,\delta).$$

By the definition of ξ given above, we have that

$$B\big(T(x),\xi\big) \subseteq T\big(B(x,\delta)\big)$$

for every $x \in X$. This inclusion is illustrated in Figure 4.2. We denote the restriction of the inverse of the map $T|_{B(x,\delta)}$ to $B(T(x),\xi)$ by

$$T_x^{-1} := (T|_{B(x,\delta)}^{-1})|_{B(T(x),\xi)} : B\big(T(x),\xi\big) \to B(x,\delta). \tag{4.30}$$

Note that T_x^{-1} is injective but not necessarily surjective. The map T_x^{-1} is the local inverse branch of T that maps $T(x)$ to x. As T expands distances by a factor $\lambda > 1$, one naturally expects the local inverse branches T_x^{-1}, $x \in X$, to contract distances by a factor λ^{-1}. This is indeed the case.

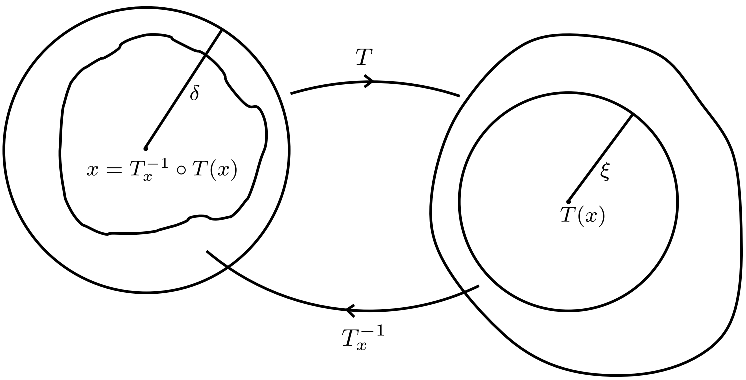

Figure 4.2: Illustration of the ball $B(T(x), \xi)$ mapped under the inverse branch T_x^{-1} inside the ball $B(x, \delta)$.

Proposition 4.2.4. *The local inverse branches T_x^{-1}, for each $x \in X$, are contractions with (contraction) ratio λ^{-1}.*

Proof. Fix $x \in X$. We aim to prove that if $y, z \in B(T(x), \xi)$, then

$$d(T_x^{-1}(y), T_x^{-1}(z)) \leq \lambda^{-1} d(y, z),$$

where λ is a constant of expansion for T. Since $T_x^{-1}(B(T(x), \xi)) \subseteq B(x, \delta)$, both $T_x^{-1}(y)$ and $T_x^{-1}(z)$ lie in $B(x, \delta)$, and hence

$$d(T_x^{-1}(y), T_x^{-1}(z)) < 2\delta.$$

Therefore, the expanding property of T guarantees that

$$d(y, z) = d(T \circ T_x^{-1}(y), T \circ T_x^{-1}(z)) \geq \lambda d(T_x^{-1}(y), T_x^{-1}(z)).$$

Consequently,

$$d(T_x^{-1}(y), T_x^{-1}(z)) \leq \lambda^{-1} d(y, z). \qquad \square$$

Now, let $w \in B(T(x), \xi)$. Proposition 4.2.4 implies that

$$d(T_x^{-1}(w), x) = d(T_x^{-1}(w), T_x^{-1}(T(x))) \leq \lambda^{-1} d(w, T(x)) < \lambda^{-1}\xi.$$

Thus

$$T_x^{-1}(B(T(x), \xi)) \subseteq B(x, \lambda^{-1}\xi) \subseteq B(x, \xi).$$

Thanks to this property, we can define the local inverse branches of the iterates of T. Let $x \in X$ and $n \in \mathbb{N}$. The local inverse branch of T^n that maps $T^n(x)$ to x is defined to be

$$T_x^{-n} := T_x^{-1} \circ T_{T(x)}^{-1} \circ \cdots \circ T_{T^{n-1}(x)}^{-1} : B(T^n(x), \xi) \to B(x, \xi). \tag{4.31}$$

This composition will henceforth be called the *inverse branch of T^n determined by the point x*. See Figure 4.3.

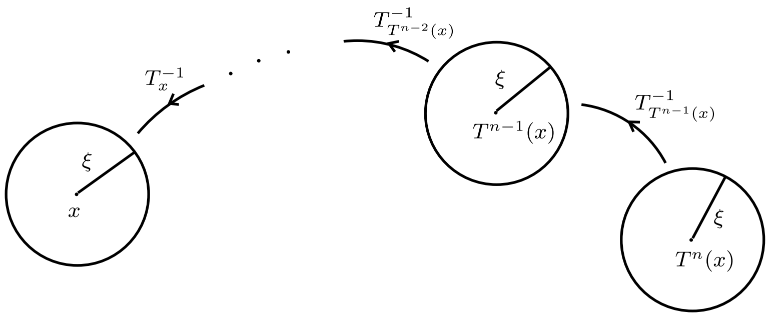

Figure 4.3: The map $T^{-1}_{T^{n-1}(x)}$ sends the ball $B(T^n(x),\xi)$ into the ball $B(T^{n-1}(x),\xi)$, which is in turn mapped into the ball $B(T^{n-2}(x),\xi)$ by $T^{-1}_{T^{n-2}(x)}$ and so on, until finally T_x^{-1} sends us back inside $B(x,\xi)$.

Remark 4.2.5. Let $x \in X$ and $y,z \in B(T^n(x),\xi)$. A successive application of Proposition 4.2.4 at the points $T^{n-1}(x)$, $T^{n-2}(x),\dots,T(x)$ and x establishes that

$$d\big(T_x^{-n}(y),T_x^{-n}(z)\big) \le \lambda^{-n}\, d(y,z).$$

In particular,

$$T_x^{-n}\big(B(T^n(x),\xi')\big) \subseteq B(x,\lambda^{-n}\xi') \subseteq B(x,\xi'), \quad \forall 0<\xi' \le \xi,\ \forall n \in \mathbb{N}, \tag{4.32}$$

and hence

$$B(T^n(x),\xi') \subseteq T^n\big(B(x,\lambda^{-n}\xi')\big) \subseteq T^n\big(B(x,\xi')\big), \quad \forall 0<\xi' \le \xi,\ \forall n \in \mathbb{N}. \tag{4.33}$$

The inverse branches of an open, distance expanding map are easiest to grasp with the aid of an example. Below, we first calculate the inverse branches for the map $T(x) := 2x \pmod 1$, and secondly give the example of a subshift of finite type.

Example 4.2.6. Consider the map $T : \mathbb{S}^1 \to \mathbb{S}^1$ defined by

$$T(x) := \begin{cases} 2x & \text{if } x \in [0,1/2] \\ 2x-1 & \text{if } x \in [1/2,1]. \end{cases}$$

Note that T is the map T_m we have seen in Example 1.1.3(b), with $m = 2$. The subscript 2 has been dropped to simplify notation in what follows. We have that T is distance expanding on neighborhoods of size $\delta = 1/4$ with expanding constant $\lambda = 2$. More precisely, if $x,y \in \mathbb{S}^1$ with $d(x,y) < 1/2$, then $d(T(x),T(y)) = 2d(x,y)$. Here, the metric d is the usual Euclidean metric on the circle. So, for instance, $d(1/8,7/8) = 1/4$, whereas

$d(T(1/8), T(7/8)) = d(1/4, 3/4) = 1/2$. Also, T is easily seen to be an open map. We can now consider the inverse branches of T. Let us start by investigating the inverse branches determined by the points 1/4 and 3/4. We know, by Proposition 4.2.1 (and, in this case, by inspection), that T is injective on any open subinterval of $\mathbb{S}^1$ with radius at most 1/4. In particular, T is injective on the intervals $(0, 1/2)$ and $(1/2, 1)$. Notice that $T(1/4) = 1/2 = T(3/4)$. Moreover,

$$T\Big(B\Big(\tfrac{1}{4}, \tfrac{1}{4}\Big)\Big) = T\Big(\Big(0, \tfrac{1}{2}\Big)\Big) = (0, 1) = B\Big(\tfrac{1}{2}, \tfrac{1}{2}\Big) = B\Big(T\Big(\tfrac{1}{4}\Big), \tfrac{1}{2}\Big)$$

and

$$T\Big(B\Big(\tfrac{3}{4}, \tfrac{1}{4}\Big)\Big) = T\Big(\Big(\tfrac{1}{2}, 1\Big)\Big) = (0, 1) = B\Big(\tfrac{1}{2}, \tfrac{1}{2}\Big) = B\Big(T\Big(\tfrac{3}{4}\Big), \tfrac{1}{2}\Big).$$

Therefore, $R(1/4, 1/4) = R(3/4, 1/4) = 1/2$. In fact, the image under T of any ball $B(x, 1/4)$ contains a ball of radius 1/2 about the point $T(x)$. Thus, in this case, $R(\delta) = R(1/4) = 1/2$ and so $\xi := \min\{\delta, R(\delta)\} = 1/4$. Hence, we obtain inverse branches $T_x^{-1} : B(T(x), 1/4) \to B(x, 1/4)$. Note that every interval $B(T(x), 1/4)$ has two inverse branches defined upon it, one taking points back to an interval around the preimage of $T(x)$ lying in $(0, 1/2)$ and the other sending points to an interval around the preimage of $T(x)$ lying in $(1/2, 1)$. For example, the two inverse branches defined on the interval $B(1/2, 1/4) = (1/4, 3/4)$ are

$$T_{\frac{1}{4}}^{-1} : \Big(\tfrac{1}{4}, \tfrac{3}{4}\Big) \to \Big(0, \tfrac{1}{2}\Big), \quad \text{defined by } T_{\frac{1}{4}}^{-1}(y) := \tfrac{y}{2}$$

and

$$T_{\frac{3}{4}}^{-1} : \Big(\tfrac{1}{4}, \tfrac{3}{4}\Big) \to \Big(\tfrac{1}{2}, 1\Big), \quad \text{defined by } T_{\frac{3}{4}}^{-1}(y) := \tfrac{y+1}{2}.$$

Let us now consider the inverse branches of the iterates of T. For each point $x \in \mathbb{S}^1$, we have the inverse branch of T^n determined by x:

$$T_x^{-n} : B\Big(T^n(x), \tfrac{1}{4}\Big) \to B\Big(x, \tfrac{1}{4}\Big).$$

Recall that this map is injective but not surjective. In particular, if $x = 2^{-(n+1)}$, the map $T_{2^{-(n+1)}}^{-n} : B(1/2, 1/4) \to B(2^{-(n+1)}, 1/4)$ turns out to be $T_{2^{-(n+1)}}^{-n}(y) = y/2^n$. For the map T, every interval $B(T^n(x), 1/4)$ has 2^n inverse branches defined upon it.

Example 4.2.7 (Subshifts of finite type). It was shown in Example 4.1.3 that the shift map $\sigma : E_A^\infty \to E_A^\infty$ is distance expanding, with $\lambda = s^{-1}$ and $\delta = 1/2$ when E_A^∞ is endowed with the metric d_s. Also, by Theorem 3.2.12, we know that the shift map is an open map. Now, let $\omega \in E_A^\infty$ and observe that $\sigma^{-1}(\omega) = \{i\omega : A_{i\omega_1} = 1\}$. We want to find the inverse branches of σ. Since $2\delta = 1$, we first describe the open ball $B(\omega, 1)$ upon which we know the shift map to be injective, due to Proposition 4.2.1:

$$B(\omega, 1) = \{\tau \in E_A^\infty : s^{|\omega\wedge\tau|} < 1\} = \{\tau \in E_A^\infty : |\omega \wedge \tau| \geq 1\} = [\omega_1].$$

In other words, $B(\omega, 1)$ is the initial cylinder set determined by the first letter of ω. Then the inverse branch of σ determined by the word ω is the map

$$\sigma_\omega^{-1} : B(\sigma(\omega), 1) \to B(\omega, 1)$$

with the property that $\sigma_\omega^{-1}(\sigma(\omega)) = \omega$, that is,

$$\begin{array}{cccc} \sigma_\omega^{-1} : & [\omega_2] & \longrightarrow & [\omega_1] \\ & \tau & \longmapsto & \omega_1\tau. \end{array}$$

In fact, note that $\sigma_\omega^{-1} : [\omega_2] \to [\omega_1\omega_2]$. Similarly, the inverse branch of the nth iterate of σ determined by the word ω is the map which adds to each word $\tau \in [\omega_{n+1}]$ a prefix (or initial block) consisting of the first n letters of ω, that is,

$$\begin{array}{cccc} \sigma_\omega^{-n} : & [\omega_{n+1}] & \longrightarrow & [\omega_1\omega_2\ldots\omega_n\omega_{n+1}] \\ & \tau & \longmapsto & \omega_1\omega_2\ldots\omega_n\tau. \end{array}$$

The next lemma states that inverse branches determined by distinct points have disjoint images.

Lemma 4.2.8. *Let $z \in X$. Let $n \in \mathbb{N}$ and $x, y \in T^{-n}(z)$ with $x \neq y$. Then*

$$T_x^{-n}(B(z, \xi)) \cap T_y^{-n}(B(z, \xi)) = \emptyset.$$

Proof. Let $1 \leq k \leq n$ be the smallest integer such that $T^k(x) = T^k(y)$ and further let $w := T^k(x) = T^k(y)$. By Remark 4.2.5, we deduce that

$$\begin{aligned} T_x^{-n}(B(z, \xi)) &= T_x^{-k} \circ T_{T^k(x)}^{-(n-k)}(B(z, \xi)) \\ &\subseteq T_x^{-k}(B(T^k(x), \xi)) = T_x^{-k}(B(w, \xi)). \end{aligned}$$

Likewise,

$$T_y^{-n}(B(z, \xi)) \subseteq T_y^{-k}(B(w, \xi)).$$

It is thus sufficient to show that the sets $T_x^{-k}(B(w, \xi))$ and $T_y^{-k}(B(w, \xi))$ are disjoint. To shorten notation in what follows, set

$$w_x := T^{k-1}(x) \text{ and } w_y := T^{k-1}(y).$$

Then

$$T_x^{-k} = T_x^{-(k-1)} \circ T_{w_x}^{-1} \text{ and } T_y^{-k} = T_y^{-(k-1)} \circ T_{w_y}^{-1}. \tag{4.34}$$

Moreover, by the definition of k and the hypothesis that $x \neq y$, we have that $T(w_x) = T(w_y) = w$ and also that $w_x \neq w_y$. Consequently, in light of Remark 4.1.2(b), it follows that $d(w_x, w_y) \geq 2\delta$. Recalling that $\xi \leq \delta$, we deduce that

$$B(w_x, \xi) \cap B(w_y, \xi) = \emptyset.$$

According to Remark 4.2.5, we know that

$$T_{w_x}^{-1}(B(w, \xi)) \subseteq B(w_x, \xi) \quad \text{and} \quad T_{w_y}^{-1}(B(w, \xi)) \subseteq B(w_y, \xi).$$

Therefore,

$$T_{w_x}^{-1}(B(w, \xi)) \cap T_{w_y}^{-1}(B(w, \xi)) = \emptyset.$$

It then follows from (4.34) that

$$\begin{aligned} T_x^{-k}(B(w, \xi)) \cap T_y^{-k}(B(w, \xi)) &= T_x^{-(k-1)}(T_{w_x}^{-1}(B(w, \xi))) \cap T_y^{-(k-1)}(T_{w_y}^{-1}(B(w, \xi))) \\ &\subseteq T^{-(k-1)}(T_{w_x}^{-1}(B(w, \xi))) \cap T^{-(k-1)}(T_{w_y}^{-1}(B(w, \xi))) \\ &= T^{-(k-1)}(T_{w_x}^{-1}(B(w, \xi)) \cap T_{w_y}^{-1}(B(w, \xi))) \\ &= T^{-(k-1)}(\emptyset) = \emptyset. \end{aligned}$$

Thus $T_x^{-n}(B(z, \xi)) \cap T_y^{-n}(B(z, \xi)) \subseteq T_x^{-k}(B(w, \xi)) \cap T_y^{-k}(B(w, \xi)) = \emptyset$. □

We now give a description of the preimage of any set of small diameter in terms of the local inverse branches of T or of one of its iterates.

Lemma 4.2.9. *For all $z \in X$, for all $A \subseteq B(z, \xi)$ and for all $n \in \mathbb{N}$, we have that*

$$T^{-n}(A) = \bigcup_{x \in T^{-n}(z)} T_x^{-n}(A).$$

Proof. Fix $z \in X$ and $A \subseteq B(z, \xi)$. Since $T^{-n}(A) \supseteq \bigcup_{x \in T^{-n}(z)} T_x^{-n}(A)$ for all $n \in \mathbb{N}$, we only need to prove the opposite inclusion. We shall do this by induction. As the basis of induction, we first do it for $n = 1$. So, let $w \in T^{-1}(A)$. We aim to show that $w \in \bigcup_{x \in T^{-1}(z)} T_x^{-1}(A)$. Since $T(w) \in A \subseteq B(z, \xi)$, we have that $z \in B(T(w), \xi)$. Now define $x := T_w^{-1}(z) \in T^{-1}(z)$. We shall show that $w \in T_x^{-1}(A)$. Recall that $T_x^{-1} : B(T(x), \xi) \to B(x, \xi)$. Since $T(x) = T(T_w^{-1}(z)) = z$, we then have that $T_x^{-1} : B(z, \xi) \to B(x, \xi)$. As $T(w) \in B(z, \xi)$, the point $w' := T_x^{-1}(T(w))$ is well-defined. Moreover, $w' \in T_x^{-1}(A)$, as $T(w) \in A$. Thus, to see that $w \in T_x^{-1}(A)$, it only remains to show that $w' = w$. We know that $T(w') = T(w)$ by definition of w'. So, according to Remark 4.1.2(b), it suffices to show that $d(w', w) < 2\delta$. Using Proposition 4.2.4, observe that

$$d(w', x) = d(T_x^{-1}(T(w)), T_x^{-1}(T(x))) \leq \lambda^{-1} d(T(w), T(x)) = \lambda^{-1} d(T(w), z) < \lambda^{-1}\xi < \delta$$

and

$$d(w, x) = d(T_w^{-1}(T(w)), T_w^{-1}(z)) \leq \lambda^{-1} d(T(w), z) < \lambda^{-1}\xi < \delta.$$

These last two inequalities combine to give

$$d(w',w) \le d(w',x) + d(x,w) < \delta + \delta = 2\delta.$$

So $w = w' := T_x^{-1}(T(w)) \in T_x^{-1}(A)$. We have thus shown that

$$T^{-1}(A) = \bigcup_{x \in T^{-1}(z)} T_x^{-1}(A).$$

For the sake of the inductive step, suppose that the assertion of our lemma holds for all $n = 1, \ldots, k$. Then

$$\begin{aligned} T^{-(k+1)}(A) = T^{-k}(T^{-1}(A)) = T^{-k}\Big(\bigcup_{x \in T^{-1}(z)} T_x^{-1}(A)\Big) \\ &= \bigcup_{x \in T^{-1}(z)} T^{-k}(T_x^{-1}(A)) \\ &= \bigcup_{x \in T^{-1}(z)} \bigcup_{y \in T^{-k}(x)} T_y^{-k}(T_x^{-1}(A)) \\ &= \bigcup_{v \in T^{-(k+1)}(z)} T_v^{-(k+1)}(A). \end{aligned}$$

This completes the proof. □

We now describe conditions under which a transitive system is very strongly transitive (see Definitions 1.5.14–1.5.15).

Lemma 4.2.10. *Every open, distance expanding and transitive dynamical system $T : X \to X$ is very strongly transitive.*

Proof. Given that T is open and X compact, it suffices to show that T is strongly transitive. Let U be an open subset of X. Let also ξ be as in (4.29). According to Theorem 1.5.11, there exists a point $x \in U$ and a number $0 < \xi' \le \xi$ such that $B(x,\xi') \subseteq U$ and $\overline{\mathcal{O}_+(x)} = X$. From (4.33), we deduce that

$$X = \overline{\mathcal{O}_+(x)} = \bigcup_{n=0}^{\infty} B(T^n(x),\xi') \subseteq \bigcup_{n=0}^{\infty} T^n(B(x,\xi')) \subseteq \bigcup_{n=0}^{\infty} T^n(U).$$

Since U is an arbitrary open set, we conclude that T is strongly transitive. □

4.3 Shadowing

Imagine that you observe the dynamics of a system $T : X \to X$ by means of some instrument which is only accurate up to a given $\alpha \ge 0$. In other words, assume that your instrument can only locate the position of a point with a precision at best α.

Then, with your instrument, you will not be able to distinguish points that are within a distance α from each other. In particular, this means that if a point x_0 lands under the map T within a distance α of a point x_1, then you will not be able to distinguish x_1 from the image of x_0 under T. Similarly, if x_1 lands under the map T within a distance α of a point x_2, then you will not be able to distinguish x_2 from the image of x_1 under T, and so on. To summarize, this sequence (x_i) can be mistaken for the orbit of the point x_0, although, in reality, it is not the orbit of x_0 and, in fact, it is not necessarily an orbit at all. The following definition and terminology make this precise.

Definition 4.3.1. Let $\alpha \geq 0$. A sequence $(x_i)_{i=0}^n$, where n can be finite or infinite, is said to be an *α-pseudo-orbit* if

$$d\big(T(x_i), x_{i+1}\big) < \alpha, \quad \forall 0 \leq i < n.$$

In particular, notice that the orbit $\mathcal{O}_+(x)$ of a point $x \in X$ can be written as $\mathcal{O}_+(x) = \{x_0 = x, x_1 = T(x), x_2 = T^2(x), \ldots\}$, which precisely means that $d(T(x_i), x_{i+1}) = 0$ for all $i \in \mathbb{Z}_+$. Thus, an orbit, when converted to the sequence of the iterates of a point, is a 0-pseudo-orbit. Inversely, a 0-pseudo-orbit is merely a sequence of successive iterates of a point.

In the following definition, we come to the important concept of *shadowing* an α-pseudo-orbit, as advertised in the title of this section.

Definition 4.3.2. A point $x \in X$ is said to *β-shadow* a pseudo-orbit $(x_i)_{i=0}^n$ if

$$d\big(T^i(x), x_i\big) < \beta, \quad \forall 0 \leq i \leq n.$$

That is, the orbit of x lies within a distance β of the pseudo-orbit $(x_i)_{i=0}^n$.

Pseudo-orbits and shadowing (along with the forthcoming closing lemma), form a long lived, important, and convenient way of studying dynamical systems exhibiting some sort of hyperbolic or expanding behavior. At the very least, they can be traced back to the breakthrough work of Anosov and Sinai (see [4, 5]). They found a mature, elegant form in [11]. Our approach follows [57], which in turn is based upon [11].

We shall now prove that any infinite sequence in X can be δ-shadowed by at most one point of the space if the dynamical system $T : X \to X$ under consideration is a map expanding balls of radius δ.

Proposition 4.3.3. *Let $T : X \to X$ be a distance expanding map with δ as a constant delimiting the neighborhoods of expansion. Then every infinite sequence of points $(x_i)_{i=0}^\infty$ in X can be δ-shadowed by at most one point of X.*

Proof. Suppose that y and z are two points which each δ-shadow the same sequence $(x_i)_{i=0}^\infty$. For all $i \geq 0$, we then have

$$d\big(T^i(y), x_i\big) < \delta \text{ and } d\big(T^i(z), x_i\big) < \delta.$$

Then, by the triangle inequality, for all $i \geq 0$ we have that

$$d(T^i(y), T^i(z)) < 2\delta.$$

By the expanding property of T, we deduce, for all $i \geq 0$, that

$$d(T^{i+1}(y), T^{i+1}(z)) \geq \lambda\, d(T^i(y), T^i(z)).$$

So, by induction, we conclude that

$$d(T^n(y), T^n(z)) \geq \lambda^n d(y, z)$$

for all $n \geq 0$. However, since the compact space X has finite diameter, this can only happen when $y = z$. □

More generally, we have the following result on the existence and uniqueness of shadowing.

Proposition 4.3.4. *Let $T : X \to X$ be an open, distance expanding map. Let $0 < \beta < \xi$, where ξ is as defined in (4.29). Let $\alpha = \min\{\xi, (\lambda-1)\beta/2\}$ and let $(x_i)_{i=0}^n$ be an α-pseudo-orbit (where n can be finite or infinite). For each $0 \leq i < n$, let $x_i' = T_{x_i}^{-1}(x_{i+1})$. Then:*

(a) *For all $0 \leq i < n$, we have that*

$$T_{x_i'}^{-1}(\overline{B(x_{i+1}, \beta/2)}) \subseteq \overline{B(x_i, \beta/2)}$$

and thus, by induction, the composite map

$$T_{x_0'}^{-1} \circ \cdots \circ T_{x_i'}^{-1} : \overline{B(x_{i+1}, \beta/2)} \to \overline{B(x_0, \beta/2)}$$

is well-defined. Henceforth, we denote this composition by T_i^{-1}.

(b) *$(T_i^{-1}(\overline{B(x_{i+1}, \beta/2)}))_{i=0}^{n-1}$ is a descending sequence of nonempty compact sets.*

(c) *The intersection $\bigcap_{i=0}^{n-1} T_i^{-1}(\overline{B(x_{i+1}, \beta/2)})$ is nonempty and all of its elements β-shadow the α-pseudo-orbit $(x_i)_{i=0}^n$.*

(d) *If $n = \infty$, then $\bigcap_{i=0}^{\infty} T_i^{-1}(\overline{B(x_{i+1}, \beta/2)})$ consists of the unique point that β-shadows the infinite α-pseudo-orbit $(x_i)_{i=0}^{\infty}$.*

Proof. Toward part (a), let $x \in \overline{B(x_{i+1}, \beta/2)}$. Since $x_i' = T_{x_i}^{-1}(x_{i+1})$, we have $T(x_i') = x_{i+1}$ and hence $x_i' = T_{x_i'}^{-1}(T(x_i')) = T_{x_i'}^{-1}(x_{i+1})$. Using Proposition 4.2.4, it follows that

$$\begin{aligned}
d(T_{x_i'}^{-1}(x), x_i) &\leq d(T_{x_i'}^{-1}(x), x_i') + d(x_i', x_i) \\
&= d(T_{x_i'}^{-1}(x), T_{x_i'}^{-1}(x_{i+1})) + d(T_{x_i}^{-1}(x_{i+1}), T_{x_i}^{-1}(T(x_i))) \\
&\leq \lambda^{-1} d(x, x_{i+1}) + \lambda^{-1} d(x_{i+1}, T(x_i)) \\
&< \lambda^{-1}(\beta/2 + \alpha) \leq \lambda^{-1}(\beta/2 + (\lambda-1)\beta/2) = \beta/2.
\end{aligned}$$

Hence $T_{x_i'}^{-1}(x) \in \overline{B(x_i, \beta/2)}$, and this proves the first assertion.

To prove part (b), notice that $T_i^{-1} = T_{i-1}^{-1} \circ T_{x_i'}^{-1}$ for every $1 \le i < n$. Using part (a), we deduce that

$$T_i^{-1}(\overline{B(x_{i+1}, \beta/2)}) = T_{i-1}^{-1} \circ T_{x_i'}^{-1}(\overline{B(x_{i+1}, \beta/2)}) \subseteq T_{i-1}^{-1}(\overline{B(x_i, \beta/2)}).$$

This proves the second assertion.

To prove part (c), recall that the intersection of a descending sequence of nonempty compact sets is a nonempty compact set. From part (b), we readily obtain that $\bigcap_{i=0}^{n-1} T_i^{-1}(\overline{B(x_{i+1}, \beta/2)}) \neq \emptyset$. Moreover, for all $0 \le j < n$, observe that

$$T^j\left(\bigcap_{i=0}^{n-1} T_i^{-1}(\overline{B(x_{i+1}, \beta/2)})\right) \subseteq T^j(T_j^{-1}(\overline{B(x_{j+1}, \beta/2)})) \subseteq \overline{B(x_{j+1}, \beta/2)}.$$

This implies that $d(T^j(x), x_{j+1}) < \beta$ for all $0 \le j < n$ and $x \in \bigcap_{i=0}^{n-1} T_i^{-1}(\overline{B(x_{i+1}, \beta/2)})$. This proves that every such x β-shadows the α-pseudo-orbit $(x_i)_{i=0}^{n}$. If $n = \infty$, Proposition 4.3.3 guarantees that only one such x exists since $\beta < \xi \le \delta$, and part (d) follows. □

We will deduce several important facts from the preceding proposition. Before doing so, we need another definition.

Definition 4.3.5. A map $T : X \to X$ satisfies the *shadowing property* if for all $\beta > 0$ there exists an $\alpha > 0$ such that every infinite α-pseudo-orbit is β-shadowed by a point of the space X.

Corollary 4.3.6 (Existence and uniqueness of shadowing). *Every open, distance expanding map satisfies the shadowing property. Moreover, if β is small enough (namely, if $\beta < \xi$), then one can choose α so that every infinite α-pseudo-orbit is β-shadowed by one and only one point of the space. In fact, α can be chosen as in Proposition* 4.3.4.

In light of the above corollary, we will say that every open, distance expanding map satisfies the *unique-shadowing property*.

Corollary 4.3.7 (Closing lemma). *For every $\beta > 0$, there exists an $\alpha > 0$ with the following property: If x is a point such that $d(T^n(x), x) < \alpha$ for some $n \in \mathbb{N}$, then there exists a periodic point of period n which β-shadows the orbit $(T^i(x))_{i=0}^{n-1}$. In fact, if $\beta < \xi$ then α can be chosen as in Proposition* 4.3.4.

Proof. First, note that if the property holds for some $\widetilde{\beta} > 0$ and a corresponding $\alpha(\widetilde{\beta})$, then it also holds for any $\beta \ge \widetilde{\beta}$ and $\alpha(\beta) = \alpha(\widetilde{\beta})$. Thus we may assume without loss of generality that $0 < \beta < \xi$. Choose α as in Proposition 4.3.4. Now, consider the infinite sequence

$$x, T(x), T^2(x), \ldots, T^{n-1}(x), x, T(x), T^2(x), \ldots, T^{n-1}(x), x, \ldots$$

This sequence can be expressed as $(x_i)_{i=0}^{\infty}$, where $x_{kn+j} = T^j(x)$ for all $k \in \mathbb{Z}_+$ and $0 \le j < n$. We claim that this sequence constitutes an α-pseudo-orbit. Indeed, when $0 \le j < n-1$, we have

$$d(T(x_{kn+j}), x_{kn+j+1}) = d(T(T^j(x)), T^{j+1}(x)) = 0 < \alpha,$$

while when $j = n-1$, we have

$$\begin{aligned} d(T(x_{kn+j}), x_{kn+j+1}) &= d(T(x_{kn+n-1}), x_{kn+n}) = d(T(x_{kn+n-1}), x_{(k+1)n}) \\ &= d(T(T^{n-1}(x)), x) = d(T^n(x), x) < \alpha. \end{aligned}$$

Thus the sequence $(x, T(x), \dots, T^{n-1}(x), x, T(x), \dots, T^{n-1}(x), x, T(x), \dots)$ is an α-pseudo-orbit, and by Corollary 4.3.6 there exists a unique point y which β-shadows it. We also notice that the point $T^n(y)$ β-shadows this infinite sequence, since $d(T^j(T^n(y)), x_j) = d(T^{n+j}(y), x_j) = d(T^{n+j}(y), x_{n+j}) < \beta$. As β-shadowing is unique, we conclude that $T^n(y) = y$. Hence y is a periodic point of period n which β-shadows the orbit $(T^i(x))_{i=0}^{n-1}$. □

From this result, we can infer that any open, distance expanding map has at least one periodic point. Let $\mathrm{Per}(T)$ denote the set of periodic points of T. We shall prove the following.

Corollary 4.3.8 (Closing lemma, existence of a periodic point). *Every open, distance expanding map of a compact metric space has a periodic point. More precisely,* $\mathrm{Per}(T) \subseteq \bigcup_{x\in X} \omega(x) \subseteq \overline{\mathrm{Per}(T)}$*, and as the middle set is nonempty, so is* $\mathrm{Per}(T)$.

Proof. The left-hand side inclusion is immediate as $x \in \omega(x)$ for all $x \in \mathrm{Per}(T)$. In order to prove the right-hand side one, choose any $x \in X$. Recall that the set $\omega(x)$, which is nonempty since X is compact, is the set of accumulation points of the sequence $(T^n(x))_{n=0}^{\infty}$ of iterates of x. Let $y \in \omega(x)$. Fix momentarily an arbitrary $\beta > 0$ and let $\alpha := \alpha(\beta) > 0$ be as in the closing lemma. Then there exists a subsequence $(T^{n_k}(x))_{k=0}^{\infty}$ such that $d(T^{n_k}(x), y) < \alpha/2$ for all $k \in \mathbb{N}$. Therefore, $d(T^{n_k}(x), T^{n_j}(x)) < \alpha$ for all $j, k \in \mathbb{N}$. Fix k and let $j := k+1$. Further, define $z := T^{n_k}(x)$. Then

$$d(T^{n_{k+1}-n_k}(z), z) = d(T^{n_{k+1}}(x), T^{n_k}(x)) < \alpha.$$

According to the closing lemma, there then exists a periodic point w of period $n_{k+1}-n_k$ which β-shadows the orbit $(T^i(z))_{i=0}^{n_{k+1}-n_k-1} = (T^i(x))_{i=n_k}^{n_{k+1}-1}$. Then

$$d(w, y) \le d(w, z) + d(z, y) \le \beta + \alpha/2.$$

This means that there is a periodic point at a distance at most $\beta + \alpha/2$ from y. As β tends to zero, we also have that α tends to zero. Hence, the point y belongs to the closure of the set of periodic points. □

As an immediate consequence of this corollary, we obtain the following result.

Corollary 4.3.9 (Density of periodic points). *The set of periodic points of an open, distance expanding map $T : X \to X$ of a compact metric space X is dense if and only if $\overline{\bigcup_{x\in X} \omega(x)} = X$.*

From the definition of transitivity (cf. Definition 1.5.6), we also obtain the following.

Corollary 4.3.10 (Density of periodic points for transitive maps). *For every transitive, open, distance expanding map of a compact metric space, the set of periodic points is dense.*

In the previous two results, we imposed some restriction on the dynamics of the map. This time we impose some conditions on the space on which the system lives.

Corollary 4.3.11 (Density of periodic points on a connected space). *For every open distance expanding map of a connected compact metric space, the set of periodic points is dense.*

Proof. Fix an arbitrary $x \in X$. We aim to demonstrate that there are periodic points arbitrarily close to x. Let $0 < \beta < \xi$, where ξ was defined in (4.29), and let $\alpha := \alpha(\beta) > 0$ be as in the closing lemma and Proposition 4.3.4. Let $\{U_1, U_2, \ldots, U_p\}$ be a finite open cover of X of diameter less than β (that is, the diameter of each U_i is less than β). Choose any $n \in \mathbb{N}$ such that $(p+1)\lambda^{-n}\beta < \alpha$. Since X is connected, there exists a β-chain of length at most $p+1$ joining x to $T^n(x)$. In other words, there exists a finite sequence

$$x =: y_0, y_1, \ldots, y_{k-1}, y_k := T^n(x)$$

such that $d(y_j, y_{j+1}) < \beta$ for each $0 \le j < k$, where $k \le p$. The elements of the β-chain are chosen to be such that $y_j, y_{j+1} \in U_{i_j}$ for all $0 \le j < k$. By applying an appropriately chosen inverse branch of T^n to this chain, we can construct a $(\lambda^{-n}\beta)$-chain of length at most $p+1$ ending at x. Indeed, let $y_k^{(n)} = T_x^{-n}(T^n(x)) = x$. By recursion on j from $k-1$ to 0, we define $y_j^{(n)} = T^{-n}_{y_{j+1}^{(n)}}(y_j)$. This results in the finite sequence

$$y_0^{(n)} = T^{-n}_{y_1^{(n)}}(y_0), \ldots, y_{k-1}^{(n)} = T^{-n}_{y_k^{(n)}}(y_{k-1}) = T_x^{-n}(y_{k-1}), y_k^{(n)} = x.$$

Observe that for all $0 \le j < k$ we have

$$\begin{aligned}
d(y_j^{(n)}, y_{j+1}^{(n)}) &= d\big(T^{-n}_{y_{j+1}^{(n)}}(y_j), T^{-n}_{y_{j+1}^{(n)}}(T^n(y_{j+1}^{(n)}))\big) \\
&\le \lambda^{-n} d(y_j, T^n(y_{j+1}^{(n)})) \\
&= \lambda^{-n} d(y_j, y_{j+1}) \\
&< \lambda^{-n}\beta.
\end{aligned}$$

Thus we have defined a $(\lambda^{-n}\beta)$-chain of length at most $p+1$ ending at x. Consequently, by the triangle inequality, we deduce that

$$d(y_0^{(n)}, x) = d(y_0^{(n)}, y_k^{(n)}) \leq (k+1)\lambda^{-n}\beta \leq (p+1)\lambda^{-n}\beta < \alpha.$$

Note also that $T^n(y_0^{(n)}) = y_0 = x$. It follows from these last two facts that the infinite sequence

$$y_0^{(n)}, T(y_0^{(n)}), \ldots, T^{n-1}(y_0^{(n)}), T^n(y_0^{(n)}) = x, y_0^{(n)}, T(y_0^{(n)}), \ldots, T^{n-1}(y_0^{(n)}), x, \ldots$$

is an infinite α-pseudo-orbit. Then, according to Proposition 4.3.4(d), there is a unique point z that β-shadows this pseudo-orbit. However, $T^n(z)$ also β-shadows this pseudo-orbit. Thus z is a periodic point of period n. This implies in particular that there exists a periodic point which is β-close to x. As $0 < \beta < \xi$ was chosen arbitrarily, we deduce that the point x is a periodic point or a point of accumulation of periodic points. As x was chosen arbitrarily in X, we conclude that the periodic points of T are dense in X. □

Note that there exist open distance expanding maps defined upon disconnected compact metric spaces whose set of periodic points is not dense (see Exercise 4.6.6).

4.4 Markov partitions

As was alluded to in Chapter 3, symbolic dynamical systems are often used to "represent" other dynamical systems. In the remainder of this chapter, we shall show that an open, expanding map $T : X \to X$ of a compact metric space X can be represented by a subshift of finite type $\sigma : F \to F$, where $F \subseteq E^\infty$ for some finite set E.

In general, one cannot expect that T and σ be topologically conjugate. For instance, T might act on a connected space X, whereas σ always acts on a totally disconnected subshift $F \subseteq E^\infty$. As continuous maps preserve connectedness, it is then out of the question that $\sigma : F \to F$ be a factor of $T : X \to X$, let alone that σ and T be topologically conjugate. In general, the best we may hope for is that T be a factor of σ and that most points of X, ideally points which form a dense G_δ-subset of X, be represented by a unique symbolic point ω in F. Ideally, F would be a subshift of finite type, that is, F would be of the form E_A^∞ for some incidence/transition matrix A. This turns out to be possible.

The construction of such representations can be roughly described as follows. Cover the space X with some special finite collection $\mathcal{R} = \{R_1, R_2, \ldots, R_p\}$ of subsets of X. The collection $\mathcal{R}$ will be called a *Markov "partition"*. We will shortly give some justification for the conditions imposed on a Markov "partition", but first we outline how the "partition" can be used to generate symbolic representations of points in X. The orbit of any point $x \in X$ may be tracked by recording the members of $\mathcal{R}$ in which each iterate of x lands. We may thereby associate to each point x at least one symbolic point $\omega = (\omega_i)_{i=0}^{\infty} \in E^{\infty} := \{1, 2, \ldots, p\}^{\infty}$ such that

$$T^i(x) \in R_{\omega_i}, \quad \forall i \geq 0.$$

Equivalently, this can be expressed by requiring that

$$x \in \bigcap_{i=0}^{\infty} T^{-i}(R_{\omega_i}).$$

However, it is often possible to associate more than one symbolic point to a given point x in this way. For instance, this occurs whenever the orbit of x falls in the nonempty intersection of two members of the Markov "partition" $\mathcal{R}$. We immediately obtain at least two representatives for such an x. In order to achieve a one-to-one association on as large a subset of X as possible (ideally on a dense G_δ-subset of X), we require that the sets R_j intersect as little as possible. Namely, we require that they only intersect within their boundaries, if they intersect at all. Recall from topology that the boundary of a closed set is a nowhere dense set. This justifies condition (b) in the definition of a Markov partition below.

On the other hand, in order that sets of the form $\bigcap_{i=0}^{\infty} T^{-i}(R_{\omega_i})$ each generate at most one point of X, we require that the sets R_j be "small," in some sense. For open expanding maps, this means that the diameters of the R_j should be small enough that the inverse branches of T^i be defined on them (i. e., they should be of diameter less than ξ), so that these inverse branches contract the R_j by a factor λ^{-i}. Moreover, to track the entire orbit of a point x, we usually track its first n iterates and then "take the limit" as n tends to infinity to track the entire orbit. This means that, should the finite intersection $\bigcap_{i=0}^{n} T^{-i}(R_{\omega_i})$ be nonempty for each $n \geq 0$, we would like the infinite intersection $\bigcap_{i=0}^{\infty} T^{-i}(R_{\omega_i})$ to be nonempty. This can be guaranteed by requiring that the R_j be closed or, equivalently, compact.

All of the above requirements can be fulfilled in any compact metric space X. Indeed, it is not too difficult to construct a finite cover $\mathcal{R} = \{R_1, R_2, \ldots, R_p\}$ consisting of closed sets of diameters as small as desired and which intersect only in their boundaries (see Exercise 4.6.10). Although the association $x \mapsto \omega$ is one-to-one on a dense set in X and the closure F of the symbolic points hence generated is a subshift of $E^{\infty} := \{1, 2, \ldots, p\}^{\infty}$, this subshift is generally not of finite type. To ensure that F be of finite type, we impose condition (c) in the definition of a Markov partition. Moreover, condition (a) ensures that closed sets R_j with the property that $R_j = \partial R_j$ are not added to the partition, since such sets only provide information about the dynamics of a negligible set of points of X.

Note that Markov partitions are generally not partitions of the space in the usual sense of disjoint sets, as this would imply that the space is disconnected.

Definition 4.4.1. A finite collection of closed sets $\mathcal{R} = \{R_1, R_2, \ldots, R_p\}$, which covers the space X is called a *Markov partition* for a dynamical system $T : X \to X$ if it satisfies the following three conditions:
(a) $R_i = \overline{\mathrm{Int}(R_i)}$ for all $1 \le i \le p$.
(b) $\mathrm{Int}(R_i) \cap \mathrm{Int}(R_j) = \emptyset$ for all $i \ne j$.
(c) If $T(\mathrm{Int}(R_i)) \cap \mathrm{Int}(R_j) \ne \emptyset$, then $T(R_i) \supseteq R_j$.

In other words, condition (a) requires that each element of a Markov partition be the closure of its interior, condition (b) states that the elements of a Markov partition can only intersect on their boundaries, and condition (c) states that if the image of the interior of an element R_i intersects the interior of an element R_j, then the image of R_i completely covers R_j. Since T is an open map, note that condition (b) is equivalent to
(b′) $R_i \cap \mathrm{Int}(R_j) = \emptyset$ for all $i \ne j$.

For the same reason, condition (c) can be replaced by either of the following conditions:
(c′) If $T(R_i) \cap \mathrm{Int}(R_j) \ne \emptyset$, then $T(R_i) \supseteq R_j$.
(c″) If $T(\mathrm{Int}(R_i)) \cap R_j \ne \emptyset$, then $T(R_i) \supseteq R_j$.

Example 4.4.2. Let T be the shift map $\sigma : E_A^\infty \to E_A^\infty$. Then the one-cylinders $\{[e]\}_{e\in E}$ form a Markov partition for σ. Indeed, every one-cylinder is both open and closed, and hence satisfies condition (a) of the definition of a Markov partition. Condition (b) is clearly satisfied, since words which begin with different letters are distinct. Finally, $\sigma([f]) \cap [e] \ne \emptyset$ means that $A_{fe} = 1$, that is, fe is an admissible word. Therefore, $\sigma([f]) \supseteq \sigma([fe]) = [e]$. Thus condition (c) is satisfied.

Example 4.4.3. Once again, let T be the shift map $\sigma : E_A^\infty \to E_A^\infty$. Fix any $n \in \mathbb{N}$. Then the n-cylinders $\{[\omega]\}_{\omega\in E_A^n}$ form a Markov partition for σ. The proof of this fact is similar to the one given in Example 4.4.2.

Example 4.4.4. Fix $m \in \mathbb{N}$. Recall the map $T_m(x) := mx \pmod 1$ from Example 1.1.3(b). The collection of closed intervals

$$\left\{R_i = \left[\frac{i}{m}, \frac{i+1}{m}\right] : 0 \le i < m\right\}$$

is a Markov partition for T_m. Indeed, one can immediately verify that the first two conditions are satisfied. Concerning condition (c’), observe that $T_m(R_i) = \mathbb{S}^1$ for all i, and thus condition (c’) is fulfilled. Indeed, $T_m(R_i) \supseteq R_j$ for all $0 \le i, j < m$.

Example 4.4.5. Fix $m \in \mathbb{N}$ and consider again the map $T_m(x) := mx \pmod 1$. Now, fix $k \in \mathbb{N}$. The collection of closed intervals

$$\left\{R_i = \left[\frac{i}{m^k}, \frac{i+1}{m^k}\right] : 0 \le i < m^k\right\}$$

is another Markov partition for T_m. Exactly as in the previous example, the first two conditions are clearly satisfied. Concerning condition (c'), observe that

$$T_m(R_i) = \left[\frac{mi}{m^k}, \frac{m(i+1)}{m^k}\right] = \left[\frac{mi}{m^k}, \frac{mi+m}{m^k}\right] = \bigcup_{j=mi}^{mi+m-1}\left[\frac{j}{m^k}, \frac{j+1}{m^k}\right]$$

for all i, and thus condition (c') is fulfilled.

We now present the main result of this section. Examples 4.4.3 and 4.4.5 show that the shift map and the m-times maps T_m, for all $m \in \mathbb{N}$, admit arbitrarily small Markov partitions. This is the case for all open, distance expanding maps, as we show in the next theorem. Part of the proof given here is due to David Simmons.

Theorem 4.4.6 (Existence of Markov partitions). *Every open, distance expanding map $T : X \to X$ of a compact metric space X admits Markov partitions of arbitrarily small diameters.*

Proof. Since T is an open, distance expanding map, it follows from Corollary 4.3.6 that T has the unique-shadowing property. Choose $0 < \beta < \xi/8$. Then there exists $\alpha > 0$ such that every α-pseudo-orbit is β-shadowed by exactly one point of X. As T is continuous on a compact metric space, it is uniformly continuous. Therefore, we can choose $0 < \gamma < \min(\beta, \alpha/2)$ such that for all $x_1, x_2 \in X$ with $d(x_1, x_2) < \gamma$, we know that

$$d\big(T(x_1), T(x_2)\big) < \alpha/2.$$

Step 1. Establishment of a factor map φ between a subshift (Ω, σ) and (T, X).

The collection $\{B(x, \gamma) : x \in X\}$ is an open cover of the compact space X. Therefore, there exists a finite set $E \subseteq X$ such that

$$X = \bigcup_{a \in E} B(a, \gamma).$$

Define the space Ω by

$$\Omega := \big\{\omega = (\omega_i)_{i=0}^{\infty} \in E^{\infty} : d\big(T(\omega_i), \omega_{i+1}\big) < \alpha \text{ for all } i \geq 0\big\}.$$

Observe that $\sigma(\Omega) \subseteq \Omega$. Hence, (Ω, σ) is a subshift of E^{∞} according to Theorem 3.2.4. By definition, each element of the space Ω is an α-pseudo-orbit and, therefore, for each $\omega \in \Omega$ there exists a unique point whose orbit β-shadows ω. Let us call this point $\varphi(\omega)$. In this way, we define a map $\varphi : \Omega \to X$, and by uniqueness of shadowing we have that

$$\varphi \circ \sigma = T \circ \varphi. \tag{4.35}$$

In order for φ to be a factor map, we hope that φ is continuous and surjective. Let us first show that φ is continuous. Let $\omega, \tau \in \Omega$. As $\varphi(\omega)$ β-shadows ω, we have that

$d(T^i(\varphi(\omega)), \omega_i) < \beta$ for all $i \geq 0$. Similarly, $d(T^i(\varphi(\tau)), \tau_i) < \beta$ for all $i \geq 0$. Since $\omega_i = \tau_i$ for all $0 \leq i < |\omega \wedge \tau|$, we can apply the triangle inequality to obtain that

$$d\big(T^i(\varphi(\omega)), T^i(\varphi(\tau))\big) < 2\beta < \xi \leq \delta, \quad \forall 0 \leq i < |\omega \wedge \tau|,$$

where $\delta > 0$ comes from the definition of T being expanding. So,

$$d\big(T^{i+1}(\varphi(\omega)), T^{i+1}(\varphi(\tau))\big) \geq \lambda\, d\big(T^i(\varphi(\omega)), T^i(\varphi(\tau))\big), \quad \forall 0 \leq i < |\omega \wedge \tau|.$$

It follows by a straightforward argument that

$$\begin{aligned} d\big(\varphi(\omega), \varphi(\tau)\big) &\leq \lambda^{-|\omega\wedge\tau|} \operatorname{diam}(X) = \big(s^{|\omega\wedge\tau|}\big)^{-\frac{\log\lambda}{\log s}} \operatorname{diam}(X) \\ &= \operatorname{diam}(X)\big(d_s(\omega, \tau)\big)^{-\frac{\log\lambda}{\log s}}. \end{aligned}$$

Thus φ is Hölder continuous with exponent $-\log\lambda/\log s$ and is therefore continuous.

In order to show that φ is surjective, let $x \in X$. Then, for all $i \geq 0$, we have that $T^i(x) \in B(\omega_i, \gamma)$ for some $\omega_i \in E$. As $d(T^i(x), \omega_i) < \gamma$, it follows from the choice of γ that $d(T^{i+1}(x), T(\omega_i)) < \alpha/2$ for all $i \geq 0$. We deduce that

$$d\big(T(\omega_i), \omega_{i+1}\big) \leq d\big(T(\omega_i), T^{i+1}(x)\big) + d\big(T^{i+1}(x), \omega_{i+1}\big) < \alpha/2 + \gamma < \alpha$$

for all $i \geq 0$. Thus $\omega = \omega_0\omega_1\omega_2\ldots \in \Omega$ and, by construction, x γ-shadows ω, that is, $\varphi(\omega) = x$. The proof of the surjectivity of φ is complete.

Step 2. A property of the images of one-cylinders.

For each $a \in E$, define the sets

$$P_a := \varphi\big([a]\big) = \varphi\big(\{\omega \in \Omega : \omega_0 = a\}\big).$$

All sets P_a are closed in X since they are the images under the factor map φ of the one-cylinder sets $[a]$, which are themselves closed in the compact space Ω. For each $a \in E$, set

$$W(a) := \big\{b \in E : d\big(T(a), b\big) < \alpha\big\}.$$

We claim that the following property is satisfied:

$$T(P_a) = \bigcup_{b\in W(a)} P_b. \tag{4.36}$$

Indeed, if $x \in P_a$, then $x = \varphi(\omega)$ for some $\omega \in \Omega$ with $\omega_0 = a$. By the definition of Ω, it follows that $\omega_1 \in W(a)$. Thus, invoking (4.35), we obtain that $T(x) = T(\varphi(\omega)) = \varphi(\sigma(\omega)) \in \varphi([\omega_1])$, and hence $T(x) \in P_{\omega_1} \subseteq \bigcup_{b\in W(a)} P_b$. Consequently, $T(P_a) \subseteq \bigcup_{b\in W(a)} P_b$. Conversely, let $y \in P_b$ for some $b \in W(a)$. Then $y = \varphi(\omega)$ for some

$\omega \in \Omega$ with $\omega_0 = b$. By the definition of $W(a)$, the concatenation $a\omega$ belongs to the set Ω and therefore, using (4.35) again, we get that

$$y = \varphi(\omega) = \varphi(\sigma(a\omega)) = T(\varphi(a\omega)) \in T(P_a).$$

Consequently, $\bigcup_{b\in W(a)} P_b \subseteq T(P_a)$. Relation (4.36) has been proved. This relation can be expressed by means of an incidence/transition matrix $A : E \times E \to \{0,1\}$ by setting

$$A_{ab} := \begin{cases} 1 & \text{if } T(P_a) \supseteq P_b \\ 0 & \text{if } T(P_a) \not\supseteq P_b. \end{cases}$$

Then (4.36) means that

$$T(P_a) = \bigcup_{\{b\in E\,:\,A_{ab}=1\}} P_b. \tag{4.37}$$

It is also worth observing that $P_a \subseteq B(a,\beta)$ for every $a \in E$. Thus, if $P_a \cap P_b \neq \emptyset$ for some $a,b \in E$, then $P_a \cup P_b \subseteq B(a,4\beta) \cap B(b,4\beta)$. Since $4\beta < \xi/2$, the restriction $T : B(a,4\beta) \to T(B(a,4\beta))$ is a homeomorphism. In particular, T is injective on $P_a \cup P_b$.

Step 3. Construction of the elements of a Markov partition.

For each nonempty subset S of E, define B_S to be

$$B_S := \Big[\bigcap_{a\in S} P_a\Big] \cap \Big[\bigcap_{b\in E\setminus S} (X \setminus P_b)\Big].$$

We claim that the family

$$\mathcal{R} := \{R_S := \overline{\operatorname{Int}(\overline{B_S})} : S \in \mathcal{P}_+(E)\}$$

forms a Markov partition, where $\mathcal{P}_+(E) := \{S \subseteq E : R_S \neq \emptyset\}$.

First, we shall show that for each nonempty subset $S \subseteq E$, the set $T(B_S)$ is a union of elements of $\mathcal{R}$. Toward this end, fix $S \in \mathcal{P}_+(E)$, pick $a_S \in S$ and define

$$E_{a_S} := \{e \in E : P_e \cap P_{a_S} \neq \emptyset\}.$$

Note that $S \subseteq E_{a_S}$. Indeed, $S \in \mathcal{P}_+(E)$ means that $R_S \neq \emptyset$. This implies that $B_S \neq \emptyset$, which in particular implies that $\bigcap_{a\in S} P_a \neq \emptyset$. It ensues that $P_a \cap P_{a_S} \neq \emptyset$ for every $a \in S$.

Moreover, recall that T is injective on $P_a \cup P_b$ whenever $P_a \cap P_b \neq \emptyset$. Then

$$\begin{aligned} T(B_S) &= T\bigg(\Big[\bigcap_{a\in S} P_a\Big] \cap \Big[\bigcap_{b\in E\setminus S} (X \setminus P_b)\Big]\bigg) \\ &= T\bigg(\Big[P_{a_S} \cap \bigcap_{a\in S} P_a\Big] \cap \Big[P_{a_S} \cap \bigcap_{b\in E\setminus S} (X \setminus P_b)\Big]\bigg) \end{aligned}$$

$$
\begin{aligned}
&= T\Big(P_{a_S} \cap \bigcap_{a\in S} P_a\Big) \cap T\Big(P_{a_S} \cap \bigcap_{b\in E\setminus S} (X \setminus P_b)\Big) \\
&= T\Big(\bigcap_{a\in S} P_{a_S} \cap P_a\Big) \cap T\Big(\bigcap_{b\in E\setminus S} P_{a_S} \cap (X \setminus P_b)\Big) \\
&= T\Big(\bigcap_{a\in S} P_{a_S} \cap P_a\Big) \cap T\Big(\bigcap_{b\in E\setminus S} (P_{a_S} \setminus P_b)\Big) \\
&= \Big[\bigcap_{a\in S} T(P_{a_S} \cap P_a)\Big] \cap \Big[\bigcap_{b\in E_{a_S}\setminus S} T(P_{a_S} \setminus P_b)\Big] \\
&\quad \cap \Big[\bigcap_{b\in E\setminus E_{a_S}} T(P_{a_S} \setminus P_b)\Big] \\
&= \Big[\bigcap_{a\in S} \big(T(P_{a_S}) \cap T(P_a)\big)\Big] \cap \Big[\bigcap_{b\in E_{a_S}\setminus S} \big(T(P_{a_S}) \setminus T(P_b)\big)\Big] \\
&\quad \cap \Big[\bigcap_{b\in E\setminus E_{a_S}} T(P_{a_S})\Big] \\
&= \Big[\bigcap_{a\in S} T(P_a)\Big] \cap \Big[\bigcap_{b\in E_{a_S}\setminus S} \big(X \setminus T(P_b)\big)\Big] \\
&= \Big[\bigcap_{a\in S} \bigcup_{\{c\in E\,:\,A_{ac}=1\}} P_c\Big] \cap \Big[\bigcap_{b\in E_{a_S}\setminus S} \bigcap_{\{d\in E\,:\,A_{bd}=1\}} (X \setminus P_d)\Big] \\
&= \Big[\bigcap_{a\in S} \bigcup_{\{c\in E\,:\,A_{ac}=1\}} P_c\Big] \cap \Big[\bigcap_{b\in \widehat{S}^c} (X \setminus P_b)\Big],
\end{aligned}
\tag{4.38}
$$

where $\widehat{S}^c := \bigcup_{b\in E_{a_S}\setminus S}\{d \in E : A_{bd} = 1\}$.

Now, let $x \in T(B_S)$ be arbitrary. Define

$$S(x) := \{b \in E \setminus \widehat{S}^c : x \in P_b\}.$$

Observe that if $e \in S(x)$, then $x \in P_e$. However, if $e \notin S(x)$ then $e \in \widehat{S}^c$ or $x \in X \setminus P_e$. In the former case, there exists $a \in E_{a_S} \setminus S$ such that $A_{ae} = 1$. This implies that $T(P_a) \supseteq P_e$. We will now show that $x \in X \setminus P_e$ also in this case. By way of contradiction, suppose that $x \in P_e$. Then there exists $y \in P_a$ such that $T(y) = x$. On the other hand, since $x \in T(B_S)$, there exists $z \in \bigcap_{i\in S} P_i \cap \bigcap_{j\in E\setminus S}(X \setminus P_j)$ such that $T(z) = x$. As $a_S \in S$, we know that $z \in P_{a_S}$. Thus we have $y, z \in P_a \cup P_{a_S}$ with $T(y) = x = T(z)$. As $a \in E_{a_S}$, we get $P_a \cap P_{a_S} \neq \emptyset$, and hence T is injective on $P_a \cup P_{a_S}$. We deduce that $y = z$. As $a \notin S$, we have $z \in X \setminus P_a$ by definition of z. So $y = z \in P_a \cap (X \setminus P_a) = \emptyset$. This contradiction shows that $x \in X \setminus P_e$. Thus, in either case, if $e \notin S(x)$, then $x \in X \setminus P_e$.

In summary, if $e \in S(x)$ then $x \in P_e$ whereas if $e \notin S(x)$ then $x \in X \setminus P_e$. Consequently,

$$x \in \left[\bigcap_{i\in S(x)} P_i\right] \cap \left[\bigcap_{j\in E\setminus S(x)} (X \setminus P_j)\right] = B_{S(x)}. \tag{4.39}$$

Next, we claim that $B_{S(x)} \subseteq T(B_S)$. Indeed, since $x \in T(B_S)$, it follows from (4.38) that for every $i \in S$, there exists $j_i \in E$ such that $A_{ij_i} = 1$ and $x \in P_{j_i}$. Since $x \in B_{S(x)}$ by (4.39), we deduce that $j_i \in S(x)$ and $P_{j_i} \supseteq B_{S(x)}$. Hence,

$$\bigcap_{i\in S} P_{j_i} \supseteq B_{S(x)}. \tag{4.40}$$

Since $\widehat{S}^c \subseteq E \setminus S(x)$ (by definition of $S(x)$), we have that

$$B_{S(x)} \subseteq \bigcap_{j\in E\setminus S(x)} (X \setminus P_j) \subseteq \bigcap_{j\in\widehat{S}^c} (X \setminus P_j).$$

In conjunction with (4.40), we therefore obtain that

$$B_{S(x)} \subseteq \left[\bigcap_{i\in S} P_{j_i}\right] \cap \left[\bigcap_{j\in\widehat{S}^c} (X \setminus P_j)\right].$$

Thus, according to (4.38),

$$B_{S(x)} \subseteq T(B_S),$$

proving the claim made. It follows immediately that

$$T(B_S) = \bigcup_{x\in T(B_S)} B_{S(x)}.$$

Keep in mind that, though the set $T(B_S)$ generally contains infinitely many points x, the sets $S(x)$ are all subsets of the finite set E and, therefore, there are only finitely many different subsets $S(x)$. Let $\widetilde{S} \subseteq \mathcal{P}(E)$ be the finite set consisting of all different subsets $S(x)$, $x \in T(B_S)$. Then

$$T(B_S) = \bigcup_{Q\in\widetilde{S}} B_Q. \tag{4.41}$$

Since $\overline{\mathrm{Int}(C)} \subseteq \overline{C} = C$ for any closed set C, we have

$$\overline{\mathrm{Int}(R_S)} \subseteq R_S.$$

On the other hand, as $\overline{\mathrm{Int}(Y)} \subseteq \overline{\mathrm{Int}(\overline{\mathrm{Int}(Y)})}$ for any set Y, we have

$$R_S = \overline{\mathrm{Int}(\overline{B_S})} \subseteq \overline{\mathrm{Int}\big(\overline{\mathrm{Int}(\overline{B_S})}\big)} = \overline{\mathrm{Int}(R_S)}.$$

So, condition (a) of Definition 4.4.1 is satisfied.

To verify condition (b), assume that S_1, S_2 are two nonempty subsets of E such that $S_1 \neq S_2$. Without loss of generality, say there exists $e \in S_1 \setminus S_2$. Then $B_{S_1} \subseteq P_e$ and $B_{S_2} \subseteq X \setminus P_e \subseteq X \setminus \mathrm{Int}(P_e)$. Hence, $R_{S_1} \subseteq \overline{B_{S_1}} \subseteq P_e$, and thus

$$\mathrm{Int}(R_{S_1}) \subseteq \mathrm{Int}(P_e).$$

Also,

$$R_{S_2} \subseteq \overline{B_{S_2}} \subseteq \overline{X \setminus \mathrm{Int}(P_e)} = X \setminus \mathrm{Int}(P_e).$$

Therefore,

$$\mathrm{Int}(R_{S_1}) \cap R_{S_2} = \emptyset, \tag{4.42}$$

which is more than enough to prove condition (b).

Aiming now to prove that condition (c) holds, we will first show that $T(R_S) = \bigcup_{Q \in \widetilde{S}} R_Q$. Using (4.41), the fact that T is a homeomorphism on $B(a_S, \xi) \supseteq P_{a_S} \supseteq \overline{B_S}$ and the fact that $\widetilde{S}$ is a finite set, we obtain that

$$\begin{aligned}
T(R_S) &= T\big(\overline{\mathrm{Int}(\overline{B_S})}\big) = \overline{T\big(\mathrm{Int}(\overline{B_S})\big)} \\
&= \overline{\mathrm{Int}\big(T(\overline{B_S})\big)} = \overline{\mathrm{Int}\big(\overline{T(B_S)}\big)} = \overline{\mathrm{Int}\Big(\overline{\bigcup_{Q \in \widetilde{S}} B_Q}\Big)} \\
&= \overline{\mathrm{Int}\Big(\bigcup_{Q \in \widetilde{S}} \overline{B_Q}\Big)} \qquad (4.43) \\
&\supseteq \overline{\bigcup_{Q \in \widetilde{S}} \mathrm{Int}(\overline{B_Q})} = \bigcup_{Q \in \widetilde{S}} \overline{\mathrm{Int}(\overline{B_Q})} \\
&= \bigcup_{Q \in \widetilde{S}} R_Q. \qquad (4.44)
\end{aligned}$$

On the other hand, if $x \in \overline{\mathrm{Int}(\bigcup_{Q \in \widetilde{S}} \overline{B_Q})}$, then for every open set G containing x, we have $H := G \cap \mathrm{Int}(\bigcup_{Q \in \widetilde{S}} \overline{B_Q}) \neq \emptyset$. This means that H is a nonempty open subset of $\bigcup_{Q \in \widetilde{S}} \overline{B_Q}$. Therefore, by virtue of the Baire category theorem, there exists $Q \in \widetilde{S}$ such that $H \cap \mathrm{Int}(\overline{B_Q}) \neq \emptyset$. Thus $G \cap \mathrm{Int}(\overline{B_Q}) \neq \emptyset$. Taking now the sets G to be open balls centered at x with radii converging to zero and recalling that the set $\widetilde{S}$ is finite, we conclude that there exists $Q_x \in \widetilde{S}$ such that $x \in \overline{\mathrm{Int}(\overline{B_{Q_x}})} = R_{Q_x}$. Hence we have shown that

$$\overline{\mathrm{Int}\Big(\bigcup_{Q \in \widetilde{S}} \overline{B_Q}\Big)} \subseteq \bigcup_{Q \in \widetilde{S}} R_Q.$$

Along with (4.43) and (4.44), this yields

$$T(R_S) = \bigcup_{Q \in \widetilde{S}} R_Q. \tag{4.45}$$

So, if $T(R_S) \cap \mathrm{Int}(R_Z) \neq \emptyset$, there exists $Q \subseteq \widetilde{S}$ such that $R_Q \cap \mathrm{Int}(R_Z) \neq \emptyset$, which, by invoking (4.42), yields that $Q = Z$. Employing (4.45), this gives that

$$T(R_S) \supseteq R_Z.$$

This establishes condition (c′).

It only remains to demonstrate that $\mathcal{R}$ is a cover of X. Indeed, $\{B_S : S \subseteq E\}$ is obviously a cover of X. Hence, $\{\overline{B_S} : S \subseteq E\}$ is also a cover of X. Thus, by the same argument as the one above based on the Baire category theorem, we get that

$$\bigcup_{S\subseteq E} R_S = \bigcup_{S\subseteq E} \overline{\mathrm{Int}(\overline{B_S})} = \overline{\bigcup_{S\subseteq E} \mathrm{Int}(\overline{B_S})} = X.$$

Therefore, $\mathcal{R}$ covers X and we are done. □

4.5 Symbolic representation generated by a Markov partition

Let $T : X \to X$ be an open, distance expanding map of a compact metric space X with constants λ and δ. Let $\mathcal{R} = \{R_1, \ldots, R_p\}$ be a Markov partition with $\mathrm{diam}(\mathcal{R}) < \delta$. This partition induces the alphabet $E := \{1, \ldots, p\}$ and an incidence/transition matrix $A : E \times E \to \{0, 1\}$ defined by

$$A_{ij} := \begin{cases} 1 & \text{if } T(\mathrm{Int}(R_i)) \cap \mathrm{Int}(R_j) \neq \emptyset \\ 0 & \text{otherwise.} \end{cases} \tag{4.46}$$

Let $\sigma : E_A^\infty \to E_A^\infty$ be the subshift of finite type induced by A.

Lemma 4.5.1. *If $\omega \in E_A^\infty$, then $\bigcap_{n=0}^\infty T^{-n}(R_{\omega_n})$ is a singleton.*

Proof. For every $i \in E$, the restriction $T|_{R_i} : R_i \to T(R_i)$ is injective since $\mathrm{diam}(R_i) < \delta$ (cf. Proposition 4.2.1). So the inverse map $T_i^{-1} : T(R_i) \to R_i$ is well-defined and is a contraction with ratio λ^{-1}. Note also that if $A_{ij} = 1$ then $T(\mathrm{Int}(R_i)) \cap \mathrm{Int}(R_j) \neq \emptyset$, and as $\mathcal{R}$ is a Markov partition, $R_j \subseteq T(R_i)$. Then $T^{-1}(R_j) \cap R_i \neq \emptyset$. Consequently, for any set $B \subseteq R_j$, we have

$$T^{-1}(B) \cap R_i = T_i^{-1}(B). \tag{4.47}$$

Now, let $\omega \in E_A^\infty$. We claim that

$$\bigcap_{k=0}^{n+1} T^{-k}(R_{\omega_k}) = T_{\omega_0}^{-1} \circ T_{\omega_1}^{-1} \circ \cdots \circ T_{\omega_n}^{-1}(R_{\omega_{n+1}}), \quad \forall n \geq 0.$$

We shall prove this claim by induction. For the case $n = 0$, we have that $R_{\omega_0} \cap T^{-1}(R_{\omega_1}) = T_{\omega_0}^{-1}(R_{\omega_1})$ using (4.47). Suppose now that the claim holds for $n = 0, \ldots, m$. Using (4.47) again, we obtain

$$\begin{aligned}
\bigcap_{k=0}^{(m+1)+1} T^{-k}(R_{\omega_k}) &= R_{\omega_0} \cap \bigcap_{k=1}^{m+2} T^{-k}(R_{\omega_k}) \\
&= R_{\omega_0} \cap T^{-1}\left(\bigcap_{j=0}^{m+1} T^{-j}(R_{\omega_{j+1}})\right) \\
&= R_{\omega_0} \cap T^{-1}\left(\bigcap_{j=0}^{m+1} T^{-j}(R_{(\sigma(\omega))_j})\right) \\
&= R_{\omega_0} \cap T^{-1}\big(T^{-1}_{(\sigma(\omega))_0} \circ \cdots \circ T^{-1}_{(\sigma(\omega))_m}(R_{(\sigma(\omega))_{m+1}})\big) \\
&= R_{\omega_0} \cap T^{-1}\big(T^{-1}_{\omega_1} \circ \cdots \circ T^{-1}_{\omega_{m+1}}(R_{\omega_{m+2}})\big) \\
&= T^{-1}_{\omega_0}\big(T^{-1}_{\omega_1} \circ \cdots \circ T^{-1}_{\omega_{m+1}}(R_{\omega_{m+2}})\big) \\
&= T^{-1}_{\omega_0} \circ T^{-1}_{\omega_1} \circ \cdots \circ T^{-1}_{\omega_{m+1}}(R_{\omega_{(m+1)+1}}).
\end{aligned}$$

So the claim is proved. The claim also shows that $(\bigcap_{k=0}^{n+1} T^{-k}(R_{\omega_k}))_{n=0}^{\infty}$ is a descending sequence of nonempty compact sets. Using the fact that each T_i^{-1} is contracting with ratio λ^{-1}, we obtain

$$\operatorname{diam}\left(\bigcap_{k=0}^{n+1} T^{-k}(R_{\omega_k})\right) \le \lambda^{-(n+1)} \operatorname{diam}(R_{\omega_{n+1}}) \le \lambda^{-(n+1)}\delta.$$

Since $\lim_{n\to\infty} \lambda^{-(n+1)}\delta = 0$, the set $\bigcap_{n=0}^{\infty} T^{-n}(R_{\omega_n})$ is a singleton. □

Thanks to this lemma, the *coding map* $\pi : E_A^{\infty} \to X$, where $\pi(\omega)$ is defined to be the singleton point in the set

$$\bigcap_{n=0}^{\infty} T^{-n}(R_{\omega_n}), \tag{4.48}$$

is well-defined.

Amongst other properties, we want to show that the coding map is Hölder continuous. Recall that on a compact metric space, a map is Hölder continuous if and only if it is locally Hölder continuous. That is, for a map $f : (Y, d_Y) \to (Z, d_Z)$ with Y compact, it is sufficient to show that there exist constants $\delta > 0$ and $C \ge 0$ such that for every $x, y \in Y$ with $d_Y(x,y) < \delta$, we have that $d_Z(f(x), f(y)) \le C(d_Y(x,y))^{\alpha}$.

Theorem 4.5.2. *The coding map $\pi : (E_A^{\infty}, d_s) \to (X, d)$ satisfies the following properties:*
(a) *The map π is Hölder continuous.*
(b) *The map π is surjective.*
(c) *The restriction of π to $\pi^{-1}(X \setminus \bigcup_{n=0}^{\infty} T^{-n}(\bigcup_{i=1}^{p} \partial R_i))$ is injective. So every point of the forward T-invariant, dense G_δ-set $X \setminus \bigcup_{n=0}^{\infty} T^{-n}(\bigcup_{i=1}^{p} \partial R_i)$ has a unique preimage under π.*

(d) *The map π makes the following diagram commutative:*

$$\begin{tikzcd} E_A^\infty \arrow[r, "\sigma"] \arrow[d, "\pi"'] & E_A^\infty \arrow[d, "\pi"] \\ X \arrow[r, "T"'] & X \end{tikzcd}$$

That is, $\pi \circ \sigma = T \circ \pi$.

In particular, π is a factor map between the symbolic system/representation $\sigma : E_A^\infty \to E_A^\infty$ and the original dynamical system $T : X \to X$.

Note that $X \setminus \bigcup_{n=0}^\infty T^{-n}(\bigcup_{i=1}^p \partial R_i)$ is the set of all points in X whose orbit under T never encounters the boundary of the elements of the Markov partition $\mathcal{R} = \{R_1, R_2, \ldots, R_p\}$.

Proof. In order to shorten the notation, in the following proof we write:

$$Z := X \setminus \bigcup_{n=0}^{\infty} T^{-n}\left(\bigcup_{i=1}^{p} \partial R_i\right).$$

(a) We will prove that π is (locally) Lipschitz continuous with respect to the metric $d_{\lambda^{-1}}$. Recall that $d_{\lambda^{-1}}(\omega,\tau) = \lambda^{-|\omega\wedge\tau|}$. Choose $\omega, \tau \in E_A^\infty$ to be such that $|\omega \wedge \tau| \geq 1$. Therefore,

$$\pi(\omega) \in \bigcap_{n=0}^{|\omega\wedge\tau|-1} T^{-n}(R_{\omega_n}) = \bigcap_{n=0}^{|\omega\wedge\tau|-1} T^{-n}(R_{\tau_n}) \ni \pi(\tau).$$

Thus,

$$\begin{aligned} d(\pi(\omega), \pi(\tau)) &\leq \operatorname{diam}\left(\bigcap_{n=0}^{|\omega\wedge\tau|-1} T^{-n}(R_{\omega_n})\right) \\ &\leq \lambda^{-(|\omega\wedge\tau|-1)} \operatorname{diam}(R_{\omega_{|\omega\wedge\tau|}}) \\ &\leq \lambda^{-|\omega\wedge\tau|} \cdot \lambda \operatorname{diam}(X) = (\lambda \operatorname{diam}(X)) d_{\lambda^{-1}}(\omega,\tau). \end{aligned}$$

So π is Lipschitz continuous (i. e., Hölder continuous with exponent $\alpha = 1$) when E_A^∞ is endowed with the metric $d_{\lambda^{-1}}$. Since the metrics d_s, $s \in (0,1)$, are Hölder equivalent (see Exercise 3.4.5), we deduce that π is Hölder continuous with respect to any metric d_s.

(b) We now show that π is surjective. For this, it suffices to show that $Z \subseteq \pi(E_A^\infty)$ and that $\overline{Z} = X$. This is because the map $\pi : E_A^\infty \to X$ is continuous, E_A^∞ is compact and so $\pi(E_A^\infty)$ is compact and thereby closed. We shall first demonstrate that Z is a dense G_δ-set in X. Notice that

$$Z = \bigcap_{n=0}^{\infty} \left[X \setminus T^{-n} \left(\bigcup_{i=1}^{p} \partial R_i \right) \right] = \bigcap_{n=0}^{\infty} T^{-n} \left(X \setminus \bigcup_{i=1}^{p} \partial R_i \right) = \bigcap_{n=0}^{\infty} T^{-n} \left(\bigcap_{i=1}^{p} (X \setminus \partial R_i) \right).$$

Now, the boundary of any closed set is a closed, nowhere dense set (for a closed set, "nowhere dense" amounts to saying that the set has empty interior). Thus the complement of any boundary is an open, dense subset of the space. This means that the sets $X \setminus \partial R_i$, $1 \le i \le p$, are all open, dense subsets of X. Consequently, their finite intersection $\bigcap_{i=1}^{p}(X \setminus \partial R_i)$ is an open, dense subset of X. As the preimage of an open, dense set under a continuous map is open and dense, we deduce that $T^{-n}(\bigcap_{i=1}^{p}(X \setminus \partial R_i))$ is an open, dense subset of X for every $n \ge 0$. Hence, Z is a countable intersection of open sets, that is, Z is a G_δ-set. Moreover, as Z is a countable intersection of open, dense subsets of X (a complete metric space), it follows from the Baire category theorem that Z is dense in X.

Second, let us show that $Z \subseteq \pi(E_A^\infty)$. Let $x \in Z$. We shall find an A-admissible word $\rho = (\rho_n)_{n=0}^{\infty}$ such that $\pi(\rho) = x$. For each $n \ge 0$, the letter ρ_n is selected among those letters of the alphabet E in such a way that $x \in T^{-n}(\mathrm{Int}(R_{\rho_n}))$. Thus

$$x \in \bigcap_{n=0}^{\infty} T^{-n}(\mathrm{Int}(R_{\rho_n})).$$

We show that $\rho \in E_A^\infty$, that is, that $A_{\rho_n \rho_{n+1}} = 1$ for each $n \ge 0$. Indeed,

$$\begin{aligned} A_{\rho_n \rho_{n+1}} = 1 &\Longleftrightarrow T(\mathrm{Int}(R_{\rho_n})) \cap \mathrm{Int}(R_{\rho_{n+1}}) \neq \emptyset \\ &\Longleftrightarrow \mathrm{Int}(R_{\rho_n}) \cap T^{-1}(\mathrm{Int}(R_{\rho_{n+1}})) \neq \emptyset. \end{aligned}$$

As $x \in \bigcap_{n=0}^{\infty} T^{-n}(\mathrm{Int}(R_{\rho_n}))$, we know that

$$\begin{aligned} x &\in T^{-n}(\mathrm{Int}(R_{\rho_n})) \cap T^{-(n+1)}(\mathrm{Int}(R_{\rho_{n+1}})) \\ &= T^{-n}(\mathrm{Int}(R_{\rho_n}) \cap T^{-1}(\mathrm{Int}(R_{\rho_{n+1}}))) \end{aligned}$$

for each $n \ge 0$. Then

$$T^n(x) \in \mathrm{Int}(R_{\rho_n}) \cap T^{-1}(\mathrm{Int}(R_{\rho_{n+1}}))$$

for all $n \ge 0$. In particular, this intersection is nonempty. So $A_{\rho_n \rho_{n+1}} = 1$ for all $n \ge 0$, and $x = \pi(\rho)$ for some $\rho \in E_A^\infty$.

(c) In the course of the proof of (b), we demonstrated that Z is a dense G_δ-set. The fact that Z is forward T-invariant is obvious since this set consists of all points in X whose orbit under T never encounters the boundary of the elements of the

Markov partition. Thus $T(Z) \subseteq Z$. We now show that $\pi^{-1}(x)$ is a singleton for every $x \in Z$. Suppose that $\omega, \tau \in \pi^{-1}(Z)$ are such that $\pi(\omega) = \pi(\tau)$. As $\{\pi(\omega)\} = \bigcap_{n=0}^{\infty} T^{-n}(R_{\omega_n})$, we have that $\pi(\omega) \in T^{-n}(R_{\omega_n})$ for every $n \geq 0$. On the other hand, since $\pi(\omega) \in Z$, we have $\pi(\omega) \notin T^{-n}(\partial R_{\omega_n})$ for every $n \geq 0$. It therefore follows that $\pi(\omega) \in T^{-n}(\mathrm{Int}(R_{\omega_n}))$ for every $n \geq 0$. Similarly, $\pi(\tau) \in T^{-n}(\mathrm{Int}(R_{\tau_n}))$ for every $n \geq 0$. Since $\pi(\omega) = \pi(\tau) =: x$, we deduce that $T^n(x) \in \mathrm{Int}(R_{\omega_n}) \cap \mathrm{Int}(R_{\tau_n})$ for every $n \geq 0$. Condition (b) imposed on the Markov partition forces $\omega_n = \tau_n$ for each $n \geq 0$, that is, $\omega = \tau$.

(d) Finally, we show that the diagram in statement (d) does indeed commute. Let $\omega \in E_A^{\infty}$. Then

$$\begin{aligned} T(\{\pi(\omega)\}) = T\left(\bigcap_{n=0}^{\infty} T^{-n}(R_{\omega_n})\right) &\subseteq \bigcap_{n=0}^{\infty} T^{-(n-1)}(R_{\omega_n}) \\ &= T(R_{\omega_0}) \cap \bigcap_{n=1}^{\infty} T^{-(n-1)}(R_{(\sigma(\omega))_{n-1}}) \\ &\subseteq \bigcap_{m=0}^{\infty} T^{-m}(R_{(\sigma(\omega))_m}) = \{\pi(\sigma(\omega))\}. \end{aligned}$$

Since both the left- and the right-hand sides are singletons, equality follows, and $T(\pi(\omega)) = \pi(\sigma(\omega))$ for all $\omega \in E_A^{\infty}$. That is, $T \circ \pi = \pi \circ \sigma$. □

Example 4.5.3. Consider again the map $T_m : \mathbb{S}^1 \to \mathbb{S}^1$, which is given by $T_m(x) := mx \pmod 1$. We have seen in Example 4.4.4 that the family of closed intervals $\{[\frac{i}{m}, \frac{i+1}{m}] : 0 \leq i < m\}$ forms a Markov partition for T_m. One can then show that the coding map generated by this partition is given, for every $\omega := (\omega_k)_{k=0}^{\infty} \in \{0, \ldots, m-1\}^{\infty}$, by

$$\pi(\omega) = \sum_{k=0}^{\infty} \frac{\omega_k}{m^{k+1}}.$$

In particular, if $m = 2$, we obtain the binary coding of each point $x \in \mathbb{S}^1$.

To end this chapter, we express basic properties of Markov partitions in symbolic terms and show that σ inherits some dynamical properties from T.

Lemma 4.5.4. *Let $\mathcal{R} = \{R_1, R_2, \ldots, R_p\}$ be a Markov partition for an open, distance expanding dynamical system $T : X \to X$. Let $E = \{1, 2, \ldots, p\}$, A as in (4.46) and $\sigma : E_A^{\infty} \to E_A^{\infty}$ be the symbolic representation of T induced by $\mathcal{R}$. Let also $n \in \mathbb{N}$.*
(a) *For every $i \in E$, it holds that $T^n(R_i) = \bigcup_{j \in E : A_{ij}^n \neq 0} R_j$.*
(b) *$\mathcal{R}$ is a Markov partition for T^n.*
(c) *If T is topologically transitive, then so is σ.*
(d) *If T is topologically exact, then so is σ.*

(e) *Let $\partial\mathcal{R} := \bigcup_{e\in E}\partial R_e$. If T is a local homeomorphism on a neighborhood of each element of $\mathcal{R}$, then the set $\partial\mathcal{R}$ is forward T-invariant while the set $\pi^{-1}(X\setminus\partial\mathcal{R})$ is backward σ-invariant.*

Proof.

(a) Fix $i\in E$. First, assume that $n=1$. Recall that

$$R_\ell = \overline{\mathrm{Int}(R_\ell)},\quad \forall \ell\in E. \tag{4.49}$$

Since T is continuous and X a compact metric space, it follows that

$$T(R_\ell) = T\big(\overline{\mathrm{Int}(R_\ell)}\big) = \overline{T\big(\mathrm{Int}(R_\ell)\big)},\quad \forall \ell\in E. \tag{4.50}$$

If $A_{ij}\neq 0$, that is, if $T(\mathrm{Int}(R_i))\cap\mathrm{Int}(R_j)\neq\emptyset$, then $T(R_i)\supseteq R_j$ since $\mathcal{R}$ is a Markov partition. Therefore, $T(R_i)\supseteq\bigcup_{j\in E:A_{ij}\neq 0}R_j$. If it turned out that $T(R_i)\neq\bigcup_{j\in E:A_{ij}\neq 0}R_j$ then we would have $T(R_i)\cap[X\setminus\bigcup_{j\in E:A_{ij}\neq 0}R_j]\neq\emptyset$. By (4.50) and the openness of $X\setminus\bigcup_{j\in E:A_{ij}\neq 0}R_j$, this would imply that $T(\mathrm{Int}(R_i))\cap[X\setminus\bigcup_{j\in E:A_{ij}\neq 0}R_j]\neq\emptyset$. Since $X\setminus\bigcup_{j\in E:A_{ij}\neq 0}R_j\subseteq\bigcup_{k\in E:A_{ik}=0}R_k$, we would deduce that $T(\mathrm{Int}(R_i))\cap[\bigcup_{k\in E:A_{ik}=0}R_k]\neq\emptyset$. By (4.49) and the openness of $T(\mathrm{Int}(R_i))$, it would ensue that $T(\mathrm{Int}(R_i))\cap\mathrm{Int}(R_{\tilde{k}})\neq\emptyset$ for some $\tilde{k}$ such that $A_{i\tilde{k}}=0$. But $T(\mathrm{Int}(R_i))\cap\mathrm{Int}(R_{\tilde{k}})\neq\emptyset$ means that $A_{i\tilde{k}}=1$. This contradiction imposes that

$$T(R_i) = \bigcup_{j\in E\,:\,A_{ij}\neq 0} R_j.$$

This is the basic step in this proof by induction. For the inductive step, suppose that the statement holds for some $n\in\mathbb{N}$, that is, $T^n(R_i)=\bigcup_{j\in E\,:\,A^n_{ij}\neq 0}R_j$. Then

$$\begin{aligned}
T^{n+1}(R_i) = T(T^n(R_i)) &= T\Big(\bigcup_{j\in E\,:\,A^n_{ij}\neq 0} R_j\Big)\\
&= \bigcup_{j\in E\,:\,A^n_{ij}\neq 0} T(R_j) = \bigcup_{j\in E\,:\,A^n_{ij}\neq 0}\ \bigcup_{k\in E\,:\,A_{jk}\neq 0} R_k \qquad (4.51)\\
&= \bigcup_{k\in E\,:\,A^{n+1}_{ik}\neq 0} R_k. \qquad (4.52)
\end{aligned}$$

(b) If $T^n(\mathrm{Int}(R_i))\cap\mathrm{Int}(R_j)\neq\emptyset$, then by (a) there is $k\in E$ such that $A^n_{ik}\neq 0$ and $R_k\cap\mathrm{Int}(R_j)\neq\emptyset$. Since $\mathcal{R}$ is a Markov partition, we infer that $R_k=R_j$, and thus $k=j$. Therefore, $A^n_{ij}\neq 0$ and, by (a) again, we conclude that $T^n(R_i)\supseteq R_j$. Consequently, $\mathcal{R}$ is a Markov partition for T^n.

(c) Let $i,j\in E$. If T is topologically transitive, then there exists $n\in\mathbb{N}$ such that $T^n(\mathrm{Int}(R_i))\cap\mathrm{Int}(R_j)\neq\emptyset$. By (a) and (b), it follows that $A^n_{ij}\neq 0$. Since $i,j\in E$ are arbitrary, this means that the matrix A is irreducible, that is, the shift map σ is transitive according to Theorem 3.2.14.

(d) If T is topologically exact, then for every open subset U of X there exists $N(U) \in \mathbb{N}$ such that $T^{N(U)}(U) = X$. In particular, for every $i \in E$ there is $N(i) \in \mathbb{N}$ such that $T^{N(i)}(\mathrm{Int}(R_i)) = X$. Let $N = \max\{N(i) : i \in E\}$. Then $T^N(\mathrm{Int}(R_i)) = X$ for all $i \in E$. So $T^N(\mathrm{Int}(R_i)) \cap \mathrm{Int}(R_j) \neq \emptyset$ for all $i, j \in E$. By (a) and (b), it follows that $A^N_{ij} \neq 0$ for all $i, j \in E$. This precisely means that $A^N > 0$, and thus the matrix A is primitive, that is, the shift map σ is topologically exact according to Theorem 3.2.16.

(e) For any $i \in E$, given that T is by hypothesis a local homeomorphism on a neighborhood of R_i, it follows from (a) that

$$T(\partial R_i) = \partial T(R_i) = \partial\Big(\bigcup_{j \in E : A_{ij} \neq 0} R_j\Big) \subseteq \bigcup_{j \in E : A_{ij} \neq 0} \partial R_j.$$

Consequently,

$$T(\partial \mathcal{R}) = \bigcup_{i \in E} T(\partial R_i) \subseteq \bigcup_{j \in E} \partial R_j = \partial \mathcal{R}.$$

That is, the set $\partial\mathcal{R}$ is forward T-invariant. According to Theorem 4.5.2, the coding map π is a factor map between $\sigma : E_A^\infty \to E_A^\infty$ and $T : X \to X$. By Remark 1.4.4, the set $\pi^{-1}(\partial\mathcal{R})$ is forward σ-invariant. Thus its complement $E_A^\infty \backslash \pi^{-1}(\partial\mathcal{R}) = \pi^{-1}(X \backslash \partial\mathcal{R})$ is backward σ-invariant. □

4.6 Exercises

Exercise 4.6.1. Let $S, T : X \to X$ be two distance expanding maps on a compact metric space (X, d). Show that $S \circ T : X \to X$ is a distance expanding map. Deduce from this that every iterate T^n, $n \in \mathbb{N}$, of T is distance expanding.

Exercise 4.6.2. Suppose that $T : X \to X$ is a continuous map on a compact metric space (X, d) whose nth iterate T^n is distance expanding with constant of expansion λ and constant delimiting the neighborhoods of expansion δ. Prove that T is distance expanding with the same constants λ and δ when X is endowed with the metric d' defined by

$$d'(x, y) = \sum_{k=0}^{n-1} \frac{1}{\lambda^k}\, d(T^k(x), T^k(y)).$$

Show also that the metrics d and d' are topologically equivalent. (Exercises 4.6.1 and 4.6.2 thus prove that T is distance expanding if and only if all its iterates are distance expanding. However, they may be expanding distances with respect to different, though topologically equivalent, metrics.)

Exercise 4.6.3. Inspiring yourself from the proof of Lemma 4.2.2, show that $p > 0$ in the proof of Theorem 4.1.5.

Exercise 4.6.4. Let $(X_n)_{n=0}^{\infty}$ be a descending sequence of compact sets in a metric space (X,d). Let $(\delta_n)_{n=0}^{\infty}$ be a sequence of positive numbers that converges to 0. Show that

$$\bigcap_{n=0}^{\infty} X_n = \bigcap_{n=0}^{\infty} B(X_n,\delta_n).$$

Exercise 4.6.5. Find an open map $T : X \to X$ that admits a subsystem $T|_F : F \to F$ which is not open, where F is a closed forward T-invariant subset of X.

Exercise 4.6.6. Find an open, distance expanding dynamical system $T : X \to X$ defined upon a disconnected, compact metric space X which does not have a dense set of periodic points.

Exercise 4.6.7. Prove that the set of conditions (a), (b), and (c') defines a Markov partition.

Exercise 4.6.8. Let T be the shift map $\sigma : E_A^{\infty} \to E_A^{\infty}$. Fix any $n \in \mathbb{N}$. Prove that the n-cylinders $\{[\omega]\}_{\omega\in E_A^n}$ form a Markov partition for σ.

Exercise 4.6.9. Show that the preimage of an open, dense set under a continuous map is open and dense.

Exercise 4.6.10. Let X be a compact metric space and let $\delta > 0$. Construct a finite cover of X consisting of closed sets of diameters less than δ which intersect only in their boundaries.

Exercise 4.6.11. Let X be an infinite compact metric space and $T : X \to X$ be a transitive, open, distance expanding map. Show that T is not minimal.

Exercise 4.6.12. Let (X,d) be a compact metric space. Define a metric ρ on $X \times \{0,1\}$ by setting

$$\rho\big((x_1,x_2),(y_1,y_2)\big) := d(x_1,y_1) + |x_2 - y_2|.$$

Show that if $T : X \to X$ is a distance expanding map, then the map $\widehat{T} : X \times \{0,1\} \to X \times \{0,1\}$ given by the formula

$$\widehat{T}(x_1,x_2) := \big(T(x_1), x_2 +_2 1\big)$$

is also distance expanding, where $+_2 : \{0,1\} \to \{0,1\}$ denotes addition modulo 2.

Exercise 4.6.13. For every $n \in \mathbb{N}$, show that a distance expanding map can have only finitely many periodic points of period n.

Exercise 4.6.14. Let $T : X \to X$ be an open distance expanding map. Show that the function $x \mapsto \#(T^{-1}(x))$ is locally constant.

Exercise 4.6.15. Suppose that $S, T : X \to X$ are two arbitrary dynamical systems. Define the distance between S and T by the formula

$$d_\infty(S, T) := \sup\{d(S(x), T(x)) : x \in X\}.$$

Now suppose that S and T are both open distance expanding maps with the same parameters δ, λ and ξ. Show that if $d_\infty(S, T) < \min\{\xi, \xi(\lambda - 1)\}$, then S and T are topologically conjugate (cf. the discussion on structural stability at the end of Section 1.2).

5 (Positively) expansive maps

There are various concepts of expansion of a map which have aroused the interest of a great many mathematicians. We have already encountered distance expanding maps; in the present chapter we will introduce positively expansive maps. The study of this class of maps goes back to the 1960s. Such maps are abundant. In particular, all distance expanding maps are expansive, and so, more particularly, all subshifts over a finite alphabet are expansive.

Amidst the large variety of dynamical behaviors which can be thought of as expansion in some sense, expansiveness has turned out to be a rather weak but nevertheless useful mathematical notion. It is a topological concept, in the sense that it is a topological conjugacy invariant. Expansive maps are important for many reasons. One of them is that the entropy function is upper semi-continuous within this class. In particular, all expansive maps admit a measure of maximal entropy and, more generally, have equilibrium states under all continuous potentials.

In Section 5.1, we introduce the concept of expansiveness. In Section 5.2, we define the notion of uniform expansiveness and prove that expansiveness and uniform expansiveness are one and the same notion on compact metrizable spaces. In Section 5.3, we demonstrate that every expansive system is in fact expanding with respect to some metric compatible with the topology on the underlying space. This important fact is due to Coven and Reddy [17] (cf. [18]). It signifies that many of the results proved in Chapter 4, such as the existence of Markov partitions and of a nice symbolic representation, the density of periodic points, the closing lemma, and shadowing, hold for all positively expansive maps. Finally, in Section 5.4 we provide a class of examples of expansive maps that are not expanding. They generate what are called parabolic Cantor sets.

5.1 Expansiveness

The notion of expansiveness was introduced by Utz [73] in 1950 for homeomorphisms. He used the term *unstable* homeomorphisms to describe these maps. Gottschalk and Hedlund [25] later suggested the term *expansive* homeomorphisms, which has been used ever since. Five years after Utz, Williams [78] investigated positive expansiveness of maps. Among other early contributors are Bryant [14], Keynes and Robertson [37], Sears [63], and Coven and Reddy [17].

Definition 5.1.1. A topological dynamical system $T : (X, d) \to (X, d)$ is said to be *(positively) expansive* provided that there exists a constant $\delta > 0$ such that for every $x \neq y$ there is $n = n(x, y) \geq 0$ with

$$d(T^n(x), T^n(y)) > \delta.$$

https://doi.org/10.1515/9783110702682-005

The constant δ is called an *expansive constant* for T, and T is then said to be δ-*expansive*. Equivalently, T is δ-expansive if

$$\sup_{n\geq 0} d\big(T^n(x), T^n(y)\big) \leq \delta \quad \Longrightarrow \quad x = y.$$

In other words, δ-expansiveness means that two forward T-orbits that remain forever within a distance δ from each other originate from the same point, and are therefore only one orbit.

We now note some important and interesting properties of expansiveness.

Remark 5.1.2.
(a) If T is δ-expansive, then T is δ'-expansive for any $0 < \delta' < \delta$.
(b) The expansiveness of T is independent of topologically equivalent metrics, although particular expansive constants generally depend on the metric chosen. See Exercise 5.5.1.
(c) In light of (b), the concept of expansiveness can be defined solely in topological terms, meaning without a reference to a metric. See Exercise 5.5.2.
(d) Unlike the expanding property examined in Chapter 4, expansiveness is a topological conjugacy invariant. See Exercise 5.5.3.

The expansiveness of a system can also be expressed in terms of the following "dynamical" metrics. These metrics are sometimes called Bowen's metrics, since Bowen [10] used them extensively in defining topological entropy for noncompact dynamical systems.

Definition 5.1.3. Let $T : (X, d) \to (X, d)$ be a dynamical system. For every $n \in \mathbb{N}$, let $d_n : X \times X \to [0, \infty)$ be the metric

$$d_n(x, y) := \max\{d\big(T^j(x), T^j(y)\big) : 0 \leq j < n\}.$$

Although the notation does not make explicit the dependence on T, it is crucial to remember that the metrics d_n arise from the dynamics of the system T. It is in this sense that they are dynamical metrics. The corresponding balls will be denoted by

$$B_n(x, \varepsilon) := \{y \in X : d_n(x, y) < \varepsilon\}.$$

Observe that $d_1 = d$ and that for each $x, y \in X$ we have $d_n(x, y) \geq d_m(x, y)$ whenever $n \geq m$. Moreover, it is worth noticing that the metrics d_n, $n \in \mathbb{N}$, are topologically equivalent. Indeed, given a sequence $(x_k)_{k=1}^{\infty}$ in X, the continuity of T ensures that

$$\begin{aligned}\lim_{k\to\infty} d(x_k, y) = 0 &\Longleftrightarrow \lim_{k\to\infty} d\big(T^j(x_k), T^j(y)\big) = 0, \quad \forall 0 \leq j < n,\ \forall n \in \mathbb{N} \\ &\Longleftrightarrow \lim_{k\to\infty} d_n(x_k, y) = 0, \quad \forall n \in \mathbb{N}.\end{aligned}$$

Furthermore, it is easy to see that a dynamical system $T : (X,d) \to (X,d)$ is δ-expansive if and only if

$$\sup_{n\in\mathbb{N}} d_n(x,y) \le \delta \quad \Longrightarrow \quad x = y.$$

Example 5.1.4. Every subshift of finite type $\sigma : E_A^\infty \to E_A^\infty$ is expansive, with any $\delta \in (0,1)$ as an expansive constant. Indeed, note that if $\omega = \omega_1\omega_2\omega_3\ldots$ and $\tau = \tau_1\tau_2\tau_3\ldots$ are any two distinct elements of E_A^∞, then there exists $n \in \mathbb{N}$ such that $\omega_n \neq \tau_n$. Hence,

$$d_s(\sigma^{n-1}(\omega), \sigma^{n-1}(\tau)) = s^{-|\omega_n\omega_{n+1}\ldots\wedge\tau_n\tau_{n+1}\ldots|} = s^0 = 1 > \delta.$$

This example is an instance of the fact that any distance expanding dynamical system is expansive.

Proposition 5.1.5. *Every distance expanding dynamical system is expansive.*

Proof. Let $T : X \to X$ be a distance expanding dynamical system on a compact metric space (X,d), and let δ and λ be constants determining neighborhoods of expansion and magnitude of that expansion, respectively, per Definition 4.1.1. We will show that any $0 < \delta' < 2\delta$ is an expansive constant for T. Let $x, y \in X$ be such that

$$\sup_{n\ge 0} d(T^n(x), T^n(y)) \le \delta'. \tag{5.1}$$

Then

$$d(T^n(x), T^n(y)) \ge \lambda\, d(T^{n-1}(x), T^{n-1}(y)), \quad \forall n \in \mathbb{N}.$$

By induction, it follows that

$$d(T^n(x), T^n(y)) \ge \lambda^n d(x,y), \quad \forall n \ge 0.$$

Therefore,

$$0 \le d(x,y) \le \limsup_{n\to\infty} \lambda^{-n} \mathrm{diam}(X) = 0.$$

Hence $d(x,y) = 0$, that is, $x = y$. So T is δ'-expansive for all $0 < \delta' < 2\delta$. $\square$

5.2 Uniform expansiveness

In this section, we introduce a certain type of uniformity for expansiveness. In 1962, Bryant [14] remarked on the uniformity in the expansiveness of compact dynamical systems. This uniform expansiveness was formalized and studied by Sears [63] eleven years later.

Definition 5.2.1. A topological dynamical system $T : (X,d) \to (X,d)$ is said to be *(positively) uniformly expansive* if there exists $\delta > 0$ with the property that for every

$0 < \zeta < \delta$ there is $N = N(\zeta) \in \mathbb{N}$ such that

$$d(x,y) > \zeta \Longrightarrow d_N(x,y) > \delta.$$

The constant δ is called a *uniformly expansive constant* for T, and T is then said to be *uniformly δ-expansive*.

It is easy to check that every uniformly δ-expansive system is δ-expansive (this is left to the reader). It turns out that the converse is also true for systems that are defined on a compact metric space.

Proposition 5.2.2. *A topological dynamical system $T : (X,d) \to (X,d)$ is δ-expansive if and only if it is uniformly δ-expansive.*

Proof. As mentioned above, it is straightforward to check that every uniformly δ-expansive system is δ-expansive. To prove the converse, suppose by way of contradiction that $T : (X,d) \to (X,d)$ is a δ-expansive system that is not uniformly δ-expansive. Then there exist $0 < \delta' < \delta$ and sequences $(x_n)_{n=0}^{\infty}$ and $(y_n)_{n=0}^{\infty}$ in X such that $d(x_n,y_n) > \delta'$ but $d_n(x_n,y_n) \le \delta$ for all $n \ge 0$. Since X is compact, we may assume (by passing to subsequences if necessary) that the sequences $(x_n)_{n=0}^{\infty}$ and $(y_n)_{n=0}^{\infty}$ converge to, say, x and y, respectively. On one hand, this implies that

$$d(x,y) = \lim_{n\to\infty} d(x_n,y_n) \ge \delta' > 0,$$

and hence $x \neq y$. On the other hand, if we fix momentarily $N \in \mathbb{N}$, for all $n \ge N$, we have that

$$d_N(x_n,y_n) \le d_n(x_n,y_n) \le \delta.$$

Therefore,

$$d_N(x,y) = \lim_{n\to\infty} d_N(x_n,y_n) \le \delta.$$

Since $d_N(x,y) \le \delta$ for every $N \in \mathbb{N}$, the δ-expansiveness of the system implies that $x = y$. This is, of course, in contradiction with our previous deduction that $x \neq y$. Thus, T is uniformly δ-expansive. □

Remark 5.2.3. Note that expansiveness and uniform expansiveness are distinct concepts in the realm of noncompact dynamical systems (see [63]).

Finally, we record the following observation, which is interesting on its own but will also be used in Section 9.6 on Brin–Katok's local entropy formula.

Observation 5.2.4. Let $T : X \to X$ be an expansive topological dynamical system and let d be a metric compatible with the topology on X. If $\delta > 0$ is an expansive constant for T corresponding to this metric, then for every $x \in X$, every $\zeta \in (0,\delta]$ and every integer $n > N(\zeta)$, note that

$$B_n(x,\delta) \subseteq B_{n-N(\zeta/2)}(x,\zeta).$$

Proof. Let $y \in X \setminus B_{n-N(\zeta/2)}(x,\zeta)$. Then there exists a least integer $0 \le k < n - N(\zeta/2)$ such that $T^k(y) \notin B(T^k(x),\zeta)$. This means that $d(T^k(y),T^k(x)) \ge \zeta > \zeta/2$. It follows from Definition 5.2.1 that $d_{N(\zeta/2)}(T^k(y),T^k(x)) > \delta$. Thus $d_{k+N(\zeta/2)}(y,x) > \delta$. Hence,

$$y \notin B_{k+N(\zeta/2)}(x,\delta) \supseteq B_{n-N(\zeta/2)+N(\zeta/2)}(x,\delta) = B_n(x,\delta).$$

So $y \in X \setminus B_n(x,\delta)$. That is, $X \setminus B_{n-N(\zeta/2)}(x,\zeta) \subseteq X \setminus B_n(x,\delta)$. □

5.3 Expansive maps are expanding with respect to an equivalent metric

The aim in this section is to provide a partial converse to Proposition 5.1.5, where it was shown that every distance expanding map is expansive.

Theorem 5.3.1. *If a topological dynamical system $T : X \to X$ is expansive, then there exists a metric, compatible with the topology of X, with respect to which T is distance expanding.*

The original proof that an expansive map defined upon a compact metric space is expanding with respect to a topologically equivalent metric is due to Coven and Reddy [17]. The proof we now present differs slightly by using uniform expansiveness. Like that of Coven and Reddy, the proof relies on a topological lemma, which we state here without proof (cf. [23]).

Frink's Metrization Lemma. *Let X be a metrizable space and let $(U_n)_{n=0}^{\infty}$ be a sequence of open neighborhoods of the diagonal $\triangle := \{(x,x) : x \in X\}$ of $X \times X$ having the following three properties:*
(a) $U_0 = X \times X$.
(b) $\bigcap_{n=0}^{\infty} U_n = \triangle$.
(c) $U_n \circ U_n \circ U_n \subseteq U_{n-1}$ *for every $n \in \mathbb{N}$, where*

$$R \circ S := \big\{(x,y) \in X \times X : \exists z \in X \text{ with } (x,z) \in R \text{ and } (z,y) \in S\big\}.$$

Then there exists a metric ρ, compatible with the topology of X, such that

$$U_n \subseteq \big\{(x,y) \in X \times X : \rho(x,y) < 2^{-n}\big\} \subseteq U_{n-1}$$

for every $n \in \mathbb{N}$.

Proof of Theorem 5.3.1. We shall construct a family of sets that satisfies the hypotheses of Frink's lemma, and then show that some iterate of T is expanding with respect to Frink's metric.

Since T is expansive, Proposition 5.2.2 implies that it is uniformly expansive. Let $3\theta > 0$ be a uniformly expansive constant for T with respect to a metric d compatible with the topology of X. For all $n \geq 0$ and all $\varepsilon > 0$, let

$$V_n(\varepsilon) := \{(x,y) \in X \times X : d_n(x,y) < \varepsilon\}.$$

Each set $V_n(\varepsilon)$ is an open neighborhood of the diagonal $\triangle$ in $X \times X$. The set $V_n(\varepsilon)$ is the collection of all couples of points whose forward orbits stay within a distance ε from each other up to and including time $n-1$. Let $M \geq 0$ be such that $d_M(x,y) > 3\theta$ whenever $d(x,y) > \theta/2$ (cf. Definition 5.2.1). Then no couple (x,y) such that $d(x,y) \geq \theta$ belongs to $V_M(3\theta)$. Hence,

$$V_M(3\theta) \subseteq \{(x,y) \in X \times X : d(x,y) < \theta\} = V_0(\theta). \tag{5.2}$$

Now, let $U_0 = X \times X$ and define

$$U_n := V_{Mn}(\theta) = \{(x,y) \in X \times X : d_{Mn}(x,y) < \theta\}$$

for each $n \in \mathbb{N}$. We shall show that the sets $(U_n)_{n=0}^{\infty}$ satisfy the three conditions of Frink's lemma. The first condition is satisfied by definition of U_0. Regarding the second condition, it is clear that $\triangle \subseteq \bigcap_{n=0}^{\infty} U_n$ because $\triangle \subseteq U_n$ for each $n \geq 0$. For the opposite inclusion, let $(x,y) \in \bigcap_{n=0}^{\infty} U_n$. Then $(x,y) \in U_n := V_{Mn}(\theta)$ for all $n \in \mathbb{N}$. Hence, $d_{Mn}(x,y) < \theta$ for all $n \in \mathbb{N}$, or, in other words, $d(T^j(x), T^j(y)) < \theta$ for all $j \geq 0$. As θ is an expansive constant for T, we deduce that $x = y$. So $\bigcap_{n=0}^{\infty} U_n \subseteq \triangle$. Since both inclusions hold, we conclude that

$$\bigcap_{n=0}^{\infty} U_n = \triangle.$$

It only remains to show that the third condition is satisfied, namely that

$$U_n \circ U_n \circ U_n \subseteq U_{n-1}, \quad \forall n \in \mathbb{N}.$$

For this, fix $n \in \mathbb{N}$ and let $(x,y) \in U_n \circ U_n \circ U_n$. Then there exist points $u, v \in X$ such that $(x,u), (u,v), (v,y) \in U_n$. Therefore, by the triangle inequality, $d_{Mn}(x,y) < 3\theta$. This means that

$$d_M(T^j(x), T^j(y)) < 3\theta$$

for all $0 \leq j \leq M(n-1)$. Thus, $(T^j(x), T^j(y)) \in V_M(3\theta) \subseteq V_0(\theta)$ for all $0 \leq j \leq M(n-1)$, by (5.2). Hence, $(x,y) \in V_{M(n-1)}(\theta) = U_{n-1}$, and we have proved that $U_n \circ U_n \circ U_n \subseteq U_{n-1}$ for any $n \in \mathbb{N}$. To summarize, the family $(U_n)_{n=0}^{\infty}$ satisfies all three hypotheses of Frink's lemma.

Therefore, we can now apply Frink's lemma to obtain a metric ρ, compatible with the topology of X, such that

$$U_n \subseteq \{(x,y) \in X \times X : \rho(x,y) < 2^{-n}\} \subseteq U_{n-1}$$

for all $n \in \mathbb{N}$.

We now show that $T^{3M} : X \to X$ is distance expanding with respect to the metric ρ. To this end, choose $x \neq y$ such that $\rho(x,y) < 2^{-4}$. Then there exists a unique $n \geq 0$ such that $(x,y) \in U_n \backslash U_{n+1}$, because $(U_n)_{n=0}^{\infty}$ is a descending sequence of sets such that $\bigcap_{n=0}^{\infty} U_n = \triangle$ and because $U_0 = X \times X$. By Frink's lemma, since $\rho(x,y) < 2^{-4}$, we have that $(x,y) \in U_3$. So $n \geq 3$.

On one hand, since (x,y) belongs to U_n, we have that

$$\rho(x,y) < 2^{-n}.$$

On the other hand, given that $(x,y) \in U_n \backslash U_{n+1} := V_{Mn}(\theta) \backslash V_{M(n+1)}(\theta)$, there exists $Mn \leq j < M(n+1)$ such that $d(T^j(x), T^j(y)) \geq \theta$. Write j in the form $j = i + 3M$. Then we have that $0 \leq M(n-3) \leq i < M(n-2)$ and

$$d(T^i(T^{3M}(x)), T^i(T^{3M}(y))) \geq \theta.$$

From this, we obtain that

$$\left(T^{3M}(x), T^{3M}(y)\right) \notin V_{M(n-2)} = U_{n-2} \supseteq \left\{(x,y) \in X \times X : \rho(x,y) < 2^{-(n-1)}\right\}.$$

Then it follows that whenever $\rho(x,y) < 2^{-4}$,

$$\rho\left(T^{3M}(x), T^{3M}(y)\right) \geq 2^{-(n-1)} = 2 \cdot 2^{-n} \geq 2\rho(x,y).$$

Hence, T^{3M} is expanding with respect to Frink's metric. It follows from Exercise 4.6.2 that the map T is expanding with respect to a metric topologically equivalent to Frink's metric, which is in turn topologically equivalent to the original metric d. □

We now draw some important conclusions from the fact that an expansive dynamical system is expanding with respect to a topologically equivalent metric. Namely, we can infer the existence of Markov partitions and of a symbolic representation.

Corollary 5.3.2. *Every open, expansive dynamical system $T : X \to X$ admits Markov partitions of arbitrarily small diameters.*

Proof. This follows immediately from Theorems 5.3.1 and 4.4.6. □

Corollary 5.3.3. *Every open, expansive dynamical system $T : X \to X$ admits a symbolic representation. More precisely, every open, expansive system is a factor of a subshift of finite type $\sigma : E_A^{\infty} \to E_A^{\infty}$ with a coding map $\pi : E_A^{\infty} \to X$ such that every point in a forward T-invariant dense G_δ-subset of X admits a unique symbolic representation.*

Proof. This follows from Theorems 5.3.1 and 4.5.2. □

Yet another repercussion of the fact that an expansive dynamical system is expanding with respect to a topologically equivalent metric is the following.

Corollary 5.3.4. *Every open, expansive, and transitive dynamical system $T : X \to X$ is very strongly transitive.*

Proof. This is a direct consequence of Theorems 5.3.1 and 4.2.10. □

5.4 Parabolic Cantor sets

In this section, we describe a family of C^1 maps defined on topological Cantor subspaces of $\mathbb{R}$, that are expansive but not distance expanding with respect to the standard Euclidean metric on $\mathbb{R}$. Of course, bearing in mind what we have just proved in the previous section, for each of these maps there exists a metric, compatible with the Euclidean topology on $\mathbb{R}$, with respect to which the map is distance expanding.

Let $I := [0,1]$. Let E be a finite set, say $E = \{0,1,\ldots,k-1\}$, and let $\varphi_e : I \to I$, $e \in E$, be C^1 maps with the following properties:

(I1) $\varphi_0'(0) = 1$.

(I2) $0 < \varphi_0'(x) < 1$ for all $x \in I \setminus \{0\}$.

(I3) $0 < |\varphi_e'(x)| < 1$ for all $x \in I$ and $e \in E \setminus \{0\}$.

(II) $\varphi_0(0) = 0$.

(III) $\varphi_e(I) \cap \varphi_f(I) = \emptyset$ for all $e \neq f$.

This setting is reminiscent of hyperbolic Cantor sets in Subsection 4.1.2, with $X_0 = I$. Indeed, conditions (I1)–(I3) concern the derivative of the generators and are the counterparts for parabolic Cantor sets of condition (i) for hyperbolic Cantor sets. In particular, they imply that the generators are invertible functions on I, so each generator is either strictly increasing or strictly decreasing. Condition (III) is identical to condition (iii). Condition (ii) is automatically fulfilled.

Conditions (I1) and (II) are the reason for calling the limit set X a *parabolic Cantor set*, as opposed to an hyperbolic one. These conditions ensure that the derivative of one of the generators is equal to 1 at a fixed point.

Under these conditions, the limit set X is constructed through the same procedure as for hyperbolic Cantor sets (see (4.8) and (4.6)). The neighborhood U can be constructed in precisely the same way as for hyperbolic Cantor sets (see (4.12)), provided that we first extend and replace the generators with C^1 diffeomorphisms of $\mathbb{R}$ as follows:

$$\overline{\varphi}_0(x) = \begin{cases} -\varphi_0(1) + \varphi_0'(1)\cdot(x+1) & \text{if } x \le -1 \\ -\varphi_0(-x) & \text{if } x \in -I \\ \varphi_0(x) & \text{if } x \in I \\ \varphi_0(1) + \varphi_0'(1)\cdot(x-1) & \text{if } x \ge 1, \end{cases}$$

whereas, for all $e \in E \setminus \{0\}$,

$$\overline{\varphi}_e(x) = \begin{cases} \varphi_e(0) + \varphi_e'(0)\cdot x & \text{if } x \le 0 \\ \varphi_e(x) & \text{if } x \in I \\ \varphi_e(1) + \varphi_e'(1)\cdot(x-1) & \text{if } x \ge 1. \end{cases}$$

The map $T : U \to \mathbb{R}$ is then defined in exactly the same manner as for hyperbolic Cantor sets (see (4.13)).

All definitions, relations, and proofs for hyperbolic Cantor sets which do not involve the constant M hold for parabolic Cantor sets. Observe that $(\varphi_e^{-1})'(\varphi_e(x)) = 1/\varphi_e'(x)$ for all $x \in I$ and $e \in E$. Condition (I1) implies that $T'(0) = 1$. The triple (X, U, T) is not an expanding repeller for parabolic Cantor sets since it does not satisfy condition (b) of the definition of a repeller (cf. Definition 4.1.4 and (4.16)). Condition (II) implies that the point 0 lies in the limit set X and prevents any iterate of T of being an expanding repeller. However, conditions (a) and (c) of a repeller are satisfied.

Now, define

$$\lambda := \min_{e\in E\setminus\{0\}} \min_{x\in\varphi_e(I)} |T'(x)| = \frac{1}{M} > 1, \quad \text{where} \quad M := \max_{e\in E\setminus\{0\}} \max_{x\in I} |\varphi_e'(x)|.$$

Theorem 5.4.1. *The map $T : X \to X$ is expansive.*

Proof. Consider $x \in \varphi_0(I) \setminus \{0\}$. By the mean value theorem, there exists $y \in \operatorname{Int}(\varphi_0(I))$ such that

$$\frac{T(x)}{x} = \frac{T(x) - T(0)}{x - 0} = T'(y) > 1. \tag{5.3}$$

Therefore, $T(x) > x$ for all $x \in \varphi_0(I) \setminus \{0\}$ and the map $T|_{\varphi_0(I)} : \varphi_0(I) \to I$ has no fixed point other than 0. Let $\overline{\delta} > 0$ be the length of the smallest gap between all $\varphi_e(I)$'s. We claim that any $0 < \delta < \overline{\delta}$ is an expansive constant for $T : X \to X$. So suppose that there exist points $z, w \in X$ such that

$$|T^n(z) - T^n(w)| \le \delta, \quad \forall n \ge 0.$$

Let $\triangle$ be the closed interval joining z and w. Then for every $n \ge 0$, there exists a unique $e \in E$ such that $T^n(\triangle) \subseteq \varphi_e(I)$. It follows from the fact that $T'(x) > 1$ for all $x \in U \setminus \{0\}$ and from the definition of λ that for all $n \ge 0$,

$$|T^{n+1}(\triangle)| > |T^n(\triangle)| \quad \text{whenever } T^n(\triangle) \subseteq \varphi_0(I) \tag{5.4}$$

and

$$|T^{n+1}(\triangle)| \ge \lambda |T^n(\triangle)| \quad \text{whenever } T^n(\triangle) \subseteq \varphi_e(I) \text{ for some } e \in E \setminus \{0\}. \tag{5.5}$$

Since $\lambda > 1$ and $|T^n(\triangle)| \le 1$ for all $n \ge 0$, relation (5.5) can only hold for finitely many n. That is, there exists $N \ge 0$ such that for all $n \ge 0$,

$$T^{n+N}(\triangle) = T^n\big(T^N(\triangle)\big) \subseteq \varphi_0(I).$$

Fix $x \in T^N(\triangle)$. If $x \ne 0$, then $T^n(x) \in \varphi_0(I) \setminus \{0\}$ for all $n \ge 0$. It follows from (5.3) that the sequence of iterates $(T^n(x))_{n=0}^{\infty}$ is (strictly) increasing. Thus, it has a limit point which, according to Lemma 1.1.4, is a fixed point for T. This contradicts the fact that $T|_{\varphi_0(I)} : \varphi_0(I) \to I$ has no fixed point but 0. Therefore, $T^N(\triangle) = \{0\}$. Hence, $\triangle = \{0\}$ and $z = w (= 0)$. □

Since T is expansive on the limit set X, according to Theorem 5.3.1 there exists a metric, compatible with the topology of X, with respect to which T is distance expanding on the limit set X. However, that metric is not the usual Euclidean metric.

Proposition 5.4.2. *The map $T : X \to X$ is not expanding with respect to the Euclidean metric.*

Proof. Let $\varepsilon > 0$. Since T is a C^1 map and $T'(0) = 1$, there exists $\eta > 0$ such that $T'(x) < 1+\varepsilon$ for all $x \in [0,\eta)$. Fix $y \in X \cap (0,\eta)$. It follows from the mean value theorem that for some $z \in (0,y)$,

$$\frac{|T(y) - T(0)|}{|y - 0|} = T'(z) < 1 + \varepsilon.$$

Therefore, T is not expanding with respect to the metric $d(x,y) = |x - y|$. □

5.5 Exercises

Exercise 5.5.1. Prove that the expansiveness of $T : X \to X$ is independent of the metric on X (though expansive constants generally depend on the metric chosen). That is, show that, given two metrics d and d', which generate the topology of the compact metrizable space X, the map T is expansive when X is equipped with the metric d if and only if T is expansive when X is endowed with the metric d'.

Exercise 5.5.2. A dynamical system $T : X \to X$ on a topological space X is said to be expansive if there exists a base $\mathcal{B}$ for the topology such that for every $x \neq y$ there is $n = n(x,y) \geq 0$ with

$$\{T^n(x), T^n(y)\} \not\subseteq U, \quad \forall U \in \mathcal{B}.$$

Note that if X is second-countable, then expansiveness is equivalent to the existence of a countable base with the above property. Show that this definition is equivalent to Definition 5.1.1 when X is a compact metrizable space.

Exercise 5.5.3. Prove that expansiveness is a topological conjugacy invariant.

Exercise 5.5.4. Prove that the metrics d_n, $n \in \mathbb{N}$, given in Definition 5.1.3 induce the same topology.

Exercise 5.5.5. Let $T : (X,d) \to (X,d)$ be a dynamical system. For every $n \geq 0$, let $d_\infty : X \times X \to [0,\infty)$ be the function

$$d_\infty(x,y) := \sup_{0 \leq j < \infty} d(T^j(x), T^j(y)) = \sup_{n \in \mathbb{N}} d_n(x,y).$$

Show that d_∞ defines a metric on X. Prove that if T is expansive on X then d_∞ generates the discrete topology on X. In particular, if X has infinite cardinality and T is expansive, then d_∞ is not topologically equivalent to any d_n, $n \in \mathbb{N}$.

Exercise 5.5.6. Let $T : X \to X$ be a topological dynamical system. Prove that the following conditions are equivalent:
(a) T is expansive.
(b) T^n is expansive for some $n \in \mathbb{N}$.
(c) T^n is expansive for all $n \in \mathbb{N}$.

Exercise 5.5.7. Show that the Cartesian product of finitely many expansive dynamical systems is an expansive system.

Exercise 5.5.8. Find two expansive maps on the same compact metric space whose composition is not expansive.

Exercise 5.5.9. Recall that the unit circle $\mathbb{S}^1$ is homeomorphic to the closed interval $[0,1]$ when 0 and 1 are identified. Define the map $T : \mathbb{S}^1 \to \mathbb{S}^1$ by the following formula:

$$T(x) = \begin{cases} x + 2x^2 & \text{if } 0 \le x \le 1/2 \\ 2x - 1 & \text{if } 1/2 \le x \le 1. \end{cases}$$

Show that T is not distance expanding with respect to the linear Euclidean metric on $\mathbb{S}^1$, but that T is an expansive map. Recall that the linear Euclidean metric on $\mathbb{S}^1$ is given by

$$d(x,y) = \min\{|x-y|, |1+x-y|, |1+y-x|\}.$$

Exercise 5.5.10. Suppose that for all $n \in \mathbb{N}$, the map $T_n : X_n \to X_n$ is a continuous map of a compact metric space X_n. Let X be the one-point (Alexandroff) compactification of the disjoint union of all spaces X_n, $n \in \mathbb{N}$. Denote the added point by ω. Define the map $T : X \to X$ by the formula

$$T(x) = \begin{cases} T_n(x) & \text{if } x \in X_n \\ \omega & \text{if } x = \omega. \end{cases}$$

Show that T is never expansive.

6 Shub expanding endomorphisms

In Section 6.2 of this chapter, we give a systematic account of Shub expanding endomorphisms. These maps are far-reaching generalizations of the expanding endomorphisms of the circle which we first introduce in Section 6.1. They constitute a large subclass of distance expanding maps. Their origins lie in the seminal papers of Epstein and Shub [22], Shub [66], and Krzyżewski and Szlenk [42]. Our exposition stems from the chapter on expanding endomorphisms in Szlenk's book [71].

Basic knowledge of algebraic geometry/topology is assumed. The first chapter of the book by Hatcher [28] is an engaging source for the reader unfamiliar with notions such as lift, deck transformation, homotopy, and the fundamental group of a topological space, notions which will be used throughout this chapter, especially in our digression into algebraic topology in Section 6.3.

Finally, in Section 6.4 we establish that Shub's expanding endomorphisms are structurally stable, form an open set in an appropriate topology of smooth maps, are topologically exact (and hence transitive), have at least one fixed point as well as a dense set of periodic points, and their universal covering space is diffeomorphic to $\mathbb{R}^n$.

6.1 Shub expanding endomorphisms of the circle

In this section, we study a special class of maps of the unit circle, the Shub expanding endomorphisms of $\mathbb{S}^1$.

Let $\gamma : \mathbb{S}^1 \to (0,\infty)$ be a C^1 function on $\mathbb{S}^1$ (recall that this means that the first derivative of γ exists and is continuous). The function γ induces the Riemannian metric $\rho_\gamma = \gamma|dx|$ on $\mathbb{S}^1$. If $\triangle$ is an arc of $\mathbb{S}^1$ and $\varphi : \triangle \to \mathbb{S}^1$ is a C^1 curve on $\mathbb{S}^1$, then the length $\ell_\gamma(\varphi)$ of φ is defined to be

$$\ell_\gamma(\varphi) := \int_\triangle |\varphi'(x)|\, \gamma(x)\, dx. \tag{6.1}$$

The Riemannian metric ρ_γ induces a distance (which, in somewhat of an abuse of notation, we will also denote by ρ_γ) on $\mathbb{S}^1$ as follows. Let $a, b \in \mathbb{S}^1$, and let $\triangle_1$ and $\triangle_2$ be the two arcs of $\mathbb{S}^1$ joining a and b. Let $\mathrm{Id}_{\triangle_i} : \triangle_i \to \triangle_i$, $i = 1, 2$, be the identity curves on these respective arcs. We define

$$\rho_\gamma(a,b) := \min_{i=1,2} \ell_\gamma(\mathrm{Id}_{\triangle_i}) = \min_{i=1,2} \int_{\triangle_i} \gamma(x)\, dx.$$

If $v \in T_x\mathbb{S}^1$, that is, if v is a vector in $\mathbb{R}^2$ tangent to $\mathbb{S}^1$ at the point $x \in \mathbb{S}^1$, then the norm of v relative to the metric ρ_γ is given by

$$\|v\|_\gamma := \gamma(x)\|v\|,$$

https://doi.org/10.1515/9783110702682-006

where $\|\cdot\|$ is the standard Euclidean norm in $\mathbb{R}^2$. We can thus rewrite (6.1) as

$$\ell_\gamma(\varphi) := \int_{\triangle} \|\varphi'(x)\|_\gamma \, dx.$$

If $T : \mathbb{S}^1 \to \mathbb{S}^1$ is a differentiable map, then for every $x \in \mathbb{S}^1$ the derivative map $T'(x)$ sends $T_x\mathbb{S}^1$ into $T_{T(x)}\mathbb{S}^1$ and

$$\|T'(x)(v)\| = \|T'(x)\| \cdot \|v\|, \quad \forall v \in T_x\mathbb{S}^1.$$

Hence,

$$\begin{aligned}\|T'(x)(v)\|_\gamma &= \gamma(T(x))\|T'(x)(v)\| = \gamma(T(x))\|T'(x)\| \cdot \|v\| \\ &= \frac{\gamma(T(x))}{\gamma(x)}\|T'(x)\| \cdot \|v\|_\gamma.\end{aligned}$$

We naturally set

$$\|T'(x)\|_\gamma := \frac{\|T'(x)(v)\|_\gamma}{\|v\|_\gamma} = \frac{\gamma(T(x))}{\gamma(x)}\|T'(x)\| \tag{6.2}$$

and call this quantity the *norm of* $T'(x)$ *with respect to the metric* ρ_γ. We call a C^1 endomorphism $T : \mathbb{S}^1 \to \mathbb{S}^1$ *Shub expanding* with respect to the metric ρ_γ if there exists some $\lambda > 1$ such that

$$\|T'(x)(v)\|_\gamma \geq \lambda\|v\|_\gamma, \quad \forall v \in T_x\mathbb{S}^1, \ \forall x \in \mathbb{S}^1.$$

Equivalently, $T : \mathbb{S}^1 \to \mathbb{S}^1$ is Shub expanding if

$$\|T'(x)\|_\gamma \geq \lambda, \quad \forall x \in \mathbb{S}^1. \tag{6.3}$$

Example 6.1.1. For every integer $|k| \geq 2$, the map $E_k : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $E_k(x) := kx \pmod 1$ is Shub expanding with respect to the standard Riemannian metric (that is, when $\gamma \equiv 1$). Indeed, $\|E_k'(x)\| = |k| \geq 2$ for every $x \in \mathbb{S}^1$.

We now show that each Shub expanding map of the circle is, up to a C^1 conjugacy, Shub expanding with respect to the standard Euclidean metric on $\mathbb{S}^1$. Indeed, in light of (6.2), we may multiply γ by a constant factor without changing $\|T'(x)\|_\gamma$ and in such a way that

$$\int_{\mathbb{S}^1} \gamma(x)\, dx = 1.$$

We say that such a Riemannian metric is normalized. Then the map $H : \mathbb{S}^1 \to \mathbb{S}^1$ given by the formula

$$H(x) = \int_0^x \gamma(t)\, dt$$

defines a C^1 diffeomorphism of $\mathbb{S}^1$ such that

$$H'(x) = \gamma(x), \quad \forall x \in \mathbb{S}^1. \tag{6.4}$$

We define the C^1 endomorphism

$$\overline{T} := H \circ T \circ H^{-1} : \mathbb{S}^1 \to \mathbb{S}^1. \tag{6.5}$$

We then obtain the following important result.

Theorem 6.1.2. *Each Shub expanding map $T : \mathbb{S}^1 \to \mathbb{S}^1$ of the unit circle is C^1 conjugate to the map $\overline{T} : \mathbb{S}^1 \to \mathbb{S}^1$, which is Shub expanding with respect to the standard Euclidean metric on $\mathbb{S}^1$.*

Proof. Suppose that $T : \mathbb{S}^1 \to \mathbb{S}^1$ is a Shub expanding map. As argued above, we may assume without loss of generality that the corresponding function $\gamma : \mathbb{S}^1 \to (0, \infty)$ is normalized. Let $\overline{T} : \mathbb{S}^1 \to \mathbb{S}^1$ be given by (6.5). Using the chain rule, (6.4) and (6.3), we obtain that

$$\begin{aligned} \|\overline{T}\,'(x)\| &= H'\big(T(H^{-1}(x))\big) \cdot \big\|T'(H^{-1}(x))\big\| \cdot (H^{-1})'(x) \\ &= \gamma\big(T(H^{-1}(x))\big)\big\|T'(H^{-1}(x))\big\|\big(H'(H^{-1}(x))\big)^{-1} \\ &= \gamma\big(T(H^{-1}(x))\big)\big\|T'(H^{-1}(x))\big\|\big(\gamma(H^{-1}(x))\big)^{-1} \\ &= \big\|T'(H^{-1}(x))\big\|_\gamma \geq \lambda > 1. \end{aligned}$$

Thus, $\overline{T}$ is Shub expanding with respect to the standard Euclidean metric on $\mathbb{S}^1$. □

Our goal now is to prove a structure theorem for Shub expanding maps of the unit circle and to demonstrate the structural stability of the maps E_k, $|k| \geq 2$, from Example 6.1.1 (see Section 1.2 for more on structural stability). Before stating that theorem, let us add one more piece of notation. Given $k \in \mathbb{Z}$, let $\mathcal{E}^k(\mathbb{S}^1)$ be the space of all Shub expanding endomorphisms of $\mathbb{S}^1$ with degree equal to k. For a review of the notions of lift and degree of a circle map, see Section 2.1.

Theorem 6.1.3. *Every Shub expanding map $T \in \mathcal{E}^k(\mathbb{S}^1)$, where $|k| \geq 2$, is topologically conjugate to the map $E_k : \mathbb{S}^1 \to \mathbb{S}^1$. In addition, the map $E_k : \mathbb{S}^1 \to \mathbb{S}^1$ is strongly structurally stable when the space $\mathcal{E}^k(\mathbb{S}^1)$ is endowed with the topology of uniform convergence.*

Proof. In light of Theorem 6.1.2, we may assume without loss of generality that T is Shub expanding with respect to the standard Euclidean metric.

The following classical argument is essentially the proof of Theorem 2.4.6 in [33]. We give the proof for any positive k and mention afterwards the modifications necessary for a negative k. Consider the arcs

$$\triangle_n^m = \pi\left(\left[\frac{m}{k^n}, \frac{m+1}{k^n}\right]\right)$$

for all $n \in \mathbb{Z}_+$ and $0 \le m < k^n$, where $\mathbb{R} \ni x \mapsto \pi(x) = x(\text{mod } 1) \in \mathbb{S}^1$. For each $n \in \mathbb{Z}_+$, the family

$$\xi_n = \{\triangle_n^0, \dots, \triangle_n^{k^n-1}\}$$

is the "partition" of $\mathbb{S}^1$ into the k^n arcs whose endpoints are consecutive rational numbers with denominator k^n. These arcs are such that

$$E_k(\triangle_n^m) = \triangle_{n-1}^{m'}, \tag{6.6}$$

where m' is the unique integer between 0 and $k^{n-1} - 1$ such that $m' = m \ (\text{mod } k^{n-1})$.

We now construct a nested sequence of "partitions"

$$\zeta_n = \{\pi(\Gamma_n^0), \dots, \pi(\Gamma_n^{k^n-1})\}$$

of $\mathbb{S}^1$ into arcs which will be in a natural, order-preserving correspondence with the standard sequence ξ_n. Let p be a fixed point for a lift $\tilde{T}$ of T as in Lemma 2.1.14. If T is close to E_k, pick p close to 0. Since $\tilde{T}(p) = p$ and T is of degree k, we know that $\tilde{T}(p+1) = p + k$. Moreover, since T is locally injective, its lift $\tilde{T}$ is a strictly monotone continuous function. Then there are unique real numbers

$$p = a_1^0 < a_1^1 < a_1^2 < \cdots < a_1^{k-1} < a_1^k = p + 1$$

such that $\tilde{T}(a_1^m) = p + m$ for each $0 \le m \le k$. Let $\Gamma_1^m = [a_1^m, a_1^{m+1}]$ for every $0 \le m < k$. Then

$$T(\pi(\Gamma_1^m)) = \pi \circ \tilde{T}([a_1^m, a_1^{m+1}]) = \pi([p+m, p+m+1]) = \mathbb{S}^1,$$

and T is injective on the arc $\pi(\Gamma_1^m)$ up to identification of the endpoints of Γ_1^m. If T is close to E_k, then clearly each number a_1^m is close to m/k.

Furthermore, since $\tilde{T}(a_1^m) = p + m$ and $\tilde{T}(a_1^{m+1}) = p + m + 1$, and since $\tilde{T}$ is a strictly monotone continuous function, there are unique real numbers

$$a_1^m = a_2^{km} < a_2^{km+1} < \cdots < a_2^{km+k-1} < a_2^{k(m+1)} = a_1^{m+1}$$

such that $\tilde{T}(a_2^{km+i}) = a_1^i \ (\text{mod } 1)$ for $0 \le i \le k$. Again, a_2^{km+i} is close to $(km+i)/k^2$ if T is close to E_k. Let $\Gamma_2^m = [a_2^m, a_2^{m+1}]$ for $0 \le m < k^2$, so that $T(\pi(\Gamma_2^m)) = \pi \circ \tilde{T}([a_2^m, a_2^{m+1}]) = \pi([a_1^{m'}, a_1^{m'+1}]) = \pi(\Gamma_1^{m'})$, where m' is the unique integer between 0 and $k-1$ such that $m' = m \ (\text{mod } k)$.

We continue inductively and for each $n \in \mathbb{N}$ we define points a_n^{km+i} for $0 \le m < k^{n-1}$ and $0 \le i \le k$ such that

$$a_{n-1}^m = a_n^{km} < a_n^{km+1} < \cdots < a_n^{km+k-1} < a_n^{k(m+1)} = a_{n-1}^{m+1}$$

and

$$\widetilde{T}(a_n^{km+i}) = a_{n-1}^{m'} \pmod 1, \tag{6.7}$$

where $0 \le m' < k^{n-1}$ and $m' = km + i \pmod{k^{n-1}}$. Let $\Gamma_n^m = [a_n^m, a_n^{m+1}]$ for $0 \le m < k^n$. Then $T(\pi(\Gamma_n^m)) = \pi(\Gamma_{n-1}^{m'})$, where $0 \le m' < k^{n-1}$ and $m = m' \pmod{k^{n-1}}$. By induction, $T^n(\pi(\Gamma_n^m)) = \mathbb{S}^1$ and T^n is injective on $\pi(\Gamma_n^m)$ up to identification of the endpoints of Γ_n^m.

So far, we have only used the facts that T is locally injective (and thus its lift $\widetilde{T}$ is strictly monotone) and that T has degree k. If T is Shub expanding, that is, if $\|T'(x)\| \ge \lambda > 1$ for all $x \in \mathbb{S}^1$, then the length of each arc $\pi(\Gamma_n^m)$ does not exceed λ^{-n}, so the set of points $\{\pi(a_n^m)\}_{n\in\mathbb{N},0\le m<k^n}$ is dense in $\mathbb{S}^1$ while $\{a_n^m\}_{n\in\mathbb{N},0\le m<k^n}$ is dense in the interval $[p, p+1]$. This is the only place in the proof where the fact that T is an expanding map is used. (In fact, the use of differentiability could be easily avoided.)

Furthermore, for any $N \in \mathbb{N}$ and $\varepsilon > 0$ one can find $\delta > 0$ such that if T is δ-close to E_k in the uniform topology, then

$$\left|a_n^m - \frac{m}{k^n}\right| < \frac{\varepsilon}{3}, \quad \forall 1 \le n \le N,\ \forall 0 \le m < k^n. \tag{6.8}$$

We define a correspondence h between the set $\{a_n^m\}_{n\in\mathbb{N},0\le m<k^n}$ and all k-ary rationals, that is, the rational numbers whose denominators are powers of k, by setting

$$h(a_n^m) = \frac{m}{k^n}.$$

This correspondence is monotone and since the set $\{a_n^m\}_{n\in\mathbb{N},0\le m<k^n}$ is dense in the interval $[p, p+1]$, it can be uniquely extended to a homeomorphism $h : [p, p+1] \to [0, 1]$. Since $h(\Gamma_n^m) = \triangle_n^m$ for all $n \in \mathbb{N}$ and $0 \le m < k^n$, relations (6.6) and (6.7) imply that T is topologically conjugate to E_k via h, that is,

$$E_k \circ h = h \circ T. \tag{6.9}$$

Assuming under the conditions of (6.8) that N and ε are chosen such that $1/k^N < \varepsilon/3$, one sees in addition that $|a_n^m - h(a_n^m)| < \varepsilon$ for $n \in \mathbb{N}$ and $0 \le m < k^n$, and hence $|h(x) - x| < \varepsilon$ for all x, that is, $h \in B_{\rho_\infty}(\mathrm{Id}_{\mathbb{S}^1}, \varepsilon)$. Recall also that if T is close enough to E_k in the topology of uniform convergence, then the degree of T is k, by Lemma 2.1.12. Thus, E_k is strongly structurally stable in the space $\mathcal{E}^k(\mathbb{S}^1)$.

The case of a negative k differs primarily in notation. The order of the real numbers a_n^{km+i} between a_{n-1}^m and a_{n-1}^{m+1} will be increasing for even n's and decreasing for odd n's, the same as the corresponding structure imposed by the map E_k on the k-ary rationals. □

Let $C^1(\mathbb{S}^1, \mathbb{S}^1)$ denote the space of all C^1 endomorphisms of $\mathbb{S}^1$ endowed with the C^1 topology.

Corollary 6.1.4. *Every Shub expanding map of the unit circle is structurally stable in* $C^1(\mathbb{S}^1, \mathbb{S}^1)$ *and every map* $E_k : \mathbb{S}^1 \to \mathbb{S}^1$, $|k| \geq 2$, *is strongly structurally stable in this class.*

Proof. This is an immediate consequence of Theorem 6.1.3 once one observes that each element of $\mathcal{E}^k(\mathbb{S}^1)$ has a neighborhood U in $C^1(\mathbb{S}^1, \mathbb{S}^1)$ such that $U \subseteq \mathcal{E}^k(\mathbb{S}^1)$. For more information, see Theorem 6.2.5. □

6.2 Definition, characterization, and properties of general Shub expanding endomorphisms

Unless stated otherwise, we shall let M be a compact connected smooth (i. e., C^∞) manifold, and ρ a Riemannian metric on M.

If $\gamma : I \to M$ is a smooth curve defined on an interval $I \subseteq \mathbb{R}$, then the length of γ with respect to the Riemannian metric ρ is defined to be

$$\ell_\rho(\gamma) := \int_I \|D_t\gamma(e_t)\|\, dt = \int_I \|D_t\gamma\|\, dt,$$

where e_t is the unit tangent vector to I at the point t. Given $x, y \in M$, let $\Gamma(x, y)$ be the collection of all smooth curves on M whose endpoints are x and y. The distance $\rho(x, y)$ between x and y is defined as

$$\rho(x, y) := \inf\{\ell_\rho(\gamma) \mid \gamma \in \Gamma(x, y)\},$$

where, as above, we shall use the same symbol ρ to denote the original Riemannian metric and the distance it induces on M. A curve γ joining x to y whose length is equal to $\rho(x, y)$ is called a *geodesic* from x to y. Although we will not rely on this fact, a geodesic joining x and y always exists. In fact, it is unique if the points x and y are sufficiently close.

Let us begin by defining the class of transformations of M that we will study.

Definition 6.2.1. A C^1 endomorphism $T : M \to M$ is *Shub expanding* if there exists $k \in \mathbb{N}$ such that

$$\|D_x T^k(v)\|_{T^k(x)} \geq 2\|v\|_x, \quad \forall x \in M,\ \forall v \in T_x M.$$

We immediately present a characterization of these maps. We invite the reader to envision the implications it will have on the dynamics of these maps.

Proposition 6.2.2. *If* $T : M \to M$ *is a* C^1 *endomorphism, then the following statements are equivalent:*

(a) *The map* $T : M \to M$ *is Shub expanding.*
(b) *There exist constants* $\mu > 1$ *and* $C > 0$ *such that for all* $n \in \mathbb{N}$,

$$\|D_x T^n(v)\|_{T^n(x)} \geq C\mu^n \|v\|_x, \quad \forall x \in M,\ \forall v \in T_x M.$$

(c) *There exist $\lambda > 1$ and a Riemannian metric ρ' on M such that*

$$\|D_x T(v)\|_{T(x),\rho'} \geq \lambda \|v\|_{x,\rho'}, \quad \forall x \in M, \ \forall v \in T_x M. \tag{6.10}$$

Proof. Let us first prove that (c)⇒(a). Since M is compact, the Riemannian norms ρ and ρ' are equivalent in the sense that there exists a constant $L \geq 1$ such that

$$L^{-1}\|v\|_{x,\rho} \leq \|v\|_{x,\rho'} \leq L\|v\|_{x,\rho}, \quad \forall x \in M, \ \forall v \in T_x M.$$

Using the chain rule repeatedly, for every $k \in \mathbb{N}$ we obtain that

$$\begin{aligned}\|D_x T^k(v)\|_{T^k(x),\rho} &\geq L^{-1}\|D_x T^k(v)\|_{T^k(x),\rho'} \\ &= L^{-1}\|D_{T^{k-1}(x)} T(D_x T^{k-1}(v))\|_{T(T^{k-1}(x)),\rho'} \\ &\geq L^{-1}\lambda \|D_x T^{k-1}(v)\|_{T^{k-1}(x),\rho'} \\ &\geq \ldots \geq L^{-1}\lambda^k \|v\|_{x,\rho'} \geq \lambda^k L^{-2}\|v\|_{x,\rho}.\end{aligned}$$

It suffices to take $k \in \mathbb{N}$ so large that $\lambda^k \geq 2L^2$.

We now prove that (a)⇒(b). To begin, it follows from the definition of a Shub expanding endomorphism that $\mathrm{Ker}(D_x T^k) = \{0\}$ for every $x \in M$. Hence, $\mathrm{Ker}(D_x T^n) = \{0\}$ for all $x \in M$ and all $n \in \mathbb{N}$ (first, use the chain rule in the form $D_x T^k = D_{T(x)} T^{k-1} \circ D_x T$ to establish the statement for $n = 1$ and then deduce it for any n). Since the tangent spaces $T_x M$ and $T_{T^n(x)} M$ are of finite dimension equal to $\dim(M)$, all the maps $D_x T^n : T_x M \to T_{T^n(x)} M$ are linear isomorphisms and thereby invertible. In particular, $\|(D_x T)^{-1}\| < \infty$ for each $x \in M$. Moreover, observe that the determinant function $x \mapsto \det(D_x T)$ is continuous on M since $T \in C^1(M, M)$, and does not vanish anywhere on M since $D_x T$ is invertible for every $x \in M$. As the entries of the inverse matrix A^{-1} of an invertible matrix A are polynomial functions of the entries of A divided by $\det(A)$, the entries of the matrix $(D_x T)^{-1}$ depend continuously on $x \in M$. Consequently, the function $x \mapsto \|(D_x T)^{-1}\|$ is continuous on M, and thus $\|(DT)^{-1}\|_\infty := \max_{x \in M} \|(D_x T)^{-1}\| < \infty$ since M is compact. Let

$$\alpha := \max\{1, \|(DT)^{-1}\|_\infty\} < \infty.$$

Fix an arbitrary $n \in \mathbb{N}$. Write $n = qk + r$, where q and r are integers such that $q \geq 0$ and $0 \leq r < k$. For every $x \in M$ and every $v \in T_x M$, we have

$$\begin{aligned}\|v\|_x &= \|(D_x T^n)^{-1}(D_x T^n(v))\|_x \leq \|(D_x T^n)^{-1}\| \cdot \|D_x T^n(v)\|_{T^n(x)} \\ &= \|(D_{T^{qk}(x)} T^r \circ D_x T^{qk})^{-1}\| \cdot \|D_x T^n(v)\|_{T^n(x)} \\ &\leq \|(D_{T^{qk}(x)} T^r)^{-1}\| \cdot \|(D_x T^{qk})^{-1}\| \cdot \|D_x T^n(v)\|_{T^n(x)} \\ &\leq \prod_{i=0}^{r-1} \|(D_{T^{qk+i}(x)} T)^{-1}\| \cdot \prod_{j=0}^{q-1} \|(D_{T^{jk}(x)} T^k)^{-1}\| \cdot \|D_x T^n(v)\|_{T^n(x)}\end{aligned}$$

$$\le \alpha^r 2^{-q} \|D_x T^n(v)\|_{T^n(x)}$$
$$\le 2\alpha^{k-1} 2^{-(q+1)} \|D_x T^n(v)\|_{T^n(x)}.$$

Given that $k(q+1) \ge n$, it follows from the above estimate that

$$\|D_x T^n(v)\|_{T^n(x)} \ge \frac{1}{2}\alpha^{1-k}(2^{1/k})^n \|v\|_x.$$

Thus, part (b) is proved with $C = \alpha^{1-k}/2 > 0$ and $\mu = 2^{1/k} > 1$.

Since the implication (b)⇒(a) is obvious, to complete the proof it suffices to show that (a)⇒(c). Define on M a new metric ρ' with scalar product on the tangent spaces given by

$$\langle v, w\rangle'_x := \sum_{j=0}^{k-1} \langle D_x T^j(v), D_x T^j(w)\rangle_{T^j(x)}.$$

Then

$$\|v\|^2_{x,\rho'} = \langle v, v\rangle'_x = \sum_{j=0}^{k-1} \|D_x T^j(v)\|^2_{T^j(x),\rho} \tag{6.11}$$

whereas

$$\|D_x T(v)\|^2_{T(x),\rho'} = \sum_{j=1}^{k} \|D_x T^j(v)\|^2_{T^j(x),\rho}. \tag{6.12}$$

For all $j = 0, 1, \ldots, k$, we have that

$$\|D_x T^j(v)\|_{T^j(x),\rho} \le \|DT\|^j_{\infty,\rho} \|v\|_{x,\rho} \le \max\{1, \|DT\|^k_{\infty,\rho}\}\|v\|_{x,\rho}.$$

Write $\beta := \max\{1, \|DT\|^k_{\infty,\rho}\}$. By (a), it then follows that

$$\|D_x T^k(v)\|^2_{T^k(x),\rho} \ge 4\|v\|^2_{x,\rho} = 2\|v\|^2_{x,\rho} + \frac{2}{k}\sum_{j=0}^{k-1} \|v\|^2_{x,\rho}$$
$$\ge 2\|v\|^2_{x,\rho} + \frac{2}{k}\sum_{j=0}^{k-1} \beta^{-2}\|D_x T^j(v)\|^2_{T^j(x),\rho}.$$

From this, (6.11) and (6.12), we deduce that

$$\|D_x T(v)\|^2_{T(x),\rho'} = \sum_{j=1}^{k-1} \|D_x T^j(v)\|^2_{T^j(x),\rho} + \|D_x T^k(v)\|^2_{T^k(x),\rho}$$
$$\ge 2\|v\|^2_{x,\rho} + (1 + 2\beta^{-2}k^{-1}) \sum_{j=1}^{k-1} \|D_x T^j(v)\|^2_{T^j(x),\rho}$$

$$\geq \left(1+\min\{1,2\beta^{-2}k^{-1}\}\right)\sum_{j=0}^{k-1}\|D_xT^j(v)\|^2_{T^j(x),\rho}$$
$$= \left(1+\min\{1,2\beta^{-2}k^{-1}\}\right)\|v\|^2_{x,\rho'}.$$

Taking $\lambda = (1+\min\{1,2\beta^{-2}k^{-1}\})^{1/2} > 1$ completes the proof that (a)⇒(c) and thereby completes the proof of the proposition. □

A Riemannian metric satisfying condition (6.10) is said to be *e-adapted* to T while a corresponding number λ is called an *expanding factor* for this metric.

Corollary 6.2.3. *Being a Shub expanding endomorphism of a compact connected smooth Riemannian manifold M is independent of the Riemannian metric that M is endowed with. More precisely, a map $T : M \to M$ which is a Shub expanding endomorphism with respect to some Riemannian metric on M is a Shub expanding endomorphism with respect to all Riemannian metrics on M.*

As observed in the proof of Proposition 6.2.2 (or as may be readily deduced from part (c) of that proposition), all the maps $D_xT : T_xM \to T_{T(x)}M$, $x \in M$, are linear isomorphisms. Therefore, by virtue of the inverse function theorem, the map T is a local diffeomorphism at every point of M. Since M is compact, T is a covering map. We have thus obtained the following important fact.

Theorem 6.2.4. *Every Shub expanding endomorphism is a covering map.*

We now turn our attention to topological properties of sets of Shub expanding endomorphisms. For every $r \geq 1$, we denote by $\mathcal{E}^r(M)$ the set of all C^r Shub expanding endomorphisms of the manifold M. This set has the remarkable property of being open.

Theorem 6.2.5. *For each $r \geq 1$, the set $\mathcal{E}^r(M)$ is an open subset of the space $C^r(M,M)$ endowed with the C^1 topology.*

Proof. Let

$$\mathcal{I}^r(M) := \{T \in C^r(M,M) \mid D_xT \text{ is invertible}, \forall x \in M\}$$
$$= \{T \in C^r(M,M) \mid \det(D_xT) \neq 0, \forall x \in M\}.$$

Since the determinant function $x \mapsto \det(D_xT)$ is continuous on M for any $T \in C^r(M,M)$ and since M is compact, we deduce that

$$\mathcal{I}^r(M) = \left\{T \in C^r(M,M) : \min_{x\in M}|\det(D_xT)| \neq 0\right\}.$$

Moreover, since the determinant function $(x,T) \mapsto \det(D_xT)$ is continuous on $M \times C^r(M,M)$, the function $T \mapsto \min_{x\in M}|\det(D_xT)|$ is continuous on $C^r(M,M)$. This guarantees that $\mathcal{I}^r(M)$ is an open subset of $C^r(M,M)$. Notice also that the continuous function $(x,T) \mapsto \det(D_xT)$ does not vanish on $M \times \mathcal{I}^r(M)$. Since the entries of the inverse matrix A^{-1} of an invertible matrix A are polynomials of the entries of A divided by $\det(A)$, the entries of the matrix $(D_xT)^{-1}$ depend continuously on $(x,T) \in M \times \mathcal{I}^r(M)$. Consequently, the function $(x,T) \mapsto \|(D_xT)^{-1}\|$ is continuous on $M \times \mathcal{I}^r(M)$. Thus, the function $T \mapsto \max_{x\in M} \|(D_xT)^{-1}\|$ is continuous on $\mathcal{I}^r(M)$. As observed in (the proof of) Proposition 6.2.2, $\mathcal{E}^r(M) \subseteq \mathcal{I}^r(M)$. Fix $S \in \mathcal{E}^r(M)$. Let ρ be a metric e-adapted to S and λ an expanding factor for this metric. Formula (6.10) implies that $\|(D_xS)^{-1}\| \le \lambda^{-1}$ for all $x \in M$. The continuity of the function $T \mapsto \max_{x\in M} \|(D_xT)^{-1}\|$ on $\mathcal{I}^r(M)$ ensures the existence of a neighborhood U of S in $\mathcal{I}^r(M)$ such that

$$\max_{x\in M} \|(D_xT)^{-1}\| \le \frac{\lambda^{-1}+1}{2}$$

for all $T \in U$. But

$$\|v\|_x = \|(D_xT)^{-1}(D_xT(v))\|_x \le \|(D_xT)^{-1}\| \cdot \|D_xT(v)\|_{T(x)}$$

for every $x \in M$. Therefore,

$$\|D_xT(v)\|_{T(x)} \ge \|(D_xT)^{-1}\|^{-1}\|v\|_x \ge \frac{2}{\lambda^{-1}+1}\|v\|_x$$

for every $x \in M$. Since $2/(\lambda^{-1}+1) > 1$, we conclude that each $T \in U$ is Shub expanding. □

One fundamental fact about continuous maps on a compact connected smooth manifold is that they are homotopic if they are sufficiently close.

Theorem 6.2.6. *If M is a compact connected smooth manifold, then any two continuous maps sufficiently close in $C(M,M)$ are homotopic.*

Proof. Let ρ be a Riemannian metric on M and let $\exp : TM \to M$ be the corresponding exponential map. Let $\exp_x := \exp|_{T_xM}$ for each $x \in M$. Since M is compact, there exists a radius $\delta > 0$ such that for every $x \in M$, the inverse map $\exp_x^{-1} : B_\rho(x,\delta) \to T_xM$ is well-defined and so diffeomorphic. Take any two elements $f, g \in C(M,M)$ such that $\rho_\infty(f,g) < \delta$. Define a map $F : M \times [0,1] \to M$ as follows:

$$F(x,t) := \exp_{f(x)}\big(t \exp_{f(x)}^{-1}(g(x))\big).$$

As a composition of continuous maps, the map F is continuous. Also, $F(x,0) = \exp_{f(x)}(0) = f(x)$ and $F(x,1) = \exp_{f(x)}(\exp_{f(x)}^{-1}(g(x))) = g(x)$. Thus, F is a homotopy from f to g. □

According to Theorems 6.2.5 and 6.2.6, in order to establish the structural stability of Shub expanding endomorphisms, it suffices to prove that any two homotopic Shub expanding endomorphisms are topologically conjugate. This feat will be achieved at the very end of this chapter. Theorem 6.2.6 also partly explains the involvement of algebraic topology, which we will now briefly investigate.

6.3 A digression into algebraic topology

In this section, we develop some algebraically topological results that will be relied upon in the rest of the chapter.

6.3.1 Deck transformations

Let M be a compact connected smooth manifold with Riemannian metric ρ. Let $\widetilde{M}$ be the universal covering space of M and $\pi : \widetilde{M} \to M$ be the canonical projection from $\widetilde{M}$ to M. Every continuous map $\overline{G} : \widetilde{M} \to \widetilde{M}$ such that the diagram

$$\begin{tikzcd} & \widetilde{M} \arrow[d, "\pi"] \\ \widetilde{M} \arrow[ur, "\overline{G}"] \arrow[r, "\pi"'] & M \end{tikzcd}$$

commutes, that is, such that

$$\pi \circ \overline{G} = \pi,$$

is called a *deck transformation* of the manifold M. We adopt the convention of denoting deck transformations with an overline. By the unique lifting property (cf. Proposition 1.34 in [28]), a deck transformation is uniquely determined by its value at any point of $\widetilde{M}$.

Moreover, given any two points $\tilde{x}, \tilde{y} \in \widetilde{M}$ such that $\pi(\tilde{x}) = \pi(\tilde{y})$, by the unique lifting property there exist unique deck transformations $\overline{G}_{\tilde{x},\tilde{y}} : \widetilde{M} \to \widetilde{M}$ and $\overline{G}_{\tilde{y},\tilde{x}} : \widetilde{M} \to \widetilde{M}$ such that

$$\overline{G}_{\tilde{x},\tilde{y}}(\tilde{x}) = \tilde{y} \quad \text{and} \quad \overline{G}_{\tilde{y},\tilde{x}}(\tilde{y}) = \tilde{x}.$$

Consequently, $\overline{G}_{\tilde{y},\tilde{x}} \circ \overline{G}_{\tilde{x},\tilde{y}}$ is a deck transformation such that $\overline{G}_{\tilde{y},\tilde{x}} \circ \overline{G}_{\tilde{x},\tilde{y}}(\tilde{x}) = \tilde{x}$. It follows from the unique lifting property that $\overline{G}_{\tilde{y},\tilde{x}} \circ \overline{G}_{\tilde{x},\tilde{y}} = \mathrm{Id}_{\widetilde{M}}$, and, by the same token, $\overline{G}_{\tilde{x},\tilde{y}} \circ \overline{G}_{\tilde{y},\tilde{x}} = \mathrm{Id}_{\widetilde{M}}$. Therefore, $\overline{G}_{\tilde{x},\tilde{y}}$ is a diffeomorphism of $\widetilde{M}$.

Furthermore, since $\pi : \widetilde{M} \to M$ is a local diffeomorphism, it induces a Riemannian metric $\tilde{\rho}$ on $\widetilde{M}$ defined as follows:

$$\langle w, v \rangle_{\tilde{x},\tilde{\rho}} := \langle D_{\tilde{x}}\pi(w), D_{\tilde{x}}\pi(v) \rangle_{\pi(\tilde{x}),\rho}, \quad \forall \tilde{x} \in \widetilde{M},\ \forall w, v \in T_{\tilde{x}}\widetilde{M}.$$

With this Riemannian metric on $\widetilde{M}$, the projection map $\pi : \widetilde{M} \to M$ is an infinitesimal and local isometry and all deck transformations of M are infinitesimal and global isometries with respect to the metric $\tilde{\rho}$.

Proposition 6.3.1. *If M is a compact connected smooth manifold with Riemannian metric ρ, then the set D_M of all deck transformations of M is a group of diffeomorphisms (with composition as group action) acting transitively on each fiber of π. Each element of D_M is uniquely determined by its value at any point of $\widetilde{M}$, is an infinitesimal and global $\widetilde{\rho}$-isometry, where $\widetilde{\rho}$ is the metric induced by π and ρ.*

Proof. The only part that remains to be proved is the transitivity. For this, let $\overline{G}, \overline{H} \in D_M$. Then

$$\pi \circ (\overline{G} \circ \overline{H}) = (\pi \circ \overline{G}) \circ \overline{H} = \pi \circ \overline{H} = \pi.$$

That is, the group D_M acts transitively on each fiber of π. □

Later on, we will also need the following result.

Proposition 6.3.2. *If X is a connected Hausdorff topological space and $f, g : X \to \widetilde{M}$ are two continuous maps such that $\pi \circ f = \pi \circ g$, then there exists a unique deck transformation $\overline{G} \in D_M$ such that $\overline{G} \circ f = g$.*

Proof. Fix $x_0 \in M$ and let $\overline{G}$ be the unique element of D_M such that $\overline{G} \circ f(x_0) = g(x_0)$. Let

$$E = \{x \in X \mid \overline{G} \circ f(x) = g(x)\}.$$

Obviously, E is nonempty. It is also closed since its complement is open in X. We shall prove that E is also open. Indeed, let $x \in E$. Since the projection $\pi : \widetilde{M} \to M$ is a local homeomorphism, there exists an open neighborhood $\widetilde{V}$ of $g(x)$ in $\widetilde{M}$ such that the map $\pi|_{\widetilde{V}}$ is one-to-one. As the maps $\overline{G} \circ f$ and g are continuous, there is an open neighborhood U of x in X such that

$$\overline{G} \circ f(U) \subseteq \widetilde{V} \quad \text{and} \quad g(U) \subseteq \widetilde{V}.$$

Let $y \in U$. Then $\overline{G} \circ f(y)$ and $g(y)$ belong to $\widetilde{V}$. Moreover, $\pi(\overline{G} \circ f(y)) = \pi(f(y)) = \pi(g(y))$. Thus $\overline{G} \circ f(y) = g(y)$ by the injectivity of $\pi|_{\widetilde{V}}$. This shows that $y \in E$ and hence $U \subseteq E$, thereby proving that the nonempty, closed set E is also open. Since X is connected, we therefore conclude that $E = X$. The uniqueness of $\overline{G}$ follows immediately from Proposition 6.3.1 since $\overline{G}$ must be the unique deck transformation satisfying $\overline{G}(f(x_0)) = g(x_0)$. □

We now point out a fascinating characterization of the convergence of sequences of deck transformations at any point of the universal covering space.

Lemma 6.3.3. *Every sequence $(\overline{G}_n)_{n=1}^{\infty}$ in D_M that converges at one point of $\widetilde{M}$ is eventually constant.*

Proof. Suppose that there exists $\tilde{x} \in \widetilde{M}$ such that $(\overline{G}_n(\tilde{x}))_{n=1}^{\infty}$ is a convergent sequence in $\widetilde{M}$. Let $x := \pi(\tilde{x}) \in M$. There is $r > 0$ such that the balls $(B_{\tilde{\rho}}(\tilde{y}, r))_{\tilde{y} \in \pi^{-1}(x)}$ are mutually disjoint. Since $\overline{G}_n \in D_M$ for all $n \in \mathbb{N}$, we have $\pi(\overline{G}_n(\tilde{x})) = \pi(\tilde{x}) = x$, that is, $\overline{G}_n(\tilde{x}) \in \pi^{-1}(x)$ for all $n \in \mathbb{N}$. Thus, if $k \in \mathbb{N}$ is so large that $\tilde{\rho}(\overline{G}_i(\tilde{x}), \overline{G}_j(\tilde{x})) < r$ for all $i, j \geq k$, then $\overline{G}_i(\tilde{x}) = \overline{G}_j(\tilde{x})$. Since deck transformations are uniquely determined by their value at any point according to Proposition 6.3.1, we conclude that $\overline{G}_i = \overline{G}_j$ for all $i, j \geq k$. □

As an immediate consequence of this lemma, we obtain the following powerful result.

Corollary 6.3.4. *A sequence $(\overline{G}_n)_{n=1}^{\infty}$ in D_M converges uniformly on compact subsets of $\widetilde{M}$ if and only if it is eventually constant.*

Proof. If a sequence $(\overline{G}_n)_{n=1}^{\infty}$ in D_M converges uniformly on compact subsets of $\widetilde{M}$, then it converges pointwise on $\widetilde{M}$. Therefore, it is eventually constant according to Lemma 6.3.3. Obviously, any eventually constant sequence converges uniformly on compact subsets. □

Theoretically, establishing uniform convergence on compact subsets may prove to be difficult. Fortunately, there exists a simpler, pointwise criterion for sequences of deck transformations.

Lemma 6.3.5. *If $(\overline{G}_n)_{n=1}^{\infty}$ is a sequence in D_M and if $(\tilde{z}_n)_{n=1}^{\infty}$ is a sequence of points in $\widetilde{M}$ converging to some point $\tilde{z} \in \widetilde{M}$ such that $\tilde{w} := \lim_{n\to\infty} \overline{G}_n(\tilde{z}_n)$ exists, then the sequence $(\overline{G}_n)_{n=1}^{\infty}$ converges uniformly on compact subsets of $\widetilde{M}$ to an element $\overline{G} \in D_M$, which is uniquely determined by the requirement that $\overline{G}(\tilde{z}) = \tilde{w}$. In fact, the sequence $(\overline{G}_n)_{n=1}^{\infty}$ is eventually constant. More precisely, its terms eventually coincide with the unique element $\overline{G} \in D_M$ such that $\overline{G}(\tilde{z}) = \tilde{w}$.*

Proof. Let

$$\xi = \max\Big\{\sup\{\tilde{\rho}(\tilde{z}_n, \tilde{z}) : n \in \mathbb{N}\}, \sup\{\tilde{\rho}\big(\overline{G}_n(\tilde{z}_n), \tilde{w}\big) : n \in \mathbb{N}\}\Big\} < \infty.$$

Fix $r > 0$ and take $\tilde{x} \in \overline{B}_{\tilde{\rho}}(\tilde{z}, r)$. Then for every $n \in \mathbb{N}$, we have that

$$\begin{aligned}\tilde{\rho}\big(\overline{G}_n(\tilde{x}), \tilde{w}\big) &\leq \tilde{\rho}\big(\overline{G}_n(\tilde{x}), \overline{G}_n(\tilde{z})\big) + \tilde{\rho}\big(\overline{G}_n(\tilde{z}), \overline{G}_n(\tilde{z}_n)\big) + \tilde{\rho}\big(\overline{G}_n(\tilde{z}_n), \tilde{w}\big) \\ &= \tilde{\rho}(\tilde{x}, \tilde{z}) + \tilde{\rho}(\tilde{z}, \tilde{z}_n) + \tilde{\rho}\big(\overline{G}_n(\tilde{z}_n), \tilde{w}\big) \leq r + \xi + \xi = 2\xi + r.\end{aligned}$$

This means that for all $n \in \mathbb{N}$,

$$\overline{G}_n\big(\overline{B}_{\tilde{\rho}}(\tilde{z}, r)\big) \subseteq \overline{B}_{\tilde{\rho}}(\tilde{w}, 2\xi + r). \tag{6.13}$$

Now let $(n_k)_{k=1}^{\infty}$ be a strictly increasing sequence of positive integers. Fix $j \in \mathbb{N}$ and suppose that we have extracted from $(n_k)_{k=1}^{\infty}$ a subsequence $(n_k^{(j)})_{k=1}^{\infty}$ such that the sequence $(\overline{G}_{n_k^{(j)}})_{k=1}^{\infty}$ converges uniformly on $\overline{B}_{\tilde{\rho}}(\tilde{z}, j)$ to some continuous map

$\overline{G}^{(j)} : \overline{B}_{\tilde{\rho}}(\tilde{z}, j) \to \overline{B}_{\tilde{\rho}}(\tilde{w}, 2\xi + j)$. The inductive step, which also provides the basis of the induction, is as follows. Since both balls $\overline{B}_{\tilde{\rho}}(\tilde{z}, j+1)$ and $\overline{B}_{\tilde{\rho}}(\tilde{w}, 2\xi + j + 1)$ are compact and since the sequence of $\tilde{\rho}$-isometries $(\overline{G}_n)_{n=1}^{\infty}$ forms an equicontinuous family of maps, Arzelà–Ascoli's theorem permits us to extract from $(n_k^{(j)})_{k=1}^{\infty}$ a subsequence $(n_k^{(j+1)})_{k=1}^{\infty}$ such that the sequence $(\overline{G}_{n_k^{(j+1)}})_{k=1}^{\infty}$ converges uniformly on $\overline{B}_{\tilde{\rho}}(\tilde{z}, j+1)$ to some continuous map $\overline{G}^{(j+1)} : \overline{B}_{\tilde{\rho}}(\tilde{z}, j+1) \to \overline{B}_{\tilde{\rho}}(\tilde{w}, 2\xi + j + 1)$.

Obviously, $\overline{G}^{(j+1)}|_{\overline{B}_{\tilde{\rho}}(\tilde{z},j)} = \overline{G}^{(j)}$ and gluing all the maps $(\overline{G}^{(j)})_{j=1}^{\infty}$ together results in a map $\overline{G} : \widetilde{M} \to \widetilde{M}$ defined as $\overline{G}(\tilde{x}) = \overline{G}^{(j)}(\tilde{x})$ if $\tilde{x} \in \overline{B}_{\tilde{\rho}}(\tilde{z}, j)$. The sequence $(\overline{G}_{n_j^{(j)}})_{j=1}^{\infty}$ is a subsequence of $(\overline{G}_n)_{n=1}^{\infty}$ that converges uniformly to $\overline{G}$ on every compact ball $\overline{B}_{\tilde{\rho}}(\tilde{z}, i)$, $i \in \mathbb{N}$. This means that $(\overline{G}_{n_j^{(j)}})_{j=1}^{\infty}$ converges to $\overline{G}$ uniformly on compact subsets of $\widetilde{M}$. So $\overline{G}$ is a continuous map from $\widetilde{M}$ to $\widetilde{M}$ and

$$\pi \circ \overline{G}(\tilde{x}) = \pi\Big(\lim_{j\to\infty} \overline{G}_{n_j^{(j)}}(\tilde{x})\Big) = \lim_{j\to\infty} \big(\pi \circ \overline{G}_{n_j^{(j)}}(\tilde{x})\big) = \lim_{j\to\infty} \pi(\tilde{x}) = \pi(\tilde{x}).$$

Hence $\overline{G} \in D_M$ and

$$\overline{G}(\tilde{z}) = \lim_{j\to\infty} \overline{G}_{n_j^{(j)}}(\tilde{z}) = \lim_{j\to\infty} \overline{G}_{n_j^{(j)}}(\tilde{z}_{n_j^{(j)}}) = \lim_{n\to\infty} \overline{G}_n(\tilde{z}_n) = \tilde{w}.$$

The uniqueness of $\overline{G}$ follows from Proposition 6.3.1. The rest of the proposition follows from Corollary 6.3.4. □

6.3.2 Lifts

We now study the concept of the *lift of a map*. We keep with the convention adopted in Chapter 2 of denoting a lift of a given map with a tilde above the map.

Proposition 6.3.6. *Let N and M be two compact connected smooth manifolds. For any continuous map $S : N \to M$ there exists a continuous map $\tilde{S} : \tilde{N} \to \widetilde{M}$ such that the following diagram commutes:*

$$\begin{array}{ccc} \widetilde{N} & \xrightarrow{\tilde{S}} & \widetilde{M} \\ {\scriptstyle \pi_N}\downarrow & & \downarrow{\scriptstyle \pi_M} \\ N & \xrightarrow[S]{} & M \end{array}$$

That is,

$$S \circ \pi_N = \pi_M \circ \tilde{S}.$$

All such maps $\tilde{S}$ are called lifts *of S.*

Proof. Denote by $\pi_1(X)$ be the fundamental group of a path-connected space X and let $(\pi_X)_* : \pi_1(\widetilde{X}) \to \pi_1(X)$ be the homomorphism induced by the canonical projection $\pi_X : \widetilde{X} \to X$, where $\widetilde{X}$ is the universal covering space of X.

Consider the diagram

$$\begin{tikzcd} & & \widetilde{M} \arrow[d, "\pi_M"] \\ \widetilde{N} \arrow[r, "\pi_N"'] & N \arrow[r, "S"'] & M \end{tikzcd}$$

Notice that $(S \circ \pi_N)_*(\pi_1(\widetilde{N})) = \{0\}_M = (\pi_M)_*(\{0\}_M) = (\pi_M)_*(\pi_1(\widetilde{M}))$. From the lifting criterion (cf. Proposition 1.33 in [28]), there thus exists a continuous map $\widetilde{S} : \widetilde{N} \to \widetilde{M}$ such that the following diagram commutes:

$$\begin{tikzcd} & & \widetilde{M} \arrow[d, "\pi_M"] \\ \widetilde{N} \arrow[r, "\pi_N"'] \arrow[urr, "\widetilde{S}"] & N \arrow[r, "S"'] & M \end{tikzcd}$$

□

One particularly interesting case is the lifting of covering maps.

Proposition 6.3.7. *If $S : N \to M$ is a covering map, then all of its lifts $\widetilde{S} : \widetilde{N} \to \widetilde{M}$ are homeomorphisms.*

Proof. Recall from the proof of Proposition 6.3.6 that

$$(S \circ \pi_N)_*\big(\pi_1(\widetilde{N})\big) = (\pi_M)_*\big(\pi_1(\widetilde{M})\big).$$

From the lifting criterion (cf. Proposition 1.33 in [28]), there then exists a continuous map $\widehat{S} : \widetilde{M} \to \widetilde{N}$ such that the following diagram commutes:

$$\begin{tikzcd} & \widetilde{N} \arrow[d, "\pi_N"] \\ & N \arrow[d, "S"] \\ \widetilde{M} \arrow[r, "\pi_M"'] \arrow[uur, "\widehat{S}"] & M \end{tikzcd}$$

Using Proposition 6.3.6, let $\widetilde{S} : \widetilde{N} \to \widetilde{M}$ be a lift of $S : N \to M$. Then

$$\pi_M \circ (\widetilde{S} \circ \widehat{S}) = (\pi_M \circ \widetilde{S}) \circ \widehat{S} = (S \circ \pi_N) \circ \widehat{S} = \pi_M.$$

Hence, in view of Proposition 6.3.1, the map $\widetilde{S} \circ \widehat{S}$ is a homeomorphism. On the other hand, $(S \circ \pi_N) \circ (\widehat{S} \circ \widetilde{S}) = (S \circ \pi_N \circ \widehat{S}) \circ \widetilde{S} = \pi_M \circ \widetilde{S} = S \circ \pi_N$. Since $S \circ \pi_N : \widetilde{N} \to M$ is a covering map, an argument similar to the one yielding Proposition 6.3.1 certifies that $\widehat{S} \circ \widetilde{S}$ is a homeomorphism. As $\widetilde{S} \circ \widehat{S}$ and $\widehat{S} \circ \widetilde{S}$ are homeomorphisms, so are $\widehat{S}$ and $\widetilde{S}$. □

Note that deck transformations of M are simply lifts of the identity covering map $\mathrm{Id}_M : M \to M$.

We now provide a characterization of lifts.

Proposition 6.3.8. *Let N and M be two compact connected smooth manifolds. A continuous map $\widetilde{S} : \widetilde{N} \to \widetilde{M}$ is a lift of some continuous map from N to M if and only if there exists a (necessarily unique) map $h : D_N \to D_M$ such that the following diagram commutes for all $\overline{F} \in D_N$:*

$$\begin{tikzcd} \widetilde{N} \arrow[r, "\overline{F}"] \arrow[d, "\widetilde{S}"'] & \widetilde{N} \arrow[d, "\widetilde{S}"] \\ \widetilde{M} \arrow[r, "h(\overline{F})"'] & \widetilde{M} \end{tikzcd}$$

In other words,

$$\widetilde{S} \circ \overline{F} = h(\overline{F}) \circ \widetilde{S}, \quad \forall \overline{F} \in D_N.$$

Moreover, $h : D_N \to D_M$ is a group homomorphism. This induced homomorphism of the groups of deck transformations will also be denoted by $\widetilde{S}^*$.

Proof. First, suppose that $S : N \to M$ is a continuous map which has for lift $\widetilde{S} : \widetilde{N} \to \widetilde{M}$. For any $\overline{F} \in D_N$, we then have that

$$\pi_M \circ (\widetilde{S} \circ \overline{F}) = (\pi_M \circ \widetilde{S}) \circ \overline{F} = (S \circ \pi_N) \circ \overline{F} = S \circ (\pi_N \circ \overline{F}) = S \circ \pi_N = \pi_M \circ \widetilde{S}.$$

It follows from Proposition 6.3.2 that there exists a unique $h(\overline{F}) \in D_M$ such that $h(\overline{F}) \circ \widetilde{S} = \widetilde{S} \circ \overline{F}$.

For the converse implication, suppose that $\widetilde{S} : \widetilde{N} \to \widetilde{M}$ is a continuous map and that there is a map $h : D_N \to D_M$ such that $\widetilde{S} \circ \overline{F} = h(\overline{F}) \circ \widetilde{S}$ for all $\overline{F} \in D_N$. Define the map $S : N \to M$ by setting

$$S(x) := \pi_M \circ \widetilde{S}(\widetilde{x}),$$

where $\widetilde{x}$ is any element of $\pi_N^{-1}(x)$. To be well-defined, we must show that $S(x)$ is independent of the element $\widetilde{x}$ chosen in $\pi_N^{-1}(x)$. For this, let $\widetilde{z} \in \pi_N^{-1}(x)$. Since $\pi_N(\widetilde{x}) = \pi_N(\widetilde{z})$, there is a unique $\overline{F} \in D_N$ such that $\widetilde{z} = \overline{F}(\widetilde{x})$. Then

$$\pi_M \circ \widetilde{S}(\widetilde{z}) = \pi_M \circ \widetilde{S} \circ \overline{F}(\widetilde{x}) = \pi_M \circ h(\overline{F}) \circ \widetilde{S}(\widetilde{x}) = \pi_M \circ \widetilde{S}(\widetilde{x}).$$

The map $S : N \to M$ is thus well-defined. It is continuous since $\pi_M \circ \widetilde{S}$ is continuous and the projection $\pi_N : \widetilde{N} \to N$ is a covering map. Furthermore, $S \circ \pi_N(\widetilde{x}) = S(x) = \pi_M \circ \widetilde{S}(\widetilde{x})$. So $\widetilde{S}$ is a lift of S.

Regarding the last assertion, for any $\overline{F}_1, \overline{F}_2 \in D_N$ we have that $\overline{F}_1 \circ \overline{F}_2 \in D_N$. As $h : D_N \to D_M$ is the unique map such that $h(\overline{F}) \circ \widetilde{S} = \widetilde{S} \circ \overline{F}$ for all $\overline{F} \in D_N$, it ensues that

$$\big(h(\overline{F}_1 \circ \overline{F}_2)\big) \circ \widetilde{S} = \widetilde{S} \circ \overline{F}_1 \circ \overline{F}_2 = h(\overline{F}_1) \circ \widetilde{S} \circ \overline{F}_2 = h(\overline{F}_1) \circ h(\overline{F}_2) \circ \widetilde{S} = \big(h(\overline{F}_1) \circ h(\overline{F}_2)\big) \circ \widetilde{S}.$$

The uniqueness of h implies that $h(\overline{F}_1 \circ \overline{F}_2) = h(\overline{F}_1) \circ h(\overline{F}_2)$. So h is a homomorphism between the groups D_N and D_M. □

Given a certain lift, we next show that the set of all lifts that share the same induced homomorphism as the given lift naturally forms a complete metric space.

Let N and M be two compact connected smooth manifolds and let $\widetilde{\alpha} : \widetilde{N} \to \widetilde{M}$ be a lift of some continuous map from N to M. Let $V_{\widetilde{\alpha}}$ denote the set of all continuous maps $\widetilde{S} : \widetilde{N} \to \widetilde{M}$ that are lifts of some continuous map from N to M and such that $\widetilde{S}^* = \widetilde{\alpha}^*$. Let $\widetilde{\rho}_\infty : V_{\widetilde{\alpha}} \times V_{\widetilde{\alpha}} \to [0, \infty]$ be the function defined by

$$\widetilde{\rho}_\infty(\widetilde{S}_1, \widetilde{S}_2) := \sup\{\widetilde{\rho}_M(\widetilde{S}_1(\widetilde{x}), \widetilde{S}_2(\widetilde{x})) \mid \widetilde{x} \in \widetilde{N}\}.$$

Lemma 6.3.9. *The function $\widetilde{\rho}_\infty$ constitutes a metric on the set $V_{\widetilde{\alpha}}$ and the metric space $(V_{\widetilde{\alpha}}, \widetilde{\rho}_\infty)$ is complete.*

Proof. The symmetry of $\widetilde{\rho}_\infty$ and the triangle inequality being obvious, we shall only demonstrate the finiteness of $\widetilde{\rho}_\infty$ to establish that $\widetilde{\rho}_\infty$ is a metric on $V_{\widetilde{\alpha}}$.

Let $R > \operatorname{diam}_{\rho_N}(N)$. We first show that

$$\pi_N\big(B_{\widetilde{\rho}_N}(\widetilde{z}, R)\big) = N, \quad \forall\, \widetilde{z} \in \widetilde{N}. \tag{6.14}$$

To that end, fix $\widetilde{z} \in \widetilde{N}$ and $x \in N$. Let γ be a smooth curve in N joining $\pi_N(\widetilde{z})$ and x whose ρ_N-length is smaller than R. Let $\widetilde{\gamma}$ be a lift of γ to $\widetilde{N}$ whose initial point is $\widetilde{z}$. Let $\widetilde{w}$ denote the other endpoint of the curve γ. Since $\pi_N : \widetilde{N} \to N$ is a local (and thus infinitesimal) isometry, we deduce that

$$\widetilde{\rho}_N(\widetilde{z}, \widetilde{w}) \le \ell_{\widetilde{\rho}_N}(\widetilde{\gamma}) = \ell_{\rho_N}(\gamma) < R.$$

Since $x = \pi_N(\widetilde{w})$, it follows that $x \in \pi_N(B_{\widetilde{\rho}_N}(\widetilde{z}, R))$, and thus (6.14) holds.

Now, let $\widetilde{S}_1, \widetilde{S}_2 \in V_{\widetilde{\alpha}}$. Since $\overline{B}_{\widetilde{\rho}_N}(\widetilde{z}, R)$ is a compact subset of $\widetilde{N}$, we have

$$A := \sup\{\widetilde{\rho}_M(\widetilde{S}_1(\widetilde{x}), \widetilde{S}_2(\widetilde{x})) \mid \widetilde{x} \in \overline{B}_{\widetilde{\rho}_N}(\widetilde{z}, R)\} < \infty.$$

Take an arbitrary point $\widetilde{w} \in \widetilde{N}$. In light of (6.14), there exists a point $\widetilde{x} \in B_{\widetilde{\rho}_N}(\widetilde{z}, R)$ such that $\pi_N(\widetilde{x}) = \pi_N(\widetilde{w})$. Hence, there exists a deck transformation $\overline{F} \in D_N$ such that $\overline{F}(\widetilde{x}) = \widetilde{w}$. Consequently,

$$\begin{aligned}
\widetilde{\rho}_M(\widetilde{S}_1(\widetilde{w}), \widetilde{S}_2(\widetilde{w})) &= \widetilde{\rho}_M(\widetilde{S}_1 \circ \overline{F}(\widetilde{x}), \widetilde{S}_2 \circ \overline{F}(\widetilde{x})) \\
&= \widetilde{\rho}_M(\widetilde{S}_1^*(\overline{F}) \circ \widetilde{S}_1(\widetilde{x}), \widetilde{S}_2^*(\overline{F}) \circ \widetilde{S}_2(\widetilde{x})) \\
&= \widetilde{\rho}_M(\widetilde{\alpha}^*(\overline{F}) \circ \widetilde{S}_1(\widetilde{x}), \widetilde{\alpha}^*(\overline{F}) \circ \widetilde{S}_2(\widetilde{x})) \\
&= \widetilde{\rho}_M(\widetilde{S}_1(\widetilde{x}), \widetilde{S}_2(\widetilde{x})) \le A.
\end{aligned}$$

Thus, $\widetilde{\rho}_\infty(\widetilde{S}_1, \widetilde{S}_2) \le A < \infty$ and thereby $\widetilde{\rho}_\infty(V_{\widetilde{\alpha}} \times V_{\widetilde{\alpha}}) \subseteq [0, \infty)$. So $\widetilde{\rho}_\infty$ is a metric.

Let us now show that $\widetilde{\rho}_\infty$ is complete. Let $(\widetilde{S}_n)_{n=1}^\infty$ be a Cauchy sequence in $V_{\widetilde{\alpha}}$. Then this sequence is also a Cauchy sequence with respect to the topology of uniform convergence on compact subsets of $\widetilde{N}$. Let $\widehat{S} : \widetilde{N} \to \widetilde{M}$ be the limit of that sequence. Let $\overline{F} \in D_N$. Since $\widetilde{S}_n \circ \overline{F} = \widetilde{\alpha}^*(\overline{F}) \circ \widetilde{S}_n$ for all $n \in \mathbb{N}$, we infer that $\widehat{S} \circ \overline{F} = \widetilde{\alpha}^*(\overline{F}) \circ \widehat{S}$. By Proposition 6.3.8, we deduce that $\widehat{S} \in V_{\widetilde{\alpha}}$. Let $\widetilde{z} \in \widetilde{N}$ and $R > \operatorname{diam}_{\rho_N}(N)$. As established above,

$$\widetilde{\rho}_\infty(\widetilde{S}_n, \widehat{S}) \leq \sup\{\widetilde{\rho}_M(\widetilde{S}_n(\widetilde{x}), \widehat{S}(\widetilde{x})) \mid \widetilde{x} \in \overline{B}_{\widetilde{\rho}_N}(\widetilde{z}, R)\}$$

for all $n \in \mathbb{N}$, whence the uniform convergence of the sequence $(\widetilde{S}_n)_{n=1}^\infty$ to $\widehat{S}$ on the compact ball $\overline{B}_{\widetilde{\rho}_N}(\widetilde{z}, R)$ implies that it also converges to $\widehat{S}$ with respect to the metric $\widetilde{\rho}_\infty$ on $V_{\widetilde{\alpha}}$. □

6.4 Dynamical properties

In the last section of this chapter, we will discover the dynamical properties of Shub expanding endomorphisms.

6.4.1 Expanding property

Shub expanding endomorphisms are distance expanding in the following sense.

Theorem 6.4.1. *Every Shub expanding endomorphism $T : M \to M$ is distance expanding with respect to the distance ρ induced on M by any Riemannian metric ρ e-adapted to T.*

Proof. According to Theorem 6.2.4, every Shub expanding endomorphism T is a covering map, and thus a local homeomorphism. Thanks to the compactness of M, there then exists some $\delta_T > 0$ such that the map $T|_{B_\rho(x, 2\delta_T)} : B_\rho(x, 2\delta_T) \to M$ is injective for all $x \in M$. Fix two points $x_1, x_2 \in M$ such that $\rho(x_1, x_2) < 2\delta_T$ and pick any smooth curve $\gamma : I \to M$ joining $T(x_1)$ and $T(x_2)$, that is, $\gamma(a) = T(x_1)$ and $\gamma(b) = T(x_2)$, where $I = [a, b] \subseteq \mathbb{R}$. Since $T : M \to M$ is a covering map, there exists a smooth curve $\widehat{\gamma} : I \to M$ such that $\widehat{\gamma}(a) = x_1$ and $T \circ \widehat{\gamma} = \gamma$. In particular, $T(\widehat{\gamma}(b)) = T(x_2)$. So if $\widehat{\gamma}(b) \notin B_\rho(x_1, 2\delta_T)$, then

$$\rho(\widehat{\gamma}(a), \widehat{\gamma}(b)) = \rho(x_1, \widehat{\gamma}(b)) \geq 2\delta_T > \rho(x_1, x_2).$$

On the other hand, if $\widehat{\gamma}(b) \in B_\rho(x_1, 2\delta_T)$, then $\widehat{\gamma}(b) = x_2$ since the map $T|_{B_\rho(x_1, 2\delta_T)}$ is injective, whence $\rho(\widehat{\gamma}(a), \widehat{\gamma}(b)) = \rho(x_1, x_2)$. In either case,

$$\ell_\rho(\widehat{\gamma}) \geq \rho(\widehat{\gamma}(a), \widehat{\gamma}(b)) \geq \rho(x_1, x_2).$$

Therefore,

$$\begin{aligned}
\ell_\rho(\gamma) &= \int_I \|D_t\gamma(e_t)\|\, dt = \int_I \|D_t(T \circ \widehat{\gamma})(e_t)\|\, dt \\
&= \int_I \|D_{\widehat{\gamma}(t)}T(D_t\widehat{\gamma}(e_t))\|\, dt \\
&\geq \lambda \int_I \|D_t\widehat{\gamma}(e_t)\|\, dt = \lambda \ell_\rho(\widehat{\gamma}) \geq \lambda\rho(x_1, x_2).
\end{aligned}$$

So, taking the infimum over all curves $\gamma \in \Gamma(T(x_1), T(x_2))$, we conclude that $\rho(T(x_1), T(x_2)) \geq \lambda\rho(x_1, x_2)$ for all $x_1, x_2 \in M$ such that $\rho(x_1, x_2) < 2\delta_T$. □

The previous theorem shows that Shub expanding endomorphisms provide a large class of distance expanding maps.

6.4.2 Topological exactness and density of periodic points

The next theorem reveals several additional dynamical properties of Shub expanding maps.

Theorem 6.4.2. *Let M be a compact connected smooth manifold. If $T : M \to M$ is a Shub expanding endomorphism, then:*
(a) *T has a fixed point.*
(b) *The universal covering manifold $\widetilde{M}$ is diffeomorphic to $\mathbb{R}^k$, where $k = \dim(M)$.*
(c) *T is topologically exact.*
(d) *The set of periodic points of T is dense in M.*

Proof. Let ρ be a Riemannian metric on M e-adapted to T, and $\widetilde{\rho}$ the Riemannian metric induced by π and ρ on $\widetilde{M}$. Recall that with these metrics, the projection map $\pi : \widetilde{M} \to M$ is an infinitesimal and local isometry and all maps in D_M are infinitesimal and global isometries with respect to the metric $\widetilde{\rho}$. Let $\widetilde{T} : \widetilde{M} \to \widetilde{M}$ be a lift of T to $\widetilde{M}$. Such a lift exists according to Proposition 6.3.6. By Theorem 6.2.4 and Proposition 6.3.7, the map $\widetilde{T}$ is a diffeomorphism. With a calculation analogous to that in the proof of Theorem 6.4.1, we can prove the following claim.

Claim. *The diffeomorphism $\widetilde{T}^{-1} : \widetilde{M} \to \widetilde{M}$ is a global contraction with respect to the metric $\widetilde{\rho}$. More precisely,*

$$\widetilde{\rho}\big(\widetilde{T}^{-1}(\widetilde{x}), \widetilde{T}^{-1}(\widetilde{y})\big) \leq \lambda^{-1}\widetilde{\rho}(\widetilde{x}, \widetilde{y}), \quad \forall \widetilde{x}, \widetilde{y} \in \widetilde{M}. \tag{6.15}$$

(a) By the Banach contraction principle, the map $\widetilde{T}^{-1} : \widetilde{M} \to \widetilde{M}$ has a unique fixed point $\widetilde{w} \in \widetilde{M}$. That is, $\widetilde{T}^{-1}(\widetilde{w}) = \widetilde{w}$, or, equivalently, $\widetilde{T}(\widetilde{w}) = \widetilde{w}$. Then $T(\pi(\widetilde{w})) = \pi(\widetilde{T}(\widetilde{w})) = \pi(\widetilde{w})$, that is, $\pi(\widetilde{w})$ is a fixed point of $T : M \to M$.

(b) Let $\widetilde{w} \in \widetilde{M}$ be the fixed point of the maps $\widetilde{T}, \widetilde{T}^{-1} : \widetilde{M} \to \widetilde{M}$. Since $\widetilde{M}$ is a smooth manifold, there exist $r > 0$ and a smooth diffeomorphism $\varphi : V \to B_{\widetilde{\rho}}(\widetilde{w}, r)$ from an open neighborhood V of the origin in $\mathbb{R}^k$ onto $B_{\widetilde{\rho}}(\widetilde{w}, r)$ and such that $\varphi(0) = \widetilde{w}$ and $\varphi'(0) : \mathbb{R}^k \to T_{\widetilde{w}}\widetilde{M}$ is an isometry. Since

$$\widetilde{T}^{-1}\big(B_{\widetilde{\rho}}(\widetilde{w}, r)\big) \subseteq B_{\widetilde{\rho}}\big(\widetilde{w}, \lambda^{-1}r\big) \subseteq B_{\widetilde{\rho}}(\widetilde{w}, r),$$

the conjugate of $\widetilde{T}^{-1}$ via φ is a well-defined diffeomorphism, namely

$$G := \varphi^{-1} \circ \widetilde{T}^{-1} \circ \varphi : V \to V.$$

Hence, for all $k \geq 0$,

$$G^k = \varphi^{-1} \circ \widetilde{T}^{-k} \circ \varphi : V \to V. \tag{6.16}$$

Notice that

$$G(0) = \varphi^{-1} \circ \widetilde{T}^{-1} \circ \varphi(0) = \varphi^{-1}(\widetilde{T}^{-1}(\widetilde{w})) = \varphi^{-1}(\widetilde{w}) = 0$$

and

$$\|G'(0)\| = \|(\varphi^{-1})'(\widetilde{w}) \circ (\widetilde{T}^{-1})'(\widetilde{w}) \circ \varphi'(0)\| = \|(\widetilde{T}^{-1})'(\widetilde{w})\| \leq \lambda^{-1} < 1.$$

Therefore, there exists $R > 0$ so small that $\overline{B}(0, R) \subseteq V$,

$$G(B(0,R)) \subseteq B\left(0, \frac{\lambda^{-1}+1}{2}R\right) \subseteq B(0,R), \tag{6.17}$$

and

$$\|G'(x)\| \leq \frac{\lambda^{-1}+1}{2}, \quad \forall x \in B(0,R). \tag{6.18}$$

In view of Exercise 6.5.5, there then exists a diffeomorphism $\widehat{G} : \mathbb{R}^k \to \mathbb{R}^k$ such that

$$\widehat{G}|_{B(0,R)} = G \quad \text{and} \quad \|\widehat{G}'(x)\| \leq \frac{\lambda^{-1}+1}{2}, \quad \forall x \in \mathbb{R}^k. \tag{6.19}$$

Let us now define a map $H : \widetilde{M} \to \mathbb{R}^k$ in the following way. For each $\widetilde{x} \in \widetilde{M}$, choose $n = n(\widetilde{x}) \geq 0$ such that $\widetilde{T}^{-n}(\widetilde{x}) \in \varphi(B(0,R))$ and declare

$$H(\widetilde{x}) = \widehat{G}^{-n} \circ \varphi^{-1} \circ \widetilde{T}^{-n}(\widetilde{x}). \tag{6.20}$$

We shall first show that $H(\widetilde{x})$ is well-defined by establishing that its definition is independent of the choice of n. Then we will proceed on showing that $H : \widetilde{M} \to \mathbb{R}^k$ is a diffeomorphism. So assume that, in addition to $\widetilde{T}^{-n}(\widetilde{x})$, the iterate $\widetilde{T}^{-j}(\widetilde{x})$ is in $\varphi(B(0,R))$. We may assume without loss of generality that $0 \leq j \leq n$. Write $\widetilde{T}^{-j}(\widetilde{x}) = \varphi(x')$, where $x' \in B(0,R) \subseteq V$. Using (6.16) with $k = n - j$, (6.17) and (6.19), we get

$$\begin{aligned}
\widehat{G}^{-n} \circ \varphi^{-1} \circ \widetilde{T}^{-n}(\widetilde{x}) &= \widehat{G}^{-n} \circ \varphi^{-1} \circ \widetilde{T}^{-(n-j)} \circ \widetilde{T}^{-j}(\widetilde{x}) \\
&= \widehat{G}^{-n} \circ \varphi^{-1} \circ \widetilde{T}^{-(n-j)} \circ \varphi(x') \\
&= \widehat{G}^{-n} \circ G^{n-j}(x') = \widehat{G}^{-j}(x') = \widehat{G}^{-j} \circ \varphi^{-1} \circ \widetilde{T}^{-j}(\widetilde{x}).
\end{aligned}$$

Thus, the map $H : \widetilde{M} \to \mathbb{R}^k$ is well-defined. Since the same n used to define H at $\widetilde{x}$ works for any point $\widetilde{y} \in B_{\widetilde{\rho}}(\widetilde{x}, \varepsilon)$ if $\varepsilon > 0$ is sufficiently small, it follows from (6.20) that the map H is smooth as a composition of smooth maps. As a composition of local diffeomorphisms, it is further a local diffeomorphism. It only remains to show that H

is globally bijective. To prove injectivity, assume that $H(\tilde{x}) = H(\tilde{y})$. Since $\tilde{T}^{-1} : \widetilde{M} \to \widetilde{M}$ is a global contraction fixing $\tilde{w} = \varphi(0)$, there exists $n \geq 0$ so large that both $\tilde{T}^{-n}(\tilde{x})$ and $\tilde{T}^{-n}(\tilde{y})$ lie in $\varphi(B(0,R))$. Then

$$\widehat{G}^{-n} \circ \varphi^{-1} \circ \tilde{T}^{-n}(\tilde{x}) = \widehat{G}^{-n} \circ \varphi^{-1} \circ \tilde{T}^{-n}(\tilde{y}).$$

Applying to this equality $\widehat{G}^n$, φ and $\tilde{T}^n$ successively, we conclude that $\tilde{x} = \tilde{y}$, thereby establishing the injectivity of H. To prove its surjectivity, take an arbitrary $y \in \mathbb{R}^k$. Since $\widehat{G} : \mathbb{R}^k \to \mathbb{R}^k$ is a global contraction according to (6.19) and has 0 for fixed point, there exists $n \geq 0$ so large that $\widehat{G}^n(y) \in B(0,R)$. Then $\varphi(\widehat{G}^n(y)) \in \varphi(B(0,R))$. It follows that

$$H\big(\tilde{T}^n(\varphi(\widehat{G}^n(y)))\big) = \widehat{G}^{-n} \circ \varphi^{-1} \circ \tilde{T}^{-n}\big(\tilde{T}^n \circ \varphi \circ \widehat{G}^n(y)\big) = y.$$

Thus, H is surjective. Because $H : \widetilde{M} \to \mathbb{R}^k$ is a diffeomorphism, $k = \dim(\widetilde{M}) = \dim(M)$.

(c) Let $R > \operatorname{diam}_\rho(M)$ and recall that (6.14) then holds with $N = M$. Let U be a nonempty, open subset of M. Fix an arbitrary $x \in U$ and any $\tilde{x} \in \pi^{-1}(x)$. Since $\pi^{-1}(U)$ is an open subset of $\widetilde{M}$ containing $\tilde{x}$, there exists some $r > 0$ such that $B_{\tilde{\rho}}(\tilde{x}, r) \subseteq \pi^{-1}(U)$. Choose $n \geq 0$ so large that $\lambda^n r \geq R$. By (6.15), we observe that

$$\tilde{T}^n\big(B_{\tilde{\rho}}(\tilde{x}, r)\big) \supseteq B_{\tilde{\rho}}\big(\tilde{T}^n(\tilde{x}), \lambda^n r\big) \supseteq B_{\tilde{\rho}}\big(\tilde{T}^n(\tilde{x}), R\big).$$

It follows from (6.14) that

$$T^n(U) \supseteq T^n\big(\pi(B_{\tilde{\rho}}(\tilde{x}, r))\big) = \pi\big(\tilde{T}^n(B_{\tilde{\rho}}(\tilde{x}, r))\big) \supseteq \pi\big(B_{\tilde{\rho}}(\tilde{T}^n(\tilde{x}), R)\big) = M.$$

Thus, T is topologically exact.

(d) As in part (c), let $R > \operatorname{diam}_\rho(M)$ and recall that (6.14) then holds. Let also U be a nonempty, open subset of M. Fix an arbitrary $x \in U$ and $\tilde{x} \in \pi^{-1}(x)$. Since $\pi^{-1}(U)$ is an open subset of $\widetilde{M}$ containing $\tilde{x}$, there exists some $0 < r \leq R$ such that $\overline{B}_{\tilde{\rho}}(\tilde{x}, r) \subseteq \pi^{-1}(U)$. Choose $n \geq 0$ so large that $\lambda^n r \geq 2R$. By (6.14), there exists $\tilde{y} \in B_{\tilde{\rho}}(\tilde{T}^n(\tilde{x}), R)$ such that $\pi(\tilde{y}) = x = \pi(\tilde{x})$ and hence there is $\overline{G} \in D_M$ such that $\overline{G}(\tilde{x}) = \tilde{y} \in B_{\tilde{\rho}}(\tilde{T}^n(\tilde{x}), R)$. Using (6.15), it follows that

$$\tilde{T}^n\big(B_{\tilde{\rho}}(\tilde{x}, r)\big) \supseteq B_{\tilde{\rho}}\big(\tilde{T}^n(\tilde{x}), \lambda^n r\big) \supseteq B_{\tilde{\rho}}\big(\tilde{T}^n(\tilde{x}), 2R\big) \supseteq B_{\tilde{\rho}}\big(\overline{G}(\tilde{x}), R\big).$$

Since $\overline{G} : \widetilde{M} \to \widetilde{M}$ is a $\tilde{\rho}$-isometry, we deduce that

$$\tilde{T}^n\big(B_{\tilde{\rho}}(\tilde{x}, r)\big) \supseteq \overline{G}\big(B_{\tilde{\rho}}(\tilde{x}, R)\big) \supseteq \overline{G}\big(B_{\tilde{\rho}}(\tilde{x}, r)\big).$$

Consequently,

$$\tilde{T}^{-n} \circ \overline{G}\big(\overline{B}_{\tilde{\rho}}(\tilde{x}, r)\big) \subseteq \overline{B}_{\tilde{\rho}}(\tilde{x}, r).$$

As $\widetilde{T}^{-n}$ is a contraction and $\overline{G}$ is an isometry, the map $\widetilde{T}^{-n} \circ \overline{G} : \overline{B}_{\widetilde{\rho}}(\widetilde{x}, r) \to \overline{B}_{\widetilde{\rho}}(\widetilde{x}, r)$ is a contraction and the Banach contraction principle asserts that $\widetilde{T}^{-n} \circ \overline{G}$ has a fixed point $\widetilde{w} \in \overline{B}_{\widetilde{\rho}}(\widetilde{x}, r)$. Therefore, $\widetilde{T}^n(\widetilde{w}) = \overline{G}(\widetilde{w})$, and hence

$$T^n\big(\pi(\widetilde{w})\big) = \pi\big(\widetilde{T}^n(\widetilde{w})\big) = \pi\big(\overline{G}(\widetilde{w})\big) = \pi(\widetilde{w}).$$

Furthermore,

$$\pi(\widetilde{w}) \in \pi\big(\overline{B}_{\widetilde{\rho}}(\widetilde{x}, r)\big) \subseteq U.$$

Thus, T^n has a fixed point in U. Since U is an arbitrary open set in M, we conclude that the set of periodic points of T is dense in M. □

6.4.3 Topological conjugacy and structural stability

In order to establish a topological conjugacy between any two Shub expanding endomorphisms that are homotopic, we shall first show that the existence of a semiconjugacy between the induced homomorphisms of lifts of two Shub expanding maps implies the existence of a semiconjugacy between the lifts themselves.

Lemma 6.4.3. *Let N and M be compact connected smooth manifolds. Let $S : N \to N$ and $T : M \to M$ be Shub expanding endomorphisms. Finally, let $\alpha : N \to M$ be a continuous map. If there exist lifts $\widetilde{S} : \widetilde{N} \to \widetilde{N}$ of S, $\widetilde{T} : \widetilde{M} \to \widetilde{M}$ of T, and $\widetilde{\alpha} : \widetilde{N} \to \widetilde{M}$ of α, such that*

$$\widetilde{T}^* \circ \widetilde{\alpha}^* = \widetilde{\alpha}^* \circ \widetilde{S}^*, \tag{6.21}$$

then there exists a unique map $\widetilde{H} \in V_{\widetilde{\alpha}}$ such that

$$\widetilde{T} \circ \widetilde{H} = \widetilde{H} \circ \widetilde{S},$$

where $V_{\widetilde{\alpha}}$ is the set of all continuous maps $\widetilde{A} : \widetilde{N} \to \widetilde{M}$ that are lifts of some continuous map from N to M and such that $\widetilde{A}^* = \widetilde{\alpha}^*$.

Proof. For every $\widetilde{A} \in V_{\widetilde{\alpha}}$, define

$$\theta(\widetilde{A}) := \widetilde{T}^{-1} \circ \widetilde{A} \circ \widetilde{S}.$$

Claim 1. *The transformation $\theta(\widetilde{A})$ is a lift of some continuous map from N to M.*

Proof of Claim 1. According to Proposition 6.3.8, for every $\overline{G} \in D_M$ we have $\widetilde{T}^*(\overline{G}) \circ \widetilde{T} = \widetilde{T} \circ \overline{G}$. Since $\widetilde{T}$ is a homeomorphism by Proposition 6.3.7 and Theorem 6.2.4, we obtain that $\overline{G} = \widetilde{T}^{-1} \circ \widetilde{T}^*(\overline{G}) \circ \widetilde{T}$. So, if $\overline{G}_1, \overline{G}_2 \in D_M$ and $\widetilde{T}^*(\overline{G}_1) = \widetilde{T}^*(\overline{G}_2)$, then $\overline{G}_1 = \overline{G}_2$. This means that the homomorphism $\widetilde{T}^* : D_M \to D_M$ is injective. Since $\widetilde{\alpha}^* \circ \widetilde{S}^* = \widetilde{T}^* \circ \widetilde{\alpha}^*$, the range of the map $\widetilde{\alpha}^* \circ \widetilde{S}^*$ is contained in the range of $\widetilde{T}^*$ and, therefore, the map $\beta := (\widetilde{T}^*)^{-1} \circ \widetilde{\alpha}^* \circ \widetilde{S}^* : D_N \to D_M$ is well-defined. We will show that

$$\theta(\widetilde{A}) \circ \overline{G} = \beta(\overline{G}) \circ \theta(\widetilde{A}), \quad \forall \widetilde{A} \in V_{\widetilde{\alpha}}, \, \forall \overline{G} \in D_N.$$

Indeed, for all $\widetilde{A} \in V_{\widetilde{\alpha}}$, all $\overline{G}_1 \in D_N$ and all $\overline{G}_2 \in D_M$, we have

$$(\widetilde{A} \circ \widetilde{S}) \circ \overline{G}_1 = (\widetilde{A}^* \circ \widetilde{S}^*)(\overline{G}_1) \circ \widetilde{A} \circ \widetilde{S} \quad \text{and} \quad \widetilde{T}^{-1} \circ \widetilde{T}^*(\overline{G}_2) \circ \widetilde{T} = \overline{G}_2.$$

Recalling that $\widetilde{A}^* = \widetilde{\alpha}^*$ by hypothesis, it follows that for all $\overline{G} \in D_N$,

$$\begin{aligned}
\theta(\widetilde{A}) \circ \overline{G} &= \widetilde{T}^{-1} \circ \widetilde{A} \circ \widetilde{S} \circ \overline{G} = \widetilde{T}^{-1} \circ (\widetilde{A}^* \circ \widetilde{S}^*)(\overline{G}) \circ \widetilde{A} \circ \widetilde{S} \\
&= \widetilde{T}^{-1} \circ \widetilde{T}^* \circ (\widetilde{T}^*)^{-1} \circ \widetilde{A}^* \circ \widetilde{S}^*(\overline{G}) \circ \widetilde{A} \circ \widetilde{S} \\
&= \widetilde{T}^{-1} \circ \widetilde{T}^*(((\widetilde{T}^*)^{-1} \circ \widetilde{A}^* \circ \widetilde{S}^*)(\overline{G})) \circ \widetilde{T} \circ \widetilde{T}^{-1} \circ \widetilde{A} \circ \widetilde{S} \\
&= ((\widetilde{T}^*)^{-1} \circ \widetilde{A}^* \circ \widetilde{S}^*)(\overline{G}) \circ \widetilde{T}^{-1} \circ \widetilde{A} \circ \widetilde{S} \\
&= \beta(\overline{G}) \circ \theta(\widetilde{A}).
\end{aligned}$$

As $\beta(\overline{G}) \in D_M$, by virtue of Proposition 6.3.8, the above equality implies that $\theta(\widetilde{A})$ is a lift of some continuous map from N to M. Thus, the proof of Claim 1 is complete. □

Claim 2. $\theta(V_{\widetilde{\alpha}}) \subseteq V_{\widetilde{\alpha}}$.

Proof of Claim 2. We aim to show that if $\widetilde{A} \in V_{\widetilde{\alpha}}$, then $(\theta(\widetilde{A}))^*(\overline{G}) = \widetilde{\alpha}^*(\overline{G})$ for all $\overline{G} \in D_N$. Recall from Proposition 6.3.8 that the map $^* : D_N \to D_M$ is a homomorphism. From this fact and from (6.21), we obtain that

$$\begin{aligned}
(\theta(\widetilde{A}))^*(\overline{G}) &= (\widetilde{T}^{-1} \circ \widetilde{A} \circ \widetilde{S})^*(\overline{G}) \\
&= (\widetilde{T}^{-1})^* \circ \widetilde{A}^* \circ \widetilde{S}^*(\overline{G}) = (\widetilde{T}^{-1})^* \circ \widetilde{\alpha}^* \circ \widetilde{S}^*(\overline{G}) \\
&= (\widetilde{T}^{-1})^* \circ \widetilde{T}^* \circ \widetilde{\alpha}^*(\overline{G}) = (\widetilde{T}^{-1} \circ \widetilde{T})^* \circ \widetilde{\alpha}^*(\overline{G}) = \widetilde{\alpha}^*(\overline{G}).
\end{aligned}$$

This establishes Claim 2. □

Claim 3. *The map $\theta : V_{\widetilde{\alpha}} \to V_{\widetilde{\alpha}}$ is a contraction with respect to the metric $\widetilde{\rho}_\infty$ on $V_{\widetilde{\alpha}}$.*

Proof of Claim 3. Let $\widetilde{A}, \widetilde{B} \in V_{\widetilde{\alpha}}$. Using (6.15), we get

$$\begin{aligned}
\widetilde{\rho}_\infty(\theta(\widetilde{A}), \theta(\widetilde{B})) &= \widetilde{\rho}_\infty(\widetilde{T}^{-1} \circ \widetilde{A} \circ \widetilde{S}, \widetilde{T}^{-1} \circ \widetilde{B} \circ \widetilde{S}) \\
&= \sup\{\widetilde{\rho}_M(\widetilde{T}^{-1} \circ \widetilde{A} \circ \widetilde{S}(\widetilde{x}), \widetilde{T}^{-1} \circ \widetilde{B} \circ \widetilde{S}(\widetilde{x})) : \widetilde{x} \in \widetilde{N}\} \\
&\le \lambda^{-1} \sup\{\widetilde{\rho}_M(\widetilde{A} \circ \widetilde{S}(\widetilde{x}), \widetilde{B} \circ \widetilde{S}(\widetilde{x})) : \widetilde{x} \in \widetilde{N}\} \\
&= \lambda^{-1} \sup\{\widetilde{\rho}_M(\widetilde{A}(\widetilde{y}), \widetilde{B}(\widetilde{y})) : \widetilde{y} \in \widetilde{N}\} \\
&= \lambda^{-1} \widetilde{\rho}_\infty(\widetilde{A}, \widetilde{B}).
\end{aligned}$$

This substantiates Claim 3.

In light of Claim 3 and Lemma 6.3.9, Banach's contraction principle affirms that the map $\theta : V_{\widetilde{\alpha}} \to V_{\widetilde{\alpha}}$ has a unique fixed point $\widetilde{H} \in V_{\widetilde{\alpha}}$. The equality $\theta(\widetilde{H}) = \widetilde{H}$ is equivalent to the equality $\widetilde{T} \circ \widetilde{H} = \widetilde{H} \circ \widetilde{S}$. □

We now demonstrate that homotopic Shub expanding endomorphisms exhibit conjugate dynamics. This generalizes Theorem 6.1.3.

Theorem 6.4.4. *Let M be a compact connected smooth manifold. If $T, S : M \to M$ are two homotopic Shub expanding endomorphisms, then T and S are topologically conjugate.*

Proof. Let $(F_t)_{0\le t\le 1}$ be a homotopy from T to S in M. Thus, $F_0 = T$ while $F_1 = S$. Let $(\widetilde{F}_t)_{0\le t\le 1}$ be a lift of $(F_t)_{0\le t\le 1}$ to $\widetilde{M}$. In particular, $\widetilde{F}_0$ is a lift of T and $\widetilde{F}_1$ is a lift of S. In light of Proposition 6.3.8, we have for every $t \in [0,1]$ that

$$\widetilde{F}_t \circ \overline{G} = \widetilde{F}_t^*(\overline{G}) \circ \widetilde{F}_t, \quad \forall \overline{G} \in D_M. \tag{6.22}$$

Claim. *The function $[0,1] \ni t \mapsto \widetilde{F}_t^*(\overline{G}) \in D_M$ is constant for every $\overline{G} \in D_M$.*

Proof of the claim. Fix $\overline{G} \in D_M$. Let $s \in [0,1]$. Choose any sequence $(s_n)_{n=1}^{\infty}$ in $[0,1]$ converging to s. Fix $\widetilde{x} \in \widetilde{M}$. Let $\widetilde{z} := \widetilde{F}_s(\widetilde{x})$ and $\widetilde{z}_n := \widetilde{F}_{s_n}(\widetilde{x})$ for all $n \in \mathbb{N}$. Then

$$\lim_{n\to\infty} \widetilde{z}_n = \lim_{n\to\infty} \widetilde{F}_{s_n}(\widetilde{x}) = \widetilde{F}_s(\widetilde{x}) = \widetilde{z}$$

and, by (6.22),

$$\lim_{n\to\infty} (\widetilde{F}_{s_n}^*(\overline{G}))(\widetilde{z}_n) = \lim_{n\to\infty} \widetilde{F}_{s_n}^*(\overline{G}) \circ \widetilde{F}_{s_n}(\widetilde{x}) = \lim_{n\to\infty} \widetilde{F}_{s_n} \circ \overline{G}(\widetilde{x}) = \widetilde{F}_s(\overline{G}(\widetilde{x})).$$

Therefore, Lemma 6.3.5 asserts that the sequence $(\widetilde{F}_{s_n}^*(\overline{G}))_{n=1}^{\infty}$ eventually coincides with the unique deck transformation $\overline{\Gamma} \in D_M$ determined by the condition $\overline{\Gamma}(\widetilde{z}) = \widetilde{F}_s(\overline{G}(\widetilde{x}))$. But $\widetilde{z} = \widetilde{F}_s(\widetilde{x})$, so

$$\overline{\Gamma}(\widetilde{F}_s(\widetilde{x})) = \widetilde{F}_s(\overline{G}(\widetilde{x})) = \widetilde{F}_s^*(\overline{G})(\widetilde{F}_s(\widetilde{x})).$$

Hence, $\overline{\Gamma} = \widetilde{F}_s^*(\overline{G})$. In summary, the sequence $(\widetilde{F}_{s_n}^*(\overline{G}))_{n=1}^{\infty}$ is eventually equal to $\widetilde{F}_s^*(\overline{G})$. Since this is true for any convergent sequence $(s_n)_{n=1}^{\infty}$ in $[0,1]$, we conclude that the function $[0,1] \ni t \mapsto \widetilde{F}_t^*(\overline{G}) \in D_M$ is constant. This confirms the claim.

Setting $\widetilde{F}_0 = \widetilde{T}$ and $\widetilde{F}_1 = \widetilde{S}$ and letting $(\widetilde{F}_t)_{0\le t\le 1}$ be a lift of $(F_t)_{0\le t\le 1}$ to $\widetilde{M}$, it follows from the claim that

$$\widetilde{T}^* = \widetilde{S}^*.$$

So we may apply Lemma 6.4.3 with $M = N$, with $\alpha = \mathrm{Id}_M$ and with $\widetilde{\alpha} = \mathrm{Id}_{\widetilde{M}}$, to obtain a unique element $\widetilde{A} \in V_{\mathrm{Id}_{\widetilde{M}}}$ such that

$$\widetilde{T} \circ \widetilde{A} = \widetilde{A} \circ \widetilde{S}. \tag{6.23}$$

By the symmetry between $\widetilde{T}$ and $\widetilde{S}$, there is also an element $\widetilde{B} \in V_{\mathrm{Id}_{\widetilde{M}}}$ such that

$$\widetilde{S} \circ \widetilde{B} = \widetilde{B} \circ \widetilde{T}.$$

Hence,

$$\widetilde{S} \circ (\widetilde{B} \circ \widetilde{A}) = (\widetilde{S} \circ \widetilde{B}) \circ \widetilde{A} = (\widetilde{B} \circ \widetilde{T}) \circ \widetilde{A} = \widetilde{B} \circ (\widetilde{T} \circ \widetilde{A}) = \widetilde{B} \circ (\widetilde{A} \circ \widetilde{S}) = (\widetilde{B} \circ \widetilde{A}) \circ \widetilde{S}.$$

Moreover,

$$\widetilde{S} \circ \mathrm{Id}_{\widetilde{M}} = \mathrm{Id}_{\widetilde{M}} \circ \widetilde{S}.$$

Therefore, the uniqueness part of Lemma 6.4.3, applied with $M = N$, $T = S$, $\alpha = \mathrm{Id}_M$ and $\widetilde{\alpha} = \mathrm{Id}_{\widetilde{M}}$, yields $\widetilde{B} \circ \widetilde{A} = \mathrm{Id}_{\widetilde{M}}$. Likewise, symmetrically, $\widetilde{A} \circ \widetilde{B} = \mathrm{Id}_{\widetilde{M}}$. Let $x \in M$ and choose an arbitrary $\widetilde{x} \in \pi^{-1}(x)$. Given that $\widetilde{A}$ and $\widetilde{B}$, as elements of $V_{\mathrm{Id}_{\widetilde{M}}}$, are lifts of some continuous maps $A : M \to M$ and $B : M \to M$, respectively, we then have that

$$A \circ B(x) = A \circ B \circ \pi(\widetilde{x}) = A \circ \pi \circ \widetilde{B}(\widetilde{x}) = \pi \circ \widetilde{A} \circ \widetilde{B}(\widetilde{x}) = \pi(\widetilde{x}) = x.$$

So, $A \circ B = \mathrm{Id}_M$ and, likewise, $B \circ A = \mathrm{Id}_M$. Thus, A and B are homeomorphisms. Furthermore, due to (6.23), we have that

$$\begin{aligned} T \circ A(x) &= T \circ A \circ \pi(\widetilde{x}) = T \circ \pi \circ \widetilde{A}(\widetilde{x}) = \pi \circ \widetilde{T} \circ \widetilde{A}(\widetilde{x}) = \pi \circ \widetilde{A} \circ \widetilde{S}(\widetilde{x}) \\ &= A \circ \pi \circ \widetilde{S}(\widetilde{x}) = A \circ S \circ \pi(\widetilde{x}) = A \circ S(x). \end{aligned}$$

This means that $T \circ A = A \circ S$ for some homeomorphism $A : M \to M$, that is, T and S are topologically conjugate. □

The crowning statement of this chapter pertains to the structural stability of Shub expanding endomorphisms. Recall that structural stability was defined in Section 1.2.

Theorem 6.4.5. *Every Shub expanding endomorphism of a compact connected smooth manifold M is structurally stable in $\mathcal{E}^1(M)$, the space of all C^1 endomorphisms of M.*

Proof. This is an immediate consequence of Theorems 6.2.5, 6.2.6, and 6.4.4. □

In Chapter 13, we will develop the theory of Gibbs states for open distance expanding systems. In conjunction with the theory of Shub expanding endomorphisms described here, we will derive in Section 13.7 the following theorem, which was first proved for C^2 maps by Krzyżewski and Szlenk [42]. It is also in this paper that the appropriate transfer (also called Ruelle or Perron–Frobenius) operator was for the first time explicitly used in dynamical systems. Our proof will be different, based on the theory of Gibbs states developed in Chapter 13; nevertheless, there will be significant similarities with that of Krzyżewski and Szlenk.

Each Riemannian metric ρ on a compact connected smooth manifold M induces a unique volume (Lebesgue) measure λ_ρ on M and the volume measures induced by various Riemannian metrics are mutually equivalent. Call this class of measures the Lebesgue measure class on M.

Theorem 6.4.6. *If $T : M \to M$ is a $C^{1+\varepsilon}$ Shub expanding endomorphism on a compact connected smooth manifold M, then there exists a unique T-invariant Borel probability measure μ on M which is absolutely continuous with respect to the Lebesgue measure class on M. In fact, μ is equivalent to the Lebesgue measure class on M, and μ is ergodic.*

6.5 Exercises

Exercise 6.5.1. Let $T : \mathbb{S}^1 \to \mathbb{S}^1$ be a Shub expanding map. Prove that $|\deg(T)| \geq 2$.

Exercise 6.5.2. Let M be a compact connected smooth manifold and $f, g : M \to M$ be Shub expanding endomorphisms with respect to some Riemannian metric ρ on M. Show that $f \circ g$ is also a Shub expanding endomorphism with respect to ρ.

Exercise 6.5.3. Suppose that $f_1, f_2 : M \to M$ are Shub expanding endomorphisms with respect to Riemannian metrics ρ_1 and ρ_2, respectively. Is there always a Riemannian metric ρ such that $f_1 \circ f_2$ is expanding with respect to ρ? You may assume that $M = \mathbb{S}^1$.

Exercise 6.5.4. Prove that the Cartesian product of finitely many Shub expanding endomorphisms is a Shub expanding endomorphism if the product manifold is endowed with the standard L^1 product metric

$$\langle v, w \rangle_x := \sum_{k=1}^{n} \langle v_k, w_k \rangle_{x_k}.$$

Exercise 6.5.5. Suppose that V is an open neighborhood of the origin in a Euclidean space $\mathbb{R}^k$ and that $G : V \to V$ is a diffeomorphism. Let $R > 0$ be such that $\overline{B}(0, R) \subseteq V$. Show that the map $\widehat{G} : \mathbb{R}^k \to \mathbb{R}^k$, where

$$\widehat{G}(x) = \begin{cases} G(x) & \text{if } x \in \overline{B}(0, R) \\ G(\frac{Rx}{\|x\|}) + [G'(\frac{Rx}{\|x\|})](x - \frac{Rx}{\|x\|}) & \text{if } x \notin \overline{B}(0, R) \end{cases}$$

is a diffeomorphic extension of $G|_{\overline{B}(0,R)}$.

7 Topological entropy

In this chapter, we study the notion of topological entropy, one of the most useful and widely-applicable topological invariant thus far discovered. It was introduced to dynamical systems by Adler, Konheim, and McAndrew [2] in 1965. Their definition was motivated by Kolmogorov and Sinai's definition of metric/measure-theoretic entropy introduced in [67] less than a decade earlier. In this book, we do not follow the historical order of discovery of these notions. It is more suitable to present topological entropy first.

Metric and topological entropies not only have related origins and similar names. There are truly significant mathematical relations between them, particularly the one given by the variational principle, which is treated at length in Chapter 12.

The topological entropy of a dynamical system $T : X \to X$, which we introduce in Section 7.2 and shall be denoted by $\mathrm{h}_{\mathrm{top}}(T)$, is a nonnegative extended real number that measures the complexity of the system. Somewhat more precisely, $\mathrm{h}_{\mathrm{top}}(T)$ is the exponential growth rate of the number of orbits separated under T. The topological entropy of a dynamical system is defined in three stages. First, we define the entropy of a cover of the underlying space. Second, we define the entropy of the system with respect to any given cover. Third, the entropy of the system is defined to be the supremum, over all covers, of the entropy of the system with respect to each of those.

Recall from Chapter 1 that a mathematical property is said to be a topological invariant for the category of topological dynamical systems if it is shared by any pair of topologically conjugate systems. For topological entropy, being an invariant means that if $T : X \to X$ and $S : Y \to Y$ are two topologically conjugate dynamical systems, then $\mathrm{h}_{\mathrm{top}}(T) = \mathrm{h}_{\mathrm{top}}(S)$. However, the converse is generally not true. That is, if $T : X \to X$ and $S : Y \to Y$ are two dynamical systems with equal topological entropy, then T and S may not be topologically conjugate. You are asked to provide such an example in Exercise 7.6.15. Therefore, topological entropy is not a complete invariant.

In Section 7.3, we treat at length Bowen's characterization of topological entropy in terms of separated and spanning sets.

In Chapter 11, we will introduce and deal with topological pressure, which is a substantial generalization of topological entropy. Our approach to topological pressure will stem from and extend that for topological entropy. In this sense, the present chapter can be viewed as a preparation to Chapter 11.

7.1 Covers of a set

Definition 7.1.1. Let X be a nonempty set. A family $\mathcal{U}$ of subsets of X is said to form a *cover* of X if

$$X \subseteq \bigcup_{U \in \mathcal{U}} U.$$

Furthermore, $\mathcal{V}$ is said to be a *subcover* of $\mathcal{U}$ if $\mathcal{V}$ is itself a cover and $\mathcal{V} \subseteq \mathcal{U}$.

https://doi.org/10.1515/9783110702682-007

We will always denote covers by calligraphic letters, $\mathcal{U}, \mathcal{V}, \mathcal{W}$, and so on.

Let us begin by introducing a useful way of obtaining a new cover from two existing covers.

Definition 7.1.2. If $\mathcal{U}$ and $\mathcal{V}$ are covers of X, then their *join*, denoted $\mathcal{U} \vee \mathcal{V}$, is the cover

$$\mathcal{U} \vee \mathcal{V} := \{U \cap V : U \in \mathcal{U}, V \in \mathcal{V}\}.$$

Remark 7.1.3. The join operation is commutative (i. e., $\mathcal{U} \vee \mathcal{V} = \mathcal{V} \vee \mathcal{U}$) and associative (in other words, $(\mathcal{U} \vee \mathcal{V}) \vee \mathcal{W} = \mathcal{U} \vee (\mathcal{V} \vee \mathcal{W})$). Thanks to this associativity, the join operation extends naturally to any finite collection $\{\mathcal{U}_j\}_{j=0}^{n-1}$ of covers of X:

$$\bigvee_{j=0}^{n-1} \mathcal{U}_j := \mathcal{U}_0 \vee \cdots \vee \mathcal{U}_{n-1} = \left\{ \bigcap_{j=0}^{n-1} U_j : U_j \in \mathcal{U}_j, \forall 0 \le j \le n-1 \right\}.$$

It is also useful to be able to compare covers. For this purpose, we introduce the following relation on the collection of all covers of a set.

Definition 7.1.4. Let $\mathcal{U}$ and $\mathcal{V}$ be covers of a set X. We say that $\mathcal{V}$ is *finer* than, or a *refinement* of, $\mathcal{U}$, and denote this by $\mathcal{U} \prec \mathcal{V}$, if every element of $\mathcal{V}$ is a subset of an element of $\mathcal{U}$. That is, for every set $V \in \mathcal{V}$ there exists a set $U \in \mathcal{U}$ such that $V \subseteq U$. It is also sometimes said that $\mathcal{V}$ is *inscribed* in $\mathcal{U}$, or that $\mathcal{U}$ is coarser than $\mathcal{V}$.

Lemma 7.1.5. *Let $\mathcal{U}$, $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{X}$ be covers of a set X. Then:*

(a) *The refinement relation $\prec$ is reflexive (i. e., $\mathcal{U} \prec \mathcal{U}$) and transitive (i. e., if $\mathcal{U} \prec \mathcal{V}$ and $\mathcal{V} \prec \mathcal{W}$, then $\mathcal{U} \prec \mathcal{W}$).*

(b) $\mathcal{U} \prec \mathcal{U} \vee \mathcal{V}$.

(c) *If $\mathcal{V}$ is a subcover of $\mathcal{U}$, then $\mathcal{U} \prec \mathcal{V}$.*

(d) *$\mathcal{U}$ is a subcover of $\mathcal{U} \vee \mathcal{U}$. Hence, from (c) and (b), we deduce that*

$$\mathcal{U} \prec \mathcal{U} \vee \mathcal{U} \prec \mathcal{U}.$$

Nevertheless, $\mathcal{U}$ is not equal to $\mathcal{U} \vee \mathcal{U}$ in general.

(e) *If $\mathcal{U} \prec \mathcal{V}$ <u>or</u> $\mathcal{U} \prec \mathcal{W}$, then $\mathcal{U} \prec \mathcal{V} \vee \mathcal{W}$.*

(f) *If $\mathcal{U} \prec \mathcal{W}$ <u>and</u> $\mathcal{V} \prec \mathcal{W}$, then $\mathcal{U} \vee \mathcal{V} \prec \mathcal{W}$.*

(g) *If $\mathcal{U} \prec \mathcal{W}$ <u>and</u> $\mathcal{V} \prec \mathcal{X}$, then $\mathcal{U} \vee \mathcal{V} \prec \mathcal{W} \vee \mathcal{X}$.*

Proof. All of these properties can be proved directly and are left to the reader. As a hint, observe that property (e) is a consequence of (b) and the transitivity part of (a), while property (g) follows upon combining (e) and (f). □

Remark 7.1.6. Although it is reflexive and transitive, the relation $\prec$ is not antisymmetric (i. e., $\mathcal{U} \prec \mathcal{V} \prec \mathcal{U}$ does not necessarily imply $\mathcal{U} = \mathcal{V}$; see Lemma 7.1.5(d)). Therefore, $\prec$ does not generate a partial order on the collection of all covers of a set X.

If X is a metric space, then it makes sense to talk about the diameter of a cover in terms of the diameter of its elements. This is the purpose of the next definition.

Definition 7.1.7. If (X, d) is a metric space, then the *diameter* of a cover $\mathcal{U}$ of X is defined by

$$\operatorname{diam}(\mathcal{U}) := \sup\{\operatorname{diam}(U) : U \in \mathcal{U}\},$$

where

$$\operatorname{diam}(U) := \sup\{d(x, y) : x, y \in U\}.$$

It is also often of interest to know that all sets of some specified diameter are each contained in at least one element of a given cover. This is made precise in the following definition.

Definition 7.1.8. A number $\delta > 0$ is said to be a *Lebesgue number* for a cover $\mathcal{U}$ of a metric space (X, d) if every set of diameter at most δ is contained in an element of $\mathcal{U}$.

It is clear that if δ_0 is a Lebesgue number for a cover $\mathcal{U}$, then so is any δ with $0 < \delta < \delta_0$. One can easily prove by contradiction that every open cover of a compact metric space admits such a number. By an open cover, we simply mean a cover whose elements are all open subsets of the space.

7.1.1 Dynamical covers

In this subsection, we add a dynamical aspect to the above discussion. Let X be a nonempty set and let $T : X \to X$ be a map. We will define covers that are induced by the dynamics of the map T. First, let us define the preimage of a cover under a map.

Definition 7.1.9. Let X and Y be nonempty sets. Let $h : X \to Y$ be a map and $\mathcal{V}$ be a cover of Y. The *preimage* of $\mathcal{V}$ under the map h is the cover of X consisting of all the preimages of the elements of $\mathcal{V}$ under h, that is,

$$h^{-1}(\mathcal{V}) := \{h^{-1}(V) : V \in \mathcal{V}\}.$$

We now show that, as far as set operations go, the operator h^{-1} behaves well with respect to cover operations.

Lemma 7.1.10. *Let $h : X \to Y$ be a map, and $\mathcal{U}$ and $\mathcal{V}$ be covers of Y. The following assertions hold:*

(a) *The map h^{-1} preserves the refinement relation, that is,*

$$\mathcal{U} \prec \mathcal{V} \implies h^{-1}(\mathcal{U}) \prec h^{-1}(\mathcal{V}).$$

Moreover, if $\mathcal{V}$ is a subcover of $\mathcal{U}$ then $h^{-1}(\mathcal{V})$ is a subcover of $h^{-1}(\mathcal{U})$.

(b) *The map h^{-1} respects the join operation, that is,*

$$h^{-1}(\mathcal{U} \vee \mathcal{V}) = h^{-1}(\mathcal{U}) \vee h^{-1}(\mathcal{V}).$$

Note that if $Y = X$, then by induction h^{-n} enjoys these properties for any $n \in \mathbb{N}$.

Proof. These assertions are straightforward to prove and are thus left to the reader. □

We now introduce covers that follow the orbits of a given map by indicating to which elements of a given cover the successive iterates of the map belong.

Definition 7.1.11. Let $T : X \to X$ be a map and $\mathcal{U}$ be a cover of X. For every $n \in \mathbb{N}$ and $0 \le m < n$, define the *dynamical cover*

$$\mathcal{U}_m^n := \bigvee_{j=m}^{n-1} T^{-j}(\mathcal{U}) = T^{-m}(\mathcal{U}) \vee T^{-(m+1)}(\mathcal{U}) \vee \cdots \vee T^{-(n-1)}(\mathcal{U}).$$

To lighten notation, we will write $\mathcal{U}^n$ in lieu of $\mathcal{U}_0^n$.

A typical element of $\mathcal{U}^n$ is of the form $U_0 \cap T^{-1}(U_1) \cap T^{-2}(U_2) \cap \ldots \cap T^{-(n-1)}(U_{n-1})$ for some $U_0, U_1, U_2, \ldots, U_{n-1} \in \mathcal{U}$. This element is the set of all points of X whose iterates under T fall successively into the elements U_0, U_1, U_2, and so on, up to U_{n-1}.

Lemma 7.1.12. *Let $\mathcal{U}$ and $\mathcal{V}$ be covers of a set X. Let $T : X \to X$ be a map. For every $k, m, n \in \mathbb{N}$, the following statements hold:*
(a) *If $\mathcal{U} \prec \mathcal{V}$, then $\mathcal{U}^n \prec \mathcal{V}^n$.*
(b) $(\mathcal{U} \vee \mathcal{V})^n = \mathcal{U}^n \vee \mathcal{V}^n$.
(c) $\mathcal{U}^n \prec \mathcal{U}^{n+1}$.
(d) $(\mathcal{U}^k)^n \prec \mathcal{U}^{n+k-1} \prec (\mathcal{U}^k)^n$.
(e) $\mathcal{U}_m^n = T^{-k}(\mathcal{U}_{m-k}^{n-k})$ *for all $k \le m < n$. In particular, $\mathcal{U}_m^n = T^{-m}(\mathcal{U}^{n-m})$.*

Proof. Property (a) follows directly from Lemmas 7.1.10(a) and 7.1.5(g). Property (b) is a consequence of Lemma 7.1.10(b) and the associativity of the join operation. As $\mathcal{U}^{n+1} = \mathcal{U}^n \vee T^{-n}(\mathcal{U})$, property (c) follows from an application of Lemma 7.1.5(b). Property (d) is a little more intricate to prove. Using Lemma 7.1.10(b) and Remark 7.1.3, we obtain that

$$\begin{aligned}
(\mathcal{U}^k)^n &= \mathcal{U}^k \vee T^{-1}(\mathcal{U}^k) \vee \cdots \vee T^{-(n-1)}(\mathcal{U}^k) \\
&= (\mathcal{U} \vee \cdots \vee T^{-(k-1)}(\mathcal{U})) \vee T^{-1}(\mathcal{U} \vee \cdots \vee T^{-(k-1)}(\mathcal{U})) \vee \cdots \\
&\quad \cdots \vee T^{-(n-1)}(\mathcal{U} \vee \cdots \vee T^{-(k-1)}(\mathcal{U})) \\
&= \mathcal{U} \vee (T^{-1}(\mathcal{U}) \vee T^{-1}(\mathcal{U})) \vee (T^{-2}(\mathcal{U}) \vee T^{-2}(\mathcal{U}) \vee T^{-2}(\mathcal{U})) \vee \cdots \\
&\quad \cdots \vee (T^{-(n+k-3)}(\mathcal{U}) \vee T^{-(n+k-3)}(\mathcal{U})) \vee T^{-(n+k-2)}(\mathcal{U}).
\end{aligned}$$

Now, according to Lemma 7.1.5(d), $T^{-j}(\mathcal{U}) \prec T^{-j}(\mathcal{U}) \vee T^{-j}(\mathcal{U}) \prec T^{-j}(\mathcal{U})$ for all $j \in \mathbb{N}$. We deduce from a repeated application of Lemma 7.1.5(g) that

$$T^{-j}(\mathcal{U}) \prec \bigvee_{l=1}^{m} T^{-j}(\mathcal{U}) \prec T^{-j}(\mathcal{U}), \quad \forall m \in \mathbb{N}.$$

Another round of repeated applications of Lemma 7.1.5(g) allows us to conclude that

$$(\mathcal{U}^k)^n \prec \mathcal{U} \vee T^{-1}(\mathcal{U}) \vee T^{-2}(\mathcal{U}) \vee \cdots \vee T^{-(n+k-3)}(\mathcal{U}) \vee T^{-(n+k-2)}(\mathcal{U}) \prec (\mathcal{U}^k)^n.$$

That is,

$$(\mathcal{U}^k)^n \prec \mathcal{U}^{n+k-1} \prec (\mathcal{U}^k)^n.$$

Finally, property (e) follows from Lemma 7.1.10(b) with $h = T^k$ since

$$\mathcal{U}_m^n = \bigvee_{j=m}^{n-1} T^{-j}(\mathcal{U}) = T^{-k}\left(\bigvee_{j=m-k}^{n-k-1} T^{-j}(\mathcal{U})\right) = T^{-k}(\mathcal{U}_{m-k}^{n-k}).$$ □

7.2 Definition of topological entropy via open covers

The definition of topological entropy via open covers only requires the underlying space to be a topological space. It need not be a metrizable space. The topological entropy of a dynamical system $T : X \to X$ is defined in three stages, which, for clarity of exposition, we split into the following three subsections.

7.2.1 First stage: entropy of an open cover

At this stage, the dynamics of the system T are not in consideration. We simply look at the difficulty of covering the underlying compact space X with open covers.

Definition 7.2.1. Let $\mathcal{U}$ be an open cover of X. Define

$$Z_1(\mathcal{U}) := \min\{\#\mathcal{V} : \mathcal{V} \text{ is a subcover of } \mathcal{U}\}.$$

That is, $Z_1(\mathcal{U})$ denotes the minimum number of elements of $\mathcal{U}$ necessary to cover X. A subcover of $\mathcal{U}$ whose cardinality equals this minimum number is called a *minimal subcover* of $\mathcal{U}$.

Every open cover admits at least one minimal subcover and any such subcover is finite since X is compact. Thus $1 \le Z_1(\mathcal{U}) < \infty$ for all open covers $\mathcal{U}$ of X.

We now observe that the function $Z_1(\cdot)$ acts as desired with respect to the refinement relation. In other words, the finer the cover, the larger the minimum number of elements required to cover the space, that is, the more difficult it is to cover the space.

Lemma 7.2.2. *If $\mathcal{U} \prec \mathcal{V}$, then $Z_1(\mathcal{U}) \le Z_1(\mathcal{V})$. In particular, this holds if $\mathcal{V}$ is a subcover of $\mathcal{U}$.*

Proof. Let $\mathcal{U} \prec \mathcal{V}$. For every $V \in \mathcal{V}$, there exists a set $i(V) \in \mathcal{U}$ such that $V \subseteq i(V)$. This defines a function $i : \mathcal{V} \to \mathcal{U}$. Let $\mathcal{W}$ be a minimal subcover of $\mathcal{V}$. Then $i(\mathcal{W}) := \{i(W) : W \in \mathcal{W}\} \subseteq \mathcal{U}$ is a cover of X since

$$X \subseteq \bigcup_{W \in \mathcal{W}} W \subseteq \bigcup_{W \in \mathcal{W}} i(W).$$

Thus $i(\mathcal{W})$ is a subcover of $\mathcal{U}$, and hence

$$Z_1(\mathcal{U}) \le \# i(\mathcal{W}) \le \# \mathcal{W} = Z_1(\mathcal{V}).$$ □

Another fundamental property of the function $Z_1(\cdot)$ is that it is submultiplicative with respect to the join operation. Recall that a sequence $(a_n)_{n=1}^{\infty}$ of real numbers is said to be submultiplicative if

$$a_{m+n} \le a_m a_n, \quad \forall m, n \in \mathbb{N}.$$

Lemma 7.2.3. *Let $\mathcal{U}$ and $\mathcal{V}$ be open covers of X. Then*

$$Z_1(\mathcal{U} \vee \mathcal{V}) \le Z_1(\mathcal{U}) \cdot Z_1(\mathcal{V}).$$

Proof. Let $\underline{\mathcal{U}}$ be a minimal subcover of $\mathcal{U}$ and $\underline{\mathcal{V}}$ be a minimal subcover of $\mathcal{V}$. Then $\underline{\mathcal{U}} \vee \underline{\mathcal{V}}$ is a subcover of $\mathcal{U} \vee \mathcal{V}$. Therefore,

$$Z_1(\mathcal{U} \vee \mathcal{V}) \le \#(\underline{\mathcal{U}} \vee \underline{\mathcal{V}}) \le \#\underline{\mathcal{U}} \cdot \#\underline{\mathcal{V}} = Z_1(\mathcal{U}) \cdot Z_1(\mathcal{V}).$$ □

We can now define the entropy of a cover.

Definition 7.2.4. Let $\mathcal{U}$ be an open cover of X. The *entropy* of $\mathcal{U}$ is defined to be

$$H(\mathcal{U}) := \log Z_1(\mathcal{U}).$$

So, the entropy of an open cover is simply the logarithm of the minimum number of elements of that cover needed to cover the space. The presence of the logarithm function shall be explained shortly. If the entropy of a given cover is to accurately reflect the complexity of that cover, that is, the number of elements necessary for covering the space, then the finer the cover, the larger its entropy should be. In other words, entropy of covers should be increasing with respect to the refinement relation. This, along with other basic properties of the entropy of covers, is shown to hold in the following lemma.

Lemma 7.2.5. *Let $\mathcal{U}$ and $\mathcal{V}$ be open covers of X. Entropy of covers satisfies the following properties:*

(a) $0 \le H(\mathcal{U}) < \infty$.
(b) $H(\mathcal{U}) = 0$ *if and only if* $X \in \mathcal{U}$.
(c) *If* $\mathcal{U} \prec \mathcal{V}$, *then* $H(\mathcal{U}) \le H(\mathcal{V})$.
(d) $H(\mathcal{U} \vee \mathcal{V}) \le H(\mathcal{U}) + H(\mathcal{V})$.

Proof. The first two properties come directly from entropy's definition. The third follows from Lemma 7.2.2 and the fact that the logarithm is an increasing function. The fourth and final property follows from Lemma 7.2.3. □

7.2.2 Second stage: entropy of a system relative to an open cover

In this second stage, we will take into account the dynamics of the topological dynamical system $T : X \to X$. Since T is continuous, every open cover $\mathcal{U}$ of X generates the sequence of dynamical covers $(\mathcal{U}^n)_{n=1}^{\infty}$, all of which are also open.

Definition 7.2.6. Let $\mathcal{U}$ be an open cover of X. For every $n \in \mathbb{N}$, define $Z_n(\mathcal{U})$ to be

$$Z_n(\mathcal{U}) := Z_1(\mathcal{U}^n) = \min\{\#\mathcal{V} : \mathcal{V} \text{ is a subcover of } \mathcal{U}^n\}.$$

Thus $Z_n(\mathcal{U})$ is the minimum number of elements of $\mathcal{U}^n$ needed to cover X. This number describes the complexity of the dynamics of T with respect to $\mathcal{U}$ from time 0 until time $n-1$. Observe also that

$$Z_n(\mathcal{U}) = \exp\big(H(\mathcal{U}^n)\big).$$

For a given open cover, the sequence $(Z_n(\cdot))_{n=1}^{\infty}$ has an interesting property.

Lemma 7.2.7. *For any open cover* $\mathcal{U}$ *of* X, *the sequence* $(Z_n(\mathcal{U}))_{n=1}^{\infty}$ *is nondecreasing.*

Proof. Since $\mathcal{U}^n \prec \mathcal{U}^{n+1}$ for all $n \in \mathbb{N}$ according to Lemma 7.1.12(c), the sequence $(Z_n(\mathcal{U}))_{n=1}^{\infty} = (Z_1(\mathcal{U}^n))_{n=1}^{\infty}$ is nondecreasing by Lemma 7.2.2. □

As the next lemma shows, like the function $Z_1(\cdot)$, the functions $Z_n(\cdot)$ respect the refinement relation.

Lemma 7.2.8. *If* $\mathcal{U} \prec \mathcal{V}$, *then* $Z_n(\mathcal{U}) \le Z_n(\mathcal{V})$, *and thus* $H(\mathcal{U}^n) \le H(\mathcal{V}^n)$ *for every* $n \in \mathbb{N}$. *In particular, these inequalities hold if* $\mathcal{V}$ *is a subcover of* $\mathcal{U}$.

Proof. If $\mathcal{U} \prec \mathcal{V}$, then Lemma 7.1.12(a) states that $\mathcal{U}^n \prec \mathcal{V}^n$ for every $n \in \mathbb{N}$. It follows from Lemma 7.2.2 that

$$Z_n(\mathcal{U}) = Z_1\big(\mathcal{U}^n\big) \le Z_1\big(\mathcal{V}^n\big) = Z_n(\mathcal{V}).$$

Since the logarithm is an increasing function, it ensues that

$$H\big(\mathcal{U}^n\big) = \log Z_n(\mathcal{U}) \le \log Z_n(\mathcal{V}) = H\big(\mathcal{V}^n\big).$$ □

Similar to the function $Z_1(\cdot)$, the functions $Z_n(\cdot)$ are submultiplicative with respect to the join operation.

Lemma 7.2.9. *Let $\mathcal{U}$ and $\mathcal{V}$ be open covers of X and let $n \in \mathbb{N}$. Then*

$$Z_n(\mathcal{U} \vee \mathcal{V}) \le Z_n(\mathcal{U})Z_n(\mathcal{V}),$$

and thus

$$H\big((\mathcal{U} \vee \mathcal{V})^n\big) \le H(\mathcal{U}^n) + H(\mathcal{V}^n).$$

Proof. Using Lemmas 7.1.12(b) and 7.2.3, we obtain that

$$\begin{aligned} Z_n(\mathcal{U} \vee \mathcal{V}) &= Z_1\big((\mathcal{U} \vee \mathcal{V})^n\big) = Z_1(\mathcal{U}^n \vee \mathcal{V}^n) \\ &\le Z_1(\mathcal{U}^n)Z_1(\mathcal{V}^n) = Z_n(\mathcal{U})Z_n(\mathcal{V}). \end{aligned}$$

Taking the logarithm of both sides gives $H((\mathcal{U} \vee \mathcal{V})^n) \le H(\mathcal{U}^n) + H(\mathcal{V}^n)$. □

We now refocus our attention on the sequence $(Z_n(\mathcal{U}))_{n=1}^{\infty}$ for a given open cover $\mathcal{U}$ of X. We have already established in Lemma 7.2.7 that this sequence is nondecreasing. Lemma 7.2.3 suggests that this sequence might be submultiplicative and might even grow exponentially with n. This explains the use of the logarithm function. By working in a logarithmic scale, we study the exponential growth rate of the numbers $(Z_n(\mathcal{U}))_{n=1}^{\infty}$. This will further ensure that the entropy of the system with respect to any specific open cover is finite.

Lemma 7.2.10. *For any open cover $\mathcal{U}$ of X, the sequence $(Z_n(\mathcal{U}))_{n=1}^{\infty}$ is submultiplicative.*

Proof. Let $m, n \in \mathbb{N}$. Choose a minimal subcover $\mathcal{A}$ of $\mathcal{U}^m$ and a minimal subcover $\mathcal{B}$ of $\mathcal{U}^n$. Using Lemma 7.1.10, we obtain that the open cover $\mathcal{A} \vee T^{-m}(\mathcal{B})$ satisfies

$$\begin{aligned} \mathcal{A} \vee T^{-m}(\mathcal{B}) &\subseteq \mathcal{U}^m \vee T^{-m}(\mathcal{U}^n) \\ &= \big(\mathcal{U} \vee \cdots \vee T^{-(m-1)}(\mathcal{U})\big) \vee T^{-m}\big(\mathcal{U} \vee \cdots \vee T^{-(n-1)}(\mathcal{U})\big) \\ &= \mathcal{U} \vee \cdots \vee T^{-(m-1)}(\mathcal{U}) \vee T^{-m}(\mathcal{U}) \vee \cdots \vee T^{-(m+n-1)}(\mathcal{U}) \\ &= \mathcal{U}^{m+n}. \end{aligned}$$

That is, $\mathcal{A} \vee T^{-m}(\mathcal{B})$ is a subcover of $\mathcal{U}^{m+n}$. Consequently,

$$Z_{m+n}(\mathcal{U}) \le \#\big(\mathcal{A} \vee T^{-m}(\mathcal{B})\big) \le \#\mathcal{A} \cdot \#\big(T^{-m}(\mathcal{B})\big) \le \#\mathcal{A} \cdot \#\mathcal{B} = Z_m(\mathcal{U})Z_n(\mathcal{U}).$$

This establishes the submultiplicativity of the sequence $(Z_n(\mathcal{U}))_{n=1}^{\infty}$. □

We immediately deduce the following.

Corollary 7.2.11. *For any open cover $\mathcal{U}$ of X, the sequence $(H(\mathcal{U}^n))_{n=1}^{\infty}$ is subadditive.*

Proof. Since $Z_{m+n}(\mathcal{U}) \le Z_m(\mathcal{U})Z_n(\mathcal{U})$ for all $m, n \in \mathbb{N}$ according to Lemma 7.2.10, taking the logarithm of both sides yields that

$$H(\mathcal{U}^{m+n}) = \log Z_{m+n}(\mathcal{U}) \le \log Z_m(\mathcal{U}) + \log Z_n(\mathcal{U}) = H(\mathcal{U}^m) + H(\mathcal{U}^n). \qquad \square$$

We are now ready to take the second step in the definition of the topological entropy of a system.

Definition 7.2.12. Let $T : X \to X$ be a dynamical system and let $\mathcal{U}$ be an open cover of X. The *topological entropy of T with respect to $\mathcal{U}$* is defined as

$$\mathrm{h_{top}}(T, \mathcal{U}) := \lim_{n\to\infty} \frac{1}{n} H(\mathcal{U}^n) = \inf_{n\in\mathbb{N}} \frac{1}{n} H(\mathcal{U}^n).$$

The existence of the limit and its equality with the infimum follow directly from combining Corollary 7.2.11 and Lemma 3.2.17. Furthermore, note that

$$\mathrm{h_{top}}(T, \mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(\mathcal{U}) = \inf_{n\in\mathbb{N}} \frac{1}{n} \log Z_n(\mathcal{U}).$$

Remark 7.2.13. Since $1 \le Z_n(\mathcal{U}) < \infty$ for all $n \in \mathbb{N}$, we readily see that

$$0 \le \mathrm{h_{top}}(T, \mathcal{U}) \le H(\mathcal{U}) < \infty.$$

Similar to the functions $Z_n(\cdot)$, the topological entropy with respect to covers respects the refinement relation. It is also subadditive with respect to the join operation, as the following proposition shows.

Proposition 7.2.14. *Let $\mathcal{U}$ and $\mathcal{V}$ be open covers of X.*
(a) *If $\mathcal{U} \prec \mathcal{V}$, then $\mathrm{h_{top}}(T, \mathcal{U}) \le \mathrm{h_{top}}(T, \mathcal{V})$. In particular, if $\mathcal{V}$ is a subcover of $\mathcal{U}$ then $\mathrm{h_{top}}(T, \mathcal{U}) \le \mathrm{h_{top}}(T, \mathcal{V})$.*
(b) $\mathrm{h_{top}}(T, \mathcal{U} \vee \mathcal{V}) \le \mathrm{h_{top}}(T, \mathcal{U}) + \mathrm{h_{top}}(T, \mathcal{V})$.

Proof. (a) By Lemma 7.2.8, we have $Z_n(\mathcal{U}) \le Z_n(\mathcal{V})$ for every $n \in \mathbb{N}$. Taking the logarithm of both sides, dividing by n and passing to the limit as n tends to infinity, we deduce that $\mathrm{h_{top}}(T, \mathcal{U}) \le \mathrm{h_{top}}(T, \mathcal{V})$ whenever $\mathcal{U} \prec \mathcal{V}$.

(b) By Lemma 7.2.9, we have $Z_n(\mathcal{U} \vee \mathcal{V}) \le Z_n(\mathcal{U})Z_n(\mathcal{V})$ for every $n \in \mathbb{N}$. Taking the logarithm of both sides, dividing by n and passing to the limit as n tends to infinity, we conclude that $\mathrm{h_{top}}(T, \mathcal{U} \vee \mathcal{V}) \le \mathrm{h_{top}}(T, \mathcal{U}) + \mathrm{h_{top}}(T, \mathcal{V})$. $\square$

An interesting property of the entropy of a system with respect to a given cover is that it remains the same for all dynamical covers generated by that cover.

Lemma 7.2.15. $\mathrm{h_{top}}(T, \mathcal{U}^k) = \mathrm{h_{top}}(T, \mathcal{U})$ *for each $k \in \mathbb{N}$.*

Proof. The case $k = 1$ is trivial. So suppose that $k \geq 2$. Let $n \in \mathbb{N}$. Lemma 7.1.12(d) asserts that $(\mathcal{U}^k)^n \prec \mathcal{U}^{n+k-1} \prec (\mathcal{U}^k)^n$. From Lemma 7.2.2, we infer that $Z_1((\mathcal{U}^k)^n) = Z_1(\mathcal{U}^{n+k-1})$, and thus

$$Z_n(\mathcal{U}^k) = Z_1((\mathcal{U}^k)^n) = Z_1(\mathcal{U}^{n+k-1}) = Z_{n+k-1}(\mathcal{U}).$$

Therefore,

$$\begin{aligned} \mathrm{h_{top}}(T, \mathcal{U}^k) &= \lim_{n\to\infty} \frac{1}{n} \log Z_n(\mathcal{U}^k) = \lim_{n\to\infty} \frac{n+k-1}{n(n+k-1)} \log Z_{n+k-1}(\mathcal{U}) \\ &= \lim_{n\to\infty} \frac{n+k-1}{n} \cdot \lim_{n\to\infty} \frac{1}{n+k-1} \log Z_{n+k-1}(\mathcal{U}) = \mathrm{h_{top}}(T, \mathcal{U}). \end{aligned}$$ □

7.2.3 Third and final stage: entropy of a system

At this point, we are in a position to give the definition of the topological entropy of a dynamical system $T : X \to X$. The topological entropy of T is defined to be the supremum over all open covers, of the entropy of the system with respect to each of these covers.

Definition 7.2.16. The *topological entropy* of T is defined to be

$$\mathrm{h_{top}}(T) := \sup\{\mathrm{h_{top}}(T, \mathcal{U}) : \mathcal{U} \text{ is an open cover of } X\}.$$

Remark 7.2.17.
(a) In view of Remark 7.2.13, we have that $0 \leq \mathrm{h_{top}}(T) \leq \infty$.
(b) The topological entropy of the identity map $\mathrm{Id}(x) = x$ is zero. Indeed, for any open cover $\mathcal{U}$ of X we have that $\mathcal{U}^n = \mathcal{U}$, and hence $Z_n(\mathcal{U}) = Z_1(\mathcal{U})$, for every $n \in \mathbb{N}$. Thus $\mathrm{h_{top}}(\mathrm{Id}, \mathcal{U}) = 0$ for all open covers $\mathcal{U}$ of X, and thereby $\mathrm{h_{top}}(\mathrm{Id}) = 0$.
(c) Despite the fact that $\mathrm{h_{top}}(T, \mathcal{U}) < \infty$ for every open cover $\mathcal{U}$ of X, there exist dynamical systems T that have infinite topological entropy.
(d) As every open cover of a compact space admits a finite subcover, it follows from Proposition 7.2.14(a) that the supremum in the definition of topological entropy can be restricted to finite open covers.

Our next aim is to address the most important and natural question: Is topological entropy a topological conjugacy invariant? Before answering this question, the reader might like to recall from Chapter 1 that if $T : X \to X$ and $S : Y \to Y$ are two dynamical systems for which there exists a continuous surjection $h : X \to Y$ such that $h \circ T = S \circ h$, then S is said to be a *factor* of T. In such a situation, it is intuitively clear that $\mathrm{h_{top}}(S) \leq \mathrm{h_{top}}(T)$ since every orbit of T is projected onto an orbit of S. Thus T may have "more" orbits (in some sense) than S and is therefore at least as complex as S.

Proposition 7.2.18. *If $T : X \to X$ and $S : Y \to Y$ are two dynamical systems such that S is a factor of T, then*

$$\mathrm{h_{top}}(S) \le \mathrm{h_{top}}(T).$$

In particular, if S and T are topologically conjugate, then $\mathrm{h_{top}}(S) = \mathrm{h_{top}}(T)$.

Proof. Let $h : X \to Y$ be a factor map, so that $h \circ T = S \circ h$. Since h is a continuous surjection, every open cover $\mathcal{V}$ of Y can be lifted to the open cover $h^{-1}(\mathcal{V})$ of X. We shall prove that $\mathrm{h_{top}}(T, h^{-1}(\mathcal{V})) = \mathrm{h_{top}}(S, \mathcal{V})$ for every open cover $\mathcal{V}$ of Y. From this, we shall conclude that

$$\begin{aligned}\mathrm{h_{top}}(T) &= \sup\{\mathrm{h_{top}}(T, \mathcal{U}) : \mathcal{U} \text{ open cover of } X\}\\ &\ge \sup\{\mathrm{h_{top}}(T, h^{-1}(\mathcal{V})) : \mathcal{V} \text{ open cover of } Y\}\\ &= \sup\{\mathrm{h_{top}}(S, \mathcal{V}) : \mathcal{V} \text{ open cover of } Y\}\\ &= \mathrm{h_{top}}(S).\end{aligned}$$

In particular, if S and T are topologically conjugate, then S is a factor of T and T is a factor of S. So $\mathrm{h_{top}}(T) \ge \mathrm{h_{top}}(S)$ and $\mathrm{h_{top}}(S) \ge \mathrm{h_{top}}(T)$, that is, $\mathrm{h_{top}}(S) = \mathrm{h_{top}}(T)$.

It remains to prove that $\mathrm{h_{top}}(T, h^{-1}(\mathcal{V})) = \mathrm{h_{top}}(S, \mathcal{V})$ for each open cover $\mathcal{V}$ of Y. Fix $n \in \mathbb{N}$ momentarily. The respective actions of the maps S and T on $\mathcal{V}$ and $h^{-1}(\mathcal{V})$ until time $n-1$ will be denoted by $\mathcal{V}_S^n$ and $(h^{-1}(\mathcal{V}))_T^n$. Then

$$\begin{aligned}(h^{-1}(\mathcal{V}))_T^n &= h^{-1}(\mathcal{V}) \vee T^{-1}(h^{-1}(\mathcal{V})) \vee \cdots \vee T^{-(n-1)}(h^{-1}(\mathcal{V}))\\ &= h^{-1}(\mathcal{V}) \vee (h \circ T)^{-1}(\mathcal{V}) \vee \cdots \vee (h \circ T^{n-1})^{-1}(\mathcal{V})\\ &= h^{-1}(\mathcal{V}) \vee (S \circ h)^{-1}(\mathcal{V}) \vee \cdots \vee (S^{n-1} \circ h)^{-1}(\mathcal{V})\\ &= h^{-1}(\mathcal{V}) \vee h^{-1}(S^{-1}(\mathcal{V})) \vee \cdots \vee h^{-1}(S^{-(n-1)}(\mathcal{V}))\\ &= h^{-1}(\mathcal{V} \vee S^{-1}(\mathcal{V}) \vee \cdots \vee S^{-(n-1)}(\mathcal{V}))\\ &= h^{-1}(\mathcal{V}_S^n).\end{aligned}$$

Therefore,

$$Z_n(T, h^{-1}(\mathcal{V})) = Z_1((h^{-1}(\mathcal{V}))_T^n) = Z_1(h^{-1}(\mathcal{V}_S^n)) \le Z_1(\mathcal{V}_S^n) = Z_n(S, \mathcal{V}).$$

Since h is surjective, $(h^{-1}(\mathcal{V}))_T^n = h^{-1}(\mathcal{V}_S^n)$ implies that $h((h^{-1}(\mathcal{V}))_T^n) = \mathcal{V}_S^n$. Thus,

$$Z_n(S, \mathcal{V}) = Z_1(\mathcal{V}_S^n) = Z_1(h((h^{-1}(\mathcal{V}))_T^n)) \le Z_1((h^{-1}(\mathcal{V}))_T^n) = Z_n(T, h^{-1}(\mathcal{V})).$$

Combining the previous two inequalities, we obtain that

$$Z_n(T, h^{-1}(\mathcal{V})) = Z_n(S, \mathcal{V}).$$

Since n was chosen arbitrarily, by successively taking the logarithm of both sides, dividing by n and passing to the limit as n tends to infinity, we conclude that

$$\mathrm{h_{top}}(T, h^{-1}(\mathcal{V})) = \mathrm{h_{top}}(S, \mathcal{V}). \qquad \square$$

We have now shown that topological entropy is indeed a topological conjugacy invariant. Let us now study its behavior with respect to the iterates of the system.

Theorem 7.2.19. *For every $k \in \mathbb{N}$, we have* $\mathrm{h}_{\mathrm{top}}(T^k) = k\,\mathrm{h}_{\mathrm{top}}(T)$.

Proof. Fix $k \in \mathbb{N}$. Let $\mathcal{U}$ be an open cover of X. The action of the map T^k on $\mathcal{U}$ until time $n-1$ will be denoted by $\mathcal{U}^n_{T^k}$. For every $n \in \mathbb{N}$, we have that

$$\begin{aligned}\mathcal{U}^n_{T^k} &= \mathcal{U} \vee T^{-k}(\mathcal{U}) \vee \cdots \vee T^{-(n-1)k}(\mathcal{U}) \\ &\prec \mathcal{U} \vee T^{-1}(\mathcal{U}) \vee T^{-2}(\mathcal{U}) \vee \cdots \vee T^{-k}(\mathcal{U}) \vee \cdots \vee T^{-(n-1)k}(\mathcal{U}) \\ &= \mathcal{U}^{(n-1)k+1}.\end{aligned}$$

By Lemma 7.2.2, it ensues that

$$Z_n(T^k,\mathcal{U}) = Z_1(\mathcal{U}^n_{T^k}) \le Z_1(\mathcal{U}^{(n-1)k+1}) = Z_{(n-1)k+1}(T,\mathcal{U}).$$

Consequently,

$$\begin{aligned}\mathrm{h}_{\mathrm{top}}(T^k,\mathcal{U}) &= \lim_{n\to\infty} \frac{1}{n} \log Z_n(T^k,\mathcal{U}) \\ &\le \lim_{n\to\infty} \frac{(n-1)k+1}{n} \frac{1}{(n-1)k+1} \log Z_{(n-1)k+1}(T,\mathcal{U}) \\ &= \lim_{n\to\infty} \frac{(n-1)k+1}{n} \cdot \lim_{n\to\infty} \frac{1}{(n-1)k+1} \log Z_{(n-1)k+1}(T,\mathcal{U}) \\ &= k \cdot \mathrm{h}_{\mathrm{top}}(T,\mathcal{U}).\end{aligned}$$

The inequality arises from the fact that there were gaps in the iterates of T in the cover $\mathcal{U}^n_{T^k}$ that are not present in $\mathcal{U}^{(n-1)k+1}$. We can fill in those gaps by considering $\mathcal{U}^k$ rather than $\mathcal{U}$. Indeed,

$$\begin{aligned}(\mathcal{U}^k)^n_{T^k} &= \mathcal{U}^k \vee T^{-k}(\mathcal{U}^k) \vee \cdots \vee T^{-(n-1)k}(\mathcal{U}^k) \\ &= (\mathcal{U} \vee T^{-1}(\mathcal{U}) \vee T^{-2}(\mathcal{U}) \vee \cdots \vee T^{-(k-1)}(\mathcal{U})) \\ &\quad \vee T^{-k}(\mathcal{U} \vee T^{-1}(\mathcal{U}) \vee T^{-2}(\mathcal{U}) \vee \cdots \vee T^{-(k-1)}(\mathcal{U})) \\ &\quad \vee \cdots \\ &\quad \vee T^{-(n-1)k}(\mathcal{U} \vee T^{-1}(\mathcal{U}) \vee T^{-2}(\mathcal{U}) \vee \cdots \vee T^{-(k-1)}(\mathcal{U})) \\ &= \mathcal{U} \vee T^{-1}(\mathcal{U}) \vee T^{-2}(\mathcal{U}) \vee \cdots \vee T^{-(k-1)}(\mathcal{U}) \\ &\quad \vee T^{-k}(\mathcal{U}) \vee T^{-(k+1)}(\mathcal{U}) \vee T^{-(k+2)}(\mathcal{U}) \vee \cdots \vee T^{-(2k-1)}(\mathcal{U}) \\ &\quad \vee \cdots \\ &\quad \vee T^{-(n-1)k}(\mathcal{U}) \vee T^{-((n-1)k+1)}(\mathcal{U}) \vee \cdots \vee T^{-(nk-1)}(\mathcal{U}) \\ &= \mathcal{U}^{nk}.\end{aligned}$$

Therefore,

$$Z_n(T^k,\mathcal{U}^k) = Z_1((\mathcal{U}^k)^n_{T^k}) = Z_1(\mathcal{U}^{nk}) = Z_{nk}(T,\mathcal{U}).$$

Consequently,

$$\mathrm{h}_{\mathrm{top}}(T^k,\mathcal{U}^k)=\lim_{n\to\infty}\frac{1}{n}\log Z_n(T^k,\mathcal{U}^k)=k\cdot\lim_{n\to\infty}\frac{1}{nk}\log Z_{nk}(T,\mathcal{U})=k\cdot\mathrm{h}_{\mathrm{top}}(T,\mathcal{U}).$$

Using Lemma 7.2.15, we conclude that

$$\begin{aligned}\mathrm{h}_{\mathrm{top}}(T^k)&=\sup\{\mathrm{h}_{\mathrm{top}}(T^k,\mathcal{U}):\mathcal{U}\text{ open cover of }X\}\\&=\sup\{\mathrm{h}_{\mathrm{top}}(T^k,\mathcal{U}^k):\mathcal{U}\text{ open cover of }X\}\\&=k\sup\{\mathrm{h}_{\mathrm{top}}(T,\mathcal{U}):\mathcal{U}\text{ open cover of }X\}=k\,\mathrm{h}_{\mathrm{top}}(T).\end{aligned}$$ □

Taking a supremum over the collection of all (finite) open covers can be inconvenient, to say the least. Indeed, this collection is usually uncountable. We would thus like to identify situations in which the topological entropy of a system is determined by the topological entropy of the system with respect to a countable family of covers, that is, with respect to a sequence of covers. If topological entropy really provides a good description of the complexity of the dynamics of the system, then it is natural to request that this sequence of covers eventually become finer and finer, and that it encompass the structure of the underlying space at increasingly small scales. In a metrizable space, this suggests looking at the diameter of the covers.

The forthcoming lemma is the first result that requires the underlying space to be metrizable. In this lemma, note that $\mathcal{U}_n$ is a general cover. In particular, it is typically not equal to the dynamical cover $\mathcal{U}^n$. So, take care not to confuse the two.

Lemma 7.2.20. *The following quantities are all equal:*
(a) $\mathrm{h}_{\mathrm{top}}(T)$.
(b) $\sup\{\mathrm{h}_{\mathrm{top}}(T,\mathcal{U}):\mathcal{U}$ *open cover with* $\mathrm{diam}(\mathcal{U})\le\delta\}$ *for any* $\delta>0$.
(c) $\lim_{\varepsilon\to 0}\mathrm{h}_{\mathrm{top}}(T,\mathcal{U}_\varepsilon)$ *for any open covers* $(\mathcal{U}_\varepsilon)_{\varepsilon\in(0,\infty)}$ *such that* $\lim_{\varepsilon\to 0}\mathrm{diam}(\mathcal{U}_\varepsilon)=0$.
(d) $\lim_{n\to\infty}\mathrm{h}_{\mathrm{top}}(T,\mathcal{U}_n)$ *for any open covers* $(\mathcal{U}_n)_{n=1}^\infty$ *such that* $\lim_{n\to\infty}\mathrm{diam}(\mathcal{U}_n)=0$.

Proof. Clearly, (a)≥(b). It is also easy to see that (b)≥(c) for any $\delta>0$ and any family $(\mathcal{U}_\varepsilon)_{\varepsilon\in(0,\infty)}$ as described, and that (b)≥(d) for any sequence $(\mathcal{U}_n)_{n=1}^\infty$ as specified. It thus suffices to prove that (c)≥(a) and (d)≥(a). Actually, we will prove that (d)=(a) and leave to the reader the task of adapting that proof to establish that (c)=(a).

Let $\mathcal{V}$ be any open cover of X, and let $\delta(\mathcal{V})$ be a Lebesgue number for $\mathcal{V}$. As $\lim_{n\to\infty}\mathrm{diam}(\mathcal{U}_n)=0$, there exists $N\in\mathbb{N}$ such that $\mathrm{diam}(\mathcal{U}_n)<\delta(\mathcal{V})$ for each $n\ge N$. Fix $n\ge N$ momentarily. As $\mathrm{diam}(\mathcal{U}_n)<\delta(\mathcal{V})$, every member of $\mathcal{U}_n$ is contained in a member of $\mathcal{V}$. Thus $\mathcal{V}\prec\mathcal{U}_n$. By Proposition 7.2.14(a), we obtain that

$$\mathrm{h}_{\mathrm{top}}(T,\mathcal{V})\le\mathrm{h}_{\mathrm{top}}(T,\mathcal{U}_n).$$

Since this is true for all $n \geq N$, we deduce that

$$\mathrm{h}_{\mathrm{top}}(T,\mathcal{V}) \leq \inf_{n\geq N} \mathrm{h}_{\mathrm{top}}(T,\mathcal{U}_n) \leq \liminf_{n\to\infty} \mathrm{h}_{\mathrm{top}}(T,\mathcal{U}_n).$$

As the open cover $\mathcal{V}$ was chosen arbitrarily, we conclude that

$$\begin{aligned}\mathrm{h}_{\mathrm{top}}(T) = \sup_{\mathcal{V}} \mathrm{h}_{\mathrm{top}}(T,\mathcal{V}) &\leq \liminf_{n\to\infty} \mathrm{h}_{\mathrm{top}}(T,\mathcal{U}_n)\\ &\leq \limsup_{n\to\infty} \mathrm{h}_{\mathrm{top}}(T,\mathcal{U}_n)\\ &\leq \sup_{\mathcal{U}} \mathrm{h}_{\mathrm{top}}(T,\mathcal{U}) = \mathrm{h}_{\mathrm{top}}(T),\end{aligned}$$

which implies that $\mathrm{h}_{\mathrm{top}}(T) = \lim_{n\to\infty} \mathrm{h}_{\mathrm{top}}(T,\mathcal{U}_n)$. □

Part (d) of Lemma 7.2.20 characterized the topological entropy of a system as the limit of the topological entropy of the system relative to a sequence of covers. An even better result would be the characterization of the topological entropy as the topological entropy with respect to a single cover. This quest suggests introducing the following notion.

Definition 7.2.21. An open cover $\mathcal{U}$ of a metric space (X,d) is said to be a *generator* for a topological dynamical system $T : X \to X$ if

$$\lim_{n\to\infty} \mathrm{diam}(\mathcal{U}^n) = 0.$$

Lemma 7.2.22. *If a system $T : X \to X$ has a generator $\mathcal{U}$, then*

$$\mathrm{h}_{\mathrm{top}}(T) = \mathrm{h}_{\mathrm{top}}(T,\mathcal{U}).$$

Proof. It follows from Lemma 7.2.20 (with $\mathcal{U}_n = \mathcal{U}^n$) and Lemma 7.2.15 that

$$\mathrm{h}_{\mathrm{top}}(T) = \lim_{n\to\infty} \mathrm{h}_{\mathrm{top}}(T,\mathcal{U}^n) = \lim_{n\to\infty} \mathrm{h}_{\mathrm{top}}(T,\mathcal{U}) = \mathrm{h}_{\mathrm{top}}(T,\mathcal{U}).$$ □

It is natural to wonder about the class(es) of systems that admit a generator.

Lemma 7.2.23. *A topological dynamical system $T : X \to X$ admits a generator if and only if it is expansive. In fact, if T is δ-expansive when the compact metric space X is equipped with a metric d, then every open cover $\mathcal{U}$ of X such that* $\mathrm{diam}(\mathcal{U}) \leq \delta$ *is a generator for T.*

Proof. Suppose that an open cover $\mathcal{V}$ is a generator for T. Let δ be a Lebesgue number for $\mathcal{V}$. If $d(T^n(x),T^n(y)) \leq \delta$ for all $n \geq 0$, then for every n there exists $V_n \in \mathcal{V}$ such that $T^n(x), T^n(y) \in V_n$. Therefore $x, y \in \bigcap_{n=0}^{\infty} T^{-n}(V_n)$. This implies that x, y lie in a common member of $\mathcal{V}^N$ for all $N \in \mathbb{N}$. Since $\lim_{N\to\infty} \mathrm{diam}(\mathcal{V}^N) = 0$, we conclude that $x = y$. So δ is an expansive constant for T.

The converse implication can be seen as an immediate consequence of uniform expansiveness. See Proposition 5.2.2. Nevertheless, we provide a direct proof. Let $\mathcal{U} = \{U_e : e \in E\}$ be a finite open cover of X with $\operatorname{diam}(\mathcal{U}) \le \delta$. This means that E is a finite index set (an alphabet), and E^∞ is compact when endowed with any of the metrics $d_s(\omega, \tau) = s^{|\omega\wedge\tau|}$ (cf. Definition 3.1.10 and Lemma 3.1.7). Let $\varepsilon > 0$. We must show that there exists $N \in \mathbb{N}$ such that $\operatorname{diam}(\mathcal{U}^n) < \varepsilon$ for all $n \ge N$. First, we claim that for every $\omega \in E^\infty$, the set

$$\bigcap_{j=0}^{\infty} T^{-j}(\overline{U_{\omega_j}})$$

comprises at most one point. To show this, let $x, y \in \bigcap_{j=0}^{\infty} T^{-j}(\overline{U_{\omega_j}})$. Then $T^j(x), T^j(y) \in \overline{U_{\omega_j}}$ for all $j \ge 0$. So, for all $j \ge 0$,

$$d(T^j(x), T^j(y)) \le \operatorname{diam}(\overline{U_{\omega_j}}) = \operatorname{diam}(U_{\omega_j}) \le \operatorname{diam}(\mathcal{U}) \le \delta.$$

By the δ-expansiveness of T, we deduce that $x = y$ and the claim is proved. It follows from this fact that

$$\lim_{n\to\infty} \operatorname{diam}\left(\bigcap_{j=0}^{n-1} T^{-j}(\overline{U_{\omega_j}})\right) = 0, \quad \forall \omega \in E^\infty.$$

Otherwise, a compactness argument shows that the set $\bigcap_{j=0}^{\infty} T^{-j}(\overline{U_{\omega_j}})$ would contain at least two points (see Exercise 7.6.6), which would contradict the claim established above.

Thus, for every $\omega \in E^\infty$ there exists a smallest $N(\omega) \in \mathbb{N}$ such that

$$\operatorname{diam}\left(\bigcap_{j=0}^{N(\omega)-1} T^{-j}(\overline{U_{\omega_j}})\right) < \varepsilon.$$

This defines a function $N : E^\infty \to \mathbb{N}$. We claim that this function is locally constant and thereby continuous. Indeed, let $\omega \in E^\infty$ and pick any $\tau \in [\omega|_{N(\omega)}] = [\omega_0\omega_1\omega_2\ldots\omega_{N(\omega)-1}]$. Then $\tau_j = \omega_j$ for all $0 \le j < N(\omega)$. Therefore, for all $1 \le n \le N(\omega)$, we have that

$$\bigcap_{j=0}^{n-1} T^{-j}(\overline{U_{\tau_j}}) = \bigcap_{j=0}^{n-1} T^{-j}(\overline{U_{\omega_j}}).$$

This implies that for all $1 \le n < N(\omega)$,

$$\operatorname{diam}\left(\bigcap_{j=0}^{n-1} T^{-j}(\overline{U_{\tau_j}})\right) = \operatorname{diam}\left(\bigcap_{j=0}^{n-1} T^{-j}(\overline{U_{\omega_j}})\right) \ge \varepsilon,$$

whereas

$$\operatorname{diam}\left(\bigcap_{j=0}^{N(\omega)-1} T^{-j}(\overline{U_{\tau_j}})\right) = \operatorname{diam}\left(\bigcap_{j=0}^{N(\omega)-1} T^{-j}(\overline{U_{\omega_j}})\right) < \varepsilon.$$

Thus $N(\tau) = N(\omega)$ for every $\tau \in [\omega|_{N(\omega)}]$. This proves that N is a locally constant function.

Since N is continuous and E^∞ is compact, the image $N(E^\infty)$ is a compact subset of $\mathbb{N}$ and is hence bounded. Set $N_{\max} := \max\{N(\omega) : \omega \in E^\infty\} < \infty$. Then for every $n \geq N_{\max}$ and for every $\omega \in E^\infty$, we have

$$\operatorname{diam}\left(\bigcap_{j=0}^{n-1} T^{-j}(U_{\omega_j})\right) \leq \operatorname{diam}\left(\bigcap_{j=0}^{N_{\max}-1} T^{-j}(\overline{U_{\omega_j}})\right) \leq \operatorname{diam}\left(\bigcap_{j=0}^{N(\omega)-1} T^{-j}(\overline{U_{\omega_j}})\right) < \varepsilon.$$

So $\operatorname{diam}(\mathcal{U}^n) < \varepsilon$ for all $n \geq N_{\max}$. As $\varepsilon > 0$ was chosen arbitrarily, we conclude that $\mathcal{U}$ is a generator for T. $\square$

In light of the previous two results, the topological entropy of an expansive system can be characterized as the topological entropy of that system with respect to a single cover.

Theorem 7.2.24. *If $T : X \to X$ is a δ-expansive dynamical system on a compact metric space (X, d), then*

$$\mathrm{h_{top}}(T) = \mathrm{h_{top}}(T, \mathcal{U})$$

for any open cover $\mathcal{U}$ of X with $\operatorname{diam}(\mathcal{U}) \leq \delta$.

Proof. This is an immediate consequence of Lemmas 7.2.22 and 7.2.23. $\square$

Remark 7.2.25. Notice that Theorem 7.2.24 immediately implies that the topological entropy of an expansive dynamical system on a compact metric space is finite.

Example 7.2.26. Let E be a finite alphabet and A be an incidence/transition matrix. Let $\sigma : E_A^\infty \to E_A^\infty$ be the corresponding subshift of finite type. We will show that

$$\mathrm{h_{top}}(\sigma) = \log r(A),$$

where $r(A)$ is the spectral radius of A. Recall from Example 5.1.4 that the shift map σ is expansive and has for an expansive constant any $0 < \delta < 1$ when E_A^∞ is endowed with any metric $d_s(\omega, \tau) = s^{|\omega \wedge \tau|}$, where $0 < s < 1$. Choose $\mathcal{U} := \{[e] : e \in E\}$ as a (finite) open cover of E_A^∞. So $\mathcal{U}$ is the open partition of E_A^∞ into initial 1-cylinders. Since $\operatorname{diam}(\mathcal{U}) = s < 1$, in light of Theorem 7.2.24 we know that $\mathrm{h_{top}}(\sigma) = \mathrm{h_{top}}(\sigma, \mathcal{U})$. In order to compute $\mathrm{h_{top}}(\sigma, \mathcal{U})$, notice that for each $n \in \mathbb{N}$ we have

$$\mathcal{U}^n = \sigma^{-(n-1)}(\mathcal{U}) = \{[\omega] : \omega \in E_A^n\}.$$

That is to say, $\mathcal{U}^n$ is the open partition of E_A^∞ into initial cylinders of length n. Since the only subcover that a partition admits is itself, we obtain that

$$Z_n(\mathcal{U}) = \#E_A^n.$$

Consequently,

$$\mathrm{h_{top}}(\sigma) = \mathrm{h_{top}}(\sigma,\mathcal{U}) = \lim_{n\to\infty}\frac{1}{n}\log Z_n(\mathcal{U}) = \lim_{n\to\infty}\frac{1}{n}\log \#E_A^n = \log r(A).$$

The last equality follows from Theorem 3.2.24.

Example 7.2.27. If E is a finite alphabet, then the topological entropy of the full E-shift is equal to $\log \#E$. This is a special case of the previous example with A as the matrix that consists only of 1's. Alternatively, notice that $\#(E^n) = (\#E)^n$.

We shall now compute the topological entropy of a particular subshift of finite type, the well-known *golden mean shift*, which was introduced in Exercise 3.4.9. Perhaps not surprisingly, given its name, it will turn out that the topological entropy of the golden mean shift is equal to the logarithm of the golden mean.

Example 7.2.28. Let $\sigma : E_A^\infty \to E_A^\infty$ be the golden mean shift, that is, the subshift of finite type induced by the incidence matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

By Example 7.2.26, the computation of the topological entropy of a subshift of finite type boils down to counting the number of initial n-cylinders for all but finitely many n's, or more simply to computing the largest eigenvalue (in absolute value) of the transition matrix A.

The reader can calculate that the golden mean $\gamma := (1+\sqrt{5})/2$ and its conjugate $\gamma^* := (1-\sqrt{5})/2$ are the eigenvalues of the matrix A. Therefore,

$$\mathrm{h_{top}}(\sigma) = \log r(A) = \log \max\{|\gamma|, |\gamma^*|\} = \log\gamma.$$

Alternatively, one can prove by induction that $\#(E_A^n) = f_{n+2}$, where f_n is the nth Fibonacci number (see Exercise 7.6.7). Then one can verify by induction that

$$f_n = \frac{1}{\sqrt{5}}[\gamma^n - (\gamma^*)^n] = \frac{\gamma^n}{\sqrt{5}}[1-(\gamma^*/\gamma)^n].$$

It follows immediately that

$$\begin{aligned}
\mathrm{h_{top}}(\sigma) &= \lim_{n\to\infty}\frac{1}{n}\log\#(E_A^n) = \lim_{n\to\infty}\frac{1}{n}\log f_{n+2} \\
&= \lim_{n\to\infty}\frac{1}{n}\log\left(\frac{\gamma^{n+2}}{\sqrt{5}}[1-(\gamma^*/\gamma)^{n+2}]\right) \\
&= \lim_{n\to\infty}\frac{1}{n}\Big((n+2)\log\gamma - \log\sqrt{5} + \log[1-(\gamma^*/\gamma)^{n+2}]\Big) \\
&= \log\gamma.
\end{aligned}$$

7.3 Bowen's definition of topological entropy

In this section, we present an alternative definition of topological entropy due to Rufus Bowen. In contrast with the definition via open covers, this definition is only valid in a metric space X. Although Bowen's definition does make sense in a noncompact metric space, we will as usual assume that X is compact throughout.

Let $T : X \to X$ be a dynamical system on a compact metric space (X, d). Recall from Definition 5.1.3 the dynamical metrics (also called Bowen's metrics), d_n, for each $n \in \mathbb{N}$:

$$d_n(x, y) = \max\{d(T^j(x), T^j(y)) : 0 \le j < n\}.$$

The open ball centered at x of radius r induced by the metric d_n is denoted by $B_n(x, r)$ and is called the dynamical (n, r)-ball at x. As $d_1 = d$, we shall denote $B_1(x, r)$ simply by $B(x, r)$. Finally, observe that

$$B_n(x, r) = \{y \in X : d_n(x, y) < r\} = \bigcap_{j=0}^{n-1} T^{-j}(B(T^j(x), r)).$$

In other words, the ball $B_n(x, r)$ consists of those points whose iterates stay within a distance r from the corresponding iterates of x until time $n-1$ at least. In the language of Chapter 4, the ball $B_n(x, r)$ is the set of all those points whose orbits are r-shadowed by the orbit of x until time $n - 1$ at least.

Definition 7.3.1. A subset E of X is said to be (n, ε)-*separated* if E is ε-separated with respect to the metric d_n, which is to say that $d_n(x, y) \ge \varepsilon$ for all $x, y \in E$ with $x \ne y$.

Remark 7.3.2.
(a) If E is an (m, ε)-separated set and $m < n$, then E is also (n, ε)-separated.
(b) If E is an (n, ε')-separated set and $\varepsilon < \varepsilon'$, then E is also (n, ε)-separated.
(c) Given that the underlying space X is compact, any (n, ε)-separated set is finite. Indeed, let E be an (n, ε)-separated set, and consider the family of balls $\{B_n(x, \varepsilon/2) : x \in E\}$. If the intersection of $B_n(x, \varepsilon/2)$ and $B_n(y, \varepsilon/2)$ is nonempty for some $x, y \in E$, then there exists $z \in B_n(x, \varepsilon/2) \cap B_n(y, \varepsilon/2)$ and it follows that

$$d_n(x, y) \le d_n(x, z) + d_n(z, y) < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

As E is an (n, ε)-separated set, this inequality implies that $x = y$. This means that the balls $\{B_n(x, \varepsilon/2) : x \in E\}$ are mutually disjoint. Hence, as X is compact, there can only be finitely many points in E.

The largest separated sets will be especially useful in describing the complexity of the dynamics that the system exhibits.

Definition 7.3.3. A subset E of X is called a *maximal* (n,ε)*-separated set* if for any (n,ε)-separated set E' with $E \subseteq E'$, we have $E = E'$. In other words, no strict superset of E is (n,ε)-separated.

The counterpart of the notion of separated set is the concept of spanning set.

Definition 7.3.4. A subset F of X is said to be an (n,ε)*-spanning set* if

$$\bigcup_{x\in F} B_n(x,\varepsilon) = X.$$

That is, the orbit of every point in the space is ε-shadowed by the orbit of a point of F until time $n-1$ at least.

The smallest spanning sets play a special role in describing the complexity of the dynamics that the system possesses. They constitute the counterpart of the maximal separated sets.

Definition 7.3.5. A subset F of X is called a *minimal* (n,ε)*-spanning set* if for any (n,ε)-spanning set F' with $F \supseteq F'$, we have $F = F'$. In other words, no strict subset of F is (n,ε)-spanning.

Remark 7.3.6.
(a) If F is an (n,ε)-spanning set and $m < n$, then F is also (m,ε)-spanning.
(b) If F is an (n,ε)-spanning set and $\varepsilon < \varepsilon'$, then F is also (n,ε')-spanning.
(c) Any minimal (n,ε)-spanning set is finite since the open cover $\{B_n(x,\varepsilon) : x \in X\}$ of the compact metric space X admits a finite subcover.

The next lemma describes two useful relations between separated and spanning sets.

Lemma 7.3.7. *The following statements hold:*
(a) *Every maximal* (n,ε)*-separated set is a minimal* (n,ε)*-spanning set.*
(b) *Every* $(n,2\varepsilon)$*-separated set is embedded into any* (n,ε)*-spanning set.*

Proof. (a) Let E be a maximal (n,ε)-separated set. First, suppose that E is not (n,ε)-spanning. Then there exists a point $y \in X\setminus\bigcup_{x\in E} B_n(x,\varepsilon)$. Observe then that the set $E \cup \{y\}$ is (n,ε)-separated, which contradicts the maximality of E. Therefore, E is (n,ε)-spanning. Suppose now that E is not a minimal (n,ε)-spanning set. Then there exists an (n,ε)-spanning set $E' \subsetneq E$. Let $y \in E\setminus E'$. Since $\bigcup_{y'\in E'} B_n(y',\varepsilon) = X$, there is $y' \in E' \subseteq E$ such that $d_n(y',y) < \varepsilon$. This implies that E is not (n,ε)-separated, a contradiction. Consequently, if E is a maximal (n,ε)-separated set, then E is a minimal (n,ε)-spanning set.

(b) Let E be an $(n,2\varepsilon)$-separated set and F an (n,ε)-spanning set. For each $x \in E$, choose $i(x) \in F$ such that $x \in B_n(i(x),\varepsilon)$. We claim that the map $i : E \to F$ is injective. To show this, let $x,y \in E$ be such that $i(x) = i(y) =: z$. Then $x,y \in B_n(z,\varepsilon)$. Therefore

$d_n(x, y) < 2\varepsilon$. Since E is a $(n, 2\varepsilon)$-separated set, we deduce that $x = y$, that is, the map i is injective and so E is embedded into F. □

The next theorem is the main result of this section. It gives us another way of defining the topological entropy of a system.

Theorem 7.3.8. *For all $\varepsilon > 0$ and $n \in \mathbb{N}$, let $E_n(\varepsilon)$ be a maximal (n, ε)-separated set in X and $F_n(\varepsilon)$ be a minimal (n, ε)-spanning set in X. Then*

$$\begin{aligned}\mathrm{h_{top}}(T) = \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log \#E_n(\varepsilon) &= \lim_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log \#E_n(\varepsilon)\\ &\le \liminf_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log \#F_n(\varepsilon).\end{aligned}$$

Proof. Fix $\varepsilon > 0$ and $n \in \mathbb{N}$ and let $E_n(\varepsilon)$ be a maximal (n, ε)-separated set in X. Let $\mathcal{U}$ be an open cover of X consisting of balls of radius $\varepsilon/2$. Let $\underline{\mathcal{U}}$ be a minimal subcover of $\mathcal{U}^n$, so that $Z_n(\mathcal{U}) = \#\underline{\mathcal{U}}$. For each $x \in E_n(\varepsilon)$, let $\underline{U}(x)$ be an element of the cover $\underline{\mathcal{U}}$ which contains x and define the function $i : E_n(\varepsilon) \to \underline{\mathcal{U}}$ by setting $i(x) = \underline{U}(x)$. We claim that this function is an injection. Indeed, suppose that $x, y \in E_n(\varepsilon)$ are such that $\underline{U}(x) = \underline{U}(y)$. Then, by the definition of $\mathcal{U}^n$, we have that

$$x, y \in \bigcap_{j=0}^{n-1} T^{-j}(U_j),$$

where $U_j = B(z_j, \varepsilon/2)$ for some $z_j \in X$. This means that both $T^j(x)$ and $T^j(y)$ belong to $B(x_j, \varepsilon/2)$ for each $0 \le j < n$. So, $d_n(x, y) < \varepsilon$, and thus $x = y$ since $E_n(\varepsilon)$ is (n, ε)-separated. This establishes that $i : E_n(\varepsilon) \to \underline{\mathcal{U}}$ is injective. Therefore, $Z_n(\mathcal{U}) = \#\underline{\mathcal{U}} \ge \#E_n(\varepsilon)$. Since $\mathcal{U}$ does not depend on n and the inequality $Z_n(\mathcal{U}) \ge \#E_n(\varepsilon)$ holds for all $n \in \mathbb{N}$, we deduce that

$$\mathrm{h_{top}}(T) \ge \mathrm{h_{top}}(T, \mathcal{U}) = \lim_{n\to\infty}\frac{1}{n}\log Z_n(\mathcal{U}) \ge \limsup_{n\to\infty}\frac{1}{n}\log \#E_n(\varepsilon).$$

Consequently,

$$\mathrm{h_{top}}(T) \ge \limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log \#E_n(\varepsilon). \tag{7.1}$$

On the other hand, if $\mathcal{V}$ is an arbitrary open cover of X, if $\delta(\mathcal{V})$ is a Lebesgue number for $\mathcal{V}$, if $0 < \varepsilon < \delta(\mathcal{V})/2$ and if $n \in \mathbb{N}$, then for all $0 \le k < n$ and all $x \in E_n(\varepsilon)$, we have that

$$T^k(B_n(x, \varepsilon)) \subseteq B(T^k(x), \varepsilon) \implies \mathrm{diam}(T^k(B_n(x, \varepsilon))) \le 2\varepsilon < \delta(\mathcal{V}).$$

Hence, for all $0 \le k < n$, the set $T^k(B_n(x, \varepsilon))$ is contained in at least one element of the cover $\mathcal{V}$. Denote one of these elements by $V_k(x)$. It follows that $B_n(x, \varepsilon) \subseteq T^{-k}(V_k(x))$

for each $0 \le k < n$. In other words, we have that $B_n(x,\varepsilon) \subseteq \bigcap_{k=0}^{n-1} T^{-k}(V_k(x))$. But this intersection is an element of $\mathcal{V}^n$. Let us denote it by $V(x)$.

Since $E_n(\varepsilon)$ is a maximal (n,ε)-separated set, Lemma 7.3.7(a) asserts that it is also (n,ε)-spanning. Thus the family $\{B_n(x,\varepsilon)\}_{x\in E_n(\varepsilon)}$ is an open cover of X. Each one of these balls is contained in the corresponding element $V(x)$ of $\mathcal{V}^n$. Hence, the family $\{V(x)\}_{x\in E_n(\varepsilon)}$ is also an open cover of X, and thus a subcover of $\mathcal{V}^n$. Consequently,

$$Z_n(\mathcal{V}) \le \#\{V(x)\}_{x\in E_n(\varepsilon)} \le \#E_n(\varepsilon).$$

Since this is true for all $n \in \mathbb{N}$, we deduce that

$$\mathrm{h_{top}}(T,\mathcal{V}) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(\mathcal{V}) \le \liminf_{n\to\infty} \frac{1}{n} \log \#E_n(\varepsilon).$$

As this inequality holds for all $0 < \varepsilon < \delta(\mathcal{V})/2$, we obtain that

$$\mathrm{h_{top}}(T,\mathcal{V}) \le \liminf_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \#E_n(\varepsilon).$$

Because $\mathcal{V}$ was chosen to be an arbitrary open cover of X, we conclude that

$$\mathrm{h_{top}}(T) \le \liminf_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \#E_n(\varepsilon). \tag{7.2}$$

The inequalities (7.1) and (7.2) combined are sufficient to establish the result for the separated sets.

Now, since every $(n,2\varepsilon)$-separated set is embedded into any (n,ε)-spanning set according to Lemma 7.3.7(b), we have that

$$\#E_n(2\varepsilon) \le \#F_n(\varepsilon). \tag{7.3}$$

The inequalities (7.2) and (7.3) suffice to deduce the result for the spanning sets. □

In Theorem 7.3.8, the topological entropy of the system is expressed in terms of a specific family of maximal separated (resp., minimal spanning) sets. However, to derive theoretical results, it is sometimes simpler to use the following quantities.

Definition 7.3.9. For all $n \in \mathbb{N}$ and $\varepsilon > 0$, let

$$r_n(\varepsilon) = \sup\{\#E_n(\varepsilon) : E_n(\varepsilon) \text{ maximal } (n,\varepsilon)\text{-separated set}\}$$

and

$$s_n(\varepsilon) = \inf\{\#F_n(\varepsilon) : F_n(\varepsilon) \text{ minimal } (n,\varepsilon)\text{-spanning set}\}.$$

Thereafter, let

$$\underline{r}(\varepsilon) = \liminf_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon) \quad \text{while} \quad \overline{r}(\varepsilon) = \limsup_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon)$$

and

$$\underline{s}(\varepsilon) = \liminf_{n\to\infty} \frac{1}{n} \log s_n(\varepsilon) \quad \text{whereas} \quad \overline{s}(\varepsilon) = \limsup_{n\to\infty} \frac{1}{n} \log s_n(\varepsilon).$$

The following are simple but key observations about these quantities.

Remark 7.3.10. For all $m < n \in \mathbb{N}$ and $0 < \varepsilon < \varepsilon'$, the following relations hold.
(a) $r_m(\varepsilon) \le r_n(\varepsilon)$ and $s_m(\varepsilon) \le s_n(\varepsilon)$ by Remarks 7.3.2 and 7.3.6.
(b) $r_n(\varepsilon) \ge r_n(\varepsilon')$ and $s_n(\varepsilon) \ge s_n(\varepsilon')$ by Remarks 7.3.2 and 7.3.6.
(c) $0 < s_n(\varepsilon) \le r_n(\varepsilon) \le s_n(\varepsilon/2) < \infty$ by Lemma 7.3.7.
(d) $\underline{r}(\varepsilon) \le \overline{r}(\varepsilon)$ and $\underline{s}(\varepsilon) \le \overline{s}(\varepsilon)$.
(e) $\underline{r}(\varepsilon) \ge \underline{r}(\varepsilon')$ and $\overline{r}(\varepsilon) \ge \overline{r}(\varepsilon')$ by (b).
(f) $\underline{s}(\varepsilon) \ge \underline{s}(\varepsilon')$ and $\overline{s}(\varepsilon) \ge \overline{s}(\varepsilon')$ by (b).
(g) $0 \le \overline{s}(\varepsilon) \le \overline{r}(\varepsilon) \le \overline{s}(\varepsilon/2) \le \infty$ by (c).
(h) $0 \le \underline{s}(\varepsilon) \le \underline{r}(\varepsilon) \le \underline{s}(\varepsilon/2) \le \infty$ by (c).

We will now prove two properties that relate r_n's, s_n's, and Z_n's.

Lemma 7.3.11. *The following relations hold:*
(a) *If $\mathcal{U}$ is an open cover of X with Lebesgue number 2δ, then*

$$Z_n(\mathcal{U}) \le s_n(\delta) \le r_n(\delta), \quad \forall n \in \mathbb{N}.$$

(b) *If $\varepsilon > 0$ and $\mathcal{V}$ is an open cover of X with* $\operatorname{diam}(\mathcal{V}) \le \varepsilon$, *then*

$$s_n(\varepsilon) \le r_n(\varepsilon) \le Z_n(\mathcal{V}), \quad \forall n \in \mathbb{N}.$$

Proof. Let $n \in \mathbb{N}$. We already know that $s_n(\delta) \le r_n(\delta)$.

(a) Let $\mathcal{U}$ be an open cover with Lebesgue number 2δ and let F be an (n,δ)-spanning set. Then the dynamic balls $\{B_n(x,\delta) : x \in F\}$ form a cover of X. For every $0 \le i < n$ the ball $B(T^i(x),\delta)$, which has diameter at most 2δ, is contained in an element of $\mathcal{U}$. Therefore, $B_n(x,\delta) = \bigcap_{i=0}^{n-1} T^{-i}(B(T^i(x),\delta))$ is contained in an element of $\mathcal{U}^n = \bigvee_{i=0}^{n-1} T^{-i}(\mathcal{U})$. That is, $\mathcal{U}^n \prec \{B_n(x,\delta) : x \in F\}$. Thus $Z_n(\mathcal{U}) = Z_1(\mathcal{U}^n) \le \#F$. Since F is an arbitrary (n,δ)-spanning set, it ensues that $Z_n(\mathcal{U}) \le s_n(\delta)$.

(b) Let $\mathcal{V}$ be an open cover with $\operatorname{diam}(\mathcal{V}) \le \varepsilon$ and let E be an (n,ε)-separated set. Then no element of the cover $\mathcal{V}^n$ contains more than one element of E. Hence, $\#E \le Z_n(\mathcal{V})$. Since E is an arbitrary (n,ε)-separated set, it follows that $r_n(\varepsilon) \le Z_n(\mathcal{V})$. □

Together, Lemmas 7.3.11 and 7.2.20 have the following immediate corollary. Unlike Theorem 7.3.8, this result is symmetric with respect to separated and spanning sets. It is the advantage of using spanning sets of minimal cardinality, rather than spanning sets that are minimal in terms of inclusion.

Corollary 7.3.12. *The following equalities hold:*

$$\mathrm{h_{top}}(T) = \lim_{\varepsilon\to 0} \underline{r}(\varepsilon) = \lim_{\varepsilon\to 0} \overline{r}(\varepsilon) = \lim_{\varepsilon\to 0} \underline{s}(\varepsilon) = \lim_{\varepsilon\to 0} \overline{s}(\varepsilon).$$

Corollary 7.3.12 is useful to derive theoretical results. Nevertheless, in practice, Theorem 7.3.8 is simpler to use, as only one family (in essence, a double sequence) of sets is needed. Sometimes a single sequence is enough.

Theorem 7.3.13. *If a topological dynamical system $T : X \to X$ admits a generator with Lebesgue number 2δ, then the following statements hold for all $0 < \varepsilon \le \delta$:*

(a) *If $(E_n(\varepsilon))_{n=1}^{\infty}$ is a sequence of maximal (n,ε)-separated sets in X, then*

$$\mathrm{h_{top}}(T) = \lim_{n\to\infty} \frac{1}{n} \log \#E_n(\varepsilon).$$

(b) *If $(F_n(\varepsilon))_{n=1}^{\infty}$ is a sequence of minimal (n,ε)-spanning sets in X, then*

$$\mathrm{h_{top}}(T) \le \liminf_{n\to\infty} \frac{1}{n} \log \#F_n(\varepsilon).$$

(c) $\mathrm{h_{top}}(T) = \lim_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon)$.

(d) $\mathrm{h_{top}}(T) = \lim_{n\to\infty} \frac{1}{n} \log s_n(\varepsilon)$.

Proof. We will prove (a) and leave it to the reader to show the other parts using similar arguments.

Let $\mathcal{U}$ be a generator for T with Lebesgue number 2δ. Set $0 < \varepsilon \le \delta$. Observe that 2ε is also a Lebesgue number for $\mathcal{U}$. Choose any sequence $(E_n(\varepsilon))_{n=1}^{\infty}$ of maximal (n,ε)-separated sets. Since maximal (n,ε)-separated sets are (n,ε)-spanning sets, it follows from Lemma 7.3.11(a) that $Z_n(\mathcal{U}) \le s_n(\varepsilon) \le \#E_n(\varepsilon)$. Therefore,

$$\mathrm{h_{top}}(T,\mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(\mathcal{U}) \le \liminf_{n\to\infty} \frac{1}{n} \log \#E_n(\varepsilon). \tag{7.4}$$

On the other hand, since $\mathcal{U}$ is a generator, there exists $k \in \mathbb{N}$ such that $\operatorname{diam}(\mathcal{U}^k) \le \varepsilon$. It ensues from Lemma 7.3.11(b) that $\#E_n(\varepsilon) \le r_n(\varepsilon) \le Z_n(\mathcal{U}^k)$. Consequently,

$$\limsup_{n\to\infty} \frac{1}{n} \log \#E_n(\varepsilon) \le \lim_{n\to\infty} \frac{1}{n} \log Z_n(\mathcal{U}^k) = \mathrm{h_{top}}(T,\mathcal{U}^k) = \mathrm{h_{top}}(T,\mathcal{U}), \tag{7.5}$$

where the last equality follows from Lemma 7.2.15. Combining (7.4) and (7.5) gives

$$\mathrm{h_{top}}(T,\mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} \log \#E_n(\varepsilon).$$

As $\mathcal{U}$ is a generator, $\mathrm{h_{top}}(T) = \mathrm{h_{top}}(T,\mathcal{U})$ by Lemma 7.2.22. □

For expansive systems, the Lebesgue number can be expressed in terms of the expansive constant.

Theorem 7.3.14. *If $T : X \to X$ is a δ_0-expansive dynamical system on a compact metric space (X,d), then Theorem 7.3.13 applies with any $0 < \delta < \delta_0/4$.*

Proof. According to Theorem 7.2.23, any open cover $\mathcal{U}$ of X with $\operatorname{diam}(\mathcal{U}) \le \delta_0$ is a generator for T. In particular, a cover composed of open balls works. Let $x_1, \ldots, x_n \in X$ be such that $X = \bigcup_{i=1}^{n} B(x_i, \delta_0/2 - 2\delta)$. Then the cover $\{B(x_i, \delta_0/2) : 1 \le i \le n\}$ has diameter at most δ_0 and admits 2δ as Lebesgue number. □

7.4 Topological degree

In this section, which in the context of this book is a preparation for the subsequent section where we will give a very useful lower bound for the topological entropy of a C^1 endomorphism, we introduce the concept of topological degree for maps of smooth compact orientable manifolds. Although this is in fact a topological concept well-defined for continuous maps of topological manifolds, we concentrate on differentiable maps. This is somewhat easier and perfectly fits the needs of the proof of the entropy bound mentioned above. For more information on these notions, please see Hirsch [30].

Definition 7.4.1. Let M and N be smooth compact orientable d-dimensional manifolds. Let $f : M \to N$ be a C^1 map. A point $x \in M$ is called a regular point for f if D_xf is invertible. A point $y \in N$ is called a regular value of f if $f^{-1}(y)$ consists of regular points. Otherwise, y is called a singular value of f.

It is obvious that the set of regular values is open. It is also easy to see that the preimage of any regular value is of finite cardinality.

Lemma 7.4.2. *If $y \in N$ is a regular value of a C^1 map $f : M \to N$, then $\#f^{-1}(y) < \infty$.*

Proof. Suppose that $\#f^{-1}(y) = \infty$. Then there exists a sequence $(x_n)_{n=1}^{\infty}$ in M such that $\lim_{n\to\infty} x_n = x$ for some $x \in M$ and $f(x_n) = y$ for all $n \in \mathbb{N}$. The continuity of f imposes that $f(x) = y$, and thus f is not injective on any neighborhood of x. By the implicit function theorem, it follows that D_xf is not invertible, that is, x is not a regular point. So y is a singular value. □

Definition 7.4.3. Suppose that $y \in N$ is a regular value of a C^1 map $f : M \to N$. For each $x \in f^{-1}(y)$, let $\epsilon_x = \pm 1$ depending on whether D_xf preserves or reverses orientation. Then the degree of f at y is defined by

$$\deg_y(f) = \sum_{x \in f^{-1}(y)} \epsilon_x.$$

The preimage of any regular value being of finite cardinality, this sum is a well-defined integer. The degree of f at y measures how many times f covers N near y, counted with appropriate positive and negative multiplicities. In fact, the degree is independent of the choice of the regular value in N. In order to show that, there exists an alternative definition based on integration.

Definition 7.4.4. A positive normalized volume element on N is a continuous d-form ω that is positive on positively oriented frames and such that $\int_M \omega = 1$. The pullback $f^*\omega$ of ω under a C^1 map $f : M \to N$ is the d-form on M given by

$$(f^*\omega)(v_1, \ldots, v_d) = \omega(Df(v_1), \ldots, Df(v_d)).$$

Definition 7.4.5. If ω is a positive normalized volume element on N, then the degree of a C^1 map $f : M \to N$ with respect to ω is defined by

$$\deg_\omega(f) = \int_M f^*\omega.$$

We now show that the two aforementioned definitions of degree are independent of y and ω, respectively.

Lemma 7.4.6. *Let $y \in N$ be a regular value of a C^1 map $f : M \to N$ and ω a positive normalized volume element on N. Then* $\deg_y(f) = \deg_\omega(f)$.

Proof. As y is a regular value, there are disjoint open neighborhoods $U_1, \dots, U_k \subseteq M$ of the points $x_1, \dots, x_k$ of $f^{-1}(y)$ such that $\bigcup_{i=1}^k U_i$ is the preimage of a neighborhood V of y and $f|_{U_i}$ is a diffeomorphism for all i. If ν is an n-form supported in V such that $\int_V \nu = \int_N \nu = 1$, then $\omega = \nu + d\alpha$ for some $(n-1)$-form α, so $\int_M f^*\omega = \int_M (f^*\nu + f^*d\alpha) = \int_M f^*\nu = \sum_{i=1}^k \int (f|_{U_i})^*\nu = \sum_{i=1}^k \int_{U_i} f^*\nu$. By the transformation rule, each of the latter integrals is ± 1 according to whether $f|_{U_i}$, or equivalently $D_{x_i}f$, preserves or reverses orientation. Consequently, $\deg_\omega(f) = \int_M f^*\omega = \deg_y(f)$. □

So we can now make the following definition.

Definition 7.4.7. The degree of a C^1 map $f : M \to N$ is defined by $\deg(f) := \deg_y(f)$ for any regular value $y \in N$.

7.5 Misiurewicz–Przytycki theorem

In this section, we shall provide a very effective lower bound for the topological entropy of C^1 endomorphisms. Its attractiveness lies in it being expressed in relatively simple terms. The much stronger theorem of Yomdin [79], commonly referred to as the *entropy conjecture* and which gives the lower bound on topological entropy in terms of the logarithm of the spectral radius of the map induced on the full homology ring, is incomparably harder to prove and, often, harder to apply. The proof below is a slight modification of the one given in [33].

Theorem 7.5.1 (Misiurewicz–Przytycki theorem). *If M is a smooth compact orientable manifold and $T : M \to M$ a C^1 endomorphism, then* $\mathrm{h_{top}}(T) \geq \log|\deg(T)|$.

Proof. Fix a volume element ω on M and $\alpha \in (0,1)$. Let $L := \sup_{x\in M} \|D_xT\|$, and ϵ be such that $2\epsilon^{1-\alpha}L^\alpha = 1$. Set $B := \{x \in M : \|D_xT\| \geq \epsilon\}$. Pick a cover of B by open sets on which T is injective and let δ be a Lebesgue number for the cover. Thus, if $x, y \in B$ and $d(x,y) \leq \delta$ then $T(x) \neq T(y)$.

For every $n \in \mathbb{N}$, let

$$A := \{x \in M : \#(B \cap \{x, T(x), \dots, T^{n-1}(x)\}) \leq [\alpha n]\},$$

where $[\cdot]$ denotes the integer part function. Observe that $\lim_{n\to\infty} \frac{[\alpha n]}{n} = \alpha$. If $x \in A$ and n is so large that $\epsilon^{1-\frac{[\alpha n]}{n}} L^{\frac{[\alpha n]}{n}} \leq 2\epsilon^{1-\alpha}L^{\alpha}$, then

$$\|D_x T^n\| = \prod_{j=0}^{n-1} \|D_{T^j(x)}T\| < \epsilon^{n-[\alpha n]}L^{[\alpha n]} = \left(\epsilon^{1-\frac{[\alpha n]}{n}} L^{\frac{[\alpha n]}{n}}\right)^n \leq \left(2\epsilon^{1-\alpha}L^{\alpha}\right)^n = 1.$$

Hence, the volume of $T^n(A)$ is less than that of M. But Sard's theorem asserts that the set of singular values has Lebesgue measure zero (for more information, see [30]). Therefore, there exists a regular value x of T^n that lies in $M \setminus T^n(A)$.

We will now extract an (n, δ)-separated set from $T^{-n}(x)$. Since x is regular for T, it has at least $N := |\deg(T)|$ preimages. If at least N of them are in B (a "good transition") then take Q_1 to consist of N such preimages. Otherwise (a "bad transition"), take Q_1 to be a single preimage outside B. Either way, $Q_1 \subseteq T^{-1}(x)$ consists of regular values of T since x is a regular value of T^n. Thus we can apply the same procedure to every $y \in Q_1$ and by collecting all of the points chosen that way obtain $Q_2 \subseteq T^{-2}(x)$, and so on. The set $Q_n \subseteq T^{-n}(x)$ we hence obtain is (n, δ)-separated. Indeed, suppose that $y_1, y_2 \in Q_n$ and $d(T^k(y_1), T^k(y_2)) < \delta$ for all $k \in \{0, \ldots, n-1\}$. Then $T^{n-1}(y_1), T^{n-1}(y_2) \in Q_1$. If $T^{n-1}(y_1) \neq T^{n-1}(y_2)$, then by construction of Q_1 we know that $T^{n-1}(y_1), T^{n-1}(y_2) \in B$. Moreover, $T(T^{n-1}(y_1)) = x = T(T^{n-1}(y_2))$ and $d(T^{n-1}(y_1), T^{n-1}(y_2)) < \delta$. By definition of δ, we deduce that $T^{n-1}(y_1) = T^{n-1}(y_2)$. This contradicts the assumption that $T^{n-1}(y_1) \neq T^{n-1}(y_2)$. So $T^{n-1}(y_1) = T^{n-1}(y_2)$. Likewise $T^{n-2}(y_1) = T^{n-2}(y_2)$, and so forth, so $y_1 = y_2$ and Q_n is (n, δ)-separated.

Now $Q_n \subseteq T^{-n}(x) \subseteq T^{-n}(M \setminus T^n(A)) \subseteq M \setminus A$, that is, $Q_n \cap A = \emptyset$. Thus, for any $y \in Q_n$ there are by definition of A more than αn numbers $k \in \{0, \ldots, n-1\}$ for which $T^k(y) \in B$. So in passing from x to any $y \in Q_n$ there are at least $m := [\alpha n] + 1$ "good transitions," and hence $\#Q_n \geq N^m \geq N^{\alpha n}$. Therefore, the maximal cardinality of an (n, δ)-separated set is at least $N^{\alpha n}$, and thus $\mathrm{h_{top}}(T) \geq \alpha \log N$ by Theorem 7.3.8. Since this holds for all $\alpha \in (0, 1)$, it ensues that $\mathrm{h_{top}}(T) \geq \log N = \log|\deg(T)|$. □

The two properties of smoothness that make the preceding proof work are boundedness of the derivative together with the fact that a smooth map is a local homeomorphism near any point where the derivative is nonzero.

There are certain classes of systems where the inequality given in Theorem 7.5.1 becomes an equality. Expanding maps of a compact manifold (e. g., the unit circle) form one such class. So do rational functions of the Riemann sphere

$$T(z) = \frac{P(z)}{Q(z)},$$

where P and Q are relatively prime polynomials. Since these maps are orientation preserving, their degree is equal to the number of preimages of a regular point $w \in \widehat{\mathbb{C}}$, that is, the number of solutions to the equation $T(z) = w$. The degree of T is therefore equal to the maximum of the algebraic degrees of P and Q.

7.6 Exercises

Exercise 7.6.1. Prove Remark 7.1.3.

Exercise 7.6.2. Prove Lemma 7.1.5.

Exercise 7.6.3. Prove Lemma 7.1.10.

Exercise 7.6.4. Fill in the details of the proof of Lemma 7.1.12.

Exercise 7.6.5. Using Lemmas 7.3.11 and 7.2.20, prove Corollary 7.3.12. Then prove Theorem 7.3.13(b,c,d).

Exercise 7.6.6. Let X be a metric space. Let $(X_n)_{n=1}^{\infty}$ be a descending sequence of nonempty compact subsets of X. Prove that $\bigcap_{n=1}^{\infty} X_n$ is a singleton if and only if

$$\lim_{n\to\infty} \operatorname{diam}(X_n) = 0.$$

Furthermore, show that this result does not generally hold if the sequence is not descending.

Exercise 7.6.7. Let $E = \{0,1\}$. Let $\sigma : E_A^{\infty} \to E_A^{\infty}$ be the golden mean shift, that is, the subshift of finite type induced by the incidence matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

In this exercise, you will prove by induction that $\#(E_A^n) = f_{n+2}$ for all $n \in \mathbb{N}$, where f_n is the nth Fibonacci number.

(a) Show that $E_A^{n+1} = (E_A^n \times \{0\}) \bigcup (E_A^{n,0} \times \{1\})$ for all $n \in \mathbb{N}$, where $E_A^{n,0}$ is the set of all words in E_A^n whose last letter is 0.
(b) Prove that $E_A^{n,0} = E_A^{n-1} \times \{0\}$ for all $n \in \mathbb{N}$.
(c) Deduce that $\#(E_A^{n+1}) = \#(E_A^n) + \#(E_A^{n-1})$ for all $n \in \mathbb{N}$.
(d) Conclude that $\#(E_A^n) = f_{n+2}$ for all $n \in \mathbb{N}$.

Exercise 7.6.8. Prove that the topological entropy of any isometry of a compact metric space is equal to zero.

Exercise 7.6.9. Show that $\mathrm{h_{top}}(T^{-1}) = \mathrm{h_{top}}(T)$ for any homeomorphism $T : X \to X$ of a compact metrizable space X.

Exercise 7.6.10. Prove that the topological entropy of any homeomorphism of the unit circle is equal to zero.

Exercise 7.6.11. Let X be a countable compact metrizable space and $T : X \to X$ a dynamical system. Show that $\mathrm{h_{top}}(T) = 0$.

Exercise 7.6.12. Prove that the topological entropy of every transitive, open, expansive dynamical system is positive.

Exercise 7.6.13. Let $T : X \to X$ be a topological dynamical system and F be a closed forward T-invariant subset of X. Show that the entropy of the subsystem $T|_F : F \to F$ satisfies $\mathrm{h}_{\mathrm{top}}(T|_F) \le \mathrm{h}_{\mathrm{top}}(T)$.

Exercise 7.6.14. Let $T : X \to X$ be a topological dynamical system and $F_1, \ldots, F_n$ be finitely many closed forward T-invariant subsets of X covering X. Prove that $\mathrm{h}_{\mathrm{top}}(T) = \max\{\mathrm{h}_{\mathrm{top}}(T|_{F_i}) : 1 \le i \le n\}$.

Exercise 7.6.15. Find two dynamical systems which have the same topological entropy but are not topologically conjugate.

Exercise 7.6.16. The formula $\mathrm{h}_{\mathrm{top}}(T^n) = n\,\mathrm{h}_{\mathrm{top}}(T)$ may suggest that $\mathrm{h}_{\mathrm{top}}(T \circ S) = \mathrm{h}_{\mathrm{top}}(T) + \mathrm{h}_{\mathrm{top}}(S)$. Show that this is not true in general even if S and T commute.

Exercise 7.6.17. Let $d \in \mathbb{N}$. Prove that the topological entropy of the map of the unit circle $z \mapsto z^d$ is equal to $\log d$.

Exercise 7.6.18. For every dynamical system $T : X \to X$, let

$$\deg(T) := \min_{x \in X} \#T^{-1}(x).$$

Show that if T is a local homeomorphism, then $\mathrm{h}_{\mathrm{top}}(T) \ge \log \deg(T)$.

8 Ergodic theory

In this chapter, we move away from the study of purely topological dynamical systems to consider instead dynamical systems that come equipped with a measure. That is, instead of self-maps acting on compact metrizable spaces, we now ask that the self-maps act upon measure spaces.

The etymology of the word *ergodic* is found in the amalgamation of the two Greek words *ergon* (meaning “work”) and *odos* (meaning “path”). This term was coined by the great physicist Ludwig Boltzmann while carrying out research in statistical mechanics. The goal of ergodic theory is to study the temporal and spatial long-term behavior of a measure-preserving dynamical system. Given such a system and a measurable subset of the space it acts on, it is natural to ask with which frequency the orbits of “typical” points visit that subset. One way to think about ergodic systems is that they are systems such that the visiting frequency of orbits is equal to the measure of the subset visited. In other words, “time averages” are equal to “space averages” for these systems. This will all be made precise shortly.

The chapter is organized as follows. Section 8.1 introduces the basic object of study in ergodic theory, namely, invariant measures. In brief, a measure is said to be invariant under a measurable self-map if the measure of the set of points that are mapped to a measurable subset is equal to the measure of that subset. Section 8.2 presents the notion of ergodicity and comprises a demonstration of Birkhoff’s ergodic theorem, proved by G. D. Birkhoff [9] in 1931. This theorem is the most fundamental result in ergodic theory. It is extremely useful in numerous applications. Birkhoff’s original proof was very involved and complex. Over time some simplifications were brought by several authors. The simple proof we provide here originates from the short and elegant one due to Katok and Hasselblatt [33]. The class of ergodic measures for a given transformation is then studied in more detail. The penultimate Section 8.3 contains an introduction to various measure-theoretic mixing properties that a system may have (which ought to be compared to the topological mixing introduced in Chapter 1). It shows that ergodicity is a very weak form of mixing. In the final Section 8.4, Rokhlin’s construction of an invertible system from any given dynamical system is described and the mixing properties of this natural extension are investigated.

The reader who is not familiar with, or desires a refresher on, measure theory is encouraged to consult Appendix A. We will repeatedly refer to it in this and subsequent chapters.

8.1 Measure-preserving transformations

Throughout, $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ will denote measurable spaces, and the transformation $T : (X, \mathcal{A}) \to (Y, \mathcal{B})$ will be measurable.

https://doi.org/10.1515/9783110702682-008

If the domain of T is endowed with a measure, then the measurable transformation T induces a measure on its codomain.

Definition 8.1.1. Let $T : (X, \mathcal{A}) \to (Y, \mathcal{B})$ be a measurable transformation and let μ be a measure on $(X, \mathcal{A})$. The set function $\mu \circ T^{-1} : \mathcal{B} \to [0, \infty]$, defined by

$$(\mu \circ T^{-1})(B) := \mu(T^{-1}(B)), \quad \forall B \in \mathcal{B},$$

is a measure on $(Y, \mathcal{B})$ called the *push-down or push-forward* of the measure μ under the transformation T.

The integration of a measurable function $f : (Y, \mathcal{B}) \to \mathbb{R}$ with respect to the measure $\mu \circ T^{-1}$ can be carried out by integrating the composition of f and T with respect to the measure μ.

Lemma 8.1.2. *If $T : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B})$ is a measurable transformation, then*

$$\int_Y f \, d(\mu \circ T^{-1}) = \int_X f \circ T \, d\mu$$

for all measurable functions $f : (Y, \mathcal{B}) \to \mathbb{R}$ such that the integral $\int_X f \circ T \, d\mu$ is defined.

Proof. It is easy to see that the equality holds for characteristic functions and, by linearity of the integral, for nonnegative measurable simple functions. The result then follows for any nonnegative measurable function by approaching it pointwise via an increasing sequence of nonnegative measurable simple functions (see Theorem A.1.17 in Appendix A) and calling upon the monotone convergence theorem (Theorem A.1.35). Finally, any measurable function can be expressed as the difference between its positive and negative parts, which are both nonnegative measurable functions. □

Measure-preserving transformations are transformations between measure spaces for which the push down of the measure on the domain coincides with the measure on the codomain.

Definition 8.1.3. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measure spaces. A measurable transformation $T : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ is said to be *measure-preserving* if $\mu \circ T^{-1} = \nu$.

Proving measure preservation for all the elements of a σ-algebra is generally an onerous task. As for equality of measures, when the measures under consideration are finite, it suffices to prove measure preservation on a π-system that generates the σ-algebra on the codomain.

Lemma 8.1.4. *Let $T : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ be a measurable transformation between probability spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$. If $\mathcal{B} = \sigma(\mathcal{P})$ is a σ-algebra generated by a π-system $\mathcal{P}$ on Y, then*

$$T \text{ is measure-preserving} \quad \Longleftrightarrow \quad \mu \circ T^{-1}(P) = \nu(P), \ \forall P \in \mathcal{P}.$$

Proof. This follows immediately from Lemma A.1.26. □

Let us now consider self-transformations, that is, transformations whose codomain coincides with their domain.

Definition 8.1.5. A measure-preserving self-transformation $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$, that is, a measurable self-transformation such that $\mu \circ T^{-1} = \mu$, is called a *measure-preserving dynamical system*. Alternatively, μ is said to be *T-invariant* or *invariant with respect to T*.

Note that if a measurable transformation $T : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ is invertible and its inverse T^{-1} is measurable, then $\mu(T^{-1}(B)) = \nu(B)$ for every $B \in \mathcal{B}$ if and only if $\mu(A) = \nu(T(A))$ for every $A \in \mathcal{A}$. In particular, if $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$, then μ is T-invariant if and only if μ is T^{-1}-invariant. This justifies the following definitions.

Definition 8.1.6. A measure-preserving transformation $T : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$, which is invertible and whose inverse is measurable, is called a *measure-preserving isomorphism*.

Definition 8.1.7. A measure-preserving dynamical system $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$, which is invertible and whose inverse is measurable, is called a *measure-preserving automorphism*.

8.1.1 Examples of invariant measures

In this section, we give several examples of invariant measures for various transformations.

Example 8.1.8. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation with a fixed point x_0, that is, $T(x_0) = x_0$. Let δ_{x_0} be the Dirac point mass supported at x_0 (cf. Example A.1.21). Then δ_{x_0} is T-invariant, that is, $\delta_{x_0}(T^{-1}(A)) = \delta_{x_0}(A)$ for each $A \in \mathcal{A}$, since $x_0 \in T^{-1}(A)$ if and only if $x_0 \in A$. This example easily generalizes to invariant measures supported on periodic orbits.

Example 8.1.9. Let $\mathbb{S}^1 = [0, 2\pi] \pmod{2\pi}$. Let $\alpha \in \mathbb{R}$ and define the map $T_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ by

$$T_\alpha(x) = x + 2\pi\alpha \pmod{2\pi}.$$

Thus T_α is the rotation of the unit circle by the angle $2\pi\alpha$. The topological dynamics of T_α are radically different depending on whether the number α is rational or irrational (see Theorem 1.5.12). So will be the ergodicity of T_α with respect to the Lebesgue measure λ. However, it is fairly easy to foresee that T_α preserves λ, irrespective of the nature of α. Indeed, $T_\alpha^{-1}(x) = x - 2\pi\alpha \pmod{2\pi}$ and so $|\det DT_\alpha^{-1}(x)| = 1$ for all $x \in \mathbb{S}^1$.

Therefore,

$$\lambda(T_\alpha^{-1}(B)) = \int_B |\det DT_\alpha^{-1}(x)|\, d\lambda(x) = \int_B d\lambda(x) = \lambda(B)$$

for all $B \in \mathcal{B}(\mathbb{S}^1)$, that is, T_α preserves λ. Since T_α is invertible and its inverse is measurable, T_α is a Lebesgue measure-preserving automorphism.

Example 8.1.10. Fix $n \in \mathbb{N}$ and consider the map $T_n : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $T_n(x) := nx \pmod 1$, where $\mathbb{S}^1$ is equipped with the Borel σ-algebra $\mathcal{B}(\mathbb{S}^1)$ and the Lebesgue measure λ. We claim that λ is T-invariant. Let I be a proper subinterval of $\mathbb{S}^1$. Then $T_n^{-1}(I)$ consists of n mutually disjoint intervals (arcs) of length $\frac{1}{n}\lambda(I)$. Consequently,

$$\lambda(T_n^{-1}(I)) = n \cdot \frac{1}{n}\lambda(I) = \lambda(I).$$

Since the family of all proper subintervals of $\mathbb{S}^1$ forms a π-system which generates $\mathcal{B}(\mathbb{S}^1)$ and since T_n preserves the Lebesgue measure of all proper subintervals, Lemma 8.1.4 asserts that T_n preserves λ.

Example 8.1.11.

(a) Recall the tent map $T : [0,1] \to [0,1]$ from Example 1.1.3:

$$T(x) := \begin{cases} 2x & \text{if } x \in [0,1/2] \\ 2-2x & \text{if } x \in [1/2,1]. \end{cases}$$

The family of all intervals $\{[a,b),(a,b) : 0 < a < b < 1\}$ forms a π-system that generates the Borel σ-algebra $\mathcal{B}([0,1])$. Since the preimage of any such interval consists of 2 disjoint subintervals (one on each side of the tent) of half the length of the original interval, one readily sees from Lemma 8.1.4 that the Lebesgue measure on $[0,1]$ is invariant under the tent map.

(b) In fact, the previous example generalizes to a much larger family of maps. Let $T : [0,1] \to [0,1]$ be a piecewise linear map of the unit interval that admits a "partition" $\mathcal{P} = \{p_j\}_{j=0}^q$, where $1 \le q < \infty$ and $0 = p_0 < p_1 < \cdots < p_{q-1} < p_q = 1$, with the following properties:

(1) $[0,1] = I_1 \cup \cdots \cup I_q$, where $I_j = [p_{j-1}, p_j]$'s are the successive intervals of monotonicity of T.

(2) $T(I_j) = [0,1]$ for all $1 \le j \le q$.

(3) T is linear on I_j for all $1 \le j \le q$.

Such a map T will be called a *full Markov map*. We claim such a T preserves the Lebesgue measure λ. Indeed, it is easy to see that the absolute value of the slope of the restriction of T to the interval I_j is $1/(p_j - p_{j-1})$. Therefore, the absolute value of the slope of the corresponding inverse branch of T is $p_j - p_{j-1}$. Let $I \subseteq (0,1)$ be any interval. Then $T^{-1}(I) = \bigcup_{j=1}^q T|_{I_j}^{-1}(I)$, where $T|_{I_j}^{-1}(I)$ is a subinterval of $\mathrm{Int}(I_j)$ of length $(p_j - p_{j-1}) \cdot \lambda(I)$. Since $\mathrm{Int}(I_j) \cap \mathrm{Int}(I_k) = \emptyset$ for all $1 \le j < k \le q$, it ensues from that set disjointness (see Lemma A.1.19(g) if necessary) that

$$\lambda(T^{-1}(I)) = \sum_{j=1}^{q} \lambda\big(T|_{I_j}^{-1}(I)\big) = \sum_{j=1}^{q}(p_j - p_{j-1}) \cdot \lambda(I) = (p_q - p_0)\lambda(I) = \lambda(I).$$

Moreover, $0 \le \lambda(T^{-1}(\{0\})) \le \lambda(\{p_j : 0 \le j \le q\}) = 0$. So $\lambda(T^{-1}(\{0\})) = 0 = \lambda(\{0\})$. Similarly, $\lambda(T^{-1}(\{1\})) = 0 = \lambda(\{1\})$. It follows that $\lambda(T^{-1}(J)) = \lambda(J)$ for every interval $J \subseteq [0,1]$. Since the family of all intervals in $[0,1]$ forms a π-system that generates $\mathcal{B}([0,1])$, Lemma 8.1.4 asserts that the Lebesgue measure is invariant under any full Markov map.

Example 8.1.12. Let $T : (X,\mathcal{A}) \to (X,\mathcal{A})$ and $S : (Y,\mathcal{B}) \to (Y,\mathcal{B})$ be measurable transformations for which there exists a measurable transformation $h : (X,\mathcal{A}) \to (Y,\mathcal{B})$ such that $h \circ T = S \circ h$. We will show that every T-invariant measure generates an S-invariant push down under h. Let μ be a T-invariant measure on $(X,\mathcal{A})$. Recall that the push down of μ under h is the measure $\mu \circ h^{-1}$ on $(Y,\mathcal{B})$. It follows from the T-invariance of μ that

$$\big(\mu \circ h^{-1}\big) \circ S^{-1} = \mu \circ (S \circ h)^{-1} = \mu \circ (h \circ T)^{-1} = \big(\mu \circ T^{-1}\big) \circ h^{-1} = \mu \circ h^{-1}.$$

That is, the push down $\mu \circ h^{-1}$ is S-invariant.

Example 8.1.13. Let $(X,\mathcal{A},\mu)$ and $(Y,\mathcal{B},\nu)$ be two probability spaces, and let $T : X \to X$ and $S : Y \to Y$ be two measure-preserving dynamical systems. The direct product of T and S is the map $T \times S : X \times Y \to X \times Y$ defined by

$$(T \times S)(x,y) = \big(T(x), S(y)\big).$$

The direct product σ-algebra $\sigma(\mathcal{A} \times \mathcal{B})$ on $X \times Y$ is the σ-algebra generated by the semialgebra of measurable rectangles

$$\mathcal{A} \times \mathcal{B} := \{A \times B : A \in \mathcal{A},\ B \in \mathcal{B}\}.$$

The direct product measure $\mu \times \nu$ on $(X \times Y, \sigma(\mathcal{A} \times \mathcal{B}))$ is uniquely determined by its values on the generating semialgebra, values which are naturally given by

$$(\mu \times \nu)(A \times B) := \mu(A)\nu(B).$$

The existence and uniqueness of this product measure can be established using Theorem A.1.27, Lemma A.1.29, and Theorem A.1.28. For more information, see Halmos [27] (pp. 157–158) or Taylor [72] (Chapter III, Section 4). We claim that the product map $T \times S : (X \times Y, \sigma(\mathcal{A} \times \mathcal{B}), \mu \times \nu) \to (X \times Y, \sigma(\mathcal{A} \times \mathcal{B}), \mu \times \nu)$ is measure-preserving.

Thanks to Lemma 8.1.4, it suffices to show that $(\mu\times\nu)\circ(T\times S)^{-1}(A\times B) = (\mu\times\nu)(A\times B)$ for all $A\times B\in\mathcal{A}\times\mathcal{B}$. And indeed,

$$\begin{aligned}(\mu\times\nu)\circ(T\times S)^{-1}(A\times B) &= (\mu\times\nu)\big(T^{-1}(A)\times S^{-1}(B)\big)\\ &= \mu\big(T^{-1}(A)\big)\nu\big(S^{-1}(B)\big)\\ &= \mu(A)\nu(B)\\ &= (\mu\times\nu)(A\times B).\end{aligned}$$

The final example pertains to the shift map introduced in Chapter 3 on symbolic dynamics. In fact, we look at this map in a more general context.

Example 8.1.14. Let $(E,\mathcal{F},P)$ be a probability space. Consider the one-sided product set $E^{\mathbb{N}} = E^{\infty} := \prod_{k=1}^{\infty} E$. The product σ-algebra $\mathcal{F}_\infty$ on E^∞ is the σ-algebra generated by the semialgebra $\mathcal{S}$ of all (finite) cylinders (also called rectangles), that is, sets of the form

$$\prod_{k=1}^{n} E_k \times \prod_{l=n+1}^{\infty} E = \{\omega = (\omega_j)_{j=1}^{\infty}\in E^\infty : \omega_k\in E_k,\ \forall 1\le k\le n\},$$

where $n\in\mathbb{N}$ and $E_k\in\mathcal{F}$ for all $1\le k\le n$. The product measure μ_P on $\mathcal{F}_\infty$ is the unique probability measure which confers to a cylinder the value

$$\mu_P\Bigg(\prod_{k=1}^{n} E_k \times \prod_{l=n+1}^{\infty} E\Bigg) = \prod_{k=1}^{n} P(E_k). \tag{8.1}$$

The existence and uniqueness of this measure can be established using Theorem A.1.27, Lemma A.1.29, and Theorem A.1.28. For more information, see Halmos [27] (pp. 157–158) or Taylor [72] (Chapter III, Section 4).

As in Chapter 3, let $\sigma : E^\infty \to E^\infty$ be the left shift map, which is defined by $\sigma((\omega_n)_{n=1}^{\infty}) := (\omega_{n+1})_{n=1}^{\infty}$. The product measure μ_P is σ-invariant. Indeed, since the cylinder sets form a semialgebra which generates the product σ-algebra, in light of Lemma 8.1.4 it is sufficient to show that $\mu_P(\sigma^{-1}(S)) = \mu_P(S)$ for all cylinder sets $S\in\mathcal{S}$. And we have

$$\begin{aligned}\mu_P\circ\sigma^{-1}\Bigg(\prod_{k=1}^{n} E_k \times \prod_{l=n+1}^{\infty} E\Bigg) &= \mu_P\Bigg(E\times\prod_{k=1}^{n} E_k \times \prod_{l=n+2}^{\infty} E\Bigg)\\ &= P(E)\prod_{k=1}^{n} P(E_k)\\ &= \prod_{k=1}^{n} P(E_k)\\ &= \mu_P\Bigg(\prod_{k=1}^{n} E_k \times \prod_{l=n+1}^{\infty} E\Bigg).\end{aligned}$$

This completes the proof that the product measure is shift-invariant.

The measure-preserving dynamical system $(\sigma : E^\infty \to E^\infty, \mu_P)$ is commonly referred to as a *one-sided Bernoulli shift* with set of states E. Of particular importance is the case when E is a finite set having at least two elements. Also, the case of a countably infinite set of states E will be of special importance in this book. This will be particularly transparent in Chapters 13 and 16 onward in the second volume, where we will consider Gibbs states of Hölder continuous potentials for which Bernoulli measures are very special cases.

More examples of invariant measures can be found in Exercises 8.5.22 and 8.5.25–8.5.33.

8.1.2 Poincaré's recurrence theorem

We now present one of the fundamental results of finite ergodic theory, namely, Poincaré's recurrence theorem. This theorem states that, in a finite measure space, almost all points of a given set return infinitely often to that set under iteration. It is worth pointing out that Poincaré's recurrence theorem is striking (and, as we will see, unusual), in that its hypotheses are so completely general.

But, first, we show that the points from a measurable set that never return to that set under iteration are negligible. That is, they form a subset of measure zero.

Lemma 8.1.15. *If $T : X \to X$ is a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, then for every set $A \in \mathcal{A}$ we have*

$$\mu(\{x \in A : T^n(x) \notin A, \ \forall n \in \mathbb{N}\}) = 0.$$

Proof. Let $A \in \mathcal{A}$ and

$$N = N(T, A) := \{x \in A : T^n(x) \notin A, \ \forall n \in \mathbb{N}\}.$$

To show that $\mu(N) = 0$, let $x \in N$. Then $T^n(x) \notin A$ for every $n \in \mathbb{N}$. Therefore, $T^n(x) \notin N$ for all $n \in \mathbb{N}$. Thus $N \cap T^{-n}(N) = \emptyset$ for all $n \in \mathbb{N}$. Now, fix $k \in \mathbb{N}$ and let $1 \le j < k$. Then

$$T^{-j}(N) \cap T^{-k}(N) = T^{-j}(N \cap T^{-(k-j)}(N)) = T^{-j}(\emptyset) = \emptyset.$$

So the preimages $\{T^{-n}(N)\}_{n=0}^\infty$ of N under the iterates of T form a pairwise disjoint family of sets. It follows that

$$1 = \mu(X) \ge \mu\left(\bigcup_{n=0}^\infty T^{-n}(N)\right) = \sum_{n=0}^\infty \mu(T^{-n}(N)) = \sum_{n=0}^\infty \mu(N),$$

where the last equality follows from the T-invariance of μ. Hence, $\mu(N) = 0$. □

Knowing this, we can now demonstrate that, in a finite measure space, almost all points of a given set return infinitely often to that set under iteration.

Theorem 8.1.16 (Poincaré's recurrence theorem). *If $T : X \to X$ is a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, then*

$$\mu(\{x \in A : T^n(x) \in A \text{ for infinitely many } n \in \mathbb{N}\}) = \mu(A)$$

for every set $A \in \mathcal{A}$.

Proof. Let

$$N(T, A) := \{x \in A : T^n(x) \notin A,\ \forall n \in \mathbb{N}\}.$$

For each $k \in \mathbb{N}$, let

$$\begin{aligned} N_k &:= \{x \in A : T^n(x) \notin A,\ \forall n \geq k\} \\ &\subseteq \{x \in A : T^{kj}(x) \notin A,\ \forall j \in \mathbb{N}\} = N(T^k, A). \end{aligned}$$

Replacing T by T^k in Lemma 8.1.15, we obtain that $\mu(N(T^k, A)) = 0$ and thereby $\mu(N_k) = 0$ for all $k \in \mathbb{N}$. It follows that

$$\mu\left(\bigcup_{k=1}^{\infty} N_k\right) = 0.$$

Observe also that

$$\{x \in A : T^n(x) \in A \text{ for infinitely many } n \in \mathbb{N}\} = A \setminus \bigcup_{k=1}^{\infty} N_k.$$

Consequently,

$$\mu(\{x \in A : T^n(x) \in A \text{ for infinitely many } n\}) = \mu(A) - \mu\left(\bigcup_{k=1}^{\infty} N_k\right) = \mu(A). \qquad \square$$

Note that this result does not generally hold in infinite measure spaces. Indeed, the simplest counterexample is a translation of the real line. For instance, take the transformation of the real line $T(x) = x + 1$, which certainly preserves the Lebesgue measure, and let $A = (0, 1)$. No point of A ever comes back to A under iteration, although the Lebesgue measure of A is evidently not equal to zero.

8.1.3 Existence of invariant measures

In general, a measurable transformation $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ may not admit any invariant measure (see Exercise 8.5.29). There are also measurable transformations that admit infinite invariant measures but not finite ones (see Exercise 8.5.30). However, if

X is a compact metrizable space and $\mathcal{A}$ is the Borel σ-algebra $\mathcal{B}(X)$, then every continuous transformation does have an invariant Borel probability measure. In other words, every topological dynamical system admits an invariant Borel probability measure. Before proving this, we will study the properties of the set of invariant probability measures.

Definition 8.1.17. Let $(X, \mathcal{A})$ be a measurable space. The set of all probability measures on $(X, \mathcal{A})$ is denoted by $M(X, \mathcal{A})$. Given a measurable transformation $T : (X, \mathcal{A}) \to (X, \mathcal{A})$, the subset of all T-invariant probability measures on $(X, \mathcal{A})$ is denoted by $M(T, \mathcal{A})$.

In particular, if $\mathcal{A}$ is the Borel σ-algebra on a topological space X, then the set of all Borel probability measures on X is simply denoted by $M(X)$ instead of $M(X, \mathcal{B}(X))$, while its subset of T-invariant measures is denoted by $M(T)$ rather than $M(T, \mathcal{B}(X))$. For more information about $M(X)$, please see Subsection A.1.8.

For topological dynamical systems, there exists a characterization of invariant Borel probability measures in terms of the way they integrate continuous functions.

Theorem 8.1.18. *Let $T : X \to X$ be a topological dynamical system. Then*

$$\mu \in M(T) \quad \Longleftrightarrow \quad \int_X f \circ T \, d\mu = \int_X f \, d\mu, \quad \forall f \in C(X).$$

Proof. This follows immediately from Lemma 8.1.2 and Corollary A.1.54. □

We will now show that the set $M(T)$ is a compact and convex subset of $M(X)$ whenever T is a topological dynamical system.

In fact, the convexity holds for all measurable transformations.

Lemma 8.1.19. *For any measurable transformation $T : (X, \mathcal{A}) \to (X, \mathcal{A})$, the set $M(T, \mathcal{A})$ is a convex subset of $M(X, \mathcal{A})$.*

Proof. Let $\mu, \nu \in M(T, \mathcal{A})$. Let m be a convex combination of μ and ν, that is, let $\alpha \in [0, 1]$ and let $m = \alpha\mu + (1 - \alpha)\nu$. By the obvious convexity of $M(X, \mathcal{A})$, we already know that $m \in M(X, \mathcal{A})$. Let $A \in \mathcal{A}$. Observe that

$$m(T^{-1}(A)) = \alpha\mu(T^{-1}(A)) + (1 - \alpha)\nu(T^{-1}(A)) = \alpha\mu(A) + (1 - \alpha)\nu(A) = m(A).$$

Thus, m is T-invariant, and hence $m \in M(T, \mathcal{A})$. □

Theorem 8.1.20. *Let $T : X \to X$ be a topological dynamical system. The set $M(T)$ is a compact convex subset of the compact convex space $M(X)$ in the* weak* *topology.*

Proof. The convexity of $M(T)$ has been established in Lemma 8.1.19. Thus, we can concentrate on demonstrating the compactness of $M(T)$. According to Theorem A.1.58, the set $M(X)$ is compact in the weak* topology of $C(X)^*$. Hence, it suffices to show that

$M(T)$ is closed in $M(X)$. As described in Subsection A.1.8, the space $M(X)$ admits a metric compatible with the weak* topology. Therefore, a set is closed in that space if and only if it is sequentially closed. Let $(\mu_n)_{n=1}^{\infty}$ be a sequence in $M(T)$ which converges to a measure μ in $M(X)$. We aim to show that $\mu \in M(T)$. To that end, let $f \in C(X)$. According to Theorem 8.1.18, it suffices to show that $\int_X f \circ T\, d\mu = \int_X f\, d\mu$. Since $(\mu_n)_{n=1}^{\infty}$ converges to μ in the weak* topology and $\mu_n \in M(T)$ for all n, we deduce that

$$\int_X f \circ T\, d\mu = \lim_{n\to\infty} \int_X f \circ T\, d\mu_n = \lim_{n\to\infty} \int_X f\, d\mu_n = \int_X f\, d\mu.$$

Thus $\mu \in M(T)$ and $M(T)$ is closed in $M(X)$. As a closed subset of the compact set $M(X)$, the set $M(T)$ is compact as well. □

We now briefly examine the map $\mu \mapsto \mu \circ T^{-1}$, which will be helpful in proving the existence of invariant measures.

Lemma 8.1.21. *Let $T : X \to X$ be a topological dynamical system. The map $S : M(X) \to M(X)$, where $S(\mu) = \mu \circ T^{-1}$, is continuous and affine.*

Proof. The proof of affinity is left to the reader. We concentrate on the continuity of S. Let $(\mu_n)_{n=1}^{\infty}$ be a sequence in $M(X)$ which weak* converges to μ. Then, for any $f \in C(X)$,

$$\begin{aligned}\lim_{n\to\infty} \int_X f\, d(S(\mu_n)) &= \lim_{n\to\infty} \int_X f\, d(\mu_n \circ T^{-1}) \\ &= \lim_{n\to\infty} \int_X f \circ T\, d\mu_n \\ &= \int_X f \circ T\, d\mu = \int_X f\, d(\mu \circ T^{-1}) \\ &= \int_X f\, d(S(\mu)).\end{aligned}$$

Since f was chosen arbitrarily in $C(X)$, the sequence $(S(\mu_n))_{n=1}^{\infty}$ weak* converges to $S(\mu)$. Thus S is continuous. □

We come now to the main result of this section, namely, showing that every topological dynamical system admits at least one invariant Borel probability measure. This theorem is not very difficult to prove, but it is obviously important. For this reason, we provide two different proofs. The first involves functional analysis, whereas the second is rather more constructive.

Theorem 8.1.22 (Krylov–Bogolyubov theorem). *Let $T : X \to X$ be a topological dynamical system. Then $M(T) \neq \emptyset$.*

Proof. By the Riesz representation theorem (Theorem A.1.53), the set $M(X)$ can be identified with a subset of the Banach space $C(X)^*$. According to Theorem A.1.58, the set $M(X)$ is compact and convex in the weak* topology of $C(X)^*$. By Lemma 8.1.21, we also know that the map $S(\mu) := \mu \circ T^{-1}$ is a continuous affine self-map of $M(X)$. Thus, by Schauder–Tychonoff's fixed-point theorem (cf. Theorem V.10.5 in Dunford and Schwartz [20]) the map S has a fixed point. In other words, there exists $\mu \in M(X)$ such that $\mu \circ T^{-1} = \mu$. Note: Alternatively, since S is affine, one may use Markov-Kakutani's fixed-point theorem. It is more elementary and proved in Theorem V.10.6 of [20].

Alternative proof. Let $\mu_0 \in M(X)$ (for example, a Dirac point mass supported at a point of X). Construct the sequence of Borel probability measures $(\mu_n)_{n=1}^{\infty}$, where

$$\mu_n = \frac{1}{n}\sum_{j=0}^{n-1} \mu_0 \circ T^{-j}.$$

Since $M(X)$ is compact in the weak* topology, the sequence $(\mu_n)_{n=1}^{\infty}$ has at least one weak* limit point. Denote such a point by μ_∞. We claim that $\mu_\infty \in M(T)$. To show this, let $(\mu_{n_k})_{k=1}^{\infty}$ be a subsequence of the sequence $(\mu_n)_{n=1}^{\infty}$ which weak* converges to μ_∞. The weak* convergence of the subsequence means that

$$\int_X f\, d\mu_\infty = \lim_{k\to\infty} \int_X f\, d\mu_{n_k}, \quad \forall f \in C(X).$$

Moreover,

$$\begin{aligned}
\left|\int_X f\circ T\, d\mu_{n_k} - \int_X f\, d\mu_{n_k}\right| &= \left|\frac{1}{n_k}\sum_{j=0}^{n_k-1}\int_X f\circ T\, d(\mu_0\circ T^{-j}) - \frac{1}{n_k}\sum_{j=0}^{n_k-1}\int_X f\, d(\mu_0\circ T^{-j})\right| \\
&= \frac{1}{n_k}\left|\sum_{j=0}^{n_k-1}\left(\int_X f\circ T^{j+1}\, d\mu_0 - \int_X f\circ T^{j}\, d\mu_0\right)\right| \\
&= \frac{1}{n_k}\left|\int_X f\circ T^{n_k}\, d\mu_0 - \int_X f\, d\mu_0\right| \\
&\le \frac{2}{n_k}\|f\|_\infty.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left|\int_X f\circ T\, d\mu_\infty - \int_X f\, d\mu_\infty\right| &= \left|\lim_{k\to\infty}\int_X f\circ T\, d\mu_{n_k} - \lim_{k\to\infty}\int_X f\, d\mu_{n_k}\right| \\
&= \lim_{k\to\infty}\left|\int_X f\circ T\, d\mu_{n_k} - \int_X f\, d\mu_{n_k}\right| \\
&\le 2\|f\|_\infty \lim_{k\to\infty}\frac{1}{n_k} = 0.
\end{aligned}$$

Hence,

$$\int_X f\, d(\mu_\infty \circ T^{-1}) = \int_X f \circ T\, d\mu_\infty = \int_X f\, d\mu_\infty, \quad \forall f \in C(X).$$

By Corollary A.1.54, we conclude that $\mu_\infty \circ T^{-1} = \mu_\infty$. □

The Krylov–Bogolyubov theorem can be restated as follows: Every topological dynamical system induces at least one measure-preserving dynamical system.

8.2 Ergodic transformations

One of the aims of the present section is to state and demonstrate the first published ergodic theorem, originally proved at the outset of the 1930s by George David Birkhoff. Of course, before setting out to prove an ergodic theorem, we must first define and investigate the notion of ergodicity.

Definition 8.2.1. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. Then T is said to be *ergodic* with respect to a measure μ on $(X, \mathcal{A})$ if all completely T-invariant sets $A \in \mathcal{A}$, that is, such that $T^{-1}(A) = A$, have the property that $\mu(A) = 0$ or $\mu(X \backslash A) = 0$. Alternatively, μ is said to be T-ergodic or ergodic with respect to T.

A system is ergodic if and only if it does not admit any nontrivial subsystem. See alternative definitions in Exercise 8.5.36.

The following is a simple but important observation.

Lemma 8.2.2. *If a measure μ is ergodic with respect to a measurable transformation $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ and if a measure ν on $(X, \mathcal{A})$ is absolutely continuous with respect to μ, then ν is also ergodic. That is,*

$$\mu \text{ ergodic } \ \& \ \nu \ll \mu \quad \Longrightarrow \quad \nu \text{ ergodic}.$$

Proof. The proof is left to the reader as an exercise. □

While complete invariance of a set is an appropriate concept for topological dynamical systems, we will see that a more suitable notion for measure-preserving dynamical systems is that of almost invariance of a set.

Definition 8.2.3. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and let μ be a measure on $(X, \mathcal{A})$. A set $A \in \mathcal{A}$ is said to be *μ-almost T-invariant* if $\mu(T^{-1}(A) \triangle A) = 0$.

Of course, any completely T-invariant set is μ-almost T-invariant.

Our next goal is to show that a measure μ is ergodic if and only if all μ-almost T-invariant sets are trivial in a measure-theoretic sense, that is, have measure zero or full measure. The proof of this characterization of ergodic measures boils down to constructing a completely T-invariant set from an almost T-invariant one. This raises the more general question: Given an arbitrary set S, how can we construct from that set a completely T-invariant one?

If a set R is forward T-invariant, that is, if $T^{-1}(R) \supseteq R$, then a related completely T-invariant set is the union of all the preimages of R, that is, the set $\bigcup_{k=0}^{\infty} T^{-k}(R)$ of all points whose orbits eventually hit R. This reduces our question to the following one: Given a set S, how can we construct from it a forward T-invariant set R? One possibility is the intersection of all the preimages of S, that is, the set $R = \bigcap_{n=0}^{\infty} T^{-n}(S)$ of all points whose orbits are trapped within S. Hence, the set

$$\bigcup_{k=0}^{\infty} T^{-k}\left(\bigcap_{n=0}^{\infty} T^{-n}(S)\right) = \bigcup_{k=0}^{\infty}\bigcap_{n=0}^{\infty} T^{-(k+n)}(S) = \bigcup_{k=0}^{\infty}\bigcap_{n=k}^{\infty} T^{-n}(S)$$

is completely T-invariant. This is the set of all points whose iterates eventually fall into S and remain trapped there forever.

Similarly, if a set R is backward T-invariant, that is, if $T^{-1}(R) \subseteq R$, then an obvious candidate for a completely T-invariant set is the intersection of all the preimages of R, namely $\bigcap_{k=0}^{\infty} T^{-k}(R)$. This reduces our original question to the following: Given a set S, how can we construct from it a backward T-invariant set R? The union of all the preimages of S, namely $R = \bigcup_{n=0}^{\infty} T^{-n}(S)$, is such a set. Hence, the set

$$\bigcap_{k=0}^{\infty} T^{-k}\left(\bigcup_{n=0}^{\infty} T^{-n}(S)\right) = \bigcap_{k=0}^{\infty}\bigcup_{n=0}^{\infty} T^{-(k+n)}(S) = \bigcap_{k=0}^{\infty}\bigcup_{n=k}^{\infty} T^{-n}(S)$$

is completely T-invariant. This is the set of all points whose orbits visit S infinitely many times.

Proposition 8.2.4. *Let $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ be a measure-preserving dynamical system. Then T is ergodic with respect to μ if and only if all μ-almost T-invariant sets $A \in \mathcal{A}$ satisfy $\mu(A) = 0$ or $\mu(X\backslash A) = 0$.*

Proof. Suppose that all μ-almost T-invariant sets $A \in \mathcal{A}$ satisfy $\mu(A) = 0$ or $\mu(X\backslash A) = 0$. Let $A' \in \mathcal{A}$ be any completely T-invariant set. Since every completely T-invariant set is μ-almost T-invariant, it ensues that $\mu(A') = 0$ or $\mu(X\backslash A) = 0$. Thus T is ergodic.

We shall now prove the converse implication. Though μ is T-invariant, it is sufficient that $\mu \circ T^{-1} \ll \mu$ in the following proof (μ is then said to be quasi-T-invariant; see Definition 10.1.1). Suppose that T is ergodic and let $A \in \mathcal{A}$ be a μ-almost T-invariant set, that is, A is such that $\mu(T^{-1}(A) \triangle A) = 0$. We must show that $\mu(A) = 0$ or $\mu(X\backslash A) = 0$.

Claim 1. $\mu(T^{-n}(A) \triangle A) = 0$ *for all* $n \geq 0$.

Proof of Claim 1. Since $\mu(T^{-1}(A) \triangle A) = 0$, since μ is T-invariant and since $f^{-1}(C \triangle D) = f^{-1}(C) \triangle f^{-1}(D)$ for any map f and any sets C and D, we have for all $k \in \mathbb{N}$ that

$$\mu\big(T^{-(k+1)}(A) \triangle T^{-k}(A)\big) = \mu\big(T^{-k}(T^{-1}(A) \triangle A)\big) = \mu\big(T^{-1}(A) \triangle A\big) = 0.$$

As $C \triangle D \subseteq (C \triangle E) \cup (E \triangle D)$ for any sets C, D and E, it follows for all $n \in \mathbb{N}$ that

$$\mu(T^{-n}(A) \triangle A) \le \mu\left(\bigcup_{k=0}^{n-1} [T^{-(k+1)}(A) \triangle T^{-k}(A)]\right)$$
$$\le \sum_{k=0}^{n-1} \mu(T^{-(k+1)}(A) \triangle T^{-k}(A)) = 0. \qquad \square$$

Claim 2. *The set*

$$B := \bigcup_{k=0}^{\infty} \bigcap_{n=k}^{\infty} T^{-n}(A)$$

is completely T-invariant, and thus $\mu(B) = 0$ or $\mu(X \backslash B) = 0$.

Proof of Claim 2. Indeed,

$$T^{-1}(B) = \bigcup_{k=0}^{\infty} \bigcap_{n=k}^{\infty} T^{-(n+1)}(A) = \bigcup_{k=0}^{\infty} \bigcap_{n=k+1}^{\infty} T^{-n}(A) = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} T^{-n}(A) = B.$$

Since T is ergodic, we deduce that $\mu(B) = 0$ or $\mu(X \backslash B) = 0$. $\square$

Claim 3. $\mu(B \triangle A) = 0 = \mu((X \backslash B) \triangle (X \backslash A))$.

Proof of Claim 3. To prove this, we will use Claim 1 and two properties of the symmetric difference operation: $(\bigcup_{i \in I} C_i) \triangle D \subseteq \bigcup_{i \in I}(C_i \triangle D)$ and $(\bigcap_{i \in I} C_i) \triangle D \subseteq \bigcup_{i \in I}(C_i \triangle D)$. Indeed,

$$\begin{aligned}
\mu(B \triangle A) &= \mu\left(\left[\bigcup_{k=0}^{\infty} \bigcap_{n=k}^{\infty} T^{-n}(A)\right] \triangle A\right) \\
&\le \mu\left(\bigcup_{k=0}^{\infty}\left[\left(\bigcap_{n=k}^{\infty} T^{-n}(A)\right) \triangle A\right]\right) \\
&\le \mu\left(\bigcup_{k=0}^{\infty} \bigcup_{n=k}^{\infty} (T^{-n}(A) \triangle A)\right) \\
&= \mu\left(\bigcup_{n=0}^{\infty} (T^{-n}(A) \triangle A)\right) \\
&\le \sum_{n=0}^{\infty} \mu(T^{-n}(A) \triangle A) \\
&= 0.
\end{aligned}$$

Since $(X \backslash B) \triangle (X \backslash A) = B \triangle A$, Claim 3 is proved. $\square$

Claim 4. $\mu(A) = \mu(B)$ *and* $\mu(X \backslash A) = \mu(X \backslash B)$.

Proof of Claim 4. Since $B \triangle A = (B \backslash A) \cup (A \backslash B)$, it immediately follows from Claim 3 that $\mu(B\backslash A) = 0 = \mu(A\backslash B)$. Therefore,

$$\mu(B) = \mu\big((B\backslash A) \cup (B \cap A)\big) = \mu(B\backslash A) + \mu(B \cap A) = \mu(B \cap A) \leq \mu(A).$$

Likewise, $\mu(A) \leq \mu(B)$. Thus $\mu(A) = \mu(B)$. Similarly, $\mu(X\backslash A) = \mu(X\backslash B)$. □

To conclude the proof of the proposition, Claims 2 and 4 assert that $\mu(A) = \mu(B) = 0$ or $\mu(X\backslash A) = \mu(X\backslash B) = 0$. Finally, note that we could equally well have chosen the set $B = \bigcap_{k=0}^{\infty} \bigcup_{n=k}^{\infty} T^{-n}(A)$ in Claim 2. □

It is not difficult to check that the family $\{A \in \mathcal{A} \mid T^{-1}(A) = A\}$ of all completely T-invariant sets forms a sub-σ-algebra of $\mathcal{A}$. So does the family $\{A \mid \mu(T^{-1}(A) \triangle A) = 0\}$ of all μ-almost T-invariant sets, a fact which we shall prove shortly. Of course, the former is a smaller σ-algebra. However, it is not sufficiently flexible for our measure-theoretic purposes, as it is only defined set-theoretically. For this reason, we shall usually work with the latter.

Definition 8.2.5. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and let μ be a measure on $(X, \mathcal{A})$. The collection of all μ-almost T-invariant sets shall be denoted by

$$\mathcal{I}_\mu := \{A \in \mathcal{A} \,\big|\, \mu\big(T^{-1}(A) \triangle A\big) = 0\}.$$

Proposition 8.2.6. *The family $\mathcal{I}_\mu$ is a sub-σ-algebra of $\mathcal{A}$. Furthermore, if μ is T-invariant then $T^{-1}(\mathcal{I}_\mu) \subseteq \mathcal{I}_\mu$.*

Proof. It is clear that $\emptyset \in \mathcal{I}_\mu$.

We now show that $\mathcal{I}_\mu$ is closed under the operation of complementation. Let $A \in \mathcal{I}_\mu$. Then

$$\begin{aligned}
T^{-1}(X\backslash A) \triangle (X\backslash A) &= \big(X\backslash T^{-1}(A)\big) \triangle (X\backslash A) \\
&= \big((X\backslash T^{-1}(A))\backslash(X\backslash A)\big) \cup \big((X\backslash A)\backslash(X\backslash T^{-1}(A))\big) \\
&= \big(A\backslash T^{-1}(A)\big) \cup \big(T^{-1}(A)\backslash A\big) \\
&= T^{-1}(A) \triangle A.
\end{aligned}$$

Thus $\mu(T^{-1}(X\backslash A) \triangle (X\backslash A)) = \mu(T^{-1}(A) \triangle A) = 0$, and hence $X\backslash A \in \mathcal{I}_\mu$.

It only remains to show that $\mathcal{I}_\mu$ is closed under countable unions, that is, we must show that if $(A_n)_{n=1}^{\infty}$ is a sequence in $\mathcal{I}_\mu$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{I}_\mu$. For this, observe that

$$\begin{aligned}
\left[T^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right)\right] \triangle \left[\bigcup_{n=1}^{\infty} A_n\right] &= \left[\bigcup_{n=1}^{\infty} T^{-1}(A_n)\right] \triangle \left[\bigcup_{n=1}^{\infty} A_n\right] \\
&= \left[\bigcup_{n=1}^{\infty} T^{-1}(A_n) \Big\backslash \bigcup_{n=1}^{\infty} A_n\right] \bigcup \left[\bigcup_{n=1}^{\infty} A_n \Big\backslash \bigcup_{n=1}^{\infty} T^{-1}(A_n)\right]
\end{aligned}$$

$$\subseteq \left[\bigcup_{n=1}^{\infty}(T^{-1}(A_n)\backslash A_n)\right]\bigcup\left[\bigcup_{n=1}^{\infty}(A_n\backslash T^{-1}(A_n))\right]$$
$$= \bigcup_{n=1}^{\infty}(T^{-1}(A_n) \triangle A_n).$$

Consequently,

$$\mu\left(\left[T^{-1}\left(\bigcup_{n=1}^{\infty}A_n\right)\right] \triangle \left[\bigcup_{n=1}^{\infty}A_n\right]\right) \le \sum_{n=1}^{\infty}\mu(T^{-1}(A_n) \triangle A_n) = \sum_{n=1}^{\infty} 0 = 0.$$

So $\bigcup_{n=1}^{\infty} A_n \in \mathcal{I}_\mu$ and $\mathcal{I}_\mu$ is a σ-algebra.

If μ is T-invariant (in fact, it suffices that μ be quasi-T-invariant, i. e. $\mu \circ T^{-1} \ll \mu$) and if $A \in \mathcal{I}_\mu$, then

$$\mu(T^{-1}(T^{-1}(A)) \triangle T^{-1}(A)) = \mu(T^{-1}(T^{-1}(A) \triangle A)) = \mu(T^{-1}(A) \triangle A) = 0.$$

That is, $T^{-1}(A) \in \mathcal{I}_\mu$. Thus $T^{-1}(\mathcal{I}_\mu) \subseteq \mathcal{I}_\mu$. □

We have already discussed invariant sets, measure-theoretically invariant sets and invariant measures. We now introduce invariant and measure-theoretically invariant functions.

Definition 8.2.7. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation, let μ be a measure on $(X, \mathcal{A})$ and let $\varphi : (X, \mathcal{A}) \to \mathbb{R}$ be a measurable function.

(a) The function φ is said to be *T-invariant* if $\varphi \circ T = \varphi$.

(b) The function φ is called *μ-a. e. T-invariant* if $\varphi \circ T = \varphi$ μ-almost everywhere. In other words, φ is μ-a. e. T-invariant if the measurable set

$$D_\varphi := \{x \in X \mid \varphi(T(x)) \neq \varphi(x)\}$$

is a null set.

Lemma 8.2.8. *Let $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ be a measure-preserving dynamical system. A measurable function $\varphi : X \to \mathbb{R}$ is μ-a. e. T-invariant if and only if φ is constant on the forward orbit of μ-a. e. $x \in X$.*

Proof. If φ is constant on the forward orbit of μ-a. e. $x \in X$, then $\varphi(T(x)) = \varphi(x)$ for μ-a. e. $x \in X$, that is, φ is μ-a. e. T-invariant.

To prove the converse, suppose that φ is μ-a. e. T-invariant. For every $n \in \mathbb{N}$, we have

$$\{y \in X : \varphi(T^n(y)) \neq \varphi(T^{n-1}(y))\} = T^{-(n-1)}(\{x \in X : \varphi(T(x)) \neq \varphi(x)\}).$$

Since μ is T-invariant (in fact, it suffices that μ be quasi-T-invariant) and since φ is μ-a. e. T-invariant, this implies that for every $n \in \mathbb{N}$,

$$\mu(\{y \in X : \varphi(T^n(y)) \neq \varphi(T^{n-1}(y))\}) = 0.$$

Therefore,

$$\begin{aligned}\mu(\{x \in X : \varphi \text{ not constant over } \mathcal{O}_+(x)\}) &= \mu\Big(\bigcup_{n=1}^{\infty}\{x \in X : \varphi(T^n(x)) \neq \varphi(T^{n-1}(x))\}\Big) \\ &\leq \sum_{n=1}^{\infty}\mu(\{x \in X : \varphi(T^n(x)) \neq \varphi(T^{n-1}(x))\}) \\ &= 0.\end{aligned}$$

So φ is constant on the forward orbit of μ-a. e. $x \in X$. □

The following lemma shows that measure-theoretically invariant functions are characterized by the fact that they are measurable with respect to the σ-algebra of measure-theoretically invariant sets.

Lemma 8.2.9. *Let $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ be a measure-preserving dynamical system. A function $\varphi : X \to \mathbb{R}$ is μ-a. e. T-invariant if and only if φ is measurable with respect to the σ-algebra $\mathcal{I}_\mu$.*

Proof. First, suppose that φ is μ-a. e. T-invariant. Let $B \subseteq \mathbb{R}$ be a Borel set. In order for φ to be $\mathcal{I}_\mu$-measurable, we need to show that $\varphi^{-1}(B) \in \mathcal{I}_\mu$. To begin, notice that if

$$x \in T^{-1}(\varphi^{-1}(B)) \triangle (\varphi^{-1}(B)) = (\varphi \circ T)^{-1}(B) \triangle (\varphi^{-1}(B)),$$

then only one of the real numbers $\varphi(T(x))$ or $\varphi(x)$ belongs to the set B. Thus $\varphi(T(x)) \neq \varphi(x)$ and $x \in D_\varphi$. This means that

$$T^{-1}(\varphi^{-1}(B)) \triangle (\varphi^{-1}(B)) \subseteq D_\varphi.$$

Consequently, $\mu(T^{-1}(\varphi^{-1}(B)) \triangle (\varphi^{-1}(B))) \leq \mu(D_\varphi) = 0$. So $\varphi^{-1}(B) \in \mathcal{I}_\mu$.

To prove the converse implication, suppose by way of contradiction that φ is $\mathcal{I}_\mu$-measurable but that $\mu(D_\varphi) > 0$. We can always write

$$D_\varphi = \bigcup_{a \in \mathbb{Q}}\Big(\{x \in X : \varphi(x) < a < \varphi(T(x))\} \cup \{x \in X : \varphi(x) > a > \varphi(T(x))\}\Big).$$

This is a countable union of $\mathcal{A}$-measurable sets with positive total measure. Hence, without loss of generality, there exists some $a \in \mathbb{Q}$ such that the set

$$B_a := \{x \in X \mid \varphi(x) < a < \varphi(T(x))\}$$

is of positive measure (if not, replace φ by $-\varphi$). Observe that

$$B_a = \varphi^{-1}((-\infty, a)) \cap (\varphi \circ T)^{-1}((a, \infty)).$$

Note that $\varphi^{-1}((-\infty, a)) \in \mathcal{I}_\mu$ since φ is $\mathcal{I}_\mu$-measurable by assumption. Moreover, $(\varphi \circ T)^{-1}((a, \infty)) = T^{-1}(\varphi^{-1}((a, \infty))) \in \mathcal{I}_\mu$ since φ is $\mathcal{I}_\mu$-measurable and $T^{-1}(\mathcal{I}_\mu) \subseteq \mathcal{I}_\mu$ per

Proposition 8.2.6. Thus $B_a \in \mathcal{I}_\mu$. Now, notice that

$$T^{-1}(B_a) = \{x \in X : T(x) \in B_a\} \subseteq \{x \in X : \varphi(T(x)) < a\} \subseteq X \backslash B_a.$$

So $T^{-1}(B_a) \cap B_a = \emptyset$ and, as $B_a \in \mathcal{I}_\mu$, we deduce that

$$0 = \mu(T^{-1}(B_a) \triangle B_a) = \mu(T^{-1}(B_a) \cup B_a) \geq \mu(B_a).$$

Consequently, $\mu(B_a) = 0$ and we have reached a contradiction. Therefore, $\mu(D_\varphi) = 0$ and hence φ is μ-a. e. T-invariant. □

In other terms, Lemma 8.2.9 asserts that φ is μ-a. e. T-invariant if and only if $E(\varphi|\mathcal{I}_\mu) = \varphi$, where $E(\varphi|\mathcal{I}_\mu)$ is the conditional expectation of φ with respect to μ.

This conditional expectation function is an intrinsic part of the most important result in ergodic theory: Birkhoff's ergodic theorem. For more information on this function, see Subsection A.1.9.

8.2.1 Birkhoff's ergodic theorem

We are almost ready to state the main result of this chapter. Before doing so, we must introduce one more notation and terminology.

Definition 8.2.10. Let $T : X \to X$ be a map and let $\varphi : X \to \mathbb{R}$ be a real-valued function. Let $n \in \mathbb{N}$. The nth *Birkhoff sum* of φ at a point $x \in X$ is defined to be

$$S_n\varphi(x) = \sum_{j=0}^{n-1} \varphi(T^j(x)).$$

In other words, $S_n\varphi(x)$ is the sum of the values of the function φ at the first n points in the orbit of x. Sometimes this is referred to as the nth *ergodic sum*. As we will see in a moment, it is also convenient to define $S_0\varphi(x) = 0$.

It is easy to see that the following recurrence formula holds:

$$S_n\varphi(x) = S_k\varphi(x) + S_{n-k}\varphi(T^k(x)), \quad \forall x \in X, \forall k, n \in \mathbb{N} \text{ with } k \leq n. \tag{8.2}$$

We now come to the most important result in ergodic theory. This theorem was originally proved by George David Birkhoff in 1931. There exists now a variety of proofs. The simple one given here originates from Katok and Hasselblatt [33].

Theorem 8.2.11 (Birkhoff's ergodic theorem). *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. If $\varphi \in L^1(X, \mathcal{A}, \mu)$, then*

$$\lim_{n\to\infty} \left\| \frac{1}{n} S_n\varphi - E(\varphi|\mathcal{I}_\mu) \right\|_1 = 0 \quad \textit{and} \quad \lim_{n\to\infty} \frac{1}{n} S_n\varphi(x) = E(\varphi|\mathcal{I}_\mu)(x) \quad \textit{for } \mu\textit{-a. e. } x \in X.$$

Proof. For the μ-a. e. pointwise convergence, it suffices to prove that

$$\limsup_{n\to\infty} \frac{1}{n} S_n\varphi(x) \leq E(\varphi|\mathcal{I}_\mu)(x) \quad \text{for } \mu\text{-a. e. } x \in X. \tag{8.3}$$

Indeed, replacing φ by $-\varphi$ in (8.3), it follows that for μ-a. e. $x \in X$,

$$\liminf_{n\to\infty} \frac{1}{n} S_n\varphi(x) = -\limsup_{n\to\infty} \frac{1}{n} S_n(-\varphi)(x) \geq -E(-\varphi|\mathcal{I}_\mu)(x) = E(\varphi|\mathcal{I}_\mu)(x). \tag{8.4}$$

If (8.3) and consequently (8.4) hold, we can conclude for μ-a. e. $x \in X$ that

$$E(\varphi|\mathcal{I}_\mu)(x) \leq \liminf_{n\to\infty} \frac{1}{n} S_n\varphi(x) \leq \limsup_{n\to\infty} \frac{1}{n} S_n\varphi(x) \leq E(\varphi|\mathcal{I}_\mu)(x),$$

and this would complete the proof. In order to prove (8.3), it is sufficient to show that for every $\varepsilon > 0$ we have

$$\limsup_{n\to\infty} \frac{1}{n} S_n\varphi(x) \leq E(\varphi|\mathcal{I}_\mu)(x) + \varepsilon \quad \text{for } \mu\text{-a. e. } x \in X. \tag{8.5}$$

Indeed, if for each $\varepsilon > 0$ relation (8.5) holds everywhere except on a set X_ε of measure zero, then relation (8.3) holds everywhere except on the set $\bigcup_{k=1}^\infty X_{1/k}$ and $\mu(\bigcup_{k=1}^\infty X_{1/k}) = 0$. So fix $\varepsilon > 0$. We claim that proving (8.5) is equivalent to showing that

$$\limsup_{n\to\infty} \frac{1}{n} S_n\varphi_\varepsilon(x) \leq 0 \quad \text{for } \mu\text{-a. e. } x \in X, \tag{8.6}$$

where

$$\varphi_\varepsilon := \varphi - E(\varphi|\mathcal{I}_\mu) - \varepsilon.$$

Indeed, since $E(\varphi|\mathcal{I}_\mu)$ is $\mathcal{I}_\mu$-measurable by definition, Lemma 8.2.9 implies that $E(\varphi|\mathcal{I}_\mu)$ is μ-a. e. T-invariant, that is, $E(\varphi|\mathcal{I}_\mu) \circ T(x) = E(\varphi|\mathcal{I}_\mu)(x)$ for μ-a. e. $x \in X$. It then follows that for μ-a. e. $x \in X$,

$$\begin{aligned}
\frac{1}{n} S_n\varphi_\varepsilon(x) &= \frac{1}{n} S_n\varphi(x) - \frac{1}{n} S_n E(\varphi|\mathcal{I}_\mu)(x) - \frac{1}{n} S_n\varepsilon(x) \\
&= \frac{1}{n} S_n\varphi(x) - \frac{1}{n} \sum_{j=0}^{n-1} E(\varphi|\mathcal{I}_\mu) \circ T^j(x) - \frac{1}{n} \sum_{j=0}^{n-1} \varepsilon \circ T^j(x) \\
&= \frac{1}{n} S_n\varphi(x) - E(\varphi|\mathcal{I}_\mu)(x) - \varepsilon.
\end{aligned}$$

Thus

$$\limsup_{n\to\infty} \frac{1}{n} S_n\varphi(x) = E(\varphi|\mathcal{I}_\mu)(x) + \varepsilon + \limsup_{n\to\infty} \frac{1}{n} S_n\varphi_\varepsilon(x)$$

for μ-a. e. $x \in X$. However, in order to prove (8.6), it suffices to show that

$$\mu(A_\varepsilon) = 0, \tag{8.7}$$

where

$$A_\varepsilon := \left\{ x \in X \mid \sup_{n\in\mathbb{N}} S_n\varphi_\varepsilon(x) = \infty \right\},$$

since for any $x \notin A_\varepsilon$ we have that $\sup_{n\in\mathbb{N}} S_n\varphi_\varepsilon(x) < \infty$ and it follows that $\limsup_{n\to\infty} \frac{1}{n}S_n\varphi_\varepsilon(x) \le 0$. Now we make a critical observation:

$$E(\varphi_\varepsilon|\mathcal{I}_\mu) = E(\varphi|\mathcal{I}_\mu) - E\big(E(\varphi|\mathcal{I}_\mu)|\mathcal{I}_\mu\big) - E(\varepsilon|\mathcal{I}_\mu) = E(\varphi|\mathcal{I}_\mu) - E(\varphi|\mathcal{I}_\mu) - \varepsilon = -\varepsilon < 0.$$

Rather than restricting our attention to φ_ε, we will prove that (8.7) holds for all $f \in L^1(X,\mathcal{A},\mu)$ such that $E(f|\mathcal{I}_\mu) < 0$, with φ_ε being one such f. The T-invariance of μ implies that $f \circ T^k \in L^1(X,\mathcal{A},\mu)$ for all $k \in \mathbb{N}$. It immediately follows that $S_kf \in L^1(X,\mathcal{A},\mu)$ for all $k \in \mathbb{N}$. For each $n \in \mathbb{N}$ and each $x \in X$ define

$$M_nf(x) = \max_{1\le k\le n} S_kf(x).$$

It is easy to deduce that $M_nf \in L^1(X,\mathcal{A},\mu)$ for all $n \in \mathbb{N}$. It is also obvious that the sequence $(M_nf(x))_{n=1}^\infty$ is nondecreasing for all $x \in X$. Moreover, the recurrence formula (8.2) between successive ergodic sums ($S_n\varphi(x) = \varphi(x)+S_{n-1}\varphi(T(x))$) suggests the existence of a recurrence formula for their successive maxima. Indeed, for all $x \in X$,

$$\begin{aligned}
M_{n+1}f(x) &= \max_{1\le k\le n+1} S_kf(x) \\
&= \max_{1\le k\le n+1} \big[f(x) + S_{k-1}f\big(T(x)\big)\big] \\
&= f(x) + \max_{0\le l\le n} S_lf\big(T(x)\big) \\
&= f(x) + \max\Big\{0, \max_{1\le l\le n} S_lf\big(T(x)\big)\Big\} \\
&= f(x) + \max\big\{0, M_nf\big(T(x)\big)\big\}.
\end{aligned}$$

Therefore, for all $x \in X$,

$$M_{n+1}f(x) - M_nf\big(T(x)\big) = f(x) + \max\big\{-M_nf\big(T(x)\big), 0\big\}. \tag{8.8}$$

Since the sequence $(M_nf(T(x)))_{n=1}^\infty$ is nondecreasing for all $x \in X$, the sequence $(\max\{-M_nf(T(x)), 0\})_{n=1}^\infty$ is nonincreasing for all $x \in X$. By (8.8), the sequence $(M_{n+1}f(x)-M_nf(T(x)))_{n=1}^\infty$ is therefore nonincreasing for all $x \in X$. In order to prove (8.7) for the function f, we will investigate the limit of this latter sequence on the set

$$A = \Big\{x \in X : \sup_{n\in\mathbb{N}} S_nf(x) = \infty\Big\} = \Big\{x \in X : \lim_{n\to\infty} M_nf(x) = \infty\Big\}.$$

Using the recurrence formula (8.2), it is easy to see that $T^{-1}(A) = A$. In particular, this implies that $A \in \mathcal{I}_\mu$. Also, if $x \in A$ then $T(x) \in A$, and thus $\lim_{n\to\infty} M_nf(T(x)) = \infty$.

According to (8.8), it ensues that

$$\lim_{n\to\infty}(M_{n+1}f(x) - M_nf(T(x))) = f(x), \quad \forall x \in A. \tag{8.9}$$

Knowing the pointwise limit of this sequence on A, we further show that this non-increasing sequence is uniformly bounded by an integrable function. For all $n \in \mathbb{N}$ and $x \in X$, we have

$$f(x) \le M_{n+1}f(x) - M_nf(T(x)) \le M_2f(x) - M_1f(T(x)) = f(x) + \max\{-f(T(x)), 0\}.$$

For all $n \in \mathbb{N}$ and $x \in X$, it follows that

$$|M_{n+1}f(x) - M_nf(T(x))| \le |f(x)| + |f(T(x))|.$$

Since $|f| + |f \circ T| \in L^1(X, \mathcal{A}, \mu)$, Lebesgue's dominated convergence theorem (Theorem A.1.38) applies. We deduce from the facts that the sequence $(M_nf)_{n=1}^{\infty}$ is non-decreasing, that μ is T-invariant and that (8.9) holds on A, that

$$\begin{aligned}
0 &\le \int_A (M_{n+1}f - M_nf)\,d\mu \\
&= \int_A M_{n+1}f\,d\mu - \int_A M_nf\,d\mu \\
&= \int_A M_{n+1}f\,d\mu - \int_A M_nf\,d(\mu \circ T^{-1}) \\
&= \int_A M_{n+1}f\,d\mu - \int_A M_nf \circ T\,d\mu \\
&= \int_A (M_{n+1}f(x) - M_nf(T(x)))\,d\mu(x) \to \int_A f(x)\,d\mu(x).
\end{aligned}$$

Hence, $\int_A f\,d\mu \ge 0$. Recall that $A \in \mathcal{I}_\mu$ and $E(f|\mathcal{I}_\mu) < 0$. If it were the case that $\mu(A) > 0$, it would follow from the definition of $E(f|\mathcal{I}_\mu)$ that

$$0 \le \int_A f\,d\mu = \int_A E(f|\mathcal{I}_\mu)\,d\mu < 0,$$

which would result in a contradiction. Thus $\mu(A) = 0$. Setting $f = \varphi_\varepsilon$, we conclude that $\mu(A_\varepsilon) = 0$, hence establishing (8.7) and the μ-a. e. pointwise convergence of the sequence $(\frac{1}{n}S_n\varphi)_{n=1}^{\infty}$ to $E(\varphi|\mathcal{I}_\mu)$.

Now, if φ is bounded then

$$\left\|\frac{1}{n}S_n\varphi\right\|_\infty \le \|\varphi\|_\infty, \quad \forall n \in \mathbb{N},$$

and thus Lebesgue's dominated convergence theorem (Theorem A.1.38) asserts that the sequence $(\frac{1}{n}S_n\varphi)_{n=1}^{\infty}$ converges in $L^1(\mu)$ to $E(\varphi|\mathcal{I}_\mu)$.

In general, since the set of bounded measurable functions is dense in $L^1(\mu)$, for every $\varepsilon > 0$ there exists a bounded measurable function $\varphi_\varepsilon : X \to \mathbb{R}$ such that

$$\|\varphi - \varphi_\varepsilon\|_1 < \frac{\varepsilon}{3}.$$

By the already proven part, there then exists $N_\varepsilon \in \mathbb{N}$ such that

$$\left\|\frac{1}{n}S_n\varphi_\varepsilon - E(\varphi_\varepsilon|\mathcal{I}_\mu)\right\|_1 < \frac{\varepsilon}{3}, \quad \forall n \geq N_\varepsilon.$$

For all $n \geq N_\varepsilon$, we then deduce that

$$\begin{aligned}\left\|\frac{1}{n}S_n\varphi - E(\varphi|\mathcal{I}_\mu)\right\|_1 &\leq \left\|\frac{1}{n}S_n(\varphi - \varphi_\varepsilon)\right\|_1 + \left\|\frac{1}{n}S_n\varphi_\varepsilon - E(\varphi_\varepsilon|\mathcal{I}_\mu)\right\|_1 \\ &\quad + \left\|E(\varphi_\varepsilon|\mathcal{I}_\mu) - E(\varphi|\mathcal{I}_\mu)\right\|_1 \\ &\leq \frac{1}{n}\sum_{j=0}^{n-1}\|(\varphi - \varphi_\varepsilon)\circ T^j\|_1 + \frac{\varepsilon}{3} + \|E(|\varphi_\varepsilon - \varphi| \,|\, \mathcal{I}_\mu)\|_1 \\ &= \frac{1}{n}\sum_{j=0}^{n-1}\|\varphi - \varphi_\varepsilon\|_1 + \frac{\varepsilon}{3} + \|\varphi_\varepsilon - \varphi\|_1 < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.\end{aligned}$$

So $\lim_{n\to\infty}\|\frac{1}{n}S_n\varphi - E(\varphi|\mathcal{I}_\mu)]\|_1 = 0$. □

Let $p \geq 1$. The set $L^p(X, \mathcal{A}, \mu)$ is the set of $\mathcal{A}$-measurable functions $\varphi : X \to \mathbb{R}$ such that $\varphi^p \in L^1(X, \mathcal{A}, \mu)$. It is well known that $L^p(X, \mathcal{A}, \mu) \subseteq L^1(X, \mathcal{A}, \mu)$. Theorem 8.2.11, along with the last part of its proof where one would use this time the density of bounded measurable functions in L^p, yield the following slight generalization.

Theorem 8.2.12 (Birkhoff's ergodic theorem in L^p). *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. If $\varphi \in L^p(X, \mathcal{A}, \mu)$, then*

$$\lim_{n\to\infty}\left\|\frac{1}{n}S_n\varphi - E(\varphi|\mathcal{I}_\mu)\right\|_p = 0 \quad \text{and} \quad \lim_{n\to\infty}\frac{1}{n}S_n\varphi(x) = E(\varphi|\mathcal{I}_\mu)(x) \quad \text{for } \mu\text{-a.e. } x \in X.$$

Remark 8.2.13. If $p = 2$, then the L^p part of Theorem 8.2.12 is commonly referred to as von Neumann's ergodic theorem, proved for the first time in [52].

If a measure-preserving dynamical system on a probability space is ergodic, then Birkhoff's ergodic theorem implies the following.

Corollary 8.2.14 (Ergodic case of Birkhoff's ergodic theorem). *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. If T is ergodic with respect to μ and $\varphi \in L^1(X, \mathcal{A}, \mu)$, then $E(\varphi|\mathcal{I}_\mu) = \int_X \varphi\, d\mu$ and*

$$\lim_{n\to\infty}\left\|\frac{1}{n}S_n\varphi - \int_X \varphi\, d\mu\right\|_1 = 0 \quad \text{and} \quad \lim_{n\to\infty}\frac{1}{n}S_n\varphi(x) = \int_X \varphi\, d\mu \quad \text{for } \mu\text{-a.e. } x \in X.$$

Proof. According to Example A.1.61, the collection of sets

$$\mathcal{N} := \{A \in \mathcal{A} : \mu(A) = 0 \text{ or } \mu(A) = 1\}$$

is a σ-algebra and $E(\varphi|\mathcal{N}) = \int_X \varphi\, d\mu$. As T is ergodic with respect to μ, we have that $\mathcal{I}_\mu \subseteq \mathcal{N}$. By Proposition A.1.60(f,e), it ensues that

$$E(\varphi|\mathcal{I}_\mu) = E\big(E(\varphi|\mathcal{N})\,\big|\,\mathcal{I}_\mu\big) = E\bigg(\int_X \varphi\, d\mu\,\bigg|\,\mathcal{I}_\mu\bigg) = \int_X \varphi\, d\mu.$$

The result follows from Birkhoff's ergodic theorem. □

When T is ergodic with respect to an invariant probability measure μ, Birkhoff's ergodic theorem asserts that the average of φ along the forward orbit of μ-almost every $x \in X$ is asymptotically equal to the average of φ over the entire space. In other words, for any "typical" point the "time average" of a μ-integrable function is equal to its "space average."

In particular, if φ is the characteristic function of a measurable set, Corollary 8.2.14 guarantees the following.

Corollary 8.2.15. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. If T is ergodic with respect to μ, then for every $A \in \mathcal{A}$,*

$$\lim_{n\to\infty} \frac{1}{n} \#\{0 \le j < n : T^j(x) \in A\} = \mu(A) \quad \textit{for } \mu\textit{-a.e. } x \in X.$$

In other terms, the average time the forward orbit of a "typical point" spends in a measurable set is asymptotically equal to the measure of that set. This provides more information than Poincaré's recurrence theorem (Theorem 8.1.16).

Birkhoff's ergodic theorem is a terrifically useful tool. It has had many applications in different areas of mathematics. In particular, it is very useful in number theory. In Exercises 8.5.42–8.5.43, you will use it to prove in a simple way various statements about real numbers whose original, nonergodic proofs were quite involved.

One intuitive understanding of ergodicity is that an ergodic system is one in which for every pair of measurable sets A and B, the sets $T^{-n}(A)$ become independent of B on average.

Lemma 8.2.16. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. Then T is ergodic with respect to μ if and only if*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu\big(T^{-j}(A) \cap B\big) = \mu(A)\mu(B), \quad \forall A, B \in \mathcal{A}. \tag{8.10}$$

Equivalently,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \big[\mu\big(T^{-j}(A) \cap B\big) - \mu(A)\mu(B)\big] = 0, \quad \forall A, B \in \mathcal{A}.$$

Proof. First, suppose that T is ergodic and let $A, B \in \mathcal{A}$. For every $n \in \mathbb{N}$, we have

$$\begin{aligned}\frac{1}{n}\sum_{j=0}^{n-1}\mu(T^{-j}(A)\cap B) &= \frac{1}{n}\sum_{j=0}^{n-1}\int_B \mathbb{1}_{T^{-j}(A)}\,d\mu \\ &= \frac{1}{n}\sum_{j=0}^{n-1}\int_B \mathbb{1}_A\circ T^j\,d\mu \\ &= \int_B \frac{1}{n}\sum_{j=0}^{n-1}\mathbb{1}_A\circ T^j\,d\mu \\ &= \int_B \frac{1}{n}\#\{0\le j<n : T^j(x)\in A\}\,d\mu(x).\end{aligned}$$

Passing to the limit and using Lebesgue's dominated convergence theorem (Theorem A.1.38) and Corollary 8.2.15, we get

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\mu(T^{-j}(A)\cap B) = \int_B \lim_{n\to\infty}\frac{1}{n}\#\{0\le j<n : T^j(x)\in A\}\,d\mu(x) = \mu(A)\mu(B).$$

For the converse implication, suppose that relation (8.10) holds true for all $A, B \in \mathcal{A}$. Let $E \in \mathcal{A}$ be a completely T-invariant set. Setting $A = B = E$ in (8.10), we obtain $\mu(E) = (\mu(E))^2$. So $\mu(E) \in \{0, 1\}$ and T is ergodic with respect to μ. □

To determine whether a measure-preserving dynamical system is ergodic, it suffices to check ergodicity on a semialgebra that generates the σ-algebra.

Lemma 8.2.17. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. If $\mathcal{B} = \sigma(\mathcal{S})$ for some semialgebra $\mathcal{S}$, then T is ergodic if and only if relation* (8.10) *holds for all $A, B \in \mathcal{S}$.*

Proof. If T is ergodic, then relation (8.10) holds for all $A, B \in \mathcal{B} \supseteq \mathcal{S}$.

For the converse implication, suppose that relation (8.10) holds for all $A, B \in \mathcal{S}$. Since each member of the algebra $\mathcal{A}(\mathcal{S})$ generated by $\mathcal{S}$ can be written as a finite disjoint union of elements of $\mathcal{S}$, a straightforward calculation shows that relation (8.10) also holds for all elements of $\mathcal{A}(\mathcal{S})$.

So, let $\varepsilon > 0$ and $A, B \in \mathcal{B} = \sigma(\mathcal{S}) = \sigma(\mathcal{A}(\mathcal{S}))$. By virtue of Lemma A.1.32, there are $A_0, B_0 \in \mathcal{A}(\mathcal{S})$ such that $\mu(A \triangle A_0) < \varepsilon$ and $\mu(B \triangle B_0) < \varepsilon$. By Exercise 8.5.11, it follows that

$$|\mu(A) - \mu(A_0)| < \varepsilon \quad \text{and} \quad |\mu(B) - \mu(B_0)| < \varepsilon. \tag{8.11}$$

Using Exercise 8.5.10, notice also that for every $j \ge 0$,

$$\begin{aligned}(T^{-j}(A)\cap B)\triangle(T^{-j}(A_0)\cap B_0) &\subseteq (T^{-j}(A)\triangle T^{-j}(A_0))\cup(B\triangle B_0) \\ &= T^{-j}(A\triangle A_0)\cup(B\triangle B_0).\end{aligned}$$

Therefore,

$$\mu\big((T^{-j}(A)\cap B)\triangle(T^{-j}(A_0)\cap B_0)\big)\le \mu(A\triangle A_0)+\mu(B\triangle B_0)<2\varepsilon,\quad \forall j\ge 0.$$

By Exercise 8.5.11 again, we deduce that

$$\big|\mu\big(T^{-j}(A)\cap B\big)-\mu\big(T^{-j}(A_0)\cap B_0\big)\big|<2\varepsilon,\quad \forall j\ge 0. \tag{8.12}$$

Using (8.12) and (8.11), we obtain for all $j\ge 0$ that

$$\begin{aligned}\mu\big(T^{-j}(A)\cap B\big)-\mu(A)\mu(B) &\le \big[\mu\big(T^{-j}(A)\cap B\big)-\mu\big(T^{-j}(A_0)\cap B_0\big)\big]\\ &\quad+\big[\mu\big(T^{-j}(A_0)\cap B_0\big)-\mu(A_0)\mu(B_0)\big]\\ &\quad+\big[\mu(A_0)\mu(B_0)-\mu(A)\mu(B_0)\big]\\ &\quad+\big[\mu(A)\mu(B_0)-\mu(A)\mu(B)\big]\\ &\le \big|\mu\big(T^{-j}(A)\cap B\big)-\mu\big(T^{-j}(A_0)\cap B_0\big)\big|\\ &\quad+\big[\mu\big(T^{-j}(A_0)\cap B_0\big)-\mu(A_0)\mu(B_0)\big]\\ &\quad+\big|\mu(A_0)-\mu(A)\big|\mu(B_0)\\ &\quad+\mu(A)\big|\mu(B_0)-\mu(B)\big|\\ &< 4\varepsilon+\big[\mu\big(T^{-j}(A_0)\cap B_0\big)-\mu(A_0)\mu(B_0)\big]\end{aligned}$$

and it follows that

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\big[\mu\big(T^{-j}(A)\cap B\big)-\mu(A)\mu(B)\big]\\ &\qquad\le 4\varepsilon+\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\big[\mu\big(T^{-j}(A_0)\cap B_0\big)-\mu(A_0)\mu(B_0)\big]=4\varepsilon,\end{aligned}\tag{8.13}$$

where the above limit is 0 as a consequence of (8.10) holding for $A_0, B_0\in\mathcal{A}(S)$.

Similarly, for all $j\ge 0$,

$$\mu(A)\mu(B)-\mu\big(T^{-j}(A)\cap B\big)<4\varepsilon+\big[\mu(A_0)\mu(B_0)-\mu\big(T^{-j}(A_0)\cap B_0\big)\big]$$

and it ensues that

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\big[\mu(A)\mu(B)-\mu\big(T^{-j}(A)\cap B\big)\big]\le 4\varepsilon.$$

Therefore,

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\big[\mu\big(T^{-j}(A)\cap B\big)-\mu(A)\mu(B)\big]\ge -4\varepsilon. \tag{8.14}$$

By (8.13) and (8.14), the limsup is at most 4ε while the liminf is at least -4ε. Since $\varepsilon>0$ was chosen arbitrarily, we deduce that the limit exists and is 0. Thus, relation (8.10) holds for all elements of $\mathcal{B}$, and T is ergodic by Lemma 8.2.16. □

We end this section with a characterization of ergodicity in terms of invariant and measure-theoretically invariant functions.

Theorem 8.2.18. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. The following statements are equivalent:*
(a) *T is ergodic with respect to μ.*
(b) *If φ is a T-invariant $L^1(X, \mathcal{A}, \mu)$-function, then φ is μ-a. e. constant.*
(c) *If φ is a μ-a. e. T-invariant $L^1(X, \mathcal{A}, \mu)$-function, then φ is μ-a. e. constant.*
(d) *If φ is a T-invariant measurable function, then φ is μ-a. e. constant.*
(e) *If φ is a μ-a. e. T-invariant measurable function, then φ is μ-a. e. constant.*

Proof. We shall first prove the chain of implications (a)⇒(c)⇒(b)⇒(a) and then show that (e)⇒(d)⇒(a)⇒(e).

To begin, suppose that T is ergodic and let φ be a μ-a. e. T-invariant $L^1(X, \mathcal{A}, \mu)$-function. As φ is μ-a. e. T-invariant and μ is T-invariant, it follows from Lemma 8.2.8 that φ is constant over the forward orbit of μ-a. e. $x \in X$. This implies that $S_n\varphi(x) = n\varphi(x)$ for all $n \in \mathbb{N}$ for μ-a. e. $x \in X$. Using this and the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14), it follows that

$$\varphi(x) = \lim_{n\to\infty} \frac{1}{n} S_n\varphi(x) = \int_X \varphi \, d\mu \quad \text{for } \mu\text{-a. e. } x \in X.$$

So φ is constant μ-almost everywhere. This proves that (a)⇒(c).

Since every T-invariant function is μ-a. e. T-invariant, it is clear that (c)⇒(b).

We now want to show that (b)⇒(a). Suppose that every T-invariant $L^1(X, \mathcal{A}, \mu)$-function is constant μ-a. e. and assume by way of contradiction that T is not ergodic with respect to μ. Then there exists a set $A \in \mathcal{A}$ such that $T^{-1}(A) = A$ with $\mu(A) > 0$ and $\mu(X \backslash A) > 0$. Since $T^{-1}(A) = A$, we have that $\mathbb{1}_A \circ T = \mathbb{1}_A$. However, $\mathbb{1}_A$ is not constant μ-a. e. since $\mu(A) > 0$ and $\mu(X \backslash A) > 0$. Thus, $\mathbb{1}_A$ is a T-invariant $L^1(X, \mathcal{A}, \mu)$-function which is not constant μ-almost everywhere. This contradiction shows that T must be ergodic.

This completes the proof of the first chain (a)⇒(c)⇒(b)⇒(a).

Accordingly, let us turn our attention to the second chain. It is clear that (e)⇒(d). The above proof that (b)⇒(a) carries over directly to show that (d)⇒(a) by simply replacing "$L^1(X, \mathcal{A}, \mu)$" by "measurable." All that is left is to establish that (a)⇒(e). So, suppose that T is ergodic and let φ be a μ-a. e. T-invariant measurable function. Then $\mu(\{x \in X \mid \varphi(T(x)) \neq \varphi(x)\}) = 0$. Assume by way of contradiction that φ is not μ-a. e. constant. Then there exists $r \in \mathbb{R}$ such that $\mu(S_r) > 0$ and $\mu(X \backslash S_r) > 0$, where $S_r := \{x \mid \varphi(x) < r\}$. The set S_r is μ-a. e. T-invariant since

$$\begin{aligned} T^{-1}(S_r) \triangle S_r &= (T^{-1}(S_r) \backslash S_r) \cup (S_r \backslash T^{-1}(S_r)) \\ &= \{x : \varphi(T(x)) < r \leq \varphi(x)\} \cup \{x : \varphi(x) < r \leq \varphi(T(x))\} \\ &\subseteq \{x : \varphi(T(x)) \neq \varphi(x)\} \end{aligned}$$

and hence

$$\mu(T^{-1}(S_r) \triangle S_r) \le \mu(\{x \in X : \varphi(T(x)) \neq \varphi(x)\}) = 0.$$

Thus $\mu(T^{-1}(S_r) \triangle S_r) = 0$. Since μ is ergodic, we deduce that either $\mu(S_r) = 0$ or $\mu(X \backslash S_r) = 0$. This contradiction implies that φ must be μ-a. e. constant. □

Remark 8.2.19. It is possible to prove that the property "$L^1(X, \mathcal{A}, \mu)$" can be replaced by "$L^p(X, \mathcal{A}, \mu)$" for any $1 \le p < \infty$.

8.2.2 Existence of ergodic measures

We will shortly embark on a study of the set of all ergodic measures for a given measurable transformation.

Definition 8.2.20. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. The set of all T-invariant probability measures that are ergodic with respect to T is denoted by $E(T, \mathcal{A})$. If $\mathcal{A}$ is the Borel σ-algebra on a topological space X, in line with the notation for the T-invariant Borel probability measures, we simply write $E(T) := E(T, \mathcal{B}(X))$.

We saw in Lemma 8.1.19 that the set of invariant probability measures $M(T, \mathcal{A})$ is convex. We shall soon prove that the ergodic measures $E(T, \mathcal{A})$ form the extreme points of $M(T, \mathcal{A})$. First, we show that any two ergodic measures are either equal or mutually singular (for more information on mutual singularity, see Subsection A.1.7).

Theorem 8.2.21. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. If $\mu_1, \mu_2 \in E(T, \mathcal{A})$ and $\mu_1 \neq \mu_2$, then $\mu_1 \perp \mu_2$.*

Proof. Since $\mu_1 \neq \mu_2$, there exists some set $A \in \mathcal{A}$ with $\mu_1(A) \neq \mu_2(A)$. By Corollary 8.2.15 of Birkhoff's ergodic theorem, for each $i = 1, 2$ there exists a set X_i of full μ_i-measure such that

$$\lim_{n\to\infty} \frac{1}{n} \#\{0 \le j < n : T^j(x) \in A\} = \mu_i(A), \quad \forall x \in X_i.$$

Consequently, $\mu_1(A) = \mu_2(A)$ on $X_1 \cap X_2$. As we know that $\mu_1(A) \neq \mu_2(A)$, we deduce that $X_1 \cap X_2 = \emptyset$. Thus $\mu_1(X_1) = 1$, $\mu_2(X_2) = 1$ and $X_1 \cap X_2 = \emptyset$. Therefore, $\mu_1 \perp \mu_2$. □

We use the above theorem to give a characterization of ergodic measures as those invariant probability measures with respect to which no other invariant probability measure is absolutely continuous.

Theorem 8.2.22. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and let $\mu \in M(T, \mathcal{A})$. Then $\mu \in E(T, \mathcal{A})$ if and only if there is no $\nu \in M(T, \mathcal{A})$ such that $\nu \ll \mu$ and $\nu \neq \mu$.*

Proof. First, suppose that $\mu \in E(T, \mathcal{A})$. Let $\nu \in M(T, \mathcal{A})$ be such that $\nu \ll \mu$. We claim that $\nu \in E(T, \mathcal{A})$, too. Indeed, suppose by way of contradiction that there exists $A \in \mathcal{A}$ such that $T^{-1}(A) = A$ with $\nu(A) > 0$ and $\nu(X \backslash A) > 0$. Since $\nu \ll \mu$, it follows that $\mu(A) > 0$ and $\mu(X \backslash A) > 0$. This contradicts the ergodicity of μ. So $\nu \in E(T, \mathcal{A})$. Now, if $\nu \neq \mu$ then Theorem 8.2.21 affirms that $\nu \perp \mu$. This contradicts the hypothesis that $\nu \ll \mu$. Hence, $\nu = \mu$.

For the converse implication, suppose that μ is not ergodic (but still T-invariant by hypothesis). Then there exists some $A \in \mathcal{A}$ such that $T^{-1}(A) = A$ with $\mu(A) > 0$ and $\mu(X \backslash A) > 0$. Let μ_A be the conditional measure of μ on A, as expressed in Definition A.1.70. Then one immediately verifies that μ_A is a T-invariant probability measure such that $\mu_A \neq \mu$ and $\mu_A \ll \mu$. □

Recall that in a vector space the extreme points of a convex set are those points which cannot be represented as a nontrivial convex combination of two distinct points of the set. In concrete terms, let V be a vector space and C be a convex subset of V. A vector $v \in C$ is an *extreme point* of C if the only combination of distinct vectors $v_1, v_2 \in C$ such that $v = \alpha v_1 + (1 - \alpha) v_2$ for some $\alpha \in [0, 1]$ is a combination with $\alpha = 0$ or $\alpha = 1$.

Theorem 8.2.23. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. The ergodic measures $E(T, \mathcal{A})$ are the extreme points of the set of invariant probability measures $M(T, \mathcal{A})$.*

Proof. Suppose that $\mu \in E(T, \mathcal{A})$ is not an extreme point of $M(T, \mathcal{A})$. Then there exist measures $\mu_1 \neq \mu_2$ in $M(T, \mathcal{A})$ and $0 < \alpha < 1$ such that $\mu = \alpha \mu_1 + (1 - \alpha)\mu_2$. It follows immediately that $\mu_1 \ll \mu$ and $\mu_2 \ll \mu$. By Theorem 8.2.22, we deduce from the ergodicity of μ that $\mu_1 = \mu = \mu_2$. This contradicts the fact that $\mu_1 \neq \mu_2$. Thus μ is an extreme point of $M(T, \mathcal{A})$.

To prove the converse implication, let $\mu \in M(T, \mathcal{A}) \backslash E(T, \mathcal{A})$. We want to show that μ is not an extreme point of $M(T, \mathcal{A})$. Since μ is not ergodic, there exists a set $A \in \mathcal{A}$ such that $T^{-1}(A) = A$ with $\mu(A) > 0$ and $\mu(X \backslash A) > 0$. Observe that μ can be written as the following nontrivial convex combination of the T-invariant conditional measures μ_A and $\mu_{X \backslash A}$: for every $B \in \mathcal{A}$,

$$\begin{aligned}\mu(B) = \mu(A \cap B) + \mu\big((X \backslash A) \cap B\big) &= \mu(A)\, \mu_A(B) + \mu(X \backslash A)\, \mu_{X \backslash A}(B) \\ &= \mu(A)\, \mu_A(B) + \big(1 - \mu(A)\big)\, \mu_{X \backslash A}(B).\end{aligned}$$

Hence, μ is a nontrivial convex combination of two distinct T-invariant probability measures and thus μ is not an extreme point of $M(T, \mathcal{A})$. □

We now invoke Krein–Milman's theorem to deduce that every topological dynamical system admits an ergodic and invariant measure. Recall that the *convex hull* of a subset S of a vector space V is the set of all convex combinations of vectors of S.

Theorem 8.2.24 (Krein–Milman's theorem). *If K is a compact subset of a locally convex topological vector space V and E is the set of its extremal points, then $\overline{co}(E) \supseteq K$,* *where* $\overline{co}(E)$ *is the closed convex hull of* E. *Consequently,* $\overline{co}(E) = \overline{co}(K)$. *In particular, if* K *is convex then* $\overline{co}(E) = K$.

Proof. See Theorem V.8.4 in Dunford and Schwartz [20]. □

Corollary 8.2.25. *Let* $T : X \to X$ *be a topological dynamical system. Then*

$$\overline{\mathrm{co}}\big(E(T)\big) = M(T) \neq \emptyset.$$

In particular, $E(T) \neq \emptyset$.

Proof. In Theorems 8.1.20 and 8.1.22, we saw that whenever $T : X \to X$ is a topological dynamical system, the set $M(T)$ is a nonempty compact convex subset of the (compact) convex space $M(X)$. Moreover, Theorem 8.2.23 established that $E(T)$ is the set of extreme points of $M(T)$. The result then follows from the application of Krein–Milman's theorem with $K = M(T)$, $V = M(X)$ and $E = E(T)$. □

This corollary can be restated as follows: Every topological dynamical system induces at least one ergodic measure-preserving dynamical system.

The compactness and convexity of $M(T)$ as a subset of the convex space $M(X)$ (equipped with the weak* topology) further allows us to use Choquet's representation theorem to express each element of $M(T)$ in terms of the elements of its set of extreme points $E(T)$. In fact, this decomposition holds in a more general case.

Theorem 8.2.26 (Ergodic decomposition). *Let* $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ *be a measure-preserving transformation of a Borel probability space* $(X, \mathcal{B}, \mu)$. *Then there is a Borel probability space* $(Y, \mathcal{B}(Y), \nu)$ *and a measurable map* $Y \ni y \mapsto \mu_y \in M(X)$ *such that*

(a) μ_y *is an ergodic* T*-invariant Borel probability measure on* X *for* ν*-almost every* $y \in Y$*; and*

(b) $\mu = \int_Y \mu_y \, d\nu(y)$.

Moreover, one may require that the map $y \mapsto \mu_y$ *be injective, or alternatively set*

$$\big(Y, \mathcal{B}(Y), \nu\big) = (X, \mathcal{B}, \mu) \quad \textit{and} \quad \mu_x = \mu_x^{\mathcal{I}_\mu},$$

where $\mathcal{I}_\mu$ *is the* σ*-algebra of* μ*-almost* T*-invariant sets (see Definition* 8.2.5*) and* $\mu_x^{\mathcal{I}}$ *is a Borel probability measure on* X *for which*

$$E(\varphi|\mathcal{I}_\mu)(x) = \int_X \varphi(z)\, d\mu_x^{\mathcal{I}_\mu}(z) \quad \textit{for} \;\; \mu\textit{-a.e.}\; x \in X$$

for all $\varphi \in L^1(X, \mathcal{B}, \mu)$.

Proof. The interested reader is invited to consult Theorem 6.2 in [21]. □

So, for any topological dynamical system, every invariant measure can be uniquely written as a generalized convex combination of ergodic invariant measures.

We end our theoretical discussion of ergodic measures with the following result. We already know from Theorem 1.5.11 that the set of transitive points for any transitive map is a dense set, which can be thought of as "topologically full." The forthcoming result asserts that a dynamical system which admits an ergodic invariant measure supported on the entire space is transitive, and its set of transitive points is "full" not only topologically but also measure-theoretically.

Theorem 8.2.27. *Let $T : X \to X$ be a topological dynamical system. If $\mu \in E(T)$ and* $\mathrm{supp}(\mu) = X$*, then μ-almost every $x \in X$ is a transitive point for T. In particular, T is transitive.*

Proof. Let $\{U_k\}_{k=1}^{\infty}$ be a base for the topology of X. For each $k \in \mathbb{N}$, let X_k be the set of points whose orbits visit U_k infinitely often; in other words,

$$X_k = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} T^{-n}(U_k).$$

We observed in the discussion preceding Proposition 8.2.4 that the set X_k is completely T-invariant. Since μ is ergodic, we deduce that $\mu(X_k) = 0$ or $\mu(X \backslash X_k) = 0$. However,

$$\begin{aligned}\mu(X_k) = \mu\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} T^{-n}(U_k)\right) &= \lim_{m\to\infty} \mu\left(\bigcup_{n=m}^{\infty} T^{-n}(U_k)\right) \\ &\geq \lim_{m\to\infty} \mu(T^{-m}(U_k)) = \lim_{m\to\infty} \mu(U_k) = \mu(U_k) > 0,\end{aligned}$$

where the last strict inequality is due to the fact that the support of the measure μ is X. Therefore, $\mu(X \backslash X_k) = 0$. Since this is true for all $k \in \mathbb{N}$, we conclude that $\mu(X \backslash \bigcap_{k=1}^{\infty} X_k) = 0$. But all points in the set $\bigcap_{k=1}^{\infty} X_k$ have an orbit that visits each basic open set U_k infinitely often. Thus all points of $\bigcap_{k=1}^{\infty} X_k$ are transitive. Hence, μ-a. e. $x \in X$ is a transitive point for T. □

8.2.3 Examples of ergodic measures

We begin this section with a simple example.

Example 8.2.28. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation with a fixed point x_0. Let δ_{x_0} be the Dirac point mass supported at x_0. We saw in Example 8.1.8 that δ_{x_0} is T-invariant. The measure δ_{x_0} is also trivially ergodic since any measurable set is of measure 0 or 1.

We now revisit the rotations of the unit circle. For a comparative perspective of the topological dynamics of these maps, see Theorem 1.5.12.

Proposition 8.2.29. *Let $T_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ be the map defined by $T_\alpha(x) := x + \alpha \pmod 1$. Then T_α is ergodic with respect to the Lebesgue measure if and only if $\alpha \in \mathbb{R}\backslash\mathbb{Q}$.*

Proof. We demonstrated in Example 8.1.9 that the Lebesgue measure λ on $\mathbb{S}^1$ is T_α-invariant for any $\alpha \in \mathbb{R}$. First, assume that $\alpha \notin \mathbb{Q}$. We want to show that λ is ergodic with respect to T_α. For this, we will show that if $f \circ T_\alpha = f$ and $f \in L^2(\lambda)$, then f is λ-a. e. constant. It will then result from Theorem 8.2.18 and Remark 8.2.19 that λ is ergodic. Consider the Fourier series representation of f, which is given by

$$f(x) = \sum_{k\in\mathbb{Z}} a_k e^{2\pi ikx}.$$

Then

$$f \circ T_\alpha(x) = \sum_{k\in\mathbb{Z}} a_k e^{2\pi ik(x+\alpha)} = \sum_{k\in\mathbb{Z}} a_k e^{2\pi ik\alpha} e^{2\pi ikx}.$$

Since we assumed that $f \circ T_\alpha = f$, we deduce from the uniqueness of the Fourier series representation that $a_k e^{2\pi ik\alpha} = a_k$ for all $k \in \mathbb{Z}$. Hence, for each k we have $a_k = 0$ or $e^{2\pi ik\alpha} = 1$. The latter equality holds if and only if $k\alpha \in \mathbb{Z}$. As $\alpha \notin \mathbb{Q}$, this occurs only when $k = 0$. Thus $f(x) = a_0$ for λ-a. e. $x \in \mathbb{S}^1$, that is, f is λ-a. e. constant. This implies that λ is ergodic.

Now, suppose that $\alpha = p/q \in \mathbb{Q}$. We may assume without loss of generality that $q > p \geq 0$. In what follows, all sets must be interpreted modulo 1. Let

$$A := \bigcup_{n=0}^{q-1} \left[\frac{n}{q}, \left(n + \frac{1}{2}\right)\frac{1}{q}\right].$$

Then

$$\begin{aligned}
T_\alpha^{-1}(A) &= \bigcup_{n=0}^{q-1} \left[\frac{n-p}{q}, \left(n - p + \frac{1}{2}\right)\frac{1}{q}\right] \\
&= \bigcup_{n=0}^{p-1} \left[\frac{n-p}{q}, \left(n - p + \frac{1}{2}\right)\frac{1}{q}\right] \cup \bigcup_{n=p}^{q-1} \left[\frac{n-p}{q}, \left(n - p + \frac{1}{2}\right)\frac{1}{q}\right] \\
&= \bigcup_{n=0}^{p-1} \left[\frac{n+q-p}{q}, \left(n + q - p + \frac{1}{2}\right)\frac{1}{q}\right] \cup \bigcup_{k=0}^{q-(p+1)} \left[\frac{k}{q}, \left(k + \frac{1}{2}\right)\frac{1}{q}\right] \\
&= \bigcup_{k=q-p}^{q-1} \left[\frac{k}{q}, \left(k + \frac{1}{2}\right)\frac{1}{q}\right] \cup \bigcup_{k=0}^{q-(p+1)} \left[\frac{k}{q}, \left(k + \frac{1}{2}\right)\frac{1}{q}\right] \\
&= \bigcup_{k=0}^{q-1} \left[\frac{k}{q}, \left(k + \frac{1}{2}\right)\frac{1}{q}\right] = A.
\end{aligned}$$

Also, one immediately verifies that $\lambda(A) = q \cdot (1/2)(1/q) = 1/2$. In summary, $T_\alpha^{-1}(A) = A$ and $\lambda(A) \notin \{0, 1\}$. Thus, λ is not ergodic with respect to T_α when $\alpha \in \mathbb{Q}$. □

We now return to the doubling map and its generalizations.

Example 8.2.30. Fix $n > 1$. Recall once more the map $T_n : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $T_n(x) = nx \pmod 1$. We claim that T_n is ergodic with respect to the Lebesgue measure λ. We saw in Example 8.1.10 that T_n preserves λ. It is possible to demonstrate the ergodicity of T_n with respect to λ in a similar way that we did for T_α in Proposition 8.2.29. However, in this example we will provide a different proof. Let $A \in \mathcal{B}(\mathbb{S}^1)$ be a set such that $T_n^{-1}(A) = A$ and $\lambda(A) > 0$. To establish ergodicity, we need to show that $\lambda(A) = 1$.

Recall that Lebesgue's density theorem (see Corollary 2.14 in Mattila [46]) states that for any Lebesgue measurable set $A \subseteq \mathbb{R}^n$, the density of A is 0 or 1 at λ-almost every point of $\mathbb{R}^n$. Moreover, the density of A is 1 at λ-almost every point of A. The density of A at $x \in \mathbb{R}^n$ is defined as

$$\lim_{r\to 0} \frac{\lambda(A \cap B(x,r))}{\lambda(B(x,r))}.$$

Given that $\lambda(A) > 0$, take x to be a Lebesgue density point of A, that is, a point where the density of A is 1, that is,

$$\lim_{r\to 0} \frac{\lambda(A \cap B(x,r))}{2r} = 1. \tag{8.15}$$

Set

$$r_k := 1/(2n^k).$$

Then T_n^k is injective on each arc of length less than $2r_k$. So $T_n^k|_{B(x,r_k)}$ is injective for each $x \in \mathbb{S}^1$. On the other hand,

$$T_n^k(B(x,r_k)) = \mathbb{S}^1 \backslash \{T_n^k(x + r_k)\}.$$

Thus

$$\lambda(T_n^k(B(x,r_k))) = 1.$$

Therefore,

$$\lambda(A) = \lambda(T_n^k(A)) \geq \frac{\lambda(T_n^k(A \cap B(x,r_k)))}{\lambda(T_n^k(B(x,r_k)))} = \frac{n^k \lambda(A \cap B(x,r_k))}{n^k \lambda(B(x,r_k))} = \frac{\lambda(A \cap B(x,r_k))}{2r_k} \underset{k\to\infty}{\longrightarrow} 1$$

by (8.15). Consequently, $\lambda(A) = 1$. This proves the ergodicity of λ.

Example 8.2.31. Recall the full Markov maps from Example 8.1.11. We claim that any such T is ergodic with respect to the Lebesgue measure λ. As we did in the previous example, we would like to use Lebesgue's density theorem to prove this. However, in contradistinction with the preceding example, for all $r > 0$ and all $k \in \mathbb{N}$, the restriction of T^k to the ball $B(p_j, r)$ is not one-to-one when p_j is a point of continuity for a full Markov map T; for example, the point 1/2 for the tent map. Despite that potential lack of injectivity, let us try to use Lebesgue's density theorem.

For each $n \in \mathbb{N}$, let $\mathcal{P}_n := \{I_j^{(n)} \mid 1 \le j \le q^n\}$ be the "partition" of $[0,1]$ into the successive intervals of monotonicity of T^n. In particular, $I_j^{(1)} = I_j$ for all $1 \le j \le q$. For each $1 \le j < q^n$, let $p_j^{(n)}$ be the unique point in $I_j^{(n)} \cap I_{j+1}^{(n)}$. For all $x \in [0,1] \setminus \{p_j^{(n)} : 1 \le j < q^n\}$, let $I^{(n)}(x)$ be the unique element of $\mathcal{P}_n$ containing x. We will need two claims.

Claim 1. *For every $n \in \mathbb{N}$, the map $T^n : [0,1] \to [0,1]$ is a full Markov map under the "partition" $\mathcal{P}_n$, and $\mathcal{P}_{n+1}$ is finer than $\mathcal{P}_n$.*

Claim 2. *If $A \in \mathcal{B}([0,1])$, then*

$$\lim_{n\to\infty} \frac{\lambda(A \cap I^{(n)}(x))}{\lambda(I^{(n)}(x))} = 1 \textit{ for } \lambda\text{-}a.e.\ x \in A.$$

Proof of ergodicity of T. For the time being, suppose that both claims hold. Let A be a Borel subset of $[0,1]$ such that $T^{-1}(A) = A$ and $\lambda(A) > 0$. By the surjectivity of T, we know that $T(A) = A$. Fix any $x \in A$ satisfying Claim 2. For each $n \in \mathbb{N}$, let m_n be the slope of $T^n|_{I^{(n)}(x)}$. Using both claims, we obtain that

$$\lambda(A) = \lambda(T^n(A)) \ge \frac{\lambda(T^n(A \cap I^{(n)}(x)))}{\lambda(T^n(I^{(n)}(x)))} = \frac{m_n \lambda(A \cap I^{(n)}(x))}{m_n \lambda(I^{(n)}(x))} = \frac{\lambda(A \cap I^{(n)}(x))}{\lambda(I^{(n)}(x))} \underset{n\to\infty}{\longrightarrow} 1.$$

Consequently, $\lambda(A) = 1$. This proves the ergodicity of λ. □

Proof of Claim 1. We proceed by induction. Suppose that T^n is a full Markov map under the "partition" $\mathcal{P}_n$. It is obvious that T^{n+1} is piecewise linear. Fix $I_j^{(n)} \in \mathcal{P}_n$. For all $1 \le i \le q$, consider

$$I_{j,i}^{(n+1)} := T|_{I_i}^{-1}(I_j^{(n)}).$$

Define

$$\widetilde{\mathcal{P}}_{n+1} := \{I_{j,i}^{(n+1)} \mid 1 \le j \le q^n, 1 \le i \le q\}.$$

Then

$$\begin{aligned}
\bigcup_{j=1}^{q^n}\bigcup_{i=1}^{q} I_{j,i}^{(n+1)} &= \bigcup_{j=1}^{q^n}\bigcup_{i=1}^{q} T|_{I_i}^{-1}(I_j^{(n)}) = \bigcup_{i=1}^{q}\bigcup_{j=1}^{q^n} T|_{I_i}^{-1}(I_j^{(n)}) \\
&= \bigcup_{i=1}^{q} T|_{I_i}^{-1}\left(\bigcup_{j=1}^{q^n} I_j^{(n)}\right) = \bigcup_{i=1}^{q} T|_{I_i}^{-1}([0,1]) \\
&= \bigcup_{i=1}^{q} I_i = [0,1].
\end{aligned}$$

Thus $\widetilde{\mathcal{P}}_{n+1}$ is a cover of $[0,1]$. Obviously, the interiors of the intervals in $\widetilde{\mathcal{P}}_{n+1}$ are mutually disjoint and $\widetilde{\mathcal{P}}_{n+1}$ is finer than $\widetilde{\mathcal{P}}_n$. For all $1 \le j \le q^n$ and all $1 \le i \le q$ we further have

$$T^{n+1}(I_{j,i}^{(n+1)}) = T^n(T(I_{j,i}^{(n+1)})) = T^n(T(T|_{I_i}^{-1}(I_j^{(n)}))) = T^n(I_j^{(n)}) = [0,1].$$

So $\mathcal{P}_{n+1} = \widetilde{\mathcal{P}}_{n+1}$ and T^{n+1} is a full Markov map under the "partition" $\mathcal{P}_{n+1}$, which is finer than $\mathcal{P}_n$. This completes the inductive step. Since the claim clearly holds when $n = 1$, Claim 1 has been established for all $n \in \mathbb{N}$. □

Proof of Claim 2. Let $\mathcal{A}_n := \sigma(\mathcal{P}_n)$ be the σ-algebra generated by $\mathcal{P}_n$. By Claim 1, $\mathcal{P}_{n+1}$ is finer than $\mathcal{P}_n$ for all $n \in \mathbb{N}$ and thus the sequence of σ-algebras $(\mathcal{A}_n)_{n=1}^{\infty}$ is ascending. Now, let

$$m := \min\{|\text{slope}(T|_{I_i})| : 1 \le i \le q\} = \{|(T|_{\text{Int}(I_i)})'| : 1 \le i \le q\} > 1.$$

Then

$$\text{diam}(\mathcal{P}_n) := \sup\{\text{diam}(I_j^{(n)}) : 1 \le j \le q^n\} \le m^{-n} \underset{n\to\infty}{\longrightarrow} 0.$$

Therefore, $\mathcal{A}_\infty := \sigma(\bigcup_{n=1}^{\infty} \mathcal{A}_n)$ contains all the subintervals of $[0,1]$. Hence, $\mathcal{A}_\infty = \mathcal{B}([0,1])$. Let $A \in \mathcal{B}([0,1])$. According to Theorem A.1.67,

$$\lim_{n\to\infty} E(\mathbb{1}_A|\mathcal{A}_n)(x) = E(\mathbb{1}_A|\mathcal{B}([0,1]))(x) = \mathbb{1}_A(x) \quad \text{for } \lambda\text{-a. e. } x \in [0,1]. \tag{8.16}$$

Now, recall that $\mathcal{P}_n := \{I_j^{(n)} : 1 \le j \le q^n\}$ is the "partition" of $[0,1]$ into the successive intervals of monotonicity of T^n. Moreover, $p_j^{(n)}$ is the unique point in $I_j^{(n)} \cap I_{j+1}^{(n)}$ for every $1 \le j < q^n$. Define

$$\mathcal{P}_n' := \{I_1^{(n)} \backslash \{p_1^{(n)}\}\} \cup \{\text{Int}(I_j^{(n)}) : 1 < j < q^n\} \cup \{I_{q^n}^{(n)} \backslash \{p_{q^n-1}^{(n)}\}\} \cup \{\{p_j^{(n)}\} : 1 \le j < q^n\}.$$

Though $\mathcal{P}_n$ is not a partition per se, the family $\mathcal{P}_n'$ is a finite partition of $[0,1]$. Moreover, it is easy to see that $\mathcal{A}_n := \sigma(\mathcal{P}_n) = \sigma(\mathcal{P}_n')$. By Example A.1.62, the conditional expectation function $E(\mathbb{1}_A|\mathcal{A}_n)$ is constant on each element of the partition $\mathcal{P}_n'$. For all x except the points $p_j^{(n)}$, by definition of the conditional expectation function we must also have

$$\int_{I^{(n)}(x)} E(\mathbb{1}_A|\mathcal{A}_n)\, d\lambda = \int_{I^{(n)}(x)} \mathbb{1}_A\, d\lambda = \lambda(A \cap I^{(n)}(x)).$$

Therefore,

$$E(\mathbb{1}_A|\mathcal{A}_n)(y) = \frac{\lambda(A \cap I^{(n)}(x))}{\lambda(I^{(n)}(x))}, \quad \forall y \in \text{Int}(I^{(n)}(x)).$$

Since $I^{(n)}(y) = I^{(n)}(x)$ for all $y \in \mathrm{Int}(I^{(n)}(x))$, it ensues that

$$E(\mathbb{1}_A|\mathcal{A}_n)(z) = \frac{\lambda(A \cap I^{(n)}(z))}{\lambda(I^{(n)}(z))} \quad \text{for } \lambda\text{-a. e. } z \in [0,1].$$

It follows from (8.16) that

$$\lim_{n\to\infty} \frac{\lambda(A \cap I^{(n)}(x))}{\lambda(I^{(n)}(x))} = 1 \quad \text{for } \lambda\text{-a. e. } x \in A.$$

Claim 2 is proved and this completes this example. □

The next example concerns the shift map.

Example 8.2.32. Recall the one-sided Bernoulli shift from Example 8.1.14 where, given a probability space $(E, \mathcal{F}, P)$, the product space $(E^\infty, \mathcal{F}_\infty, \mu_P)$ is a probability space and the product measure μ_P is invariant under the left shift map σ. We now demonstrate that μ_P is also ergodic with respect to σ. To prove this, we use Lemmas 8.2.16–8.2.17. Let A, B be cylinder sets of length M and N, respectively. Since cylinder A depends on the first M coordinates, cylinder $\sigma^{-j}(A)$ depends on coordinates $j+1$ to $j+M$. Consequently, cylinders $\sigma^{-j}(A)$ and B depend on different coordinates as soon as $j \geq N$. Since μ_P is a product measure and is σ-invariant, we deduce that

$$\mu_P(\sigma^{-j}(A) \cap B) = \mu_P(\sigma^{-j}(A))\mu_P(B) = \mu_P(A)\mu_P(B), \quad \forall j \geq N.$$

It follows immediately that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu_P(\sigma^{-j}(A) \cap B) = \mu_P(A)\mu_P(B).$$

The ergodicity of σ ensues from Lemmas 8.2.16 and 8.2.17. In Example 8.3.13, we shall see that the shift map enjoys an even stronger property than ergodicity.

Our final example pertains to ordinary normal numbers.

Example 8.2.33. Let $n \geq 2$. On one hand, consider the probability space $(E, \mathcal{F}, P)$, with set $E = \{0, 1, \ldots, n-1\}$, σ-algebra $\mathcal{F} = \mathcal{P}(E)$ and probability measure $P = (1/n)\sum_{k=0}^{n-1} \delta_k$. According to Example 8.2.32, the product space $(E^\infty, \mathcal{F}_\infty, \mu_P)$ is a probability space and the measure μ_P is ergodic with respect to the shift map σ.

On the other hand, consider the map $T_n : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $T_n(x) = nx \pmod 1$. In Example 8.2.30, we learned that T_n is ergodic with respect to the Lebesgue measure λ. This map is also distance expanding, and thus admits a Markov partition. In Examples 4.4.4 and 4.4.5, explicit partitions were given. In Example 4.5.3, the coding map generated by the partition $\mathcal{R} = \{R_i = [\frac{i}{n}, \frac{i+1}{n}] : 0 \leq i < n\}$ was identified as

$$\begin{array}{rccc} \pi : & E^\infty & \longrightarrow & \mathbb{S}^1 \\ & \omega = (\omega_k)_{k=1}^\infty & \longmapsto & \pi(\omega) = \displaystyle\sum_{k=1}^\infty \frac{\omega_k}{n^k}. \end{array}$$

Properties of π were given in Theorem 4.5.2. Among others, π is continuous, surjective, and its restriction to the set $Z := \pi^{-1}(\mathbb{S}^1 \backslash \bigcup_{k=0}^{\infty} T_n^{-k}(\bigcup_{i=0}^{n-1} \partial R_i))$ is injective. The set Z consists of all $\omega \in E^\infty$ whose coordinates are not eventually constant and equal to 0 or $n-1$, i. e.

$$E^\infty \backslash Z = \bigcup_{\tau \in E^*} \{\tau 0^\infty, \tau(n-1)^\infty\}.$$

The coding map π is two-to-one on $E^\infty \backslash Z$, which is a countable set and thus has μ_P-measure zero. So $\mu_P(Z) = 1$.

The coding map $\pi : (E^\infty, \mathcal{F}_\infty, \mu_P) \to (\mathbb{S}^1, \mathcal{B}(\mathbb{S}^1), \lambda)$ is measure-preserving. Indeed, it is easy to show that the family

$$\mathcal{P} := \left\{ \left[\frac{i}{n^k}, \frac{i+1}{n^k} \right] : 0 \le i < n^k,\ k \in \mathbb{N} \right\} \bigcup \left\{ \frac{i}{n^k} : 0 \le i < n^k,\ k \in \mathbb{N} \right\}$$

is a π-system that generates $\mathcal{B}(\mathbb{S}^1)$ and $\mu_P \circ \pi^{-1}(P) = \lambda(P)$ for all $P \in \mathcal{P}$. Hence the coding map is measure-preserving according to Lemma 8.1.4.

Let $\varphi : E^\infty \to E \subseteq \mathbb{R}$ be the function $\varphi(\omega) = \omega_1$. Clearly, $\varphi \in L^1(E^\infty, \mathcal{F}_\infty, \mu_P)$. Furthermore, $\varphi \circ \sigma^j(\omega) = \omega_{j+1}$ for all $j \ge 0$. Since μ_P is ergodic with respect to the shift map σ, the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14) asserts that there is $M \in \mathcal{F}_\infty$ such that $\mu_P(M) = 1$ and

$$\begin{aligned}
\lim_{k\to\infty} \frac{1}{k} \sum_{j=1}^{k} \omega_j = \lim_{k\to\infty} \frac{1}{k} S_k\varphi(\omega) &= \int_{E^\infty} \varphi(\omega)\, d\mu_P(\omega) \\
&= \sum_{i=0}^{n-1} \int_{[i]} i\, d\mu_P = \sum_{i=0}^{n-1} i\, \mu_P([i]) = \sum_{i=0}^{n-1} i \cdot \frac{1}{n} \\
&= \frac{n-1}{2}, \quad \forall \omega \in M.
\end{aligned}$$

As π is continuous, it is Borel measurable and thus $\pi(M)$ is Lebesgue measurable. Since π is measure-preserving, we have that $\lambda(\pi(M)) = \mu_P \circ \pi^{-1}(\pi(M)) \ge \mu_P(M) = 1$. We infer that the digits of the n-adic expansion of λ-almost every number between 0 and 1 average $(n-1)/2$ asymptotically.

Now, fix any $0 \le i < n$. Since μ_P is ergodic with respect to σ, Corollary 8.2.15 of Birkhoff's ergodic theorem affirms that for μ_P-a. e. $\omega \in E^\infty$,

$$\lim_{k\to\infty} \frac{1}{k} \#\{1 \le j \le k : \omega_j = i\} = \lim_{k\to\infty} \frac{1}{k} \#\{0 \le j < k : \sigma^j(\omega) \in [i]\} = \mu_P([i]) = \frac{1}{n}.$$

Since the coding map is measure-preserving, we deduce that λ-almost every number between 0 and 1 has a n-adic expansion whose digits are equal to i with a frequency of $1/n$. This frequency is independent of the digit i, as one naturally expects.

8.2.4 Uniquely ergodic transformations

As mentioned in Subsection 8.1.3, there are measurable transformations that do not admit any invariant measure. However, in Theorem 8.1.22 we showed that every topological dynamical system carries invariant probability measures. Among those systems, some have only one such measure. Per Corollary 8.2.25, this happens precisely when there is a unique ergodic invariant measure. These maps deserve a special name.

Definition 8.2.34. A measurable transformation $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ is said to be *uniquely ergodic* if $E(T, \mathcal{A})$ is a singleton.

Lemma 8.2.35. *A topological dynamical system $T : X \to X$ is uniquely ergodic if and only if $M(T)$ is a singleton.*

Proof. According to Corollary 8.2.25, $M(T)$ is the closed convex hull of $E(T)$. Therefore, $M(T)$ is a singleton precisely when $E(T)$ is. □

Recall that by Riesz' representation theorem (Theorem A.1.53), whenever X is a compact metrizable space, every $\mu \in M(X)$ is uniquely determined by a normalized positive continuous linear functional $\ell_\mu \in C(X)^*$, namely

$$\ell_\mu(f) = \int_X f \, d\mu, \quad \forall f \in C(X).$$

If $T : X \to X$ is a topological dynamical system, then by Theorem 8.1.18 each $\mu \in M(T)$ corresponds to a T-invariant functional in the sense that

$$\ell_\mu(f \circ T) = \ell_\mu(f), \quad \forall f \in C(X).$$

We will show that if T is uniquely ergodic, then all T-invariant continuous linear functionals are scalar multiples of ℓ_{μ_0}, where μ_0 is the unique ergodic T-invariant measure.

First, we introduce the variation of a functional.

Definition 8.2.36. Let $\ell \in C(X)^*$. The *variation* of ℓ is the functional $\mathrm{var}(\ell) : C(X) \to \mathbb{R}$ defined as follows: For any function $f \in C(X)$, $f \geq 0$, set

$$\mathrm{var}(\ell)(f) := \sup_{\substack{g \in C(X) \\ 0 \leq g \leq f}} \ell(g),$$

and for any other $f \in C(X)$ let

$$\mathrm{var}(\ell)(f) := \mathrm{var}(\ell)(f_+) - \mathrm{var}(\ell)(f_-).$$

Lemma 8.2.37. *Let $\ell \in C(X)^*$. Set $\Delta\ell := \mathrm{var}(\ell) - \ell$. Then $\mathrm{var}(\ell), \Delta\ell \in C(X)^*$ and both are positive. In addition, if ℓ is T-invariant then so are $\mathrm{var}(\ell)$ and $\Delta\ell$.*

Proof. Let $f \in C(X), f \geq 0$. The positivity of $\mathrm{var}(\ell)$ is obvious since $\mathrm{var}(\ell)(f) \geq \ell(0) = 0$. It is also easy to see that $\mathrm{var}(\ell)(cf) = c \cdot \mathrm{var}(\ell)(f)$ for all $c \geq 0$. Now, let $f_1, f_2 \in C(X)$, $f_1, f_2 \geq 0$. On one hand, if $g_1, g_2 \in C(X)$ satisfy $0 \leq g_1 \leq f_1$ and $0 \leq g_2 \leq f_2$, then $0 \leq g_1 + g_2 \leq f_1 + f_2$, and hence

$$\ell(g_1) + \ell(g_2) = \ell(g_1 + g_2) \leq \mathrm{var}(\ell)(f_1 + f_2).$$

Taking the supremum over all such g_1, g_2, we get

$$\mathrm{var}(\ell)(f_1) + \mathrm{var}(\ell)(f_2) \leq \mathrm{var}(\ell)(f_1 + f_2). \tag{8.17}$$

On the other hand, let $g \in C(X)$ be such that $0 \leq g \leq f_1 + f_2$. Define $g_1 = \min\{f_1, g\}$ and $g_2 = g - g_1$. Then $g_1, g_2 \in C(X)$ and satisfy $0 \leq g_1 \leq f_1$ and $0 \leq g_2 \leq f_2$. It follows that

$$\ell(g) = \ell(g_1 + g_2) = \ell(g_1) + \ell(g_2) \leq \mathrm{var}(\ell)(f_1) + \mathrm{var}(\ell)(f_2).$$

Taking the supremum over all such g, we get

$$\mathrm{var}(\ell)(f_1 + f_2) \leq \mathrm{var}(\ell)(f_1) + \mathrm{var}(\ell)(f_2). \tag{8.18}$$

By (8.17) and (8.18),

$$\mathrm{var}(\ell)(f_1 + f_2) = \mathrm{var}(\ell)(f_1) + \mathrm{var}(\ell)(f_2).$$

This proves the linearity for nonnegative functions. The linearity for other functions follows directly from the definition of $\mathrm{var}(\ell)$ for such functions.

Now, suppose that ℓ is T-invariant. Given $f \in C(X), f \geq 0$, first notice that

$$\mathrm{var}(\ell)(f \circ T) = \sup_{\substack{g \in C(X) \\ 0 \leq g \leq f \circ T}} \ell(g) \geq \sup_{\substack{h \in C(X) \\ 0 \leq h \leq f}} \ell(h \circ T) = \sup_{\substack{h \in C(X) \\ 0 \leq h \leq f}} \ell(h) = \mathrm{var}(\ell)(f). \tag{8.19}$$

We shall prove that this inequality implies the desired equality. To do this, let us pass from functionals to measures. Let ν be the Borel measure corresponding to $\mathrm{var}(\ell)$. We shall show that (8.19) implies that $\nu \circ T^{-1}(B) \geq \nu(B)$ for all Borel sets B. Since ν is a regular measure (by Theorem A.1.24), it suffices to prove the inequality for closed sets. Let F be a closed set in X. By Urysohn's lemma (see 15.6 in Willard [77]), there is a descending sequence of open sets $(U_n)_{n=1}^{\infty}$ whose intersection is F and to which corresponds a sequence $(f_n)_{n=1}^{\infty}$ of continuous functions on X such that $0 \leq f_n \leq 1$, $f_n = 0$ on $X \backslash U_n$ and $f_n = 1$ on F. Then, for all $n \in \mathbb{N}$,

$$\int_X f_n \, d(\nu \circ T^{-1}) = \int_X f_n \circ T \, d\nu = \mathrm{var}(\ell)(f_n \circ T) \geq \mathrm{var}(\ell)(f_n) = \int_X f_n \, d\nu.$$

Since $\lim_{n\to\infty} f_n = \mathbb{1}_F$ and $\|f_n\|_\infty \leq 1$ for all $n \in \mathbb{N}$, Lebesgue's dominated convergence theorem (Theorem A.1.38) implies that

$$\nu \circ T^{-1}(F) = \int_X \mathbb{1}_F \, d(\nu \circ T^{-1}) = \lim_{n\to\infty} \int_X f_n \, d(\nu \circ T^{-1}) \geq \lim_{n\to\infty} \int_X f_n \, d\nu = \int_X \mathbb{1}_F \, d\nu = \nu(F).$$

As the closed set F was arbitrarily chosen and the Borel measure ν is regular, the measure of any Borel set B is equal to the supremum of the measures of all closed sets contained in B. It immediately follows that

$$\nu \circ T^{-1}(B) \geq \nu(B), \quad \forall B \in \mathcal{B}(X). \tag{8.20}$$

Replacing B by $X \backslash B$ in (8.20), we obtain that

$$\nu \circ T^{-1}(X \backslash B) \geq \nu(X \backslash B), \quad \forall B \in \mathcal{B}(X). \tag{8.21}$$

Since $\nu(B) + \nu(X\backslash B) = \nu(X) = \nu \circ T^{-1}(B) + \nu \circ T^{-1}(X\backslash B)$, inequality (8.21) implies that

$$\nu \circ T^{-1}(B) \leq \nu(B), \quad \forall B \in \mathcal{B}(X). \tag{8.22}$$

From (8.20) and (8.22), we deduce that $\nu \circ T^{-1}(B) = \nu(B)$ for all $B \in \mathcal{B}(X)$, that is, ν is T-invariant. It follows that $\mathrm{var}(\ell)$ is T-invariant since

$$\mathrm{var}(\ell)(f \circ T) = \int_X f \circ T \, d\nu = \int_X f \, d(\nu \circ T^{-1}) = \int_X f \, d\nu = \mathrm{var}(\ell)(f).$$

For any $f \in C(X)$, it ensues that

$$\mathrm{var}(\ell)(f \circ T) = \mathrm{var}(\ell)(f_+ \circ T) - \mathrm{var}(\ell)(f_- \circ T) = \mathrm{var}(\ell)(f_+) - \mathrm{var}(\ell)(f_-) = \mathrm{var}(\ell)(f).$$

The proof of the statements on $\Delta\ell$ are left to the reader. □

We use the variation functional to demonstrate that uniquely ergodic systems admit only T-invariant functionals that are multiples of the ergodic T-invariant measure.

Lemma 8.2.38. *Let $T : X \to X$ be a uniquely ergodic topological dynamical system. Let μ_0 be the unique ergodic T-invariant measure and ℓ_{μ_0} its corresponding T-invariant normalized positive continuous linear functional. Then any (not necessarily positive or normalized) T-invariant $\ell \in C(X)^*$ is of the form*

$$\ell = c \cdot \ell_{\mu_0},$$

where $c \in \mathbb{R}$.

Proof. Assume that $\ell \in C(X)^*$ is T-invariant. Lemma 8.2.37 then says that $\mathrm{var}(\ell)$ and $\Delta\ell$ are T-invariant positive continuous linear functionals on $C(X)$. Since T is uniquely ergodic, Lemma 8.2.35 implies that there must exist $C, \widetilde{C} \geq 0$ such that $\mathrm{var}(\ell) = C\,\ell_{\mu_0}$ and $\Delta\ell = \widetilde{C}\,\ell_{\mu_0}$. It then follows that $\ell = \mathrm{var}(\ell) - \Delta\ell = (C - \widetilde{C})\ell_{\mu_0}$. □

We aim to show that a stronger variant of Birkhoff's ergodic theorem (Theorem 8.2.11) holds for uniquely ergodic dynamical systems. More precisely, the Birkhoff averages converge uniformly and thereby everywhere. The proof relies upon a deep result in functional analysis called the Hahn–Banach theorem. The statement and proof of this theorem can be found as Theorem II.3.10 in Dunford and Schwartz [20].

In light of the existence of an invariant Borel probability measure for every topological dynamical system, we will first introduce a set of functions which will play an important role on multiple occasions in the sequel.

Definition 8.2.39. Let $T : X \to X$ be a topological dynamical system. A function $f \in C(X)$ is said to be *cohomologous to zero* in the additive group $C(X)$ if

$$f = g \circ T - g$$

for some $g \in C(X)$. The set of such functions will be denoted by $C_0(T)$.

These functions have the property that their integral with respect to any invariant measure is equal to zero. Moreover, when the system is uniquely ergodic, every function whose integral is equal to zero can be approximated by functions that are cohomologous to zero, as the following lemma shows.

Lemma 8.2.40. *Let $T : X \to X$ be a topological dynamical system and $\mu \in M(T)$. Let*

$$C_0(\mu) := \left\{ f \in C(X) \;\middle|\; \int_X f \, d\mu = 0 \right\}.$$

Then $C_0(T)$ and $C_0(\mu)$ are vector subspaces of $C(X)$. Moreover, $C_0(\mu)$ is closed in $C(X)$ and $\overline{C_0(T)} \subseteq C_0(\mu)$. In addition, if T is uniquely ergodic then $\overline{C_0(T)} = C_0(\mu)$.

Proof. It is easy to see that $C_0(T)$ and $C_0(\mu)$ are vector subspaces of $C(X)$, and that $C_0(\mu)$ is closed in $C(X)$ (recall that this latter is endowed with the topology of uniform convergence). Let $f \in C_0(T)$. Then $f = g \circ T - g$ for some $g \in C(X)$. The T-invariance of μ yields

$$\int_X f \, d\mu = \int_X g \circ T \, d\mu - \int_X g \, d\mu = \int_X g \, d(\mu \circ T^{-1}) - \int_X g \, d\mu = 0.$$

So $C_0(T) \subseteq C_0(\mu)$, and hence $\overline{C_0(T)} \subseteq \overline{C_0(\mu)} = C_0(\mu)$.

Assume now that T is uniquely ergodic. Suppose by way of contradiction that there exists $f_0 \in C_0(\mu) \backslash \overline{C_0(T)}$. According to the Hahn–Banach theorem, there is $\ell \in C(X)^*$ such that $\ell(f) = 0$ for all $f \in \overline{C_0(T)}$ whereas $\ell(f_0) = 1$. By Lemma 8.2.38, there exists $c \in \mathbb{R}$ such that $\ell = c \cdot \ell_\mu$. But this is impossible since $\ell_\mu(f_0) = 0$ while $\ell(f_0) = 1$. This contradiction implies that $\overline{C_0(T)} = C_0(\mu)$ when T is uniquely ergodic. □

We can now state a stronger version of the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14) for uniquely ergodic dynamical systems.

Theorem 8.2.41. *Let $T : X \to X$ be a topological dynamical system and $\mu \in M(T)$. The following statements are equivalent:*

(a) *T is uniquely ergodic.*
(b) *For every $f \in C(X)$, the Birkhoff averages $\frac{1}{n}S_n f(x)$ converge to $\int_X f\,d\mu$ for all $x \in X$.*
(c) *For every $f \in C(X)$, the Birkhoff averages $\frac{1}{n}S_n f$ converge uniformly to $\int_X f\,d\mu$.*

Proof. The structure of the proof will be the following sequence of implications: (a)⇒(c)⇒(b)⇒(a).
[(a)⇒(c)] Suppose that T is uniquely ergodic. For any $f \in C_0(T)$ and any $x \in X$ we have

$$\left|\frac{1}{n}S_n f(x)\right| = \left|\frac{1}{n}S_n g(T(x)) - \frac{1}{n}S_n g(x)\right| = \frac{1}{n}\left|g(T^n(x)) - g(x)\right| \le \frac{2}{n}\|g\|_\infty.$$

Thus

$$\lim_{n\to\infty}\left\|\frac{1}{n}S_n f - 0\right\|_\infty = 0, \quad \forall f \in C_0(T). \tag{8.23}$$

In Lemma 8.2.40, we observed that $C_0(T) \subseteq C_0(\mu)$. Therefore, the sequence $(\frac{1}{n}S_n f)_{n=1}^{\infty}$ converges uniformly on X to $\int_X f\,d\mu$ for every $f \in C_0(T)$.

Now, suppose that $f \in C_0(\mu)$. Since T is uniquely ergodic, Lemma 8.2.40 asserts that $\overline{C_0(T)} = C_0(\mu)$. Let $\varepsilon > 0$ and choose $f_\varepsilon \in C_0(T)$ such that $\|f - f_\varepsilon\|_\infty \le \varepsilon$. Then

$$\begin{aligned}\left\|\frac{1}{n}S_n f - \int_X f\,d\mu\right\|_\infty &\le \left\|\frac{1}{n}S_n f - \frac{1}{n}S_n f_\varepsilon\right\|_\infty + \left\|\frac{1}{n}S_n f_\varepsilon - \int_X f_\varepsilon\,d\mu\right\|_\infty + \left\|\int_X f_\varepsilon\,d\mu - \int_X f\,d\mu\right\|_\infty \\ &\le \frac{1}{n}\sum_{k=0}^{n-1}\|f\circ T^k - f_\varepsilon \circ T^k\|_\infty + \left\|\frac{1}{n}S_n f_\varepsilon - \int_X f_\varepsilon\,d\mu\right\|_\infty + 0 \\ &\le \|f - f_\varepsilon\|_\infty + \left\|\frac{1}{n}S_n f_\varepsilon - \int_X f_\varepsilon\,d\mu\right\|_\infty \\ &\le \varepsilon + \left\|\frac{1}{n}S_n f_\varepsilon - \int_X f_\varepsilon\,d\mu\right\|_\infty.\end{aligned}$$

As $f_\varepsilon \in C_0(T)$, relation (8.23) guarantees that $\lim_{n\to\infty}\|\frac{1}{n}S_n f_\varepsilon - \int_X f_\varepsilon\,d\mu\|_\infty = 0$ and we deduce that

$$\limsup_{n\to\infty}\left\|\frac{1}{n}S_n f - \int_X f\,d\mu\right\|_\infty \le \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the implication (a)⇒(c) is proved for any function $f \in C_0(\mu)$ and, therefore, for any $f \in C(X)$ by replacing f by $f - \int_X f\,d\mu$.
[(c)⇒(b)] This is obvious.
[(b)⇒(a)] Let $\nu \in M(T)$ and $f \in C(X)$. Since $\lim_{n\to\infty}\frac{1}{n}S_n f(x) = \int_X f\,d\mu$ for all $x \in X$ and $\|\frac{1}{n}S_n f\|_\infty \le \|f\|_\infty < \infty$ for all $n \in \mathbb{N}$, Lebesgue's dominated convergence theorem (Theorem A.1.38) asserts that

$$\lim_{n\to\infty}\int_X \frac{1}{n}S_n f(x)\,d\nu(x) = \int_X\left(\int_X f\,d\mu\right)d\nu(x) = \int_X f\,d\mu.$$

On the other hand, the T-invariance of ν means that $\nu = \nu \circ T^{-k}$ for all $k \in \mathbb{N}$. Thus for all $n \in \mathbb{N}$ we have that $\nu = \frac{1}{n}\sum_{k=0}^{n-1} \nu \circ T^{-k}$, and hence

$$\int_X \frac{1}{n} S_n f(x)\, d\nu(x) = \int_X \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \, d\nu = \frac{1}{n} \sum_{k=0}^{n-1} \int_X f \, d(\nu \circ T^{-k}) = \int_X f \, d\nu.$$

From the last two formulas, it follows that $\int_X f \, d\nu = \int_X f \, d\mu$ for all $f \in C(X)$. By Corollary A.1.54, we conclude that $\nu = \mu$. So μ is the unique T-invariant measure. □

The previous theorem has the following consequence in topological dynamics. Compare this with Theorem 8.2.27.

Corollary 8.2.42. *Let $T : X \to X$ be a uniquely ergodic topological dynamical system. If* $\operatorname{supp}(\mu) = X$ *for the unique $\mu \in M(T) = E(T)$, then T is minimal.*

Proof. Let U be a nonempty open set in X. Then $\mu(U) > 0$ because $\operatorname{supp}(\mu) = X$. By Theorem A.1.24, μ is regular. Hence, $\mu(U) = \sup\{\mu(F) : F \subseteq U,\ F \text{ closed}\} > 0$ and thus there is a closed set $F_0 \subseteq U$ such that $\mu(F_0) > 0$. The sets F_0 and $X \backslash U$ are disjoint closed sets and Urysohn's lemma (see 15.6 in Willard [77]) states that there is a nonnegative function $f \in C(X)$ with the properties that $f = 1$ on F_0 and $f = 0$ on $X \backslash U$. Notice that $\int_X f \, d\mu \geq \mu(F_0) > 0$. Consequently, for any $x \in X$, Theorem 8.2.41 asserts that

$$\lim_{n\to\infty} \frac{1}{n} S_n f(x) = \int_X f \, d\mu > 0.$$

Hence, there exists $n \in \mathbb{N}$ such that $f(T^n(x)) > 0$ and so $T^n(x) \in U$. But since U was chosen arbitrarily, we conclude that the orbit of x visits every open set in X; in other words, the orbit of x is dense in X. Since x was chosen arbitrarily, every orbit is dense, and thus the system T is minimal according to Theorem 1.5.4. □

We now revisit the rotations (sometimes called translations) of the torus. Recall that these rotations are a case of rotations of topological groups (see Subsection 1.6.1). According to Haar's theorem, there is, up to a positive multiplicative constant, a unique measure μ on the Borel subsets of a topological group G satisfying the following properties:

(a) μ is left-translation-invariant: $\mu(gS) = \mu(S)$ for every $g \in G$ and all Borel sets $S \subseteq G$.
(b) μ is finite on every compact set: $\mu(K) < \infty$ for all compact $K \subseteq G$.
(c) μ is outer regular on Borel sets $S \subseteq G$: $\mu(S) = \inf\{\mu(U) : S \subseteq U,\ U \text{ open}\}$.
(d) μ is inner regular on open sets $U \subseteq G$: $\mu(U) = \sup\{\mu(K) : K \subseteq U,\ K \text{ compact}\}$.

Such a measure is called a left Haar measure. As a consequence of the above properties, it also turns out that $\mu(U) > 0$ for every nonempty open subset $U \subseteq G$. In particular, if G is compact then $0 < \mu(G) < \infty$. Thus, we can uniquely specify a left Haar measure on G by adding the normalization condition $\mu(G) = 1$. This is obviously the case for the n-dimensional torus, where μ will be denoted by λ_n. This is the n-dimensional Lebesgue measure on the torus.

Proposition 8.2.43. *Let $L_\gamma : \mathbb{T}^n \to \mathbb{T}^n$ be a translation of the torus, where $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n) \in \mathbb{T}^n$. The following statements are equivalent:*
(a) *The numbers $1, \gamma_1, \gamma_2, \ldots, \gamma_n$ are linearly independent over $\mathbb{Q}$.*
(b) *L_γ is minimal.*
(c) *L_γ is transitive.*
(d) *L_γ is ergodic with respect to λ_n.*
(e) *L_γ is uniquely ergodic.*

Proof. Theorem 1.6.3 already asserted the equivalencies (a)⇔(b)⇔(c). It is also obvious that (e)⇒(d). Implication (d)⇒(c) follows from Theorem 8.2.27. It remains to prove that (a)⇒(e). To that end, assume that $1, \gamma_1, \ldots, \gamma_n$ are linearly independent over $\mathbb{Q}$. We will first show that if $f \circ L_\gamma = f$ for some $f \in L^2(\lambda_n)$, then f is λ_n-a. e. constant. It will then follow from Theorem 8.2.18 and Remark 8.2.19 that λ_n is ergodic with respect to L_γ. Consider the Fourier series representation of f:

$$f(x) = \sum_{k\in\mathbb{Z}^n} a_k e^{2\pi i\langle k,x\rangle}, \quad \text{where } \langle k, x\rangle = \sum_{j=1}^{n} k_j x_j.$$

Then

$$f \circ L_\gamma(x) = \sum_{k\in\mathbb{Z}^n} a_k e^{2\pi i\langle k,x+\gamma\rangle} = \sum_{k\in\mathbb{Z}^n} a_k e^{2\pi i\langle k,\gamma\rangle} e^{2\pi i\langle k,x\rangle}.$$

The above equalities are understood to hold in $L^2(\lambda_n)$ and hold only for λ_n-a. e. $x \in \mathbb{T}^n$. As $f \circ L_\gamma = f$, we deduce from the uniqueness of the Fourier series representation that

$$a_k e^{2\pi i\langle k,\gamma\rangle} = a_k, \quad \forall k \in \mathbb{Z}^n.$$

Hence, for each $k \in \mathbb{Z}^n$ we have $a_k = 0$ or $e^{2\pi i\langle k,\gamma\rangle} = 1$. The latter condition holds if and only if $\langle k, \gamma\rangle \in \mathbb{Z}$. As $1, \gamma_1, \ldots, \gamma_n$ are linearly independent over $\mathbb{Q}$, this happens if and only if $k_1 = k_2 = \cdots = k_n = 0$. Thus $f(x) = a_{(0,\ldots,0)}$ for λ_n-a. e. $x \in \mathbb{T}^n$. That is, f is λ_n-a. e. constant. This implies that λ_n is ergodic. It only remains to show that λ_n is the unique ergodic L_γ-invariant measure. Let $\varphi \in C(\mathbb{T}^n)$. By the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14), for λ_n-a. e. $x \in \mathbb{T}^n$ we have

$$\lim_{k\to\infty} \frac{1}{k} S_k\varphi(x) = \int_{\mathbb{T}^n} \varphi \, d\lambda_n. \tag{8.24}$$

Let $x_0 \in \mathbb{T}^n$ and $\varepsilon > 0$. Since φ is uniformly continuous on the compact metric space $\mathbb{T}^n$, there exists $\delta > 0$ such that

$$\|x' - x\| < \delta \Longrightarrow |\varphi(x') - \varphi(x)| < \varepsilon.$$

Furthermore, as $\operatorname{supp}(\lambda_n) = \mathbb{T}^n$, there exists $x_1 \in \mathbb{T}^n$ such that (8.24) holds for x_1 and $\|x_0 - x_1\| < \delta$. Bearing in mind that L_γ is an isometry, for all $k \in \mathbb{N}$ it follows that

$$\begin{aligned}\left|\frac{1}{k}S_k\varphi(x_0) - \int_{\mathbb{T}^n} \varphi\, d\lambda_n\right| &\leq \left|\frac{1}{k}S_k\varphi(x_0) - \frac{1}{k}S_k\varphi(x_1)\right| + \left|\frac{1}{k}S_k\varphi(x_1) - \int_{\mathbb{T}^n} \varphi\, d\lambda_n\right| \\ &\leq \frac{1}{k}\sum_{j=0}^{k-1}\left|\varphi(L_\gamma^j(x_0)) - \varphi(L_\gamma^j(x_1))\right| + \left|\frac{1}{k}S_k\varphi(x_1) - \int_{\mathbb{T}^n} \varphi\, d\lambda_n\right| \\ &< \frac{1}{k}\sum_{j=0}^{k-1}\varepsilon + \left|\frac{1}{k}S_k\varphi(x_1) - \int_{\mathbb{T}^n} \varphi\, d\lambda_n\right| \\ &= \varepsilon + \left|\frac{1}{k}S_k\varphi(x_1) - \int_{\mathbb{T}^n} \varphi\, d\lambda_n\right|.\end{aligned}$$

Letting $k \to \infty$, we deduce that $\left|\lim_{k\to\infty} \frac{1}{k}S_k\varphi(x_0) - \int_{\mathbb{T}^n} \varphi\, d\lambda_n\right| \leq \varepsilon$. As $\varepsilon > 0$ was arbitrary, we conclude that $\lim_{k\to\infty} \frac{1}{k}S_k\varphi(x_0) = \int_{\mathbb{T}^n} \varphi\, d\lambda_n$. But x_0 was chosen arbitrarily in $\mathbb{T}^n$ and so L_γ is uniquely ergodic by Theorem 8.2.41. □

Remark 8.2.44.
(a) This proof in fact shows that any isometry T on a compact metrizable space X that admits an invariant probability measure which is ergodic and of full topological support, is uniquely ergodic.
(b) Being an isometry can be weakened by requiring only that the iterates $\{T^n\}_{n=1}^{\infty}$ form an equicontinuous family.

8.3 Mixing transformations

In the penultimate section of this chapter, we introduce various notions of mixing for measure-preserving dynamical systems. These should be contrasted with topological mixing, which was introduced in Section 1.5. These measure-theoretical mixing forms are stronger than ergodicity in the sense that they all imply ergodicity, and are important from a statistical viewpoint (for instance, for decay of correlations and similar questions).

8.3.1 Weak mixing

Definition 8.3.1. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. The system T is said to be *weakly mixing* if

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1}\left|\mu(T^{-j}(A) \cap B) - \mu(A)\mu(B)\right| = 0, \quad \forall A, B \in \mathcal{B}. \tag{8.25}$$

Like for ergodicity, to find out if a system is weakly mixing it suffices to check weak mixing on a semialgebra that generates the σ-algebra.

Lemma 8.3.2. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. If $\mathcal{B} = \sigma(\mathcal{S})$ for some semialgebra $\mathcal{S}$, then T is weakly mixing if and only if relation* (8.25) *holds for all $A, B \in \mathcal{S}$.*

Proof. The proof is nearly identical to that of Lemma 8.2.17. Simply replace the square brackets by absolute values. □

Weak mixing is a stronger property than ergodicity. This is not surprising if you compare the definition of weak mixing with the characterization of ergodicity given in Lemma 8.2.16. Nevertheless, we will give a more direct proof of that fact.

Lemma 8.3.3. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. If T is weakly mixing, then T is ergodic.*

Proof. Suppose by way of contradiction that T is weakly mixing but not ergodic. Thus there exists a completely T-invariant set $E \in \mathcal{B}$ with $\mu(E) > 0$ and $\mu(X \backslash E) > 0$. Then $T^{-j}(E) \cap (X \backslash E) = E \cap (X \backslash E) = \emptyset$ for all $j \in \mathbb{N}$. Setting $A = E$ and $B = X \backslash E$ in (8.25), we deduce that $\mu(E)\mu(X \backslash E) = 0$. So $\mu(E) = 0$ or $\mu(X \backslash E) = 0$. This contradiction shows that T is ergodic. □

The converse of this lemma is not true; that is to say, there exist dynamical systems that are ergodic but not weakly mixing. We now provide such an example.

Example 8.3.4. Let $\alpha \in \mathbb{R} \backslash \mathbb{Q}$ and consider again the map $T_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $T_\alpha(x) := \langle x + \alpha \rangle$, where $\langle r \rangle$ denotes the fractional part of r. This is the rotation of the unit circle $\mathbb{S}^1$ by the angle $2\pi\alpha$. We saw in Proposition 8.2.29 that this map is ergodic with respect to the Lebesgue measure λ on $\mathbb{S}^1$. We shall now show that it is not weakly mixing.

By Corollary 8.2.15, for any interval $I \subseteq \mathbb{S}^1$ and λ-almost every $x \in \mathbb{S}^1$,

$$\lim_{M \to \infty} \frac{1}{M} \#\{0 \le n < M : T_\alpha^n(x) \in I\} = \lambda(I).$$

In other words, for λ-almost every $x \in \mathbb{S}^1$, the sequence $(\langle x + n\alpha \rangle)_{n=1}^\infty$ is uniformly distributed in $\mathbb{S}^1$. It follows, upon rotating by $-x$, that the sequence $(\langle n\alpha \rangle)_{n=1}^\infty$ is uniformly distributed in $\mathbb{S}^1$. Let $A = B = (0, 1/2)$ and let $(n_i)_{i=1}^\infty$ be the subsequence of $\mathbb{N}$ such that $\langle n_i \alpha \rangle \in (0, 1/10)$. Then $T_\alpha^{-n_i}(A) \supseteq [0, 4/10]$ and hence

$$\lambda\big(T_\alpha^{-n_i}(A) \cap B\big) - \lambda(A)\lambda(B) \ge \lambda\big((0, 4/10]\big) - \big(\lambda\big((0, 1/2)\big)\big)^2 = \frac{4}{10} - \frac{1}{4} = \frac{3}{20}.$$

Consequently,

$$\begin{aligned}
\liminf_{N \to \infty} \frac{1}{n_N} \sum_{k=0}^{n_N} \big|\lambda\big(T_\alpha^{-k}(A) \cap B\big) - \lambda(A)\lambda(B)\big| &\ge \liminf_{N \to \infty} \frac{1}{n_N} \sum_{i=1}^{N} \big|\lambda\big(T_\alpha^{-n_i}(A) \cap B\big) - \lambda(A)\lambda(B)\big| \\
&\ge \liminf_{N \to \infty} \frac{1}{n_N} \cdot N \cdot \frac{3}{20}
\end{aligned}$$

$$\geq \frac{3}{20} \liminf_{M\to\infty} \frac{1}{M} \#\{0 \leq n < M : \langle n\alpha \rangle \in (0, 1/10)\}$$
$$= \frac{3}{20} \cdot \frac{1}{10} > 0.$$

Therefore, T_α cannot be weakly mixing.

The following lemma provides a characterization of weakly mixing dynamical systems.

Lemma 8.3.5. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. The following statements are equivalent:*
(a) *T is weakly mixing.*
(b) *For all $f, g \in L^2(\mu)$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \left| \int_X (f \circ T^j) \cdot g \, d\mu - \int_X f \, d\mu \int_X g \, d\mu \right| = 0.$$

Proof. That (b) implies (a) follows upon choosing the functions $f = \mathbb{1}_A$ and $g = \mathbb{1}_B$. For the converse, a straightforward argument involving approximation by simple functions is enough to complete the proof. We leave the details as an exercise. □

The following result will be used to give alternative formulations of weak mixing. A subset J of $\mathbb{Z}_+$ is said to have *density zero* if

$$\lim_{n\to\infty} \frac{\#(J \cap \{0, 1, \ldots, n-1\})}{n} = 0.$$

Theorem 8.3.6. *If $(a_n)_{n=0}^\infty$ is a bounded sequence in $\mathbb{R}$, then the following statements are equivalent:*
(a) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0$.
(b) *There exists a set $J \subseteq \mathbb{Z}_+$ of density zero such that* $\lim_{J \not\ni n\to\infty} a_n = 0$.
(c) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} a_i^2 = 0$.

Proof. We shall first prove that (a)⇔(b) and then (b)⇔(c).

[(a)⇒(b)] Suppose that $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0$. To lighten notation, let $c_J(n) = \#(J \cap \{0, 1, \ldots, n-1\})$. For each $k \in \mathbb{N}$, define

$$J_k := \left\{ i \geq 0 : |a_i| \geq \frac{1}{k} \right\}.$$

Then $(J_k)_{k=1}^\infty$ is an ascending sequence of sets. We claim that each J_k has density zero. Indeed, for each $k \in \mathbb{N}$,

$$\frac{1}{n} \sum_{i=0}^{n-1} |a_i| \geq \frac{1}{n} \sum_{\substack{i=0 \\ i \in J_k}}^{n-1} |a_i| \geq \frac{1}{n} \frac{1}{k} c_{J_k}(n).$$

Thus $\lim_{n\to\infty}\frac{1}{n}\frac{1}{k}c_{J_k}(n)=0$. This implies that $\lim_{n\to\infty}\frac{1}{n}c_{J_k}(n)=0$, that is, each J_k has density zero. Therefore, there exists a strictly increasing sequence $(\ell_k)_{k=1}^{\infty}$ in $\mathbb{N}$ such that for every $k\in\mathbb{N}$,

$$\frac{1}{n}c_{J_k}(n)<\frac{1}{k},\quad \forall n\geq \ell_k.$$

Set

$$J:=\bigcup_{k=1}^{\infty}J_k\cap[\ell_k,\ell_{k+1}).$$

We claim that J has density zero. Indeed, since the sets $(J_k)_{k=1}^{\infty}$ form an ascending sequence, for every $\ell_k\leq n<\ell_{k+1}$ we have

$$J\cap[0,n)\subseteq J_k\cap[0,n)$$

and so

$$\frac{1}{n}c_J(n)\leq\frac{1}{n}c_{J_k}(n)<\frac{1}{k}.$$

Letting $n\to\infty$ imposes $k\to\infty$ and hence $\lim_{n\to\infty}\frac{1}{n}c_J(n)=0$. So J has density zero, as claimed.

Moreover, if $n\geq\ell_k$ and $n\notin J$, then $n\notin J_k$ and thus $|a_n|<1/k$. Therefore,

$$\lim_{J\not\ni n\to\infty}|a_n|=0.$$

[(b)⇒(a)] For the opposite implication, suppose that $\lim_{J\not\ni n\to\infty}|a_n|=0$ for some set $J\subseteq\mathbb{Z}_+$ of density zero. Since $(a_n)_{n=0}^{\infty}$ is bounded, let $B\geq 0$ be such that $|a_n|\leq B$ for all $n\geq 0$. Fix $\varepsilon>0$. There exists $N(\varepsilon)\in\mathbb{N}$ such that $|a_n|<\varepsilon$ whenever $n\geq N(\varepsilon)$ and $n\notin J$, and such that $c_J(n)/n<\varepsilon$ for all $n\geq N(\varepsilon)$. Then, for all $n\geq N(\varepsilon)$, we have that

$$\frac{1}{n}\sum_{i=0}^{n-1}|a_i|=\frac{1}{n}\Bigg[\sum_{i\in J\cap\{0,1,\dots,n-1\}}|a_i|\quad+\quad\sum_{i\in\{0,1,\dots,n-1\}\setminus J}|a_i|\Bigg]<\frac{1}{n}\big[c_J(n)B+n\varepsilon\big]<(B+1)\varepsilon.$$

Thus

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}|a_i|=0.$$

[(b)⇔(c)] Using the fact that (b)⇔(a), it suffices to note that $\lim_{J\not\ni n\to\infty}a_i=0$ if and only if $\lim_{J\not\ni n\to\infty}a_i^2=0$. □

This theorem allows us to reformulate weak mixing in the following alternative ways.

Corollary 8.3.7. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. The following statements are equivalent:*

(a) *T is weakly mixing.*

(b) *For every $A, B \in \mathcal{B}$, there is a set $J(A,B) \subseteq \mathbb{Z}_+$ of density zero such that*

$$\lim_{J(A,B)\not\ni n\to\infty} \mu\big(T^{-n}(A) \cap B\big) = \mu(A)\mu(B).$$

(c) *For every $A, B \in \mathcal{B}$,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \big[\mu\big(T^{-j}(A) \cap B\big) - \mu(A)\mu(B)\big]^2 = 0.$$

Proof. Apply Theorem 8.3.6 with $a_n = \mu(T^{-n}(A) \cap B) - \mu(A)\mu(B)$. □

Corollary 8.3.7 offers an intuitive view of weakly mixing systems: a system is weakly mixing if for every measurable set A, the events $T^{-n}(A)$, $n \in \mathbb{N}$, become asymptotically independent of any other measurable set B, as long as we overlook a few instances of time. The avoided times naturally depend on both A and B as well as T and μ.

Let us finish with the relation between a weakly mixing system and the product of that system with itself.

Theorem 8.3.8. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. The following statements are equivalent:*

(a) *T is weakly mixing.*

(b) *$T \times T$ is ergodic.*

(c) *$T \times T$ is weakly mixing.*

Proof. Let us first show that (a)⇒(c). To that end, let $A, B, C, D \in \mathcal{B}$ and, using Corollary 8.3.7, let J_1 and J_2 be sets of density zero such that

$$\lim_{J_1\not\ni n\to\infty} \mu\big(T^{-n}(A \cap B)\big) = \mu(A)\mu(B) \quad \text{and} \quad \lim_{J_2\not\ni n\to\infty} \mu\big(T^{-n}(C \cap D)\big) = \mu(C)\mu(D).$$

Then

$$\begin{aligned}
&\lim_{J_1\cup J_2\not\ni n\to\infty} (\mu\times\mu)\big((T\times T)^{-n}(A\times C) \cap (B\times D)\big) \\
&\quad = \lim_{J_1\cup J_2\not\ni n\to\infty} (\mu\times\mu)\big(\big(T^{-n}(A)\times T^{-n}(C)\big) \cap (B\times D)\big) \\
&\quad = \lim_{J_1\cup J_2\not\ni n\to\infty} (\mu\times\mu)\big(\big(T^{-n}(A)\cap B\big) \times \big(T^{-n}(C)\cap D\big)\big) \\
&\quad = \lim_{J_1\cup J_2\not\ni n\to\infty} \mu\big(T^{-n}(A)\cap B\big)\cdot \mu\big(T^{-n}(C)\cap D\big) \\
&\quad = \mu(A)\mu(B)\cdot\mu(C)\mu(D) = \mu(A)\mu(C)\cdot\mu(B)\mu(D) \\
&\quad = (\mu\times\mu)(A\times C)\cdot(\mu\times\mu)(B\times D).
\end{aligned}$$

Thanks to Theorem 8.3.6, we deduce that

$$\lim_{n\to\infty} \sum_{j=0}^{n-1} \big|(\mu\times\mu)\big((T\times T)^{-j}(A\times C)\cap(B\times D)\big) \\ -(\mu\times\mu)(A\times C)\cdot(\mu\times\mu)(B\times D)\big| = 0.$$

Since the collection of measurable rectangles $\{E\times F : E,F\in\mathcal{B}\}$ forms a semialgebra that generates $\mathcal{B}\times\mathcal{B}$, Lemma 8.3.2 allows us to conclude that $T\times T$ is weakly mixing.

That (c)⇒(b) is an immediate consequence of Lemma 8.3.3.

It only remains to show that (b)⇒(a). To that end, let $A,B\in\mathcal{B}$. We aim to show that $\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}[\mu(T^{-j}(A)\cap B)-\mu(A)\mu(B)]^2 = 0$. Applying Lemma 8.2.16 to $T\times T$ and the rectangles $A\times X$ and $B\times X$, we get

$$\begin{aligned}\frac{1}{n}\sum_{j=0}^{n-1}\mu\big(T^{-j}(A)\cap B\big) &= \frac{1}{n}\sum_{j=0}^{n-1}(\mu\times\mu)\big((T\times T)^{-j}(A\times X)\cap(B\times X)\big)\\ &\underset{n\to\infty}{\longrightarrow} (\mu\times\mu)(A\times X)\cdot(\mu\times\mu)(B\times X) = \mu(A)\mu(B).\end{aligned}$$

Applying the same lemma to the rectangles $A\times A$ and $B\times B$, we obtain

$$\begin{aligned}\frac{1}{n}\sum_{j=0}^{n-1}\big[\mu\big(T^{-j}(A)\cap B\big)\big]^2 &= \frac{1}{n}\sum_{j=0}^{n-1}(\mu\times\mu)\big((T\times T)^{-j}(A\times A)\cap(B\times B)\big)\\ &\underset{n\to\infty}{\longrightarrow} (\mu\times\mu)(A\times A)\cdot(\mu\times\mu)(B\times B) = \mu(A)^2\mu(B)^2.\end{aligned}$$

Thus

$$\begin{aligned}&\frac{1}{n}\sum_{j=0}^{n-1}\big[\mu\big(T^{-j}(A)\cap B\big)-\mu(A)\mu(B)\big]^2\\ &\quad= \frac{1}{n}\sum_{j=0}^{n-1}\Big(\big[\mu\big(T^{-j}(A)\cap B\big)\big]^2 - 2\mu\big(T^{-j}(A)\cap B\big)\mu(A)\mu(B)+\mu(A)^2\mu(B)^2\Big)\\ &\quad\underset{n\to\infty}{\longrightarrow} \mu(A)^2\mu(B)^2 - 2\mu(A)^2\mu(B)^2+\mu(A)^2\mu(B)^2 = 0.\end{aligned}$$

Therefore, T is weakly mixing according to Corollary 8.3.7. □

8.3.2 Mixing

We now investigate a stronger mixing form.

Definition 8.3.9. Let $T : X\to X$ be a measure-preserving dynamical system on a probability space $(X,\mathcal{B},\mu)$. The system T is said to be *mixing* if

$$\lim_{n\to\infty}\mu\big(T^{-n}(A)\cap B\big) = \mu(A)\mu(B), \quad \forall A,B\in\mathcal{B}. \tag{8.26}$$

Like for ergodicity and weak mixing, to ascertain whether a system is mixing it suffices to check that it is mixing on a semialgebra that generates the σ-algebra.

Lemma 8.3.10. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. If $\mathcal{B} = \sigma(\mathcal{S})$ for some semialgebra $\mathcal{S}$, then T is mixing if and only if relation* (8.26) *holds for all $A, B \in \mathcal{S}$.*

Proof. The proof, which goes along similar lines to that of Lemma 8.2.17, is left as an exercise. □

Lemma 8.3.11. *If $T : X \to X$ is a mixing transformation on a probability space $(X, \mathcal{B}, \mu)$, then T is weakly mixing (and therefore ergodic).*

Proof. This is immediate from the definitions of weak mixing and mixing. □

Just as was the case for weakly mixing systems, we have the following characterization of mixing systems.

Lemma 8.3.12. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. The following statements are equivalent:*
(a) *T is mixing.*
(b) *For all $f, g \in L^2(\mu)$,*

$$\lim_{n\to\infty} \int_X (f \circ T^n) \cdot g \, d\mu = \int_X f \, d\mu \int_X g \, d\mu.$$

Proof. The proof is almost identical to that of Lemma 8.3.5 and is left to the reader. □

Example 8.3.13. In Example 8.2.32, we proved that the shift map σ is ergodic with respect to the product measure μ_P. In particular, we showed that for every pair A, B of cylinder sets, $\mu_P(\sigma^{-n}(A) \cap B) = \mu_P(A)\mu_P(B)$ as long as n is large enough. It is thus clear that σ is mixing, according to Lemma 8.3.10.

Remark 8.3.14. There are several examples of dynamical systems which are weakly mixing but not mixing. For instance, Katok [32] showed that all *interval exchange transformations* are not mixing, whereas Avila and Forni [6] later proved that almost all of these transformations are weakly mixing. Interval exchange transformations are a very nice class of examples which were first introduced by Ja. G. Sinai in a series of lectures in Russian at Erivan State University in 1973 and were introduced in a published paper in English by Keane in [34]. These maps are simple to define, exhibit interesting ergodic properties, and turn up in many seemingly surprising areas of mathematics. The basic idea is the following: Partition a bounded interval $I \subseteq \mathbb{R}$ into finitely many subintervals and define a bijective map from I to I that is a translation on each subinterval. The idea is best grasped with the aid of an illustration. See Figure 8.1.

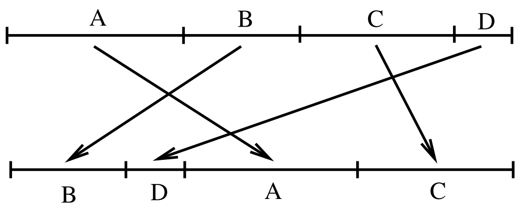

Figure 8.1: An example of an interval exchange transformation.

These maps are discontinuous at finitely many points. We will not delve into their dynamical properties. In addition to the papers mentioned above, the interested reader might consult [24, 45, 53, 74] and the references therein.

8.3.3 K-mixing

Before defining K-mixing, the reader who needs a quick refresher about the conditional expectation function with respect to the σ-algebra generated by a countable measurable partition is invited to consult Example A.1.62. As partitions are covers, the concepts, operations, and properties outlined in Section 7.1 will all be relevant here. It is worth noticing that all the operations introduced in that section result in countable measurable partitions. For instance, the join of two countable measurable partitions is a countable measurable partition; likewise, the preimage of a countable measurable partition is a countable measurable partition. Furthermore, note that for partitions, the relation $\prec$ is antisymmetric, that is, $\alpha \prec \beta \prec \alpha \iff \alpha = \beta$ (cf. Remark 7.1.6).

Definition 8.3.15. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and let α be a countable measurable partition of X. For every $n \in \mathbb{N}$, define

$$\alpha_n^\infty := \bigcup_{m>n} \alpha_n^m.$$

Given a probability measure μ on $(X, \mathcal{A})$, denote by $\sigma_c(\alpha_n^\infty)$ the completed σ-algebra generated by α_n^∞. (For more information about the completion of a σ-algebra, see Exercises 8.5.7–8.5.8.) The tail σ-algebra of α with respect to T is defined as

$$\mathrm{Tail}_T(\alpha) := \bigcap_{n=0}^{\infty} \sigma_c(\alpha_n^\infty).$$

Definition 8.3.16. A measurable transformation $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ is said to be K-*mixing* if

$$\lim_{n\to\infty} \sup_{A \in \sigma_c(\alpha_n^\infty)} \big|\mu(A \cap B) - \mu(A)\mu(B)\big| = 0$$

for every set $B \in \mathcal{A}$ and every finite measurable partition α of X.

The letter K is in honor of Kolmogorov, who first introduced this concept. K-mixing systems are simply referred to as K-*systems*. K-mixing is the strongest of all mixing properties discussed in this chapter.

Theorem 8.3.17. *Each* K-*system is mixing, and hence weakly mixing and ergodic.*

Proof. Suppose that $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ is a K-mixing transformation. Fix $A, B \in \mathcal{A}$ and consider the measurable partition $\alpha = \{A, X\backslash A\}$. Then $T^{-n}(A) \in \alpha_n^\infty$ for every $n \in \mathbb{N}$ and, therefore, by the K-mixing property,

$$\lim_{n\to\infty} \left|\mu\big(T^{-n}(A) \cap B\big) - \mu(A)\mu(B)\right| \le \lim_{n\to\infty} \sup_{F\in\alpha_n^\infty} \left|\mu(F \cap B) - \mu(F)\mu(B)\right| = 0.$$

Hence, T is mixing. □

We now give a characterization of K-systems which, as well as being interesting in its own right, will be used later to show that Rokhlin's natural extension of every metrically exact system is K-mixing.

Theorem 8.3.18. *Let $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ be a measurable transformation and $\mathcal{N}$ the σ-algebra of all sets of null or full μ-measure. Then T is* K-*mixing if and only if* $\mathrm{Tail}_T(\alpha) \subseteq \mathcal{N}$ *for every finite measurable partition α of X.*

Proof. Let $B \in \mathcal{A}$ and α a finite measurable partition of X. Suppose that $\mathrm{Tail}_T(\alpha) \subseteq \mathcal{N}$. Combining Example A.1.61 with Proposition A.1.60(f,e) reveals that

$$E\big(\mathbb{1}_B|\mathrm{Tail}_T(\alpha)\big) = E\big(E(\mathbb{1}_B|\mathcal{N}) \mid \mathrm{Tail}_T(\alpha)\big) = E\big(\mu(B) \mid \mathrm{Tail}_T(\alpha)\big) = \mu(B).$$

Fix $n \ge 0$. For every $A \in \sigma_c(\alpha_n^\infty)$, we have

$$\begin{aligned}\left|\mu(A \cap B) - \mu(A)\mu(B)\right| &= \left|\int_A \mathbb{1}_B\,d\mu - \int_A \mu(B)\,d\mu\right| \\ &= \left|\int_A \left[E\big(\mathbb{1}_B|\sigma_c(\alpha_n^\infty)\big) - \mu(B)\right] d\mu\right| \\ &\le \int_X \left|E\big(\mathbb{1}_B|\sigma_c(\alpha_n^\infty)\big) - \mu(B)\right| d\mu.\end{aligned}$$

So

$$\sup_{A\in\sigma_c(\alpha_n^\infty)} \left|\mu(A \cap B) - \mu(A)\mu(B)\right| \le \int_X \left|E\big(\mathbb{1}_B|\sigma_c(\alpha_n^\infty)\big) - \mu(B)\right| d\mu. \tag{8.27}$$

As $(\sigma_c(\alpha_n^\infty))_{n=0}^\infty$ is a descending sequence of σ-algebras whose intersection is $\mathrm{Tail}_T(\alpha)$, Theorem A.1.68 affirms that the sequence $\big(E(\mathbb{1}_B|\sigma_c(\alpha_n^\infty))\big)_{n=0}^\infty$ converges in $L^1(\mu)$ and pointwise μ-a.e. to $E(\mathbb{1}_B|\mathrm{Tail}_T(\alpha))$, which equals $\mu(B)$. It follows from (8.27) and Lebesgue's dominated convergence theorem (Theorem A.1.38) that

$$\begin{aligned}\lim_{n\to\infty}\sup_{A\in\sigma_c(\alpha_n^\infty)}\big|\mu(A\cap B)-\mu(A)\mu(B)\big| &\le \lim_{n\to\infty}\int_X\big|E(\mathbb{1}_B|\sigma_c(\alpha_n^\infty))-\mu(B)\big|\,d\mu\\ &=\int_X\Big|\lim_{n\to\infty}E(\mathbb{1}_B|\sigma_c(\alpha_n^\infty))-\mu(B)\Big|\,d\mu\\ &=0.\end{aligned}$$

Since B and α are arbitrary, T is K-mixing and one implication is proved.

In order to prove the converse, fix $C\in\mathrm{Tail}_T(\alpha)$. Then $C\in\sigma_c(\alpha_n^\infty)$ for every $n\ge 0$ and employing the definition of K-mixing with $B=C$, we obtain that

$$\big|\mu(C)-\mu(C)^2\big|=\big|\mu(C\cap C)-\mu(C)\mu(C)\big|\le\sup_{A\in\sigma_c(\alpha_n^\infty)}\big|\mu(A\cap C)-\mu(A)\mu(C)\big|\underset{n\to\infty}{\longrightarrow}0.$$

Hence, $\mu(C)=\mu(C)^2$ and it follows that either $\mu(C)=0$ or $\mu(C)=1$. So $\mathrm{Tail}_T(\alpha)\subseteq\mathcal{N}$. □

We will use this theorem in a moment. First, we recall the notion of Lebesgue space.

Definition 8.3.19. A Borel probability space $(X,\mathcal{B}(X),\mu)$ is said to be a *Lebesgue space* if X is a *Polish space* (i.e., X is completely metrizable and separable) and if μ is a complete measure.

Proposition 8.3.20. *Let $T:X\to X$ be a measure-preserving automorphism on a Lebesgue space $(X,\mathcal{B}(X),\mu)$. Assume that there exists a σ-algebra $\mathcal{B}\subseteq\mathcal{B}(X)$ with the following properties:*
(a) $T^{-1}(\mathcal{B})\subseteq\mathcal{B}$.
(b) $\sigma_c\big(\bigcup_{n=0}^\infty T^n(\mathcal{B})\big)=\mathcal{B}(X)$.
(c) $\bigcap_{n=0}^\infty T^{-n}(\mathcal{B})\subseteq\mathcal{N}$, *where $\mathcal{N}$ is the σ-algebra of all sets of null or full μ-measure.*

Then T is K*-mixing.*

Before proving this proposition, we need the following fairly nontrivial lemma, which can be found as Lemma 1 in Section 8 of Cornfeld, Fomin, and Sinai [16].

Lemma 8.3.21. *Let $T:X\to X$ be a measure-preserving automorphism of a Lebesgue space $(X,\mathcal{B}(X),\mu)$. If $\mathcal{A}$ and $\mathcal{B}$ are sub-σ-algebras of $\mathcal{B}(X)$ such that $\mathcal{A}\subseteq\sigma_c\big(\bigcup_{n=-\infty}^\infty T^n(\mathcal{B})\big)$, then* $\mathrm{Tail}_T(\mathcal{A})\subseteq\mathrm{Tail}_T(\mathcal{B})$, *where*

$$\mathrm{Tail}_T(\mathcal{C}):=\bigcap_{n=0}^\infty\sigma_c\left(\bigcup_{k=n}^\infty T^{-k}(\mathcal{C})\right)$$

for any sub-σ-algebra $\mathcal{C}$ of $\mathcal{B}(X)$.

Proof of Proposition 8.3.20. We will use Theorem 8.3.18 to establish the K-mixing property of T. Let α be a finite Borel partition of X. Then α generates a finite sub-σ-algebra $\mathcal{A}$ of $\mathcal{B}(X)$. By hypotheses (b) and (a), we have $\mathcal{A} \subseteq \sigma_c(\bigcup_{n=0}^{\infty} T^n(\mathcal{B})) = \sigma_c(\bigcup_{n=-\infty}^{\infty} T^n(\mathcal{B}))$. Using Lemma 8.3.21 and hypotheses (a) and (c), we obtain that

$$\mathrm{Tail}_T(\mathcal{A}) \subseteq \mathrm{Tail}_T(\mathcal{B}) = \bigcap_{n=0}^{\infty} \sigma_c\left(\bigcup_{k=n}^{\infty} T^{-k}(\mathcal{B})\right) = \bigcap_{n=0}^{\infty} \sigma_c(T^{-n}(\mathcal{B})) \subseteq \mathcal{N}.$$

The result follows from Theorem 8.3.18. □

We will see later that all two-sided Bernoulli shifts are K-mixing (cf. Example 8.1.14 and Exercise 8.5.22). This will be shown in Subsection 13.9.5, where we will treat the much more general case of Gibbs measures for Hölder continuous potentials. The first step in this direction is provided in the following section. In the meantime, see Exercise 8.5.23 for a direct proof based on Proposition 8.3.20.

8.4 Rokhlin's natural extension

Let $T : X \to X$ be a surjective measure-preserving dynamical system on a Lebesgue space $(X, \mathcal{F}, \mu)$ (see Definition 8.3.19). Consider the set of sequences

$$\widetilde{X} := \{(x_n)_{n=0}^{\infty} \in X^{\infty} : T(x_{n+1}) = x_n, \ \forall n \geq 0\} \subseteq X^{\infty}.$$

For every $k \geq 0$, let $\pi_k : \widetilde{X} \to X$ denote the projection onto the kth coordinate of $\widetilde{X}$, that is,

$$\pi_k((x_n)_{n=0}^{\infty}) := x_k.$$

Observe that $T \circ \pi_{k+1} = \pi_k$ for all $k \geq 0$. Equip the set $\widetilde{X}$ with the smallest σ-algebra $\widetilde{\mathcal{F}}$ that makes every projection $\pi_k : \widetilde{X} \to X$ continuous.

Note that in this construction the surjectivity of T is not really an essential assumption. Indeed, since X is a Lebesgue space, the sets $T^n(X)$, $n \geq 0$, are measurable. Because μ is T-invariant, it turns out that $\mu(T^n(X)) = 1$ for all $n \geq 0$. As the sets $T^n(X)$, $n \geq 0$, form a descending sequence, it follows that $\mu(\bigcap_{n=0}^{\infty} T^n(X)) = 1$. Finally, the map $T : \bigcap_{n=0}^{\infty} T^n(X) \to \bigcap_{n=0}^{\infty} T^n(X)$ is clearly surjective.

Definition 8.4.1. *Rokhlin's natural extension of* T is the measurable transformation $\widetilde{T} : \widetilde{X} \to \widetilde{X}$ defined by

$$\widetilde{T}((x_n)_{n=0}^{\infty}) := (T(x_0), x_0, x_1, x_2, \ldots).$$

Theorem 8.4.2. *Rokhlin's natural extension has the following properties:*

(a) *The transformation* $\widetilde{T} : \widetilde{X} \to \widetilde{X}$ *is invertible and its inverse* $\widetilde{T}^{-1} : \widetilde{X} \to \widetilde{X}$ *is (the restriction of) the left shift map*

$$\widetilde{T}^{-1}\big((x_n)_{n=0}^{\infty}\big) := (x_{n+1})_{n=0}^{\infty}.$$

(b) *For each* $n \geq 0$, *the following diagram commutes:*

$$\begin{array}{ccc} \widetilde{X} & \xrightarrow{\widetilde{T}} & \widetilde{X} \\ {\scriptstyle \pi_n}\downarrow & & \downarrow{\scriptstyle \pi_n} \\ X & \xrightarrow[T]{} & X \end{array}$$

(c) *There exists a unique probability measure* $\widetilde{\mu}$ *on the space* $(\widetilde{X}, \widetilde{\mathcal{F}})$ *such that*

$$\widetilde{\mu} \circ \pi_n^{-1} = \mu, \quad \forall n \geq 0.$$

(d) *The probability measure* $\widetilde{\mu}$ *is* $\widetilde{T}$*-invariant.*

Proof. Proof of properties (a) and (b) is left as an exercise. Property (c) follows directly from the Daniel–Kolmogorov consistency theorem (see Theorem 3.6.4 in Parthasarathy [55]). Regarding property (d), let $A \in \mathcal{F}$. For every $n \geq 0$, it follows from (b) and (c) that

$$\begin{aligned} \widetilde{\mu} \circ \widetilde{T}^{-1}\big(\pi_n^{-1}(A)\big) &= \widetilde{\mu} \circ (\pi_n \circ \widetilde{T})^{-1}(A) = \widetilde{\mu} \circ (T \circ \pi_n)^{-1}(A) \\ &= \widetilde{\mu} \circ \pi_n^{-1} \circ T^{-1}(A) = \mu \circ T^{-1}(A) \\ &= \mu(A) = \widetilde{\mu}\big(\pi_n^{-1}(A)\big). \end{aligned}$$

The family $\{\pi_n^{-1}(A) : A \in \mathcal{F}, n \geq 0\}$ forms a π-system that generates $\widetilde{\mathcal{F}}$. It ensues from Lemma A.1.26 that $\widetilde{\mu} \circ \widetilde{T}^{-1} = \widetilde{\mu}$. □

This theorem sometimes allows us to replace the μ-measure-preserving dynamical system T, which is not necessarily invertible, with the $\widetilde{\mu}$-measure-preserving automorphism $\widetilde{T} : \widetilde{X} \to \widetilde{X}$. This turns out to be of great advantage in some proofs, since dealing with invertible transformations is frequently easier than dealing with noninvertible ones. Natural extensions share many properties with their original maps. An example of this is given in the following theorem.

Theorem 8.4.3. *The natural extension measure* $\widetilde{\mu}$ *on* $\widetilde{X}$ *from Theorem* 8.4.2 *is ergodic with respect to* $\widetilde{T}$ *if and only if the measure* μ *is ergodic with respect to* T.

Proof. Suppose first that μ is not ergodic with respect to $T : X \to X$. Then there exists a set $A \in \mathcal{F}$ such that $T^{-1}(A) = A$ and $0 < \mu(A) < 1$. It follows from Theorem 8.4.2(c) that $\widetilde{\mu}(\pi_0^{-1}(A)) = \mu(A) \in (0, 1)$. Furthermore, it ensues from Theorem 8.4.2(b) that

$$\widetilde{T}^{-1}\big(\pi_0^{-1}(A)\big) = (\pi_0 \circ \widetilde{T})^{-1}(A) = (T \circ \pi_0)^{-1}(A) = \pi_0^{-1}\big(T^{-1}(A)\big) = \pi_0^{-1}(A).$$

Therefore, $\widetilde{\mu}$ is not ergodic with respect to $\widetilde{T}$.

Now assume that $T : X \to X$ is ergodic with respect to μ. We want to show that $\widetilde{T} : \widetilde{X} \to \widetilde{X}$ is then ergodic with respect to $\widetilde{\mu}$. Let $F \in L^1(\widetilde{X}, \widetilde{\mathcal{F}}, \widetilde{\mu})$ be a $\widetilde{T}$-invariant function. According to Theorem 8.2.18, it suffices to demonstrate that F is $\widetilde{\mu}$-a. e. constant.

As T is ergodic, the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14) yields that (to lighten notation, we use $\mu(g) := \int_X g\, d\mu$)

$$\lim_{k\to\infty} \left\| \frac{1}{k} \sum_{j=0}^{k-1} g \circ T^j - \mu(g) \right\|_{L^1(\mu)} = 0$$

for every function $g \in L^1(X, \mathcal{F}, \mu)$. Invoking Theorem 8.4.2, this implies that

$$\lim_{k\to\infty} \left\| \frac{1}{k} \sum_{j=0}^{k-1} G \circ \widetilde{T}^j - \widetilde{\mu}(G) \right\|_{L^1(\widetilde{\mu})} = 0 \tag{8.28}$$

for every $G \in L^1(\widetilde{X}, \widetilde{\mathcal{F}}, \widetilde{\mu})$ of the form $g \circ \pi_n$, where $g \in L^1(X, \mathcal{F}, \mu)$ and $n \geq 0$.

Fix $n \geq 0$ momentarily. The function $E(F|\widetilde{\mathcal{F}}_n)$ depends only on the nth coordinate of a point in $\widetilde{X}$ and can thus be expressed as $f_n \circ \pi_n$ for some $f_n \in L^1(X, \mathcal{F}, \mu)$. Setting $g = f_n$ and $G = f_n \circ \pi_n = E(F|\widetilde{\mathcal{F}}_n)$, it follows that $\widetilde{\mu}(G) = \widetilde{\mu}(F)$ and from (8.28) that

$$\lim_{k\to\infty} \left\| \frac{1}{k} \sum_{j=0}^{k-1} E(F|\widetilde{\mathcal{F}}_n) \circ \widetilde{T}^j - \widetilde{\mu}(F) \right\|_{L^1(\widetilde{\mu})} = 0. \tag{8.29}$$

Now, for every $n \geq 0$, let $\widetilde{\mathcal{F}}_n := \pi_n^{-1}(\mathcal{F})$. Since $T \circ \pi_{n+1} = \pi_n$, it turns out that

$$\widetilde{\mathcal{F}}_{n+1} = \pi_{n+1}^{-1}(\mathcal{F}) \supseteq \pi_{n+1}^{-1}\big(T^{-1}(\mathcal{F})\big) = (T \circ \pi_{n+1})^{-1}(\mathcal{F}) = \pi_n^{-1}(\mathcal{F}) = \widetilde{\mathcal{F}}_n.$$

Thus $(\widetilde{\mathcal{F}}_n)_{n=0}^{\infty}$ is an ascending sequence of sub-σ-algebras of $\widetilde{\mathcal{F}}$. By definition, $\widetilde{\mathcal{F}}$ is the σ-algebra generated by that sequence and we know that $F = E(F|\widetilde{\mathcal{F}})$. The martingale convergence theorem (Theorem A.1.67) then affirms that

$$\lim_{n\to\infty} \|F - E(F|\widetilde{\mathcal{F}}_n)\|_{L^1(\widetilde{\mu})} = 0. \tag{8.30}$$

But for every $j \geq 0$ and every $n \in \mathbb{N}$, it ensues from the $\widetilde{T}$-invariance of $\widetilde{\mu}$ that

$$\|F \circ \widetilde{T}^j - E(F|\widetilde{\mathcal{F}}_n) \circ \widetilde{T}^j\|_{L^1(\widetilde{\mu})} = \|F - E(F|\widetilde{\mathcal{F}}_n)\|_{L^1(\widetilde{\mu})}. \tag{8.31}$$

Fix $\varepsilon > 0$. By virtue of (8.30)–(8.31), there exists $N \in \mathbb{N}$ such that

$$\|F \circ \widetilde{T}^j - E(F|\widetilde{\mathcal{F}}_n) \circ \widetilde{T}^j\|_{L^1(\widetilde{\mu})} \leq \varepsilon$$

for every $j \geq 0$ and every $n \geq N$. Therefore, by the triangle inequality,

$$\left\| \frac{1}{k} \sum_{j=0}^{k-1} F \circ \widetilde{T}^j - \frac{1}{k} \sum_{j=0}^{k-1} E(F|\widetilde{\mathcal{F}}_n) \circ \widetilde{T}^j \right\|_{L^1(\widetilde{\mu})} \leq \varepsilon$$

for every $k \in \mathbb{N}$ and every $n \geq N$. Given that F is $\widetilde{T}$-invariant, this reduces to

$$\left\| F - \frac{1}{k} \sum_{j=0}^{k-1} E(F|\widetilde{\mathcal{F}}_n) \circ \widetilde{T}^j \right\|_{L^1(\widetilde{\mu})} \leq \varepsilon$$

for every $k \in \mathbb{N}$ and every $n \geq N$.

Fixing $n \geq N$ and letting $k \to \infty$, we deduce from this and (8.29) that $\|F - \widetilde{\mu}(F)\|_{L^1(\widetilde{\mu})} \leq \varepsilon$. Therefore $F = \widetilde{\mu}(F)$ in $L^1(\widetilde{\mu})$. Consequently, $F = \widetilde{\mu}(F)$ $\widetilde{\mu}$-almost everywhere. That is, F is $\widetilde{\mu}$-a. e. constant. □

We now introduce the concept of metric exactness.

Definition 8.4.4. A measure-preserving dynamical system $T : X \to X$ on a Lebesgue space $(X, \mathcal{F}, \mu)$ is said to be *metrically exact* if for each $A \in \mathcal{F}$ such that $\mu(A) > 0$ we have

$$\lim_{n\to\infty} \mu\big(T^n(A)\big) = 1.$$

Note that each set $T^n(A)$ is measurable since T is a measurable transformation of a Lebesgue space.

Metric exactness of a system can be characterized in terms of the tail σ-algebra of the system.

Proposition 8.4.5. *Let $T : X \to X$ be a measure-preserving dynamical system on a Lebesgue space $(X, \mathcal{F}, \mu)$. Then T is metrically exact if and only if the tail σ-algebra* $\mathrm{Tail}_T(\mathcal{F}) = \bigcap_{n=0}^{\infty} T^{-n}(\mathcal{F})$ *is contained in the σ-algebra $\mathcal{N}$ of all sets of null or full μ-measure.*

Proof. Suppose that T is metrically exact and let $F \in \mathrm{Tail}_T(\mathcal{F})$. By definition of the tail σ-algebra, there exists a sequence of sets $(F_n)_{n=0}^{\infty}$ in $\mathcal{F}$ such that $F = T^{-n}(F_n)$ for each $n \geq 0$. Suppose that $\mu(F) > 0$. Then

$$1 = \lim_{n\to\infty} \mu\big(T^n(F)\big) = \lim_{n\to\infty} \mu\big(T^n\big(T^{-n}(F_n)\big)\big) = \lim_{n\to\infty} \mu(F_n) = \lim_{n\to\infty} \mu \circ T^{-n}(F_n) = \mu(F).$$

Thus $\mathrm{Tail}_T(\mathcal{F})$ consists only of sets of measure zero and one. This proves one implication.

Now suppose that $\mathrm{Tail}_T(\mathcal{F}) \subseteq \mathcal{N}$. Fix $F \in \mathcal{F}$ with $\mu(F) > 0$. For every $k \geq 0$, consider the measurable sets

$$F_k := \bigcup_{n=k}^{\infty} T^{-n}\big(T^n(F)\big) \supseteq F \quad \text{and} \quad F_\infty := \bigcap_{k=0}^{\infty} F_k \supseteq F.$$

We claim that $F_\infty \in \mathrm{Tail}_T(\mathcal{F})$. Indeed, by definition, the sequence of sets $(F_k)_{k=0}^{\infty}$ is descending and, therefore,

$$\mu(F_\infty) = \lim_{k\to\infty} \mu(F_k) \quad \text{and} \quad F_\infty = \bigcap_{k=j}^{\infty} F_k, \quad \forall j \geq 0.$$

If $k \geq j \geq 0$, then

$$F_k = \bigcup_{n=k}^{\infty} T^{-n}(T^n(F)) = \bigcup_{i=0}^{\infty} T^{-k}(T^{-i}(T^{k+i}(F))) = T^{-j}\Bigg(T^{-(k-j)}\Bigg(\bigcup_{i=0}^{\infty} T^{-i}(T^{k+i}(F)) \Bigg)\Bigg).$$

To shorten notation, let $A_{j,k} := T^{-(k-j)}(\bigcup_{i=0}^{\infty} T^{-i}(T^{k+i}(F)))$. Then

$$F_\infty = \bigcap_{k=j}^{\infty} T^{-j}(A_{j,k}) = T^{-j}\Bigg(\bigcap_{k=j}^{\infty} A_{j,k} \Bigg).$$

Hence, for every $j \geq 0$, it follows that $F_\infty \in T^{-j}(\mathcal{F})$, or, equivalently, $F_\infty \in \mathrm{Tail}_T(\mathcal{F})$. Since $F_\infty \supseteq F$ and $\mu(F) > 0$, our hypothesis that $\mathrm{Tail}_T(\mathcal{F}) \subseteq \mathcal{N}$ implies $\mu(F_\infty) = 1$. So

$$\mu(F_0) = 1. \tag{8.32}$$

However, as $T^{-1}(T(A)) \supseteq A$ for every subset A of X, we observe that

$$T^{-(n+1)}(T^{n+1}(F)) = T^{-n}(T^{-1}(T(T^n(F)))) \supseteq T^{-n}(T^n(F)),$$

that is, the sequence of sets $(T^{-n}(T^n(F)))_{n=0}^{\infty}$ is ascending to their union F_0. Using this, (8.32) and the T-invariance of μ, we deduce that

$$1 = \mu(F_0) = \lim_{n\to\infty} \mu(T^{-n}(T^n(F))) = \lim_{n\to\infty} \mu(T^n(F)).$$

As $F \in \mathcal{F}$ was chosen arbitrarily, the transformation T is metrically exact. □

We can now demonstrate that Rokhlin's natural extension of any metrically exact system is K-mixing.

Theorem 8.4.6. *Let $T : X \to X$ be a measure-preserving dynamical system on a Lebesgue space $(X, \mathcal{F}, \mu)$. If T is metrically exact, then Rokhlin's natural extension $\widetilde{T} : \widetilde{X} \to \widetilde{X}$ is* K*-mixing.*

Proof. Let

$$\mathcal{B} := \pi_0^{-1}(\mathcal{F}) \subseteq \widetilde{\mathcal{F}}.$$

We shall verify that all the hypotheses of Proposition 8.3.20 hold. First,

$$\widetilde{T}^{-1}(\mathcal{B}) = \widetilde{T}^{-1}(\pi_0^{-1}(\mathcal{F})) = (\pi_0 \circ \widetilde{T})^{-1}(\mathcal{F}) = (T \circ \pi_0)^{-1}(\mathcal{F}) = \pi_0^{-1}(T^{-1}(\mathcal{F})) \subseteq \pi_0^{-1}(\mathcal{F}) = \mathcal{B}.$$

So condition (a) of Proposition 8.3.20 is satisfied. In order to show that condition (b) holds, note that $\widetilde{T}^n(\mathcal{B}) = \widetilde{T}^n \circ \pi_0^{-1}(\mathcal{F}) = \pi_n^{-1}(\mathcal{F})$ for all $n \geq 0$. Recall also that $\widetilde{\mathcal{F}}$ is the smallest σ-algebra containing all of the σ-algebras $\pi_n^{-1}(\mathcal{F})$. Thus condition (b) is satisfied. Finally, with $\mathcal{N}$ (resp., $\widetilde{\mathcal{N}}$) denoting the sub-σ-algebra consisting of the null and full μ-measure sets (resp., $\tilde{\mu}$-measure sets), we obtain that

$$\begin{aligned}\bigcap_{n=0}^{\infty}\widetilde{T}^{-n}(\mathcal{B}) &= \bigcap_{n=0}^{\infty}\widetilde{T}^{-n}(\pi_0^{-1}(\mathcal{F})) = \bigcap_{n=0}^{\infty}(\pi_0\circ\widetilde{T}^n)^{-1}(\mathcal{F})\\ &= \bigcap_{n=0}^{\infty}(T^n\circ\pi_0)^{-1}(\mathcal{F}) = \bigcap_{n=0}^{\infty}\pi_0^{-1}\circ T^{-n}(\mathcal{F})\\ &= \pi_0^{-1}\left(\bigcap_{n=0}^{\infty}T^{-n}(\mathcal{F})\right) = \pi_0^{-1}(\mathrm{Tail}_T(\mathcal{F})) \subseteq \pi_0^{-1}(\mathcal{N}) \subseteq \widetilde{\mathcal{N}},\end{aligned}$$

where the first set inclusion comes from Proposition 8.4.5 and the second from Theorem 8.4.2(c). So condition (c) holds. Apply Proposition 8.3.20 to conclude. □

Finally, we provide an explicit description of the Rokhlin's natural extensions of a very important class of noninvertible measure-preserving dynamical systems, namely the one-sided Bernoulli shifts with finite sets of states, which were introduced in Example 8.1.14.

Let E be a finite set. Define the map $h : \widetilde{E^{\mathbb{N}}} \to E^{\mathbb{Z}}$ by

$$(h(\omega))_n = \begin{cases} (\omega_0)_n & \text{if } n \geq 0 \\ (\omega_{-n})_0 & \text{if } n < 0 \end{cases}$$

A straightforward inspection shows that h is bijective and that the following diagram commutes:

$$\begin{array}{ccc} \widetilde{E^{\mathbb{N}}} & \xrightarrow{\tilde{\sigma}} & \widetilde{E^{\mathbb{N}}} \\ h\big\downarrow & & \big\downarrow h \\ E^{\mathbb{Z}} & \xrightarrow[\sigma]{} & E^{\mathbb{Z}} \end{array}$$

In addition, if $\widetilde{E^{\mathbb{N}}}$ and $E^{\mathbb{Z}}$ are endowed with their respective product (Tychonov) topologies, then the map h is a homeomorphism, and thus is a measurable isomorphism if $\widetilde{E^{\mathbb{N}}}$ and $E^{\mathbb{Z}}$ are equipped with the corresponding Borel σ-algebras.

Furthermore, if E has at least two elements and $P : E \to [0,1]$ is a probability vector, let μ_P^+ be the corresponding one-sided Bernoulli measure on $E^{\mathbb{N}}$ introduced in Example 8.1.14 and denoted there just by μ_P. Likewise, let μ_P be the two-sided Bernoulli measure on $E^{\mathbb{Z}}$ introduced in Exercise 8.5.22.

For every $k \leq 0$ and $n \geq 0$, the cylinder $[\omega_k\omega_{k+1}\ldots\omega_{-1}\omega_0\omega_1\ldots\omega_n] \subseteq E^{\mathbb{Z}}$ satisfies

$$h^{-1}([\omega_k\ldots\omega_n]) = [\omega_0\ldots\omega_n]\times[\omega_{-1}]\times[\omega_{-2}]\times\cdots\times[\omega_k]\times\prod_{j=-k+1}^{\infty}E.$$

Consequently,

$$\widetilde{\mu_P^+}(h^{-1}([\omega_k \dots \omega_n])) = \prod_{i=k}^{n} P_i = \mu_P([\omega_k \dots \omega_n]).$$

Since all such cylinders form a π-system generating the Borel σ-algebra on $E^{\mathbb{Z}}$, we conclude that

$$\widetilde{\mu_P^+} \circ h^{-1} = \mu_P.$$

We have therefore proved the following.

Theorem 8.4.7. *If E is a finite set having at least two elements and $P : E \to [0,1]$ is a probability vector, then the Rokhlin's natural extension of the one-sided Bernoulli shift $(\sigma : E^{\mathbb{N}} \to E^{\mathbb{N}}, \mu_P)$ is metrically (i. e. measure-theoretically) isomorphic to the two-sided Bernoulli shift $(\sigma : E^{\mathbb{Z}} \to E^{\mathbb{Z}}, \mu_P)$.*

8.5 Exercises

Note: Exercises 8.5.1–8.5.19 pertain to measure theory. They use the terminology and notation introduced in Appendix A. The reader who would rather just concentrate on ergodic theory may skip those exercises.

Exercise 8.5.1. Let X be a set and $\mathcal{C}$ be a finite collection of subsets of X.
(a) Show that the algebra $\mathcal{A}(\mathcal{C})$ generated by $\mathcal{C}$ is finite.

Hint: First, identify the elements of $\mathcal{A}(\mathcal{C})$ when $\mathcal{C}$ consists of two disjoint sets. Generalize your argument to the case in which $\mathcal{C}$ consists of a finite number of disjoint sets. Then reduce the general case of a finite collection $\mathcal{C}$ to the case of an equivalent finite collection of disjoint sets.

(b) Deduce that $\sigma(\mathcal{C}) = \mathcal{A}(\mathcal{C})$.

Exercise 8.5.2. Let X be a set and $Y \subseteq X$. What is $\mathcal{A}(\{Y\})$? And $\sigma(\{Y\})$?

Exercise 8.5.3. Let X be an infinite set. Show that

$$\{Y \subseteq X : \text{either } Y \text{ or } X \setminus Y \text{ is finite}\}$$

is an algebra but not a σ-algebra on X.

Exercise 8.5.4. Let X be an uncountable set. Show that

$$\{Y \subseteq X : \text{either } Y \text{ or } X \setminus Y \text{ is countable}\}$$

is a σ-algebra on X. It is often called the *countable-cocountable* σ-algebra.

Exercise 8.5.5. Let X be a set and $Y \subseteq X$. If $\mathcal{B}$ is a σ-algebra on X, prove that

$$\mathcal{B}|_Y = \{B \cap Y : B \in \mathcal{B}\}$$

is a σ-algebra on Y.

Exercise 8.5.6. Let $(X, \mathcal{A}, \mu)$ be a measure space. Show that

$$\{A \in \mathcal{A} : \mu(A) = 0 \text{ or } \mu(X \backslash A) = 0\}$$

is a sub-σ-algebra of $\mathcal{A}$.

Exercise 8.5.7. Let $(X, \mathcal{B}, \mu)$ be a measure space and

$$\mathcal{N} := \{N \in \mathcal{B} \mid \mu(N) = 0\}$$

be the collection of all sets of measure zero, sometimes called null sets. Define

$$\overline{\mathcal{N}} := \{\overline{N} \subseteq X \mid \exists\, N \in \mathcal{N} \text{ such that } \overline{N} \subseteq N\} = \bigcup_{N \in \mathcal{N}} \mathcal{P}(N).$$

The completion of $(X, \mathcal{B}, \mu)$ is the measure space $(X, \overline{\mathcal{B}}, \overline{\mu})$, where

$$\overline{\mathcal{B}} := \{B \cup \overline{N} \,|\, B \in \mathcal{B},\, \overline{N} \in \overline{\mathcal{N}}\}$$

and

$$\overline{\mu}(\overline{B}) = \mu(B) \text{ whenever } \overline{B} = B \cup \overline{N} \text{ for some } B \in \mathcal{B} \text{ and } \overline{N} \in \overline{\mathcal{N}}.$$

(a) Prove that the space $(X, \overline{\mathcal{B}}, \overline{\mu})$ is well-defined (namely, that $\overline{\mathcal{B}}$ is a σ-algebra on X and that $\overline{\mu}$ is well-defined).
(b) Show that $(X, \overline{\mathcal{B}}, \overline{\mu})$ is an extension of the space $(X, \mathcal{B}, \mu)$ (i. e., $\overline{\mathcal{B}} \supseteq \mathcal{B}$ and $\overline{\mu} = \mu$ on $\mathcal{B}$).
(c) Observe that the space $(X, \overline{\mathcal{B}}, \overline{\mu})$ is complete.

Exercise 8.5.8. Let $(X, \mathcal{B}, \mu)$ be a measure space and let

$$\mathcal{B}^* := \{E \subseteq X \mid \exists\, A, B \in \mathcal{B} \text{ such that } A \subseteq E \subseteq B \text{ and } \mu(B \backslash A) = 0\}$$

be the collection of all subsets of X that are squeezed by some measurable sets whose difference is of measure zero. Define

$$\mu^*(E) = \mu(A) \text{ whenever } \exists\, A, B \in \mathcal{B} \text{ such that } A \subseteq E \subseteq B \text{ and } \mu(B \backslash A) = 0.$$

Prove that $(X, \mathcal{B}^*, \mu^*)$ is the completion of $(X, \mathcal{B}, \mu)$ (cf. Exercise 8.5.7).

Exercise 8.5.9. Show that Lemma A.1.26 does not hold for infinite measures in general.

Hint: Consider the family of all Borel subsets of $\mathbb{R}$ which do not have 0 for element.

Exercise 8.5.10. In this exercise, we look at the set-theoretic properties of the symmetric difference operation. Let X be a set. Let $A, B, C \subseteq X$. Let $T : X \to X$ be a map. Prove the following statements:

(a) $A \triangle B = B \triangle A$.
(b) If $A \cap B = \emptyset$ then $A \triangle B = A \cup B$.
(c) $(X\backslash A) \triangle (X\backslash B) = A \triangle B$.
(d) $T^{-1}(A \triangle B) = T^{-1}(A) \triangle T^{-1}(B)$.
(e) $A \triangle C \subseteq (A \triangle B) \cup (B \triangle C)$.
This statement generalizes to any finite number of intermediaries, that is, $A_n \triangle A_0 \subseteq \bigcup_{k=0}^{n-1}(A_{k+1} \triangle A_k)$.
(f) $A \triangle (B \cup C) \subseteq (A \triangle B) \cup (A \triangle C)$.
More generally, $(\bigcup_{i\in I} A_i) \triangle (\bigcup_{i\in I} B_i) \subseteq \bigcup_{i\in I}(A_i \triangle B_i)$ for any index set I.
(g) $A \triangle (B \cap C) \subseteq (A \triangle B) \cup (A \triangle C)$.
More generally, $(\bigcap_{i\in I} A_i) \triangle (\bigcap_{i\in I} B_i) \subseteq \bigcup_{i\in I}(A_i \triangle B_i)$ for any index set I.

Exercise 8.5.11. Let $(X, \mathcal{A}, \mu)$ be a probability space and $A, B \in \mathcal{A}$. Prove the following statements:

(a) $|\mu(A) - \mu(B)| \le \mu(A \triangle B)$.
(b) If $\mu(A \triangle B) = 0$ then $\mu(A) = \mu(B)$.

Exercise 8.5.12. Let $(X, \mathcal{A}, \mu)$ be a probability space and $A, B \in \mathcal{A}$. Prove that $\mu(A \cap B) \ge \mu(A) + \mu(B) - 1$.

Exercise 8.5.13. Let $(X, \mathcal{A}, \mu)$ be a measure space. Show that if $(f_n)_{n=1}^{\infty}$ is a sequence of nonnegative measurable functions, then

$$\int_X \sum_{n=1}^{\infty} f_n \, d\mu = \sum_{n=1}^{\infty} \int_X f_n \, d\mu.$$

Exercise 8.5.14. The purpose of this exercise is to establish that a sequence of L^1 functions that converges pointwise need not converge in L^1. Construct a sequence of functions $f_n : [0, 1] \to [0, \infty)$, $n \in \mathbb{N}$, with the following properties:

(a) Each function is continuous.
(b) The sequence converges pointwise to the constant function 0.
(c) $\int_{[0,1]} f_n \, d\lambda = 1$ for all $n \in \mathbb{N}$, where λ is the Lebesgue measure on $[0, 1]$.

Deduce that the sequence does not converge in $L^1([0, 1], \mathcal{B}([0, 1]), \lambda)$.

Exercise 8.5.15. Find a sequence $(f_n)_{n=1}^{\infty}$ of $L^1(X, \mathcal{A}, \mu)$ functions with the following properties:

(a) The sequence converges pointwise to a function f.
(b) $\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu$.
(c) $\lim_{n\to\infty} \int_X |f_n| \, d\mu \neq \int_X |f| \, d\mu$.

Exercise 8.5.16. Show that on finite measure spaces Theorem A.1.41 is a generalization of Lebesgue's dominated convergence theorem (Theorem A.1.38).

Exercise 8.5.17. Construct a sequence of integrable functions that converges in L^1 to an integrable function but that does not converge pointwise almost everywhere.
Hint: Use indicator functions of carefully selected subintervals of $[0,1]$.

Exercise 8.5.18. Let $(X, \mathcal{A}, \mu)$ be a probability space, and let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Let $\overline{\mathcal{B}}$ be the completion of $\mathcal{B}$ in $\mathcal{A}$, that is,

$$\overline{\mathcal{B}} := \{A \in \mathcal{A} \mid \exists B \in \mathcal{B} \text{ such that } \mu(A \triangle B) = 0\}.$$

Show that $\overline{\mathcal{B}}$ is a sub-σ-algebra of $\mathcal{A}$ for which the following two properties hold:
(a) $\overline{\mathcal{B}} \supseteq \mathcal{B}$.
(b) $E(\varphi|\overline{\mathcal{B}}) = E(\varphi|\mathcal{B})$.

Exercise 8.5.19. Let $(X, \mathcal{A}, \mu)$ be a probability space, and let $\mathcal{B}$ and $\mathcal{C}$ be sub-σ-algebras of $\mathcal{A}$. Say that $\mathcal{B} \approx \mathcal{C}$ if $\mathcal{B} \subseteq \overline{\mathcal{C}}$ and $\mathcal{C} \subseteq \overline{\mathcal{B}}$ (cf. Exercise 8.5.18). Show that $E(\varphi|\mathcal{B}) = E(\varphi|\mathcal{C})$ if $\mathcal{B} \approx \mathcal{C}$.

Exercise 8.5.20. Let $\mathcal{B}(\mathbb{R})$ denote the Borel σ-algebra of $\mathbb{R}$. The collection

$$\mathcal{B} := \{B \in \mathcal{B}(\mathbb{R}) \mid B = -B\}$$

of all Borel sets that are symmetric with respect to the origin forms a sub-σ-algebra of $\mathcal{B}(\mathbb{R})$. Let λ denote the Lebesgue measure on $\mathbb{R}$ and let $\varphi \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. Prove that

$$E(\varphi|\mathcal{B})(x) = \frac{1}{2}[\varphi(x) + \varphi(-x)], \quad \forall x \in \mathbb{R}.$$

Hint: First show that $E(\varphi|\mathcal{B})$ must be an even function. Then use the fact that the transformation $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = -x$ is λ-invariant.

Exercise 8.5.21. Let $T : (X, \mathcal{A}) \to (Y, \mathcal{B})$ be a measurable transformation and μ be a measure on $(X, \mathcal{A})$. Show that the set function $\mu \circ T^{-1}$ is a measure on $(Y, \mathcal{B})$.

Exercise 8.5.22. This exercise pertains to a two-sided version of Example 8.1.14. Let $(E, \mathcal{F}, P)$ be a probability space. The product set $E^{\mathbb{Z}} := \prod_{k=-\infty}^{\infty} E$ is commonly equipped with the product σ-algebra $\mathcal{F}_{\mathbb{Z}}$ generated by the semialgebra of all (finite) cylinders (also called rectangles), namely, given $m, n \in \mathbb{Z}$, with $m \le n$, and $E_m, E_{m+1}, \ldots, E_n \in \mathcal{F}$, the set

$$\prod_{k=-\infty}^{m-1} E \times E_m \times E_{m+1} \times \cdots \times E_n \times \prod_{k=n+1}^{\infty} E = \{\tau \in E^{\mathbb{Z}} : \tau_k \in E_k, \forall m \le k \le n\}$$

is called a cylinder. The product measure μ_P on $\mathcal{F}_{\mathbb{Z}}$ is the unique probability measure which confers to a cylinder the value

$$\mu_P\left(\prod_{k=-\infty}^{m-1} E \times E_m \times E_{m+1} \times \cdots \times E_n \times \prod_{k=n+1}^{\infty} E\right) := \prod_{k=m}^{n} P(E_k). \tag{8.33}$$

The existence and uniqueness of this measure can be established using Theorem A.1.27, Lemma A.1.29 and Theorem A.1.28 successively. For more information, see Halmos [27] (pp. 157–158) or Taylor [72] (Chapter III, Section 4).

It is easy to show that the left shift map $\sigma : E^{\mathbb{Z}} \to E^{\mathbb{Z}}$, which is defined by $\sigma\big((\tau_n)_{n=-\infty}^{\infty}\big) := (\tau_{n+1})_{n=-\infty}^{\infty}$, preserves the product measure μ_P. The measure-preserving dynamical system $(\sigma : E^{\mathbb{Z}} \to E^{\mathbb{Z}}, \mu_P)$ is commonly referred to as a *two-sided Bernoulli shift* with set of states E. In this book, primarily focused on noninvertible dynamical systems, one-sided Bernoulli shifts will be of primer importance. However, in abstract ergodic theory, two-sided Bernoulli shifts seem to have played a more prominent role and they are encountered several times in this book, particularly in Sections 8.3.3, 8.4, and 13.9.5.

From this point on, we assume that the original probability space is of the form $(E, \mathcal{F}, P)$, where E is countable and $\mathcal{F}$ is the σ-algebra $\mathcal{P}(E)$ of all subsets of E.

Let $E^{\mathbb{Z}_-} = \prod_{k=-\infty}^{-1} E$ and $E^{\mathbb{Z}_+} = \prod_{k=0}^{\infty} E$, so $E^{\mathbb{Z}} = E^{\mathbb{Z}_-} \times E^{\mathbb{Z}_+}$. Accordingly, for any $\tau \in E^{\mathbb{Z}}$, write $\tau_- := \cdots \tau_{-2}\tau_{-1} \in E^{\mathbb{Z}_-}$ and $\tau_+ := \tau_0\tau_1\tau_2 \ldots \in E^{\mathbb{Z}_+}$.

(a) Show that the family $C = \{E^{\mathbb{Z}_-} \times \{\omega\}\}_{\omega \in E^{\mathbb{Z}_+}}$, that is, the family of sets consisting of double-sided sequences having a common positive part, constitutes an uncountable measurable partition of $E^{\mathbb{Z}}$.

Let $\mathcal{C} := \sigma(C)$ be the sub-σ-algebra of $\mathcal{F}_{\mathbb{Z}}$ generated by the partition C. We aim to calculate $E(\varphi|\mathcal{C})$ for any function $\varphi \in L^1(E^{\mathbb{Z}}, \mathcal{F}_{\mathbb{Z}}, \mu_P)$.

For each $n \geq 0$, consider the family of all n-cylinders

$$C_n := \left\{E^{\mathbb{Z}_-} \times \{e_0\} \times \cdots \times \{e_{n-1}\} \times \prod_{k=n}^{\infty} E : e_k \in E,\ \forall 0 \leq k \leq n-1\right\}.$$

(b) Prove that $(C_n)_{n=0}^{\infty}$ is an ascending sequence of countable measurable partitions of $E^{\mathbb{Z}}$.
(c) For every $n \geq 0$, let $\mathcal{C}_n := \sigma(C_n)$. Demonstrate that $(\mathcal{C}_n)_{n=0}^{\infty}$ is an ascending sequence of sub-σ-algebras of $\mathcal{F}_{\mathbb{Z}}$.
(d) Let $\mathcal{C}_\infty := \sigma(\bigcup_{n=0}^{\infty} C_n)$. Show that $\mathcal{C} = \mathcal{C}_\infty$.
(e) Deduce that $E(\varphi|\mathcal{C}) = \lim_{n\to\infty} E(\varphi|\mathcal{C}_n)$.

For any given $\tau \in E^{\mathbb{Z}}$ and $m, n \in \mathbb{Z}$ such that $m \leq n$, let

$$[\tau]_m^n = \prod_{k=-\infty}^{m-1} E \times \{\tau_m\} \times \{\tau_{m+1}\} \times \cdots \times \{\tau_n\} \times \prod_{k=n+1}^{\infty} E.$$

(f) Establish that $E(\varphi|\mathcal{C})(\tau) = \lim_{n\to\infty} E(\varphi|[\tau]_0^{n-1})$ for every $\tau \in E^{\mathbb{Z}}$.
(g) Prove that the sequence of cylinder sets $([\tau]_0^{n-1})_{n=1}^{\infty}$ is descending and that $\bigcap_{n=1}^{\infty}[\tau]_0^{n-1} = E^{\mathbb{Z}_-} \times \{\tau_+\}$.
(h) Deduce that $\lim_{n\to\infty} \mu_P([\tau]_0^{n-1}) = \mu_P(E^{\mathbb{Z}_-} \times \{\tau_+\})$.
(i) Deduce further that $\lim_{n\to\infty} \int_{[\tau]_0^{n-1}} \varphi\, d\mu_P = \int_{E^{\mathbb{Z}_-}\times\{\tau_+\}} \varphi\, d\mu_P$.
(j) Conclude that $E(\varphi|\mathcal{C})(\tau) = E(\varphi|E^{\mathbb{Z}_-} \times \{\tau_+\})$ for all $\tau \in E^{\mathbb{Z}}$.

That is, you have showed that $E(\varphi|\mathcal{C})$ is constant, and is equal to the mean value of φ, on each set of the form $E^{\mathbb{Z}_-} \times \{\omega\}$, where $\omega \in E^{\mathbb{Z}_+}$.

Exercise 8.5.23. Let E be a finite set having at least two elements, let $P : E \to [0,1]$ be a probability vector, and let $(\sigma : E^{\mathbb{Z}} \to E^{\mathbb{Z}}, \mu_P)$ be the corresponding two-sided Bernoulli shift described in Exercise 8.5.22. Provide a direct proof, based on Proposition 8.3.20, that the automorphism $(\sigma : E^{\mathbb{Z}} \to E^{\mathbb{Z}}, \mu_P)$ is K-mixing.

Hint: Consider the projection $\pi_+ : E^{\mathbb{Z}} \to E^{\mathbb{N}}$ defined by the formula $\pi_+(\omega) = \omega|_0^{\infty}$.

Exercise 8.5.24. Let $\mathcal{B}_+$ be the standard Borel σ-algebra on $E^{\infty} = E^{\mathbb{N}}$. Show that the σ-algebra $\mathcal{B} := \pi_+^{-1}(\mathcal{B}_+)$ on $E^{\mathbb{Z}}$ satisfies all the hypotheses of Proposition 8.3.20.

Exercise 8.5.25. The map $G : [0,1] \to [0,1]$ defined by

$$G(x) := \begin{cases} 0 & \text{if } x = 0 \\ \langle \frac{1}{x} \rangle & \text{if } x > 0, \end{cases}$$

where $\langle r \rangle$ denotes the fractional part of r, is called the *Gauss map*.
(a) Show that this map is not invariant under the Lebesgue measure λ on $[0,1]$.
(b) Prove that the Borel probability measure

$$\mu_G(B) := \frac{1}{\log 2} \int_B \frac{1}{1+x}\, dx$$

is G-invariant. The measure μ_G is known as the *Gauss measure*.

Exercise 8.5.26. Recall the Farey map $F : [0,1] \to [0,1]$ from Example 1.2.3. Show that the Borel probability measure

$$\mu_F(B) := \int_B \frac{1}{x}\, dx$$

is F-invariant, while the Lebesgue measure λ is not.

Exercise 8.5.27. Show that the Dirac point-mass δ_0 is the only T-invariant Borel probability measure for the doubling map on the entire real line, that is, for $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = 2x$.

Exercise 8.5.28. Find all T-invariant Borel probability measures for the squaring map on the entire real line, that is, for $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = x^2$.

Exercise 8.5.29. Prove that the continuous transformation $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = x^2 + 1$ does not admit any (finite or infinite) T-invariant Borel measure.

Exercise 8.5.30. Let $b \neq 0$. Prove that the translation $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = x+b$ admits a σ-finite T-invariant Borel measure but not a finite one.

Exercise 8.5.31. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. Suppose that $B \in \mathcal{A}$ is forward T-invariant, that is, $T(B) \subseteq B$. Let $\mathcal{A}|_B := \{A \cap B : A \in \mathcal{A}\}$ be the projection of $\mathcal{A}$ onto B. Suppose that μ is a probability measure on $(B, \mathcal{A}|_B)$. Define the set function $\widehat{\mu} : \mathcal{A} \to [0, 1]$ by the formula

$$\widehat{\mu}(A) := \mu(A \cap B).$$

Show that $\widehat{\mu}$ is a probability measure on $(X, \mathcal{A})$. Furthermore, prove that if μ is $T|_B$-invariant then $\widehat{\mu}$ is T-invariant.

Exercise 8.5.32. Find a nontrivial measure-preserving dynamical system $T : X \to X$ on a probability space $(X, \mathcal{B}, \mu)$ for which there exist at least three completely T-invariant measurable sets of positive μ-measure.

Exercise 8.5.33. Identify a nontrivial measure-preserving dynamical system $T : X \to X$ on a probability space $(X, \mathcal{B}, \mu)$ for which there exist uncountably many measurable sets of positive measure and the symmetric difference of any two of these sets has positive measure.

Exercise 8.5.34. Show that the inverse of a measure-preserving isomorphism is a measure-preserving isomorphism.

Exercise 8.5.35. Let $T : X \to X$ be a map and let $\varphi : X \to \mathbb{R}$ be a real-valued function. Let $x \in X$. If $k, n \in \mathbb{N}$ are such that $k < n$, prove that

$$S_n\varphi(x) = S_k\varphi(x) + S_{n-k}\varphi(T^k(x)).$$

Exercise 8.5.36. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. Prove the following statements:
(a) If $Y \subseteq X$ is forward T-invariant, then $\mu(Y) = \mu(\bigcup_{n=0}^{\infty} T^{-n}(Y))$.
(b) If $Z \subseteq X$ is backward T-invariant, then $\mu(Z) = \mu(\bigcap_{n=0}^{\infty} T^{-n}(Z))$.
(c) If T is ergodic with respect to μ and $W \subseteq X$ is forward or backward T-invariant, then $\mu(W) \in \{0, 1\}$.

N. B.: Part (c) means that the concept of ergodicity might alternatively be defined in terms of forward invariant sets or in terms of backward invariant sets.

Exercise 8.5.37. Find a topological dynamical system that admits a measure which is ergodic but not invariant.

Exercise 8.5.38. Fix $n > 1$. Consider the map $T_n : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $T_n(x) = nx \pmod 1$. Using the uniqueness of the Fourier series representation of functions in $L^2(\lambda)$ (like in the proof of Proposition 8.2.29), show that T_n is ergodic with respect to the Lebesgue measure λ.

Exercise 8.5.39. Recalling Exercise 8.5.25, show that the Gauss measure is ergodic for the Gauss map.

Exercise 8.5.40. Suppose that X is a countable set, $\mathcal{P}(X)$ is the σ-algebra of all subsets of X and μ is a probability measure on $(X, \mathcal{P}(X))$. Show that if $T : X \to X$ is ergodic with respect to μ then there exists a periodic point y of T such that $\mu(\{T^n(y) : n \geq 0\}) = 1$.

Exercise 8.5.41. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. Let also $f \in L^1(\mu)$. Show that

$$\lim_{n\to\infty} \frac{1}{n} f(T^n(x)) = 0 \quad \text{for } \mu\text{-a. e. } x \in X.$$

Exercise 8.5.42. Let $n \geq 2$. Every number $x \in [0, 1]$ has a n-adic expansion, that is,

$$x = \sum_{i=1}^{\infty} \frac{\omega_i}{n^i}$$

for some $\omega = (\omega_i)_{i=1}^{\infty} \in \{0, 1, \ldots, n-1\}^{\infty}$. Let $p, q \in \{0, 1, \ldots, n-1\}$. Show that

$$\lim_{k\to\infty} \frac{1}{k} \#\{1 \leq j \leq k : \omega_j = p \text{ and } \omega_{j+3} = q\} = \frac{1}{n^2} \quad \text{for } \lambda\text{-a. e. } x \in [0, 1].$$

Exercise 8.5.43. Let $n = 2$. Using the same notation as in Exercise 8.5.42, show that

$$\lim_{k\to\infty} \frac{1}{k} \sum_{j=0}^{k-1} (\omega_j^2 + \omega_{j+1}^2) = 1 \quad \text{for } \lambda\text{-a. e. } x \in [0, 1].$$

Exercise 8.5.44. Let $\ell \in C(X)^*$. Set $\Delta\ell := \operatorname{var}(\ell) - \ell$, where $\operatorname{var}(\ell)$ comes from Definition 8.2.36. Prove that $\Delta\ell \in C(X)^*$ and is positive. In addition, show that if ℓ is T-invariant then so is $\Delta\ell$.

Exercise 8.5.45. Prove Theorem 8.4.2(a,b,c).

Exercise 8.5.46. Referring back to the proof of Theorem 8.4.3, show that $\widetilde{\mathcal{F}}_n = \widetilde{T}^{-1}(\widetilde{\mathcal{F}}_{n+1})$ for all $n \geq 0$.

Exercise 8.5.47. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{B}, \mu)$. A measurable function $g : X \to \mathbb{R}$ is said to be μ-a. e. T-subinvariant if $g(T(x)) \leq g(x)$ for μ-a. e. $x \in X$. Show that if T is ergodic and $g \in L^1(\mu)$ is μ-a. e. T-subinvariant, then g is μ-a. e. constant (cf. Theorem 8.2.18).

Exercise 8.5.48. In this exercise, we introduce and discuss Markov chains, which are in a sense a generalization of the Bernoulli shifts studied in Examples 8.1.14 and 8.2.32. Let E be a countable alphabet with at least two letters. Let $A : E \times E \to (0,1)$ be a stochastic matrix, that is, a matrix such that

$$\sum_{j\in E} A_{ij} = 1, \quad \forall i \in E.$$

(a) Prove (you may use the classical Perron–Frobenius theorem for positive matrices) that there exists a unique probability vector $P : E \to [0,1]$ such that $PA = P$, that is,

$$\sum_{i\in E} P_i A_{ij} = P_j, \quad \forall j \in E.$$

(b) Further show that $P_j \in (0,1)$ for all $j \in E$.

For every $\omega \in E^*$, say $\omega \in E^n$, set

$$\mu_A([\omega]) = P_{\omega_1} \prod_{k=1}^{n-1} A_{\omega_k \omega_{k+1}}.$$

In a similar way to Examples 8.1.14 and 8.2.32:

(c) Prove that μ_A uniquely extends to a Borel probability measure on $E^{\mathbb{N}}$. (In the sequel, we keep the same symbol μ_A for this measure.)

(d) Show that the measure μ_A is shift-invariant and ergodic.

The dynamical system $(\sigma : E^{\mathbb{N}} \to E^{\mathbb{N}}, \mu_A)$ is called the one-sided *Markov chain* generated by the stochastic matrix A.

Assuming that the alphabet E is finite:

(e) Prove that the one-cylinder partition $\{[e]\}_{e\in E}$ is a weak Bernoulli generator for the measure-preserving dynamical system $(\sigma : E^{\mathbb{N}} \to E^{\mathbb{N}}, \mu_A)$. Conclude that this system is weak Bernoulli and that its Rokhlin's natural extension is isomorphic to a two-sided Bernoulli shift.

(f) Wanting an explicit representation of the Rokhlin's natural extension of the system $(\sigma : E^{\mathbb{N}} \to E^{\mathbb{N}}, \mu_A)$, prove that, as in the case of Bernoulli shifts, this extension is isomorphic to the naturally defined two-sided Markov chain $(\sigma : E^{\mathbb{Z}} \to E^{\mathbb{Z}}, \mu_A)$ generated by the stochastic matrix A and the unique probability vector P.

Note that when E is a countable set, every Bernoulli shift generated by a probability vector $P : E \to (0,1)$ is the Markov chain generated by the stochastic matrix A whose columns are all equal to that vector P. In this sense, Markov chains are generalizations of Bernoulli shifts. We established this for countable alphabets E. But in fact, even more generally, when E is an arbitrary set as in Examples 8.1.14 and 8.2.32, one can define appropriate Markov chains that generalize the Bernoulli shifts considered in those examples.

Finally, as for Bernoulli shifts, Markov chains will be shown to be Gibbs and equilibrium states for appropriately chosen Hölder continuous potentials in Exercises 13.11.11–13.11.12 and 17.9.13–17.9.14.

9 Measure-theoretic entropy

In Chapter 7, we studied the topological entropy of a topological dynamical system. We now study its measure-theoretic counterpart. Measure-theoretic entropy is also sometimes known as metric entropy or Kolmogorov–Sinai metric entropy. It was introduced by A. Kolmogorov and Ya. Sinai in the late 1950s; see [67]. Since then, its account has been presented in virtually every textbook on ergodic theory. Its introduction to dynamical systems was motivated by Ludwig Boltzmann's concept of entropy in statistical mechanics and Claude Shannon's work on information theory; see [64, 65].

As for topological entropy, there are three stages in the definition of metric entropy. Recall that topological entropy is defined by **covering** the underlying topological space with basic sets in that space, that is, open sets; metric entropy, on the other hand, is defined by **partitioning** the underlying measurable space with basic sets in that space, namely, measurable sets. Indeed, whereas one cannot generally partition a topological space into open sets (this is only possible in a disconnected space), it is generally possible to partition a measurable space into measurable sets. Accordingly, we first study measurable partitions in Section 9.2. Then we examine the concepts of information and conditional information in Section 9.3. In Section 9.4, we finally define metric entropy. And in Section 9.5, we formulate and prove the full version of Shannon–McMillan–Breiman's characterization of metric entropy. This characterization depicts what metric entropy really is. Finally, in Section 9.6 we shed further light on the nature of entropy by proving the Brin–Katok local entropy formula. Like the Shannon–McMillan–Breiman theorem, the Brin–Katok local entropy formula is very useful in applications.

9.1 An excursion into the origins of entropy

This exposition is inspired by [80].

The concept of metric entropy arose from the creation of information theory by Shannon [64, 65]. That notion was adapted from Boltzmann's advances on entropy in statistical mechanics.

Contemplate the conduct of a random experiment (for instance, the rolling of a die) with a finite number of possible outcomes $(A_1, A_2, \ldots, A_n)$ with respective probabilities $(p_1, p_2, \ldots, p_n)$. Naturally, we would like to ascribe to this experiment a quantity (a number) that indicates the level of uncertainty associated with the outcome of the experiment. For example, if a six-face unfair die has the outcomes $(1, 2, 3, 4, 5, 6)$ with associated probabilities $(95\,\%, 1\,\%, 1\,\%, 1\,\%, 1\,\%, 1\,\%)$, then the level of uncertainty of the outcome is much smaller than the level of uncertainty in the throwing of a fair die, that is, a die with an equal probability of 1/6 of falling on any of its 6 faces. We aim at

https://doi.org/10.1515/9783110702682-009

finding a real-valued function

$$H(p_1, p_2, \dots, p_n)$$

that describes the level of uncertainty. Obviously, this nonnegative function is defined on the n-tuples $(p_1, p_2, \dots, p_n)$ such that $0 \le p_i \le 1$ for all $1 \le i \le n$ and $\sum_{i=1}^n p_i = 1$. We now provide a rationale for the type of function that naturally emerges in this context.

Intuitively, the level of uncertainty of the outcome reaches...

– a minimum of 0 when one of the outcomes is absolutely certain, that is, has a probability of $100\,\% = 1$ of occurring; this means that $H(p_1, p_2, \dots, p_n) = 0$ when $p_i = 1$ for some i (and thus $p_j = 0$ for all $j \ne i$).
– a maximum when the n outcomes have equal probability $1/n$ of taking place, that is, $\max H(p_1, p_2, \dots, p_n) = H(1/n, 1/n, \dots, 1/n)$.

Speaking of the equiprobable case, it is interesting and even crucial to look at the behavior of the function

$$f(n) = H(1/n, 1/n, \dots, 1/n).$$

As the number n of outcomes grows, the level of uncertainty grows: it is easier to predict the outcome of casting a six-side fair die than a twenty-side fair die. So we posit that:

(A1) The function f must be strictly increasing.

Consider now two independent experiments, a first one with n_1 equiprobable outcomes and a second one with n_2 equiprobable outcomes. Running both experiments simultaneously results in a “product” experiment which consists of $n_1 n_2$ equiprobable outcomes. Knowledge of the outcome of the first experiment does not affect the uncertainty surrounding the outcome of the second one, and vice versa. Accordingly, it is natural to expect that once the uncertainty of the outcome of the first experiment, $f(n_1)$, is subtracted from the uncertainty on the entire experiment, $f(n_1 n_2)$, the remaining uncertainty coincides with that of the conduct of the second experiment, $f(n_2)$, that is,

$$f(n_1 n_2) - f(n_1) = f(n_2).$$

Equivalently, we posit that

(A2) $f(n_1 n_2) = f(n_1) + f(n_2)$, $\forall n_1, n_2 \in \mathbb{N}$.

Let us return now to a general experiment with n outcomes. Partition these outcomes into two subsets A and B, with respective total probabilities $p_A = p_1 + \dots + p_k$ and $p_B = p_{k+1} + \dots + p_n$. For instance, in the experiment of throwing a six-face die, we might be interested not in the face number the die falls on but rather in whether it settles on an even face or an odd face. Naturally, we would like to relate the uncertainty of the original experiment with that of the "simplified" experiment where the outcome is perceived as A or B. If the outcome of the simplified experiment is A, then the remaining uncertainty about the outcome of the original experiment is $\mathrm{H}(p_1/p_A, \dots, p_k/p_A)$. Similarly, if the outcome of the simplified experiment is B, then the remaining uncertainty about the outcome of the original experiment is $\mathrm{H}(p_{k+1}/p_B, \dots, p_n/p_B)$. Since the two outcomes in the simplified experiment occur with probabilities p_A and p_B respectively, we posit that the level of uncertainty about the original experiment can be expressed as

(A3) $\mathrm{H}(p_1, \dots, p_n) = \mathrm{H}(p_A, p_B) + p_A \mathrm{H}(\frac{p_1}{p_A}, \dots, \frac{p_k}{p_A}) + p_B \mathrm{H}(\frac{p_{k+1}}{p_B}, \dots, \frac{p_n}{p_B})$.

Finally, it is reasonable to assume that the function H is continuous, that is, a small change in the probabilities of the outcomes of an experiment, results in a small change in the level of uncertainty besetting the experiment. Because of axiom (A3), it suffices to make this assumption in the case of a binary outcome experiment:

(A4) The function $p \mapsto \mathrm{H}(p, 1-p)$ is continuous on $(0, 1)$.

Theorem 9.1.1. *The only functions satisfying axioms* (A1)–(A4) *are the functions of the form*

$$\mathrm{H}(p_1, \dots, p_n) = -C \sum_{i=1}^{n} p_i \log p_i$$

for some constant $C > 0$.

Proof. See Exercise 9.7.1. □

This will be the form of the entropy function. Mathematically, the various events that can be witnessed in an experiment constitute a measurable partition of the measurable space of all outcomes of the experiment. This explains why we study measurable partitions in the next section.

9.2 Partitions of a measurable space

Definition 9.2.1. Let $(X, \mathcal{A})$ be a measurable space. A countable measurable *partition* of X is a family $\alpha = \{A_k\}_{k=1}^{\infty}$ such that:
(a) $A_k \in \mathcal{A}$ for all $k \in \mathbb{N}$;
(b) $A_i \cap A_j = \emptyset$ for all $i \neq j$; and
(c) $\bigcup_{k=1}^{\infty} A_k = X$.

The individual sets A_k, $k \in \mathbb{N}$, making up the partition α are called *atoms* of α. For each $x \in X$ the unique atom of the partition α which contains the point x will be denoted by $\alpha(x)$. Finally, we shall denote the set of all countable measurable partitions on the space $(X, \mathcal{A})$ by $\mathrm{Part}(X, \mathcal{A})$.

In the sequel, it will always be implicitly understood that partitions are countable (finite or infinite) and measurable.

In Definition 7.1.4, we introduced a refinement relation for covers of a space. As partitions of a space constitute a special class of covers of that space, it is natural to examine the restriction of the refinement relation to partitions. In contrast with the relation for covers, the restriction turns out to be a relation of partial order on the set of all partitions.

Definition 9.2.2. Let $(X, \mathcal{A})$ be a measurable space and $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. We say that partition β is *finer* than partition α, or that α is *coarser* than β, which will be denoted by $\alpha \leq \beta$, if for every atom $B \in \beta$ there exists some atom $A \in \alpha$ such that $B \subseteq A$. In other words, each atom of α is a union of atoms of β.

Equivalently, β is a refinement of α if $\beta(x) \subseteq \alpha(x)$ for all $x \in X$. See also Exercises 9.7.2–9.7.5.

We now introduce for partitions the analogue of the join of two covers.

Definition 9.2.3. Given $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$, the partition

$$\alpha \vee \beta := \{A \cap B \mid A \in \alpha, B \in \beta\}$$

is called the *join* of α and β.

The basic properties of the join are given in the following lemma. Their proofs are left to the reader as an exercise.

Lemma 9.2.4. *Since partitions are covers, the relation $\leq$ and the operation $\vee$ enjoy all the properties of $\prec$ and $\vee$ described in Remark* 7.1.3 *and Lemma* 7.1.5. *Moreover, if $\alpha, \beta, \gamma \in \mathrm{Part}(X, \mathcal{A})$, then:*

(a) *$\leq$ is a relation of partial order on $\mathrm{Part}(X, \mathcal{A})$, that is, it is...*
   – *reflexive ($\alpha \leq \alpha$);*
   – *transitive ($\alpha \leq \beta$ and $\beta \leq \gamma \Longrightarrow \alpha \leq \gamma$); and*
   – *antisymmetric ($\alpha \leq \beta \leq \alpha \Longleftrightarrow \alpha = \beta$).*

(b) $\alpha \leq \beta \iff \alpha \vee \beta = \beta$.

(c) $\alpha \vee \alpha = \alpha$.

(d) $\alpha \vee \{X\} = \alpha$.

9.3 Information and conditional information functions

Let $(X, \mathcal{A}, \mu)$ be a probability space. The set X may be construed as the set of all possible states (or outcomes) of an experiment, while the σ-algebra $\mathcal{A}$ is the set of all possible events, and $\mu(A)$ is the probability that event $A \in \mathcal{A}$ take place. Imagine that this experiment is conducted using an instrument which, due to some limitation, can only provide measurements accurate up to the atoms of a partition $\alpha = \{A_k\}_{k=1}^{\infty} \in \mathrm{Part}(X, \mathcal{A})$. In other words, this instrument can only tell us which atom of α the outcome of the experiment falls into. Any observation made through this instrument will therefore be of the form A_k for a unique k. If the experiment were conducted today, the probability that its outcome belongs to A_k, that is, the probability that the experiment results in observing event A_k with our instrument, would be $\mu(A_k)$.

We would like to introduce a function that describes the information that our instrument would give us about the outcome of the experiment. So, let $x \in X$. Intuitively, the smaller the atom of the partition to which x belongs, the more information our instrument provides us about x. In particular, if x lies in an atom of full measure, then our instrument gives us essentially no information about x. Moreover, because our instrument cannot distinguish points which belong to a common atom of the partition, the sought-after information function must be constant on every atom. In light of Theorem 9.1.1, the following definition is natural (think about the relation between information and uncertainty on the outcome of an experiment).

Definition 9.3.1. Let $(X, \mathcal{A}, \mu)$ be a probability space and $\alpha \in \mathrm{Part}(X, \mathcal{A})$. The nonnegative function $I_\mu(\alpha) : X \to [0, \infty]$ defined by

$$I_\mu(\alpha)(x) := -\log \mu\big(\alpha(x)\big)$$

is called the *information function* of the partition α. By convention, $\log 0 = -\infty$.

As the function $t \mapsto -\log t$ is strictly decreasing, for any $x \in X$ the smaller $\mu(\alpha(x))$ is, the larger $I_\mu(\alpha)(x)$ is, that is, the smaller the measure of the atom $\alpha(x)$ is, the more information the partition α gives us about x. In particular, the finer the partition, the more information it gives us about every point in the space.

In the next lemma, we collect some of the basic properties of the information function. Their proofs are straightforward and are left to the reader.

Lemma 9.3.2. *Let $(X, \mathcal{A}, \mu)$ be a probability space and $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. Then:*
(a) *If $\alpha \leq \beta$, then $I_\mu(\alpha) \leq I_\mu(\beta)$.*
(b) *$I_\mu(\alpha)(x) = 0$ if and only if $\mu(\alpha(x)) = 1$.*
(c) *$I_\mu(\alpha)(x) = \infty$ if and only if $\mu(\alpha(x)) = 0$.*
(d) *If $\alpha(x) = \alpha(y)$, then $I_\mu(\alpha)(x) = I_\mu(\alpha)(y)$, that is, $I_\mu(\alpha)$ is constant over each atom of α.*

More advanced properties of the information function will be presented below. Meanwhile, we introduce a function which describes the information gathered from a partition α given that a partition β has already been applied.

Definition 9.3.3. Let $(X, \mathcal{A}, \mu)$ be a probability space and $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. The *conditional information function* of partition α given partition β is the nonnegative function

$I_\mu(\alpha|\beta) : X \to [0,\infty]$ defined by

$$I_\mu(\alpha|\beta)(x) := -\log \mu_{\beta(x)}(\alpha(x)),$$

where μ_B is the conditional measure from Definition A.1.70 in Appendix A. Observe that

$$I_\mu(\alpha|\beta)(x) := -\log \frac{\mu\big(\alpha(x)\cap\beta(x)\big)}{\mu(\beta(x))} = -\log \frac{\mu\big((\alpha\vee\beta)(x)\big)}{\mu(\beta(x))} = I_\mu(\alpha\vee\beta)(x) - I_\mu(\beta)(x).$$

By convention, $\frac{0}{0} = 0$ and $\infty - \infty = \infty$.

For any partition α, notice that $I_\mu(\alpha|\{X\}) = I_\mu(\alpha)$, that is, the information function coincides with the conditional information function with respect to the trivial partition $\{X\}$. Note further that $I_\mu(\alpha|\beta)$ is constant over each atom of $\alpha \vee \beta$.

Our next aim is to give some advanced properties of the conditional information function. Notice that some of these properties hold pointwise, while others hold atomwise only, that is, after integrating over atoms. In particular, the reader should compare statements (e–h) in the next theorem. First, though, we make one further definition, which is related to our excursion in Section 9.1.

Definition 9.3.4. Let the function $k : [0,1] \to [0,1]$ be defined by

$$k(t) = -t\log t,$$

where it is understood that $0\cdot(-\infty) = 0$.

The function k is continuous, strictly increasing on the interval $[0, e^{-1}]$, strictly decreasing on the interval $[e^{-1}, 1]$, and concave. See Figure 9.1. Recall that a function $k : I \to \mathbb{R}$, where $I \subseteq \mathbb{R}$ is an interval, is *concave* on I if

$$k\big(tx + (1-t)y\big) \geq tk(x) + (1-t)k(y), \quad \forall t \in [0,1], \forall x, y \in I,$$

that is, the line segment joining any two points on the curve lies under the curve.

Theorem 9.3.5. *Let $(X, \mathcal{A}, \mu)$ be a probability space and $\alpha, \beta, \gamma \in \mathrm{Part}(X, \mathcal{A})$. The following statements hold:*

(a) $I_\mu(\alpha\vee\beta|\gamma) = I_\mu(\alpha|\gamma) + I_\mu(\beta|\alpha\vee\gamma)$.
(b) $I_\mu(\alpha\vee\beta) = I_\mu(\alpha) + I_\mu(\beta|\alpha)$.
(c) *If $\alpha \leq \beta$, then $I_\mu(\alpha|\gamma) \leq I_\mu(\beta|\gamma)$.*
(d) *If $\alpha \leq \beta$, then $I_\mu(\alpha) \leq I_\mu(\beta)$.*
(e) *If $\beta \leq \gamma$, then for all $A \in \alpha$ and all $B \in \beta$,*

$$\int_{A\cap B} I_\mu(\alpha|\beta)\, d\mu \geq \int_{A\cap B} I_\mu(\alpha|\gamma)\, d\mu.$$

Note: In general, $\beta \leq \gamma \not\Rightarrow I_\mu(\alpha|\beta) \geq I_\mu(\alpha|\gamma)$.

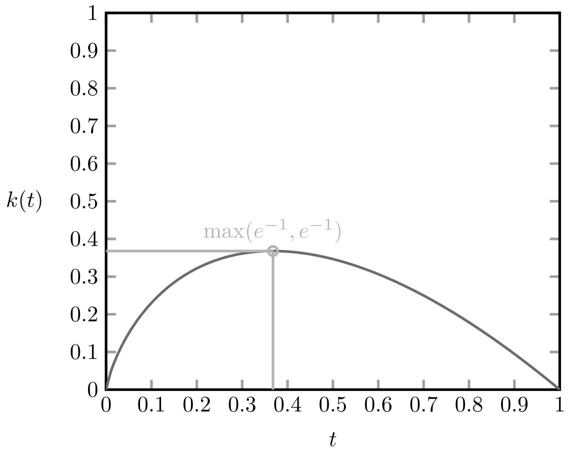

Figure 9.1: The function underlying entropy: $k(t) = -t \log t$.

(f) *For all $C \in \gamma$,*

$$\int_C I_\mu(\alpha \vee \beta|\gamma)\, d\mu \leq \int_C I_\mu(\alpha|\gamma)\, d\mu + \int_C I_\mu(\beta|\gamma)\, d\mu.$$

Note: In general, $I_\mu(\alpha \vee \beta|\gamma) \nleq I_\mu(\alpha|\gamma) + I_\mu(\beta|\gamma)$.

(g)

$$\int_X I_\mu(\alpha \vee \beta)\, d\mu \leq \int_X I_\mu(\alpha)\, d\mu + \int_X I_\mu(\beta)\, d\mu.$$

(h) *For all $A \in \alpha$ and all $B \in \beta$,*

$$\int_{A\cap B} I_\mu(\alpha|\gamma)\, d\mu \leq \int_{A\cap B} I_\mu(\alpha|\beta)\, d\mu + \int_{A\cap B} I_\mu(\beta|\gamma)\, d\mu.$$

Note: In general, $I_\mu(\alpha|\gamma) \nleq I_\mu(\alpha|\beta) + I_\mu(\beta|\gamma)$.

(i) $I_\mu(\alpha) \leq I_\mu(\alpha|\beta) + I_\mu(\beta)$.

Proof. (a) Let $x \in X$. Then

$$\begin{aligned}
I_\mu(\alpha \vee \beta|\gamma)(x) &= -\log \frac{\mu\big((\alpha \vee \beta \vee \gamma)(x)\big)}{\mu(\gamma(x))} \\
&= -\log \frac{\mu\big(\beta(x) \cap (\alpha \vee \gamma)(x)\big)}{\mu(\gamma(x))} \\
&= -\log\left(\frac{\mu\big(\beta(x) \cap (\alpha \vee \gamma)(x)\big)}{\mu\big((\alpha \vee \gamma)(x)\big)} \cdot \frac{\mu\big((\alpha \vee \gamma)(x)\big)}{\mu(\gamma(x))} \right) \\
&= -\log \frac{\mu\big(\beta(x) \cap (\alpha \vee \gamma)(x)\big)}{\mu\big((\alpha \vee \gamma)(x)\big)} - \log \frac{\mu\big((\alpha \vee \gamma)(x)\big)}{\mu(\gamma(x))} \\
&= I_\mu(\beta|\alpha \vee \gamma)(x) + I_\mu(\alpha|\gamma)(x).
\end{aligned}$$

(b) Setting $\gamma = \{X\}$ in (a) results in

$$I_\mu(\alpha \vee \beta)(x) = I_\mu(\alpha \vee \beta|\{X\})(x) = I_\mu(\beta|\alpha \vee \{X\})(x) + I_\mu(\alpha|\{X\})(x) = I_\mu(\beta|\alpha)(x) + I_\mu(\alpha)(x).$$

(c) Notice that if $\alpha \leq \beta$, then $\alpha \vee \beta = \beta$. It then follows from (a) that

$$I_\mu(\beta|\gamma) = I_\mu(\alpha \vee \beta|\gamma) = I_\mu(\alpha|\gamma) + I_\mu(\beta|\alpha \vee \gamma) \geq I_\mu(\alpha|\gamma).$$

(d) Setting $\gamma = \{X\}$ in (c) leads to (d).
(e) Suppose that $\beta \leq \gamma$. Let $A \in \alpha$ and $B \in \beta$. The downward concavity of the function k from Definition 9.3.4 means that

$$k\left(\sum_{n=1}^{\infty} a_n b_n\right) \geq \sum_{n=1}^{\infty} a_n k(b_n)$$

whenever $a_n, b_n \in [0,1]$ for all $n \in \mathbb{N}$ and $\sum_{n=1}^{\infty} a_n = 1$. Therefore,

$$k\left(\sum_{C \in \gamma} \mu_B(C) \frac{\mu(A \cap C)}{\mu(C)}\right) \geq \sum_{C \in \gamma} \mu_B(C) k\left(\frac{\mu(A \cap C)}{\mu(C)}\right). \tag{9.1}$$

Since $\beta \leq \gamma$, each atom of β is a union of atoms of γ. So, either $C \cap B = C$ or $C \cap B = \emptyset$. Thus, either $\mu_B(C) = \frac{\mu(C)}{\mu(B)}$ or $\mu_B(C) = 0$, and the left-hand side of (9.1) simplifies to

$$k\left(\sum_{C \in \gamma} \mu_B(C) \frac{\mu(A \cap C)}{\mu(C)}\right) = k\left(\sum_{C \subseteq B} \frac{\mu(A \cap C)}{\mu(B)}\right) = k\left(\frac{\mu(A \cap B)}{\mu(B)}\right) = -\frac{\mu(A \cap B)}{\mu(B)} \log \frac{\mu(A \cap B)}{\mu(B)}.$$

The right-hand side of (9.1) reduces to

$$\sum_{C \in \gamma} \mu_B(C) k\left(\frac{\mu(A \cap C)}{\mu(C)}\right) = \sum_{C \subseteq B} \frac{\mu(C)}{\mu(B)} k\left(\frac{\mu(A \cap C)}{\mu(C)}\right) = \sum_{C \subseteq B} -\frac{\mu(A \cap C)}{\mu(B)} \log \frac{\mu(A \cap C)}{\mu(C)}.$$

Hence inequality (9.1) becomes

$$-\frac{\mu(A \cap B)}{\mu(B)} \log \frac{\mu(A \cap B)}{\mu(B)} \geq \sum_{C \subseteq B} -\frac{\mu(A \cap C)}{\mu(B)} \log \frac{\mu(A \cap C)}{\mu(C)}.$$

Multiplying both sides by $\mu(B)$ yields

$$-\mu(A \cap B) \log \frac{\mu(A \cap B)}{\mu(B)} \geq \sum_{C \subseteq B} -\mu(A \cap C) \log \frac{\mu(A \cap C)}{\mu(C)}.$$

Then

$$\begin{aligned} \int_{A \cap B} I_\mu(\alpha|\beta)\, d\mu &= -\mu(A \cap B) \log \frac{\mu(A \cap B)}{\mu(B)} \\ &\geq \sum_{C \subseteq B} -\mu(A \cap C) \log \frac{\mu(A \cap C)}{\mu(C)} \end{aligned}$$

$$= \sum_{C \subseteq B} \int_{A \cap C} I_\mu(\alpha|\gamma)\, d\mu$$
$$= \int_{A \cap B} I_\mu(\alpha|\gamma)\, d\mu.$$

(f) Since $\gamma \le \alpha \vee \gamma$, this statement follows directly from combining (a) and (e).
(g) Setting $\gamma = \{X\}$ in (f) gives (g).
(h) Using part (c) and then part (a), we obtain

$$I_\mu(\alpha|\gamma) \le I_\mu(\alpha \vee \beta|\gamma) = I_\mu(\beta|\gamma) + I_\mu(\alpha|\beta \vee \gamma).$$

Since $\beta \le \beta \vee \gamma$, part (e) ensures that for all $A \in \alpha$ and all $B \in \beta$,

$$\int_{A \cap B} I_\mu(\alpha|\beta)\, d\mu \ge \int_{A \cap B} I_\mu(\alpha|\beta \vee \gamma)\, d\mu.$$

Therefore, for all $A \in \alpha$ and $B \in \beta$, we have that

$$\int_{A \cap B} I_\mu(\alpha|\gamma)\, d\mu \le \int_{A \cap B} I_\mu(\beta|\gamma)\, d\mu + \int_{A \cap B} I_\mu(\alpha|\beta)\, d\mu.$$

(i) Using parts (d) and (b) in succession, we deduce that

$$I_\mu(\alpha) \le I_\mu(\alpha \vee \beta) = I_\mu(\alpha|\beta) + I_\mu(\beta).$$ □

9.4 Definition of measure-theoretic entropy

The entropy of a measure-preserving dynamical system $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ is defined in three stages, which, for clarity of exposition, we split into the following three subsections.

9.4.1 First stage: entropy and conditional entropy for partitions

At this stage, the behavior of the system T is not in consideration. We solely look at the absolute and relative information provided by partitions.

The information function associated with a partition gives us the amount of information that can be gathered from the partition about each and every outcome of the experiment. It is useful to encompass the information given by a partition within a single number rather than a function. A natural way to achieve this is to calculate the average information given by the partition. This means integrating the information function over the entire space. The resulting integral is called the entropy of the partition.

Definition 9.4.1. Let $(X, \mathcal{A}, \mu)$ be a probability space and $\alpha \in \operatorname{Part}(X, \mathcal{A})$. The *entropy of* α with respect to the measure μ is defined to be the nonnegative extended number

$$\mathrm{H}_\mu(\alpha) := \int_X I_\mu(\alpha)\, d\mu = \sum_{A\in\alpha} -\mu(A)\log\mu(A),$$

where it is still understood that $0 \cdot (-\infty) = 0$, since null sets do not contribute to the integral.

The entropy of a partition is equal to zero if and only if the partition has an atom of full measure (which implies that all other atoms are of null measure). In particular, $\mathrm{H}_\mu(\{X\}) = 0$. Moreover, the entropy of a partition is small if the partition contains one atom with nearly full measure (so all other atoms have small measure). If the partition α is finite, it is possible, using calculus, to show that

$$0 \le \mathrm{H}_\mu(\alpha) \le \log \#\alpha \tag{9.2}$$

and that

$$\mathrm{H}_\mu(\alpha) = \log \#\alpha \quad \Longleftrightarrow \quad \mu(A) = \frac{1}{\#\alpha}, \quad \forall A \in \alpha.$$

In other words, on average we gain the most information from carrying out an experiment when the potential events are equiprobable of occurring.

Definition 9.4.2. Let $(X, \mathcal{A}, \mu)$ be a probability space and $\alpha, \beta \in \operatorname{Part}(X, \mathcal{A})$. The *conditional entropy of* α *given* β is defined to be

$$\mathrm{H}_\mu(\alpha|\beta) := \int_X I_\mu(\alpha|\beta)\, d\mu = \sum_{A\in\alpha}\sum_{B\in\beta} -\mu(A\cap B)\log\frac{\mu(A\cap B)}{\mu(B)}.$$

Note that $\mathrm{H}_\mu(\alpha) = \mathrm{H}_\mu(\alpha|\{X\})$. Recalling the measure μ_B from Definition A.1.70 and defining a partition $\alpha|_B$ of B by $\alpha|_B := \{A \cap B : A \in \alpha\}$, it follows that

$$\begin{aligned}
\mathrm{H}_\mu(\alpha|\beta) &= \sum_{B\in\beta}\sum_{A\in\alpha} -\frac{\mu(A\cap B)}{\mu(B)}\log\frac{\mu(A\cap B)}{\mu(B)}\cdot\mu(B)\\
&= \sum_{B\in\beta}\sum_{A\in\alpha} -\mu_B(A)\log\mu_B(A)\cdot\mu(B)\\
&= \sum_{B\in\beta} \mathrm{H}_{\mu_B}(\alpha|_B)\cdot\mu(B).
\end{aligned}$$

Hence, the conditional entropy of α given β is the weighted average of the entropies of the partitions of each atom $B \in \beta$ into the sets $\{A \cap B : A \in \alpha\}$.

Of course, the properties of entropy (resp. conditional entropy) are inherited from the properties of the information function (resp., the conditional information function) via integration, as the following theorem shows.

Theorem 9.4.3. *Let* $(X, \mathcal{A}, \mu)$ *be a probability space and* $\alpha, \beta, \gamma \in \mathrm{Part}(X, \mathcal{A})$. *The following statements hold:*
(a) $\mathrm{H}_\mu(\alpha \vee \beta|\gamma) = \mathrm{H}_\mu(\alpha|\gamma) + \mathrm{H}_\mu(\beta|\alpha \vee \gamma)$.
(b) $\mathrm{H}_\mu(\alpha \vee \beta) = \mathrm{H}_\mu(\alpha) + \mathrm{H}_\mu(\beta|\alpha)$.
(c) *If* $\alpha \leq \beta$, *then* $\mathrm{H}_\mu(\alpha|\gamma) \leq \mathrm{H}_\mu(\beta|\gamma)$.
(d) *If* $\alpha \leq \beta$, *then* $\mathrm{H}_\mu(\alpha) \leq \mathrm{H}_\mu(\beta)$.
(e) *If* $\beta \leq \gamma$, *then* $\mathrm{H}_\mu(\alpha|\beta) \geq \mathrm{H}_\mu(\alpha|\gamma)$.
(f) $\mathrm{H}_\mu(\alpha \vee \beta|\gamma) \leq \mathrm{H}_\mu(\alpha|\gamma) + \mathrm{H}_\mu(\beta|\gamma)$.
(g) $\mathrm{H}_\mu(\alpha \vee \beta) \leq \mathrm{H}_\mu(\alpha) + \mathrm{H}_\mu(\beta)$.
(h) $\mathrm{H}_\mu(\alpha|\gamma) \leq \mathrm{H}_\mu(\alpha|\beta) + \mathrm{H}_\mu(\beta|\gamma)$.
(i) $\mathrm{H}_\mu(\alpha) \leq \mathrm{H}_\mu(\alpha|\beta) + \mathrm{H}_\mu(\beta)$.

Proof. All the statements follow from their counterparts in Theorem 9.3.5 after integration or summation over atoms. For instance, let us prove (e). If $\beta \leq \gamma$, then it follows from Theorem 9.3.5(e) that

$$\mathrm{H}_\mu(\alpha|\beta) = \int_X I_\mu(\alpha|\beta)\, d\mu = \sum_{A\in\alpha} \sum_{B\in\beta} \int_{A\cap B} I_\mu(\alpha|\beta)\, d\mu \geq \sum_{A\in\alpha} \sum_{B\in\beta} \int_{A\cap B} I_\mu(\alpha|\gamma)\, d\mu = \mathrm{H}_\mu(\alpha|\gamma). \qquad \square$$

9.4.2 Second stage: entropy of a system relative to a partition

In this second stage, we take into account the behavior of a measure-preserving dynamical system relative to a given partition. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$ and $\alpha \in \mathrm{Part}(X, \mathcal{A})$. Observe that $T^{-1}\alpha := \{T^{-1}(A) : A \in \alpha\} \in \mathrm{Part}(X, \mathcal{A})$, too.

Recall that the set X represents the set of all possible outcomes (or states) of an experiment, while the σ-algebra $\mathcal{A}$ consists of the set of all possible events, and $\mu(A)$ is the probability that event A happen. Recall also that a partition $\alpha = \{A_k\}$ can be thought of as the set of all observations that can be made with a given instrument. The action of T on $(X, \mathcal{A}, \mu)$ may be conceived as the passage of one unit of time (for instance, a day). Today would naturally be taken as reference point for time 0. Suppose that we conduct the experiment with our instrument tomorrow. The resulting observation would be one of the atoms of α, say A_{k_1}, on day 1. Due to the passage of time (in other words, one iteration of T), in order to make observation A_{k_1} at time 1, our measure-preserving system would have to be, today, in one of the states of $T^{-1}(A_{k_1})$. The probability of making observation A_{k_1} on day 1 is thus $\mu(T^{-1}(A_{k_1}))$. Assume now that we conduct the same experiment for n consecutive days, starting today. What is the probability that we make the sequence of observations $A_{k_0}, A_{k_1}, \ldots, A_{k_{n-1}}$ on those successive days? We would make those observations precisely if our system is, today, in one of the states of $\bigcap_{m=0}^{n-1} T^{-m}(A_{k_m})$. Therefore, the probability that our observations are respectively $A_{k_0}, A_{k_1}, \ldots, A_{k_{n-1}}$ on n successive days starting today is $\mu(\bigcap_{m=0}^{n-1} T^{-m}(A_{k_m}))$. It is thus natural to consider for all $0 \leq m < n$ the partitions

$$\alpha_m^n := \bigvee_{i=m}^{n-1} T^{-i}\alpha = T^{-m}\alpha \vee \cdots \vee T^{-(n-1)}\alpha.$$

If $m \geq n$, we define α_m^n to be the trivial partition $\{X\}$. To shorten notation, we shall write α^n in lieu of α_0^n and $T^{-i}\alpha$ rather than $T^{-i}(\alpha)$. In the following lemma, we list some of the basic properties of the operator T^{-1} on partitions.

Lemma 9.4.4. *Let $T : X \to X$ be a measurable transformation of a measurable space $(X, \mathcal{A})$, and let $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. The following statements hold:*
(a) $T^{-1}(\alpha \vee \beta) = (T^{-1}\alpha) \vee (T^{-1}\beta)$.
(b) $T^{-1}(\alpha_m^n) = (T^{-1}\alpha)_m^n$ *for all* $m, n \geq 0$.
(c) $(\alpha \vee \beta)_m^n = \alpha_m^n \vee \beta_m^n$ *for all* $m, n \geq 0$.
(d) $(\alpha_k^l)_m^n = \alpha_{k+m}^{l+n-1}$.
(e) T^{-1} *preserves the partial order* $\leq$*, that is, if* $\alpha \leq \beta$ *then* $T^{-1}\alpha \leq T^{-1}\beta$.
(f) *More generally, if* $\alpha \leq \beta$ *then* $\alpha_m^n \leq \beta_m^n$ *for all* $m, n \geq 0$.
(g) $(T^{-1}\alpha)(x) = T^{-1}(\alpha(T(x)))$ *for all* $x \in X$.

Proof. The proof of assertions (a) and (e) are left to the reader.

(b) Using (a) repeatedly, we have that

$$T^{-1}(\alpha_m^n) = T^{-1}\left(\bigvee_{i=m}^{n-1} T^{-i}\alpha\right) = \bigvee_{i=m}^{n-1} T^{-1}(T^{-i}\alpha) = \bigvee_{i=m}^{n-1} T^{-i}(T^{-1}\alpha) = (T^{-1}\alpha)_m^n.$$

(c) Again by using (a) repeatedly, we obtain that

$$(\alpha \vee \beta)_m^n = \bigvee_{i=m}^{n-1} T^{-i}(\alpha \vee \beta) = \bigvee_{i=m}^{n-1}(T^{-i}\alpha \vee T^{-i}\beta) = \left(\bigvee_{i=m}^{n-1} T^{-i}\alpha\right) \vee \left(\bigvee_{i=m}^{n-1} T^{-i}\beta\right) = \alpha_m^n \vee \beta_m^n.$$

(d) Using (a), it follows that

$$(\alpha_k^l)_m^n = \bigvee_{j=m}^{n-1} T^{-j}(\alpha_k^l) = \bigvee_{j=m}^{n-1} T^{-j}\left(\bigvee_{i=k}^{l-1} T^{-i}\alpha\right) = \bigvee_{j=m}^{n-1}\bigvee_{i=k}^{l-1} T^{-(i+j)}\alpha = \bigvee_{s=k+m}^{l+n-2} T^{-s}\alpha = \alpha_{k+m}^{l+n-1}.$$

(f) Suppose that $\alpha \leq \beta$. Using (e) repeatedly and Lemma 7.1.5(g), we obtain that

$$\alpha_m^n = \bigvee_{i=m}^{n-1} T^{-i}\alpha \leq \bigvee_{i=m}^{n-1} T^{-i}\beta = \beta_m^n.$$

(g) Let $x \in X$. Choose $A \in \alpha$ such that $x \in T^{-1}(A)$. Then $T(x) \in A$, that is, $A = \alpha(T(x))$. Hence, $(T^{-1}\alpha)(x) = T^{-1}(A) = T^{-1}(\alpha(T(x)))$. □

We now describe the behavior of the operator T^{-1} with respect to the information function for any measure-preserving dynamical system T.

Lemma 9.4.5. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. Then*

$$I_\mu(T^{-1}\alpha|T^{-1}\beta) = I_\mu(\alpha|\beta) \circ T.$$

In particular,

$$I_\mu(T^{-1}\alpha) = I_\mu(\alpha) \circ T.$$

Proof. Let $x \in X$. By Lemma 9.4.4(a) and (g) and the assumption that μ is T-invariant, we have that

$$\begin{aligned} I_\mu(T^{-1}\alpha|T^{-1}\beta)(x) &= -\log \frac{\mu\big((T^{-1}\alpha \vee T^{-1}\beta)(x)\big)}{\mu((T^{-1}\beta)(x))} \\ &= -\log \frac{\mu\big((T^{-1}(\alpha \vee \beta))(x)\big)}{\mu((T^{-1}\beta)(x))} \\ &= -\log \frac{\mu\big(T^{-1}((\alpha \vee \beta)(T(x)))\big)}{\mu\big(T^{-1}(\beta(T(x)))\big)} \\ &= -\log \frac{\mu\big((\alpha \vee \beta)(T(x))\big)}{\mu(\beta(T(x)))} \\ &= I_\mu(\alpha|\beta)\big(T(x)\big) = I_\mu(\alpha|\beta) \circ T(x). \end{aligned}$$

Set $\beta = \{X\}$ to get the particular, unconditional case. □

A more intricate property of the information function is the following.

Lemma 9.4.6. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha \in \mathrm{Part}(X, \mathcal{A})$. For all $n \in \mathbb{N}$,*

$$I_\mu(\alpha^n) = \sum_{j=1}^{n} I_\mu(\alpha|\alpha_1^j) \circ T^{n-j}.$$

Proof. We will prove this lemma by induction. For $n = 1$, since α_1^1 is by definition equal to the trivial partition $\{X\}$, we have

$$I_\mu(\alpha^1) = I_\mu(\alpha) = I_\mu(\alpha|\{X\}) = I_\mu(\alpha|\alpha_1^1) = I_\mu(\alpha|\alpha_1^1) \circ T^{1-1}.$$

Now suppose that the lemma holds for some $n \in \mathbb{N}$. Then, in light of Theorem 9.3.5(b) and Lemma 9.4.5, we obtain that

$$\begin{aligned} I_\mu(\alpha^{n+1}) &= I_\mu(\alpha \vee \alpha_1^{n+1}) = I_\mu(\alpha_1^{n+1}) + I_\mu(\alpha|\alpha_1^{n+1}) \\ &= I_\mu(T^{-1}(\alpha^n)) + I_\mu(\alpha|\alpha_1^{n+1}) = I_\mu(\alpha^n) \circ T + I_\mu(\alpha|\alpha_1^{n+1}) \end{aligned}$$

$$= \sum_{j=1}^{n} I_\mu(\alpha|\alpha_1^j) \circ T^{n-j} \circ T + I_\mu(\alpha|\alpha_1^{n+1})$$
$$= \sum_{j=1}^{n} I_\mu(\alpha|\alpha_1^j) \circ T^{n+1-j} + I_\mu(\alpha|\alpha_1^{n+1}) \circ T^{n+1-(n+1)}$$
$$= \sum_{j=1}^{n+1} I_\mu(\alpha|\alpha_1^j) \circ T^{n+1-j}. \qquad \square$$

We now turn our attention to the effect that a measure-preserving dynamical system has on entropy. In particular, observe that because the system is measure-preserving, conducting the experiment today or tomorrow (or at any time in the future) gives us the same amount of average information about the outcome. This is the meaning of the second of the following properties of entropy.

Lemma 9.4.7. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. The following statements hold:*
(a) $\mathrm{H}_\mu(T^{-1}\alpha|T^{-1}\beta) = \mathrm{H}_\mu(\alpha|\beta)$.
(b) $\mathrm{H}_\mu(T^{-1}\alpha) = \mathrm{H}_\mu(\alpha)$.
(c) $\mathrm{H}_\mu(\alpha^n|\beta^n) \le n\,\mathrm{H}_\mu(\alpha|\beta)$ *for all* $n \in \mathbb{N}$.

Proof. (a) Using Lemma 9.4.5 and the T-invariance of μ, we have that

$$\mathrm{H}_\mu(T^{-1}\alpha|T^{-1}\beta) = \int_X I_\mu(T^{-1}\alpha|T^{-1}\beta)\,d\mu = \int_X I_\mu(\alpha|\beta) \circ T\,d\mu = \int_X I_\mu(\alpha|\beta)\,d\mu = \mathrm{H}_\mu(\alpha|\beta).$$

(b) Set $\beta = \{X\}$ in (a) to obtain (b).
(c) We first prove that $\mathrm{H}_\mu(\alpha^n|\beta^n) \le \sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-j}\alpha|T^{-j}\beta)$. This statement clearly holds when $n = 1$. Suppose that it holds for some $n \in \mathbb{N}$. Using Theorem 9.4.3(a) and (e), we have that

$$\begin{aligned}
\mathrm{H}_\mu(\alpha^{n+1}|\beta^{n+1}) &= \mathrm{H}_\mu(\alpha^n \vee T^{-n}\alpha|\beta^n \vee T^{-n}\beta) \\
&= \mathrm{H}_\mu(\alpha^n|\beta^n \vee T^{-n}\beta) + \mathrm{H}_\mu(T^{-n}\alpha|\alpha^n \vee \beta^n \vee T^{-n}\beta) \\
&\le \mathrm{H}_\mu(\alpha^n|\beta^n) + \mathrm{H}_\mu(T^{-n}\alpha|T^{-n}\beta) \\
&\le \sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-j}\alpha|T^{-j}\beta) + \mathrm{H}_\mu(T^{-n}\alpha|T^{-n}\beta) \\
&= \sum_{j=0}^{n} \mathrm{H}_\mu(T^{-j}\alpha|T^{-j}\beta).
\end{aligned}$$

By induction, the above statement holds for all $n \in \mathbb{N}$. By (a), we obtain that

$$\mathrm{H}_\mu(\alpha^n|\beta^n) \le \sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-j}\alpha|T^{-j}\beta) = \sum_{j=0}^{n-1} \mathrm{H}_\mu(\alpha|\beta) = n\,\mathrm{H}_\mu(\alpha|\beta). \qquad \square$$

The average information gained by conducting an experiment on n consecutive days using the partition α is given by the entropy $\mathrm{H}_\mu(\alpha^n)$ since α^n has for atoms the sets $\bigcap_{m=0}^{n-1} T^{-m}(A_{k_m})$, where $A_{k_m} \in \alpha$ for all m. Not surprisingly, the average information gained by conducting the experiment on n consecutive days using the partition α is equal to the sum of the average conditional information gained by performing α on day $j+1$ given that the outcome of performing α over the previous j days is known, summing from the first day to the last day. This is formalized in the next lemma.

Lemma 9.4.8. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha \in \mathrm{Part}(X, \mathcal{A})$. Then for all $n \in \mathbb{N}$,*

$$\mathrm{H}_\mu(\alpha^n) = \sum_{j=1}^{n} \mathrm{H}_\mu(\alpha|\alpha_1^j).$$

Proof. We deduce from Lemma 9.4.6 and the T-invariance of μ that

$$\mathrm{H}_\mu(\alpha^n) = \int_X I_\mu(\alpha^n)\, d\mu = \sum_{j=1}^{n} \int_X I_\mu(\alpha|\alpha_1^j) \circ T^{n-j}\, d\mu = \sum_{j=1}^{n} \int_X I_\mu(\alpha|\alpha_1^j)\, d\mu = \sum_{j=1}^{n} \mathrm{H}_\mu(\alpha|\alpha_1^j). \quad \square$$

Below is an alternative expression for the entropy $\mathrm{H}_\mu(\alpha^n)$.

Lemma 9.4.9. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha \in \mathrm{Part}(X, \mathcal{A})$. Then for all $n \in \mathbb{N}$,*

$$\mathrm{H}_\mu(\alpha^n) = \sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-j}\alpha|\alpha^j).$$

Proof. We prove this result by induction. The statement is trivial when $n = 1$. Suppose that it holds for $n-1$. Using Theorem 9.4.3(b), we get

$$\begin{aligned}\mathrm{H}_\mu(\alpha^n) &= \mathrm{H}_\mu(\alpha^{n-1} \vee T^{-(n-1)}\alpha) = \mathrm{H}_\mu(\alpha^{n-1}) + \mathrm{H}_\mu(T^{-(n-1)}\alpha|\alpha^{n-1}) \\ &= \sum_{j=0}^{n-2} \mathrm{H}_\mu(T^{-j}\alpha|\alpha^j) + \mathrm{H}_\mu(T^{-(n-1)}\alpha|\alpha^{n-1}) = \sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-j}\alpha|\alpha^j).\end{aligned}$$

So the statement holds for all $n \in \mathbb{N}$. $\square$

Returning to Lemma 9.4.8, since $\alpha_1^{j+1} \geq \alpha_1^j$ observe that $\mathrm{H}_\mu(\alpha|\alpha_1^{j+1}) \leq \mathrm{H}_\mu(\alpha|\alpha_1^j)$ by Theorem 9.4.3(e). So the sequence $\big(\mathrm{H}_\mu(\alpha|\alpha_1^j)\big)_{j=1}^\infty$ decreases to some limit which we shall denote by $\mathrm{h}_\mu(T, \alpha)$. Consequently, the corresponding sequence of Cesàro averages $\big(\frac{1}{n}\sum_{j=1}^n \mathrm{H}_\mu(\alpha|\alpha_1^j)\big)_{n=1}^\infty = \big(\frac{1}{n}\mathrm{H}_\mu(\alpha^n)\big)_{n=1}^\infty$ decreases to the same limit. Thus the following definition makes sense. This is the second step in the definition of the measure-theoretic entropy of a system.

Definition 9.4.10. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha \in \operatorname{Part}(X, \mathcal{A})$. The *entropy of T with respect to α*, denoted $\mathrm{h}_\mu(T, \alpha)$, is defined by

$$\begin{aligned}\mathrm{h}_\mu(T, \alpha) &:= \lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\alpha_1^n) = \lim_{n\to\infty} \frac{1}{n}\mathrm{H}_\mu(\alpha^n)\\ &= \inf_{n\to\infty} \mathrm{H}_\mu(\alpha|\alpha_1^n) = \inf_{n\to\infty} \frac{1}{n}\mathrm{H}_\mu(\alpha^n).\end{aligned}$$

The following theorem lists some of the basic properties of $\mathrm{h}_\mu(T, \cdot)$.

Theorem 9.4.11. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha, \beta \in \operatorname{Part}(X, \mathcal{A})$. The following statements hold:*
(a) $\mathrm{h}_\mu(T, \alpha) \le \mathrm{H}_\mu(\alpha)$.
(b) $\mathrm{h}_\mu(T, \alpha) \le \mathrm{H}_\mu(T^{-1}\alpha|\alpha)$.
(c) $\mathrm{h}_\mu(T, \alpha \vee \beta) \le \mathrm{h}_\mu(T, \alpha) + \mathrm{h}_\mu(T, \beta)$.
(d) *If $\alpha \le \beta$, then* $\mathrm{h}_\mu(T, \alpha) \le \mathrm{h}_\mu(T, \beta)$.
(e) $\mathrm{h}_\mu(T, \alpha) \le \mathrm{h}_\mu(T, \beta) + \mathrm{H}_\mu(\alpha|\beta)$.
(f) $\mathrm{h}_\mu(T, T^{-1}\alpha) = \mathrm{h}_\mu(T, \alpha)$.
(g) $\mathrm{h}_\mu(T, \alpha^k) = \mathrm{h}_\mu(T, \alpha)$ *for all $k \in \mathbb{N}$.*
(h) $\mathrm{h}_\mu(T^k, \alpha^k) = k \cdot \mathrm{h}_\mu(T, \alpha)$ *for all $k \in \mathbb{N}$.*
(i) *If T is invertible, then* $\mathrm{h}_\mu(T, \alpha) = \mathrm{h}_\mu(T, \bigvee_{i=-k}^{k} T^i\alpha)$ *for all $k \in \mathbb{N}$.*
(j) *If $(\beta_n)_{n=1}^\infty$ is a sequence in* $\operatorname{Part}(X, \mathcal{A})$ *such that* $\lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\beta_n) = 0$, *then*

$$\mathrm{h}_\mu(T, \alpha) \le \liminf_{n\to\infty} \mathrm{h}_\mu(T, \beta_n).$$

(k) *If* $\lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\beta^n) = 0$, *then* $\mathrm{h}_\mu(T, \alpha) \le \mathrm{h}_\mu(T, \beta)$.

Proof. (a) This follows from the fact that $\mathrm{h}_\mu(T, \alpha) = \inf_{n\in\mathbb{N}} \frac{1}{n}\mathrm{H}_\mu(\alpha^n)$.
(b) Using Lemmas 9.4.9 and 9.4.7(a) and Theorem 9.4.3(e), we have

$$\begin{aligned}\mathrm{h}_\mu(T, \alpha) &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-j}\alpha|\alpha^j)\\ &\le \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-j}\alpha\big|T^{-(j-1)}\alpha)\\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-(j-1)}(T^{-1}\alpha)\big|T^{-(j-1)}\alpha)\\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \mathrm{H}_\mu(T^{-1}\alpha|\alpha) = \mathrm{H}_\mu(T^{-1}\alpha|\alpha).\end{aligned}$$

(c) Using Lemma 9.4.4(c) and Theorem 9.4.3(g), we get

$$\begin{aligned}\mathrm{h}_\mu(T, \alpha \vee \beta) &= \lim_{n\to\infty} \frac{1}{n}\mathrm{H}_\mu((\alpha \vee \beta)^n)\\ &= \lim_{n\to\infty} \frac{1}{n}\mathrm{H}_\mu(\alpha^n \vee \beta^n)\end{aligned}$$

$$\begin{aligned}
&\le \lim_{n\to\infty}\frac{1}{n}\big[\mathrm{H}_\mu(\alpha^n)+\mathrm{H}_\mu(\beta^n)\big]\\
&= \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\alpha^n)+\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\beta^n)\\
&= \mathrm{h}_\mu(T,\alpha)+\mathrm{h}_\mu(T,\beta).
\end{aligned}$$

(d) If $\alpha \le \beta$, then $\alpha^n \le \beta^n$ and hence $\mathrm{H}_\mu(\alpha^n) \le \mathrm{H}_\mu(\beta^n)$ for all $n \in \mathbb{N}$. Therefore,

$$\mathrm{h}_\mu(T,\alpha) = \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\alpha^n) \le \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\beta^n) = \mathrm{h}_\mu(T,\beta).$$

(e) Calling upon Theorem 9.4.3(i) and Lemma 9.4.7(c), we obtain that

$$\begin{aligned}
\mathrm{h}_\mu(T,\alpha) &= \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\alpha^n)\\
&\le \liminf_{n\to\infty}\frac{1}{n}\big[\mathrm{H}_\mu(\alpha^n|\beta^n)+\mathrm{H}_\mu(\beta^n)\big]\\
&= \liminf_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\alpha^n|\beta^n)+\lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\beta^n)\\
&\le \mathrm{H}_\mu(\alpha|\beta)+\mathrm{h}_\mu(T,\beta).
\end{aligned}$$

(f) By Lemma 9.4.4(b), we know that $(T^{-1}\alpha)^n = T^{-1}(\alpha^n)$ for all $n \in \mathbb{N}$. Then, using Lemma 9.4.7(b), we deduce that

$$\mathrm{h}_\mu(T,T^{-1}\alpha) = \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu\big((T^{-1}\alpha)^n\big) = \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(T^{-1}(\alpha^n)) = \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\alpha^n) = \mathrm{h}_\mu(T,\alpha).$$

(g) By Lemma 9.4.4(d), we know that $(\alpha^k)^n = \alpha^{n+k-1}$ and hence

$$\begin{aligned}
\mathrm{h}_\mu(T,\alpha^k) &= \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu\big((\alpha^k)^n\big) = \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\alpha^{n+k-1})\\
&= \lim_{n\to\infty}\frac{n+k-1}{n}\cdot\frac{1}{n+k-1}\mathrm{H}_\mu(\alpha^{n+k-1})\\
&= \lim_{n\to\infty}\frac{n+k-1}{n}\cdot\lim_{n\to\infty}\frac{1}{n+k-1}\mathrm{H}_\mu(\alpha^{n+k-1})\\
&= \lim_{m\to\infty}\frac{1}{m}\mathrm{H}_\mu(\alpha^m) = \mathrm{h}_\mu(T,\alpha).
\end{aligned}$$

(h) Let $k \in \mathbb{N}$. Then

$$\begin{aligned}
\mathrm{h}_\mu(T^k,\alpha^k) &= \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu\Bigg(\bigvee_{j=0}^{n-1}T^{-kj}(\alpha^k)\Bigg)\\
&= \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu\Bigg(\bigvee_{j=0}^{n-1}T^{-kj}\Bigg(\bigvee_{i=0}^{k-1}T^{-i}\alpha\Bigg)\Bigg)\\
&= \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu\Bigg(\bigvee_{l=0}^{kn-1}T^{-l}\alpha\Bigg)\\
&= k\lim_{n\to\infty}\frac{1}{kn}\mathrm{H}_\mu(\alpha^{kn}) = k\,\mathrm{h}_\mu(T,\alpha).
\end{aligned}$$

(i) The proof is similar to that of part (f) and is thus left to the reader.
(j) Let $(\beta_n)_{n=1}^{\infty}$ be a sequence of partitions such that $\lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\beta_n) = 0$. By part (e),

$$\mathrm{h}_\mu(T,\alpha) \le \mathrm{h}_\mu(T,\beta_n) + \mathrm{H}_\mu(\alpha|\beta_n), \quad \forall n \in \mathbb{N}.$$

Consequently,

$$\begin{aligned}\mathrm{h}_\mu(T,\alpha) &\le \liminf_{n\to\infty}\big[\mathrm{h}_\mu(T,\beta_n) + \mathrm{H}_\mu(\alpha|\beta_n)\big] \\ &= \liminf_{n\to\infty} \mathrm{h}_\mu(T,\beta_n) + \lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\beta_n) = \liminf_{n\to\infty} \mathrm{h}_\mu(T,\beta_n).\end{aligned}$$

(k) Suppose that $\lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\beta^n) = 0$. By parts (j) and (g), we have

$$\mathrm{h}_\mu(T,\alpha) \le \liminf_{n\to\infty} \mathrm{h}_\mu(T,\beta^n) = \mathrm{h}_\mu(T,\beta).$$ □

9.4.3 Third and final stage: entropy of a system

The measure-theoretic entropy of a system is defined in a similar way to topological entropy. The third and last step in the definition consists in passing to a supremum.

Definition 9.4.12. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. The *measure-theoretic entropy* of T, denoted $\mathrm{h}_\mu(T)$, is defined by

$$\mathrm{h}_\mu(T) := \sup\{\mathrm{h}_\mu(T,\alpha) : \alpha \in \mathrm{Part}_{\mathrm{Fin}}(X,\mathcal{A})\},$$

where

$$\mathrm{Part}_{\mathrm{Fin}}(X,\mathcal{A}) := \{\alpha \in \mathrm{Part}(X,\mathcal{A}) : \#\alpha < \infty\}.$$

The following theorem is a useful tool for calculating the measure-theoretic entropy of a system. The first part is analogous to Theorem 7.2.19.

Theorem 9.4.13. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. The following statements hold:*
(a) $\mathrm{h}_\mu(T^k) = k \cdot \mathrm{h}_\mu(T)$ *for all* $k \in \mathbb{N}$.
(b) *If T is invertible, then* $\mathrm{h}_\mu(T^{-1}) = \mathrm{h}_\mu(T)$.

Proof. (a) Let $k \in \mathbb{N}$. Using Theorem 9.4.11(h), we get

$$\begin{aligned}k\,\mathrm{h}_\mu(T) &= \sup\{k\,\mathrm{h}_\mu(T,\alpha) : \alpha \in \mathrm{Part}_{\mathrm{Fin}}(X,\mathcal{A})\} \\ &= \sup\{\mathrm{h}_\mu(T^k,\alpha^k) : \alpha \in \mathrm{Part}_{\mathrm{Fin}}(X,\mathcal{A})\} \\ &\le \sup\{\mathrm{h}_\mu(T^k,\beta) : \beta \in \mathrm{Part}_{\mathrm{Fin}}(X,\mathcal{A})\} = \mathrm{h}_\mu(T^k).\end{aligned}$$

On the other hand, since $\alpha \le \alpha^k$ for all $k \in \mathbb{N}$, Theorem 9.4.11(d) and (h) give

$$\mathrm{h}_\mu(T^k, \alpha) \le \mathrm{h}_\mu(T^k, \alpha^k) = k\, \mathrm{h}_\mu(T, \alpha).$$

Passing to the supremum over all finite partitions α of X on both sides, we obtain the desired inequality, namely, $\mathrm{h}_\mu(T^k) \le k\, \mathrm{h}_\mu(T)$.

(b) To distinguish the action of T from the action of T^{-1} on a partition, we shall use the respective notation α_T^n and $\alpha_{T^{-1}}^n$. Using Lemmas 9.4.7(b) and 9.4.4(b) in turn, we deduce that

$$\begin{aligned}\mathrm{H}_\mu(\alpha_{T^{-1}}^n) = \mathrm{H}_\mu\left(\bigvee_{i=0}^{n-1} (T^{-1})^{-i}\alpha\right) &= \mathrm{H}_\mu\left(\bigvee_{i=0}^{n-1} T^i\alpha\right) \\ &= \mathrm{H}_\mu\left(T^{-(n-1)}\left(\bigvee_{i=0}^{n-1} T^i\alpha\right)\right) \\ &= \mathrm{H}_\mu\left(\bigvee_{i=0}^{n-1} T^{-(n-1-i)}\alpha\right) \\ &= \mathrm{H}_\mu\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) = \mathrm{H}_\mu(\alpha_T^n).\end{aligned}$$

It follows that $\mathrm{h}_\mu(T^{-1}, \alpha) = \mathrm{h}_\mu(T, \alpha)$ for every partition α, and thus, passing to the supremum on both sides, we conclude that $\mathrm{h}_\mu(T^{-1}) = \mathrm{h}_\mu(T)$. □

In Theorem 7.2.24, we observed that the topological entropy of an expansive dynamical system can be determined by simply calculating the entropy of that system with respect to any cover of sufficiently small diameter. We intend to prove the corresponding result for measure-theoretic entropy by the end of this section. We begin the journey to that destination with a purely measure-theoretical lemma. It says that given a finite Borel partition α of a compact metric space X and given any Borel partition β of X of sufficiently small diameter, we can group the atoms of β together in such a way that we nearly reconstruct the partition α. Notice that β may be countably infinite.

To simplify notation, we shall write $\mathrm{Part}(X) := \mathrm{Part}(X, \mathcal{B}(X))$.

Lemma 9.4.14. *Let X be a compact metric space and $\mu \in M(X)$. Let also $\alpha = \{A_1, A_2, \dots, A_n\} \in \mathrm{Part}_{\mathrm{Fin}}(X)$. Then for all $\varepsilon > 0$ there exists $\delta > 0$ so that for every $\beta \in \mathrm{Part}(X)$ with $\mathrm{diam}(\beta) < \delta$ there is $\beta' = \{B_1', B_2', \dots, B_n'\} \in \mathrm{Part}_{\mathrm{Fin}}(X)$ such that*

$$\beta' \le \beta \quad \textit{and} \quad \mu(B_i' \triangle A_i) < \varepsilon, \quad \forall 1 \le i \le n.$$

Proof. Fix $\varepsilon > 0$. Since μ is regular, for each $1 \le i \le n$ there exists a compact set $K_i \subseteq A_i$ such that

$$\mu(A_i \backslash K_i) < \frac{\varepsilon}{n}.$$

As usual, let d denote the metric on X and set

$$\theta = \min\{d(K_i, K_j) : i \neq j\}.$$

Then $\theta > 0$, as the sets K_i are compact and mutually disjoint. Let $\delta = \theta/2$ and let β be a partition with $\operatorname{diam}(\beta) < \delta$. For each $1 \leq i \leq n$, let

$$B_i' = \bigcup_{B \in \beta : B \cap K_i \neq \emptyset} B.$$

Then each B_i' is a Borel set such that $B_i' \supseteq K_i$. Furthermore, due to the choice of δ,

$$B_i' \cap B_j' = \emptyset, \quad \forall i \neq j.$$

However, the family of pairwise disjoint Borel sets $\{B_i'\}_{i=1}^n$ may not cover X completely. Indeed, there may be some sets $C \in \beta$ such that $C \cap \bigcup_{i=1}^n K_i = \emptyset$. Take all those sets and put them into B_1'. Then $\beta' = \{B_i'\}_{i=1}^n$ is a Borel partition of X such that $\beta' \leq \beta$. Moreover, since $B_j' \supseteq K_j$ for all $1 \leq j \leq n$, we get

$$\begin{aligned}
\mu(B_i' \triangle A_i) &= \mu(B_i' \backslash A_i) + \mu(A_i \backslash B_i') \\
&= \mu\big((X \backslash \cup_{j \neq i} B_j') \backslash A_i\big) + \mu(A_i \backslash B_i') \\
&\leq \mu\big((X \backslash \cup_{j \neq i} K_j) \backslash A_i\big) + \mu(A_i \backslash K_i) \\
&= \mu\big((\cup_{k=1}^n A_k \backslash \cup_{j \neq i} K_j) \backslash A_i\big) + \mu(A_i \backslash K_i) \\
&= \mu(\cup_{k \neq i} A_k \backslash \cup_{j \neq i} K_j) + \mu(A_i \backslash K_i) \\
&\leq \mu(\cup_{j \neq i} A_j \backslash K_j) + \mu(A_i \backslash K_i) \\
&= \sum_{j=1}^n \mu(A_j \backslash K_j) < n \cdot \frac{\varepsilon}{n} = \varepsilon.
\end{aligned}$$ □

From the above result, we will show that the conditional entropy of a partition α given a partition β can be made as small as desired provided that β has a small enough diameter. Indeed, from Theorem 9.4.3(e), given partitions α, β and β' as in the above lemma, we have that $\mathrm{H}_\mu(\alpha|\beta) \leq \mathrm{H}_\mu(\alpha|\beta')$, where the partition β' is designed to resemble the partition α. In order to estimate the conditional entropy $\mathrm{H}_\mu(\alpha|\beta')$, we must estimate the contribution of all atoms of the partition $\alpha \vee \beta'$. There are essentially two kinds of atoms to be taken into account, namely, atoms of the form $A_i \cap B_i'$ and atoms of the form $A_i \cap B_j'$ with $i \neq j$. Intuitively, because A_i is more or less equal to B_i' (after all, $\mu(A_i \triangle B_i')$ is small), the information provided by A_i assuming that measurement β' resulted in B_i' is small. On the other hand, since A_i is nearly disjoint from B_j' when $i \neq j$ (given that A_i is close to B_i' and $B_i' \cap B_j' = \emptyset$), the information obtained from getting A_i given that observation B_j' occurred is also small. This is what we now prove rigorously. The proof will make use of the function k from Definition 9.3.4.

Lemma 9.4.15. *Let X be a compact metric space and $\mu \in M(X)$. Let also $\alpha \in \mathrm{Part}_{\mathrm{Fin}}(X)$. For every $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\beta \in \mathrm{Part}(X),\ \mathrm{diam}(\beta) < \delta \Longrightarrow \mathrm{H}_\mu(\alpha|\beta) < \varepsilon.$$

Proof. Let $\alpha = \{A_1, A_2, \ldots, A_n\}$ be a finite Borel partition of X. As atoms of measure zero do not affect conditional entropy, we may assume that $\mu(A_i) > 0$ for all $1 \le i \le n$. Fix $\varepsilon > 0$ and let $0 < \overline{\varepsilon} < e^{-1}$ be so small that

$$\max\{k(\overline{\varepsilon}), k(1-\overline{\varepsilon})\} < \frac{\varepsilon}{2n}.$$

Then there exists $\widehat{\varepsilon} > 0$ such that

$$0 < \frac{\widehat{\varepsilon}}{\mu(A_i) - \widehat{\varepsilon}} < \overline{\varepsilon} \quad \text{and} \quad \frac{\mu(A_i) - \widehat{\varepsilon}}{\mu(A_i) + \widehat{\varepsilon}} > 1 - \overline{\varepsilon} \tag{9.3}$$

for all $1 \le i \le n$. In particular, the left relation in (9.3) imposes that $\mu(A_i) > \widehat{\varepsilon}$ for all i. Let $\delta > 0$ be the number ascribed to $\widehat{\varepsilon}$ in Lemma 9.4.14. Let β be a partition with $\mathrm{diam}(\beta) < \delta$, and let $\beta' = \{B_1', B_2', \ldots, B_n'\} \le \beta$ be such that $\mu(A_i \triangle B_i') \le \widehat{\varepsilon}$ for all $1 \le i \le n$, also as prescribed in Lemma 9.4.14. Since $\mu(A_i) > \widehat{\varepsilon}$ for all i, this implies that $\mu(B_i') > 0$ for all i. Moreover,

$$\left|\mu(A_i) - \mu(B_i')\right| \le \mu(A_i \triangle B_i') \le \widehat{\varepsilon}. \tag{9.4}$$

Therefore,

$$0 < \mu(A_i) - \widehat{\varepsilon} \le \mu(A_i) - \mu(A_i \triangle B_i') \le \mu(B_i') \le \mu(A_i) + \mu(A_i \triangle B_i') \le \mu(A_i) + \widehat{\varepsilon}. \tag{9.5}$$

Hence,

$$\mu(A_i \cap B_i') = \mu(A_i) - \mu(A_i \backslash B_i') \ge \mu(A_i) - \mu(A_i \triangle B_i') \ge \mu(A_i) - \widehat{\varepsilon} > 0.$$

Using this, (9.5) and (9.3), we get

$$\frac{\mu(A_i \cap B_i')}{\mu(B_i')} \ge \frac{\mu(A_i) - \widehat{\varepsilon}}{\mu(A_i) + \widehat{\varepsilon}} > 1 - \overline{\varepsilon}.$$

By choice of $\overline{\varepsilon}$, the function k is decreasing on the interval $[1-\overline{\varepsilon}, 1] \subseteq [1-e^{-1}, 1] \subseteq [e^{-1}, 1]$, and thus

$$k\left(\frac{\mu(A_i \cap B_i')}{\mu(B_i')}\right) \le k(1-\overline{\varepsilon}) < \frac{\varepsilon}{2n} \tag{9.6}$$

for all i.

Now, suppose that $i \neq j$. Since $\alpha = \{A_k\}_{k=1}^n$ is a partition of X, we know that $A_i \cap B_j' \subseteq B_j' \backslash A_j \subseteq A_j \triangle B_j'$. Using this, (9.5), (9.4) and (9.3), we infer that

$$\frac{\mu(A_i \cap B_j')}{\mu(B_j')} \leq \frac{\mu(A_j \triangle B_j')}{\mu(A_j) - \mu(A_j \triangle B_j')} \leq \frac{\widehat{\varepsilon}}{\mu(A_j) - \widehat{\varepsilon}} < \overline{\varepsilon}.$$

By choice of $\overline{\varepsilon}$, the function k is increasing on the interval $[0, \overline{\varepsilon}] \subseteq [0, e^{-1}]$, and hence

$$k\left(\frac{\mu(A_i \cap B_j')}{\mu(B_j')}\right) \leq k(\overline{\varepsilon}) < \frac{\varepsilon}{2n} \tag{9.7}$$

for all $i \neq j$. Then, by Theorem 9.4.3(e) and (9.6)–(9.7), we have

$$\begin{aligned}
\mathrm{H}_\mu(\alpha|\beta) \leq \mathrm{H}_\mu(\alpha|\beta') &= \sum_{A\in\alpha}\sum_{B'\in\beta'} -\mu(A\cap B')\log\frac{\mu(A\cap B')}{\mu(B')} \\
&= \sum_{i,j=1}^{n} \mu(B_j')k\left(\frac{\mu(A_i \cap B_j')}{\mu(B_j')}\right) \\
&= \sum_{i=1}^{n} \mu(B_i')k\left(\frac{\mu(A_i \cap B_i')}{\mu(B_i')}\right) + \sum_{\substack{i,j=1\\ i\neq j}}^{n} \mu(B_j')k\left(\frac{\mu(A_i \cap B_j')}{\mu(B_j')}\right) \\
&< \sum_{i=1}^{n} \mu(B_i')\frac{\varepsilon}{2n} + \sum_{i=1}^{n}\sum_{j=1}^{n} \mu(B_j')\frac{\varepsilon}{2n} = \frac{\varepsilon}{2n} + \sum_{i=1}^{n}\frac{\varepsilon}{2n} = \frac{\varepsilon}{2n} + n\cdot\frac{\varepsilon}{2n} \\
&\leq \varepsilon. \qquad \square
\end{aligned}$$

From the above lemma, we can infer that any sequence of partitions whose diameters tend to 0 provide asymptotically as much information as any given finite partition can.

Corollary 9.4.16. *Let X be a compact metric space and $\mu \in M(X)$. Let also $(\alpha_n)_{n=1}^\infty$ be a sequence in* $\mathrm{Part}(X)$ *such that* $\lim_{n\to\infty} \mathrm{diam}(\alpha_n) = 0$. *Then*

$$\lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\alpha_n) = 0$$

for every $\alpha \in \mathrm{Part}_{\mathrm{Fin}}(X)$.

Proof. Let α be a finite Borel partition of X. By Lemma 9.4.15, for every $\varepsilon > 0$ there exists a $\delta > 0$ such that if $\mathrm{diam}(\beta) < \delta$ then $\mathrm{H}_\mu(\alpha|\beta) < \varepsilon$. Since $\mathrm{diam}(\alpha_n) \to 0$, it follows that $\mathrm{H}_\mu(\alpha|\alpha_n) \to 0$ as $n \to \infty$. □

The above corollary about conditional entropy of partitions allows us to deduce the following result on the measure-theoretic entropy of a system. This is the counterpart of Lemma 7.2.20.

Theorem 9.4.17. *Let X be a compact metric space and $\mu \in M(X)$. Let also $T : X \to X$ be a measure-preserving dynamical system on $(X, \mathcal{B}(X), \mu)$ and $(\alpha_n)_{n=1}^{\infty}$ be a sequence in* $\mathrm{Part}_{\mathrm{Fin}}(X)$ *such that* $\lim_{n\to\infty} \mathrm{diam}(\alpha_n) = 0$. *Then*

$$\mathrm{h}_\mu(T) = \lim_{n\to\infty} \mathrm{h}_\mu(T, \alpha_n).$$

Proof. Let α be a finite partition of X consisting of Borel sets. By Corollary 9.4.16, we know that $\lim_{n\to\infty} \mathrm{H}_\mu(\alpha|\alpha_n) = 0$. By Theorem 9.4.11(j), it follows that

$$\mathrm{h}_\mu(T, \alpha) \le \liminf_{n\to\infty} \mathrm{h}_\mu(T, \alpha_n) \le \limsup_{n\to\infty} \mathrm{h}_\mu(T, \alpha_n) \le \mathrm{h}_\mu(T).$$

Since this is true for any finite Borel partition α, we deduce from a passage to the supremum that

$$\mathrm{h}_\mu(T) \le \liminf_{n\to\infty} \mathrm{h}_\mu(T, \alpha_n) \le \limsup_{n\to\infty} \mathrm{h}_\mu(T, \alpha_n) \le \mathrm{h}_\mu(T).$$

Hence, $\mathrm{h}_\mu(T) = \lim_{n\to\infty} \mathrm{h}_\mu(T, \alpha_n)$. □

We can easily deduce a counterpart to Lemma 7.2.22.

Corollary 9.4.18. *Let X be a compact metric space and $\mu \in M(X)$. Let also $T : X \to X$ be a measure-preserving dynamical system on $(X, \mathcal{B}(X), \mu)$ and $\alpha \in$* $\mathrm{Part}_{\mathrm{Fin}}(X)$ *be such that* $\lim_{n\to\infty} \mathrm{diam}(\alpha^n) = 0$. *Then*

$$\mathrm{h}_\mu(T) = \mathrm{h}_\mu(T, \alpha).$$

Proof. By Theorems 9.4.17 and 9.4.11(g), we have that

$$\mathrm{h}_\mu(T) = \lim_{n\to\infty} \mathrm{h}_\mu(T, \alpha^n) = \lim_{n\to\infty} \mathrm{h}_\mu(T, \alpha) = \mathrm{h}_\mu(T, \alpha).$$ □

Partitions α such that $\lim_{n\to\infty} \mathrm{diam}(\alpha^n) = 0$ allow us (when they exist) to find the entropy of a transformation by simply computing the entropy of the transformation with respect to one such partition. As in Definition 7.2.21, we give them a special name.

Definition 9.4.19. Let X be a compact metric space and $\mu \in M(X)$. Let also $T : X \to X$ be a measure-preserving dynamical system on $(X, \mathcal{B}(X), \mu)$. Any $\alpha \in \mathrm{Part}_{\mathrm{Fin}}(X)$ such that $\lim_{n\to\infty} \mathrm{diam}(\alpha^n) = 0$ is called a *generator* for T.

We have already seen in Lemma 7.2.23 that expansive dynamical systems admit generators.

Theorem 9.4.20. *Let $T : X \to X$ be an expansive dynamical system preserving a Borel probability measure μ. If $\alpha \in$* $\mathrm{Part}_{\mathrm{Fin}}(X)$ *satisfies* $\mathrm{diam}(\alpha) < \delta(T)$, *where $\delta(T)$ is an expansive constant for T, then α is a generator for T and* $\mathrm{h}_\mu(T) = \mathrm{h}_\mu(T, \alpha)$.

Proof. Let $\alpha = \{A_k\}_{k=1}^m$ be a finite Borel partition with $\operatorname{diam}(\alpha) < \delta(T)$. Define $\delta = (\delta(T) - \operatorname{diam}(\alpha))/2$. The finite open cover $\widetilde{\alpha} = \{B(A_k, \delta)\}_{k=1}^m$ has diameter $\operatorname{diam}(\widetilde{\alpha}) \leq \delta(T)$. Lemma 7.2.23 asserts that $\lim_{n\to\infty} \operatorname{diam}(\widetilde{\alpha}^n) = 0$. As $\operatorname{diam}(\alpha^n) \leq \operatorname{diam}(\widetilde{\alpha}^n)$ for all $n \in \mathbb{N}$, it ensues that $\lim_{n\to\infty} \operatorname{diam}(\alpha^n) = 0$ and the result thus follows from Corollary 9.4.18. □

Let us now give some examples, all but the first one of which are applications of Theorem 9.4.17, Corollary 9.4.18 and/or Theorem 9.4.20.

Example 9.4.21. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. If there exists a finite measurable set $Y \subseteq X$ of full measure, then $\mathrm{h}_\mu(T) = 0$. Indeed, for any $\beta \in \operatorname{Part}(X, \mathcal{A})$ we have

$$\mathrm{H}_\mu(\beta) = -\sum_{B\in\beta} \mu(B) \log \mu(B) = -\sum_{B\in\beta} \mu(B \cap Y) \log \mu(B \cap Y).$$

This means that the entropy of a partition of X is equal to the entropy of the projection of that partition onto Y. In other words, the entropy of a partition of X coincides with the entropy of a partition of Y. Since Y is finite, there are only finitely many such partitions. Therefore, $\mathrm{H}_\mu(\beta)$ can only take finitely many values. Consequently, the entropies $\mathrm{H}_\mu(\beta^n)$, $n \in \mathbb{N}$, can also only take finitely many values. Thus

$$\mathrm{h}_\mu(T, \beta) = \lim_{n\to\infty} \frac{1}{n} \mathrm{H}_\mu(\beta^n) = 0.$$

Since β was arbitrary, we conclude that

$$\mathrm{h}_\mu(T) = \sup\{\mathrm{h}_\mu(T, \beta) : \beta \in \operatorname{Part}_{\mathrm{Fin}}(X, \mathcal{A})\} = 0.$$

Example 9.4.22. The entropy of any homeomorphism $T : \mathbb{S}^1 \to \mathbb{S}^1$ of the unit circle $\mathbb{S}^1$ is equal to 0 with respect to any T-invariant Borel probability measure.

Indeed, let μ be any T-invariant Borel probability measure. Let α and β be finite partitions of $\mathbb{S}^1$ into intervals. Then $\alpha \vee \beta$ is a partition of $\mathbb{S}^1$ into at most $(\#\alpha + \#\beta)$ intervals since $\#(\alpha \vee \beta)$ is equal to the number of endpoints of the intervals in $\alpha \vee \beta$, which is bounded above by the sum of the number of endpoints of the intervals in α and the number of endpoints of the intervals in β. Moreover, since T is a homeomorphism, we know that $T^{-k}\alpha$ is a partition of $\mathbb{S}^1$ into $\#\alpha$ intervals for every $k \in \mathbb{N}$. Therefore, α^n is a partition of $\mathbb{S}^1$ into at most $\#(\alpha^n) \leq n \cdot \#\alpha$ intervals. Consequently,

$$0 \leq \mathrm{H}_\mu(\alpha^n) \leq \log \#(\alpha^n) \leq \log n + \log \#\alpha.$$

We deduce that

$$0 \leq \mathrm{h}_\mu(T, \alpha) = \lim_{n\to\infty} \frac{1}{n} \mathrm{H}_\mu(\alpha^n) \leq \lim_{n\to\infty} \frac{1}{n}(\log n + \log \#\alpha) = 0.$$

Now, for every $m \in \mathbb{N}$, let α_m be a partition of $\mathbb{S}^1$ into m intervals of equal length. Then $\lim_{m\to\infty} \operatorname{diam}(\alpha_m) = 0$ and, by the above result, $\mathrm{h}_\mu(T, \alpha_m) = 0$ for all $m \in \mathbb{N}$. It follows from Theorem 9.4.17 that

$$\mathrm{h}_\mu(T) = \lim_{m\to\infty} \mathrm{h}_\mu(T, \alpha_m) = 0.$$

Example 9.4.23. Let E be a finite set and $A : E \times E \to \{0, 1\}$ be an incidence matrix. We proved in Example 4.1.3 that the shift map $\sigma : E_A^\infty \to E_A^\infty$ is expanding, and hence is expansive. More precisely, any $0 < \delta < 1$ is an expansive constant when E_A^∞ is endowed with the metric $d_s(\omega, \tau) = s^{|\omega\wedge\tau|}$, for any $0 < s < 1$. Let

$$\alpha = \{[e] : e \in E\}$$

be the partition of E_A^∞ into its initial 1-cylinders. Then

$$\operatorname{diam}(\alpha) = s < 1.$$

If μ is any σ-invariant Borel probability measure, then

$$\mathrm{h}_\mu(\sigma) = \mathrm{h}_\mu(\sigma, \alpha)$$

according to Theorem 9.4.20.

In particular, let us consider the full E-shift. Let μ be the product measure determined by its value on the cylinder sets; in other words,

$$\mu([\omega_1\omega_2 \ldots \omega_n]) = \prod_{k=1}^{n} P(\omega_k),$$

where P is a probability measure on the σ-algebra of all subsets of E and $P(e) := P(\{e\})$. It was shown in Example 8.1.14 that μ is σ-invariant (and σ-ergodic per Example 8.2.32). Furthermore, it is possible to show by induction that

$$\mathrm{H}_\mu(\alpha^n) = -n \sum_{e\in E} P(e) \log P(e).$$

Thus

$$\mathrm{h}_\mu(\sigma) = \mathrm{h}_\mu(\sigma, \alpha) = \lim_{n\to\infty} \frac{1}{n} \mathrm{H}_\mu(\alpha^n) = -\sum_{e\in E} P(e) \log P(e).$$

Example 9.4.24. Let $T_n : \mathbb{S}^1 \to \mathbb{S}^1$ be the n-fold map defined by $T_n(x) = nx \pmod 1$, where $\mathbb{S}^1$ is equipped with the σ-algebra of Borel sets and with the Lebesgue measure λ. We have already seen in Example 8.1.10 that T_n preserves λ (we also showed in Example 8.2.30 that T_n is ergodic with respect to λ). Consider the partition

$$\alpha = \left\{ \left[\frac{j}{n}, \frac{j+1}{n} \right) : 0 \le j < n \right\}.$$

The map T_n is expanding, and thus expansive, with any $0 < \delta < 1/n$ as expansive constant. As $\operatorname{diam}(\alpha) = 1/n$, Theorem 9.4.20 does not apply to α directly. Nevertheless, observe that

$$\alpha^k = \left\{ \left[\frac{j}{n^k}, \frac{j+1}{n^k} \right) : 0 \le j < n^k \right\}$$

for all $k \in \mathbb{N}$. Thus $\operatorname{diam}(\alpha^k) = 1/n^k < 1/n$ for all $k \geq 2$. Using Theorems 9.4.20 and 9.4.11(g), we deduce that

$$\mathrm{h}_\lambda(T_n) = \mathrm{h}_\lambda(T_n, \alpha^k) = \mathrm{h}_\lambda(T_n, \alpha).$$

Moreover,

$$\begin{aligned}
\mathrm{H}_\lambda(\alpha^k) &= \sum_{I \in \alpha^k} -\lambda(I) \log \lambda(I) \\
&= \sum_{j=0}^{n^k-1} -\lambda\left(\left[\frac{j}{n^k}, \frac{j+1}{n^k} \right) \right) \log \lambda\left(\left[\frac{j}{n^k}, \frac{j+1}{n^k} \right) \right) \\
&= \sum_{j=0}^{n^k-1} -\frac{1}{n^k} \log \frac{1}{n^k} = -\log \frac{1}{n^k} = k \log n.
\end{aligned}$$

Consequently,

$$\mathrm{h}_\lambda(T_n) = \mathrm{h}_\lambda(T_n, \alpha) = \lim_{k \to \infty} \frac{1}{k} \mathrm{H}_\lambda(\alpha^k) = \log n.$$

9.5 Shannon–McMillan–Breiman theorem

The Shannon–McMillan–Breiman theorem is a central result in information theory and can be thought of as a sort of ergodic theorem for measure-theoretic entropy. Indeed, the proof relies heavily on Birkhoff's ergodic theorem (Theorem 8.2.11). It also uses the following result.

Lemma 9.5.1. *Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. Let $\alpha \in \operatorname{Part}(X, \mathcal{A})$. Let also*

$$f_n := I_\mu(\alpha | \alpha_1^n) \quad \textit{and} \quad f^* := \sup_{n \in \mathbb{N}} f_n.$$

Then, for all $r \in \mathbb{R}$ and all $A \in \alpha$, we have

$$\mu(\{x \in A : f^*(x) > r\}) \le \min\{\mu(A), e^{-r}\}.$$

Proof. Let $A \in \alpha$ and fix $n \in \mathbb{N}$. To shorten notation, let

$$f_n^A = -\log E(\mathbb{1}_A | \sigma(\alpha_1^n)),$$

where $\sigma(\alpha_1^n)$ is the sub-σ-algebra generated by the countable partition α_1^n. Fix $x \in A$. Then, using Example A.1.62 in Appendix A, we get

$$\begin{aligned} f_n^A(x) &= -\log E(\mathbb{1}_A | \sigma(\alpha_1^n))(x) = -\log\Big[\frac{1}{\mu(\alpha_1^n(x))} \int_{\alpha_1^n(x)} \mathbb{1}_A \, d\mu\Big] \\ &= -\log \frac{\mu(A \cap \alpha_1^n(x))}{\mu(\alpha_1^n(x))} = -\log \frac{\mu(\alpha(x) \cap \alpha_1^n(x))}{\mu(\alpha_1^n(x))} = I_\mu(\alpha|\alpha_1^n)(x) \\ &= f_n(x) \end{aligned}$$

for all $x \in A$. Hence,

$$f_n = \sum_{A \in \alpha} \mathbb{1}_A \cdot f_n^A.$$

Now, for $n \in \mathbb{N}$ and $r \in \mathbb{R}$ consider the set

$$B_n^{A,r} = \Big\{x \in X : \max_{1 \le i < n} f_i^A(x) \le r \text{ while } f_n^A(x) > r\Big\}.$$

The family $\{B_n^{A,r}\}_{n=1}^\infty$ consists of mutually disjoint sets. Also, recall that $\alpha_1^n \le \alpha_1^{n+1}$, and thus $\sigma(\alpha_1^n) \subseteq \sigma(\alpha_1^{n+1})$ for each $n \in \mathbb{N}$. By definition, each f_n^A is measurable with respect to $\sigma(\alpha_1^n)$. Consequently, $B_n^{A,r} \in \sigma(\alpha_1^n)$. Then

$$\begin{aligned} \mu(A \cap B_n^{A,r}) &= \int_{B_n^{A,r}} \mathbb{1}_A \, d\mu = \int_{B_n^{A,r}} E(\mathbb{1}_A | \sigma(\alpha_1^n)) \, d\mu \\ &= \int_{B_n^{A,r}} \exp(-f_n^A) \, d\mu \le \int_{B_n^{A,r}} e^{-r} \, d\mu = e^{-r} \mu(B_n^{A,r}). \end{aligned}$$

But

$$\begin{aligned} \{x \in A : f^*(x) > r\} &= \{x \in A : \exists\, n \in \mathbb{N} \text{ such that } f_n(x) > r\} \\ &= \{x \in A : \exists\, n \in \mathbb{N} \text{ such that } f_n^A(x) > r\} = A \cap \bigcup_{n=1}^\infty B_n^{A,r}. \end{aligned}$$

Using the disjointness of the $B_n^{A,r}$'s, it ensues that

$$\mu(\{x \in A : f^*(x) > r\}) = \sum_{n=1}^\infty \mu(A \cap B_n^{A,r}) \le \sum_{n=1}^\infty e^{-r} \mu(B_n^{A,r}) = e^{-r} \mu\Big(\bigcup_{n=1}^\infty B_n^{A,r}\Big) \le e^{-r}. \quad \square$$

Corollary 9.5.2. *In addition to the hypotheses of Lemma* 9.5.1, *assume that* $\mathrm{H}_\mu(\alpha) < \infty$. *Then* $f^* \in L^1(X, \mathcal{A}, \mu)$ *and* $\|f^*\|_1 \le \mathrm{H}_\mu(\alpha) + 1$.

Proof. Since $f^* \ge 0$, we have $\int_X |f^*| \, d\mu = \int_X f^* \, d\mu$. Using Lemmas 9.5.1 and A.1.37, we obtain

$$\begin{aligned} \|f^*\|_1 &= \int_X f^* \, d\mu = \sum_{A \in \alpha} \int_A f^* \, d\mu \\ &= \sum_{A \in \alpha} \int_0^\infty \mu(\{x \in A : f^*(x) > r\}) \, dr \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{A\in\alpha}\int_0^\infty \min\{\mu(A), e^{-r}\}\,dr \\
&= \sum_{A\in\alpha}\left(\int_0^{-\log\mu(A)} \mu(A)\,dr + \int_{-\log\mu(A)}^{\infty} e^{-r}\,dr\right) \\
&= \sum_{A\in\alpha}\left(-\mu(A)\log\mu(A) + \left[-e^{-r}\right]_{-\log\mu(A)}^{\infty}\right) \\
&= \sum_{A\in\alpha} -\mu(A)\log\mu(A) + \sum_{A\in\alpha}\mu(A) \\
&= \mathrm{H}_\mu(\alpha) + 1 < \infty. \qquad \square
\end{aligned}$$

Corollary 9.5.3. *Under the same hypotheses as Corollary* 9.5.2, *the sequence* $(f_n)_{n=1}^\infty$ *converges* μ*-a. e. and in* $L^1(X,\mathcal{A},\mu)$.

Proof. Recall that $\alpha_1^n \leq \alpha_1^{n+1}$, and thus $\sigma(\alpha_1^n) \subseteq \sigma(\alpha_1^{n+1})$ for each $n \in \mathbb{N}$. For any $x \in A \in \alpha$, we have $f_n(x) = f_n^A(x) = -\log E(\mathbb{1}_A|\sigma(\alpha_1^n))(x)$ and Doob's martingale convergence theorem for conditional expectations (Theorem A.1.67) guarantees that $\lim_{n\to\infty} E(\mathbb{1}_A|\sigma(\alpha_1^n))$ exists μ-almost everywhere. Hence, the sequence of nonnegative functions $(f_n)_{n=1}^\infty$ converges μ-a. e. to some limit function $g \geq 0$. Since $|f_n| = f_n \leq f^*$ for all n, we have $|g| = g \leq f^*$, and thus $|f_n - g| \leq 2f^*$ μ-almost everywhere. Applying Lebesgue's dominated convergence theorem (Theorem A.1.38) to the sequence $(|f_n - g|)_{n=1}^\infty$, we obtain

$$\lim_{n\to\infty}\|f_n - g\|_1 = \lim_{n\to\infty}\int_X |f_n - g|\,d\mu = \int_X \lim_{n\to\infty}|f_n - g|\,d\mu = 0.$$

In other words, $f_n \to g$ in $L^1(X,\mathcal{A},\mu)$. $\square$

We are finally in a position to prove the main result of this section and chapter.

Theorem 9.5.4 (Shannon–McMillan–Breiman theorem). *Let* $T : X \to X$ *be a measure-preserving dynamical system on a probability space* $(X,\mathcal{A},\mu)$ *and let* $\alpha \in \mathrm{Part}(X,\mathcal{A})$ *be such that* $\mathrm{H}_\mu(\alpha) < \infty$. *Then the following limits exist:*

$$f := \lim_{n\to\infty} I_\mu(\alpha|\alpha_1^n) \quad \textit{and} \quad \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} f\circ T^j = E(f|\mathcal{I}_\mu) \quad \mu\textit{-a. e.},$$

where $\mathcal{I}_\mu$ *is the sub-σ-algebra of all μ-almost T-invariant sets (see Definition* 8.2.5*).*

Moreover, the following statements hold:

(a) $\lim_{n\to\infty}\frac{1}{n}I_\mu(\alpha^n) = E(f|\mathcal{I}_\mu)$ *μ-a. e. and in* $L^1(\mu)$.

(b) $\mathrm{h}_\mu(T,\alpha) = \lim_{n\to\infty}\frac{1}{n}\mathrm{H}_\mu(\alpha^n) = \int_X E(f|\mathcal{I}_\mu)\,d\mu = \int_X f\,d\mu$.

Proof. According to Corollary 9.5.3, the first sequence of functions $(f_n)_{n=1}^{\infty} = (I_\mu(\alpha|\alpha_1^n))_{n=1}^{\infty}$ converges μ-a. e. to an integrable function f. The second limit exists by virtue of Birkhoff's ergodic theorem (Theorem 8.2.11). Note also that all functions $(f_n)_{n=1}^{\infty}$ are nonnegative, and thus so are f and $E(f|\mathcal{I}_\mu)$. To prove the remaining two statements, let us first assume that (a) holds and derive (b) from it. Then we will prove (a).

Let us assume that (a) holds. Using Scheffé's lemma (Lemma A.1.39) and the fact that $(I_\mu(\alpha^n))_{n=1}^{\infty}$ and $E(f|\mathcal{I}_\mu)$ are nonnegative, we obtain that

$$\mathrm{h}_\mu(T,\alpha) = \lim_{n\to\infty} \frac{1}{n}\mathrm{H}_\mu(\alpha^n) = \lim_{n\to\infty} \frac{1}{n}\int_X I_\mu(\alpha^n)\,d\mu = \int_X E(f|\mathcal{I}_\mu)\,d\mu = \int_X f\,d\mu.$$

This establishes (b).

In order to prove (a), first notice that by Lemma 9.4.6 we have

$$I_\mu(\alpha^n) = \sum_{k=1}^{n} I_\mu(\alpha|\alpha_1^k)\circ T^{n-k} = \sum_{j=0}^{n-1} I_\mu(\alpha|\alpha_1^{n-j})\circ T^j = \sum_{j=0}^{n-1} f_{n-j}\circ T^j.$$

Then, by the triangle inequality,

$$\begin{aligned}
\left|\frac{1}{n}I_\mu(\alpha^n) - E(f|\mathcal{I}_\mu)\right| &= \left|\frac{1}{n}\sum_{j=0}^{n-1}(f_{n-j}\circ T^j - f\circ T^j) + \frac{1}{n}\sum_{j=0}^{n-1} f\circ T^j - E(f|\mathcal{I}_\mu)\right| \\
&\le \left|\frac{1}{n}\sum_{j=0}^{n-1}(f_{n-j} - f)\circ T^j\right| + \left|\frac{1}{n}S_n f - E(f|\mathcal{I}_\mu)\right| \\
&\le \frac{1}{n}\sum_{j=0}^{n-1}\left|f_{n-j} - f\right|\circ T^j + \left|\frac{1}{n}S_n f - E(f|\mathcal{I}_\mu)\right|. \qquad (9.8)
\end{aligned}$$

Birkhoff's ergodic theorem (Theorem 8.2.11) asserts that the second term on the right-hand side tends to 0 in $L^1(\mu)$. Let us now investigate the first term on that right-hand side. Set $g_n = |f_n - f|$. Since $(f_n)_{n=1}^{\infty}$ converges to f in $L^1(\mu)$ according to Corollary 9.5.3, the sequence $(g_n)_{n=1}^{\infty}$ converges to 0 in $L^1(\mu)$. So do its Cesàro averages $(\frac{1}{n}\sum_{i=1}^{n} g_i)_{n=1}^{\infty}$. Then

$$\begin{aligned}
\lim_{n\to\infty}\left\|\frac{1}{n}\sum_{j=0}^{n-1} g_{n-j}\circ T^j\right\|_1 &= \lim_{n\to\infty}\int_X \left|\frac{1}{n}\sum_{j=0}^{n-1} g_{n-j}\circ T^j\right| d\mu \\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\int_X g_{n-j}\circ T^j\,d\mu \\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\int_X g_{n-j}\,d\mu
\end{aligned}$$

$$= \lim_{n\to\infty} \int_X \frac{1}{n} \sum_{i=1}^{n} g_i \, d\mu$$
$$= 0.$$

That is, the functions $\frac{1}{n}\sum_{j=0}^{n-1} g_{n-j} \circ T^j$ converge to 0 in $L^1(\mu)$. Thus the first term on the right-hand side of (9.8), like the second term, converges to 0 in $L^1(\mu)$. This implies that the sequence $(\frac{1}{n}I_\mu(\alpha^n))_{n=1}^{\infty}$ converges to $E(f|\mathcal{I}_\mu)$ in $L^1(\mu)$.

It only remains to show convergence μ-a. e. of that same sequence. To this end, for each $N \in \mathbb{N}$ let $G_N = \sup_{n\geq N} g_n$. The sequence of functions $(G_N)_{N=1}^{\infty}$ is decreasing and bounded below by 0, so it converges to some function. As $f_n \to f$ μ-a. e., we know that $g_n = |f_n - f| \to 0$ μ-almost everywhere. It follows that $G_N \searrow 0$ μ-almost everywhere. Also, the functions G_N are uniformly bounded above by an integrable function since

$$0 \leq G_N \leq G_1 = \sup_{n\in\mathbb{N}} g_n \leq \sup_{n\in\mathbb{N}}(|f_n| + |f|) \leq f^* + f \in L^1(\mu),$$

where $f^*, f \in L^1(\mu)$ according to Corollaries 9.5.2–9.5.3. So $G_N \in L^1(\mu)$ for all $N \in \mathbb{N}$ and Lebesgue's dominated convergence theorem affirms that

$$\lim_{N\to\infty} \int_X E(G_N|\mathcal{I}_\mu) \, d\mu = \lim_{N\to\infty} \int_X G_N \, d\mu = \int_X \lim_{N\to\infty} G_N \, d\mu = 0.$$

Moreover, according to Proposition A.1.60, since $(G_N)_{N=1}^{\infty}$ is decreasing and bounded below by 0, so is the sequence of conditional expectations $(E(G_N|\mathcal{I}_\mu))_{N=1}^{\infty}$ μ-almost everywhere. Summarizing, we have established that $E(G_N|\mathcal{I}_\mu) \searrow$ μ-a. e., $E(G_N|\mathcal{I}_\mu) \geq 0$ μ-a. e. and $\int_X E(G_N|\mathcal{I}_\mu)\, d\mu \searrow 0$ as $N \to \infty$. It ensues that $E(G_N|\mathcal{I}_\mu) \searrow 0$ μ-a. e. as $N \to \infty$.

Fix temporarily $N \in \mathbb{N}$. Then for any $n > N$, we have

$$\begin{aligned}\frac{1}{n}\sum_{j=0}^{n-1} g_{n-j} \circ T^j &= \frac{1}{n}\sum_{j=0}^{n-N} g_{n-j} \circ T^j + \frac{1}{n}\sum_{j=n-N+1}^{n-1} g_{n-j} \circ T^j \\ &\leq \frac{n-N}{n} \cdot \frac{1}{n-N}\sum_{j=0}^{n-N} G_N \circ T^j + \frac{1}{n}\sum_{j=n-N+1}^{n-1} G_1 \circ T^j.\end{aligned}$$

Let $F_N = \sum_{j=0}^{N-2} G_1 \circ T^j$. Using Birkhoff's ergodic theorem (Theorem 8.2.11), we deduce that

$$\begin{aligned}0 \leq \limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} g_{n-j} \circ T^j &\leq \lim_{n\to\infty} \frac{1}{n-N}\sum_{j=0}^{n-N} G_N \circ T^j + \limsup_{n\to\infty} \frac{1}{n} F_N \circ T^{n-N+1} \\ &= E(G_N|\mathcal{I}_\mu) + \limsup_{n\to\infty} \frac{1}{n} F_N \circ T^{n-N+1} \quad \mu\text{-a. e.} \\ &= E(G_N|\mathcal{I}_\mu) \quad \mu\text{-a. e..}\end{aligned}$$

Since N was chosen arbitrarily and we have showed earlier that $E(G_N|\mathcal{I}_\mu) \to 0$ μ-a. e. as $N \to \infty$, we conclude that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} g_{n-j} \circ T^j = 0 \quad \mu\text{-a. e..}$$

This establishes the μ-a. e. convergence of the first term on the right-hand side of (9.8). The μ-a. e. convergence of the second term on that right-hand side follows from Birkhoff's ergodic theorem (Theorem 8.2.11). Therefore, the sequence $(\frac{1}{n} I_\mu(\alpha^n))_{n=1}^\infty$ converges to $E(f|\mathcal{I}_\mu)$ μ-almost everywhere. □

Corollary 9.5.5 (Ergodic case of Shannon–McMillan–Breiman theorem). *Let $T : X \to X$ be an ergodic measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$ and let $\alpha \in \mathrm{Part}(X, \mathcal{A})$ be such that $\mathrm{H}_\mu(\alpha) < \infty$. Then*

$$\mathrm{h}_\mu(T, \alpha) = \lim_{n\to\infty} \frac{1}{n} I_\mu(\alpha^n)(x) \quad \textit{for } \mu\textit{-a. e. } x \in X.$$

Proof. This follows immediately from Shannon–McMillan–Breiman theorem (Theorem 9.5.4) and the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14). □

When $\mathrm{h}_\mu(T, \alpha) > 0$, this corollary reveals that $\mu(\alpha^n(x))$ tends to 0 with exponential rate $e^{-\mathrm{h}_\mu(T,\alpha)}$ for μ-a. e. $x \in X$ (see Exercise 9.7.11).

The right-hand side in the above equality can be viewed as a local entropy at x. The corollary then states that at almost every x the local entropy exists and is equal to the entropy of the transformation relative to the partition. Another approach to local entropy is discussed next.

9.6 Brin–Katok local entropy formula

We now derive the celebrated Brin–Katok local entropy formula.

In preparation for this, we show that given any Borel probability measure μ there exist finite Borel partitions of arbitrarily small diameters whose atoms have negligible boundaries with respect to μ.

Lemma 9.6.1. *Let (X, d) be a compact metric space and $\mu \in M(X)$. For every $\varepsilon > 0$, there exists a finite Borel partition α of X such that $\mathrm{diam}(\alpha) < \varepsilon$ and $\mu(\partial A) = 0$ for all $A \in \alpha$.*

Proof. Let $\varepsilon > 0$ and let $\{x_1, \ldots, x_n\}$ be an $(\varepsilon/4)$-spanning set of X. For each $1 \le i \le n$, the sets $\{x \in X : d(x, x_i) = r\}$, where $\varepsilon/4 < r < \varepsilon/2$, are mutually disjoint, and thus only countably many of them may have positive μ-measure. Hence, there exists $\varepsilon/4 < t < \varepsilon/2$ such that

$$\mu(\{x \in X : d(x, x_i) = t\}) = 0, \quad \forall 1 \le i \le n. \tag{9.9}$$

Define the sets A_i, $1 \le i \le n$, inductively by

$$A_i := \{x \in X : d(x, x_i) \le t\} \backslash (\cup_{j=1}^{i-1} A_j).$$

Since $t < \varepsilon/2$, the family $\alpha := \{A_1, \dots, A_n\}$ is a Borel partition of X with diameter smaller than ε. Noting that $\partial(A\backslash B) \subseteq \partial A \cup \partial B$ and $\partial(A \cup B) \subseteq \partial A \cup \partial B$, it follows from (9.9) that $\mu(\partial A_i) = 0$ for all $1 \le i \le n$. □

We now recall the concept, frequently used in coding theory, of *Hamming metric*. Let E be a nonempty finite set and $n \in \mathbb{N}$. The Hamming metric $\rho_{E,n}^{(H)}$ on E^n is defined by

$$\rho_{E,n}^{(H)}(\omega, \tau) = \frac{1}{n} \sum_{k=1}^{n} (1 - \delta_{\omega_k \tau_k}),$$

where δ_{ab} is the Kronecker delta symbol, that is,

$$\delta_{ab} = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \ne b. \end{cases}$$

Equivalently,

$$\rho_{E,n}^{(H)}(\omega, \tau) = \frac{1}{n} \#\{1 \le k \le n : \omega_k \ne \tau_k\}. \tag{9.10}$$

It is well known and a straightforward exercise to check that $\rho_{E,n}^{(H)}$ is a metric on E^n. Given $\omega \in E^n$ and $r \ge 0$, we naturally denote by $B_{E,n}^{(H)}(\omega, r)$ the open ball, in the Hamming metric $\rho_{E,n}^{(H)}$, centered at ω and of radius r. Formally,

$$B_{E,n}^{(H)}(\omega, r) = \{\tau \in E^n : \rho_{E,n}^{(H)}(\omega, \tau) < r\}.$$

Standard combinatorial considerations show that the number of elements in the ball $B_{E,n}^{(H)}(\omega, r)$ depends only on $\#E$, n, and r, and is equal to

$$\#B_{E,n}^{(H)}(\omega, r) = \sum_{k=0}^{[rn]} (\#E - 1)^k \binom{n}{k}.$$

As Katok writes in [31], using this and Stirling's formula, it is easy to verify that for every $r \in (0, \frac{\#E-1}{\#E})$ we have

$$\lim_{n\to\infty} \frac{1}{n} \log \#B_{E,n}^{(H)}(\omega, r) = r \log(\#E - 1) - r \log r - (1 - r) \log(1 - r) =: g(r). \tag{9.11}$$

Observe that

$$\lim_{r\to 0} g(r) = 0, \tag{9.12}$$

and thus for every $r \in (0, \frac{\#E-1}{\#E})$ there is $N(r) \in \mathbb{N}$ such that

$$\#B_{E,n}^{(H)}(\omega, r) \le e^{g(r)n}, \quad \forall n \ge N(r). \tag{9.13}$$

Returning to dynamics, let (X,d) be a metric space, let $T : X \to X$ be a Borel measurable self-transformation and let μ be a Borel probability measure on X. Let also α be a Borel partition of X.

As the symbol α^n might be interpreted in two different ways in the proof of the forthcoming Theorem 9.6.2, we introduce further notation. As before, the nth refined partition of α with respect to the map T will be denoted by $\alpha^n := \bigvee_{i=0}^{n-1} T^{-i}\alpha$. The n-folded Cartesian product $\alpha \times \alpha \ldots \times \alpha$ will be denoted by $\widehat{\alpha}^n$.

In the proof of Theorem 9.6.2, we will work with the Hamming metrics $\rho_{E,n}^{(H)}$ on the sets $\widehat{\alpha}^n$, $n \in \mathbb{N}$. We introduce two mappings.

First, we define the map

$$\alpha^n \ni A \longmapsto \widehat{A} \in \widehat{\alpha}^n$$

as follows. Given that α is a partition, every $A \in \alpha^n$ is uniquely represented as

$$A = \bigcap_{i=0}^{n-1} T^{-i}(A_i),$$

where $A_i \in \alpha$ for all $0 \le i < n$. We naturally set

$$\widehat{A} := (A_0, A_1, \ldots, A_{n-1}) \in \widehat{\alpha}^n$$

and we note that the map $\alpha^n \ni A \longmapsto \widehat{A} \in \widehat{\alpha}^n$ is one-to-one.

Second, we define the map

$$\widehat{\alpha}^n \ni A = (A_1, A_2, \ldots, A_n) \longmapsto \check{A} \in \alpha^n \cup \{\emptyset\}$$

by the formula

$$\check{A} := \bigcap_{i=0}^{n-1} T^{-i}(A_{i+1})$$

and we note that the map $\widehat{\alpha}^n \ni A \longmapsto \check{A} \in \alpha^n \cup \{\emptyset\}$ is one-to-one on ${}^{\vee -1}(\alpha^n) = \{A \in \widehat{\alpha}^n : \check{A} \ne \emptyset\}$ and by restricting the first mapping to that set and the second mapping to the image of that set, the two restricted mappings are inverse of one another.

Introducing more notation, we denote

$$\widehat{G} := \{\widehat{g} \mid g \in G\} \quad \text{and} \quad \check{H} := \{\check{h} \mid h \in H\}$$

for all $G \subseteq \alpha^n$ and $H \subseteq \widehat{\alpha}^n$. We abbreviate

$$\check{B}_{\alpha,n}^{(H)}(A, r) := \widecheck{B_{\alpha,n}^{(H)}(A, r)} \quad \text{and} \quad \widehat{\alpha}^n(x) := \widehat{\alpha^n(x)}$$

for every $A \in \widehat{\alpha}^n$, $r > 0$ and $x \in X$. Finally, we denote

$$\cup\beta := \bigcup_{B\in\beta} B$$

for every $\beta \subseteq \alpha^n$.

We now present and prove the ergodic version of Brin and Katok's formula from [12]. The general (i. e. nonergodic) case is considerably more complicated and rarely needed in applications. However, unlike [12], we do not assume that the map $T : X \to X$ is continuous but merely that it is Borel measurable. The proof we provide is motivated by Pesin's relevant considerations in [56].

Theorem 9.6.2 (Brin–Katok local entropy formula). *Let (X, d) be a compact metric space and let $T : X \to X$ be a Borel measurable map. If μ is an ergodic T-invariant Borel probability measure on X, then for μ-almost every $x \in X$ we have*

$$\mathrm{h}_\mu(T) = \lim_{\delta\to 0}\liminf_{n\to\infty}\frac{-\log\mu(B_n(x,\delta))}{n} = \lim_{\delta\to 0}\limsup_{n\to\infty}\frac{-\log\mu(B_n(x,\delta))}{n},$$

where $B_n(x,\delta)$ is the dynamical (n,δ)-ball at x (see Section 7.3).

Proof. It suffices to prove that for μ-almost every $x \in X$,

$$\mathrm{h}_\mu(T) \le \lim_{\delta\to 0}\liminf_{n\to\infty}\frac{-\log\mu(B_n(x,\delta))}{n} \le \lim_{\delta\to 0}\limsup_{n\to\infty}\frac{-\log\mu(B_n(x,\delta))}{n} \le \mathrm{h}_\mu(T). \quad (9.14)$$

The middle inequality being obvious, we start with the right inequality, as it is simpler to establish.

Temporarily fix $\delta > 0$. Since X is a compact metric space, there is a finite Borel partition α_δ of X such that $\mathrm{diam}(\alpha_\delta) < \delta$. Then $\alpha_\delta^n(x) \subseteq B_n(x,\delta)$ for all $x \in X$ and all $n \in \mathbb{N}$. By the ergodic case of Shannon–McMillan–Breiman theorem (Corollary 9.5.5), we know that there exists a Borel set $X_1(\alpha_\delta)$ such that

$$\mu(X_1(\alpha_\delta)) = 1 \quad (9.15)$$

and

$$\lim_{n\to\infty}\frac{-\log\mu(\alpha_\delta^n(x))}{n} = \mathrm{h}_\mu(T,\alpha_\delta), \quad \forall x \in X_1(\alpha_\delta).$$

As $\alpha^n(x) \subseteq B_n(x,\delta)$, we deduce that

$$\limsup_{n\to\infty}\frac{-\log\mu(B_n(x,\delta))}{n} \le \mathrm{h}_\mu(T,\alpha_\delta) \le \mathrm{h}_\mu(T), \quad \forall x \in X_1(\alpha_\delta). \quad (9.16)$$

It follows from this and (9.15) that the set

$$X_1 := \bigcap_{k=1}^{\infty} X_1(\alpha_{1/k})$$

satisfies

$$\mu(X_1) = 1 \tag{9.17}$$

and

$$\limsup_{n\to\infty} \frac{-\log \mu(B_n(x,1/k))}{n} \le \mathrm{h}_\mu(T), \quad \forall x \in X_1,\ \forall k \in \mathbb{N}. \tag{9.18}$$

Observing that the left-hand sides of (9.16) and (9.18) are a decreasing function of δ and an increasing function of k, respectively, we conclude that

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \frac{-\log \mu(B_n(x,\delta))}{n} = \lim_{k\to\infty} \limsup_{n\to\infty} \frac{-\log \mu(B_n(x,1/k))}{n} \le \mathrm{h}_\mu(T),\ \forall x \in X_1.$$

This is the right inequality in (9.14).

To achieve the left inequality in (9.14), temporarily fix $\varepsilon \in \left(0, \frac{\#\alpha-1}{\#\alpha}\right)$ with α a finite Borel partition of X such that

$$\mu(\partial\alpha) = 0, \tag{9.19}$$

where $\partial\alpha$ denotes the boundary of the partition α. For any $\eta > 0$, set

$$U_\eta(\alpha) := \{x \in X : B(x,\eta) \not\subseteq \alpha(x)\}.$$

Since $\bigcap_{\eta>0} U_\eta(\alpha) = \partial\alpha$ and $U_{\eta_1}(\alpha) \subseteq U_{\eta_2}(\alpha)$ whenever $\eta_1 \le \eta_2$, it follows from (9.19) that

$$\lim_{\eta\to 0} \mu(U_\eta(\alpha)) = 0.$$

Consequently, there exists $\eta_\varepsilon > 0$ such that $\mu(U_\eta(\alpha)) < \varepsilon$ for every $0 < \eta \le \eta_\varepsilon$. By the ergodic case of Birkhoff's ergodic theorem for an indicator function (Corollary 8.2.15) and by Egorov's theorem (Theorem A.1.44), for every $\eta \in (0, \eta_\varepsilon]$ there exist a Borel set $X(\varepsilon,\eta) \subseteq X$ and an integer $M(\varepsilon,\eta) \in \mathbb{N}$ such that

$$\mu(X(\varepsilon,\eta)) \ge 1 - \varepsilon \tag{9.20}$$

and

$$\frac{1}{n}\#\{0 \le i < n : T^i(x) \in U_\eta(\alpha)\} < \varepsilon, \quad \forall x \in X(\varepsilon,\eta),\ \forall n \ge M(\varepsilon,\eta).$$

Now, observe that if $y \in B_n(x,\eta)$ then for each $0 \le i < n$,

$$\text{either} \quad \alpha(T^i(x)) = \alpha(T^i(y)) \quad \text{or} \quad T^i(x) \in U_\eta(\alpha).$$

So, if $x \in X(\varepsilon,\eta)$ and $y \in B_n(x,\eta)$ for some $n \ge M(\varepsilon,\eta)$, then

$$\rho^{(H)}_{\alpha,n}(\widehat{\alpha}^n(x), \widehat{\alpha}^n(y)) < \varepsilon.$$

(See (9.10) for the definition of $\rho^{(H)}_{\alpha,n}$.) Equivalently,

$$B_n(x,\eta) \subseteq \breve{B}^{(H)}_{\alpha,n}(\widehat{\alpha}^n(x),\varepsilon), \quad \forall x \in X(\varepsilon,\eta),\ \forall n \geq M(\varepsilon,\eta). \tag{9.21}$$

We thus need an upper estimate on $\mu(\breve{B}^{(H)}_{\alpha,n}(\widehat{\alpha}^n(x),\varepsilon))$. For every $n \in \mathbb{N}$ define

$$Z_n := \{A \in \alpha^n : \mu(A) \geq \exp((-\mathrm{h}_\mu(T,\alpha) + 3g(\varepsilon))n)\},$$

with $g(\cdot)$ from (9.11). As sets in Z_n are mutually disjoint and $\mu(X) = 1$, we deduce that

$$\#Z_n \leq \exp((\mathrm{h}_\mu(T,\alpha) - 3g(\varepsilon))n). \tag{9.22}$$

To get an appropriate upper estimate, there are "good" and "bad" atoms in α^n. Let

$$\mathrm{Bad}(\alpha^n,\varepsilon) := \{A \in \alpha^n : B^{(H)}_{\alpha,n}(\widehat{A},\varepsilon) \cap \widehat{Z_n} \neq \emptyset\} \subseteq \breve{B}^{(H)}_{\alpha,n}(\widehat{Z_n},\varepsilon). \tag{9.23}$$

Using (9.13), if $n \geq N(\varepsilon)$ and $A \in \alpha^n \setminus \mathrm{Bad}(\alpha^n,\varepsilon)$ then we obtain

$$\begin{aligned}\mu(\breve{B}^{(H)}_{\alpha,n}(\widehat{A},\varepsilon)) &\leq \#B^{(H)}_{\alpha,n}(\widehat{A},\varepsilon)\exp((-\mathrm{h}_\mu(T,\alpha) + 3g(\varepsilon))n)\\ &\leq \exp((-\mathrm{h}_\mu(T,\alpha) + 4g(\varepsilon))n).\end{aligned} \tag{9.24}$$

Along with (9.21), this implies that

$$\mu(B_n(x,\eta)) \leq \exp((-\mathrm{h}_\mu(T,\alpha) + 4g(\varepsilon))n) \tag{9.25}$$

if $x \in X(\varepsilon,\eta) \setminus \cup\mathrm{Bad}(\alpha^n,\varepsilon)$ for some $n \geq \max\{N(\varepsilon), M(\varepsilon,\eta)\}$. Hence,

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\frac{-\log\mu(B_n(x,\delta))}{n} \geq \liminf_{n\to\infty}\frac{-\log\mu(B_n(x,\eta))}{n} \geq \mathrm{h}_\mu(T,\alpha) - 4g(\varepsilon) \tag{9.26}$$

for all $x \in \bigcup_{i=1}^\infty \bigcap_{j=i}^\infty (X(\varepsilon,\eta) \setminus \cup\mathrm{Bad}(\alpha^j,\varepsilon)) = X(\varepsilon,\eta) \cap \bigcup_{i=1}^\infty \bigcap_{j=i}^\infty (X \setminus \cup\mathrm{Bad}(\alpha^j,\varepsilon))$.

We need to show that this latter set is big measurewise, i. e. that its measure is ε-close to 1. For this, it suffices to show that one of its subsets is big. This subset has the similar form $X(\varepsilon,\eta) \cap \bigcup_{i=1}^\infty \bigcap_{j=i}^\infty (\cup\beta_j(\varepsilon) \setminus \cup\mathrm{Bad}(\alpha^j,\varepsilon))$. We define $\beta_j(\varepsilon)$ now and then estimate the measure of the said subset in several steps. This is the most arduous task.

By the ergodic case of the Shannon–McMillan–Breiman theorem (Corollary 9.5.5) and Egorov's theorem (Theorem A.1.44), there exists a Borel set $Y(\varepsilon) \subseteq X$ and an integer $N_1(\varepsilon) \geq N(\varepsilon)$ such that

$$\mu(Y(\varepsilon)) > 1 - \varepsilon \tag{9.27}$$

and

$$\frac{-\log\mu(\alpha^n(x))}{n} \geq \mathrm{h}_\mu(T,\alpha) - g(\varepsilon), \quad \forall x \in Y(\varepsilon),\ \forall n \geq N_1(\varepsilon).$$

Equivalently,

$$\mu(\alpha^n(x)) \le \exp(-(\mathrm{h}_\mu(T,\alpha) - g(\varepsilon))n), \quad \forall x \in Y(\varepsilon),\ \forall n \ge N_1(\varepsilon).$$

Let

$$\beta_n(\varepsilon) := \{\alpha^n(x) : x \in Y(\varepsilon)\}.$$

Fix temporarily $n \ge N_1(\varepsilon)$. Then

$$\mu(A) \le \exp(-(\mathrm{h}_\mu(T,\alpha) - g(\varepsilon))n), \quad \forall A \in \beta_n(\varepsilon). \tag{9.28}$$

Let

$$D_n(\varepsilon) := \{A \in \beta_n(\varepsilon) : \widehat{A} \in B^{(H)}_{\alpha,n}(\widehat{Z_n},\varepsilon)\} = \beta_n(\varepsilon) \cap \breve{B}^{(H)}_{\alpha,n}(\widehat{Z_n},\varepsilon).$$

Using (9.28) as well as (9.13) and (9.22), we get

$$\begin{aligned}\mu(\cup D_n(\varepsilon)) &\le \exp(-(\mathrm{h}_\mu(T,\alpha) - g(\varepsilon))n)\#(\beta_n(\varepsilon) \cap \breve{B}^{(H)}_{\alpha,n}(\widehat{Z_n},\varepsilon))\\ &\le \exp(-(\mathrm{h}_\mu(T,\alpha) - g(\varepsilon))n)\#B^{(H)}_{\alpha,n}(\widehat{Z_n},\varepsilon)\\ &\le \exp(-(\mathrm{h}_\mu(T,\alpha) - g(\varepsilon))n)\#Z_n \exp(g(\varepsilon)n)\\ &\le \exp(-g(\varepsilon)n).\end{aligned} \tag{9.29}$$

Using (9.23), we obtain that

$$\begin{aligned}\cup\beta_n(\varepsilon) \setminus \cup\mathrm{Bad}(\alpha^n,\varepsilon) &= \cup\beta_n(\varepsilon) \setminus (\cup\beta_n(\varepsilon) \cap \cup\mathrm{Bad}(\alpha^n,\varepsilon))\\ &\supseteq \cup\beta_n(\varepsilon) \setminus (\cup\beta_n(\varepsilon) \cap \cup\breve{B}^{(H)}_{\alpha,n}(\widehat{Z_n},\varepsilon))\\ &= \cup\beta_n(\varepsilon) \setminus \cup(\beta_n(\varepsilon) \cap \breve{B}^{(H)}_{\alpha,n}(\widehat{Z_n},\varepsilon))\\ &= \cup\beta_n(\varepsilon) \setminus \cup D_n(\varepsilon).\end{aligned}$$

Therefore, for every $k \ge N_1(\varepsilon)$ we obtain that

$$\begin{aligned}\bigcap_{n=k}^{\infty}(\cup\beta_n(\varepsilon) \setminus \cup\mathrm{Bad}(\alpha^n,\varepsilon)) &\supseteq \bigcap_{n=k}^{\infty}(\cup\beta_n(\varepsilon) \setminus \cup D_n(\varepsilon)) \supseteq \bigcap_{n=k}^{\infty}\cup\beta_n(\varepsilon) \Big\backslash \bigcup_{n=k}^{\infty}\cup D_n(\varepsilon)\\ &\supseteq \bigcap_{n=k}^{\infty} Y(\varepsilon) \Big\backslash \bigcup_{n=k}^{\infty}\cup D_n(\varepsilon) = Y(\varepsilon) \Big\backslash \bigcup_{n=k}^{\infty}\cup D_n(\varepsilon).\end{aligned}$$

From this, (9.27) and (9.29), we deduce that

$$\begin{aligned}\mu\left(\bigcap_{n=k}^{\infty}(\cup\beta_n(\varepsilon) \setminus \cup\mathrm{Bad}(\alpha^n,\varepsilon))\right) &\ge \mu(Y(\varepsilon)) - \mu\left(\bigcup_{n=k}^{\infty}\cup D_n(\varepsilon)\right)\\ &\ge \mu(Y(\varepsilon)) - \sum_{n=k}^{\infty}\mu(\cup D_n(\varepsilon))\end{aligned}$$

$$\begin{aligned} &\geq 1-\varepsilon-\sum_{n=k}^{\infty} e^{-g(\varepsilon)n} \\ &= 1-\varepsilon-\frac{e^{-g(\varepsilon)k}}{1-e^{-g(\varepsilon)}}. \end{aligned} \tag{9.30}$$

Consequently,

$$\mu\left(\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}(\cup\beta_n(\varepsilon)\setminus\cup\mathrm{Bad}(\alpha^n,\varepsilon))\right)=\lim_{k\to\infty}\mu\left(\bigcap_{n=k}^{\infty}(\cup\beta_n(\varepsilon)\setminus\cup\mathrm{Bad}(\alpha^n,\varepsilon))\right)\geq 1-\varepsilon. \tag{9.31}$$

It follows from (9.31) and (9.20) that

$$\mu\left(X(\varepsilon,\eta)\cap\left(\bigcup_{i=1}^{\infty}\bigcap_{j=i}^{\infty}(\cup\beta_j(\varepsilon)\setminus\cup\mathrm{Bad}(\alpha^j,\varepsilon))\right)\right)\geq 1-2\varepsilon.$$

Thus

$$\mu\left(\bigcup_{q=k}^{\infty}\left(X(1/q,\eta_{1/q})\cap\left(\bigcup_{i=1}^{\infty}\bigcap_{j=i}^{\infty}(\cup\beta_j(1/q)\setminus\cup\mathrm{Bad}(\alpha^j,1/q))\right)\right)\right)=1,\quad\forall k\in\mathbb{N}.$$

If

$$\widehat{X}_2(\alpha):=\bigcap_{k=1}^{\infty}\bigcup_{q=k}^{\infty}\left(X(1/q,\eta_{1/q})\cap\left(\bigcup_{i=1}^{\infty}\bigcap_{j=i}^{\infty}(\cup\beta_j(1/q)\setminus\cup\mathrm{Bad}(\alpha^j,1/q))\right)\right),$$

then

$$\mu(\widehat{X}_2(\alpha))=1 \tag{9.32}$$

and, by virtue of (9.26) and (9.12), we deduce that

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\frac{-\log\mu(B_n(x,\delta))}{n}\geq \mathrm{h}_\mu(T,\alpha),\quad\forall x\in\widehat{X}_2(\alpha). \tag{9.33}$$

As the metric space X is compact, it follows from Lemma 9.6.1 and Theorem 9.4.17 that there exists a sequence $(\alpha_k)_{k=1}^{\infty}$ of finite Borel partitions of X such that $\mu(\partial\alpha_k)=0$ for every $k\in\mathbb{N}$ and

$$\lim_{k\to\infty}\mathrm{h}_\mu(T,\alpha_k)=\mathrm{h}_\mu(T).$$

Setting

$$\widehat{X}_2:=\bigcap_{k=1}^{\infty}\widehat{X}_2(\alpha_k),$$

we have by (9.32) that

$$\mu(\widehat{X}_2) = 1 \tag{9.34}$$

and by (9.33) that

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\frac{-\log\mu(B_n(x,\delta))}{n} \geq \mathrm{h}_\mu(T), \quad \forall x \in \widehat{X}_2.$$

This is the left inequality in (9.14).

As $\mu(X_1 \cap \widehat{X}_2) = 1$ by (9.17) and (9.34), and as all three inequalities in (9.14) are valid on $X_1 \cap \widehat{X}_2$, the result ensues. □

Remark 9.6.3. As a matter of fact, only total boundedness of the metric d is needed for Theorem 9.6.2 to hold. More precisely, in Lemma 9.6.1 total boundedness is sufficient, and compactness has not been used anywhere in the proof of Theorem 9.6.2.

In the case of an expansive system T, we have the following stronger and simpler version of Theorem 9.6.2.

Theorem 9.6.4 (Brin–Katok local entropy formula for expansive maps). *Let $T : X \to X$ be an expansive topological dynamical system and let d be a metric compatible with the topology on X. If $\delta > 0$ is an expansive constant for T corresponding to this metric, then for every $\zeta \in (0, \delta]$, every ergodic T-invariant Borel probability measure μ on X and μ-almost every $x \in X$, we have*

$$\mathrm{h}_\mu(T) = \lim_{n\to\infty}\frac{-\log\mu(B_n(x,\zeta))}{n}.$$

Proof. For every $x \in X$, denote

$$\overline{h}_\mu(T,\zeta,x) := \limsup_{n\to\infty}\frac{-\log\mu(B_n(x,\zeta))}{n} \quad \text{and} \quad \underline{h}_\mu(T,\zeta,x) := \liminf_{n\to\infty}\frac{-\log\mu(B_n(x,\zeta))}{n}.$$

Since $B_n(x,\zeta) \subseteq B_n(x,\delta)$ for every $n \in \mathbb{N}$, it is clear that

$$\overline{h}_\mu(T,\delta,x) \leq \overline{h}_\mu(T,\zeta,x). \tag{9.35}$$

On the other hand, using Observation 5.2.4 we obtain that

$$\begin{aligned}
\underline{h}_\mu(T,\delta,x) &= \liminf_{n\to\infty}\frac{-\log\mu(B_n(x,\delta))}{n} \\
&\geq \liminf_{n\to\infty}\frac{-\log\mu(B_{n-N(\zeta/2)}(x,\zeta))}{n} \\
&= \liminf_{n\to\infty}\frac{n-N(\zeta/2)}{n}\cdot\frac{-\log\mu(B_{n-N(\zeta/2)}(x,\zeta))}{n-N(\zeta/2)}
\end{aligned}$$

$$= \liminf_{n\to\infty} \frac{-\log \mu(B_{n-N(\zeta/2)}(x,\zeta))}{n - N(\zeta/2)}$$
$$= \underline{h}_\mu(T,\zeta,x). \tag{9.36}$$

Since, by Theorem 9.6.2,

$$\lim_{\zeta\to 0} \underline{h}_\mu(T,\zeta,x) = \mathrm{h}_\mu(T) = \lim_{\zeta\to 0} \overline{h}_\mu(T,\zeta,x) \quad \text{for } \mu\text{-a.e. } x \in X,$$

we infer from (9.36) and (9.35) that $\underline{h}_\mu(T,\delta,x) \geq \mathrm{h}_\mu(T)$ and $\overline{h}_\mu(T,\delta,x) \leq \mathrm{h}_\mu(T)$ for μ-almost every $x \in X$. As it is always true that $\underline{h}_\mu(T,\delta,x) \leq \overline{h}_\mu(T,\delta,x)$, we deduce that $\underline{h}_\mu(T,\delta,x) = \mathrm{h}_\mu(T) = \overline{h}_\mu(T,\delta,x)$ and the result holds when $\zeta = \delta$. Since any $\zeta \in (0,\delta]$ is also an expansive constant for T, the theorem is validated. □

9.7 Exercises

Exercise 9.7.1. The objective of this exercise is to prove Theorem 9.1.1. Using axioms (A1)–(A4), proceed as follows.

(a) Given $n \in \mathbb{N}$, prove by induction that $f(n^k) = kf(n)$ for all $k \geq 0$. (Think about the meaning of this relationship.)

(b) Given $n \geq 2$, for every $r \in \mathbb{N}$ there exists a unique $k \geq 0$ such that $n^k \leq 2^r < n^{k+1}$. Show that

$$kf(n) \leq rf(2) < (k+1)f(n).$$

(c) Prove that (b) holds with f replaced by log.

(d) Given $n \geq 2$, deduce that

$$\left|\frac{f(2)}{f(n)} - \frac{\log 2}{\log n}\right| < \frac{1}{r}, \quad \forall r \in \mathbb{N}.$$

(e) Conclude that $f(n) = (f(2)/\log 2)\log n$ for all $n \in \mathbb{N}$.

(f) Let $C = f(2)/\log 2$. Observe that in order to establish that

$$\mathrm{H}(p, 1-p) = -Cp\log p - C(1-p)\log(1-p), \tag{9.37}$$

it suffices to prove that this relation holds for all rational $p \in (0,1)$.

(g) Accordingly, let $p = r/s \in \mathbb{Q} \cap (0,1)$. By partitioning some experiment appropriately, show that

$$f(s) = \mathrm{H}\left(\frac{r}{s}, \frac{s-r}{s}\right) + \frac{r}{s}f(r) + \frac{s-r}{s}f(s-r).$$

(h) Deduce (9.37) and observe that the function H extends continuously to $[0,1]$.

(i) Hence, the formula in Theorem 9.1.1 holds when $n = 2$. Prove by induction that it holds for any $n \in \mathbb{N}$.

Exercise 9.7.2. Let $(X, \mathcal{A})$ be a measurable space. Show that $\leq$ is a partial order relation on the set $\mathrm{Part}(X, \mathcal{A})$.

Exercise 9.7.3. Let $(X, \mathcal{A})$ be a measurable space and $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$. Prove that $\alpha \leq \beta$ if and only if $\beta(x) \subseteq \alpha(x)$ for all $x \in X$.

Exercise 9.7.4. Show that $\alpha \leq \beta$ if and only if $A = \bigcup\{B \in \beta : B \cap A \neq \emptyset\}$ for all $A \in \alpha$.

Exercise 9.7.5. Prove that $\alpha \leq \beta$ if and only if $A = \bigcup\{B \in \beta : B \subseteq A\}$ for all $A \in \alpha$.

Exercise 9.7.6. Find a probability space $(X, \mathcal{A}, \mu)$ and $\alpha, \beta, \gamma \in \mathrm{Part}(X, \mathcal{A})$ such that $\beta \leq \gamma$ but $I_\mu(\alpha|\beta)(x) \ngeq I_\mu(\alpha|\gamma)(x)$ for some $x \in X$.

Exercise 9.7.7. Find a probability space $(X, \mathcal{A}, \mu)$ and $\alpha, \beta \in \mathrm{Part}(X, \mathcal{A})$ such that $I_\mu(\alpha \vee \beta)(x) \nleq I_\mu(\alpha)(x) + I_\mu(\beta)(x)$ for some $x \in X$.

Exercise 9.7.8. Find a probability space $(X, \mathcal{A}, \mu)$ and $\alpha, \beta, \gamma \in \mathrm{Part}(X, \mathcal{A})$ such that $I_\mu(\alpha|\gamma)(x) \nleq I_\mu(\alpha|\beta)(x) + I_\mu(\beta|\gamma)(x)$ for some $x \in X$.

Exercise 9.7.9. Let $T : X \to X$ be a measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. Given $\alpha \in \mathrm{Part}(X, \mathcal{A})$, show that the sequence $(\mathrm{H}_\mu(\alpha^n))_{n=1}^\infty$ is subadditive. Then deduce from Lemma 3.2.17 that the limit $\lim_{n\to\infty} \frac{1}{n}\mathrm{H}_\mu(\alpha^n) = \mathrm{h}_\mu(T, \alpha)$ exists and is nonnegative.

Exercise 9.7.10. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. Show that the mapping $\mu \mapsto \mathrm{h}_\mu(T)$ is affine on the set $M(T, \mathcal{A})$ of all T-invariant probability measures on $(X, \mathcal{A})$. In other words, show that if $T : X \to X$ is a dynamical system preserving two probability measures μ and ν on the measurable space $(X, \mathcal{A})$, then

$$\mathrm{h}_{s\mu+(1-s)\nu}(T) = s\mathrm{h}_\mu(T) + (1-s)\mathrm{h}_\nu(T)$$

for all $0 \leq s \leq 1$.

Exercise 9.7.11. Let $T : X \to X$ be an ergodic measure-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, and let $\alpha \in \mathrm{Part}(X, \mathcal{A})$ be such that $\mathrm{H}_\mu(\alpha) < \infty$. According to Corollary 9.5.5,

$$\mathrm{h}_\mu(T, \alpha) = \lim_{n\to\infty} \frac{1}{n} I_\mu(\alpha^n)(x) \quad \text{for } \mu\text{-a. e. } x \in X.$$

Let $0 < \varepsilon < 1$ and for each $n \in \mathbb{N}$ let $N_n(\varepsilon)$ be the minimum number of atoms of α^n needed to construct a set of measure at least $1 - \varepsilon$. Show that

$$\mathrm{h}_\mu(T, \alpha) = \lim_{n\to\infty} \frac{1}{n} \log N_n(\varepsilon).$$

Exercise 9.7.12. Prove that for the full E-shift equipped with the product measure μ as in Example 9.4.23, and the partition $\alpha := \{[e] : e \in E\}$,

$$\mathrm{H}_\mu(\alpha^n) = -n \sum_{e\in E} P(e) \log P(e), \quad \forall n \in \mathbb{N}.$$

Exercise 9.7.13. Let $T : [0,1] \to [0,1]$ be the tent map (see Example 1.1.3). Show that T preserves the Lebesgue measure on $[0,1]$ and that its entropy with respect to Lebesgue measure is equal to $\log 2$.

Exercise 9.7.14. Let $(X, \mathcal{A})$ be a measurable space and $(\alpha_n)_{n=1}^\infty$ be a sequence of increasingly finer countable measurable partitions of $(X, \mathcal{A})$ which generates the σ-algebra $\mathcal{A}$, that is, such that

$$\alpha_n \le \alpha_{n+1}, \quad \forall n \in \mathbb{N} \quad \text{and} \quad \sigma\left(\bigcup_{n=1}^\infty \mathcal{A}_n\right) = \mathcal{A}.$$

Suppose that μ and ν are probability measures on $(X, \mathcal{A})$ such that $\mu \ll \nu$. Let $\rho = d\mu/d\nu$ be the Radon–Nikodym derivative of μ with respect to ν (cf. Theorem A.1.50). Using Example A.1.62 and the martingale convergence theorem for conditional expectations (Theorem A.1.67), show that

$$\rho(x) = \lim_{n\to\infty} \frac{\mu(\alpha_n(x))}{\nu(\alpha_n(x))} \text{ for } \nu\text{-a. e. } x \in X.$$

Exercise 9.7.15. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces and $T : X \to X$ and $S : Y \to Y$ be measurable transformations. A measurable transformation $\pi : X \to Y$ is called a *factor map* between T and S if $\pi \circ T = S \circ \pi$.

If $(X, \mathcal{A}, \mu)$ is a measure space, then recall that π induces the push down measure $\mu \circ \pi^{-1}$ on the measurable space $(Y, \mathcal{B})$.

Let $\alpha, \beta \in \mathrm{Part}(Y, \mathcal{B})$. Prove the following statements:

(a) $\pi^{-1}(\alpha \vee \beta) = (\pi^{-1}\alpha) \vee (\pi^{-1}\beta)$.

(b) $\pi^{-1}(\alpha_m^n) = (\pi^{-1}\alpha)_m^n$ for all $m, n \ge 0$.

(c) π^{-1} preserves the partial order $\le$, that is, if $\alpha \le \beta$ then $\pi^{-1}\alpha \le \pi^{-1}\beta$.

(d) $(\pi^{-1}\alpha)(x) = \pi^{-1}(\alpha(\pi(x)))$ for all $x \in X$.

(e) $I_\mu(\pi^{-1}\alpha|\pi^{-1}\beta) = I_{\mu\circ\pi^{-1}}(\alpha|\beta) \circ \pi$.

(f) $I_\mu(\pi^{-1}\alpha) = I_{\mu\circ\pi^{-1}}(\alpha) \circ \pi$.

(g) $\mathrm{H}_\mu(\pi^{-1}\alpha|\pi^{-1}\beta) = \mathrm{H}_{\mu\circ\pi^{-1}}(\alpha|\beta)$.

(h) $\mathrm{H}_\mu(\pi^{-1}\alpha) = \mathrm{H}_{\mu\circ\pi^{-1}}(\alpha)$.

(i) $\mathrm{h}_\mu(T, \pi^{-1}\alpha) = \mathrm{h}_{\mu\circ\pi^{-1}}(S, \alpha)$.

(j) $\mathrm{h}_\mu(T) \ge \mathrm{h}_{\mu\circ\pi^{-1}}(S)$.

(k) If π is bimeasurable (i. e., measurable, bijective and its inverse is measurable), then $\mathrm{h}_\mu(T) = \mathrm{h}_{\mu\circ\pi^{-1}}(S)$.

10 Infinite invariant measures

In this chapter, we deal with measurable transformations preserving measures that are no longer assumed to be finite. The outlook is then substantially different than in the case of finite measures. As far as we know, only J. Aaronson's book [1] is entirely dedicated to infinite ergodic theory.

In Section 10.1, we introduce and investigate in detail the notions of quasi-invariant measures, ergodicity, and conservativity. We also prove Halmos' recurrence theorem, which is a generalization of Poincaré's recurrence theorem for quasi-invariant measures that are not necessarily finite.

In Section 10.2, we discuss first return times, first return maps, and induced systems. We further establish relations between invariant measures for the original transformation and the induced transformation.

In Section 10.3, we study implications of Birkhoff's ergodic theorem for finite and infinite measure spaces. Among others, we demonstrate Hopf's ergodic theorem, which applies to measure-preserving transformations of σ-finite spaces.

Finally, in Section 10.4, we seek a condition under which, given a quasi-invariant probability measure, one can construct a σ-finite invariant measure which is absolutely continuous with respect to the original measure. To this end, we introduce a class of transformations, called Martens maps, that have this feature and even more. In fact, these maps have the property that any quasi-invariant probability measure admits an equivalent σ-finite invariant one.

Applications of these concepts and results can be found in Chapters 13–14 of the second volume and Chapters 29–32 of the third volume.

10.1 Quasi-invariant measures, ergodicity and conservativity

By definition, quasi-invariant measures preserve sets of measure zero.

Definition 10.1.1. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. A measure μ on $(X, \mathcal{A})$ is called *quasi-T-invariant* if $\mu \circ T^{-1} \ll \mu$.

Obviously, invariant measures are quasi-invariant but the converse statement does not hold in general.

The concept of ergodicity defined in Chapter 8 for transformations of probability spaces readily generalizes to transformations of arbitrary measure spaces.

Definition 10.1.2. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. Then T is *ergodic* with respect to μ if

$$T^{-1}(A) = A \quad \Longrightarrow \quad \mu(A) = 0 \text{ or } \mu(X \backslash A) = 0.$$

Alternatively, μ is said to be ergodic with respect to T.

https://doi.org/10.1515/9783110702682-010

The next result states that, like a T-invariant measure, a quasi-T-invariant measure μ is ergodic if and only if every μ-a. e. T-invariant set is trivial in a measure-theoretic sense, that is, has measure zero or its complement is of measure zero. This is a generalization of Proposition 8.2.4.

Proposition 10.1.3. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. Then T is ergodic with respect to μ if and only if*

$$\mu\big(T^{-1}(A) \triangle A\big) = 0 \quad \Longrightarrow \quad \mu(A) = 0 \ \textit{or} \ \mu(X \backslash A) = 0.$$

Proof. The proof goes along similar lines to that of Proposition 8.2.4 and is left to the reader. □

In Section 1.4, we studied the concept of wandering points. We revisit this dynamical behavior from a measure-theoretic standpoint.

Definition 10.1.4. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. A set $W \in \mathcal{A}$ is a *wandering set for T* if its preimages $(T^{-n}(W))_{n=0}^{\infty}$ are mutually disjoint.

One way of constructing wandering sets is now described.

Lemma 10.1.5. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. For every $A \in \mathcal{A}$, the set $W_A := A \backslash \bigcup_{n=1}^{\infty} T^{-n}(A)$ is a wandering set for T.*

To lighten notation, define

$$A^{-} := \bigcup_{n=1}^{\infty} T^{-n}(A) \quad \text{and thus} \quad W_A := A \backslash A^{-}.$$

Proof. Suppose for a contradiction that W_A is not wandering for T, that is, $T^{-k}(W_A) \cap T^{-l}(W_A) \neq \emptyset$ for some $0 \le k < l$. This means that $T^{-k}\big(W_A \cap T^{-(l-k)}(W_A)\big) \neq \emptyset$, and thus $W_A \cap T^{-(l-k)}(W_A) \neq \emptyset$. Set $j = l - k \in \mathbb{N}$. Fix $x \in W_A \cap T^{-j}(W_A)$. On one hand, $x \in W_A$ implies that $x \notin A^{-}$. So $x \notin T^{-j}(A)$. On the other hand, $W_A \subseteq A$ and $x \in T^{-j}(W_A)$ imply that $x \in T^{-j}(A)$. This is a contradiction and W_A must therefore be a wandering set. □

Next, we introduce the notion of conservativity.

Definition 10.1.6. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. Then T is *conservative* with respect to μ if $\mu(W) = 0$ for every wandering set W for T.

Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. For every $B \in \mathcal{A}$, define the set $B_\infty \in \mathcal{A}$ to be

$$B_\infty := \{x \in X : T^n(x) \in B \text{ for infinitely many } n \in \mathbb{N}\} = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} T^{-n}(B).$$

Clearly, $T^{-1}(B_\infty) = B_\infty$ and so $T^{-1}(X\backslash B_\infty) = X\backslash B_\infty$. Notice also that if W is a wandering set for T, then $W \cap \bigcup_{n=1}^\infty T^{-n}(W) = \emptyset$. In particular, $W \cap W_\infty = \emptyset$.

Poincaré's recurrence theorem (Theorem 8.1.16) asserts that if μ is T-invariant and finite, then $\mu(B\backslash B_\infty) = 0$. We shall now prove its generalization in two respects: namely, by assuming only (1) that μ is quasi-T-invariant and (2) that μ may be infinite.

Theorem 10.1.7 (Halmos' recurrence theorem). *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. For every $A \in \mathcal{A}$, the following equivalence holds: $\mu(B\backslash B_\infty) = 0$ for all measurable sets $B \subseteq A$ if and only if $\mu(W) = 0$ for all wandering sets $W \subseteq A$.*

Proof. Fix $A \in \mathcal{A}$. If $\mu(A) = 0$, then $\mu(B\backslash B_\infty) = 0$ for all measurable sets $B \subseteq A$ and $\mu(W) = 0$ for all wandering sets $W \subseteq A$. Thus the equivalence is trivially satisfied when $\mu(A) = 0$, and we may assume in the sequel that $\mu(A) > 0$.

[$\Rightarrow$] We will prove the contrapositive statement. Suppose that $\mu(W) > 0$ for some wandering set $W \subseteq A$. Then $W \cap W_\infty = \emptyset$. Therefore, $\mu(W\backslash W_\infty) > 0$. Thus W is a measurable set $B \subseteq A$ such that $\mu(B\backslash B_\infty) > 0$.

[$\Leftarrow$] Let us now prove the converse implication. Assume that $\mu(W) = 0$ for all wandering sets $W \subseteq A$. Fix a measurable set $B \subseteq A$ and for all $n \geq 0$ let

$$B_n := B \cap T^{-n}(B) \Big\backslash \bigcup_{\ell=n+1}^{\infty} T^{-\ell}(B),$$

that is, B_n is the set of points in B that return to B at time n but never again thereafter. So

$$B\backslash B_\infty = \bigcup_{n\geq 0} B_n.$$

Suppose for a contradiction that $\mu(B\backslash B_\infty) > 0$. That is, suppose there is $n \geq 0$ such that $\mu(B_n) > 0$. Lemma 10.1.5 asserts that W_{B_n} is a wandering set for T. Since $B_n \subseteq A$, the hypothesis implies that $\mu(W_{B_n}) = 0$. This means that $\mu(B_n\backslash B_n^-) = 0$. Since $\mu(B_n) > 0$, there thus exists $x \in B_n \cap B_n^- = B_n \cap \bigcup_{k=1}^\infty T^{-k}(B_n)$. So $x \in B_n$. There is also $k \in \mathbb{N}$ such that $x \in T^{-k}(B_n)$. As $T^{-k}(B_n) \subseteq T^{-(n+k)}(B)$, it turns out that $x \in T^{-\ell}(B)$, where $\ell = n + k$. So $x \notin B_n$. This is a contradiction. Consequently, $\mu(B\backslash B_\infty) = 0$ for any measurable set $B \subseteq A$. □

Taking $A = X$ in Theorem 10.1.7, we get the following special case.

Corollary 10.1.8. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. Then T is conservative if and only if $\mu(B\backslash B_\infty) = 0$ for all $B \in \mathcal{A}$.*

In particular, Poincaré's recurrence theorem (Theorem 8.1.16) confirms that every measure-preserving transformation of a finite measure space is conservative.

Corollary 10.1.9. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a finite T-invariant measure. Then T is conservative.*

Conservativity also has the following consequence.

Corollary 10.1.10. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. If T is conservative, then*

$$\sum_{n=0}^{\infty} \mu\big(T^{-n}(A)\big) = \infty$$

for all sets $A \in \mathcal{A}$ such that $\mu(A) > 0$.

Proof. Let $A \in \mathcal{A}$. If $\sum_{n=0}^{\infty} \mu(T^{-n}(A)) < \infty$, then $\mu(A_\infty) = 0$ by Borel–Cantelli Lemma (Lemma A.1.20). Moreover, $\mu(A \backslash A_\infty) = 0$ according to Corollary 10.1.8. Therefore, we conclude that $\mu(A) = \mu(A \cap A_\infty) + \mu(A \backslash A_\infty) = 0$. □

We now prove a characterization of ergodicity + conservativity.

Theorem 10.1.11. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. Then T is conservative and ergodic if and only if $\mu(X \backslash A_\infty) = 0$ for every $A \in \mathcal{A}$ such that $\mu(A) > 0$.*

Proof. Assume first that $T : X \to X$ is conservative and ergodic. Fix $A \in \mathcal{A}$ with $\mu(A) > 0$. We have earlier observed that $T^{-1}(A_\infty) = A_\infty$. Then, due to the ergodicity of μ, either $\mu(A_\infty) = 0$ or $\mu(X \backslash A_\infty) = 0$. In the former case, $\mu(A) = \mu(A \backslash A_\infty)$ and Corollary 10.1.8 implies that $\mu(A) = 0$. As $\mu(A) > 0$ by assumption, this means that this case never happens. Only the latter case $\mu(X \backslash A_\infty) = 0$ occurs, and this proves the implication $\Rightarrow$.

We now prove the converse implication.

Let us first show conservativity. Let $A \in \mathcal{A}$. If $\mu(A) = 0$ then obviously $\mu(A \backslash A_\infty) = 0$. If $\mu(A) > 0$, then by assumption $\mu(X \backslash A_\infty) = 0$ and again $\mu(A \backslash A_\infty) = 0$. Corollary 10.1.8 then confirms the conservativity of T.

Now we establish ergodicity. Let $A \in \mathcal{A}$ be such that $T^{-1}(A) = A$. Then $A_\infty = A$. We must show that either $\mu(A) = 0$ or $\mu(X \backslash A) = 0$. Suppose that $\mu(A) > 0$. By assumption, we then have $\mu(X \backslash A_\infty) = 0$. Since $A_\infty = A$, this means that $\mu(X \backslash A) = 0$. □

For finite invariant measures, Theorem 10.1.11, in conjunction with Corollary 10.1.9, provides yet another characterization of ergodicity.

Corollary 10.1.12. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a finite T-invariant measure. Then T is ergodic if and only if $\mu(X \backslash A_\infty) = 0$ for every $A \in \mathcal{A}$ such that $\mu(A) > 0$.*

Sets that are visited by almost every point in the space will also play a crucial role. Accordingly, we make the following definition.

Definition 10.1.13. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. A set $A \in \mathcal{A}$ is said to be *absorbing* with respect to μ if $0 < \mu(A) < \infty$ and $\mu(X \setminus \bigcup_{k=0}^{\infty} T^{-k}(A)) = 0$.

Notice that any invariant measure which admits an absorbing set is σ-finite.

Obviously, a set $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$ and $\mu(X \setminus A_\infty) = 0$, is absorbing with respect to μ. Therefore, we have the following corollary to Theorem 10.1.11.

Corollary 10.1.14. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a quasi-T-invariant measure. If T is conservative and ergodic, then every $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$ is absorbing.*

To end this section, we briefly return to transformations of probability spaces and introduce a concept of weak metric exactness. Recall that the notion of metric exactness was described in Definition 8.4.4.

Definition 10.1.15. Let $(X, \mathcal{F}, \mu)$ be a Lebesgue probability space and $T : X \to X$ a transformation for which μ is a quasi-T-invariant measure. Then T is said to be *weakly metrically exact* if for each $A \in \mathcal{F}$ such that $\mu(A) > 0$ we have

$$\limsup_{n\to\infty} \mu\big(T^n(A)\big) = 1.$$

Note that each set $T^n(A)$ is measurable since T is a measurable transformation of a Lebesgue space.

Theorem 10.1.16. *Every weakly metrically exact transformation $T : (X, \mathcal{F}, \mu) \to (X, \mathcal{F}, \mu)$ is conservative and ergodic.*

Proof. We first prove ergodicity. Let $A \in \mathcal{A}$ be such that $T^{-1}(A) = A$. Then $T^n(A) \subseteq A$ for all $n \in \mathbb{N}$. Since μ is a probability measure, we must show that $\mu(A) \in \{0, 1\}$. Suppose that $\mu(A) > 0$. The weak metric exactness of T implies that

$$1 = \limsup_{n\to\infty} \mu\big(T^n(A)\big) \le \mu(A) \le 1.$$

So $\mu(A) = 1$ and T is ergodic.

To prove the conservativity of T, suppose for a contradiction that there exists a wandering set W such that $\mu(W) > 0$. Then $W \cap \bigcup_{n=1}^{\infty} T^{-n}(W) = \emptyset$. Therefore $\bigcup_{n=1}^{\infty} T^n(W) \subseteq X \setminus W$, and the weak metric exactness of T yields that

$$1 = \limsup_{n\to\infty} \mu\big(T^n(W)\big) \le \mu(X \setminus W) = 1 - \mu(W).$$

So $\mu(W) = 0$. This is a contradiction. Consequently, T is conservative. □

10.2 Invariant measures and inducing

Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a T-invariant measure. Fix $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$ and $\mu(A \setminus A_\infty) = 0$. Then $\mu(A \cap A_\infty) = \mu(A)$. Let

$A'_\infty = A \cap A_\infty$. The function $\tau_{A'_\infty} : A'_\infty \to \mathbb{N}$ given by the formula

$$\tau_{A'_\infty}(x) = \min\{n \in \mathbb{N} : T^n(x) \in A'_\infty\} \tag{10.1}$$

is well-defined and measurable when A'_∞ is endowed with the σ-algebra $\mathcal{A}|_{A'_\infty} := \{B \subseteq A'_\infty : B \in \mathcal{A}\}$ and $\mathbb{N}$ is equipped with the σ-algebra $\mathcal{P}(\mathbb{N})$ of all subsets of $\mathbb{N}$. Consequently, the map $T_{A'_\infty} : A'_\infty \to A'_\infty$ defined by

$$T_{A'_\infty}(x) = T^{\tau_{A'_\infty}(x)}(x) \tag{10.2}$$

is well-defined and measurable. The number $\tau_{A'_\infty}(x) \in \mathbb{N}$ is called the *first return time* of x to the set A'_∞ and, accordingly, the map $T_{A'_\infty}$ is called the *first return map* or the *induced map*. Given that $A'_\infty \subseteq A$ and $\mu(A'_\infty) = \mu(A)$, without loss of generality we will assume that $A'_\infty = A$ from the outset, hence alleging notation to τ_A and T_A.

Finally, in a similar way to Definition A.1.70, let μ_A be the conditional probability measure on $(A, \mathcal{A}|_A)$ defined by

$$\mu_A(B) = \frac{\mu(B)}{\mu(A)}, \quad \forall B \in \mathcal{A}|_A.$$

This measure has the following properties.

Theorem 10.2.1. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$.*

If μ is T-invariant and $\mu(A \backslash A_\infty) = 0$, then μ_A is T_A-invariant and T_A is conservative.

Conversely, if ν is a T_A-invariant probability measure on $(A, \mathcal{A}|_A)$, then there exists a T-invariant measure $\tilde{\mu}$ on $(X, \mathcal{A})$ such that $\tilde{\mu}_A = \nu$ on $(A, \mathcal{A}|_A)$. In fact, $\tilde{\mu}$ may be constructed as follows: for any $B \in \mathcal{A}$, let

$$\tilde{\mu}(B) = \sum_{n=0}^{\infty} \nu\left(A \cap T^{-n}(B) \setminus \bigcup_{k=1}^{n} T^{-k}(A) \right) = \sum_{n=0}^{\infty} \nu(\{x \in A \cap T^{-n}(B) : \tau_A(x) > n\}). \tag{10.3}$$

In particular, the set A is absorbing with respect to $\tilde{\mu}$ and $\tilde{\mu}$ is σ-finite. The measure $\tilde{\mu}$ is said to be induced *by ν.*

Proof. First, suppose that μ is T-invariant and that $0 < \mu(A) < \infty$ and $\mu(A \backslash A_\infty) = 0$. Let $B \in \mathcal{A}|_A$. Then

$$\begin{aligned} \mu(T_A^{-1}(B)) &= \sum_{n=1}^{\infty} \mu(T_A^{-1}(B) \cap \tau_A^{-1}(n)) \\ &= \sum_{n=1}^{\infty} \mu(A \cap T^{-n}(B) \cap \tau_A^{-1}(n)) \\ &= \sum_{n=1}^{\infty} \mu\left(A \cap T^{-n}(B) \setminus \bigcup_{k=1}^{n-1} T^{-k}(A) \right) \\ &= \sum_{n=1}^{\infty} \mu(A \cap T^{-1}(B_{n-1})), \end{aligned} \tag{10.4}$$

where

$$B_0 := B \quad \text{and} \quad B_n := T^{-n}(B) \Big\backslash \bigcup_{k=0}^{n-1} T^{-k}(A), \quad \forall n \in \mathbb{N}.$$

Observe that $\mu(B_n) \le \mu(T^{-n}(B)) = \mu(B) \le \mu(A) < \infty$ for every $n \ge 0$. Since $T^{-1}(B_{n-1}) = (A \cap T^{-1}(B_{n-1})) \cup B_n$ and $(A \cap T^{-1}(B_{n-1})) \cap B_n = \emptyset$ for all $n \in \mathbb{N}$, we obtain that

$$\mu(A \cap T^{-1}(B_{n-1})) = \mu(T^{-1}(B_{n-1})) - \mu(B_n) = \mu(B_{n-1}) - \mu(B_n)$$

for all $n \in \mathbb{N}$. Therefore, by (10.4),

$$\mu(T_A^{-1}(B)) = \lim_{n\to\infty}(\mu(B) - \mu(B_n)) \le \mu(B). \tag{10.5}$$

This relation also holds for $A\backslash B$, that is,

$$\mu(T_A^{-1}(A\backslash B)) \le \mu(A\backslash B).$$

Since $A = T_A^{-1}(A) = T_A^{-1}(B) \cup T_A^{-1}(A\backslash B)$ and $T_A^{-1}(B) \cap T_A^{-1}(A\backslash B) = \emptyset$, it follows that

$$\mu(T_A^{-1}(B)) = \mu(A) - \mu(T_A^{-1}(A\backslash B)) \ge \mu(A) - \mu(A\backslash B) = \mu(B). \tag{10.6}$$

It ensues from (10.5) and (10.6) that $\mu(T_A^{-1}(B)) = \mu(B)$. Thus $\mu_A \circ T_A^{-1} = \mu_A$. So μ_A is T_A-invariant. The conservativity of T_A with respect to μ_A is a direct consequence of Corollary 10.1.9.

To prove the converse implication, assume that ν is a probability measure on $(A, \mathcal{A}|_A)$ such that $\nu \circ T_A^{-1} = \nu$. We shall first show that the measure $\widetilde{\mu}$ given by (10.3) is T-invariant. Indeed, let $B \in \mathcal{A}$. Then

$$\begin{aligned}
\widetilde{\mu}(T^{-1}(B)) &= \sum_{n=0}^{\infty} \nu(\{x \in A \cap T^{-n}(T^{-1}(B)) : \tau_A(x) > n\}) \\
&= \sum_{n=0}^{\infty} \nu(\{x \in A \cap T^{-(n+1)}(B) : \tau_A(x) > n+1\}) \\
&\quad + \sum_{n=0}^{\infty} \nu(\{x \in A \cap T^{-(n+1)}(B) : \tau_A(x) = n+1\}) \\
&= \widetilde{\mu}(B) - \nu(A \cap B) + \sum_{n=1}^{\infty} \nu(\{x \in A \cap T^{-n}(B) : \tau_A(x) = n\}) \\
&= \widetilde{\mu}(B) - \nu(A \cap B) + \sum_{n=1}^{\infty} \nu(\{x \in A \cap T^{-n}(A \cap B) : \tau_A(x) = n\}) \\
&= \widetilde{\mu}(B) - \nu(A \cap B) + \sum_{n=1}^{\infty} \nu(T_A^{-1}(A \cap B) \cap \tau_A^{-1}(n)) \\
&= \widetilde{\mu}(B) - \nu(A \cap B) + \nu(T_A^{-1}(A \cap B)) \\
&= \widetilde{\mu}(B).
\end{aligned}$$

Thus $\widetilde{\mu}$ is T-invariant when ν is T_A-invariant.

Moreover, if $B \subseteq A$ then (10.3) reduces to $\widetilde{\mu}(B) = \nu(A \cap B) = \nu(B)$. In particular, $\widetilde{\mu}(A) = \nu(A) = 1$, and hence $\widetilde{\mu}_A(B) = \frac{\widetilde{\mu}(B)}{\widetilde{\mu}(A)} = \widetilde{\mu}(B) = \nu(B)$. That is, $\widetilde{\mu} = \widetilde{\mu}_A = \nu$ on $(A, \mathcal{A}|_A)$.

Furthermore, (10.3) gives

$$\widetilde{\mu}\left(X \Big\backslash \bigcup_{k=1}^{\infty} T^{-k}(A)\right) = \sum_{n=0}^{\infty} \nu\left(\left\{x \in A \Big\backslash \bigcup_{k=1}^{\infty} T^{-(n+k)}(A) : \tau_A(x) > n\right\}\right) = \sum_{n=0}^{\infty} \nu(\emptyset) = 0.$$

Thus A is absorbing with respect to $\widetilde{\mu}$. In addition, this shows that $\widetilde{\mu}$ is σ-finite when $\widetilde{\mu}(T^{-k}(A)) < \infty$ for all $k \in \mathbb{N}$. Since $\widetilde{\mu}$ is T-invariant, this condition is equivalent to $\widetilde{\mu}(A) < \infty$. And we saw above that $\widetilde{\mu}(A) = 1$. Thus $\widetilde{\mu}$ is σ-finite. □

Remark 10.2.2. It follows from (10.5) and (10.6) that $\lim_{n\to\infty} \mu(B_n) = 0$. In particular, taking $B = A$, we get that

$$\lim_{n\to\infty} \mu\left(T^{-n}(A) \Big\backslash \bigcup_{k=0}^{n-1} T^{-k}(A)\right) = 0.$$

Theorem 10.2.1 raises some interesting questions. Among others, in the second part of the theorem, is the induced measure $\widetilde{\mu}$ unique? In Exercise 10.5.1, you will learn that this is generally not the case. However, we now prove that uniqueness prevails when the backward orbit of the set A covers the space almost everywhere.

Proposition 10.2.3. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. If μ is a T-invariant measure that admits an absorbing set A, then $\mu = \widetilde{\mu}$ if ν is replaced by μ in formula* (10.3).
(*Formally, $\mu = \mu(A) \cdot \widetilde{\mu}$, where $\widetilde{\mu}$ is given by* (10.3) *with $\nu = \mu_A$.*)

Proof. Let $B \in \mathcal{A}$ be such that $\mu(B) < \infty$. For every $j \geq 0$, let

$$B_j := T^{-j}(B) \Big\backslash \bigcup_{k=0}^{j} T^{-k}(A).$$

Observe that $\mu(B_j) \leq \mu(T^{-j}(B)) = \mu(B) < \infty$ for every $j \geq 0$. As μ is T-invariant and $T^{-1}(B_{j-1}) \supseteq B_j$ for all $j \in \mathbb{N}$, we get for every $n \in \mathbb{N}$ that

$$\begin{aligned}
\mu(B\backslash A) - \mu(B_n) &= \sum_{j=1}^{n} [\mu(B_{j-1}) - \mu(B_j)] = \sum_{j=1}^{n} [\mu(T^{-1}(B_{j-1})) - \mu(B_j)] \\
&= \sum_{j=1}^{n} \mu(T^{-1}(B_{j-1})\backslash B_j) \\
&= \sum_{j=1}^{n} \mu\left(\left(T^{-j}(B) \Big\backslash \bigcup_{k=1}^{j} T^{-k}(A)\right) \Big\backslash \left(T^{-j}(B) \Big\backslash \bigcup_{k=0}^{j} T^{-k}(A)\right)\right) \\
&= \sum_{j=1}^{n} \mu\left(A \cap T^{-j}(B) \Big\backslash \bigcup_{k=1}^{j} T^{-k}(A)\right).
\end{aligned}$$

Then

$$\mu(B) - \mu(B_n) = \mu(B \cap A) + \mu(B \backslash A) - \mu(B_n) = \sum_{j=0}^{n} \mu\left(A \cap T^{-j}(B) \backslash \bigcup_{k=1}^{j} T^{-k}(A) \right).$$

Replacing ν by μ in formula (10.3), we deduce that

$$\widetilde{\mu}(B) = \sum_{j=0}^{\infty} \mu\left(A \cap T^{-j}(B) \backslash \bigcup_{k=1}^{j} T^{-k}(A) \right) = \mu(B) - \lim_{n\to\infty} \mu(B_n).$$

Therefore, in order to complete the proof, we need to show that

$$\lim_{n\to\infty} \mu(B_n) = 0.$$

Fix $\varepsilon > 0$. Since A is absorbing, we have that $\mu(X \backslash \bigcup_{k=0}^{\infty} T^{-k}(A)) = 0$. Equivalently,

$$\mu\left(X \Big\backslash \bigcup_{k=0}^{\infty} A^{(k)} \right) = 0, \tag{10.7}$$

where

$$A^{(k)} := T^{-k}(A) \Big\backslash \bigcup_{j=0}^{k-1} T^{-j}(A).$$

The usefulness of the $A^{(k)}$s lies in their mutual disjointness. Indeed, relation (10.7) implies that $\mu(B \backslash \bigcup_{k=0}^{\infty} A^{(k)}) = 0$. Then

$$\mu(B) = \mu\left(B \cap \bigcup_{k=0}^{\infty} A^{(k)} \right) = \mu\left(\bigcup_{k=0}^{\infty} (B \cap A^{(k)}) \right) = \sum_{k=0}^{\infty} \mu(B \cap A^{(k)}). \tag{10.8}$$

Since $\mu(B) < \infty$, there exists $\ell_\varepsilon \in \mathbb{N}$ so large that

$$\sum_{\ell=\ell_\varepsilon+1}^{\infty} \mu(B \cap A^{(\ell)}) < \frac{\varepsilon}{2}. \tag{10.9}$$

Relation (10.7) also ensures that $\mu(B_n \backslash \bigcup_{k=0}^{\infty} A^{(k)}) = 0$ for every $n \geq 0$. So, like for B,

$$\mu(B_n) = \sum_{k=0}^{\infty} \mu(B_n \cap A^{(k)}).$$

But

$$\begin{aligned} B_n \cap A^{(k)} &= \left[T^{-n}(B) \Big\backslash \bigcup_{i=0}^{n} T^{-i}(A) \right] \cap \left[T^{-k}(A) \Big\backslash \bigcup_{j=0}^{k-1} T^{-j}(A) \right] \\ &= [T^{-n}(B) \cap T^{-k}(A)] \Big\backslash \bigcup_{i=0}^{\max\{n,k-1\}} T^{-i}(A) \end{aligned}$$

$$= \begin{cases} \emptyset & \text{if } k \le n \\ T^{-n}(B \cap T^{-(k-n)}(A)) \setminus \bigcup_{i=0}^{k-1} T^{-i}(A) & \text{if } k > n \end{cases}$$

$$= \begin{cases} \emptyset & \text{if } k \le n \\ [T^{-n}(B \cap T^{-(k-n)}(A)) \setminus \bigcup_{i=n}^{k-1} T^{-i}(A)] \setminus \bigcup_{i=0}^{n-1} T^{-i}(A) & \text{if } k > n \end{cases}$$

$$= \begin{cases} \emptyset & \text{if } k \le n \\ T^{-n}(B \cap A^{(k-n)}) \setminus \bigcup_{i=0}^{n-1} T^{-i}(A) & \text{if } k > n. \end{cases}$$

Consequently,

$$\mu(B_n) = \sum_{k>n} \mu(B_n \cap A^{(k)}) = \sum_{k=n+1}^{\infty} \mu\left(T^{-n}(B \cap A^{(k-n)}) \Big\backslash \bigcup_{i=0}^{n-1} T^{-i}(A) \right) = \sum_{\ell=1}^{\infty} \mu(B_n^{(\ell)}),$$

where

$$B_n^{(\ell)} := T^{-n}(B \cap A^{(\ell)}) \Big\backslash \bigcup_{i=0}^{n-1} T^{-i}(A) \subseteq T^{-(n+\ell)}(A) \Big\backslash \bigcup_{i=0}^{n+\ell-1} T^{-i}(A) = A^{(n+\ell)}.$$

Remark 10.2.2 asserts that $\lim_{N\to\infty} \mu(A^{(N)}) = 0$. Then

$$\lim_{n\to\infty} \mu(B_n^{(\ell)}) = 0, \quad \forall \ell \in \mathbb{N}. \tag{10.10}$$

As $\mu(B_n^{(\ell)}) \le \mu(T^{-n}(B \cap A^{(\ell)})) = \mu(B \cap A^{(\ell)})$ for each $\ell, n \in \mathbb{N}$, it follows from (10.9) that

$$\sum_{\ell=\ell_\varepsilon+1}^{\infty} \mu(B_n^{(\ell)}) < \frac{\varepsilon}{2}, \quad \forall n \in \mathbb{N}.$$

By (10.10), there exists $n_\varepsilon \in \mathbb{N}$ so large that for all $1 \le \ell \le \ell_\varepsilon$ and all $n \ge n_\varepsilon$,

$$\mu(B_n^{(\ell)}) \le \frac{\varepsilon}{2\ell_\varepsilon}.$$

So, for all $n \ge n_\varepsilon$, we have

$$\mu(B_n) = \sum_{\ell=1}^{\infty} \mu(B_n^{(\ell)}) = \sum_{\ell=1}^{\ell_\varepsilon} \mu(B_n^{(\ell)}) + \sum_{\ell=\ell_\varepsilon+1}^{\infty} \mu(B_n^{(\ell)}) \le \sum_{\ell=1}^{\ell_\varepsilon} \frac{\varepsilon}{2\ell_\varepsilon} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus $\lim_{n\to\infty} \mu(B_n) = 0$ and the proof is complete for sets B of finite measure.

Now, let $B \in \mathcal{A}$ be any set. Since $\mu(A^{(k)}) \le \mu(T^{-k}(A)) = \mu(A) < \infty$ for all $k \ge 0$, the sets $(B \cap A^{(k)})_{k=0}^{\infty}$ are of finite measure. Then the first part of this proof shows that $\mu(B \cap A^{(k)}) = \widetilde{\mu}(B \cap A^{(k)})$. By (10.8) and the mutual disjointness of the $A^{(k)}$s, we conclude that

$$\mu(B) = \sum_{k=0}^{\infty} \mu(B \cap A^{(k)}) = \sum_{k=0}^{\infty} \widetilde{\mu}(B \cap A^{(k)}) = \widetilde{\mu}\left(B \cap \bigcup_{k=0}^{\infty} A^{(k)} \right) = \widetilde{\mu}(B).$$

The last equality follows from the fact that $\widetilde{\mu}(X \setminus \bigcup_{k=0}^{\infty} A^{(k)}) = 0$ according to Theorem 10.2.1. So $\mu = \widetilde{\mu}$. □

Corollary 10.2.4. *Let $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ be an ergodic conservative measure-preserving transformation and any $A \in \mathcal{A}$ with $0 < \mu(A) < \infty$. Then $\mu = \widetilde{\mu}$ if ν is replaced by μ in formula* (10.3).

Proof. According to Corollary 10.1.14, any $A \in \mathcal{A}$ with $0 < \mu(A) < \infty$ is absorbing with respect to μ and Proposition 10.2.3 applies to any such A. □

Let $\varphi : X \to \mathbb{R}$ be a measurable function and $n \in \mathbb{N}$. Recall from Definition 8.2.10 that the nth *Birkhoff sum* of φ at a point $x \in X$ is

$$S_n\varphi(x) = \sum_{j=0}^{n-1} \varphi(T^j(x)).$$

Let $A \in \mathcal{A}$. Define the function $\varphi_A : A \to \mathbb{R}$ by the formula

$$\varphi_A(x) = S_{\tau_A(x)}\varphi(x). \tag{10.11}$$

In the next proposition, we describe properties that φ_A inherits from φ.

Proposition 10.2.5. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. Let also $\varphi : X \to \mathbb{R}$ be a measurable function. If μ is a T-invariant measure and A is an absorbing set with respect to μ, then:*
(a) *If $\varphi \in L^1(\mu)$, then $\varphi_A \in L^1(\mu_A)$.*
(b) *If $\varphi \geq 0$ or $\varphi \in L^1(\mu)$, then*

$$\int_A \varphi_A \, d\mu_A = \frac{1}{\mu(A)} \int_X \varphi \, d\mu.$$

If, in addition, T is conservative and ergodic, then the above two statements apply to all sets $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$.

Proof. Suppose first that $\varphi = \mathbb{1}_B$ for some $B \in \mathcal{A}$ such that $0 < \mu(B) < \infty$. In view of Proposition 10.2.3, we have

$$\begin{aligned}
\int_X \mathbb{1}_B \, d\mu = \mu(B) &= \sum_{k=0}^{\infty} \mu(\{x \in A \cap T^{-k}(B) : \tau_A(x) > k\}) \\
&= \sum_{n=1}^{\infty} \sum_{j=0}^{n-1} \mu(\{x \in A \cap T^{-j}(B) : \tau_A(x) = n\}) \\
&= \sum_{n=1}^{\infty} \sum_{j=0}^{n-1} \int_{\tau_A^{-1}(n)} \mathbb{1}_{T^{-j}(B)} \, d\mu \\
&= \sum_{n=1}^{\infty} \sum_{j=0}^{n-1} \int_{\tau_A^{-1}(n)} \mathbb{1}_B \circ T^j \, d\mu
\end{aligned}$$

$$= \sum_{n=1}^{\infty} \int_{\tau_A^{-1}(n)} S_n \mathbb{1}_B \, d\mu$$

$$= \sum_{n=1}^{\infty} \int_{\tau_A^{-1}(n)} \varphi_A \, d\mu$$

$$= \int_A \varphi_A \, d\mu = \mu(A) \int_A \varphi_A \, d\mu_A,$$

and we are done with this case. If $\varphi : X \to \mathbb{R}$ is a simple measurable function, that is, $\varphi = \sum_{i=1}^n a_i \varphi^{(i)}$, where all $a_i \in \mathbb{R}$, $1 \le i \le n$, and all $\varphi^{(i)}$, $1 \le i \le n$, are characteristic functions of some measurable sets with positive and finite measures, then

$$\int_X \varphi \, d\mu = \sum_{i=1}^n a_i \int_X \varphi^{(i)} \, d\mu$$

$$= \mu(A) \sum_{i=1}^n a_i \int_A \varphi_A^{(i)} \, d\mu_A$$

$$= \mu(A) \int_A \sum_{i=1}^n a_i \varphi_A^{(i)} \, d\mu_A$$

$$= \mu(A) \int_A \left(\sum_{i=1}^n a_i \varphi^{(i)} \right)_A d\mu_A$$

$$= \mu(A) \int_A \varphi_A \, d\mu_A,$$

and we are done in this case as well. The next case is to consider an arbitrary nonnegative measurable function $\varphi : X \to [0, \infty)$. Then φ is the pointwise limit of an increasing sequence of nonnegative simple measurable functions $(\varphi^{(n)})_{n=1}^{\infty}$. It is easy to see that φ_A is the pointwise limit of the increasing sequence of nonnegative measurable functions $(\varphi_A^{(n)})_{n=1}^{\infty}$. Applying twice the monotone convergence theorem (Theorem A.1.35), we then get that

$$\int_X \varphi \, d\mu = \lim_{n\to\infty} \int_X \varphi^{(n)} \, d\mu = \lim_{n\to\infty} \mu(A) \int_A \varphi_A^{(n)} \, d\mu_A = \mu(A) \int_A \lim_{n\to\infty} \varphi_A^{(n)} \, d\mu_A = \mu(A) \int_A \varphi_A \, d\mu_A.$$

We are also done in this case. Since $|\varphi_A| \le |\varphi|_A$, this in particular shows that if $\varphi \in L^1(\mu)$, then $\varphi_A \in L^1(\mu_A)$. Moreover, writing $\varphi = \varphi^+ - \varphi^-$, where $\varphi^+ = \max\{\varphi, 0\}$ and $\varphi^- = \max\{-\varphi, 0\}$, both functions φ^+ and φ^- are in $L^1(\mu)$ when φ is and

$$\int_X \varphi \, d\mu = \int_X (\varphi^+ - \varphi^-) \, d\mu = \mu(A) \int_A (\varphi_A^+ - \varphi_A^-) \, d\mu_A = \mu(A) \int_A \varphi_A \, d\mu_A. \qquad \square$$

Observe that if $\varphi \equiv 1$, then $\varphi_A \equiv \tau_A$. As an immediate consequence of the previous proposition, we obtain that the average of the first return time to a set is inversely proportional to the relative measure of that set in the space.

Lemma 10.2.6 (Kac's lemma). *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. If μ is a T-invariant measure and A is an absorbing set with respect to μ, then*

$$\int_A \tau_A \, d\mu_A = \frac{\mu(X)}{\mu(A)}.$$

In particular:

(a) *The measure μ is finite ($\mu(X) < \infty$) if and only if*

$$\int_A \tau_A \, d\mu_A < \infty.$$

(b) *If μ is a probability measure, then*

$$\int_A \tau_A \, d\mu_A = \frac{1}{\mu(A)}.$$

If, in addition, T is conservative and ergodic, then the above statements apply to every set $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$.

We now study the transmission of ergodicity between T and T_A.

Proposition 10.2.7. *Let $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ be a measure-preserving transformation and $A \in \mathcal{A}$ with $0 < \mu(A) < \infty$.*

If $T : X \to X$ is ergodic and conservative with respect to μ, then $T_A : A \to A$ is ergodic with respect to μ_A.

Conversely, if A is absorbing with respect to μ and $T_A : A \to A$ is ergodic with respect to μ_A, then $T : X \to X$ is ergodic with respect to μ.

Proof. First, suppose that $T : X \to X$ is ergodic and conservative. Let $C \subseteq A$ be completely T_A-invariant and assume that $\mu_A(C) > 0$. This latter assumption implies that $\mu(C) > 0$. By Theorem 10.1.11, we know that $\mu(X \backslash C_\infty) = 0$. But since $T_A^{-1}(C) = C$, we also have that $T_A^{-1}(A \backslash C) = A \backslash C$. Therefore, $A \backslash C \subseteq X \backslash C_\infty$, and hence $\mu(A \backslash C) = 0$. So $\mu_A(A \backslash C) = 0$ and T_A is ergodic.

In order to prove the converse, suppose that $T_A : A \to A$ is ergodic with respect to μ_A and let $B \in \mathcal{A}$ be such that $T^{-1}(B) = B$ and $\mu(B) > 0$. Suppose also that A is absorbing with respect to μ. Then $\mu(X \backslash \bigcup_{n=0}^{\infty} T^{-n}(A)) = 0$ and there exists $k \geq 0$ such that $\mu(B \cap T^{-k}(A)) > 0$. Therefore,

$$\mu(B \cap A) = \mu(T^{-k}(B \cap A)) = \mu(T^{-k}(B) \cap T^{-k}(A)) = \mu(B \cap T^{-k}(A)) > 0.$$

Recall that $(B\cap A)_\infty = \bigcap_{k=1}^\infty \bigcup_{n=k}^\infty T^{-n}(B\cap A)$ and write $(B\cap A)_\infty^{T_A} := \bigcap_{k=1}^\infty \bigcup_{n=k}^\infty T_A^{-n}(B\cap A)$. Since μ_A is a probability measure and $T_A : A \to A$ is ergodic, Corollary 10.1.12 states that $\mu_A(A\backslash(B\cap A)_\infty^{T_A}) = 0$ and this implies that

$$1 = \mu_A\big((B\cap A)_\infty^{T_A}\big) \le \mu_A\left(\bigcup_{n=0}^\infty T_A^{-n}(B\cap A)\right) \le \mu_A\left(A\cap \bigcup_{n=0}^\infty T^{-n}(B\cap A)\right) \le 1.$$

This means that $\mu(A\backslash \bigcup_{n=0}^\infty T^{-n}(B\cap A)) = 0$. Using the T-invariance of μ, it follows that $\mu(\bigcup_{k=0}^\infty T^{-k}(A)\backslash \bigcup_{n=0}^\infty T^{-n}(B\cap A)) = 0$. By hypothesis, $\mu(X\backslash \bigcup_{k=0}^\infty T^{-k}(A)) = 0$. Then $\mu(X\backslash \bigcup_{n=0}^\infty T^{-n}(B\cap A)) = 0$. Consequently,

$$\mu(X\backslash B) = \mu\left(X\backslash \bigcup_{n=0}^\infty T^{-n}(B)\right) \le \mu\left(X\backslash \bigcup_{n=0}^\infty T^{-n}(B\cap A)\right) = 0.$$

Hence, $\mu(X\backslash B) = 0$, and thus T is ergodic. □

10.3 Ergodic theorems

Birkhoff's ergodic theorem (Theorem 8.2.11 and Corollaries 8.2.14–8.2.15) concerns measure-preserving dynamical systems acting on probability spaces. Its ramifications are manifold. We studied some of them in the last two chapters. In this section, we use it with the inducing procedure described in the previous section to prove Hopf's ergodic theorem, which holds for measure-preserving transformations of σ-finite measure spaces.

But first, as a straightforward consequence of Birkhoff's ergodic theorem, we have the following two useful facts.

Proposition 10.3.1. *Let $T : X \to X$ be an ergodic measure-preserving transformation of a probability space $(X, \mathcal{A}, \mu)$. Fix $A \in \mathcal{A}$ such that $\mu(A) > 0$. For every $x \in X$, let $(k_n(x))_{n=1}^\infty$ be the sequence of successive times at which the iterates of x visit the set A. Then*

$$\lim_{n\to\infty} \frac{k_{n+1}(x)}{k_n(x)} = 1 \quad \textit{for } \mu\textit{-a. e. } x \in X.$$

Proof. Note that $S_{k_n(x)}\mathbb{1}_A(x) = n$. It follows from the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14) that for μ-a. e. $x \in X$,

$$\begin{aligned}\lim_{n\to\infty} \frac{k_{n+1}(x)}{k_n(x)} &= \lim_{n\to\infty}\left(\frac{n}{k_n(x)}\cdot\frac{k_{n+1}(x)}{n+1}\right)\\ &= \lim_{n\to\infty}\frac{1}{k_n(x)}S_{k_n(x)}\mathbb{1}_A(x)\cdot\frac{1}{\lim_{n\to\infty}\frac{1}{k_{n+1}(x)}S_{k_{n+1}(x)}\mathbb{1}_A(x)}\\ &= \mu(A)\cdot\frac{1}{\mu(A)} = 1.\end{aligned}$$ □

Proposition 10.3.2. *Let $T : X \to X$ be an ergodic measure-preserving transformation of a probability space $(X, \mathcal{A}, \mu)$. If $f \in L^1(\mu)$, then*

$$\lim_{n\to\infty} \frac{1}{n} f(T^n(x)) = 0 \quad \textit{for } \mu\textit{-a.e. } x \in X.$$

Proof. It follows from the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14) that for μ-a. e. $x \in X$,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} f(T^n(x)) &= \lim_{n\to\infty} \frac{1}{n+1} f(T^n(x)) = \lim_{n\to\infty} \frac{1}{n+1} (S_{n+1} f(x) - S_n f(x)) \\ &= \lim_{n\to\infty} \frac{1}{n+1} S_{n+1} f(x) - \lim_{n\to\infty} \frac{1}{n+1} S_n f(x) \\ &= \lim_{n\to\infty} \frac{1}{n+1} S_{n+1} f(x) - \lim_{n\to\infty} \frac{1}{n} S_n f(x) = 0. \end{aligned}$$ □

The following result is an application of Birkhoff's ergodic theorem to the ergodic theory of transformations preserving σ-finite measures.

Theorem 10.3.3 (Hopf's ergodic theorem). *Let $T : X \to X$ be an ergodic and conservative measure-preserving transformation of a σ-finite measure space $(X, \mathcal{A}, \mu)$. If $f, g \in L^1(\mu)$ and $\int_X g \, d\mu \neq 0$, then*

$$\lim_{n\to\infty} \frac{S_n f(x)}{S_n g(x)} = \frac{\int_X f \, d\mu}{\int_X g \, d\mu} \quad \textit{for } \mu\textit{-a.e. } x \in X.$$

Proof. (Note: We strongly recommend that the reader work on Exercise 10.5.2 before examining this proof.) Since μ is σ-finite, there are mutually disjoint sets $\{X_j\}_{j=1}^{\infty}$ such that $0 < \mu(X_j) < \infty$ for all $j \in \mathbb{N}$ and $\bigcup_{j=1}^{\infty} X_j = X$.

Fix $j \in \mathbb{N}$. Since T is measure-preserving and conservative, Corollary 10.1.8 affirms that $\mu(X_j \backslash (X_j)_\infty) = 0$. Thus the first return time to X_j and the first return map to X_j are well-defined by (10.1) and (10.2), respectively. Let $\tau_j := \tau_{X_j}$ and $T_j := T_{X_j}$. Let also $\varphi : X \to \mathbb{R}$ and set $\varphi_j := \varphi_{X_j}$ per (10.11). Given $x \in X_j$, let $S^j_n \varphi_j(x) := \sum_{i=0}^{n-1} \varphi_j(T^i_j(x))$. For every $k \in \mathbb{N}$, let $j_k(x)$ be the largest integer $n \geq 0$ such that $\sum_{i=0}^{n-1} \tau_j(T^i_j(x)) \leq k$. Then

$$S_k \varphi(x) = S^j_{j_k(x)} \varphi_j(x) + S_{\Delta k(x)} \varphi\big(T_j^{j_k(x)}(x)\big),$$

where $\Delta k(x) := k - \sum_{i=0}^{j_k(x)-1} \tau_j(T^i_j(x)) \geq 0$. Consequently,

$$\frac{1}{j_k(x)} S_k \varphi(x) = \frac{1}{j_k(x)} S^j_{j_k(x)} \varphi_j(x) + \frac{1}{j_k(x)} S_{\Delta k(x)} \varphi\big(T_j^{j_k(x)}(x)\big). \tag{10.12}$$

But

$$\left| \frac{1}{j_k(x)} S_{\Delta k(x)} \varphi\big(T_j^{j_k(x)}(x)\big) \right| \leq \frac{1}{j_k(x)} S_{\Delta k(x)} |\varphi|\big(T_j^{j_k(x)}(x)\big) \leq \frac{1}{j_k(x)} |\varphi|_j\big(T_j^{j_k(x)}(x)\big). \tag{10.13}$$

Let $\mu_j := \mu_{X_j}$. Since T is measure-preserving, ergodic and conservative, Theorem 10.2.1 and Proposition 10.2.7 assert that T_j is measure-preserving, conservative and ergodic with respect to μ_j. Moreover, Proposition 10.2.5 states that if $\varphi \in L^1(\mu)$, then $|\varphi|_j \in L^1(\mu_j)$. It follows from Proposition 10.3.2 (with $T = T_j$, $\mu = \mu_j$, and $f = |\varphi|_j$) that the right-hand side of (10.13) approaches 0 for μ-a. e. $x \in X_j$, and so does the left-hand side:

$$\lim_{k\to\infty}\left|\frac{1}{j_k(x)}S_{\Delta k(x)}\varphi(T_j^{j_k(x)}(x))\right| = 0 \quad \text{for } \mu\text{-a. e. } x \in X_j.$$

It ensues from this, (10.12), the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14) (with $T = T_j$, $\mu = \mu_j$ and $\varphi = f_j, g_j$) and Proposition 10.2.5 that for μ-a. e. $x \in X_j$,

$$\frac{S_k f(x)}{S_k g(x)} = \frac{\frac{1}{j_k(x)}S_k f(x)}{\frac{1}{j_k(x)}S_k g(x)} = \frac{\frac{1}{j_k(x)}S^j_{j_k(x)}f_j(x) + \frac{1}{j_k(x)}S_{\Delta k(x)}f(T_j^{j_k(x)}(x))}{\frac{1}{j_k(x)}S^j_{j_k(x)}g_j(x) + \frac{1}{j_k(x)}S_{\Delta k(x)}g(T_j^{j_k(x)}(x))} \underset{k\to\infty}{\longrightarrow} \frac{\int_{X_j} f_j\,d\mu_j}{\int_{X_j} g_j\,d\mu_j} = \frac{\int_X f\,d\mu}{\int_X g\,d\mu}.$$

Since $\bigcup_{j=1}^{\infty} X_j = X$, the conclusion holds for μ-a. e. $x \in X$. □

The following result rules out any hope for an ergodic theorem closer to Birkhoff's ergodic theorem in the case of infinite measures.

Corollary 10.3.4. *Let $T : X \to X$ be an ergodic and conservative measure-preserving transformation of a σ-finite and infinite measure space $(X, \mathcal{A}, \mu)$. If $f \in L^1(\mu)$, then*

$$\lim_{n\to\infty}\frac{1}{n}S_n f(x) = 0 \quad \text{for } \mu\text{-a. e. } x \in X.$$

Proof. Since μ is σ-finite and $\mu(X) = \infty$, there exists a sequence $(A_k)_{k=1}^{\infty}$ of measurable sets such that

$$0 < \mu(A_k) < \infty,\ \forall k \in \mathbb{N} \quad \text{and} \quad \lim_{k\to\infty}\mu(A_k) = \infty. \tag{10.14}$$

As $|f| \in L^1(\mu)$, we deduce from Hopf's ergodic theorem (Theorem 10.3.3) that for every $k \in \mathbb{N}$,

$$\limsup_{n\to\infty}\frac{1}{n}S_n|f| \le \limsup_{n\to\infty}\frac{S_n|f|}{S_n\mathbb{1}_{A_k}} = \frac{\int_X |f|\,d\mu}{\int_X \mathbb{1}_{A_k}\,d\mu} = \frac{\|f\|_1}{\mu(A_k)} \quad \mu\text{-a. e.}$$

So, by (10.14), $\limsup_{n\to\infty}\frac{1}{n}S_n|f| = 0$ μ-a. e. Since $|\frac{1}{n}S_n f| \le \frac{1}{n}S_n|f|$, we are done. □

As a matter of fact, as the next two results show, Hopf's ergodic theorem precludes the existence of even weaker forms of Birkhoff's ergodic theorem in the case of infinite invariant measures. Indeed, for all $f \in L^1(\mu)$ Corollary 10.3.4 of Hopf's ergodic theorem implies that

$$\lim_{n\to\infty}\frac{1}{a_n}S_n f(x) = 0 \quad \text{for } \mu\text{-a. e. } x \in X$$

if there exists $C > 0$ such that $a_n \geq Cn$ for all n. We will now show that there are no constants $a_n > 0$ such that

$$\lim_{n\to\infty} \frac{1}{a_n} S_n f(x) = \int_X f\, d\mu \quad \text{for } \mu\text{-a. e. } x \in X, \quad \forall f \in L^1(\mu).$$

We will accomplish this in two steps. The first step will require the following proposition.

Proposition 10.3.5. *Let $T : X \to X$ be an ergodic and conservative measure-preserving transformation of a probability space $(X, \mathcal{A}, \mu)$. Let $a : [0, \infty) \to [0, \infty)$ be continuous, strictly increasing, and satisfying $\frac{a(x)}{x} \searrow 0$ as $x \nearrow \infty$. If $\int_X a(|f|)\, d\mu < \infty$, then*

$$\lim_{n\to\infty} \frac{a(|S_n f(x)|)}{n} = 0 \quad \text{for } \mu\text{-a. e. } x \in X.$$

Proof. An outline of a proof can be found in Exercise 10.5.3. □

In the first step, a sequence $(a(n))_{n=1}^{\infty}$ will be imposed properties that mimic those of the function a in Proposition 10.3.5. We will show that the outcome takes the form of a dichotomy: $\liminf_{n\to\infty} \frac{1}{a(n)} S_n f(x)$ is either 0 or ∞, for μ-a. e. $x \in X$ for all $f \in L^1_+(\mu) := \{f \in L^1(\mu) : f \geq 0 \text{ and } \int_X f\, d\mu > 0\}$.

Theorem 10.3.6. *Let $T : X \to X$ be an ergodic and conservative measure-preserving transformation of a σ-finite and infinite measure space $(X, \mathcal{A}, \mu)$. Let $(a(n))_{n=1}^{\infty}$ be a sequence such that*

$$a(n) \nearrow \infty \quad \text{and} \quad \frac{a(n)}{n} \searrow 0 \quad \text{as } n \nearrow \infty.$$

(a) *If there exists $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$ and $\int_A a(\tau_A(x))\, d\mu(x) < \infty$, then*

$$\lim_{n\to\infty} \frac{1}{a(n)} S_n f(x) = \infty \quad \text{for } \mu\text{-a. e. } x \in X, \quad \forall f \in L^1_+(\mu).$$

(b) *Otherwise,*

$$\liminf_{n\to\infty} \frac{1}{a(n)} S_n f(x) = 0 \quad \text{for } \mu\text{-a. e. } x \in X, \quad \forall f \in L^1_+(\mu).$$

Proof. (a) Suppose that $A \in \mathcal{A}$ satisfies $0 < \mu(A) < \infty$ and $\int_A a(\tau_A(x))\, d\mu(x) < \infty$. Clearly, the set

$$I := \left\{x \in X : \lim_{n\to\infty} \frac{S_n \mathbb{1}_A(x)}{a(n)} = \infty\right\}$$

is T-invariant. As T is ergodic, either $\mu(I) = 0$ or $\mu(X \backslash I) = 0$. We will show that I contains μ-a. e. $x \in A$. This will allow us to conclude that $\mu(I) \geq \mu(A) > 0$, and hence $\mu(X \backslash I) = 0$.

Since T is measure-preserving and conservative, Corollary 10.1.8 affirms that $\mu(A\backslash A_\infty) = 0$. Thus the nth return time to A and the nth return map to A are well-defined for all $n \in \mathbb{N}$. See Exercise 10.5.2. Since T is measure-preserving, ergodic, and conservative, Theorem 10.2.1 and Proposition 10.2.7 assert that T_A is measure-preserving, conservative and ergodic with respect to μ_A. As $\int_A a(\tau_A(x))\,d\mu_A(x) < \infty$, it follows from Proposition 10.3.5 (with $T = T_A$, $\mu = \mu_A$, and $f = \tau_A$) that

$$\lim_{n\to\infty} \frac{a(\tau_A^n(x))}{n} = \lim_{n\to\infty} \frac{a(S_n^{T_A}\tau_A(x))}{n} = 0 \quad \text{for } \mu\text{-a. e. } x \in A. \tag{10.15}$$

Since T is measure-preserving, ergodic and conservative, Theorem 10.1.11 states that $\mu(X\backslash A_\infty) = 0$. For every $x \in A_\infty$ and $n \in \mathbb{N}$, let $k_n(x)$ be the largest integer $k \geq 0$ such that $\tau_A^k(x) \leq n$. Since $S_{\tau_A^n}\mathbb{1}_A \equiv n$, we then have

$$\frac{S_n\mathbb{1}_A(x)}{a(n)} \geq \frac{S_{\tau_A^{k_n(x)}}\mathbb{1}_A(x)}{a(\tau_A^{k_n(x)+1}(x))} = \frac{k_n(x)}{a(\tau_A^{k_n(x)+1}(x))} = \frac{k_n(x)+1}{a(\tau_A^{k_n(x)+1}(x))}\frac{k_n(x)}{k_n(x)+1}. \tag{10.16}$$

Using (10.15), we deduce from (10.16) that the set I contains μ-a. e. $x \in A$. As pointed out at the beginning of the proof, this implies that $\mu(X\backslash I) = 0$. Let $f \in L_+^1(\mu)$. It ensues from Hopf's ergodic theorem (Theorem 10.3.3) that

$$\lim_{n\to\infty} \frac{S_nf(x)}{a(n)} = \lim_{n\to\infty} \frac{S_nf(x)}{S_n\mathbb{1}_A(x)} \lim_{n\to\infty} \frac{S_n\mathbb{1}_A(x)}{a(n)} = \frac{\int_X f\,d\mu}{\mu(A)} \lim_{n\to\infty} \frac{S_n\mathbb{1}_A(x)}{a(n)} = \infty$$

for μ-a. e. $x \in I$, that is, for μ-a. e. $x \in X$ since $\mu(X\backslash I) = 0$.

(b) Assume that there does not exist a set $B \in \mathcal{A}$ such that $0 < \mu(B) < \infty$ and $\int_B a(\tau_B(x))\,d\mu(x) < \infty$. Suppose for a contradiction that the conclusion of (b) is not satisfied. That is, suppose that there exist a set $A \in \mathcal{A}$ with $0 < \mu(A) < \infty$ and a function $f \in L_+^1(\mu)$ such that

$$F(x) := \liminf_{n\to\infty} \frac{1}{a(n)} S_nf(x) > 0 \quad \text{for } \mu\text{-a. e. } x \in A. \tag{10.17}$$

By Hopf's ergodic theorem, this actually holds for every $f \in L_+^1(\mu)$, with the same set A.

Moreover, since $\frac{a(n)}{n} \searrow$ as $n \nearrow \infty$ and $f \geq 0$, we have

$$\frac{S_nf}{a(n)} \circ T \leq \frac{1}{n} \cdot \frac{n}{a(n)} S_{n+1}f \leq \frac{1}{n} \cdot \frac{n+1}{a(n+1)} S_{n+1}f = \left(1 + \frac{1}{n}\right)\frac{S_{n+1}f}{a(n+1)}.$$

Using this inequality and Corollary 10.3.4, the function F satisfies

$$F \circ T = \sup_{n\in\mathbb{N}} \inf_{k\geq n} \frac{S_kf \circ T}{a(k)} \leq \sup_{n\in\mathbb{N}} \inf_{k\geq n} \left(1 + \frac{1}{k}\right)\frac{S_{k+1}f}{a(k+1)} = \sup_{n\in\mathbb{N}} \inf_{k\geq n} \frac{S_{k+1}f}{a(k+1)} = F$$

μ-a. e. on X. That is, F is μ-a. e. T-subinvariant, and hence F is constant μ-a. e. on X by the ergodicity of T (Exercise 8.5.47 generalizes to any measure space). By (10.17), we conclude that for every $f \in L^1_+(\mu)$ there is a constant $c = c(f) > 0$ such that

$$\liminf_{n\to\infty} \frac{1}{a(n)} S_n f(x) = c \quad \text{for } \mu\text{-a. e. } x \in A.$$

In particular, this applies to the function $\mathbb{1}_A$ on the set A, that is,

$$\liminf_{n\to\infty} \frac{1}{a(n)} S_n \mathbb{1}_A(x) = c > 0 \quad \text{for } \mu\text{-a. e. } x \in A.$$

According to Egorov's theorem (Theorem A.1.44), there is $B \in \mathcal{A}$ such that $B \subseteq A$, $\mu(B) > 0$, and over which the sequence $\left(\inf_{k\geq n} \frac{S_k \mathbb{1}_A}{a(k)}\right)_{k=1}^{\infty}$ converges uniformly to c. Consequently, there is $K \in \mathbb{N}$ such that

$$S_k \mathbb{1}_A(x) \geq \frac{c}{2} a(k), \quad \forall x \in B,\ \forall k \geq K.$$

Since $S_1 \mathbb{1}_A(x) = 1$ for all $x \in B$ and $f \geq 0$, we know that $S_k \mathbb{1}_A(x) \geq 1$ for all $k \in \mathbb{N}$ and all $x \in B$. Then there is $0 < \tilde{c} \leq c/2$ such that

$$S_k \mathbb{1}_A(x) \geq \tilde{c}\, a(k), \quad \forall x \in B,\ \forall k \in \mathbb{N}.$$

In particular,

$$S_{\tau_B(x)} \mathbb{1}_A(x) \geq \tilde{c}\, a(\tau_B(x)), \quad \forall x \in B.$$

It ensues from the ergodic case of Birkhoff's ergodic theorem (Corollary 8.2.14) (with $T = T_B$, $\mu = \mu_B$, and $\varphi = (\mathbb{1}_A)_B$) and Proposition 10.2.5 that for μ-a. e. $x \in B$,

$$\begin{aligned}
\frac{1}{n}\sum_{k=0}^{n-1} a\big(\tau_B(T_B^k(x))\big) &\leq \frac{1}{n\tilde{c}}\sum_{k=0}^{n-1} S_{\tau_B(T_B^k(x))}\mathbb{1}_A\big(T_B^k(x)\big) \\
&= \frac{1}{\tilde{c}\, n} S_{\tau_B^n(x)}\mathbb{1}_A(x) = \frac{1}{\tilde{c}} \cdot \frac{1}{n} S_n^{T_B}(\mathbb{1}_A)_B(x) \\
&\underset{n\to\infty}{\longrightarrow} \frac{1}{\tilde{c}} \int_B (\mathbb{1}_A)_B \, d\mu_B = \frac{1}{\tilde{c}} \frac{1}{\mu(B)} \int_X \mathbb{1}_A \, d\mu = \frac{1}{\tilde{c}} \frac{\mu(A)}{\mu(B)}.
\end{aligned}$$

Since μ_B is T_B-invariant, it follows that

$$\int_B a(\tau_B)\, d\mu = \frac{1}{n} \sum_{k=0}^{n-1} \int_B a(\tau_B) \circ T_B^k \, d\mu$$

for every $n \in \mathbb{N}$. Passing to the limit $n \to \infty$, we conclude that $\int_B a(\tau_B)\, d\mu \leq \frac{\mu(A)}{\tilde{c}\,\mu(B)} < \infty$, hence contradicting the hypothesis in (b). $\square$

In the second step, no restriction other than strict positivity will be put on the sequence $(a_n)_{n=1}^{\infty}$. We will construct a related sequence $(a(n))_{n=1}^{\infty}$ that satisfies the first step. We will again show that there are only two possibilities: either $\liminf_{n\to\infty} \frac{1}{a_n} S_n f(x) = 0$ or $\limsup_{n\to\infty} \frac{1}{a_n} S_n f(x) = \infty$, for μ-a. e. $x \in X$ for all $f \in L^1_+(\mu)$.

Theorem 10.3.7. *Let $T : X \to X$ be an ergodic and conservative measure-preserving transformation of a σ-finite and infinite measure space $(X, \mathcal{A}, \mu)$. Let $(a_n)_{n=1}^{\infty}$ be a sequence such that $a_n > 0$ for all $n \in \mathbb{N}$. Then*

(a) *either* $\liminf\limits_{n\to\infty} \dfrac{S_n f(x)}{a_n} = 0$ *for μ-a. e. $x \in X$, $\forall f \in L^1_+(\mu)$;*

(b) *or there is $n_k \nearrow \infty$ such that* $\lim\limits_{k\to\infty} \dfrac{S_{n_k} f(x)}{a_{n_k}} = \infty$ *for μ-a. e. $x \in X$, $\forall f \in L^1_+(\mu)$.*

Proof. If $(a_n)_{n=1}^{\infty}$ is bounded, then (b) holds. Indeed, let $f \in L^1_+(\mu)$. Since $\int_X f\,d\mu > 0$, there exist $\varepsilon > 0$ and $B \in \mathcal{A}$ such that $\mu(B) > 0$ and $f \geq \varepsilon$ on B. As T is conservative and ergodic, Theorem 10.1.11 affirms that $\mu(X \backslash B_\infty) = 0$. Therefore, for μ-a. e. $x \in X$ there exists a sequence $(n_k(x))_{k=1}^{\infty}$ such that $n_k(x) \nearrow \infty$ and $T^{n_k(x)}(x) \in B$. Hence, $S_{n_k(x)} f(x) \geq k\varepsilon$. In fact, $S_n f(x) \geq k\varepsilon$ for any $n \geq n_k(x)$ since $f \in L^1_+(\mu)$. Therefore, $\lim_{n\to\infty} S_n f(x) = \infty$. As $(a_n)_{n=1}^{\infty}$ is bounded, it follows that $\lim_{n\to\infty} \frac{S_n f(x)}{a_n} = \infty$ for μ-a. e. $x \in X$. So (b) holds with $(n_k)_{k=1}^{\infty} = (n)_{n=1}^{\infty}$ when $(a_n)_{n=1}^{\infty}$ is bounded.

We can thereby restrict our attention to the case $\limsup_{n\to\infty} a_n = \infty$. Suppose that (a) does not hold. That is, there exist a set $A \in \mathcal{A}$ with $\mu(A) > 0$ and a function $f \in L^1_+(\mu)$ such that

$$F(x) := \liminf_{n\to\infty} \frac{S_n f(x)}{a_n} > 0 \quad \text{for } \mu\text{-a. e. } x \in A. \tag{10.18}$$

By Hopf's ergodic theorem (Theorem 10.3.3), this actually holds for every $f \in L^1_+(\mu)$, with the same set A. Then for μ-a. e. $x \in A$,

$$0 \leq \limsup_{n\to\infty} \frac{a_n}{n} \leq \limsup_{n\to\infty} \frac{a_n}{S_n f(x)} \limsup_{n\to\infty} \frac{S_n f(x)}{n} = \left[\liminf_{n\to\infty} \frac{S_n f(x)}{a_n}\right]^{-1} \lim_{n\to\infty} \frac{S_n f(x)}{n} = 0$$

by (10.18) and Corollary 10.3.4. Thus $a_n = o(n)$ as $n \to \infty$. For every $n \in \mathbb{N}$, set

$$\overline{a}_n = \max_{1\leq k\leq n} a_k.$$

Clearly, $a_n \leq \overline{a}_n$ for all $n \in \mathbb{N}$ and $\overline{a}_n \nearrow \infty$ as $n \nearrow \infty$. Moreover, for each $n \in \mathbb{N}$ there is $1 \leq k(n) \leq n$ such that $\overline{a}_n = a_{k(n)}$. Note that $k(n) \to \infty$ as $n \to \infty$. Then

$$\liminf_{n\to\infty} \frac{S_n f}{a_n} \geq \liminf_{n\to\infty} \frac{S_n f}{\overline{a}_n} = \liminf_{n\to\infty} \frac{S_n f}{a_{k(n)}} \geq \liminf_{n\to\infty} \frac{S_{k(n)} f}{a_{k(n)}} \geq \liminf_{n\to\infty} \frac{S_n f}{a_n},$$

where the last inequality follows from the fact that the lim inf of a subsequence of a sequence is greater than or equal to the lim inf of the full sequence. By this and (10.18),

$$\liminf_{n\to\infty} \frac{S_n f}{\overline{a}_n} = \liminf_{n\to\infty} \frac{S_n f}{a_n} > 0 \quad \mu\text{-a. e. on } A, \quad \forall f \in L^1_+(\mu). \tag{10.19}$$

Next, set $b_n = \frac{\overline{a}_n}{n}$, and let $1 = n_0 < n_1 < \cdots$ be defined by

$$\{n_k\}_{k\in\mathbb{N}} = \{j \geq 2 : b_i > b_j,\ \forall 1 \leq i \leq j-1\}.$$

For every $k \geq 0$,

$$b_{n_k} > b_{n_{k+1}} \quad \text{and} \quad n_k b_{n_k} \leq n_{k+1} b_{n_{k+1}},$$

whence

$$0 < \frac{n_k}{n_{k+1}} \leq \frac{b_{n_{k+1}}}{b_{n_k}} < 1.$$

Thus there exists $\alpha_k \in (0,1]$ such that

$$\left(\frac{n_k}{n_{k+1}}\right)^{\alpha_k} = \frac{b_{n_{k+1}}}{b_{n_k}}.$$

Define

$$b(x) = \frac{b_{n_k} n_k^{\alpha_k}}{x^{\alpha_k}}, \quad x \in [n_k, n_{k+1}],\ k \in \mathbb{N}, \quad \textit{and} \quad a(x) = xb(x).$$

Evidently,

$$a(n_k) = \overline{a}_{n_k}, \quad \forall k \in \mathbb{N}.$$

By definition of the n_k's, we have that for $k \in \mathbb{N}$, $n \in [n_k, n_{k+1})$,

$$b_{n_k} \leq b_n \quad \text{and hence} \quad b(n) \leq b_n,$$

whereby

$$a(n) \leq \overline{a}_n, \quad \forall n \in \mathbb{N}.$$

Hence, following (10.19),

$$\liminf_{n\to\infty} \frac{S_n f}{a(n)} > 0 \quad \mu\text{-a. e. on } A, \quad \forall f \in L^1_+(\mu).$$

It is evident that

$$a(n) \nearrow \infty \quad \text{and} \quad \frac{a(n)}{n} \searrow 0 \quad \text{as } n \nearrow \infty.$$

So by Theorem 10.3.6,

$$\lim_{n\to\infty} \frac{S_n f}{a(n)} = \infty \quad \mu\text{-a. e. on } X, \quad \forall f \in L^1_+(\mu).$$

Then (b) follows since $a_{n_k} \leq \overline{a}_{n_k} = a(n_k)$. □

10.4 Absolutely continuous σ-finite invariant measures

In this section, we establish a very useful, relatively easy to verify, sufficient condition for a quasi-invariant probability measure to admit an absolutely continuous σ-finite invariant measure. This condition actually provides a σ-finite invariant measure equivalent to the original quasi-invariant probability measure. It goes back to the work of Marco Martens [44] and has been used many times, notably in [40]. It obtained its nearly final form in [70]. In contrast to Martens, where σ-compact metric spaces form the setting, the sufficient condition in [70] is stated for abstract measure spaces, and the proof uses the concept of Banach limit rather than weak convergence. In this section, we somewhat strengthen the assertions in [70].

We first identify the possible relations between σ-finite invariant measures with respect to which a system is ergodic and conservative. The following result is an analogue of Theorem 8.2.21.

Theorem 10.4.1. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. If μ_1 and μ_2 are σ-finite T-invariant measures and if T is ergodic and conservative with respect to both measures, then either μ_1 and μ_2 are mutually singular or else they coincide up to a positive multiplicative constant.*

Proof. We may assume that neither $\mu_1 \equiv 0$ nor $\mu_2 \equiv 0$. Since both measures are σ-finite, there is a sequence $(Y_n)_{n=1}^{\infty}$ of mutually disjoint measurable sets such that $\max\{\mu_1(Y_n), \mu_2(Y_n)\} < \infty$ for all $n \in \mathbb{N}$ and $X = \bigcup_{n=1}^{\infty} Y_n$.

First, suppose that $\mu_1(Y_n) > 0$ for some $n \in \mathbb{N}$ and that μ_1 and μ_2 coincide on Y_n up to a positive multiplicative constant. Without loss of generality, assume that $n = 1$ and that $\mu_1|_{Y_1} = \mu_2|_{Y_1}$. It immediately follows from Corollary 10.2.4 that $\mu_1 = \mu_2$, and we are done in this case.

Now, assume that μ_1 and μ_2 do not coincide on X up to any positive multiplicative constant. For each $n \in \mathbb{N}$, select a set Z_n in the following way:

(1) If $\mu_1(Y_n)\cdot\mu_2(Y_n) > 0$, it ensues from the previous case that $\mu_1|_{Y_n}$ and $\mu_2|_{Y_n}$ cannot be equal up to any positive multiplicative constant. Hence, $\mu_1|_{Y_n} \neq \mu_2|_{Y_n}$. Combining Proposition 10.2.7 and Theorem 8.2.21, we deduce that the measures $\mu_1|_{Y_n}$ and $\mu_2|_{Y_n}$ are mutually singular, that is, there is a measurable set $Z_n \subseteq Y_n$ such that $\mu_1(Z_n) = 0$ and $\mu_2(Y_n \backslash Z_n) = 0$.
(2) If $\mu_1(Y_n) = 0$, set $Z_n = Y_n$.
(3) Otherwise, set $Z_n = \emptyset$. (In this case, $\mu_2(Y_n) = 0$.)

Observe that $Z_n \subseteq Y_n$ for all $n \in \mathbb{N}$. Therefore, the sets $(Z_n)_{n=1}^{\infty}$ are mutually disjoint. Moreover, setting $Z := \bigcup_{n=1}^{\infty} Z_n \in \mathcal{A}$, it turns out that

$$\mu_1(Z) = \sum_{n=1}^{\infty} \mu_1(Z_n) = 0$$

while

$$\mu_2(X\backslash Z) = \sum_{n=1}^{\infty} \mu_2(Y_n\backslash Z) = \sum_{n=1}^{\infty} \mu_2(Y_n\backslash Z_n) = 0.$$

So the measures μ_1 and μ_2 are mutually singular. □

The preceding theorem allows us to derive the uniqueness of a σ-finite invariant measure which is absolutely continuous with respect to a given quasi-invariant measure, assuming that the transformation is ergodic and conservative with respect to the quasi-invariant measure.

Theorem 10.4.2. *Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and m a σ-finite quasi-T-invariant measure. If T is ergodic and conservative with respect to m then, up to a positive multiplicative constant, there exists at most one nonzero σ-finite T-invariant measure μ which is absolutely continuous with respect to m.*

Proof. Suppose that μ_1 and μ_2 are nonzero σ-finite T-invariant measures absolutely continuous with respect to m. Since m is ergodic and conservative, so are the measures μ_1 and μ_2. It follows from Theorem 10.4.1 that if μ_1 and μ_2 do not coincide up to a positive multiplicative constant, then they are mutually singular. But this means that there exists a measurable set $Y \subseteq X$ such that $\mu_1(Y) = 0$ and $\mu_2(X\backslash Y) = 0$. So

$$0 \le \mu_1\left(\bigcup_{n=0}^{\infty} T^{-n}(Y)\right) \le \sum_{n=0}^{\infty} \mu_1(T^{-n}(Y)) = \sum_{n=0}^{\infty} \mu_1(Y) = 0. \tag{10.20}$$

On the other hand, $\mu_2(Y) > 0$. Since $\mu_2 \ll m$, this implies that $m(Y) > 0$. Thus $m(X\backslash \bigcup_{n=0}^{\infty} T^{-n}(Y)) = 0$ by virtue of Theorem 10.1.11. Since $\mu_1 \ll m$, this forces $\mu_1(X\backslash \bigcup_{n=0}^{\infty} T^{-n}(Y)) = 0$. Along with (10.20), this gives that $\mu_1(X) = 0$. This contradicts the assumption that $\mu_1(X) \neq 0$. □

We now introduce the concept of Martens map.

Definition 10.4.3. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation. Let also m be a quasi-T-invariant probability measure. The transformation T is called a *Martens map* if it admits a countable family $\{X_n\}_{n=0}^{\infty}$ of subsets of X with the following properties:

(a) $X_n \in \mathcal{A}$, $\forall n \ge 0$.
(b) $m(X\backslash \bigcup_{n=0}^{\infty} X_n) = 0$.
(c) For all $m, n \ge 0$, there exists $j \ge 0$ such that $m(X_m \cap T^{-j}(X_n)) > 0$.
(d) For all $j \ge 0$, there exists $K_j \ge 1$ such that for all $A, B \in \mathcal{A}$ with $A \cup B \subseteq X_j$ and for all $n \ge 0$,

$$m(T^{-n}(A))m(B) \le K_j m(A) m(T^{-n}(B)).$$

(e) $\sum_{n=0}^{\infty} m(T^{-n}(X_0)) = \infty$.

(f) $T(\bigcup_{j=l}^{\infty} Y_j) \in \mathcal{A}$ for all $l \geq 0$, where $Y_j := X_j \setminus \bigcup_{i<j} X_i$.
(g) $\lim\limits_{l\to\infty} m(T(\bigcup_{j=l}^{\infty} Y_j)) = 0$.

The family $\{X_n\}_{n=0}^{\infty}$ is called a *Martens cover*.

Remark 10.4.4.
(1) Without loss of generality, condition (b) can be replaced by $\bigcup_{n=0}^{\infty} X_n = X$.
(2) Condition (c) imposes that $m(X_n) > 0$ for all $n \geq 0$.
(3) In light of Corollary 10.1.10, if T is conservative with respect to μ then condition (e) is fulfilled.
(4) In conditions (f-g), note that $\bigcup_{j=l}^{\infty} Y_j = \bigcup_{j=0}^{\infty} X_j \setminus \bigcup_{i<l} X_i \subseteq X \setminus \bigcup_{i<l} X_i$.
(5) If the map $T : X \to X$ is finite-to-one, then condition (g) is satisfied. For then, $\bigcap_{l=1}^{\infty} T(\bigcup_{j=l}^{\infty} Y_j) = \emptyset$.

Let l^{∞} denote the Banach space of all bounded real-valued sequences $x = (x_n)_{n=1}^{\infty}$ with norm $\|x\|_{\infty} := \sup_{n\in\mathbb{N}} |x_n|$. Recall that a *Banach limit* is a shift-invariant positive continuous linear functional $l_B : l^{\infty} \to \mathbb{R}$ which extends the usual limits. More precisely, for all sequences $x = (x_n)_{n=1}^{\infty}$, $y = (y_n)_{n=1}^{\infty} \in l^{\infty}$ and $\alpha, \beta \in \mathbb{R}$, the following properties hold:
(a) $l_B(\alpha x + \beta y) = \alpha\, l_B(x) + \beta\, l_B(y)$ (linearity).
(b) $\|l_B\| := \sup\{|l_B(x)| : \|x\|_{\infty} \leq 1\} < \infty$ (continuity/boundedness).
(c) If $x \geq 0$, that is, if $x_n \geq 0$ for all $n \in \mathbb{N}$, then $l_B(x) \geq 0$ (positivity).
(d) $l_B(\sigma(x)) = l_B(x)$, where $\sigma : l^{\infty} \to l^{\infty}$ is the (left) shift map defined by $(\sigma(x))_n = x_{n+1}$ for all $n \in \mathbb{N}$ (shift-invariance).
(e) If x is a convergent sequence, then $l_B(x) = \lim\limits_{n\to\infty} x_n$.

It follows from properties (a), (c) and (e) that a Banach limit also satisfies:
(f) $\liminf\limits_{n\to\infty} x_n \leq l_B(x) \leq \limsup\limits_{n\to\infty} x_n$.
(g) If $x \leq y$, that is, if $x_n \leq y_n$ for all $n \in \mathbb{N}$, then $l_B(x) \leq l_B(y)$.

As already announced, the main result of this section is the following.

Theorem 10.4.5. *Let $(X, \mathcal{A}, m)$ be a probability space and $T : X \to X$ a Martens map with Martens cover $\{X_j\}_{j=0}^{\infty}$ and for which m is quasi-T-invariant. Then there exists a σ-finite T-invariant measure μ equivalent to m on X. In addition, $0 < \mu(X_j) < \infty$, $\forall j \geq 0$.*

A measure μ with the above properties can be constructed as follows. Let $l_B : l^{\infty} \to \mathbb{R}$ be a Banach limit and let $Y_j := X_j \setminus \bigcup_{i<j} X_i$ for every $j \geq 0$. For each $A \in \mathcal{A}$, set

$$m_n(A) := \frac{\sum_{k=0}^{n} m(T^{-k}(A))}{\sum_{k=0}^{n} m(T^{-k}(X_0))}. \tag{10.21}$$

If $A \in \mathcal{A}$ and $A \subseteq Y_j$ for some $j \geq 0$, then $(m_n(A))_{n=1}^{\infty} \in l^{\infty}$ and set

$$\mu(A) := l_B((m_n(A))_{n=1}^{\infty}). \tag{10.22}$$

For a general $A \in \mathcal{A}$, set

$$\mu(A) := \sum_{j=0}^{\infty} \mu(A \cap Y_j).$$

If $(m_n(A))_{n=1}^{\infty} \in l^{\infty}$ for some $A \in \mathcal{A}$, then

$$\mu(A) = l_B\big((m_n(A))_{n=1}^{\infty}\big) - \lim_{l\to\infty} l_B\left(\left(m_n\Big(A \cap \bigcup_{j=l}^{\infty} Y_j\Big)\right)_{n=0}^{\infty}\right). \tag{10.23}$$

In particular, if $A \in \mathcal{A}$ is contained in a finite union of sets X_j, $j \geq 0$, then

$$\mu(A) = l_B\big((m_n(A))_{n=1}^{\infty}\big).$$

Finally, if T is ergodic and conservative with respect to m, then μ is unique up to a positive multiplicative constant and T is ergodic and conservative with respect to μ.

In order to prove Theorem 10.4.5, we need several lemmas.

Lemma 10.4.6. *Let $(Z, \mathcal{F})$ be a measurable space such that:*
(a) $Z = \bigcup_{j=0}^{\infty} Z_j$ *for some mutually disjoint sets $Z_j \in \mathcal{F}$; and*
(b) ν_j *is a finite measure on Z_j for each $j \geq 0$.*

Then the set function $\nu : \mathcal{F} \to [0, \infty]$ defined by

$$\nu(F) := \sum_{j=0}^{\infty} \nu_j(F \cap Z_j)$$

is a σ-finite measure on Z.

Proof. Clearly, $\nu(\emptyset) = 0$. Let $F \in \mathcal{F}$ and $\{F_n\}_{n=1}^{\infty}$ a partition of F into sets in $\mathcal{F}$. Then

$$\nu(F) = \sum_{j=0}^{\infty} \nu_j(F \cap Z_j) = \sum_{j=0}^{\infty} \nu_j\left(\bigcup_{n=1}^{\infty} (F_n \cap Z_j)\right) = \sum_{j=0}^{\infty}\sum_{n=1}^{\infty} \nu_j(F_n \cap Z_j) = \sum_{n=1}^{\infty}\sum_{j=0}^{\infty} \nu_j(F_n \cap Z_j) = \sum_{n=1}^{\infty} \nu(F_n),$$

where the order of summation could be changed since all terms involved are non-negative. Thus ν is a measure. Moreover, by definition, $Z = \bigcup_{j=0}^{\infty} Z_j$ and $\nu(Z_j) = \nu_j(Z_j) < \infty$ for all $j \geq 0$. Therefore, ν is σ-finite. □

From this point on, all lemmas rely on the same main hypotheses as Theorem 10.4.5.

Lemma 10.4.7. *For all $n, j \geq 0$ and all $A, B \in \mathcal{A}$ with $A \cup B \subseteq X_j$, we have*

$$m_n(A) m(B) \leq K_j \, m(A) \, m_n(B).$$

Proof. This follows directly from the definition of m_n and condition (d) of Definition 10.4.3. □

Lemma 10.4.8. *For every $j \geq 0$, we have $(m_n(X_j))_{n=1}^{\infty} \in l^{\infty}$ and $\mu(Y_j) \leq \mu(X_j) < \infty$.*

Proof. Fix $j \geq 0$. In virtue of condition (c) of Definition 10.4.3, there exists $q \geq 0$ such that $m(X_j \cap T^{-q}(X_0)) > 0$. By Lemma 10.4.7 and the definition of m_n, for all $n \geq 0$ we have that

$$\begin{aligned} m_n(Y_j) \leq m_n(X_j) &\leq K_j \frac{m(X_j)}{m(X_j \cap T^{-q}(X_0))} m_n(X_j \cap T^{-q}(X_0)) \\ &\leq K_j \frac{m(X_j)}{m(X_j \cap T^{-q}(X_0))} m_n(T^{-q}(X_0)) \\ &\leq K_j \frac{m(X_j)}{m(X_j \cap T^{-q}(X_0))} \frac{\sum_{k=0}^{n+q} m(T^{-k}(X_0))}{\sum_{k=0}^{n} m(T^{-k}(X_0))} \\ &= K_j \frac{m(X_j)}{m(X_j \cap T^{-q}(X_0))} \left[1 + \frac{\sum_{k=n+1}^{n+q} m(T^{-k}(X_0))}{\sum_{k=0}^{n} m(T^{-k}(X_0))}\right] \\ &\leq K_j \frac{m(X_j)}{m(X_j \cap T^{-q}(X_0))} \left[1 + \frac{q}{m(X_0)}\right]. \end{aligned} \tag{10.24}$$

Consequently, $(m_n(X_j))_{n=1}^{\infty} \in l^{\infty}$ and properties (g) and (e) of a Banach limit yield that

$$\mu(Y_j) \leq K_j \frac{m(X_j)}{m(X_j \cap T^{-q}(X_0))} \left[1 + \frac{q}{m(X_0)}\right] < \infty.$$

Since $X_j = \bigcup_{i=0}^{j} Y_i$ and the Y_i's are mutually disjoint, we deduce that

$$\mu(Y_j) \leq \sum_{i=0}^{j} \mu(X_j \cap Y_i) = \sum_{i=0}^{\infty} \mu(X_j \cap Y_i) =: \mu(X_j) \leq \sum_{i=0}^{j} \mu(Y_i) < \infty.$$ □

For every $j \geq 0$, set $\mu_j := \mu|_{Y_j}$.

Lemma 10.4.9. *For every $j \geq 0$ such that $\mu(Y_j) > 0$ and for every measurable set $A \subseteq Y_j$, we have*

$$K_j^{-1} \frac{\mu(Y_j)}{m(Y_j)} m(A) \leq \mu_j(A) \leq K_j \frac{\mu(Y_j)}{m(Y_j)} m(A).$$

Proof. This follows from the definition of μ, and by setting $B = Y_j$ in Lemma 10.4.7 and using properties (a) and (g) of a Banach limit. □

Lemma 10.4.10. *For each $j \geq 0$, μ_j is a finite measure on Y_j.*

Proof. Let $j \geq 0$. If $\mu_j(Y_j) = 0$, then the result is trivial. So assume that $\mu_j(Y_j) > 0$. Let $A \subseteq Y_j$ be a measurable set and $(A_k)_{k=1}^{\infty}$ a countable measurable partition of A. Using termwise operations on sequences, for every $l \in \mathbb{N}$ we have

$$\left(\sum_{k=1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty} - \sum_{k=1}^{l} (m_n(A_k))_{n=1}^{\infty} = \left(\sum_{k=1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty} - \left(\sum_{k=1}^{l} m_n(A_k)\right)_{n=1}^{\infty}$$
$$= \left(\sum_{k=l+1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty}.$$

It therefore follows from Lemma 10.4.7 (with $A = A_k$ and $B = Y_j$) that

$$\begin{aligned}\left\| \left(\sum_{k=1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty} - \sum_{k=1}^{l} (m_n(A_k))_{n=1}^{\infty} \right\|_{\infty} &= \left\| \left(\sum_{k=l+1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty} \right\|_{\infty} \\ &\leq \left\| \frac{K_j}{m(Y_j)} \left(m_n(Y_j) \sum_{k=l+1}^{\infty} m(A_k)\right)_{n=1}^{\infty} \right\|_{\infty} \\ &= \frac{K_j}{m(Y_j)} \left\| \left(m_n(Y_j) \sum_{k=l+1}^{\infty} m(A_k)\right)_{n=1}^{\infty} \right\|_{\infty}.\end{aligned}$$

Since $(m_n(Y_j))_{n=1}^{\infty} \in l^{\infty}$ by Lemma 10.4.8 and since $\lim_{l\to\infty} \sum_{k=l+1}^{\infty} m(A_k) = 0$, we conclude that

$$\lim_{l\to\infty} \left\| \left(\sum_{k=1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty} - \sum_{k=1}^{l} (m_n(A_k))_{n=1}^{\infty} \right\|_{\infty} = 0.$$

This means that

$$\left(\sum_{k=1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty} = \sum_{k=1}^{\infty} (m_n(A_k))_{n=1}^{\infty} \quad \text{in } l^{\infty}.$$

Hence, using the continuity of the Banach limit $l_B : l^{\infty} \to \mathbb{R}$, we get

$$\begin{aligned}\mu(A) = l_B\big((m_n(A))_{n=1}^{\infty}\big) &= l_B\left(\left(m_n\Big(\bigcup_{k=1}^{\infty} A_k\Big)\right)_{n=1}^{\infty}\right) = l_B\left(\left(\sum_{k=1}^{\infty} m_n(A_k)\right)_{n=1}^{\infty}\right) \\ &= \sum_{k=1}^{\infty} l_B\big((m_n(A_k))_{n=1}^{\infty}\big) = \sum_{k=1}^{\infty} \mu(A_k).\end{aligned}$$

So μ_j is countably additive. Also, $\mu_j(\emptyset) = 0$. Thus μ_j is a measure. By Lemma 10.4.8, this measure μ_j is finite. □

Combining Lemmas 10.4.6, 10.4.8, 10.4.9, and 10.4.10, and condition (b) of Definition 10.4.3, we obtain the following.

Lemma 10.4.11. *The set function* μ *is a* σ*-finite measure on* X *equivalent to* m*. Moreover,* $\mu(Y_j) \le \mu(X_j) < \infty$ *and* $\mu(X_j) > 0$ *for all* $j \ge 0$.

Lemma 10.4.12. *Formula* (10.23) *holds.*

Proof. Fix $A \in \mathcal{A}$ such that $(m_n(A))_{n=1}^{\infty} \in l^{\infty}$. For every $l \in \mathbb{N}$, we then have

$$l_B((m_n(A))_{n=1}^{\infty}) = l_B\left(\sum_{j=0}^{l}(m_n(A \cap Y_j))_{n=1}^{\infty}\right) + l_B\left(\left(m_n\Big(\bigcup_{j=l+1}^{\infty} A \cap Y_j\Big)\right)_{n=1}^{\infty}\right)$$

$$= \sum_{j=0}^{l} l_B((m_n(A \cap Y_j))_{n=1}^{\infty}) + l_B\left(\left(m_n\Big(A \cap \bigcup_{j=l+1}^{\infty} Y_j\Big)\right)_{n=1}^{\infty}\right).$$

Letting $l \to \infty$, we deduce that

$$l_B((m_n(A))_{n=1}^{\infty}) = \sum_{j=0}^{\infty} l_B((m_n(A \cap Y_j))_{n=1}^{\infty}) + \lim_{l\to\infty} l_B\left(\left(m_n\Big(A \cap \bigcup_{j=l+1}^{\infty} Y_j\Big)\right)_{n=1}^{\infty}\right)$$

$$= \sum_{j=0}^{\infty} \mu(A \cap Y_j) + \lim_{l\to\infty} l_B\left(\left(m_n\Big(A \cap \bigcup_{j=l}^{\infty} Y_j\Big)\right)_{n=1}^{\infty}\right)$$

$$= \mu(A) + \lim_{l\to\infty} l_B\left(\left(m_n\Big(A \cap \bigcup_{j=l}^{\infty} Y_j\Big)\right)_{n=1}^{\infty}\right).$$

This establishes formula (10.23). In particular, if $A \subseteq \bigcup_{j=0}^{k} X_j$ for some $k \in \mathbb{N}$, then $A \cap \bigcup_{j=l}^{\infty} Y_j \subseteq (\bigcup_{j=0}^{k} X_j) \cap (X \setminus \bigcup_{i<l} X_i) = \emptyset$ for all $l > k$. In that case, the equation above reduces to

$$l_B((m_n(A))_{n=1}^{\infty}) = \mu(A). \qquad \square$$

Lemma 10.4.13. *The* σ*-finite measure* μ *is* T*-invariant.*

Proof. Let $i \ge 0$ be such that $m(Y_i) > 0$. Fix a measurable set $A \subset Y_i$. By definition, $\mu(A) = l_B((m_n(A))_{n=1}^{\infty})$. Furthermore, for all $n \ge 0$ notice that

$$\left|m_n(T^{-1}(A)) - m_n(A)\right| = \frac{\left|m(T^{-(n+1)}(A)) - m(A)\right|}{\sum_{k=0}^{n} m(T^{-k}(X_0))} \le \frac{1}{\sum_{k=0}^{n} m(T^{-k}(X_0))}.$$

Thus $(m_n(T^{-1}(A)))_{n=1}^{\infty} \in l^{\infty}$ because $(m_n(A))_{n=1}^{\infty} \in l^{\infty}$. Moreover, by condition (e) of Definition 10.4.3, it follows from the above and properties (a), (e), and (g) of a Banach limit that $l_B((m_n(T^{-1}(A)))_{n=1}^{\infty}) = l_B((m_n(A))_{n=1}^{\infty}) = \mu(A)$.

Keep A a measurable subset of Y_i. Fix $l \in \mathbb{N}$. We then have

$$m_n\left(T^{-1}(A) \cap \bigcup_{j=l}^{\infty} Y_j\right) = \frac{\sum_{k=0}^{n} m(T^{-k}(T^{-1}(A) \cap \bigcup_{j=l}^{\infty} Y_j))}{\sum_{k=0}^{n} m(T^{-k}(X_0))}$$

$$\le \frac{\sum_{k=0}^{n} m(T^{-(k+1)}(A \cap T(\bigcup_{j=l}^{\infty} Y_j)))}{\sum_{k=0}^{n} m(T^{-k}(X_0))}$$

$$\leq m_{n+1}\Big(A \cap T\Big(\bigcup_{j=l}^{\infty} Y_j\Big)\Big) \cdot \frac{\sum_{k=0}^{n+1} m(T^{-k}(X_0))}{\sum_{k=0}^{n} m(T^{-k}(X_0))}$$

$$\leq K_i \frac{m_{n+1}(Y_i)}{m(Y_i)} \cdot m\Big(A \cap T\Big(\bigcup_{j=l}^{\infty} Y_j\Big)\Big) \cdot \frac{\sum_{k=0}^{n+1} m(T^{-k}(X_0))}{\sum_{k=0}^{n} m(T^{-k}(X_0))},$$

where the last inequality sign holds by Lemma 10.4.7 since $A \subseteq Y_i$. When $n \to \infty$, the last quotient on the right-hand side approaches 1. Therefore,

$$0 \leq l_B\Big(\Big(m_n\Big(T^{-1}(A) \cap \bigcup_{j=l}^{\infty} Y_j\Big)\Big)_{n=1}^{\infty}\Big) \leq K_i \frac{\mu(Y_i)}{m(Y_i)} m\Big(T\Big(\bigcup_{j=l}^{\infty} Y_j\Big)\Big).$$

Hence, by virtue of condition (g) of Definition 10.4.3,

$$0 \leq \lim_{l\to\infty} l_B\Big(\Big(m_n\Big(T^{-1}(A) \cap \bigcup_{j=l}^{\infty} Y_j\Big)\Big)_{n=1}^{\infty}\Big) \leq K_i \frac{\mu(Y_i)}{m(Y_i)} \lim_{l\to\infty} m\Big(T\Big(\bigcup_{j=l}^{\infty} Y_j\Big)\Big) = 0.$$

So

$$\lim_{l\to\infty} l_B\Big(\Big(m_n\Big(T^{-1}(A) \cap \bigcup_{j=l}^{\infty} Y_j\Big)\Big)_{n=1}^{\infty}\Big) = 0.$$

It thus follows from Lemma 10.4.12 that

$$\mu(T^{-1}(A)) = l_B((m_n(T^{-1}(A)))_{n=1}^{\infty}) = l_B((m_n(A))_{n=1}^{\infty}) = \mu(A).$$

For an arbitrary $A \in \mathcal{A}$, write $A = \bigcup_{j=0}^{\infty}(A \cap Y_j)$ and observe that

$$\mu(T^{-1}(A)) = \mu\Big(\bigcup_{j=0}^{\infty} T^{-1}(A \cap Y_j)\Big) = \sum_{j=0}^{\infty} \mu(T^{-1}(A \cap Y_j)) = \sum_{j=0}^{\infty} \mu(A \cap Y_j) = \mu(A). \qquad \square$$

Proof of Theorem 10.4.5. Combining Lemmas 10.4.8, 10.4.11, 10.4.12, and 10.4.13, with Theorems 10.4.2 and 10.1.11, we obtain Theorem 10.4.5. □

Remark 10.4.14. In the course of the proof of Theorem 10.4.5, we have shown that

$$0 < \inf\{m_n(A) : n \in \mathbb{N}\} \leq \sup\{m_n(A) : n \in \mathbb{N}\} < \infty$$

for all $j \geq 0$ and all measurable sets $A \subseteq X_j$ such that $m(A) > 0$.

10.5 Exercises

Exercise 10.5.1. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation for which there is a completely T-invariant set $A \in \mathcal{A}\backslash\{\emptyset, X\}$. Assume also that there exist a $T|_A$-invariant probability measure ν on $(A, \mathcal{A}|_A)$ and a $T|_{X\backslash A}$-invariant probability measure κ on $(X\backslash A, \mathcal{A}|_{X\backslash A})$. Set $\mu(B) = \nu(B \cap A) + \kappa(B\backslash A)$ for all $B \in \mathcal{A}$.

(a) Prove that μ is a T-invariant measure on $(X, \mathcal{A})$.
(b) Show that $0 < \mu(A) < \infty$ and $\mu(A \backslash A_\infty) = 0$.
(c) Deduce that $\mu_A = \nu$ and is T_A-invariant.
(d) Show that $\mu(X \backslash \bigcup_{k=0}^{\infty} T^{-k}(A)) \neq 0$.
(e) Let $\widetilde{\mu}$ be the measure induced by $\nu = \mu_A$.
  (1) Prove that $\mu(A) = 1 = \widetilde{\mu}(A)$.
  (2) Prove that $\mu(X \backslash A) = 1$ whereas $\widetilde{\mu}(X \backslash A) = 0$.
  (3) Conclude that there is no $c \in \mathbb{R}$ such that $\widetilde{\mu} = c\mu$. In particular, $\widetilde{\mu} \neq \mu$.

Exercise 10.5.2. Let $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ be a measurable transformation and μ a T-invariant measure. Fix $A \in \mathcal{A}$ such that $0 < \mu(A) < \infty$ and $\mu(A \backslash A_\infty) = 0$. So, from this point on, A will be identified with $A \cap A_\infty$.

The first return time function $\tau_A : A \to \mathbb{N}$ was defined in (10.1) while the first return map $T_A : A \to A$ was introduced in (10.2).

Similarly, for every $n \in \mathbb{N}$ the nth return time function $\tau_A^n : A \to \mathbb{N}$ is defined by

$$\tau_A^n(x) := \min\{k \in \mathbb{N} : \#\{1 \le j \le k : T^j(x) \in A\} = n\}.$$

The nth return map $T_A^n : A \to A$ is subsequently defined as

$$T_A^n(x) = T^{\tau_A^n(x)}(x).$$

Finally, let $\varphi : X \to \mathbb{R}$ be a measurable function. The function $\varphi_A : A \to \mathbb{R}$ was defined in (10.11). The nth *Birkhoff sum* of φ_A under T_A at a point $x \in A$ is denoted by $S_n^{T_A}\varphi_A : A \to \mathbb{R}$ and is naturally given by

$$S_n^{T_A}\varphi_A(x) = \sum_{i=0}^{n-1} \varphi_A(T_A^i(x)).$$

(a) Show that $T_A^n = T_A \circ T_A \ldots \circ T_A$, with n copies of T_A in the composition. In other words, show that the nth return map T_A^n is the usual n-time composition of the first return map T_A.
(b) Prove that $\tau_A^n(x) = \sum_{i=0}^{n-1} \tau_A(T_A^i(x)) = S_n^{T_A}\tau_A(x)$. In other terms, the nth return time is the sum of the first return times of the first n iterates of x that fall into A.
(c) Deduce that $\tau_A^{n+1}(x) - \tau_A^n(x) = \tau_A(T_A^n(x))$.
(d) Show that $S_{\tau_A^n(x)}\varphi(x) = \sum_{k=0}^{n-1} S_{\tau_A(T_A^k(x))}\varphi(T_A^k(x)) = S_n^{T_A}\varphi_A(x)$.
(e) Given $x \in A$ and $k \in \mathbb{N}$, let $n(x)$ be the largest integer $n \ge 0$ such that $\tau_A^n(x) \le k$. In other words, $n(x)$ is the number of times that the iterates of x visit A by time k, and $\tau_A^{n(x)}(x)$ is the last time at which an iterate of x falls into A prior to or at time k. Demonstrate that

$$S_k\varphi(x) = S_{n(x)}^{T_A}\varphi_A(x) + S_{\Delta k(x)}\varphi(T_A^{n(x)}(x)),$$

where $\Delta k(x) := k - \tau_A^{n(x)}(x) \ge 0$.

(f) Show that $|\varphi_A| \leq |\varphi|_A$.
(g) Prove that $S_{\Delta k(x)}\varphi(T_A^{n(x)}(x)) \leq |\varphi|_A(T_A^{n(x)}(x))$.

Exercise 10.5.3. In this exercise, you will give a proof of Proposition 10.3.5. You will first establish the proposition under the additional assumption that $a(0) = 0$.
(a) Prove that $a(x+y) \leq a(x) + a(y)$ for all $x, y \in [0, \infty)$.
(b) Show that

$$\lim_{M\to\infty} a\big(|f|\mathbb{1}_{\{|f|\geq M\}}(x)\big) = 0 \quad \text{for } \mu\text{-a.e. } x \in X.$$

(c) Deduce that

$$\lim_{M\to\infty} \int_X a\big(|f|\mathbb{1}_{\{|f|\geq M\}}\big)\, d\mu = 0.$$

(d) Fix $\varepsilon > 0$. From (c), identify two nonnegative functions $g, h \in L^1(\mu)$ such that

$$|f| = g + h, \quad \sup_{x\in X} g(x) < \infty, \quad \text{and} \quad \int_X a(h)\, d\mu < \varepsilon.$$

(e) Using (a) and (d), show that

$$\limsup_{n\to\infty} \frac{a\big(|S_n f(x)|\big)}{n} \leq \int_X a(h)\, d\mu.$$

(f) Conclude that

$$\lim_{n\to\infty} \frac{a\big(|S_n f(x)|\big)}{n} = 0 \quad \text{for } \mu\text{-a.e. } x \in X.$$

You will now establish the proposition without the assumption that $a(0) = 0$. Choose $m > 0$ such that $a(m) = \alpha m$ for some $0 < \alpha < 1$, and define

$$\tilde{a}(x) = \begin{cases} \alpha x & \text{if } 0 \leq x \leq m \\ a(x) & \text{if } x \geq m. \end{cases}$$

(g) Show that $\tilde{a}$ is continuous, strictly increasing, $\tilde{a} \equiv a$ on $[m, \infty)$, $\frac{\tilde{a}(x)}{x} \searrow 0$ as $x \nearrow \infty$, and $\tilde{a}(0) = 0$.
(h) Deduce that

$$\lim_{n\to\infty} \frac{\tilde{a}\big(|S_n f(x)|\big)}{n} = 0 \quad \text{for } \mu\text{-a.e. } x \in X.$$

(i) Conclude that

$$\lim_{n\to\infty} \frac{a\big(|S_n f(x)|\big)}{n} = 0 \quad \text{for } \mu\text{-a.e. } x \in X.$$

Exercise 10.5.4. Let $X = \overline{\mathbb{R}}$, λ be the Lebesgue measure on $\overline{\mathbb{R}}$, and $T : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ be Boole's transformation defined by $T(x) = x - \frac{1}{x}$.
(a) Prove that λ is T-invariant.
(b) Show that T is conservative.

Exercise 10.5.5. For each $i = 1, 2$, let $T_i : (X_i, \mathcal{A}_i, \mu_i) \to (X_i, \mathcal{A}_i, \mu_i)$ be a measurable transformation of a probability space $(X_i, \mathcal{A}_i, \mu_i)$ such that μ_i is quasi-T_i-invariant. The Cartesian product transformation $T = T_1 \times T_2$ is defined on the product space $(X, \mathcal{A}, \mu) := (X_1 \times X_2, \mathcal{A}_1 \times \mathcal{A}_2, m_1 \times m_2)$ by $T(x_1, x_2) = (T_1(x_1), T_2(x_2))$ (cf. Example 8.1.13). Prove that if T_1 is measure-preserving and T_2 is conservative, then $T_1 \times T_2$ is conservative.

Exercise 10.5.6. Generalize Exercise 8.5.47 to every measure space.

Exercise 10.5.7. Let $S : (X_S, \mathcal{A}_S, \mu_S) \to (X_S, \mathcal{A}_S, \mu_S)$ be a measurable transformation of a σ-finite measure space $(X_S, \mathcal{A}_S, \mu_S)$. Let $\varphi : X_S \to \mathbb{N}$ be a measurable function. The *Kakutani tower* over T with height function φ is the transformation T of the σ-finite measure space $(X_T, \mathcal{A}_T, \mu_T)$ defined as follows:

$$X_T = \{(x, n) : x \in X_S \text{ and } n \le \varphi(x)\},$$
$$\mathcal{A}_T = \sigma(\{A \times \{n\} : n \in \mathbb{N} \text{ and } A = A_S \cap \varphi^{-1}([n, \infty)) \text{ for some } A_S \in \mathcal{A}_S\}),$$
$$\mu_T(A \times \{n\}) = \mu_S(A),$$

and

$$T(x, n) = \begin{cases} (S(x), \varphi(S(x))) & \text{if } n = 1, \\ (x, n - 1) & \text{if } n \ge 2. \end{cases}$$

Prove the following statements:
(a) If μ_S is quasi-S-invariant, then μ_T is quasi-T-invariant.
(b) If S is conservative, then so is T.
(c) $T_{X_S \times \{1\}}(x, 1) \equiv (S(x), 1)$ and $\varphi_{X_S \times \{1\}}(x, 1) \equiv \varphi(x)$.
(d) If μ_S is S-invariant, then μ_T is T-invariant.
(e) If S is ergodic, then so is T.

11 Topological pressure

In the forthcoming three chapters (the third one being part of the second volume), we introduce and extensively deal with the fundamental concepts and results of thermodynamic formalism, including topological pressure, the so-called variational principle, equilibrium states, and Gibbs states.

Thermodynamic formalism originated from the works of David Ruelle in the late 1960s. Ruelle's motivation came from statistical mechanics, particularly glass lattices. The foundations, classical concepts, and theorems of thermodynamic formalism were developed throughout the 1970s in the early works of Ruelle [59, 60], Rufus Bowen [10], Peter Walters [75], and Yakov Sinai [68]. More recent and modern expositions can be found in [57, 61, 76], among others. Also worthy of mention is Michal Misiurewicz's paper [49], where an elegant, short, and simple proof of the variational principle was provided. This is the proof we shall reproduce in Chapter 12.

In Chapter 11, we define and investigate the properties of topological pressure. Like topological entropy, this is a topological concept and a topological conjugacy invariant. We further give Bowen's characterization of pressure in terms of separated and spanning sets, which however requires a metric.

In Chapter 12, we relate topological pressure with Kolmogorov–Sinai metric entropy by proving the variational principle, the very cornerstone of thermodynamic formalism. This principle naturally leads to the concepts of equilibrium states and measures of maximal entropy. We deal with those at length in that chapter, particularly through the problem of existence of equilibrium states. Among others, we will show that under a continuous potential every expansive system admits an equilibrium state.

Whereas in Chapters 11 and 12 we consider general topological dynamical systems, in Chapter 13 we will restrict our attention to transitive open distance expanding maps. We will introduce therein the concept of Gibbs measures for such maps. We will also prove their existence and uniqueness for Hölder continuous potentials. We will further demonstrate that they coincide with equilibrium states for these potentials, concomitantly establishing the uniqueness of equilibrium states.

Gibbs states are measures with particularly fine and transparent stochastic properties, such as the central limit theorem, the law of the iterated logarithm, and exponential decay of correlations. They are used to describe long-term unstable behaviors of typical orbits of a given dynamical system having Gibbs states. Apart from such direct application to dynamical systems, thermodynamic formalism was used to provided a full account of SRB (Sinai–Ruelle–Bowen) measures for Axiom A diffeomorphisms and flows, and many other dynamical systems. Via Bowen's formula, thermodynamic formalism is also an indispensable tool for studying the fractal geometry of nonlinear smooth dynamical systems, particularly conformal and holomorphic ones. This is at the heart of Chapter 16 and a guiding theme in all subsequent chapters.

The main tool for dealing with Gibbs states is the transfer, also frequently called Perron–Frobenius, operator. We prove many of its functional analytic properties;

https://doi.org/10.1515/9783110702682-011

among them its almost periodicity if acting on the Banach space of continuous functions and its quasi–compactness if acting on the Banach space of Hölder continuous functions. In addition to quasi-compactness, we show that this operator has only one (and real) eigenvalue of maximal modulus and this eigenvalue is simple. The corresponding eigenfunction turns out to be the Radon–Nikodym derivative of the invariant Gibbs state with respect to the eigenmeasure of the dual transfer operator.

11.1 Definition of topological pressure via open covers

Recall that a topological dynamical system $T : X \to X$ is a self-transformation T of a compact metrizable space X. Let $\varphi : X \to \mathbb{R}$ be a real-valued continuous function. In the context of topological pressure (for historical, physical reasons), such a function is usually referred to as a *potential*.

The topological pressure of a potential is defined in two stages. This may seem surprising since topological entropy was defined in three stages in Chapter 7. However, the main reason for defining topological entropy in three stages was so as to later mirror it in the definition of measure-theoretic entropy in Chapter 9. Indeed, the first stage might just as well have been omitted and we would then have proceeded immediately to the second stage by defining $Z_n(\mathcal{U})$ directly and deriving its properties without relying upon the fact that $Z_n(\mathcal{U}) = Z_1(\mathcal{U}^n)$. The first stage for topological entropy proves to be useless when defining topological pressure since, as we will shortly discover, $Z_n(\varphi,\mathcal{U}) \neq Z_1(\varphi,\mathcal{U}^n)$ in general.

11.1.1 First stage: pressure of a potential relative to an open cover

Let us first recall the notion of Birkhoff (or ergodic) sum (cf. Definition 8.2.10). The nth *Birkhoff sum* of a potential φ at a point $x \in X$ is given by

$$S_n\varphi(x) := \sum_{j=0}^{n-1} \varphi(T^j(x)).$$

This is the sum of the values of the potential φ at the first n iterates of x under T.

Definition 11.1.1. For every $Y \subseteq X$ and $n \in \mathbb{N}$, define

$$\overline{S}_n\varphi(Y) := \sup_{y\in Y} S_n\varphi(y) \quad \text{and} \quad \underline{S}_n\varphi(Y) := \inf_{y\in Y} S_n\varphi(y).$$

Now, let $\mathcal{U}$ be an open cover of X. The minimum number $Z_n(\mathcal{U})$ of elements of $\mathcal{U}^n$ required to cover X (cf. Definition 7.2.6) generalises to the real numbers $Z_n(\varphi,\mathcal{U})$ and $z_n(\varphi,\mathcal{U})$ as follows.

Definition 11.1.2. Let $T : X \to X$ be a topological dynamical system and let $\varphi : X \to \mathbb{R}$ be a potential. Let $\mathcal{U}$ be an open cover of X. For each $n \in \mathbb{N}$, define the nth level functions (sometimes called *partition functions*) of $\mathcal{U}$ with respect to the potential φ by

$$Z_n(\varphi,\mathcal{U}) := \inf\left\{\sum_{V\in\mathcal{V}} e^{\overline{S}_n\varphi(V)} : \mathcal{V} \text{ is a subcover of } \mathcal{U}^n\right\}$$

and

$$z_n(\varphi,\mathcal{U}) := \inf\left\{\sum_{V\in\mathcal{V}} e^{\underline{S}_n\varphi(V)} : \mathcal{V} \text{ is a subcover of } \mathcal{U}^n\right\}.$$

Remark 11.1.3.
(a) It is sufficient to take the infimum over all **finite** subcovers since the exponential function takes only positive values and every subcover has itself a finite subcover. However, this infimum may not be achieved if $\mathcal{U}$ is infinite.
(b) In general, $Z_n(\varphi,\mathcal{U}) \neq Z_1(\varphi,\mathcal{U}^n)$ and $z_n(\varphi,\mathcal{U}) \neq z_1(\varphi,\mathcal{U}^n)$.
(c) If $\varphi \equiv 0$, then $Z_n(0,\mathcal{U}) = z_n(0,\mathcal{U}) = Z_n(\mathcal{U})$ for all $n \in \mathbb{N}$ and any open cover $\mathcal{U}$ of X.
(d) If $\varphi \equiv c$ for some $c \in \mathbb{R}$, then $Z_n(c,\mathcal{U}) = z_n(c,\mathcal{U}) = e^{nc}Z_n(\mathcal{U})$ for all $n \in \mathbb{N}$ and every open cover $\mathcal{U}$ of X.
(e) For all open covers $\mathcal{U}$ of X and all $n \in \mathbb{N}$, we have

$$e^{n\inf\varphi}Z_n(\mathcal{U}) \leq Z_n(\varphi,\mathcal{U}) \leq e^{n\sup\varphi}Z_n(\mathcal{U})$$

and

$$e^{n\inf\varphi}Z_n(\mathcal{U}) \leq z_n(\varphi,\mathcal{U}) \leq e^{n\sup\varphi}Z_n(\mathcal{U}).$$

We have seen in Chapter 7 that the functions $Z_n(\cdot)$, $n \in \mathbb{N}$, behave well with respect to all cover operations. In particular, it was observed in Lemma 7.2.8 that they respect the refinement relation, that is, if $\mathcal{U} \prec \mathcal{V}$ then $Z_n(\mathcal{U}) \leq Z_n(\mathcal{V})$ for every $n \in \mathbb{N}$. This is not necessarily true for the partition functions $Z_n(\varphi,\cdot)$, $n \in \mathbb{N}$. The corresponding inequality is more intricate. It involves the concept of oscillation.

Definition 11.1.4. The *oscillation* of φ with respect to an open cover $\mathcal{U}$ is defined to be

$$\operatorname{osc}(\varphi,\mathcal{U}) := \sup_{U\in\mathcal{U}} \sup_{x,y\in U} \left|\varphi(y) - \varphi(x)\right|.$$

Note that $\operatorname{osc}(\varphi,\cdot) \leq 2\|\varphi\|_\infty$. Also, $\operatorname{osc}(c,\cdot) = 0$ for all $c \in \mathbb{R}$.

Lemma 11.1.5. *For every $n \in \mathbb{N}$ and every open cover $\mathcal{U}$ of X,*

$$\operatorname{osc}(S_n\varphi,\mathcal{U}^n) \leq n\operatorname{osc}(\varphi,\mathcal{U}).$$

Proof. Let $V := U_0 \cap \cdots \cap T^{-(n-1)}(U_{n-1}) \in \mathcal{U}^n$ and $x, y \in V$. For each $0 \le j < n$, we have that $T^j(x), T^j(y) \in U_j \in \mathcal{U}$. Hence, for all $0 \le j < n$,

$$\left|\varphi(T^j(x)) - \varphi(T^j(y))\right| \le \operatorname{osc}(\varphi, \mathcal{U}).$$

Therefore,

$$\left|S_n\varphi(x) - S_n\varphi(y)\right| \le \sum_{j=0}^{n-1}\left|\varphi(T^j(x)) - \varphi(T^j(y))\right| \le n \operatorname{osc}(\varphi, \mathcal{U}).$$

Since this is true for all $x, y \in V$ and all $V \in \mathcal{U}^n$, the result follows. □

We now look at the relationship between the Z_n's and the z_n's.

Lemma 11.1.6. *For all $n \in \mathbb{N}$ and all open covers U of X, the following inequalities hold:*

$$z_n(\varphi, \mathcal{U}) \le Z_n(\varphi, \mathcal{U}) \le e^{n \operatorname{osc}(\varphi, \mathcal{U})} z_n(\varphi, \mathcal{U}).$$

Proof. The left inequality is obvious. To ascertain the right one, let $\mathcal{W}$ be a subcover of $\mathcal{U}^n$. Then

$$\begin{aligned}\sum_{W \in \mathcal{W}} e^{\overline{S}_n\varphi(W)} &\le \exp\Big(\sup_{W \in \mathcal{W}}\big[\,\overline{S}_n\varphi(W) - \underline{S}_n\varphi(W)\,\big]\Big) \sum_{W \in \mathcal{W}} e^{\underline{S}_n\varphi(W)} \\ &\le e^{\operatorname{osc}(S_n\varphi, \mathcal{U}^n)} \sum_{W \in \mathcal{W}} e^{\underline{S}_n\varphi(W)} \le e^{n \operatorname{osc}(\varphi, \mathcal{U})} \sum_{W \in \mathcal{W}} e^{\underline{S}_n\varphi(W)}.\end{aligned}$$

Taking the infimum over all subcovers of $\mathcal{U}^n$ on both sides results in the right inequality. □

In the next few results, we will see that the Z_n's and the z_n's have distinct properties.

Lemma 11.1.7. *If $\mathcal{U} \prec \mathcal{V}$, then for all $n \in \mathbb{N}$ we have that*

$$Z_n(\varphi, \mathcal{U}) e^{-n \operatorname{osc}(\varphi, \mathcal{U})} \le Z_n(\varphi, \mathcal{V}) \quad \textit{while} \quad z_n(\varphi, \mathcal{U}) \le z_n(\varphi, \mathcal{V}).$$

Proof. Fix $n \in \mathbb{N}$. Let $i : \mathcal{V} \to \mathcal{U}$ be a map such that $V \subseteq i(V)$ for all $V \in \mathcal{V}$. The map i induces a map $i_n : \mathcal{V}^n \to \mathcal{U}^n$ in the following way. For every $W := V_0 \cap \cdots \cap T^{-(n-1)}(V_{n-1}) \in \mathcal{V}^n$, define

$$i_n(W) := i(V_0) \cap \cdots \cap T^{-(n-1)}\big(i(V_{n-1})\big).$$

Observe that $W \subseteq i_n(W) \in \mathcal{U}^n$ for all $W \in \mathcal{V}^n$. Moreover, if $x \in W$ and $y \in i_n(W)$, then for each $0 \le j < n$ we have that $T^j(x) \in V_j \subseteq i(V_j) \ni T^j(y)$. So $T^j(x), T^j(y) \in i(V_j)$ for all $0 \le j < n$. Hence, $x, y \in i_n(W) \in \mathcal{U}^n$, and thus

$$S_n\varphi(x) \ge S_n\varphi(y) - \operatorname{osc}(S_n\varphi, \mathcal{U}^n).$$

Taking the supremum over all $x \in W$ on the left-hand side and over all $y \in i_n(W)$ on the right-hand side yields

$$\overline{S}_n\varphi(W) \geq \overline{S}_n\varphi(i_n(W)) - \operatorname{osc}(S_n\varphi, \mathcal{U}^n).$$

Now, let $\mathcal{W}$ be a subcover of $\mathcal{V}^n$. Then $i_n(\mathcal{W}) := \{i_n(W) : W \in \mathcal{W}\}$ is a subcover of $\mathcal{U}^n$. Therefore,

$$\begin{aligned}\sum_{W\in\mathcal{W}} e^{\overline{S}_n\varphi(W)} &\geq e^{-\operatorname{osc}(S_n\varphi,\mathcal{U}^n)} \sum_{W\in\mathcal{W}} e^{\overline{S}_n\varphi(i_n(W))} \\ &\geq e^{-\operatorname{osc}(S_n\varphi,\mathcal{U}^n)} \sum_{Y\in i_n(\mathcal{W})} e^{\overline{S}_n\varphi(Y)} \\ &\geq e^{-\operatorname{osc}(S_n\varphi,\mathcal{U}^n)} Z_n(\varphi,\mathcal{U}).\end{aligned}$$

Taking the infimum over all subcovers $\mathcal{W}$ of $\mathcal{V}^n$ on the left-hand side and using Lemma 11.1.5, we conclude that

$$Z_n(\varphi,\mathcal{V}) \geq e^{-\operatorname{osc}(S_n\varphi,\mathcal{U}^n)} Z_n(\varphi,\mathcal{U}) \geq e^{-n\operatorname{osc}(\varphi,\mathcal{U})} Z_n(\varphi,\mathcal{U}).$$

The proof of the inequality for the z_n's is left to the reader. □

In Lemma 7.2.9, we saw that the functions $Z_n(\cdot)$ are submultiplicative with respect to the join operation; in other words, $Z_n(\mathcal{U} \vee \mathcal{V}) \leq Z_n(\mathcal{U})Z_n(\mathcal{V})$ for all $n \in \mathbb{N}$. The corresponding property for the functions $Z_n(\varphi,\cdot)$ and $z_n(\varphi,\cdot)$ is the following.

Lemma 11.1.8. *Let $\mathcal{U}$ and $\mathcal{V}$ be open covers of X and let $n \in \mathbb{N}$. Then*

$$Z_n(\varphi,\mathcal{U} \vee \mathcal{V}) \leq \min\{Z_n(\varphi,\mathcal{U}) \cdot Z_n(\mathcal{V}), Z_n(\mathcal{U}) \cdot Z_n(\varphi,\mathcal{V})\}$$

and

$$z_n(\varphi,\mathcal{U} \vee \mathcal{V}) \leq \min\{e^{n\operatorname{osc}(\varphi,\mathcal{U})} z_n(\varphi,\mathcal{U}) \cdot Z_n(\mathcal{V}), Z_n(\mathcal{U}) \cdot e^{n\operatorname{osc}(\varphi,\mathcal{V})} z_n(\varphi,\mathcal{V})\}.$$

Proof. The proof is left to the reader as an exercise. □

We have also seen in Lemma 7.2.10 that the sequence $(Z_n(\mathcal{U}))_{n=1}^\infty$ is submultiplicative. This property is retained by the $Z_n(\varphi,\mathcal{U})$'s but generally not by the $z_n(\varphi,\mathcal{U})$'s (see Exercise 11.5.2).

Lemma 11.1.9. *Given an open cover $\mathcal{U}$ of X, the sequence $(Z_n(\varphi,\mathcal{U}))_{n=1}^\infty$ is submultiplicative.*

Proof. Fix $m, n \in \mathbb{N}$, let $\mathcal{V}$ be a subcover of $\mathcal{U}^m$ and $\mathcal{W}$ a subcover of $\mathcal{U}^n$. Note that $\mathcal{V} \vee T^{-m}(\mathcal{W})$ is a subcover of $\mathcal{U}^{m+n}$ since it is a cover and

$$\mathcal{V} \vee T^{-m}(\mathcal{W}) \subseteq \mathcal{U}^m \vee T^{-m}(\mathcal{U}^n) = \mathcal{U}^{m+n}.$$

Take arbitrary $V \in \mathcal{V}$ and $W \in \mathcal{W}$. Then for every $x \in V \cap T^{-m}(W)$, we have $x \in V$ and $T^m(x) \in W$, and hence

$$S_{m+n}\varphi(x) = S_m\varphi(x) + S_n\varphi\big(T^m(x)\big) \le \overline{S}_m\varphi(V) + \overline{S}_n\varphi(W).$$

Taking the supremum over all $x \in V \cap T^{-m}(W)$, we deduce that

$$\overline{S}_{m+n}\varphi\big(V \cap T^{-m}(W)\big) \le \overline{S}_m\varphi(V) + \overline{S}_n\varphi(W).$$

Therefore,

$$\begin{aligned} Z_{m+n}(\varphi,\mathcal{U}) &\le \sum_{E\in\mathcal{V}\vee T^{-m}(\mathcal{W})} e^{\overline{S}_{m+n}\varphi(E)} \\ &\le \sum_{V\in\mathcal{V}}\sum_{W\in\mathcal{W}} e^{\overline{S}_{m+n}\varphi(V\cap T^{-m}(W))} \\ &\le \sum_{V\in\mathcal{V}}\sum_{W\in\mathcal{W}} e^{\overline{S}_m\varphi(V)}e^{\overline{S}_n\varphi(W)} \\ &= \sum_{V\in\mathcal{V}} e^{\overline{S}_m\varphi(V)} \sum_{W\in\mathcal{W}} e^{\overline{S}_n\varphi(W)}. \end{aligned}$$

Taking the infimum of the right-hand side over all subcovers $\mathcal{V}$ of $\mathcal{U}^m$ and over all subcovers $\mathcal{W}$ of $\mathcal{U}^n$ gives

$$Z_{m+n}(\varphi,\mathcal{U}) \le Z_m(\varphi,\mathcal{U})Z_n(\varphi,\mathcal{U}). \qquad \square$$

We immediately deduce the following fact.

Corollary 11.1.10. *The sequence* $(\log Z_n(\varphi,\mathcal{U}))_{n=1}^{\infty}$ *is subadditive for every open cover* $\mathcal{U}$ *of* X.

Thanks to this fact, we can define the topological pressure of a potential with respect to an open cover. This constitutes the first step in the definition of the topological pressure of a potential.

Definition 11.1.11. The *topological pressure* of a potential $\varphi : X \to \mathbb{R}$ with respect to an open cover $\mathcal{U}$ of X, denoted by $\mathrm{P}(T,\varphi,\mathcal{U})$, is defined to be

$$\mathrm{P}(T,\varphi,\mathcal{U}) := \lim_{n\to\infty}\frac{1}{n}\log Z_n(\varphi,\mathcal{U}) = \inf_{n\in\mathbb{N}}\frac{1}{n}\log Z_n(\varphi,\mathcal{U}).$$

The existence of the limit and its equality with the infimum follow from Lemma 3.2.17 and Corollary 11.1.10, just as in the corresponding Definition 7.2.12 for topological entropy.

It is also possible to define similar quantities using the $z_n(\varphi,\mathcal{U})$'s rather than the $Z_n(\varphi,\mathcal{U})$'s.

Definition 11.1.12. Given a potential $\varphi : X \to \mathbb{R}$ and an open cover $\mathcal{U}$ of X, let

$$\underline{p}(T,\varphi,\mathcal{U}) := \liminf_{n\to\infty} \frac{1}{n}\log z_n(\varphi,\mathcal{U}) \quad \text{and} \quad \overline{p}(T,\varphi,\mathcal{U}) := \limsup_{n\to\infty} \frac{1}{n}\log z_n(\varphi,\mathcal{U}).$$

Remark 11.1.13. Let $\mathcal{U}$ be an open cover of X.
(a) $\mathrm{P}(T,0,\mathcal{U}) = \underline{p}(T,0,\mathcal{U}) = \overline{p}(T,0,\mathcal{U}) = \mathrm{h}_{\mathrm{top}}(T,\mathcal{U})$ by Remark 11.1.3(c).
(b) By Remark 11.1.3(e),

$$-\infty < \mathrm{h}_{\mathrm{top}}(T,\mathcal{U}) + \inf\varphi \le \mathrm{P}(T,\varphi,\mathcal{U}) \le \mathrm{h}_{\mathrm{top}}(T,\mathcal{U}) + \sup\varphi < \infty.$$

These inequalities also hold with $\mathrm{P}(T,\varphi,\mathcal{U})$ replaced by $\underline{p}(T,\varphi,\mathcal{U})$ and $\overline{p}(T,\varphi,\mathcal{U})$, respectively.
(c) Using Lemma 11.1.6,

$$\underline{p}(T,\varphi,\mathcal{U}) \le \overline{p}(T,\varphi,\mathcal{U}) \le \mathrm{P}(T,\varphi,\mathcal{U}) \le \underline{p}(T,\varphi,\mathcal{U}) + \mathrm{osc}(\varphi,\mathcal{U}).$$

We have seen in Proposition 7.2.14 that the topological entropy relative to covers respects the refinement relation and is subadditive with respect to the join operation. The topological pressure satisfies the following similar properties.

Proposition 11.1.14. *Let $\mathcal{U}$ and $\mathcal{V}$ be open covers of X.*
(a) *If $\mathcal{U} \prec \mathcal{V}$, then* $\mathrm{P}(T,\varphi,\mathcal{U}) - \mathrm{osc}(\varphi,\mathcal{U}) \le \mathrm{P}(T,\varphi,\mathcal{V})$ *while*

$$\underline{p}(T,\varphi,\mathcal{U}) \le \underline{p}(T,\varphi,\mathcal{V}) \quad \textit{and} \quad \overline{p}(T,\varphi,\mathcal{U}) \le \overline{p}(T,\varphi,\mathcal{V}).$$

(b) $\mathrm{P}(T,\varphi,\mathcal{U}\vee\mathcal{V}) \le \min\{\mathrm{P}(T,\varphi,\mathcal{U}) + \mathrm{h}_{\mathrm{top}}(T,\mathcal{V}), \mathrm{P}(T,\varphi,\mathcal{V}) + \mathrm{h}_{\mathrm{top}}(T,\mathcal{U})\}$ *whereas*

$$\begin{aligned}\overline{p}(T,\varphi,\mathcal{U}\vee\mathcal{V}) \le \min\{&\overline{p}(T,\varphi,\mathcal{U}) + \mathrm{osc}(\varphi,\mathcal{U}) + \mathrm{h}_{\mathrm{top}}(T,\mathcal{V}),\\ &\overline{p}(T,\varphi,\mathcal{V}) + \mathrm{osc}(\varphi,\mathcal{V}) + \mathrm{h}_{\mathrm{top}}(T,\mathcal{U})\}\end{aligned}$$

and a similar inequality with $\overline{p}$ replaced by $\underline{p}$.

Proof. Part (a) is an immediate consequence of Lemma 11.1.7 while (b) follows directly from Lemma 11.1.8. □

We have proved in Lemma 7.2.15 that the entropy of a system relative to covers remains the same for all dynamical covers generated by a given cover. The topological pressure of a potential has a similar property.

Lemma 11.1.15. *If $\mathcal{U}$ is an open cover of X, then*

$$\underline{p}(T,\varphi,\mathcal{U}^n) = \underline{p}(T,\varphi,\mathcal{U}) \quad \textit{and} \quad \overline{p}(T,\varphi,\mathcal{U}^n) = \overline{p}(T,\varphi,\mathcal{U})$$

whereas $\mathrm{P}(T,\varphi,\mathcal{U}^n) \le \mathrm{P}(T,\varphi,\mathcal{U})$ *for all $n \in \mathbb{N}$. In addition, if $\mathcal{U}$ is an open partition of X, then* $\mathrm{P}(T,\varphi,\mathcal{U}^n) = \mathrm{P}(T,\varphi,\mathcal{U})$ *for all $n \in \mathbb{N}$.*

Proof. Fix $n \in \mathbb{N}$. For all $k \in \mathbb{N}$ and all $x \in X$, we already know that

$$S_{k+n-1}\varphi(x) = S_k\varphi(x) + S_{n-1}\varphi(T^k(x)).$$

Therefore,

$$S_k\varphi(x) - \|S_{n-1}\varphi\|_\infty \le S_{k+n-1}\varphi(x) \le S_k\varphi(x) + \|S_{n-1}\varphi\|_\infty.$$

Hence, for any subset Y of X,

$$\overline{S}_k\varphi(Y) - \|S_{n-1}\varphi\|_\infty \le \overline{S}_{k+n-1}\varphi(Y) \le \overline{S}_k\varphi(Y) + \|S_{n-1}\varphi\|_\infty \tag{11.1}$$

and

$$\underline{S}_k\varphi(Y) - \|S_{n-1}\varphi\|_\infty \le \underline{S}_{k+n-1}\varphi(Y) \le \underline{S}_k\varphi(Y) + \|S_{n-1}\varphi\|_\infty. \tag{11.2}$$

We claim that

$$e^{-\|S_{n-1}\varphi\|_\infty} Z_k(\varphi, \mathcal{U}^n) \le Z_{k+n-1}(\varphi, \mathcal{U}) \tag{11.3}$$

and

$$e^{-\|S_{n-1}\varphi\|_\infty} z_k(\varphi, \mathcal{U}^n) \le z_{k+n-1}(\varphi, \mathcal{U}) \le e^{\|S_{n-1}\varphi\|_\infty} z_k(\varphi, \mathcal{U}^n). \tag{11.4}$$

Let us first prove (11.3). Recall that $(\mathcal{U}^n)^k \prec \mathcal{U}^{k+n-1} \prec (\mathcal{U}^n)^k$ for all $k \in \mathbb{N}$ (cf. Lemma 7.1.12(d)). However, this is insufficient to declare that a subcover of $\mathcal{U}^{k+n-1}$ is also a subcover of $(\mathcal{U}^n)^k$, or vice versa. We need to remember that $\mathcal{U} \vee \mathcal{U} \supseteq \mathcal{U}$, and thus $(\mathcal{U}^n)^k \supseteq \mathcal{U}^{k+n-1}$, that is, $\mathcal{U}^{k+n-1}$ is a subcover of $(\mathcal{U}^n)^k$. Let $\mathcal{V}$ be a subcover of $\mathcal{U}^{k+n-1}$. Then $\mathcal{V}$ is a subcover of $(\mathcal{U}^n)^k$. Using the left inequality in (11.1) with Y replaced by each $V \in \mathcal{V}$ successively, we obtain

$$e^{-\|S_{n-1}\varphi\|_\infty} Z_k(\varphi, \mathcal{U}^n) \le e^{-\|S_{n-1}\varphi\|_\infty} \sum_{V\in\mathcal{V}} e^{\overline{S}_k\varphi(V)} \le \sum_{V\in\mathcal{V}} e^{\overline{S}_{k+n-1}\varphi(V)}.$$

Taking the infimum over all subcovers $\mathcal{V}$ of $\mathcal{U}^{k+n-1}$ yields (11.3). Similarly, using the left inequality in (11.2), we get that

$$e^{-\|S_{n-1}\varphi\|_\infty} z_k(\varphi, \mathcal{U}^n) \le e^{-\|S_{n-1}\varphi\|_\infty} \sum_{V\in\mathcal{V}} e^{\underline{S}_k\varphi(V)} \le \sum_{V\in\mathcal{V}} e^{\underline{S}_{k+n-1}\varphi(V)}.$$

Taking the infimum over all subcovers $\mathcal{V}$ of $\mathcal{U}^{k+n-1}$ yields the left inequality in (11.4). Regarding the right inequality, since $\mathcal{U}^{k+n-1} \prec (\mathcal{U}^n)^k$, there exists a map $i : (\mathcal{U}^n)^k \to \mathcal{U}^{k+n-1}$ such that $W \subseteq i(W)$ for all $W \in (\mathcal{U}^n)^k$. Let $\mathcal{W}$ be a subcover of $(\mathcal{U}^n)^k$. Then $i(\mathcal{W})$

is a subcover of $\mathcal{U}^{k+n-1}$ and, using the right inequality in (11.2), we deduce that

$$\begin{aligned}
\sum_{W\in\mathcal{W}} e^{\underline{S}_k\varphi(W)} &\geq \sum_{W\in\mathcal{W}} e^{\underline{S}_k\varphi(i(W))} \geq \sum_{Z\in i(\mathcal{W})} e^{\underline{S}_k\varphi(Z)} \\
&\geq \sum_{Z\in i(\mathcal{W})} e^{\underline{S}_{k+n-1}\varphi(Z)-\|S_{n-1}\varphi\|_\infty} \\
&\geq e^{-\|S_{n-1}\varphi\|_\infty} z_{k+n-1}(\varphi,\mathcal{U}).
\end{aligned}$$

Taking the infimum over all subcovers of $(\mathcal{U}^n)^k$ on the left-hand side gives the right inequality in (11.4). So (11.3) and (11.4) always hold.

Moreover, if $\mathcal{U}$ is a partition then $\mathcal{U}\vee\mathcal{U}=\mathcal{U}$, and thus $(\mathcal{U}^n)^k=\mathcal{U}^{k+n-1}$ for all $k\in\mathbb{N}$. Let $\mathcal{W}$ be a subcover of $(\mathcal{U}^n)^k$. Using the right inequality in (11.1), we conclude that

$$\sum_{W\in\mathcal{W}} e^{\overline{S}_k\varphi(W)} \geq \sum_{W\in\mathcal{W}} e^{\overline{S}_{k+n-1}\varphi(W)-\|S_{n-1}\varphi\|_\infty} \geq e^{-\|S_{n-1}\varphi\|_\infty} Z_{k+n-1}(\varphi,\mathcal{U}).$$

Taking the infimum over all subcovers of $(\mathcal{U}^n)^k$ on the left-hand side gives

$$Z_k(\varphi,\mathcal{U}^n) \geq e^{-\|S_{n-1}\varphi\|_\infty} Z_{k+n-1}(\varphi,\mathcal{U}). \tag{11.5}$$

Finally, for the passage from the z_n's to $\overline{p}$, it follows from (11.4) that

$$\frac{k}{k+n-1}\cdot\frac{1}{k}\log z_k(\varphi,\mathcal{U}^n) - \frac{\|S_{n-1}\varphi\|_\infty}{k+n-1} \leq \frac{1}{k+n-1}\log z_{k+n-1}(\varphi,\mathcal{U})$$

and

$$\frac{1}{k+n-1}\log z_{k+n-1}(\varphi,\mathcal{U}) \leq \frac{k}{k+n-1}\cdot\frac{1}{k}\log z_k(\varphi,\mathcal{U}^n) + \frac{\|S_{n-1}\varphi\|_\infty}{k+n-1}.$$

Taking the lim sup as $k\to\infty$ in these two relations yields

$$\overline{p}(T,\varphi,\mathcal{U}^n) \leq \overline{p}(T,\varphi,\mathcal{U}) \leq \overline{p}(T,\varphi,\mathcal{U}^n).$$

Taking the lim inf instead, results in a corresponding conclusion for $\underline{p}$. Similarly, one deduces from (11.3) that $\mathrm{P}(T,\varphi,\mathcal{U}^n) \leq \mathrm{P}(T,\varphi,\mathcal{U})$ and, when $\mathcal{U}$ is a partition, it ensues from (11.5) that $\mathrm{P}(T,\varphi,\mathcal{U}^n) \geq \mathrm{P}(T,\varphi,\mathcal{U})$. □

11.1.2 Second stage: the pressure of a potential

Recall that the topological entropy of a system is defined to be the supremum over all open covers of the entropy of the system with respect to an open cover (cf. Definition 7.2.16). However, due to Proposition 11.1.14(a), taking the supremum of the pressure relative to all covers does not always lead to a quantity that has natural properties. Instead, we take the supremum of the difference between the pressure relative to a cover minus the oscillation of the potential with respect to that cover. This definition is purely topological.

Definition 11.1.16. Let $T : X \to X$ be a topological dynamical system and $\varphi : X \to \mathbb{R}$ a potential. The *topological pressure* of the potential φ, denoted $\mathrm{P}(T,\varphi)$, is defined by

$$\mathrm{P}(T,\varphi) := \sup\{\mathrm{P}(T,\varphi,\mathcal{U}) - \mathrm{osc}(\varphi,\mathcal{U}) : \mathcal{U} \text{ is an open cover of } X\}.$$

In light of Proposition 11.1.14(a), we may define the counterparts $\underline{p}(T,\varphi)$ and $\overline{p}(T,\varphi)$ of $\mathrm{P}(T,\varphi)$ by simply taking the supremum over all covers.

Definition 11.1.17. Let $T : X \to X$ be a topological dynamical system and $\varphi : X \to \mathbb{R}$ a potential. Define

$$\underline{p}(T,\varphi) := \sup\{\underline{p}(T,\varphi,\mathcal{U}) : \mathcal{U} \text{ is an open cover of } X\},$$

and

$$\overline{p}(T,\varphi) := \sup\{\overline{p}(T,\varphi,\mathcal{U}) : \mathcal{U} \text{ is an open cover of } X\}.$$

Clearly, $\underline{p}(T,\varphi) \le \overline{p}(T,\varphi)$. In fact, $\underline{p}(T,\varphi)$ and $\overline{p}(T,\varphi)$ are just other expressions of the topological pressure.

Theorem 11.1.18. *For any topological dynamical system $T : X \to X$ and potential $\varphi : X \to \mathbb{R}$, it turns out that $\underline{p}(T,\varphi) = \overline{p}(T,\varphi) = \mathrm{P}(T,\varphi)$.*

Proof. From a rearrangement of the right inequality in Remark 11.1.13(c), it follows that $\mathrm{P}(T,\varphi) \le \underline{p}(T,\varphi) \le \overline{p}(T,\varphi)$.

To prove that $\overline{p}(T,\varphi) \le \mathrm{P}(T,\varphi)$, let $(\mathcal{U}_n)_{n=1}^{\infty}$ be a sequence of open covers such that $\lim_{n\to\infty} \overline{p}(T,\varphi,\mathcal{U}_n) = \overline{p}(T,\varphi)$. Each open cover $\mathcal{U}_n$ has a Lebesgue number $\delta_n > 0$. The compactness of X guarantees that there are finitely many open balls of radius $\min\{\delta_n, 1/(2n)\}$ that cover X. These balls thereby constitute a refinement of $\mathcal{U}_n$ of diameter at most $1/n$. Thanks to Proposition 11.1.14(a), this means that we may assume without loss of generality that the sequence $(\mathcal{U}_n)_{n=1}^{\infty}$ is such that $\lim_{n\to\infty} \mathrm{diam}(\mathcal{U}_n) = 0$. Since φ is uniformly continuous, it ensues that $\lim_{n\to\infty} \mathrm{osc}(\varphi,\mathcal{U}_n) = 0$. Consequently, using the left inequality in Remark 11.1.13(c), we conclude that

$$\begin{aligned}\mathrm{P}(T,\varphi) &\ge \sup_{n\in\mathbb{N}}[\mathrm{P}(T,\varphi,\mathcal{U}_n) - \mathrm{osc}(\varphi,\mathcal{U}_n)]\\ &\ge \sup_{n\in\mathbb{N}}[\overline{p}(T,\varphi,\mathcal{U}_n) - \mathrm{osc}(\varphi,\mathcal{U}_n)]\\ &\ge \lim_{n\in\mathbb{N}}[\overline{p}(T,\varphi,\mathcal{U}_n) - \mathrm{osc}(\varphi,\mathcal{U}_n)] = \overline{p}(T,\varphi).\end{aligned}$$ □

Remark 11.1.19.

(a) $\mathrm{P}(T,0) = \mathrm{h_{top}}(T)$. Thus topological pressure generalizes topological entropy. This is a consequence of Remark 11.1.13(a) and the fact that $\mathrm{osc}(0,\mathcal{U}) = 0$ for all $\mathcal{U}$.

(b) By Remark 11.1.13(b),

$$\mathrm{h_{top}}(T) + \inf\varphi - \mathrm{osc}(\varphi,X) \le \mathrm{P}(T,\varphi) \le \mathrm{h_{top}}(T) + \sup\varphi.$$

(c) $\mathrm{P}(T,\varphi) = \infty$ if and only if $\mathrm{h_{top}}(T) = \infty$, according to part (b).

We saw in Proposition 7.2.18 that topological entropy cannot be greater for a factor of a given map than for the original map. We expect the same for its generalization, topological pressure, with a small twist, namely that the potentials to which the dynamical systems are subject must correspond.

Proposition 11.1.20. *Suppose that $S : Y \to Y$ is a factor of $T : X \to X$ via the factor map $h : X \to Y$. Then for every potential $\varphi : Y \to \mathbb{R}$, we have that $\mathrm{P}(S,\varphi) \leq \mathrm{P}(T,\varphi\circ h)$.*

Proof. Let $\mathcal{V}$ be an open cover of Y. Recall (cf. proof of Proposition 7.2.18) that $h^{-1}(\mathcal{V}_S^n) = (h^{-1}(\mathcal{V}))_T^n$ for all $n \in \mathbb{N}$. Without loss of generality, we may restrict our attention to nondegenerate subcovers, that is, subcovers whose members are all different from one another. Letting C be the collection of all nondegenerate subcovers of $\mathcal{V}_S^n$, the map $\mathcal{C} \mapsto h^{-1}(\mathcal{C}), \mathcal{C} \in C$, defines a bijection between the nondegenerate subcovers of $\mathcal{V}_S^n$ and the nondegenerate subcovers of $h^{-1}(\mathcal{V}_S^n) = (h^{-1}(\mathcal{V}))_T^n$, since h is a surjection. We leave it to the reader to show that

$$\overline{S}_n^T(\varphi\circ h)(h^{-1}(Z)) = \overline{S}_n^S\varphi(Z), \quad \forall Z \subseteq Y.$$

It then follows that (again this is left to the reader)

$$Z_n(T,\varphi\circ h, h^{-1}(\mathcal{V})) = Z_n(S,\varphi,\mathcal{V}).$$

Therefore,

$$\mathrm{P}(T,\varphi\circ h, h^{-1}(\mathcal{V})) = \mathrm{P}(S,\varphi,\mathcal{V}).$$

Observe further that $\mathrm{osc}(\varphi\circ h, h^{-1}(\mathcal{V})) = \mathrm{osc}(\varphi,\mathcal{V})$. Then

$$\begin{aligned}\mathrm{P}(T,\varphi\circ h) &\geq \mathrm{P}(T,\varphi\circ h, h^{-1}(\mathcal{V})) - \mathrm{osc}(\varphi\circ h, h^{-1}(\mathcal{V}))\\ &= \mathrm{P}(S,\varphi,\mathcal{V}) - \mathrm{osc}(\varphi,\mathcal{V}).\end{aligned}$$

Taking the supremum over all open covers $\mathcal{V}$ of Y yields $\mathrm{P}(T,\varphi\circ h) \geq \mathrm{P}(S,\varphi)$. □

An immediate but important consequence of this lemma is the following.

Corollary 11.1.21. *If $T : X \to X$ and $S : Y \to Y$ are topologically conjugate dynamical systems via a conjugacy $h : X \to Y$, then $\mathrm{P}(S,\varphi) = \mathrm{P}(T,\varphi\circ h)$ for all potentials $\varphi : Y \to \mathbb{R}$.*

We now study the behavior of topological pressure with respect to the iterates of the system. This is a generalization of Theorem 7.2.19.

Theorem 11.1.22. *For every $n \in \mathbb{N}$, we have that $\mathrm{P}(T^n, S_n\varphi) = n\,\mathrm{P}(T,\varphi)$.*

Proof. Fix $n \in \mathbb{N}$. Let $\mathcal{U}$ be an open cover of X. The action of the map T^n on $\mathcal{U}$ until time $j-1$ will be denoted by $\mathcal{U}_{T^n}^j$. Recall (cf. proof of Theorem 7.2.19) that $\mathcal{U}^{mn} = (\mathcal{U}^n)_{T^n}^m$ for all $m \in \mathbb{N}$. Furthermore, for all $x \in X$,

$$S_{mn}\varphi(x) = \sum_{k=0}^{mn-1}\varphi\circ T^k(x) = \sum_{j=0}^{m-1}(S_n\varphi)\circ T^{jn}(x) = \sum_{j=0}^{m-1}(S_n\varphi)\circ (T^n)^j(x) = S_m^{T^n}(S_n\varphi)(x),$$

where $S_m^{T^n}\psi(x) = \sum_{j=0}^{m-1}\psi((T^n)^j(x))$. Hence, $\underline{S}_{mn}\varphi(Y) = \underline{S}_m^{T^n}(S_n\varphi)(Y)$ for all subsets Y of X, and in particular for all $Y \in \mathcal{U}^{mn} = (\mathcal{U}^n)_{T^n}^m$. Thus,

$$z_{mn}(T,\varphi,\mathcal{U}) = z_m(T^n, S_n\varphi, \mathcal{U}^n), \quad \forall m \in \mathbb{N}.$$

Using this and Lemma 11.1.14(a), we get

$$\begin{aligned}\overline{p}(T,\varphi,\mathcal{U}) = \limsup_{m\to\infty}\frac{1}{m}\log z_m(T,\varphi,\mathcal{U}) &\geq \limsup_{m\to\infty}\frac{1}{mn}\log z_{mn}(T,\varphi,\mathcal{U})\\ &= \frac{1}{n}\limsup_{m\to\infty}\frac{1}{m}\log z_m(T^n,S_n\varphi,\mathcal{U}^n)\\ &= \frac{1}{n}\overline{p}(T^n,S_n\varphi,\mathcal{U}^n) \geq \frac{1}{n}\overline{p}(T^n,S_n\varphi,\mathcal{U}).\end{aligned}$$

Taking the supremum over all open covers $\mathcal{U}$ of X yields

$$\overline{p}(T,\varphi) \geq \frac{1}{n}\overline{p}(T^n,S_n\varphi).$$

Similarly,

$$\underline{p}(T,\varphi) \leq \frac{1}{n}\underline{p}(T^n,S_n\varphi).$$

The result ensues from the previous two relations and Theorem 11.1.18. □

As a generalization of topological entropy, in a metrizable space topological pressure is determined by any sequence of covers whose diameters tend to zero. The next result is an extension of Lemma 7.2.20.

Lemma 11.1.23. *The following quantities are all equal:*
(a) $\mathrm{P}(T,\varphi)$.
(b) $\overline{p}(T,\varphi)$.
(c) $\lim_{\varepsilon\to 0}[\sup\{\mathrm{P}(T,\varphi,\mathcal{U}) : \mathcal{U}$ *open cover with* $\mathrm{diam}(\mathcal{U}) \leq \varepsilon\}]$.
(d) $\sup\{\overline{p}(T,\varphi,\mathcal{U}) : \mathcal{U}$ *open cover with* $\mathrm{diam}(\mathcal{U}) \leq \delta\}$ *for any* $\delta > 0$.
(e) $\lim_{\varepsilon\to 0}\mathrm{P}(T,\varphi,\mathcal{U}_\varepsilon)$ *for any open covers* $(\mathcal{U}_\varepsilon)_{\varepsilon\in(0,\infty)}$ *such that* $\lim_{\varepsilon\to 0}\mathrm{diam}(\mathcal{U}_\varepsilon) = 0$.
(f) $\lim_{\varepsilon\to 0}\overline{p}(T,\varphi,\mathcal{U}_\varepsilon)$ *for any open covers* $(\mathcal{U}_\varepsilon)_{\varepsilon\in(0,\infty)}$ *such that* $\lim_{\varepsilon\to 0}\mathrm{diam}(\mathcal{U}_\varepsilon) = 0$.
(g) $\lim_{n\to\infty}\mathrm{P}(T,\varphi,\mathcal{U}_n)$ *for any open covers* $(\mathcal{U}_n)_{n=1}^\infty$ *such that* $\lim_{n\to\infty}\mathrm{diam}(\mathcal{U}_n) = 0$.
(h) $\lim_{n\to\infty}\overline{p}(T,\varphi,\mathcal{U}_n)$ *for any open covers* $(\mathcal{U}_n)_{n=1}^\infty$ *such that* $\lim_{n\to\infty}\mathrm{diam}(\mathcal{U}_n) = 0$.

Note that $\overline{p}$ *can be replaced by* $\underline{p}$ *in the statements above.*

Proof. We already know that (a)=(b) by Lemma 11.1.18. It is clear that (b)≥(d). It is also obvious that (d)≥(f) and (c)≥(e) for any family $(\mathcal{U}_\varepsilon)_{\varepsilon\in(0,\infty)}$ as described, and that (d)≥(h) and (c)≥(g) for any sequence $(\mathcal{U}_n)_{n=1}^\infty$ as specified. It thus suffices to prove that (f)≥(b), that (h)≥(b), that (e)≥(a), that (g)≥(a), and that (b)≥(c).

We will prove that (g)≥(a). The proofs of the other inequalities are similar. Let $\mathcal{V}$ be any open cover of X. Since $\lim_{n\to\infty}\mathrm{diam}(\mathcal{U}_n) = 0$, there exists $N \in \mathbb{N}$ such that $\mathcal{V} \prec \mathcal{U}_n$

for all $n \geq N$ (cf. proof of Lemma 7.2.20). By Proposition 11.1.14(a), we obtain that for all sufficiently large n,

$$\mathrm{P}(T,\varphi,\mathcal{U}_n) \geq \mathrm{P}T,\varphi,\mathcal{V}) - \mathrm{osc}(\varphi,\mathcal{V}).$$

We immediately deduce that

$$\liminf_{n\to\infty} \mathrm{P}(T,\varphi,\mathcal{U}_n) \geq \mathrm{P}(T,\varphi,\mathcal{V}) - \mathrm{osc}(\varphi,\mathcal{V}).$$

As the open cover $\mathcal{V}$ was chosen arbitrarily, passing to the supremum over all open covers allows us to conclude that

$$\liminf_{n\to\infty} \mathrm{P}(T,\varphi,\mathcal{U}_n) \geq \mathrm{P}(T,\varphi).$$

But $\lim_{n\to\infty} \mathrm{osc}(\varphi,\mathcal{U}_n) = 0$ since $\lim_{n\to\infty} \mathrm{diam}(\mathcal{U}_n) = 0$ and φ is uniformly continuous. Therefore,

$$\begin{aligned}
\mathrm{P}(T,\varphi) &= \sup_{\mathcal{V}}\big[\mathrm{P}(T,\varphi,\mathcal{V}) - \mathrm{osc}(\varphi,\mathcal{V})\big] \\
&\geq \limsup_{n\to\infty}\big[\mathrm{P}(T,\varphi,\mathcal{U}_n) - \mathrm{osc}(\varphi,\mathcal{U}_n)\big] \\
&= \limsup_{n\to\infty} \mathrm{P}(T,\varphi,\mathcal{U}_n) - \lim_{n\to\infty} \mathrm{osc}(\varphi,\mathcal{U}_n) \\
&= \limsup_{n\to\infty} \mathrm{P}(T,\varphi,\mathcal{U}_n) \geq \liminf_{n\to\infty} \mathrm{P}(T,\varphi,\mathcal{U}_n) \geq \mathrm{P}(T,\varphi).
\end{aligned}$$

Hence, $\mathrm{P}(T,\varphi) = \lim_{n\to\infty} \mathrm{P}(T,\varphi,\mathcal{U}_n)$. □

We can now obtain a slightly stronger estimate than Remark 11.1.19(b) for the difference between topological entropy and topological pressure when the underlying space is metrizable.

Corollary 11.1.24. $\mathrm{h}_{\mathrm{top}}(T) + \inf\varphi \leq \mathrm{P}(T,\varphi) \leq \mathrm{h}_{\mathrm{top}}(T) + \sup\varphi$.

Proof. The upper bound was already mentioned in Remark 11.1.19(b). In order to derive the lower bound, we return to Remark 11.1.13(b). Let $(\mathcal{U}_n)_{n=1}^{\infty}$ be a sequence of open covers of X such that $\lim_{n\to\infty} \mathrm{diam}(\mathcal{U}_n) = 0$. According to Remark 11.1.13(b), for each $n \in \mathbb{N}$ we have

$$\mathrm{h}_{\mathrm{top}}(T,\mathcal{U}_n) + \inf\varphi \leq \mathrm{P}(T,\varphi,\mathcal{U}_n).$$

Passing to the limit $n \to \infty$ and using Lemmas 7.2.20 and 11.1.23, we conclude that

$$\mathrm{h}_{\mathrm{top}}(T) + \inf\varphi \leq \mathrm{P}(T,\varphi).$$ □

The preceding lemma characterized the topological pressure of a potential as the limit of the topological pressure of the potential relative to a sequence of covers. An even better result would be the characterization of the topological pressure of a potential as the topological pressure of that potential with respect to a single cover. As might by now be expected, such a characterization exists when the system has a generator. This is a generalization of Lemma 7.2.22 (see also Definition 7.2.21).

Lemma 11.1.25. *If a system $T : X \to X$ has a generator $\mathcal{U}$, then*

$$\mathrm{P}(T,\varphi) = \underline{p}(T,\varphi,\mathcal{U}) = \overline{p}(T,\varphi,\mathcal{U}).$$

Moreover, if the generator $\mathcal{U}$ is a partition, then

$$\mathrm{P}(T,\varphi) = \mathrm{P}(T,\varphi,\mathcal{U}).$$

Proof. It follows from Lemmas 11.1.23 and 11.1.15 that

$$\mathrm{P}(T,\varphi) = \lim_{n\to\infty} \overline{p}(T,\varphi,\mathcal{U}^n) = \lim_{n\to\infty} \overline{p}(T,\varphi,\mathcal{U}) = \overline{p}(T,\varphi,\mathcal{U}).$$

A similar argument leads to the statements for $\underline{p}$ and for a generating partition. □

We then have the following generalization of Theorem 7.2.24.

Theorem 11.1.26. *If $T : X \to X$ is a δ-expansive dynamical system on a compact metric space (X,d), then*

$$\mathrm{P}(T,\varphi) = \underline{p}(T,\varphi,\mathcal{U}) = \overline{p}(T,\varphi,\mathcal{U})$$

for any open cover $\mathcal{U}$ of X with $\mathrm{diam}(\mathcal{U}) \le \delta$. *Moreover,*

$$\mathrm{P}(T,\varphi) = \mathrm{P}(T,\varphi,\mathcal{U})$$

for any open partition $\mathcal{U}$ of X with $\mathrm{diam}(\mathcal{U}) \le \delta$.

Proof. This is an immediate consequence of Lemmas 11.1.25 and 7.2.23. □

11.2 Bowen's definition of topological pressure

We have seen in Theorem 7.3.8 and Corollary 7.3.12 that topological entropy can also be defined using separated or spanning sets. This definition may be generalized to yield a definition of topological pressure, which coincides with the one from the previous section. To lighten notation, for any $n \in \mathbb{N}$ and $Y \subseteq X$, let

$$\Sigma_n(Y) = \sum_{x\in Y} e^{S_n\varphi(x)}.$$

Theorem 11.2.1. *For all $n \in \mathbb{N}$ and all $\varepsilon > 0$, let $E_n(\varepsilon)$ be a maximal (n,ε)-separated set and $F_n(\varepsilon)$ be a minimal (n,ε)-spanning set. Then*

$$\mathrm{P}(T,\varphi) = \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\Sigma_n(E_n(\varepsilon)) = \lim_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\Sigma_n(E_n(\varepsilon))$$
$$\leq \liminf_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\Sigma_n(F_n(\varepsilon)).$$

Proof. Fix $\varepsilon > 0$. Let $\mathcal{U}_\varepsilon$ be an open cover of X consisting of balls of radius $\varepsilon/2$. Fix $n \in \mathbb{N}$. Let $\mathcal{U}$ be a subcover of $\mathcal{U}_\varepsilon^n$ such that $Z_n(\varphi,\mathcal{U}_\varepsilon) \geq e^{-1}\sum_{U\in\mathcal{U}}\exp(\overline{S}_n\varphi(U))$. For each $x \in E_n(\varepsilon)$, let $U(x)$ be an element of the cover $\mathcal{U}$ which contains x and define the function $i : E_n(\varepsilon) \to \mathcal{U}$ by setting $i(x) = U(x)$. We have already shown in the proof of Theorem 7.3.8 that this function is an injection. Therefore,

$$Z_n(\varphi,\mathcal{U}_\varepsilon) \geq e^{-1}\sum_{U\in\mathcal{U}} e^{\overline{S}_n\varphi(U)} \geq e^{-1}\sum_{x\in E_n(\varepsilon)} e^{\overline{S}_n\varphi(U(x))} \geq e^{-1}\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)}.$$

Since this is true for all $n \in \mathbb{N}$, we deduce that

$$\mathrm{P}(T,\varphi,\mathcal{U}_\varepsilon) = \lim_{n\to\infty}\frac{1}{n}\log Z_n(\varphi,\mathcal{U}_\varepsilon) \geq \limsup_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)}.$$

Letting $\varepsilon \to 0$ and using Lemma 11.1.23 yields that

$$\mathrm{P}(T,\varphi) \geq \limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)}. \tag{11.6}$$

On the other hand, if $\mathcal{V}$ is an arbitrary open cover of X, if $\delta(\mathcal{V})$ is a Lebesgue number for $\mathcal{V}$, if $0 < \varepsilon < \delta(\mathcal{V})/2$ and if $n \in \mathbb{N}$, then for all $0 \leq k < n$ and all $x \in E_n(\varepsilon)$ we have

$$T^k(B_n(x,\varepsilon)) \subseteq B(T^k(x),\varepsilon) \implies \mathrm{diam}(T^k(B_n(x,\varepsilon))) \leq 2\varepsilon < \delta(\mathcal{V}).$$

Hence, for all $0 \leq k < n$, the set $T^k(B_n(x,\varepsilon))$ is contained in at least one element of $\mathcal{V}$. Denote one such element by $V_k(x)$. Then $B_n(x,\varepsilon) \subseteq \bigcap_{k=0}^{n-1} T^{-k}(V_k(x))$. But this latter intersection is simply an element of $\mathcal{V}^n$. Let us denote it by $V(x)$.

Since $E_n(\varepsilon)$ is a maximal (n,ε)-separated set, by Lemma 7.3.7 it is also (n,ε)-spanning, so the family $\{B_n(x,\varepsilon)\}_{x\in E_n(\varepsilon)}$ is an open cover of X. Each of these balls is contained in the corresponding set $V(x)$. Hence, the family $\{V(x)\}_{x\in E_n(\varepsilon)}$ is also an open cover of X. Therefore, it is a subcover of $\mathcal{V}^n$. Consequently,

$$Z_n(\varphi,\mathcal{V}) \leq \sum_{x\in E_n(\varepsilon)} e^{\overline{S}_n\varphi(V(x))} \leq e^{n\,\mathrm{osc}(\varphi,\mathcal{V})}\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)},$$

where the last inequality is due to Lemma 11.1.5. It follows that

$$\mathrm{P}(T,\varphi,\mathcal{V}) \leq \mathrm{osc}(\varphi,\mathcal{V}) + \liminf_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)}.$$

Since $\mathcal{V}$ is independent of $\varepsilon > 0$, we deduce that

$$\mathrm{P}(T,\varphi,\mathcal{V}) - \mathrm{osc}(\varphi,\mathcal{V}) \le \liminf_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)}.$$

Then, as $\mathcal{V}$ was chosen to be an arbitrary open cover of X, we conclude that

$$\mathrm{P}(T,\varphi) \le \liminf_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)}. \tag{11.7}$$

Inequalities (11.6)–(11.7) establish the result for the separated sets. We can deduce the result for the spanning sets as in the proof of Theorem 7.3.8. □

In Theorem 11.2.1, the topological pressure of the system is expressed in terms of a specific family of maximal separated (resp. minimal spanning) sets. However, to derive theoretical results, it is sometimes simpler to use the following quantities.

Definition 11.2.2. For all $n \in \mathbb{N}$ and $\varepsilon > 0$, let

$$P_n(T,\varphi,\varepsilon) = \sup\{\Sigma_n(E_n(\varepsilon)) \;:\; E_n(\varepsilon) \text{ maximal } (n,\varepsilon)\text{-separated set}\}$$
$$Q_n(T,\varphi,\varepsilon) = \inf\{\Sigma_n(F_n(\varepsilon)) \;:\; F_n(\varepsilon) \text{ minimal } (n,\varepsilon)\text{-spanning set}\}.$$

Thereafter, let

$$\underline{P}(T,\varphi,\varepsilon) = \liminf_{n\to\infty}\frac{1}{n}\log P_n(T,\varphi,\varepsilon), \qquad \overline{P}(T,\varphi,\varepsilon) = \limsup_{n\to\infty}\frac{1}{n}\log P_n(T,\varphi,\varepsilon)$$
$$\underline{Q}(T,\varphi,\varepsilon) = \liminf_{n\to\infty}\frac{1}{n}\log Q_n(T,\varphi,\varepsilon), \qquad \overline{Q}(T,\varphi,\varepsilon) = \limsup_{n\to\infty}\frac{1}{n}\log Q_n(T,\varphi,\varepsilon).$$

The following are key observations constitute a generalization of Remark 7.3.10.

Remark 11.2.3. Let $m \le n \in \mathbb{N}$ and $0 < \varepsilon < \varepsilon'$. The following relations hold:

(a) $P_m(T,\varphi,\varepsilon) \le P_n(T,\varphi,\varepsilon)e^{(n-m)\|\varphi\|_\infty}$ by Remark 7.3.2(a).

(b) $e^{-n\|\varphi\|_\infty} \le P_n(T,\varphi,\varepsilon) \le r_n(\varepsilon)e^{n\|\varphi\|_\infty}$ and $P_n(T,0,\varepsilon) = r_n(\varepsilon)$.

(c) $Q_m(T,\varphi,\varepsilon) \le Q_n(T,\varphi,\varepsilon)e^{(n-m)\|\varphi\|_\infty}$ by Remark 7.3.6(a).

(d) $e^{-n\|\varphi\|_\infty} \le Q_n(T,\varphi,\varepsilon) \le s_n(\varepsilon)e^{n\|\varphi\|_\infty}$ and $Q_n(T,0,\varepsilon) = s_n(\varepsilon)$.

(e) $P_n(T,\varphi,\varepsilon) \ge P_n(T,\varphi,\varepsilon')$ and $Q_n(T,\varphi,\varepsilon) \ge Q_n(T,\varphi,\varepsilon')$ by Remarks 7.3.2 and 7.3.6(b).

(f) $0 < Q_n(T,\varphi,\varepsilon) \le P_n(T,\varphi,\varepsilon) < \infty$ by Lemma 7.3.7.

(g) $\underline{P}(T,\varphi,\varepsilon) \le \overline{P}(T,\varphi,\varepsilon)$ and $\underline{Q}(T,\varphi,\varepsilon) \le \overline{Q}(T,\varphi,\varepsilon)$.

(h) $-\|\varphi\|_\infty \le \underline{P}(T,\varphi,\varepsilon) \le \underline{r}(\varepsilon) + \|\varphi\|_\infty$ and $-\|\varphi\|_\infty \le \overline{P}(T,\varphi,\varepsilon) \le \overline{r}(\varepsilon) + \|\varphi\|_\infty$ by (b).

(i) $-\|\varphi\|_\infty \le \underline{Q}(T,\varphi,\varepsilon) \le \underline{s}(\varepsilon) + \|\varphi\|_\infty$ and $-\|\varphi\|_\infty \le \overline{Q}(T,\varphi,\varepsilon) \le \overline{s}(\varepsilon) + \|\varphi\|_\infty$ by (d).

(j) $\underline{P}(T,\varphi,\varepsilon) \ge \underline{P}(T,\varphi,\varepsilon')$ and $\overline{P}(T,\varphi,\varepsilon) \ge \overline{P}(T,\varphi,\varepsilon')$ by (e).

(k) $\underline{Q}(T,\varphi,\varepsilon) \ge \underline{Q}(T,\varphi,\varepsilon')$ and $\overline{Q}(T,\varphi,\varepsilon) \ge \overline{Q}(T,\varphi,\varepsilon')$ by (e).

(l) $-\|\varphi\|_\infty \le \overline{Q}(T,\varphi,\varepsilon) \le \overline{P}(T,\varphi,\varepsilon) \le \infty$ and $-\|\varphi\|_\infty \le \underline{Q}(T,\varphi,\varepsilon) \le \underline{P}(T,\varphi,\varepsilon) \le \infty$ by (f).

We now describe a relationship between the P_n's, the Q_n's and the cover-related quantities Z_n's and z_n's. This is the counterpart of Lemma 7.3.11.

Lemma 11.2.4. *The following relations hold:*
(a) *If $\mathcal{U}$ is an open cover of X with Lebesgue number 2δ, then*

$$z_n(T,\varphi,\mathcal{U}) \le Q_n(T,\varphi,\delta) \le P_n(T,\varphi,\delta).$$

(b) *If $\varepsilon > 0$ and $\mathcal{V}$ is an open cover of X with* $\operatorname{diam}(\mathcal{V}) \le \varepsilon$, *then*

$$Q_n(T,\varphi,\varepsilon) \le P_n(T,\varphi,\varepsilon) \le Z_n(T,\varphi,\mathcal{V}).$$

Proof. We already know that $Q_n(T,\varphi,\delta) \le P_n(T,\varphi,\delta)$.

(a) Let $\mathcal{U}$ be an open cover with Lebesgue number 2δ and let F be an (n,δ)-spanning set. Then the dynamic balls $\{B_n(x,\delta) : x \in F\}$ form a cover of X. For every $0 \le i < n$, the ball $B(T^i(x),\delta)$, which has diameter at most 2δ, is contained in an element of $\mathcal{U}$. Therefore $B_n(x,\delta) = \bigcap_{i=0}^{n-1} T^{-i}(B(T^i(x),\delta))$ is contained in an element of $\mathcal{U}^n = \bigvee_{i=0}^{n-1} T^{-i}(\mathcal{U})$. That is, $\mathcal{U}^n \prec \{B_n(x,\delta) : x \in F\}$. Then there exists a map $i : \{B_n(x,\delta) : x \in F\} \to \mathcal{U}^n$ such that $B_n(x,\delta) \subseteq i(B_n(x,\delta))$ for every $x \in F$. Let $\mathcal{W}$ be a subcover of $\{B_n(x,\delta) : x \in F\}$. Then $i(\mathcal{W})$ is a subcover of $\mathcal{U}^n$ and thus

$$\begin{aligned}\Sigma_n(F) = \sum_{x\in F} e^{S_n\varphi(x)} \ge \sum_{x\in F} e^{\underline{S}_n\varphi(B_n(x,\delta))} &\ge \sum_{W\in\mathcal{W}} e^{\underline{S}_n\varphi(W)} \ge \sum_{W\in\mathcal{W}} e^{\underline{S}_n\varphi(i(W))}\\ &\ge \sum_{Z\in i(\mathcal{W})} e^{\underline{S}_n\varphi(Z)} \ge z_n(T,\varphi,\mathcal{U}).\end{aligned}$$

Since F is an arbitrary (n,δ)-spanning set, it ensues that $Q_n(T,\varphi,\delta) \ge z_n(T,\varphi,\mathcal{U})$.

(b) Let $\mathcal{V}$ be an open cover with $\operatorname{diam}(\mathcal{V}) \le \varepsilon$ and let E be an (n,ε)-separated set. Let $\mathcal{W}$ be a subcover of $\mathcal{V}^n$. Let $i : E \to \mathcal{W}$ be such that $x \in i(x)$ for all $x \in E$. This map is injective as no element of the cover $\mathcal{V}^n$ can contain more than one element of E. Then

$$\Sigma_n(E) = \sum_{x\in E} e^{S_n\varphi(x)} \le \sum_{x\in E} e^{\overline{S}_n\varphi(i(x))} = \sum_{W\in i(E)} e^{\overline{S}_n\varphi(W)} \le \sum_{W\in\mathcal{W}} e^{\overline{S}_n\varphi(W)}.$$

As $\mathcal{W}$ is an arbitrary subcover of $\mathcal{V}^n$, it follows that $\Sigma_n(E) \le Z_n(T,\varphi,\mathcal{V})$. Since E is an arbitrary (n,ε)-separated set, we deduce that $P_n(T,\varphi,\varepsilon) \le Z_n(T,\varphi,\mathcal{V})$. □

These inequalities have the following immediate consequences.

Corollary 11.2.5. *The following relations hold:*
(a) *If $\mathcal{U}$ is an open cover of X with Lebesgue number 2δ, then*

$$\underline{p}(T,\varphi,\mathcal{U}) \le \underline{Q}(T,\varphi,\delta) \le \underline{P}(T,\varphi,\delta).$$

(b) *If $\varepsilon > 0$ and $\mathcal{V}$ is an open cover of X with* $\operatorname{diam}(\mathcal{V}) \le \varepsilon$, *then*

$$\overline{Q}(T,\varphi,\varepsilon) \le \overline{P}(T,\varphi,\varepsilon) \le \mathrm{P}(T,\varphi,\mathcal{V}).$$

We can then surmise new expressions for the topological pressure (cf. Corollary 7.3.12).

Corollary 11.2.6. *The following equalities hold:*

$$\mathrm{P}(T,\varphi) = \lim_{\varepsilon\to 0} \underline{P}(T,\varphi,\varepsilon) = \lim_{\varepsilon\to 0} \overline{P}(T,\varphi,\varepsilon) = \lim_{\varepsilon\to 0} \underline{Q}(T,\varphi,\varepsilon) = \lim_{\varepsilon\to 0} \overline{Q}(T,\varphi,\varepsilon).$$

Proof. Let $(\mathcal{U}_\varepsilon)_{\varepsilon\in(0,\infty)}$ be a family of open covers such that $\lim_{\varepsilon\to\infty} \operatorname{diam}(\mathcal{U}_\varepsilon) = 0$. Let δ_ε be a Lebesgue number for $\mathcal{U}_\varepsilon$. Then $\lim_{\varepsilon\to\infty} \delta_\varepsilon = 0$, as $\delta_\varepsilon \le \operatorname{diam}(\mathcal{U}_\varepsilon)$. Using Lemma 11.1.23 and Corollary 11.2.5(a), we deduce that

$$\mathrm{P}(T,\varphi) = \lim_{\varepsilon\to\infty} \underline{p}(T,\varphi,\mathcal{U}_\varepsilon) \le \lim_{\varepsilon\to 0} \underline{Q}(T,\varphi,\varepsilon) \le \lim_{\varepsilon\to 0} \underline{P}(T,\varphi,\varepsilon). \tag{11.8}$$

On the other hand, using Lemma 11.1.23 and Corollary 11.2.5(b), we obtain

$$\lim_{\varepsilon\to 0} \overline{Q}(T,\varphi,\varepsilon) \le \lim_{\varepsilon\to 0} \overline{P}(T,\varphi,\varepsilon) \le \lim_{\varepsilon\to 0} \sup_{\operatorname{diam}(\mathcal{V})\le\varepsilon} \mathrm{P}(T,\varphi,\mathcal{V}) = \mathrm{P}(T,\varphi). \tag{11.9}$$

Combining (11.8) and (11.9) allows us to conclude. □

Corollary 11.2.6 is useful to derive theoretical results. Nevertheless, in practice, Theorem 11.2.1 is simpler to use, as only one family of sets is needed. Sometimes a single sequence of sets is enough (cf. Theorem 7.3.13).

Theorem 11.2.7. *If a topological dynamical system $T : X \to X$ admits a generator with Lebesgue number 2δ, then the following statements hold for all $0 < \varepsilon \le \delta$:*

(a) *If $(E_n(\varepsilon))_{n=1}^\infty$ is a sequence of maximal (n,ε)-separated sets in X, then*

$$\mathrm{P}(T,\varphi) = \lim_{n\to\infty} \frac{1}{n} \log \Sigma_n(E_n(\varepsilon)).$$

(b) *If $(F_n(\varepsilon))_{n=1}^\infty$ is a sequence of minimal (n,ε)-spanning sets in X, then*

$$\mathrm{P}(T,\varphi) \le \liminf_{n\to\infty} \frac{1}{n} \log \Sigma_n(F_n(\varepsilon)).$$

(c) $\mathrm{P}(T,\varphi) = \lim_{n\to\infty} \frac{1}{n} \log P_n(T,\varphi,\varepsilon)$.

(d) $\mathrm{P}(T,\varphi) = \lim_{n\to\infty} \frac{1}{n} \log Q_n(T,\varphi,\varepsilon)$.

Proof. We will prove (a) and leave it to the reader to show the other parts using similar arguments.

Let $\mathcal{U}$ be a generator with Lebesgue number 2δ. Then $\mathrm{P}(T,\varphi) = \underline{p}(T,\varphi,\mathcal{U})$ by Lemma 11.1.25. Set $0 < \varepsilon \le \delta$. Observe that 2ε is also a Lebesgue number for $\mathcal{U}$. Choose any sequence $(E_n(\varepsilon))_{n=1}^\infty$ of maximal (n,ε)-separated sets. Since maximal (n,ε)-separated sets are (n,ε)-spanning sets, it follows from Lemma 11.2.4(a) that $z_n(T,\varphi,\mathcal{U}) \le Q_n(T,\varphi,\varepsilon) \le \Sigma_n(E_n(\varepsilon))$. Therefore,

$$\mathrm{P}(T,\varphi) = \underline{p}(T,\varphi,\mathcal{U}) = \liminf_{n\to\infty} \frac{1}{n} \log z_n(T,\varphi,\mathcal{U}) \le \liminf_{n\to\infty} \frac{1}{n} \log \Sigma_n(E_n(\varepsilon)). \tag{11.10}$$

On the other hand, since $\mathcal{U}$ is a generator, there exists $K \in \mathbb{N}$ such that $\operatorname{diam}(\mathcal{U}^k) \leq \varepsilon$ for all $k \geq K$. It ensues from Lemma 11.2.4(b) that $\Sigma_n(E_n(\varepsilon)) \leq P_n(T, \varphi, \varepsilon) \leq Z_n(T, \varphi, \mathcal{U}^k)$ for all $k \geq K$. Consequently,

$$\limsup_{n\to\infty} \frac{1}{n} \log \Sigma_n(E_n(\varepsilon)) \leq \lim_{n\to\infty} \frac{1}{n} \log Z_n(T, \varphi, \mathcal{U}^k) = \mathrm{P}(T, \varphi, \mathcal{U}^k)$$

for all $k \geq K$. It follows from Lemma 11.1.23(g) that

$$\limsup_{n\to\infty} \frac{1}{n} \log \Sigma_n(E_n(\varepsilon)) \leq \lim_{k\to\infty} \mathrm{P}(T, \varphi, \mathcal{U}^k) = \mathrm{P}(T, \varphi). \tag{11.11}$$

Combining (11.10) and (11.11) gives (a). □

Recall that for expansive systems, the Lebesgue number can be expressed in terms of the expansive constant.

Theorem 11.2.8. *If $T : X \to X$ is a δ_0-expansive dynamical system on a compact metric space (X, d), then Theorem* 11.2.7 *applies with any* $0 < \delta < \delta_0/4$.

Proof. See the proof of Theorem 7.3.14. □

11.3 Basic properties of topological pressure

In this section, we give some of the most basic properties of topological pressure. First, we show that the addition or subtraction of a constant to the potential increases or decreases the pressure of the potential by that same constant.

Proposition 11.3.1. *Let $T : X \to X$ be a topological dynamical system and $\varphi : X \to \mathbb{R}$ a potential. For any constant $c \in \mathbb{R}$, we have* $\mathrm{P}(T, \varphi + c) = \mathrm{P}(T, \varphi) + c$.

Proof. For each $n \in \mathbb{N}$ and $\varepsilon > 0$, let $E_n(\varepsilon)$ be a maximal (n, ε)-separated set. By Theorem 11.2.1,

$$\begin{aligned}
\mathrm{P}(T, \varphi + c) &= \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \sum_{x\in E_n(\varepsilon)} e^{S_n(\varphi+c)(x)} \\
&= \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)} e^{nc} \\
&= \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \left[\log\left(\sum_{x\in E_n(\varepsilon)} e^{S_n\varphi(x)} \right) + nc \right] \\
&= \mathrm{P}(T, \varphi) + c.
\end{aligned}$$
□

Next, we show that the pressure, as a function of the potential, is increasing.

Proposition 11.3.2. *Let $T : X \to X$ be a topological dynamical system and $\varphi, \psi : X \to \mathbb{R}$ be potentials. If $\varphi \leq \psi$, then $\mathrm{P}(T, \varphi) \leq \mathrm{P}(T, \psi)$. In particular,*

$$\mathrm{h}_{\mathrm{top}}(T) + \inf \varphi \leq \mathrm{P}(T, \varphi) \leq h_{top}(T) + \sup \varphi.$$

Proof. That $\mathrm{P}(T,\varphi) \le \mathrm{P}(T,\psi)$ whenever $\varphi \le \psi$ is obvious from Theorem 11.2.1. The second statement was proved in Corollary 11.1.24 but also follows from the first statement, Proposition 11.3.1, and the fact that $0 + \inf \varphi \le \varphi \le 0 + \sup \varphi$ and $\mathrm{P}(T,0) = \mathrm{h}_{\mathrm{top}}(T)$. □

In general, it is not the case that $\mathrm{P}(T, c\,\varphi) = c\mathrm{P}(T,\varphi)$. For example, suppose that $\mathrm{P}(T,0) \ne 0$. Then the equation $\mathrm{P}(T, c\,0) = c\,\mathrm{P}(T,0)$ only holds when $c = 1$.

11.4 Examples

Example 11.4.1. Let E be a finite alphabet and let $\sigma : E^\infty \to E^\infty$ be the full E-shift map. Let $\widetilde{\varphi} : E \to \mathbb{R}$ be a function. Then the function $\varphi : E^\infty \to \mathbb{R}$ defined by $\varphi(\omega) := \widetilde{\varphi}(\omega_1)$ is a continuous function on E^∞ which depends only upon the first coordinate ω_1 of the word $\omega \in E^\infty$. We will show that

$$\mathrm{P}(\sigma,\varphi) = \log \sum_{e\in E} \exp(\widetilde{\varphi}(e)).$$

According to Example 5.1.4, the shift map σ is δ-expansive for any $0 < \delta < 1$ when E^∞ is endowed with the metric $d_s(\omega,\tau) = s^{|\omega\wedge\tau|}$, where $0 < s < 1$. Choose $\mathcal{U} = \{[e] : e \in E\}$ as (finite) open cover of E^∞. So $\mathcal{U}$ is the partition of E^∞ into its initial 1-cylinders. Since $\operatorname{diam}(\mathcal{U}) = s < 1$, Theorem 11.1.26 states that $\mathrm{P}(\sigma,\varphi) = \mathrm{P}(\sigma,\varphi,\mathcal{U})$.

In order to compute $\mathrm{P}(\sigma,\varphi,\mathcal{U})$, observe that $\mathcal{U}^n = \{[\omega] : \omega \in E^n\}$ is the partition of E^∞ into its initial n-cylinders. Then

$$\begin{aligned}
\mathrm{P}(\sigma,\varphi) = \mathrm{P}(\sigma,\varphi,\mathcal{U}) &= \lim_{n\to\infty} \frac{1}{n} \log Z_n(\varphi,\mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{U\in\mathcal{U}^n} e^{\overline{S}_n\varphi(U)} \\
&= \lim_{n\to\infty} \frac{1}{n} \log \sum_{\omega\in E^n} e^{\overline{S}_n\varphi([\omega])} \\
&= \lim_{n\to\infty} \frac{1}{n} \log \sum_{\omega_1\ldots\omega_n\in E^n} \exp(\widetilde{\varphi}(\omega_1) + \cdots + \widetilde{\varphi}(\omega_n)) \\
&= \lim_{n\to\infty} \frac{1}{n} \log\Big(\sum_{\omega_1\in E} \exp(\widetilde{\varphi}(\omega_1)) \cdots \sum_{\omega_n\in E} \exp(\widetilde{\varphi}(\omega_n)) \Big) \\
&= \lim_{n\to\infty} \frac{1}{n} \log\Big(\sum_{e\in E} \exp(\widetilde{\varphi}(e)) \Big)^n \\
&= \log \sum_{e\in E} \exp(\widetilde{\varphi}(e)).
\end{aligned}$$

Example 11.4.2. Let E be a finite alphabet and let $\sigma : E^\infty \to E^\infty$ be the full E-shift map. Let $\widetilde{\varphi} : E^2 \to \mathbb{R}$ be a function. Then the function $\varphi : E^\infty \to \mathbb{R}$ defined by $\varphi(\omega) = \widetilde{\varphi}(\omega_1,\omega_2)$ is a continuous function on E^∞ which depends only upon the first two coordinates of the word $\omega \in E^\infty$.

As in the previous example, $\mathrm{P}(\sigma,\varphi) = \mathrm{P}(\sigma,\varphi,\mathcal{U})$, where $\mathcal{U} = \{[e] : e \in E\}$ is the (finite) open partition of E^∞ into its initial 1-cylinders and

$$\mathrm{P}(\sigma,\varphi) = \mathrm{P}(\sigma,\varphi,\mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{\omega\in E^n} e^{\overline{S}_n\varphi([\omega])}.$$

But in this case

$$\begin{aligned}\sum_{\omega\in E^n} e^{\overline{S}_n\varphi([\omega])} &= \sum_{\omega\in E^n} \exp\begin{pmatrix} \widetilde{\varphi}(\omega_1,\omega_2) + \widetilde{\varphi}(\omega_2,\omega_3) + \cdots \\ + \widetilde{\varphi}(\omega_{n-1},\omega_n) + \max_{e\in E} \widetilde{\varphi}(\omega_n, e) \end{pmatrix} \\ &= \sum_{\omega_1\in E}\sum_{\omega_2\in E} e^{\widetilde{\varphi}(\omega_1,\omega_2)} \sum_{\omega_3\in E} e^{\widetilde{\varphi}(\omega_2,\omega_3)} \cdots \sum_{\omega_n\in E} e^{\widetilde{\varphi}(\omega_{n-1},\omega_n)} \cdot \max_{e\in E} \exp(\widetilde{\varphi}(\omega_n, e)).\end{aligned}$$

Since

$$m := \min_{e,f\in E} \exp(\widetilde{\varphi}(f,e)) \le \max_{e\in E} \exp(\widetilde{\varphi}(\omega_n,e)) \le \max_{e,f\in E} \exp(\widetilde{\varphi}(f,e)) =: M$$

for all $n \in \mathbb{N}$ and all $\omega_n \in E$, we have that

$$\sum_{\omega\in E^n} e^{\overline{S}_n\varphi([\omega])} \asymp \sum_{\omega_1\in E}\sum_{\omega_2\in E} e^{\widetilde{\varphi}(\omega_1,\omega_2)} \sum_{\omega_3\in E} e^{\widetilde{\varphi}(\omega_2,\omega_3)} \cdots \sum_{\omega_n\in E} e^{\widetilde{\varphi}(\omega_{n-1},\omega_n)}$$

for all n, with uniform constant of comparability $C = \max\{m^{-1}, M\}$.

Let $A : E^2 \to \mathbb{R}_+$ be the positive matrix whose entries are $A_{ef} = \exp(\widetilde{\varphi}(e,f))$. Equip this matrix with the norm $\|A\| = \sum_{e\in E}\sum_{f\in E} A_{ef}$. It is easy to prove by induction that

$$\|A^{n-1}\| = \sum_{\omega_1\in E}\sum_{\omega_2\in E} e^{\widetilde{\varphi}(\omega_1,\omega_2)} \sum_{\omega_3\in E} e^{\widetilde{\varphi}(\omega_2,\omega_3)} \cdots \sum_{\omega_n\in E} e^{\widetilde{\varphi}(\omega_{n-1},\omega_n)}$$

for all $n \ge 2$, and hence

$$\sum_{\omega\in E^n} e^{\overline{S}_n\varphi([\omega])} \asymp \|A^{n-1}\|.$$

Therefore,

$$\begin{aligned}\mathrm{P}(T,\varphi) &= \lim_{n\to\infty} \frac{1}{n} \log \sum_{\omega\in E^n} e^{\overline{S}_n\varphi([\omega])} \\ &= \lim_{n\to\infty} \frac{1}{n} \log\|A^{n-1}\| = \log \lim_{n\to\infty} \|A^n\|^{1/n} = \log r(A),\end{aligned}$$

where $r(A)$ is the spectral radius of A, that is, the largest eigenvalue of A (in absolute value).

11.5 Exercises

Exercise 11.5.1. Let $T : X \to X$ be a dynamical system and $\mathcal{U}$ be an open cover of X. Show that $Z_n(\varphi,\mathcal{U}) \neq Z_1(\varphi,\mathcal{U}^n)$ in general. That is, find a potential φ such that $Z_n(\varphi,\mathcal{U}) \neq Z_1(\varphi,\mathcal{U}^n)$ for some $n \in \mathbb{N}$.

Note: It is possible to find a potential for which the above nonequality holds for any $n > 1$.

Exercise 11.5.2. Using a symbolic dynamical system, give an example of a sequence $(z_n(\mathcal{U}))_{n=1}^{\infty}$ which is not submultiplicative.

Exercise 11.5.3. Prove Lemma 11.1.8.

Exercise 11.5.4. Show that for every $t \geq 0$ there exists a dynamical system $T : X \to X$ whose topological entropy is equal to t.

Exercise 11.5.5. Consider the full shift $\sigma : \{0,1\}^{\infty} \to \{0,1\}^{\infty}$. Let $\varphi : \{0,1\}^{\infty} \to \mathbb{R}$ be given by the formula

$$\varphi(\omega_1\omega_2\ldots) := \begin{cases} -\log 4 & \text{if } \omega_1 = 0 \\ \log 3 - \log 4 & \text{if } \omega_1 = 1. \end{cases}$$

Show that $\mathrm{P}(\sigma,\varphi) = 0$.

Exercise 11.5.6. Let $T : X \to X$ be a dynamical system. Show that the following are equivalent:
(a) $\mathrm{h}_{\mathrm{top}}(T)$ is finite.
(b) There exists a continuous function $\varphi : X \to \mathbb{R}$ such that $\mathrm{P}(T,\varphi)$ is finite.
(c) $\mathrm{P}(T,\varphi)$ is finite for every continuous function $\varphi : X \to \mathbb{R}$.

Exercise 11.5.7. Let $T : X \to X$ be a dynamical system such that $\mathrm{h}_{\mathrm{top}}(T) < \infty$. Prove that the topological pressure function $\mathrm{P}(T,\bullet) : C(X) \to \mathbb{R}$ is Lipschitz continuous with Lipschitz constant 1, and convex.

Exercise 11.5.8. Generalize Examples 11.4.1 and 11.4.2 to the case where $k \in \mathbb{N}$ and φ depends on k coordinates.

Exercise 11.5.9. Show that the pressure function is not linear; more precisely, in general $\mathrm{P}(T,t\varphi) \neq t\mathrm{P}(T,\varphi)$.

Exercise 11.5.10. Show that if $\inf\varphi \leq 0$, then the pressure function $\mathbb{R} \ni t \mapsto \mathrm{P}(T,t\varphi)$ is convex, that is,

$$\mathrm{P}\big(T,(st_1 + (1-s)t_2)\varphi\big) \leq s\,\mathrm{P}(T,t_1\varphi) + (1-s)\,\mathrm{P}(T,t_2\varphi), \quad \forall s \in [0,1],\ \forall t_1,t_2 \in \mathbb{R}.$$

Conclude that the function $\mathbb{R} \ni t \mapsto \mathrm{P}(T,t\varphi)$ is differentiable at all but at most countably many t's.

12 The variational principle and equilibrium states

In Section 12.1, we state and prove a fundamental result of thermodynamic formalism known as the *variational principle*. This deep result establishes a crucial relationship between topological dynamics and ergodic theory, by way of a formula linking topological pressure and measure-theoretic entropy. The variational principle in its classical form and full generality was proved in [75] and [10]. The proof we present follows that of Michal Misiurewicz [49], which is particularly elegant, short, and simple.

In Section 12.2, we introduce the concept of equilibrium states, give sufficient conditions for their existence, such as the upper semicontinuity of the metric entropy function (which prevails under any expansive system). We single out a special class of equilibrium states, those corresponding to a potential identically equal to zero, and following tradition, call them measures of maximal entropy. We do not deal in this chapter with the issue of the uniqueness of equilibrium states. Nevertheless, we provide an example of a topological dynamical system with positive and finite topological entropy which does not have any measure of maximal entropy.

12.1 The variational principle

For any topological dynamical system $T : X \to X$, subject to a potential $\varphi : X \to \mathbb{R}$ and equipped with a T-invariant measure μ, the quantity $\mathrm{h}_\mu(T) + \int \varphi\, d\mu$ is called the *free energy* of the system T with respect to μ under the potential φ. The variational principle states that the topological pressure of a system is the supremum of the free energy generated by that system.

Recall that $M(T)$ is the set of all T-invariant Borel probability measures on X and that by definition any potential φ is continuous.

Theorem 12.1.1 (Variational principle). *Let $T : X \to X$ be a topological dynamical system and $\varphi : X \to \mathbb{R}$ a potential. Then*

$$\mathrm{P}(T, \varphi) = \sup\left\{\mathrm{h}_\mu(T) + \int_X \varphi\, d\mu : \mu \in M(T)\right\}.$$

Remark 12.1.2. In fact, as we shall see in Corollary 12.1.10, the supremum can be restricted to the subset $E(T)$ of ergodic measures in $M(T)$.

The proof of the variational principle will be given in two parts. In Part I, we will show that $\mathrm{P}(T, \varphi) \geq \mathrm{h}_\mu(T) + \int \varphi\, d\mu$ for every measure $\mu \in M(T)$. Part II consists in the proof of the inequality $\sup\{\mathrm{h}_\mu(T) + \int \varphi\, d\mu : \mu \in M(T)\} \geq \mathrm{P}(T, \varphi)$.

The first part is relatively easier to prove than the second one. In the proof of Part I, we will need Jensen's inequality. Recall that a function $k : I \to \mathbb{R}$, where $I \subseteq \mathbb{R}$ is an

https://doi.org/10.1515/9783110702682-012

interval, is convex on I if

$$\psi(tx + (1-t)y) \le t\psi(x) + (1-t)\psi(y), \quad \forall t \in [0,1], \forall x, y \in I.$$

Theorem 12.1.3 (Jensen's inequality). *Let $(X, \mathcal{A}, \mu)$ be a probability space. Let $-\infty \le a < b \le \infty$ and $\psi : (a,b) \to \mathbb{R}$ be a convex function. If $f \in L^1(\mu)$ and $f(X) \subseteq (a,b)$, then*

$$\psi\Big(\int_X f\, d\mu\Big) \le \int_X \psi \circ f\, d\mu.$$

Proof. See, for instance, [58]. □

We shall also need the following lemma, which states that any finite Borel partition α of X can be, from a measure-theoretic entropy viewpoint, approximated as closely as desired by a finite Borel partition β whose elements are compact and are, with one exception, contained in those of α.

Lemma 12.1.4. *Let $\mu \in M(X)$, let $\alpha := \{A_1, \dots, A_n\}$ be a finite Borel partition of X, and let $\varepsilon > 0$. Then there exist compact sets $B_i \subseteq A_i$, $1 \le i \le n$, such that the partition $\beta := \{B_1, \dots, B_n, X\backslash(B_1 \cup \cdots \cup B_n)\}$ satisfies*

$$\mathrm{H}_\mu(\alpha|\beta) \le \varepsilon.$$

Proof. Let the measure μ and the partition α be as stated and let $\varepsilon > 0$. Recall from Definition 9.3.4 the nonnegative continuous function $k : [0,1] \to [0,1]$ defined by

$$k(t) = -t \log t,$$

where it is understood that $0 \cdot (-\infty) = 0$. The continuity of k at 0 implies that there exists $\delta > 0$ such that $k(t) < \varepsilon/n$ when $0 \le t < \delta$. Since μ is regular and X is compact, for each $1 \le i \le n$ there exists a compact set $B_i \subseteq A_i$ such that $\mu(A_i \backslash B_i) < \delta$. Then $k(\mu(A_i \backslash B_i)) < \varepsilon/n$ for all $1 \le i \le n$. Observe further that $X \backslash \bigcup_{j=1}^n B_j = \bigcup_{j=1}^n A_j \backslash B_j$. By Definition 9.4.2 of conditional entropy, it follows that

$$\begin{aligned}
\mathrm{H}_\mu(\alpha|\beta) &= \sum_{j=1}^n \sum_{i=1}^n -\mu(A_i \cap B_j) \log \frac{\mu(A_i \cap B_j)}{\mu(B_j)} \\
&\quad + \sum_{i=1}^n -\mu(A_i \cap (X\backslash \cup_{j=1}^n B_j)) \log \frac{\mu(A_i \cap (X\backslash \cup_{j=1}^n B_j))}{\mu(X\backslash \cup_{j=1}^n B_j)} \\
&= \sum_{j=1}^n -\mu(B_j) \log \frac{\mu(B_j)}{\mu(B_j)} + \sum_{i=1}^n -\mu(A_i \cap (\cup_{j=1}^n A_j \backslash B_j)) \log \frac{\mu(A_i \cap (\cup_{j=1}^n A_j \backslash B_j))}{\mu(\cup_{j=1}^n A_j \backslash B_j)} \\
&= 0 + \sum_{i=1}^n -\mu(A_i \backslash B_i) \log \frac{\mu(A_i \backslash B_i)}{\mu(\cup_{j=1}^n A_j \backslash B_j)}
\end{aligned}$$

$$= \sum_{i=1}^{n} -\mu(A_i \backslash B_i)\big[\log \mu(A_i \backslash B_i) - \log \mu\big(\cup_{j=1}^{n} A_j \backslash B_j\big)\big]$$
$$= \sum_{i=1}^{n} k\big(\mu(A_i \backslash B_i)\big) + \sum_{i=1}^{n} \mu(A_i \backslash B_i) \log \mu\big(\cup_{j=1}^{n} A_j \backslash B_j\big)$$
$$\leq \sum_{i=1}^{n} k\big(\mu(A_i \backslash B_i)\big) \leq n \cdot \frac{\varepsilon}{n} = \varepsilon. \qquad \square$$

We are now in a position to begin the proof of the first part of the variational principle.

Proof of Part I. Recall that our aim is to establish the inequality

$$\mathrm{P}(T, \varphi) \geq \mathrm{h}_\mu(T) + \int_X \varphi \, d\mu, \quad \forall \mu \in M(T). \tag{12.1}$$

We claim that it is sufficient to prove that there exists a constant $C \in \mathbb{R}$, independent of T, φ and μ, such that

$$\mathrm{P}(T, \varphi) \geq \mathrm{h}_\mu(T) + \int_X \varphi \, d\mu + C. \tag{12.2}$$

Indeed, suppose that such a constant exists. In particular, this means that this constant works not only for the system (X, T) under the potential φ and a measure $\mu \in M(T)$ but also for any higher-iterate system (X, T^n) under the potential $S_n\varphi = \sum_{k=0}^{n-1} \varphi \circ T^k$ and the same measure μ, since any T-invariant measure is T^n-invariant. Fix temporarily $n \in \mathbb{N}$. Using successively Theorem 11.1.22, inequality (12.2) with the quadruple $(X, T^n, S_n\varphi, \mu)$ instead of (X, T, φ, μ), and Theorems 9.4.13 and 8.1.18, we then obtain that

$$n\,\mathrm{P}(T, \varphi) = \mathrm{P}(T^n, S_n\varphi) \geq \mathrm{h}_\mu(T^n) + \int_X S_n\varphi \, d\mu + C = n\,\mathrm{h}_\mu(T) + n \int_X \varphi \, d\mu + C.$$

Dividing by n and letting n tend to infinity yields inequality (12.1).

Of course, to obtain (12.2) it suffices to show that

$$\mathrm{P}(T, \varphi) \geq \mathrm{h}_\mu(T, \alpha) + \int_X \varphi \, d\mu + C \tag{12.3}$$

for all finite Borel partitions α of X (see Definition 9.4.12). So let α be any such partition and let $\varepsilon > 0$. To obtain (12.3), it is enough to prove that

$$\mathrm{P}(T, \varphi) \geq \mathrm{h}_\mu(T, \alpha) + \int_X \varphi \, d\mu + C - 2\varepsilon. \tag{12.4}$$

By Theorem 11.2.1, it suffices to demonstrate that

$$\limsup_{n\to\infty} \frac{1}{n} \log \sum_{y \in E_n(\delta)} e^{S_n\varphi(y)} \geq \mathrm{h}_\mu(T, \alpha) + \int_X \varphi \, d\mu + C - 2\varepsilon \tag{12.5}$$

for all sufficiently small $\delta > 0$ and some family $\{E_n(\delta) : n \in \mathbb{N}, \delta > 0\}$ of (n, δ)-separated sets. In light of Definition 9.4.10 and of Theorem 8.1.18, it is sufficient to prove that

$$\frac{1}{n} \log \sum_{y \in E_n(\delta)} e^{S_n\varphi(y)} \geq \frac{1}{n} \mathrm{H}_\mu(\alpha^n) + \frac{1}{n} \int_X S_n\varphi \, d\mu + C - 2\varepsilon \tag{12.6}$$

for all sufficiently small $\delta > 0$, all large enough $n \in \mathbb{N}$ and all (n, δ)-separated sets $E_n(\delta)$.

To this end, let β be the finite Borel partition given by Lemma 12.1.4. Then $\mathrm{H}_\mu(\alpha|\beta) \leq \varepsilon$. Momentarily fix $n \in \mathbb{N}$. By Theorem 9.4.3(i) and Lemma 9.4.7(c), we know that

$$\mathrm{H}_\mu(\alpha^n) \leq \mathrm{H}_\mu(\beta^n) + \mathrm{H}_\mu(\alpha^n|\beta^n) \leq \mathrm{H}_\mu(\beta^n) + n\,\mathrm{H}_\mu(\alpha|\beta) \leq \mathrm{H}_\mu(\beta^n) + n\varepsilon. \tag{12.7}$$

From (12.6) and (12.7), it thus suffices to establish that

$$\log \sum_{y \in E_n(\delta)} e^{S_n\varphi(y)} \geq \mathrm{H}_\mu(\beta^n) + \int_X S_n\varphi \, d\mu + (C - \varepsilon)n \tag{12.8}$$

for all sufficiently small $\delta > 0$, all large enough $n \in \mathbb{N}$ and all (n, δ)-separated sets $E_n(\delta)$. To prove this inequality, we will estimate the term $\mathrm{H}_\mu(\beta^n) + \int S_n\varphi \, d\mu$ from above. Since the logarithm function is concave (so its negative is convex), Jensen's inequality (Theorem 12.1.3) implies that

$$\begin{aligned}
\mathrm{H}_\mu(\beta^n) + \int_X S_n\varphi \, d\mu &\leq \sum_{B \in \beta^n} \mu(B)[-\log \mu(B) + S_n\varphi(B)] \\
&= \sum_{B \in \beta^n} \mu(B) \log \frac{\exp(S_n\varphi(B))}{\mu(B)} \\
&= \int_X \log \frac{\exp(S_n\varphi(\beta^n(x)))}{\mu(\beta^n(x))} \, d\mu(x) \\
&\leq \log \int_X \frac{\exp(S_n\varphi(\beta^n(x)))}{\mu(\beta^n(x))} \, d\mu(x) \\
&= \log \sum_{B \in \beta^n} e^{S_n\varphi(B)}.
\end{aligned} \tag{12.9}$$

Since each set $B_i \in \beta$ is compact, it follows that $d(B_i, B_j) > 0$ for all $i \neq j$. As φ is uniformly continuous, let $0 < \delta < \frac{1}{2} \min\{d(B_i, B_j) : i \neq j\}$ be such that

$$d(x, y) < \delta \Longrightarrow |\varphi(x) - \varphi(y)| < \varepsilon. \tag{12.10}$$

Now, consider an arbitrary maximal (n,δ)-separated set $E_n(\delta)$ and fix temporarily $B \in \beta$. According to Lemma 7.3.7, each maximal (n,δ)-separated set is an (n,δ)-spanning set. So for every $x \in B$, there exists $y \in E_n(\delta)$ such that $x \in B_n(y,\delta)$ and, therefore, $|S_n\varphi(x) - S_n\varphi(y)| < n\varepsilon$ by (12.10). As the set $E_n(\delta)$ is finite, there is $y_B \in E_n(\delta)$ such that

$$S_n\varphi(B) \le S_n\varphi(y_B) + n\varepsilon \quad \text{and} \quad B \cap B_n(y_B,\delta) \ne \emptyset. \tag{12.11}$$

Moreover, since $d(B_i,B_j) > 2\delta$ for each $i \ne j$, any ball $B(z,\delta)$, $z \in X$, intersects at most one B_i and perhaps $X \setminus \bigcup_j B_j$. Hence,

$$\#\{B \in \beta : B \cap B(z,\delta) \ne \emptyset\} \le 2 \tag{12.12}$$

for all $z \in X$. Thus,

$$\#\{B \in \beta^n : B \cap B_n(z,\delta) \ne \emptyset\} \le 2^n \tag{12.13}$$

for all $z \in X$. So the function $f : \beta^n \to E_n(\delta)$ defined by $f(B) = y_B$ is at most 2^n-to-one. Consequently, by (12.11) we obtain that

$$2^n \sum_{y \in E_n(\delta)} e^{S_n\varphi(y)} \ge \sum_{B \in \beta^n} e^{S_n\varphi(y_B)} \ge \sum_{B \in \beta^n} e^{S_n\varphi(B)} \cdot e^{-n\varepsilon}.$$

Multiplying both sides by 2^{-n}, then taking the logarithm of both sides and applying (12.9) yields

$$\begin{aligned}\log \sum_{y \in E_n(\delta)} e^{S_n\varphi(y)} &\ge \log \sum_{B \in \beta^n} e^{S_n\varphi(B)} - n\varepsilon - n\log 2 \\ &\ge \mathrm{H}_\mu(\beta^n) + \int_X S_n\varphi \, d\mu + n(-\log 2 - \varepsilon).\end{aligned}$$

This inequality, which is nothing other than the sought inequality (12.8) with $C = -\log 2$, holds for all $0 < \delta < \frac{1}{2}\min\{d(B_i,B_j) : i \ne j\}$, all $n \in \mathbb{N}$ and all maximal (n,δ)-separated sets $E_n(\delta)$. This concludes the proof of Part I. □

Remark 12.1.5. Observe that the constant $C = -\log 2$ originates from relation (12.12), and thus depends solely on the **existence** of the Borel partition β, which is ensured by Lemma 12.1.4.

Let us move on to the proof of Part II of the variational principle. In addition to Lemma 9.6.1, we shall need the following three lemmas.

The first of those states that given any finite Borel partition α whose atoms have boundaries with zero μ-measure, the entropy of α, as a function of the underlying Borel probability measure, is continuous at μ.

Lemma 12.1.6. *Let $\mu \in M(X)$. If α is a finite Borel partition of X such that $\mu(\partial A) = 0$ for all $A \in \alpha$, then the function*

$$\begin{array}{rcccl} \mathrm{H}_{\bullet}(\alpha) & : & M(X) & \longrightarrow & [0,\infty] \\ & & \nu & \longmapsto & \mathrm{H}_{\nu}(\alpha) \end{array}$$

is continuous at μ.

Proof. This follows directly from the fact that according to the Portmanteau theorem (Theorem A.1.56), a sequence of Borel probability measures $(\mu_n)_{n=1}^{\infty}$ converges weakly* to a measure μ if and only if $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for every Borel set A with $\mu(\partial A) = 0$. Recall that $\mathrm{H}_{\nu}(\alpha) = -\sum_{A\in\alpha} \nu(A)\log\nu(A)$. □

In the second lemma, we show that the entropy of α, as a function of the underlying Borel probability measure, is concave.

Lemma 12.1.7. *For any finite Borel partition α of X, the function $\mathrm{H}_{\bullet}(\alpha)$ is concave.*

Proof. Let α be a finite Borel partition of X, and μ and ν Borel probability measures on X. Let also $t \in (0,1)$. Since the function $k(x) = -x\log x$ is concave, for each $A \in \alpha$ we have

$$k\big(t\mu(A) + (1-t)\nu(A)\big) \geq t\,k\big(\mu(A)\big) + (1-t)k\big(\nu(A)\big).$$

Therefore,

$$\begin{aligned} \mathrm{H}_{t\mu+(1-t)\nu}(\alpha) &= \sum_{A\in\alpha} k\big(t\mu(A) + (1-t)\nu(A)\big) \\ &\geq t\sum_{A\in\alpha} k\big(\mu(A)\big) + (1-t)\sum_{A\in\alpha} k\big(\nu(A)\big) \\ &= t\mathrm{H}_{\mu}(\alpha) + (1-t)\mathrm{H}_{\nu}(\alpha). \end{aligned}$$ □

Finally, the third lemma is a generalization of the Krylov–Bogolyubov theorem (Theorem 8.1.22).

Lemma 12.1.8. *Let $T : X \to X$ be a dynamical system. If $(\mu_n)_{n=1}^{\infty}$ is a sequence of measures in $M(X)$, then every* weak* *limit point of the sequence $(m_n)_{n=1}^{\infty}$, where*

$$m_n := \frac{1}{n}\sum_{i=0}^{n-1} \mu_n \circ T^{-i},$$

is a T-invariant measure.

Proof. By the compactness of $M(X)$, the sequence $(m_n)_{n=1}^{\infty}$ has accumulation points. Let $(m_{n_j})_{j=1}^{\infty}$ be a subsequence which converges weakly* to, say, $m \in M(X)$. Let $f \in C(X)$.

Using Lemma 8.1.2, we obtain that

$$\begin{aligned}\left|\int_X f\circ T\,dm-\int_X f\,dm\right| &= \lim_{j\to\infty}\left|\int_X f\circ T\,dm_{n_j}-\int_X f\,dm_{n_j}\right|\\ &= \lim_{j\to\infty}\left|\frac{1}{n_j}\int_X\sum_{i=0}^{n_j-1}(f\circ T^{i+1}-f\circ T^i)\,d\mu_{n_j}\right|\\ &= \lim_{j\to\infty}\frac{1}{n_j}\left|\int_X(f\circ T^{n_j}-f)\,d\mu_{n_j}\right|\\ &\le \lim_{j\to\infty}\frac{2\|f\|_\infty}{n_j}=0.\end{aligned}$$

Thus, by Theorem 8.1.18 the measure m is T-invariant. □

We are now ready to prove Part II of the variational principle.

Proof of Part II. We aim to show that

$$\sup\left\{\mathrm{h}_\mu(T)+\int_X\varphi\,d\mu:\mu\in M(T)\right\}\ge \mathrm{P}(T,\varphi).$$

Fix $\varepsilon>0$. Let $(E_n(\varepsilon))_{n=1}^\infty$ be a sequence of maximal (n,ε)-separated sets in X. For every $n\in\mathbb{N}$, define the measures μ_n and m_n by

$$\mu_n:=\frac{\sum_{x\in E_n(\varepsilon)}e^{S_n\varphi(x)}\delta_x}{\sum_{x\in E_n(\varepsilon)}e^{S_n\varphi(x)}}\quad\text{and}\quad m_n:=\frac{1}{n}\sum_{k=0}^{n-1}\mu_n\circ T^{-k},$$

where δ_x denotes the Dirac measure concentrated at the point x. Let $(n_i)_{i=1}^\infty$ be a strictly increasing sequence in $\mathbb{N}$ such that $(m_{n_i})_{i=1}^\infty$ converges weakly* to, say, m, and such that

$$\lim_{i\to\infty}\frac{1}{n_i}\log\sum_{x\in E_{n_i}(\varepsilon)}e^{S_n\varphi(x)}=\limsup_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)}e^{S_n\varphi(x)}.\tag{12.14}$$

For ease of exposition, define

$$s_n:=\sum_{x\in E_n(\varepsilon)}e^{S_n\varphi(x)}\quad\text{and}\quad \mu(x):=\mu(\{x\}).$$

From Lemma 12.1.8, the limit measure m belongs to $M(T)$. Also, in view of Lemma 9.6.1, there exists a finite Borel partition α such that $\mathrm{diam}(\alpha)<\varepsilon$ and $m(\partial A)=0$ for all $A\in\alpha$. Since $\#(A\cap E_n(\varepsilon))\le 1$ for all $A\in\alpha^n$, we obtain that

$$\begin{aligned}\mathrm{H}_{\mu_n}(\alpha^n)+\int_X S_n\varphi\,d\mu_n &= \sum_{x\in E_n(\varepsilon)}\mu_n(x)[-\log\mu_n(x)+S_n\varphi(x)]\\ &= \sum_{x\in E_n(\varepsilon)}\frac{e^{S_n\varphi(x)}}{s_n}\left[-\log\frac{e^{S_n\varphi(x)}}{s_n}+S_n\varphi(x)\right]\end{aligned}$$

$$= \frac{1}{s_n} \sum_{x \in E_n(\varepsilon)} e^{S_n\varphi(x)} [-S_n\varphi(x) + \log s_n + S_n\varphi(x)]$$
$$= \log s_n = \log \sum_{x \in E_n(\varepsilon)} e^{S_n\varphi(x)}. \tag{12.15}$$

Now, fix $M \in \mathbb{N}$ and let $n \geq 2M$. For $j = 0, 1, \ldots, M-1$, define $s(j) := \lfloor \frac{n-j}{M} \rfloor - 1$, where $\lfloor r \rfloor$ denotes the integer part of r. Note that

$$\bigvee_{k=0}^{s(j)} T^{-(kM+j)}(\alpha^M) = \bigvee_{\ell=j}^{(s(j)+1)M+j-1} T^{-\ell}(\alpha)$$

and

$$(s(j)+1)M + j - 1 = \left\lfloor \frac{n-j}{M} \right\rfloor M + j - 1 \leq n - j + j - 1 = n - 1.$$

Observe also that

$$(n-1) - ((s(j)+1)M + j) = n - 1 - \left(\left\lfloor \frac{n-j}{M} \right\rfloor M + j \right)$$
$$\leq n - 1 - \left(\frac{n-j}{M} - 1 \right) M - j = M - 1.$$

Setting $R_j := \{0, 1, \ldots, j-1\} \cup \{(s(j)+1)M + j, \ldots, n-1\}$, we have $\#R_j \leq 2M$ and

$$\alpha^n = \bigvee_{k=0}^{s(j)} T^{-(kM+j)}(\alpha^M) \vee \bigvee_{i \in R_j} T^{-i}(\alpha).$$

Hence, using Theorem 9.4.3(g) and (9.2), we get that

$$\mathrm{H}_{\mu_n}(\alpha^n) \leq \sum_{k=0}^{s(j)} \mathrm{H}_{\mu_n}(T^{-(kM+j)}(\alpha^M)) + \mathrm{H}_{\mu_n}\left(\bigvee_{i \in R_j} T^{-i}(\alpha) \right)$$
$$\leq \sum_{k=0}^{s(j)} \mathrm{H}_{\mu_n \circ T^{-(kM+j)}}(\alpha^M) + \log \# \left(\bigvee_{i \in R_j} T^{-i}(\alpha) \right)$$
$$\leq \sum_{k=0}^{s(j)} \mathrm{H}_{\mu_n \circ T^{-(kM+j)}}(\alpha^M) + \log(\#\alpha)^{\#R_j}$$
$$\leq \sum_{k=0}^{s(j)} \mathrm{H}_{\mu_n \circ T^{-(kM+j)}}(\alpha^M) + 2M \log \#\alpha.$$

Summing over all $j = 0, 1, \ldots, M-1$ and using Lemma 12.1.7, we obtain

$$M\mathrm{H}_{\mu_n}(\alpha^n) \leq \sum_{j=0}^{M-1} \sum_{k=0}^{s(j)} \mathrm{H}_{\mu_n \circ T^{-(kM+j)}}(\alpha^M) + 2M^2 \log \#\alpha$$
$$\leq \sum_{l=0}^{n-1} \mathrm{H}_{\mu_n \circ T^{-l}}(\alpha^M) + 2M^2 \log \#\alpha$$

$$\begin{aligned}&\le n\mathrm{H}_{\frac{1}{n}\sum_{l=0}^{n-1}\mu_n\circ T^{-l}}(\alpha^M)+2M^2\log\#\alpha\\&=n\mathrm{H}_{m_n}(\alpha^M)+2M^2\log\#\alpha.\end{aligned}$$

Adding $M\int_X S_n\varphi\,d\mu_n$ to both sides and applying (12.15) yields

$$M\log\sum_{x\in E_n(\varepsilon)}e^{S_n\varphi(x)}\le n\mathrm{H}_{m_n}(\alpha^M)+M\int_X S_n\varphi\,d\mu_n+2M^2\log\#\alpha.$$

As $\frac{1}{n}\int_X S_n\varphi\,d\mu_n=\int_X\varphi\,dm_n$ by Lemma 8.1.2, dividing both sides of the above inequality by Mn gives us that

$$\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)}e^{S_n\varphi(x)}\le\frac{1}{M}\mathrm{H}_{m_n}(\alpha^M)+\int_X\varphi\,dm_n+\frac{2M}{n}\log\#\alpha.$$

Since $\partial T^{-1}(A)\subseteq T^{-1}(\partial A)$ for every set $A\subseteq X$ and $\partial(A\cap B)\subset\partial A\cup\partial B$ for all sets $A,B\in X$, the m-measure of the boundary of each atom of the partition α^M is, as for α, equal to zero. Therefore, upon letting n tend to infinity along the subsequence $(n_i)_{i=1}^\infty$, we know that $(m_{n_i})_{i=1}^\infty$ converges weakly* to m and that (12.14) holds, so we infer from the above inequality and from Lemma 12.1.6 that

$$\limsup_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)}e^{S_n\varphi(x)}\le\frac{1}{M}\mathrm{H}_m(\alpha^M)+\int_X\varphi\,dm.$$

Letting $M\to\infty$, we obtain by Definition 9.4.10 that

$$\limsup_{n\to\infty}\frac{1}{n}\log\sum_{x\in E_n(\varepsilon)}e^{S_n\varphi(x)}\le\mathrm{h}_m(T,\alpha)+\int_X\varphi\,dm\le\sup\Big\{\mathrm{h}_\mu(T)+\int_X\varphi\,d\mu:\mu\in M(T)\Big\}.$$

As $\varepsilon>0$ is arbitrary, Theorem 11.2.1 yields the desired inequality. This completes the proof of Part II. □

12.1.1 Consequences of the variational principle

Let us now state some immediate consequences of the variational principle. A first consequence concerns the topological entropy of the system. The topological entropy of a system is the supremum of all measure-theoretic entropies of the system.

Corollary 12.1.9. $\mathrm{h}_{\mathrm{top}}(T)=\sup\{\mathrm{h}_\mu(T):\mu\in M(T)\}$.

Proof. This follows directly upon letting $\varphi\equiv 0$. □

Furthermore, the pressure of the system is determined by the supremum of the free energy of the system with respect to its ergodic measures. Recall that $E(T)$ denotes the subset of ergodic measures in $M(T)$.

Corollary 12.1.10. *For every $\mu \in M(T)$, there exists $\nu \in E(T)$ such that $\mathrm{h}_\nu(T) + \int_X \varphi\, d\nu \geq \mathrm{h}_\mu(T) + \int_X \varphi\, d\mu$. Consequently,*

$$\mathrm{P}(T,\varphi) = \sup\Big\{\mathrm{h}_\nu(T) + \int_X \varphi\, d\nu : \nu \in E(T)\Big\}.$$

Proof. Let $\mu \in M(T)$. According to Theorem 8.2.26, the measure μ has a decomposition into ergodic measures. More precisely, there exists a Borel probability space $(Y, \mathcal{B}(Y), \tau)$ and a measurable map $Y \ni y \mapsto \mu_y \in M(X)$ such that $\mu_y \in E(T)$ for τ-almost every $y \in Y$ and $\mu = \int_Y \mu_y\, d\tau(y)$. Then

$$\int_X \varphi\, d\mu = \int_Y \Big(\int_X \varphi\, d\mu_y\Big)\, d\tau(y).$$

Moreover, using a generalization of Exercise 9.7.10, we have that

$$\mathrm{h}_\mu(T) = \mathrm{h}_{\int_Y \mu_y\, d\tau(y)}(T) = \int_Y \mathrm{h}_{\mu_y}(T)\, d\tau(y).$$

It follows that

$$\mathrm{h}_\mu(T) + \int_X \varphi\, d\mu = \int_Y \Big[\mathrm{h}_{\mu_y}(T) + \int_X \varphi\, d\mu_y\Big]\, d\tau(y).$$

It is a well-known fact from measure theory (a simple consequence of Lemma A.1.34(a)) that there is $Z \in \mathcal{B}(Y)$ such that $\tau(Z) > 0$ and $\mathrm{h}_{\mu_z}(T) + \int_X \varphi\, d\mu_z \geq \mathrm{h}_\mu(T) + \int_X \varphi\, d\mu$ for every $z \in Z$. Given that $\tau(E(T)) = 1$, it follows that $\tau(Z \cap E(T)) > 0$. So there exists $\nu \in E(T)$ such that $\mathrm{h}_\nu(T) + \int_X \varphi\, d\nu \geq \mathrm{h}_\mu(T) + \int_X \varphi\, d\mu$. □

We now show that the pressure function is Lipschitz continuous.

Corollary 12.1.11. *If $T : X \to X$ is a dynamical system such that $\mathrm{h}_{\mathrm{top}}(T) < \infty$, then the pressure function $\mathrm{P}(T, \bullet) : C(X) \to \mathbb{R}$ is Lipschitz continuous with Lipschitz constant 1.*

Proof. Let $\psi, \varphi \in C(X)$. Let also $\varepsilon > 0$. By the variational principle, there exists $\mu \in M(T)$ such that

$$\mathrm{P}(T,\psi) \leq \mathrm{h}_\mu(T) + \int_X \psi\, d\mu + \varepsilon.$$

Then, using the variational principle once again, we get

$$\mathrm{P}(T,\psi) \leq \mathrm{h}_\mu(T) + \int_X \varphi\, d\mu + \int_X (\psi - \varphi)\, d\mu + \varepsilon \leq \mathrm{P}(T,\varphi) + \|\psi - \varphi\|_\infty + \varepsilon.$$

Since this is true for all $\varepsilon > 0$, we conclude that

$$\mathrm{P}(T,\psi) - \mathrm{P}(T,\varphi) \leq \|\psi - \varphi\|_\infty.$$ □

Finally, we show that the pressure of any subsystem is at most the pressure of the entire system.

Corollary 12.1.12. *If $T : X \to X$ is a topological dynamical system, $\varphi : X \to \mathbb{R}$ a potential and Y a closed T-invariant subset of X, then* $\mathrm{P}(T|_Y, \varphi|_Y) \leq \mathrm{P}(T, \varphi)$.

Proof. Each $T|_Y$-invariant measure μ on Y generates the T-invariant measure $\overline{\mu}(B) = \mu(B \cap Y)$ on X and $\overline{\mu}$ is such that $\mathrm{h}(T, \overline{\mu}) = \mathrm{h}(T|_Y, \mu)$ and $\int_X \varphi \, d\overline{\mu} = \int_Y \varphi \, d\mu$. ☐

12.2 Equilibrium states

In light of the variational principle, the measures that maximize the free energy of the system, that is, the measures which respect to which the free energy of the system coincides with its pressure, are given a special name.

Definition 12.2.1. Let $T : X \to X$ be a topological dynamical system and $\varphi : X \to \mathbb{R}$ a potential. A measure $\mu \in M(T)$ is called an *equilibrium state* for φ provided that

$$\mathrm{P}(T, \varphi) = \mathrm{h}_\mu(T) + \int_X \varphi \, d\mu.$$

Notice that if a given potential φ has an equilibrium state, then φ has an ergodic equilibrium state according to Corollary 12.1.10. When $\varphi \equiv 0$, the equilibrium states are also called *measures of maximal entropy*, that is, measures for which $\mathrm{h}_\mu(T) = \mathrm{h}_{\mathrm{top}}(T)$. In particular, if $\mathrm{h}_{\mathrm{top}}(T) = 0$, then every invariant measure is a measure of maximal entropy for T. Recall that this is the case for homeomorphisms of the unit circle (see Exercise 7.6.10), among other examples.

A simple consequence of the variational principle is the following.

Theorem 12.2.2. *If $T : X \to X$ is a topological dynamical system and $\varphi : X \to \mathbb{R}$ is a Hölder continuous potential such that* $\mathrm{P}(T, \varphi) > \sup \varphi$*, then*

$$\mathrm{h}_\mu(T) > 0$$

for every equilibrium state μ of φ.

Proof. Since μ is an equilibrium state for φ, we have that

$$\mathrm{P}(T, \varphi) = \mathrm{h}_\mu(T) + \int_X \varphi \, d\mu \leq \mathrm{h}_\mu(T) + \sup \varphi.$$

Rearranging the terms,

$$\mathrm{h}_\mu(T) \geq \mathrm{P}(T, \varphi) - \sup \varphi > 0. \qquad \square$$

It is natural to wonder whether equilibrium states exist for all topological dynamical systems. As the following example demonstrates, the answer is negative.

Example 12.2.3. We construct a system with positive, finite topological entropy but without any measure of maximal entropy. Let $(T_n : X_n \to X_n)_{n=1}^{\infty}$ be a sequence of topological dynamical systems with the property that

$$\mathrm{h}_{\mathrm{top}}(T_n) < \mathrm{h}_{\mathrm{top}}(T_{n+1}), \ \forall n \in \mathbb{N} \quad \text{and} \quad \sup_{n\in\mathbb{N}} \mathrm{h}_{\mathrm{top}}(T_n) < \infty.$$

Let $\bigsqcup_{n=1}^{\infty} X_n$ denote the disjoint union of the spaces X_n, and let $X = \{\omega\} \cup \bigsqcup_{n=1}^{\infty} X_n$ be the one-point compactification of $\bigsqcup_{n=1}^{\infty} X_n$. Define the map $T : X \to X$ by

$$T(x) := \begin{cases} T_n(x) & \text{if } x \in X_n \\ \omega & \text{if } x = \omega. \end{cases}$$

Then T is continuous. Suppose that μ is an ergodic measure of maximal entropy for T. Then $\mu(\{\omega\}) \in \{0,1\}$ since $T^{-1}(\{\omega\}) = \{\omega\}$. But if $\mu(\{\omega\}) = 1$, then we would have $\mu(\bigsqcup_{n=1}^{\infty} X_n) = 0$. Hence, on one hand, we would have $\mathrm{h}_\mu(T) = 0$, while, on the other hand, $\mathrm{h}_\mu(T) = \mathrm{h}_{\mathrm{top}}(T) \geq \sup_{n\in\mathbb{N}} \mathrm{h}_{\mathrm{top}}(T_n) > 0$. This contradiction imposes that $\mu(\{\omega\}) = 0$. Similarly, $\mu(X_n) \in \{0,1\}$ for all $n \in \mathbb{N}$ since $T^{-1}(X_n) = X_n$. Therefore, there exists a unique $N \in \mathbb{N}$ such that $\mu(X_N) = 1$. It follows that

$$\mathrm{h}_{\mathrm{top}}(T) = \mathrm{h}_\mu(T) = \mathrm{h}_\mu(T_N) \leq \mathrm{h}_{\mathrm{top}}(T_N) < \sup_{n\in\mathbb{N}} \mathrm{h}_{\mathrm{top}}(T_n) \leq \mathrm{h}_{\mathrm{top}}(T).$$

This contradiction implies that there is no measure of maximal entropy for the system T.

Given that equilibrium states do not always exist, we would like to find conditions under which they do exist. But since the function $\mu \mapsto \int_X \varphi \, d\mu$ is continuous in the weak* topology on the compact space $M(T)$, the function $\mu \mapsto \mathrm{h}_\mu(T)$ cannot be continuous in general. Otherwise, the sum of these last two functions would be continuous and would hence attain a maximum on the compact space $M(T)$, that is, equilibrium states would always exist. Nevertheless, the function $\mu \mapsto \mathrm{h}_\mu(T)$ is sometimes upper semicontinuous and this is sufficient to ensure the existence of an equilibrium state. Let us first recall the notion of upper (and lower) semicontinuity.

Definition 12.2.4. Let X be a topological space. A function $f : X \to [-\infty, \infty]$ is *upper semicontinuous* if for all $x \in X$,

$$\limsup_{y\to x} f(y) \leq f(x).$$

Equivalently, f is upper semicontinuous if the set $\{x \in X : f(x) < r\}$ is open in X for all $r \in \mathbb{R}$. A function $f : X \to [-\infty, \infty]$ is *lower semicontinuous* if $-f$ is upper semicontinuous.

Evidently, a function $f : X \to [-\infty, \infty]$ is continuous if and only if it is both upper and lower semicontinuous. Like continuous functions, upper semicontinuous functions attain their upper bound (while lower semicontinuous functions reach their lower bound) on every compact set.

One class of dynamical systems for which the function $\mu \mapsto \mathrm{h}_\mu(T)$ is upper semicontinuous are the expansive maps T.

Theorem 12.2.5. *If $T : X \to X$ is expansive, then the function*

$$\begin{array}{cccc} \mathrm{h}_\bullet(T) & : & M(T) & \longrightarrow & [0,\infty] \\ & & \mu & \longmapsto & \mathrm{h}_\mu(T) \end{array}$$

is upper semicontinuous. Hence, each potential $\varphi : X \to \mathbb{R}$ has an equilibrium state.

Proof. Fix $\delta > 0$ an expansive constant for T and let $\mu \in M(T)$. According to Lemma 9.6.1, there exists a finite Borel partition α of X such that $\mathrm{diam}(\alpha) < \delta$ and $\mu(\partial A) = 0$ for each $A \in \alpha$. Let $\varepsilon > 0$. As $\mathrm{h}_\mu(T) \geq \mathrm{h}_\mu(T, \alpha) = \inf_{n\in\mathbb{N}} \frac{1}{n}\mathrm{H}_\mu(\alpha^n)$ by Definitions 9.4.12 and 9.4.10, there exists $m \in \mathbb{N}$ such that

$$\frac{1}{m}\mathrm{H}_\mu(\alpha^m) \leq \mathrm{h}_\mu(T) + \frac{\varepsilon}{2}.$$

Let $(\mu_n)_{n=1}^\infty$ be a sequence of measures in $M(T)$ converging weakly* to μ. Since $\mathrm{diam}(\alpha) < \delta$, it follows from Theorem 9.4.20 that

$$\mathrm{h}_{\mu_n}(T) = \mathrm{h}_{\mu_n}(T, \alpha)$$

for all $n \in \mathbb{N}$. Moreover, by Lemma 12.1.6 (with α replaced by α^m), we have

$$\lim_{n\to\infty} \mathrm{H}_{\mu_n}(\alpha^m) = \mathrm{H}_\mu(\alpha^m).$$

Therefore, there exists $N \in \mathbb{N}$ such that

$$\frac{1}{m}\left|\mathrm{H}_{\mu_n}(\alpha^m) - \mathrm{H}_\mu(\alpha^m)\right| \leq \frac{\varepsilon}{2}$$

for all $n \geq N$. Hence, for all $n \geq N$, we deduce that

$$\mathrm{h}_{\mu_n}(T) = \mathrm{h}_{\mu_n}(T, \alpha) \leq \frac{1}{m}\mathrm{H}_{\mu_n}(\alpha^m) \leq \frac{1}{m}\mathrm{H}_\mu(\alpha^m) + \frac{\varepsilon}{2} \leq \mathrm{h}_\mu(T) + \varepsilon.$$

Consequently, $\limsup_{n\to\infty} \mathrm{h}_{\mu_n}(T) \leq \mathrm{h}_\mu(T)$ for any sequence $(\mu_n)_{n=1}^\infty$ in $M(T)$ converging weakly* to μ. Thus $\limsup_{\nu\to\mu} \mathrm{h}_\nu(T) \leq \mathrm{h}_\mu(T)$, or, in other words, $\mu \mapsto \mathrm{h}_\mu(T)$ is upper semicontinuous.

Since the function $\mu \mapsto \int_X \varphi\, d\mu$ is continuous in the weak* topology on the compact space $M(T)$, it follows that the function $\mu \mapsto \mathrm{h}_\mu(T) + \int_X \varphi\, d\mu$ is upper semicontinuous. Since upper semicontinuous functions attain their upper bound on any compact set, we conclude from the variational principle that each potential φ admits an equilibrium state. □

Recall the class of piecewise monotone continuous maps of the interval. These are not necessarily expansive maps (e. g., the tent map is not expansive). Nonetheless, the function $\mu \mapsto \mathrm{h}_\mu(T)$ is upper semicontinuous for any such map.

Theorem 12.2.6. *If $T : X \to X$ is a piecewise monotone continuous map of the interval, then the function $\mu \mapsto \mathrm{h}_\mu(T)$ is upper semicontinuous. Hence, each potential $\varphi : X \to \mathbb{R}$ has an equilibrium state.*

Proof. The avid reader is referred to [50]. □

12.3 Examples of equilibrium states

Example 12.3.1. By Corollary 12.1.10, any uniquely ergodic system has a unique equilibrium state for every continuous potential. This unique equilibrium state is obviously the unique ergodic invariant measure of the system.

For instance, recall that (cf. Proposition 8.2.43) a translation of the torus $L_\gamma : \mathbb{T}^n \to \mathbb{T}^n$, where $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n) \in \mathbb{T}^n$, is uniquely ergodic if and only if the numbers $1, \gamma_1, \gamma_2, \ldots, \gamma_n$ are linearly independent over $\mathbb{Q}$. Such a translation has a unique equilibrium state.

Let us now look at a symbolic example.

Example 12.3.2. We revisit Example 11.4.1, where E is a finite alphabet and $\sigma : E^\infty \to E^\infty$ is the one-sided full E-shift map. Recall that any function $\widetilde{\varphi} : E \to \mathbb{R}$ generates a continuous potential $\varphi : E^\infty \to \mathbb{R}$ defined by $\varphi(\omega) := \widetilde{\varphi}(\omega_1)$ on E^∞. This potential depends only on the first coordinate ω_1 of the word $\omega \in E^\infty$.

Let $\mathcal{F}$ be the σ-algebra $\mathcal{P}(E)$ of all subsets of E and let P be a probability measure/vector on E, that is, $\sum_{e\in E} P(\{e\}) = 1$. Recall from Examples 8.1.14 and 8.2.32 that the one-sided Bernoulli shift $(\sigma : E^\infty \to E^\infty, \mu_P)$ is an ergodic measure-preserving system.

Let $S = \sum_{e\in E} \exp(\widetilde{\varphi}(e))$. Note that $0 < S < \infty$. We will show that μ_P is an equilibrium state for $\sigma : E^\infty \to E^\infty$ when

$$P(\{e\}) = \frac{1}{S}\exp(\widetilde{\varphi}(e)), \quad \forall e \in E. \tag{12.16}$$

First, let us consider $\mathrm{h}_{\mu_P}(\sigma)$. Let $\alpha := \{[e]\}_{e\in E}$ be the partition of E^∞ into its initial 1-cylinders. It is easy to see that $\alpha^n = \{[\omega]\}_{\omega\in E^n}$, that is, α^n is the partition of E^∞ into its initial n-cylinders. Recall that α is a generator for σ (see Definition 9.4.19 and Example 9.4.23). By Theorem 9.4.20 and Definition 9.4.10, we know that

$$\mathrm{h}_{\mu_P}(\sigma) = \mathrm{h}_{\mu_P}(\sigma, \alpha) = \inf_{n\in\mathbb{N}} \frac{1}{n}\mathrm{H}_{\mu_P}(\alpha^n).$$

By induction on n, it is not difficult to establish that $\mathrm{H}_{\mu_P}(\alpha^n) = n\mathrm{H}_{\mu_P}(\alpha)$. Therefore,

$$\begin{aligned}
\mathrm{h}_{\mu_P}(\sigma) &= \mathrm{H}_{\mu_P}(\alpha) \\
&= -\sum_{e\in E} \mu_P([e]) \log \mu_P([e]) = -\sum_{e\in E} P(\{e\}) \log P(\{e\}) \\
&= -\sum_{e\in E} \frac{1}{S} \exp(\widetilde{\varphi}(e)) \log\left[\frac{1}{S} \exp(\widetilde{\varphi}(e))\right] \\
&= -\frac{1}{S} \sum_{e\in E} \exp(\widetilde{\varphi}(e))[\widetilde{\varphi}(e) - \log S] \\
&= -\frac{1}{S} \sum_{e\in E} \widetilde{\varphi}(e) \exp(\widetilde{\varphi}(e)) + \log S.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\int_{E^\infty} \varphi \, d\mu_P &= \sum_{e\in E} \int_{[e]} \varphi(\omega)\, d\mu_P(\omega) = \sum_{e\in E} \int_{[e]} \widetilde{\varphi}(\omega_1)\, d\mu_P(\omega) \\
&= \sum_{e\in E} \widetilde{\varphi}(e) \mu_P([e]) = \sum_{e\in E} \widetilde{\varphi}(e) P(\{e\}) \\
&= \frac{1}{S} \sum_{e\in E} \widetilde{\varphi}(e) \exp(\widetilde{\varphi}(e)).
\end{aligned}$$

It ensues that

$$\mathrm{h}_{\mu_P}(\sigma) + \int_{E^\infty} \varphi \, d\mu_P = \log S = \mathrm{P}(\sigma, \varphi)$$

where the last equality was derived in Example 11.4.1. Hence, μ_P is an equilibrium state for the potential φ when P satisfies (12.16).

Further examples will be given in Subsection 13.7.3 and in Chapters 17 and 27 onward.

12.4 Exercises

Exercise 12.4.1. Generalize Example 12.2.3 to show that if in addition $\varphi : X \to \mathbb{R}$ is a potential such that

$$\mathrm{P}(T_n, \varphi|_{X_n}) < \mathrm{P}(T_{n+1}, \varphi|_{X_{n+1}})$$

for all $n \in \mathbb{N}$, then φ has no equilibrium state.

Exercise 12.4.2. In the context of Exercise 11.5.8, give an explicit description of the equilibrium states of φ when $k = 1$ and $k = 2$.

Exercise 12.4.3. Let $T : X \to X$ be a topological dynamical system and $\varphi : X \to \mathbb{R}$ a potential. For any $n \in \mathbb{N}$, prove that if μ is an equilibrium state for the couple (T, φ), then μ is an equilibrium state for the couple $(T^n, S_n\varphi)$, too.

Exercise 12.4.4. Let $T : X \to X$ be a topological dynamical system and $\varphi : X \to \mathbb{R}$ a continuous potential. Show that the set of all equilibrium states for φ is a convex subset of $M(T)$. Deduce that if φ has a unique ergodic equilibrium state, then it has unique equilibrium state. Conclude also that if φ has two different equilibrium states, then it has uncountably many (in fact, a continuum of) equilibrium states.

Exercise 12.4.5. Let $T : X \to X$ be a topological dynamical system. Two continuous functions $\varphi, \psi : X \to \mathbb{R}$ are said to be cohomologous modulo a constant (or, equivalently, $\varphi - \psi$ is cohomologous to a constant) in the additive group $C(X)$ if there exist a continuous function $u : X \to \mathbb{R}$ and a constant $c \in \mathbb{R}$ such that

$$\varphi - \psi = u \circ T - u + c.$$

Show that such potentials φ and ψ share the same equilibrium states.

Exercise 12.4.6. Going beyond Example 12.2.3, give an example of a transitive topological dynamical system which does not have any measure of maximal entropy.

Exercise 12.4.7. Using Example 12.2.3, give an example of a topological dynamical system which admits infinitely many equilibrium states under a certain potential.

Exercise 12.4.8. Give an example of a transitive topological dynamical system which has infinitely many measures of maximal entropy.

Exercise 12.4.9. If $T : X \to X$ is a dynamical system such that $\mathrm{h_{top}}(T) < \infty$, then deduce from the variational principle that the pressure function $\mathrm{P}(T, \bullet) : C(X) \to \mathbb{R}$ is convex.

Appendix A – A selection of classical results

This appendix lists classical definitions and results that will be used in this volume. Several of these results are stated without proofs. We sometimes complemented them with classical examples.

A.1 Measure theory

Let us begin by gathering together some of the standard results from measure theory that will be needed in this book. Measure theory is one of the main tools in ergodic theory, so it is important to be familiar with it. Proofs and further explanations of the results can be found in many books on measure theory, for instance, Billingsley [7, 8] and Rudin [58].

A.1.1 Collections of sets and measurable spaces

Given a set X, we shall denote the set of all subsets of X by $\mathcal{P}(X)$. Let us recall the definitions of some important collections of subsets of a set. The most basic collection is called a π-system.

Definition A.1.1. Let X be a set. A nonempty family $\mathcal{P} \subseteq \mathcal{P}(X)$ is a *π-system* on X if $P_1 \cap P_2 \in \mathcal{P}$ for all $P_1, P_2 \in \mathcal{P}$.

In other words, a π-system is a collection that is closed under finite intersections. For example, the family of open intervals $\{(a,\infty) : a \in \mathbb{R}\}$ constitutes a π-system on $\mathbb{R}$. So does the family of closed intervals $\{[a,\infty) : a \in \mathbb{R}\}$. Other examples are the families $\{(-\infty, b) : b \in \mathbb{R}\}$ and $\{(-\infty, b] : b \in \mathbb{R}\}$.

A "slightly" more complex collection is a semialgebra.

Definition A.1.2. Let X be a set. A family $\mathcal{S} \subseteq \mathcal{P}(X)$ is called a *semialgebra* on X if it satisfies the following three conditions:
(a) $\emptyset \in \mathcal{S}$.
(b) $\mathcal{S}$ is a π-system.
(c) If $S \in \mathcal{S}$, then $X \setminus S$ can be written as a finite union of mutually disjoint sets in $\mathcal{S}$. That is, $X \setminus S = \bigcup_{i=1}^{n} S_i$ for some $n \in \mathbb{N}$ and $S_1, S_2, \ldots, S_n \in \mathcal{S}$ with $S_i \cap S_j = \emptyset$ whenever $i \neq j$.

Every semialgebra is a π-system but the converse is not true in general. For instance, none of the π-systems described above is a semialgebra. However, the collection of all intervals forms a semialgebra on $\mathbb{R}$.

An even more intricate collection is an algebra.

https://doi.org/10.1515/9783110702682-013

Definition A.1.3. Let X be a set. A family $\mathcal{A} \subseteq \mathcal{P}(X)$ is said to be an *algebra* on X if it satisfies the following three conditions:
(a) $\emptyset \in \mathcal{A}$.
(b) $\mathcal{A}$ is a π-system.
(c) If $A \in \mathcal{A}$, then $X \setminus A \in \mathcal{A}$.

Every algebra is a semialgebra, although the converse is not true in general. For instance, the semialgebra outlined earlier is not an algebra. Nevertheless, as we will observe in the next lemma, the collection of all subsets of $\mathbb{R}$ that can be expressed as a finite union of intervals is an algebra on $\mathbb{R}$.

The fact that an algebra is stable under finite intersections and complementation implies that an algebra is stable under finitely many set operations (e. g., unions, intersections, differences, symmetric differences, complementation, and combinations thereof).

Note that $\{\emptyset, X\}$ and $\mathcal{P}(X)$ are trivial algebras on X. Since the intersection of any family of algebras is an algebra, the algebra $\mathcal{A}(\mathcal{C})$ generated by any collection $\mathcal{C}$ of subsets of X is well-defined as the smallest, in the sense of set inclusion, of all algebras on X that contain $\mathcal{C}$. If the collection $\mathcal{C}$ is a semialgebra, then it is easy to describe the algebra it generates.

Lemma A.1.4. *Let $\mathcal{S}$ be a semialgebra on X. The algebra $\mathcal{A}(\mathcal{S})$ generated by $\mathcal{S}$ consists of those subsets A of X, which can be written as a finite union of mutually disjoint sets in $\mathcal{S}$, that is, all sets $A \subseteq X$ such that $A = \bigcup_{i=1}^{n} S_i$ for some $S_1, S_2, \ldots, S_n \in \mathcal{S}$ with $S_i \cap S_j = \emptyset$ whenever $i \neq j$.*

Proof. Let

$$\mathcal{A} := \left\{ A \subseteq X \mid \exists S_1, \ldots, S_n \in \mathcal{S},\ S_i \cap S_j = \emptyset, \forall i \neq j \text{ such that } A = \bigcup_{i=1}^{n} S_i \right\}.$$

It is easy to see that $\mathcal{A}$ is an algebra containing $\mathcal{S}$. Therefore, $\mathcal{A} \supseteq \mathcal{A}(\mathcal{S})$. On the other hand, since any algebra is closed under finite unions, any algebra containing $\mathcal{S}$ must contain $\mathcal{A}$. Thus $\mathcal{A}(\mathcal{S}) \supseteq \mathcal{A}$. Hence, $\mathcal{A} = \mathcal{A}(\mathcal{S})$. □

In measure theory, the most important type of collection of subsets of a given set is a σ-algebra.

Definition A.1.5. Let X be a set. A family $\mathcal{B} \subseteq \mathcal{P}(X)$ is called a *σ-algebra* on X if it satisfies the following three conditions:
(a) $\emptyset \in \mathcal{B}$.
(b) $\bigcap_{n=1}^{\infty} B_n \in \mathcal{B}$ for every sequence $(B_n)_{n=1}^{\infty}$ of sets in $\mathcal{B}$.
(c) If $B \in \mathcal{B}$ then $X \setminus B \in \mathcal{B}$.

Note that condition (b) can be replaced by:

(b′) $\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}$ for every sequence $(B_n)_{n=1}^{\infty}$ of sets in $\mathcal{B}$.

A σ-algebra on X is thus a family of subsets of X which is closed under countably many set operations. Clearly, any σ-algebra is an algebra, though the converse is not true in general.

Note that $\{\emptyset, X\}$ and $\mathcal{P}(X)$ are trivial σ-algebras on X. They are respectively called the *indiscrete* and *discrete* σ-algebras. As the intersection of any family of σ-algebras is itself a σ-algebra, the σ-algebra $\sigma(\mathcal{C})$ generated by any collection $\mathcal{C}$ of subsets of X is well-defined as the smallest σ-algebra that contains $\mathcal{C}$. In particular, if the collection $\mathcal{C}$ is finite then the algebra $\mathcal{A}(\mathcal{C})$ it generates is also finite, and thus $\sigma(\mathcal{C}) = \mathcal{A}(\mathcal{C})$ (see Exercise 8.5.1).

A set X equipped with a σ-algebra $\mathcal{B}$ is called a *measurable space* and the elements of $\mathcal{B}$ are accordingly called *measurable sets*.

Example A.1.6. Let X be a topological space and let $\mathcal{T}$ be the topology of X, that is, the collection of all open subsets of X. Then $\sigma(\mathcal{T})$ is a σ-algebra on X called the *Borel σ-algebra* of X. Henceforth, we will denote this latter by $\mathcal{B}(X)$. In particular, $\mathcal{B}(X)$ contains all open sets and closed sets, as well as all countable unions of closed sets and all countable intersections of open sets, that is, all F_σ- and G_δ-sets, respectively. Note that $\mathcal{T}$ is a π-system but not a semialgebra in general.

In the Euclidean space $\mathbb{R}$, the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is generated by the even simpler π-system of open intervals $\{(a,\infty) : a \in \mathbb{R}\}$. Similarly, it is generated by the semialgebra comprising all intervals.

Sometimes the functions considered take values in the extended real numbers $\overline{\mathbb{R}} := [-\infty,\infty]$. A base for the order topology of $\overline{\mathbb{R}}$ is the π-system of all open intervals, namely $\{[-\infty,b) : b \in \overline{\mathbb{R}}\} \cup \{(a,b) : a,b \in \overline{\mathbb{R}}\} \cup \{(a,\infty] : a \in \overline{\mathbb{R}}\} \cup \{[-\infty,\infty]\}$. The Borel σ-algebra $\mathcal{B}(\overline{\mathbb{R}})$ is generated by the even simpler π-system of open intervals $\{(a,\infty] : a \in \mathbb{R}\}$. Note that

$$\mathcal{B}(\overline{\mathbb{R}}) = \big\{B, \{-\infty\} \cup B, B \cup \{\infty\}, \{-\infty\} \cup B \cup \{\infty\} \mid B \in \mathcal{B}(\mathbb{R})\big\}.$$

More examples of algebras and σ-algebras are presented in Exercises 8.5.2–8.5.4.

We now introduce λ-systems, also called Dynkin systems. These collections of sets are closed under complementation and countable disjoint unions.

Definition A.1.7. Let X be a set. A family $\mathcal{L} \subseteq \mathcal{P}(X)$ is called a *λ-system* on X if it satisfies the following three conditions:

(a) $X \in \mathcal{L}$.

(b) $\bigcup_{n=1}^{\infty} L_n \in \mathcal{L}$ for every sequence $(L_n)_{n=1}^{\infty}$ of sets in $\mathcal{L}$ such that $L_n \cap L_m = \emptyset$ for all $m \neq n$.

(c) If $L \in \mathcal{L}$, then $X \setminus L \in \mathcal{L}$.

Note that condition (c) can be replaced by:

(c′) If $K, L \in \mathcal{L}$ and $K \subseteq L$, then $L \setminus K \in \mathcal{L}$.

Every σ-algebra is a λ-system but the converse is not true in general. Nevertheless, it is not difficult to see how these two concepts are related.

Lemma A.1.8. *A collection of sets forms a σ-algebra if and only if it is both a λ-system and a π-system.*

Proof. We have already observed that every σ-algebra is a λ-system and a π-system. So suppose that $\mathcal{B}$ is both a λ-system and a π-system on a set X. Since $\mathcal{B}$ is a π-system that enjoys properties (a) and (c) of a λ-system, it is clear that $\mathcal{B}$ is an algebra. Therefore, it just remains to prove that $\mathcal{B}$ satisfies condition (b') of Definition A.1.5. Let $(B_n)_{n=1}^{\infty}$ be a sequence of sets in $\mathcal{B}$. For every $n \in \mathbb{N}$, let $B_n' = \bigcup_{k=1}^{n} B_k$. As $\mathcal{B}$ is an algebra, $B_n' \in \mathcal{B}$ for all $n \in \mathbb{N}$. The sequence $(B_n')_{n=1}^{\infty}$ is ascending and is such that $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} B_n'$. Thus it suffices to prove condition (b') for ascending sequences in $\mathcal{B}$. Moreover, observe that $\bigcup_{n=1}^{\infty} B_n' = B_1' \cup \bigcup_{n=1}^{\infty} (B_{n+1}' \setminus B_n')$. By condition (c') of a λ-system, we know that $B_{n+1}' \setminus B_n' \in \mathcal{B}$ for each $n \in \mathbb{N}$. Furthermore, the sets B_1' and $B_{n+1}' \setminus B_n'$, $n \in \mathbb{N}$, are mutually disjoint. By condition (b) of a λ-system, it follows that

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} B_n' = B_1' \cup \bigcup_{n=1}^{\infty} (B_{n+1}' \setminus B_n') \in \mathcal{B}.$$

Hence, $\mathcal{B}$ is an algebra satisfying condition (b') of Definition A.1.5. So $\mathcal{B}$ is a σ-algebra. □

The importance and usefulness of λ-systems mostly lie in the following theorem.

Theorem A.1.9 (Dynkin's π-λ theorem). *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system such that $\mathcal{P} \subseteq \mathcal{L}$, then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.*

Proof. See Theorem 3.2 in Billingsley [7]. □

Furthermore, as the intersection of any family of λ-systems is a λ-system, the λ-system $\mathcal{L}(\mathcal{C})$ generated by any collection $\mathcal{C} \subseteq \mathcal{P}(X)$ is well-defined as the intersection of all λ-systems that comprise $\mathcal{C}$. When $\mathcal{C}$ is a π-system, the λ-system and the σ-algebra that are generated by $\mathcal{C}$ are one and the same.

Corollary A.1.10. *If $\mathcal{P}$ is a π-system, then $\sigma(\mathcal{P}) = \mathcal{L}(\mathcal{P})$.*

Proof. This immediately follows from Lemma A.1.8 and Theorem A.1.9. □

Finally, let us recall yet another type of collection of sets named, for obvious reasons, a monotone class.

Definition A.1.11. Let X be a set. A family $\mathcal{M} \subseteq \mathcal{P}(X)$ is called a *monotone class* on X if it is stable under countable monotone unions and countable monotone intersections.

In other words,

$$\text{if } (M_n)_{n=1}^{\infty} \subseteq \mathcal{M} \text{ is such that } M_1 \subseteq M_2 \subseteq M_3 \subseteq \cdots, \quad \text{then } \bigcup_{n=1}^{\infty} M_n \in \mathcal{M}$$

and

$$\text{if } (M_n)_{n=1}^{\infty} \subseteq \mathcal{M} \text{ is such that } M_1 \supseteq M_2 \supseteq M_3 \supseteq \cdots, \quad \text{then } \bigcap_{n=1}^{\infty} M_n \in \mathcal{M}.$$

Every σ-algebra is a monotone class but the converse is not true in general. Nevertheless, a monotone class which is an algebra is a σ-algebra. There is an analogue of Dynkin's theorem for monotone classes.

Theorem A.1.12 (Halmos' monotone class theorem). *If $\mathcal{A}$ is an algebra and $\mathcal{M}$ is a monotone class such that $\mathcal{A} \subseteq \mathcal{M}$, then $\sigma(\mathcal{A}) \subseteq \mathcal{M}$.*

Proof. See Theorem 3.4 in Billingsley [7]. □

Because the intersection of any family of monotone classes is a monotone class, the monotone class $\mathcal{M}(\mathcal{C})$ generated by any collection $\mathcal{C} \subseteq \mathcal{P}(X)$ is well-defined as the intersection of all monotone classes that comprise $\mathcal{C}$. When $\mathcal{C}$ is a semialgebra, the σ-algebra and the monotone class generated by $\mathcal{C}$ coincide.

Theorem A.1.13. *If $\mathcal{S}$ is a semialgebra, then $\sigma(\mathcal{S}) = \sigma(\mathcal{A}(\mathcal{S})) = \mathcal{M}(\mathcal{A}(\mathcal{S})) = \mathcal{M}(\mathcal{S})$.*

Proof. The equalities $\sigma(\mathcal{S}) = \sigma(\mathcal{A}(\mathcal{S}))$ and $\mathcal{M}(\mathcal{A}(\mathcal{S})) = \mathcal{M}(\mathcal{S})$ are obvious. Since a σ-algebra is a monotone class and $\sigma(\mathcal{A}(\mathcal{S})) \supseteq \mathcal{A}(\mathcal{S})$, it is evident that $\sigma(\mathcal{A}(\mathcal{S})) \supseteq \mathcal{M}(\mathcal{A}(\mathcal{S}))$. The opposite inclusion is the object of Halmos' monotone class theorem. □

A.1.2 Measurable transformations

We now look at maps between measurable spaces. Recall that a measurable space is a set X equipped with a σ-algebra $\mathcal{A}$. The elements of $\mathcal{A}$ are the measurable sets in that space.

Definition A.1.14. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. A transformation $T : X \to Y$ is said to be *measurable* provided that $T^{-1}(B) \in \mathcal{A}$ for every $B \in \mathcal{B}$.

We have earlier mentioned that σ-algebras are the most important collections of sets in measure theory. However, it is generally impossible to describe, in a simple form, the sets in a σ-algebra. Luckily, σ-algebras that are generated by smaller and simpler structures like π-systems, semialgebras, or algebras, are much easier to cope with. In this situation, proving that some interesting property is satisfied for the sets in these smaller and simpler structures is often sufficient to guarantee that that property holds for all sets in the σ-algebra. This is the case for the measurability of transformations.

Theorem A.1.15. *Let $T : (X, \mathcal{A}) \to (Y, \mathcal{B})$ be a transformation. If $\mathcal{B} = \sigma(\mathcal{C})$ is a σ-algebra generated by a collection $\mathcal{C} \subseteq \mathcal{P}(Y)$, then T is measurable if and only if $T^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{C}$.*

Proof. It is clear that if T is measurable, then $T^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{C}$ since $\mathcal{C} \subseteq \sigma(\mathcal{C}) = \mathcal{B}$. Conversely, suppose that $T^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{C}$. Consider the collection of sets $\mathcal{B}' = \{B \subseteq Y : T^{-1}(B) \in \mathcal{A}\}$. By assumption, $\mathcal{B}' \supseteq \mathcal{C}$. It is also easy to see that $\mathcal{B}'$ is a σ-algebra. Thus $\mathcal{B}' \supseteq \sigma(\mathcal{C}) = \mathcal{B}$, and hence T is measurable. □

If the range Y of a transformation is a Borel subset of $\overline{\mathbb{R}}$, then unless otherwise stated Y will be assumed to be endowed with its Borel σ-algebra, which is just the projection of $\mathcal{B}(\overline{\mathbb{R}})$ onto Y (see Exercise 8.5.5). In this context, we will use the term function instead of transformation.

Example A.1.16. Let $(X, \mathcal{A})$ be a measurable space.

(a) Let $A \subseteq X$. The *indicator function* $\mathbb{1}_A : X \to \{0, 1\}$ (also called *characteristic function*) defined by

$$\mathbb{1}_A(x) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

is measurable if and only if A is measurable, that is, $A \in \mathcal{A}$.

(b) A function $s : X \to \mathbb{R}$ which takes only finitely many values is called a *simple function*. Such a function can be expressed in the form

$$s = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i},$$

where $A_i = \{x \in X : s(x) = \alpha_i\}$ and the α_i's are the values of the function s. Such a function is measurable if and only if each set A_i is measurable.

The following theorem shows the utility of simple functions. It states that any nonnegative measurable function is the pointwise limit of a nondecreasing sequence of nonnegative measurable simple functions.

Theorem A.1.17. *Let $(X, \mathcal{A})$ be a measurable space and $f : X \to [0, \infty]$ be a measurable function. Then there exists a sequence $(s_n)_{n=1}^{\infty}$ of measurable simple functions on X such that*

(a) $0 \le s_1 \le s_2 \le \cdots \le f$.

(b) $\lim_{n\to\infty} s_n(x) = f(x)$, $\forall x \in X$.

Proof. See Theorem 1.17 in Rudin [58]. □

A.1.3 Measure spaces

The concept of measure is obviously central to measure theory.

Definition A.1.18. Let $(X, \mathcal{A})$ be a measurable space. A set function $\mu : \mathcal{A} \to [0, \infty]$ is said to be a *measure* on X provided that

(a) $\mu(\emptyset) = 0$.

(b) μ is *countably additive*, that is, for each sequence $(A_n)_{n=1}^{\infty}$ of pairwise disjoint sets belonging to $\mathcal{A}$, the function μ is such that

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

The triple $(X, \mathcal{A}, \mu)$ is called a *measure space*.

If $\mu(X) < \infty$, then μ is said to be a *finite measure*. If $\mu(X) = 1$, then μ is a *probability measure*. Finally, μ is said to be *σ-finite* if there exists a sequence $(A_n)_{n=1}^{\infty}$ of sets in $\mathcal{A}$ such that $\mu(A_n) < \infty$ for all $n \in \mathbb{N}$ and $\bigcup_{n=1}^{\infty} A_n = X$.

Here are a few basic properties of measures.

Lemma A.1.19. *Let $(X, \mathcal{A}, \mu)$ be a measure space and $A, B \in \mathcal{A}$.*

(a) *If $A \subseteq B$, then $\mu(A) \leq \mu(B)$.*

(b) *If $A \subseteq B$ and $\mu(B \setminus A) < \infty$, then $\mu(A) = \mu(B) - \mu(B \setminus A)$.*

(c) $\mu(A \cup B) \leq \mu(A) + \mu(B)$.

(d) *If $(A_n)_{n=1}^{\infty}$ is an ascending sequence in $\mathcal{A}$ (i. e., $A_n \subseteq A_{n+1}, \forall n \in \mathbb{N}$), then*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n) = \sup_{n\in\mathbb{N}} \mu(A_n).$$

(e) *If $(B_n)_{n=1}^{\infty}$ is a descending sequence in $\mathcal{A}$ (i. e., $A_n \supseteq A_{n+1}, \forall n \in \mathbb{N}$) and if $\mu(B_1) < \infty$, then*

$$\mu\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mu(B_n) = \inf_{n\in\mathbb{N}} \mu(B_n).$$

(f) *If $(C_n)_{n=1}^{\infty}$ is any sequence in $\mathcal{A}$, then*

$$\mu\left(\bigcup_{n=1}^{\infty} C_n\right) \leq \sum_{n=1}^{\infty} \mu(C_n).$$

(g) *If $(D_n)_{n=1}^{\infty}$ is a sequence in $\mathcal{A}$ such that $\mu(D_m \cap D_n) = 0$ for all $m \neq n$, then*

$$\mu\left(\bigcup_{n=1}^{\infty} D_n\right) = \sum_{n=1}^{\infty} \mu(D_n).$$

Proof. The proof is left to the reader as an exercise. □

A more intricate property of measures deserves a special name.

Lemma A.1.20 (Borel–Cantelli lemma). *Let $(X, \mathcal{A}, \mu)$ be a measure space and $(A_n)_{n=1}^{\infty}$ a sequence in $\mathcal{A}$. If $\sum_{n=1}^{\infty} \mu(A_n) < \infty$, then $\mu(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n) = 0$.*

Proof. The proof is left to the reader as an exercise. □

We now provide two simple examples of measures. The first of these may seem insignificant at first glance but turns out to be very useful in practice.

Example A.1.21. Let X be a nonempty set and $\mathcal{P}(X)$ be the discrete σ-algebra on X.

(a) Choose a point $x \in X$. Define the set function $\delta_x : \mathcal{P}(X) \to \{0, 1\}$ by setting

$$\delta_x(A) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

One readily verifies that δ_x is a probability measure. It is referred to as the *Dirac point mass* or *Dirac measure* concentrated at the point x.

(b) For any $A \subseteq X$ define $m(A)$ to be the number of elements in the set A if A is finite and set $m(A) = \infty$ if the set A is infinite. Then $m : \mathcal{P}(X) \to [0, \infty]$ is called the *counting measure* on X.

There is a notion of completeness for measure spaces.

Definition A.1.22. A measure space $(X, \mathcal{A}, \mu)$ is said to be *complete* if every subset of a set of measure zero is measurable. That is, if $A \subseteq X$ and there is $B \in \mathcal{A}$ such that $A \subseteq B$ and $\mu(B) = 0$, then $A \in \mathcal{A}$.

Note that any measure space can be extended to a complete one (see Exercises 8.5.7–8.5.8).

In Example A.1.6, we introduced the concept of Borel σ-algebra. We now introduce Borel measures and describe different forms of regularity for these measures.

Definition A.1.23. Let X be a topological space and $\mathcal{B}(X)$ be the Borel σ-algebra on X.

(a) A *Borel measure* μ on X is a measure defined on the Borel σ-algebra $\mathcal{B}(X)$ of X. The resulting measure space $(X, \mathcal{B}(X), \mu)$ is called a Borel measure space. In particular, if μ is a probability measure then $(X, \mathcal{B}(X), \mu)$ is called a Borel probability space.

(b) A Borel measure μ is said to be *inner regular* if

$$\mu(B) = \sup\{\mu(K) : K \subseteq B,\ K \text{ compact}\}, \quad \forall B \in \mathcal{B}(X).$$

(c) A Borel measure μ is said to be *outer regular* if

$$\mu(B) = \inf\{\mu(G) : B \subseteq G,\ G \text{ open}\}, \quad \forall B \in \mathcal{B}(X).$$

(d) A Borel measure is called *regular* if it is both inner and outer regular.

Theorem A.1.24. *Every Borel probability measure on a separable, completely metrizable space is regular.*

Proof. See Theorem 17.11 in Kechris [35]. □

Example A.1.25. There exists a complete, regular measure λ defined on a σ-algebra $\mathcal{L}$ on $\mathbb{R}^k$ with the following properties:
(a) $\lambda(R) = \mathrm{Vol}(R)$ for every k-rectangle $R \subseteq \mathbb{R}^k$, where Vol denotes the usual k-dimensional volume in $\mathbb{R}^k$.
(b) $\mathcal{L}$ is the completion of the Borel σ-algebra $\mathcal{B}(\mathbb{R}^k)$; more precisely, $E \in \mathcal{L}$ if and only if there exist an F_σ-set F and a G_δ-set G such that $F \subseteq E \subseteq G$ and $\lambda(G \setminus F) = 0$.
(c) λ is *translation invariant*, that is, $\lambda(E + x) = \lambda(E)$ for all $E \in \mathcal{L}$ and all $x \in \mathbb{R}^k$.

Moreover, up to a multiplicative constant, λ is the only translation-invariant Borel measure on $\mathbb{R}^k$ that gives finite measure to all compact sets.

The measure λ is called the *Lebesgue measure on* $\mathbb{R}^k$ and, accordingly, the sets in $\mathcal{L}$ are said to be *Lebesgue measurable*. Unlike in Example A.1.21, not all subsets of $\mathbb{R}^k$ are Lebesgue measurable. Indeed, Vitali (see 5.7, The Vitali Monsters on p. 120 of [29]) showed that it is impossible to construct a measure having properties (a)–(c) on the set of all subsets of $\mathbb{R}^k$.

The following lemma is one more eloquent manifestation of the relevance of structures simpler than σ-algebras.

Lemma A.1.26. *Let X be a set and $\mathcal{P}$ be a π-system on X. Let μ and ν be probability measures on $(X, \sigma(\mathcal{P}))$. Then*

$$\mu = \nu \iff \mu(P) = \nu(P), \quad \forall P \in \mathcal{P}.$$

Proof. The implication $\Rightarrow$ is trivial. For the opposite one $\Leftarrow$, suppose that $\mu(P) = \nu(P)$ for all $P \in \mathcal{P}$. Consider the collection of sets $\mathcal{A} := \{A \subseteq X : \mu(A) = \nu(A)\}$. By assumption, $\mathcal{P} \subseteq \mathcal{A}$. It is also easy to see that $\mathcal{A}$ is a λ-system. Per Corollary A.1.10, it follows that $\sigma(\mathcal{P}) = \mathcal{L}(\mathcal{P}) \subseteq \mathcal{A}$. □

In summary, two probability measures that agree on a π-system are equal on the σ-algebra generated by that π-system. However, this result does not generally hold for infinite measures (see Exercise 8.5.9).

A.1.4 Extension of set functions to measures

The main shortcoming of the preceding lemma lies in the assumption that the set functions μ and ν are measures defined on a σ-algebra. In particular, this means that they are countably additive on that entire σ-algebra. However, we frequently define a set function $\mu : \mathcal{C} \to [0, \infty]$ on a collection $\mathcal{C}$ of subsets of a set X on which the values of μ are naturally determined but it may be unclear whether μ may be extended to a measure on $\sigma(\mathcal{C})$. The forthcoming results are extremely useful in that regard. The first one concerns the extension of a finitely/countably additive set function from a semialgebra to an algebra.

Theorem A.1.27. *Let $\mathcal{S}$ be a semialgebra on a set X and let $\mu : \mathcal{S} \to [0, \infty]$ be a finitely additive set function, that is, a function such that*

$$\mu\left(\bigcup_{i=1}^{n} S_i\right) = \sum_{i=1}^{n} \mu(S_i)$$

for every finite family $(S_i)_{i=1}^{n}$ of mutually disjoint sets in $\mathcal{S}$ such that $\bigcup_{i=1}^{n} S_i \in \mathcal{S}$. Then there exists a unique finitely additive set function $\overline{\mu} : \mathcal{A}(\mathcal{S}) \to [0, \infty]$ which is an extension of μ to $\mathcal{A}(\mathcal{S})$, the algebra generated by $\mathcal{S}$. Moreover, the extension $\overline{\mu}$ is countably additive whenever the original set function μ is.

Proof. This directly follows from Lemma A.1.4. For more detail, see Theorems 3.4 and 3.5 in Kingman and Taylor [38]. □

The second result concerns the extension of a countably additive set function from an algebra to a σ-algebra.

Theorem A.1.28 (Carathéodory's extension theorem). *Let $\mathcal{A}$ be an algebra on a set X and let $\mu : \mathcal{A} \to [0, 1]$ be a countably additive set function, that is, a function such that*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

for every sequence $(A_i)_{i=1}^{\infty}$ of mutually disjoint sets in $\mathcal{A}$ such that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. Suppose also that $\mu(X) = 1$. Then there exists a unique probability measure $\overline{\mu} : \sigma(\mathcal{A}) \to [0, 1]$, which is an extension of $\mu : \mathcal{A} \to [0, 1]$.

Proof. See Theorem 3.1 in Billingsley [7] or Theorem 4.2 in Kingman and Taylor [38]. □

Carathéodory's extension theorem thus reduces the problem to demonstrating that a set function is countably additive on an algebra. This can be hard to prove, so we sometimes rely upon the following result.

Lemma A.1.29. *Let $\mathcal{A}$ be an algebra on a set X and $\mu : \mathcal{A} \to [0, \infty)$ be a finitely additive function. Then μ is countably additive if and only if*

$$\lim_{i\to\infty} \mu(A_i) = 0 \tag{A.1}$$

for every descending sequence $(A_i)_{i=1}^{\infty}$ of sets in $\mathcal{A}$ such that $\bigcap_{i=1}^{\infty} A_i = \emptyset$.

Furthermore, if (A.1) *holds, then μ has a unique extension to a σ-additive function (a measure) from* $\sigma(\mathcal{A})$*, the σ-algebra generated by* $\mathcal{A}$*, to* $[0, \infty)$.

Proof. Recall that measures enjoy property (d) of Lemma A.1.19. However, the set function μ considered here is not a measure since it is defined on an algebra rather than on a σ-algebra. Nevertheless, we will show that μ satisfies property (d) on the algebra $\mathcal{A}$ because of hypothesis (A.1). Let $(A_i)_{i=1}^{\infty}$ be an ascending sequence of sets in $\mathcal{A}$ such that $A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. Since $\mathcal{A}$ is an algebra, we know that $A \setminus A_i \in \mathcal{A}$ for all $i \in \mathbb{N}$. Therefore, the sequence $(A \setminus A_i)_{i=1}^{\infty}$ is a descending sequence of sets in $\mathcal{A}$ such that

$$\bigcap_{i=1}^{\infty}(A \setminus A_i) = A \setminus \left(\bigcup_{i=1}^{\infty} A_i\right) = \emptyset.$$

By hypothesis (A.1), we infer that

$$\lim_{i\to\infty} \mu(A \setminus A_i) = 0.$$

Moreover, since $A = (A \setminus A_i) \cup A_i$ and μ is finitely additive on $\mathcal{A}$ and finite, we deduce that

$$\mu(A \setminus A_i) = \mu(A) - \mu(A_i), \quad \forall i \in \mathbb{N}.$$

Hence,

$$\lim_{i\to\infty} \big(\mu(A) - \mu(A_i)\big) = 0.$$

Thus

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mu(A_i). \tag{A.2}$$

Now, let $(B_i)_{i=1}^{\infty}$ be any sequence of mutually disjoint sets in $\mathcal{A}$ such that $\bigcup_{i=1}^{\infty} B_i \in \mathcal{A}$. For every $i \in \mathbb{N}$, let $B_i' = \bigcup_{j=1}^{i} B_j$. As $\mathcal{A}$ is an algebra, $B_i' \in \mathcal{A}$ for all $i \in \mathbb{N}$. The sequence $(B_i')_{i=1}^{\infty}$ is ascending and is such that $\bigcup_{i=1}^{\infty} B_i' = \bigcup_{i=1}^{\infty} B_i \in \mathcal{A}$. By (A.2), we know that

$$\mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \mu\left(\bigcup_{i=1}^{\infty} B_i'\right) = \lim_{i\to\infty} \mu(B_i').$$

Using this, the fact that the B_i's are mutually disjoint and that μ is finitely additive, we conclude that

$$\mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \lim_{i\to\infty} \mu(B_i') = \lim_{i\to\infty} \mu\left(\bigcup_{j=1}^{i} B_j\right) = \lim_{i\to\infty} \sum_{j=1}^{i} \mu(B_j) = \sum_{i=1}^{\infty} \mu(B_i).$$

That is, μ is countably additive on $\mathcal{A}$.

The second part of the statement directly follows from the first one and Theorem A.1.28. □

As an immediate consequence of Lemma A.1.29 we get the following.

Lemma A.1.30. *Let $\mathcal{A}$ be an algebra on a set X, let $\nu : \sigma(\mathcal{A}) \to [0, \infty)$ be a finite measure, and let $\mu : \mathcal{A} \to [0, \infty)$ be a finitely additive function which is absolutely continuous with respect to ν, meaning that $\mu(A) = 0$ whenever $A \in \mathcal{A}$ and $\nu(A) = 0$. Then μ is countably additive and has a unique extension to a σ-additive function (a measure) from $\sigma(\mathcal{A})$ to $[0, \infty)$.*

Finally, as a straightforward consequence of Lemma A.1.29 we have the following.

Lemma A.1.31. *Let $\mathcal{A}$ be a σ-algebra on a set X, let $\nu : \mathcal{A} \to [0, \infty)$ be a finite measure and let $\mu : \mathcal{A} \to [0, \infty)$ be a finitely additive function which is absolutely continuous with respect to ν on some algebra generating $\mathcal{A}$, meaning that $\mu(A) = 0$ whenever A belongs to this algebra and $\nu(A) = 0$. Then μ is a measure.*

Another feature of an algebra is that any element of a σ-algebra generated by an algebra can be approximated as closely as desired by an element of the algebra. Before stating the precise result, recall that the *symmetric difference* of two sets A and B is denoted by $A \triangle B$ and is the set of all points that belong to exactly one of those two sets. That is,

$$A \triangle B := (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B).$$

Properties of the symmetric difference are examined in Exercises 8.5.10–8.5.11.

Lemma A.1.32. *Let $\mathcal{A}$ be an algebra on a set X and μ be a probability measure on $(X, \sigma(\mathcal{A}))$. Then for every $\varepsilon > 0$ and $B \in \sigma(\mathcal{A})$ there exists some $A \in \mathcal{A}$ such that $\mu(A \triangle B) < \varepsilon$.*

Proof. See Theorem 4.4 in Kingman and Taylor [38]. □

A.1.5 Integration

Let us now briefly recollect some facts about integration. First, the definition of the integral of a measurable function with respect to a measure.

Definition A.1.33. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $A \in \mathcal{A}$.

(a) If $s : X \to [0, \infty)$ is a measurable simple function of the form,

$$s = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i},$$

then the integral of the function s over the set A with respect to the measure μ is defined as

$$\int_A s \, d\mu := \sum_{i=1}^{n} \alpha_i \mu(A_i \cap A).$$

We use the convention that $0 \cdot \infty = 0$ in case it happens that $\alpha_i = 0$ and $\mu(A_i \cap A) = \infty$ for some $1 \le i \le n$.

(b) If $f : X \to [0, \infty]$ is a measurable function, then the integral of the function f over the set A with respect to the measure μ is defined as

$$\int_A f \, d\mu := \sup \int_A s \, d\mu,$$

where the supremum is taken over all measurable simple functions $0 \le s \le f$. Note that if f is simple, then definitions (a) and (b) coincide.

(c) If $f : X \to \overline{\mathbb{R}}$ is a measurable function, then the integral of the function f over the set A with respect to the measure μ is defined as

$$\int_A f \, d\mu := \int_A f_+ \, d\mu - \int_A f_- \, d\mu,$$

as long as $\min\{\int_A f_+ \, d\mu, \int_A f_- \, d\mu\} < \infty$, where f_+ and f_- respectively denote the positive and negative parts of f. That is, $f_+(x) := \max\{f(x), 0\}$ whereas $f_-(x) := \max\{-f(x), 0\}$.

(d) A measurable function $f : X \to \overline{\mathbb{R}}$ is said to be *integrable* if $\int_X |f| \, d\mu < \infty$. We denote this by $f \in L^1(X, \mathcal{A}, \mu)$. If there can be no confusion, we simply write $f \in L^1(\mu)$.

(e) A property is said to hold *μ-almost everywhere* (sometimes abbreviated μ-a. e.) if the property holds on the entire space except possibly on a set of μ-measure zero.

The following properties follow from this definition.

Lemma A.1.34. *Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f, g \in L^1(X, \mathcal{A}, \mu)$, $A, B \in \mathcal{A}$, and $a, b \in \mathbb{R}$.*

(a) *If $f \le g$ μ-a. e., then $\int_A f \, d\mu \le \int_A g \, d\mu$. Also, if $f < g$ μ-a. e., then $\int_A f \, d\mu < \int_A g \, d\mu$.*

(b) *If $A \subseteq B$ and $0 \le f$ μ-a. e., then $0 \le \int_A f \, d\mu \le \int_B f \, d\mu$.*

(c)

$$\left| \int_A f \, d\mu \right| \le \int_A |f| \, d\mu.$$

(d) *Linearity:*

$$\int_A (af + bg) \, d\mu = a \int_A f \, d\mu + b \int_A g \, d\mu.$$

(e) *If $(A_n)_{n=1}^{\infty}$ is a sequence of mutually disjoint measurable sets, then*

$$\int_{\bigcup_{n=1}^{\infty} A_n} f \, d\mu = \sum_{n=1}^{\infty} \int_{A_n} f \, d\mu.$$

(f) *$f = g$ μ-a. e. $\Longleftrightarrow \int_A f\,d\mu = \int_A g\,d\mu$, $\forall A \in \mathcal{A}$.*
(g) *The relation $f = g$ μ-a. e. is an equivalence relation on the set $L^1(X, \mathcal{A}, \mu)$. The equivalence classes generated by this relation form a Banach space also denoted by $L^1(X, \mathcal{A}, \mu)$ (or $L^1(\mu)$, for short) with norm*

$$\|f\|_1 := \int_X |f|\,d\mu < \infty.$$

A sequence $(f_n)_{n=1}^{\infty}$ in $L^1(\mu)$ is said to converge to f in $L^1(\mu)$ if $\lim_{n\to\infty} \|f_n - f\|_1 = 0$.

A.1.6 Convergence theorems

In measure theory, there are fundamental theorems that are especially helpful for finding the integral of functions that are the pointwise limits of sequences of functions.

The first of these results applies to monotone sequences of functions. A sequence of functions $(f_n)_{n=1}^{\infty}$ is monotone if it is increasing pointwise ($f_{n+1}(x) \geq f_n(x)$ for all $x \in X$ and all $n \in \mathbb{N}$) or decreasing pointwise ($f_{n+1}(x) \leq f_n(x)$ for all $x \in X$ and all $n \in \mathbb{N}$). We state the theorem for increasing sequences, but its counterpart for decreasing sequences can be easily deduced from it.

Theorem A.1.35 (Monotone convergence theorem). *Let $(X, \mathcal{A}, \mu)$ be a measure space. If $(f_n)_{n=1}^{\infty}$ is an increasing sequence of nonnegative measurable functions, then the integral of their pointwise limit is equal to the limit of their integrals, that is,*

$$\int_X \lim_{n\to\infty} f_n\,d\mu = \lim_{n\to\infty} \int_X f_n\,d\mu.$$

Proof. See Theorem 1.26 in Rudin [58]. □

Note that this theorem holds for almost everywhere increasing sequences of almost everywhere nonnegative measurable functions with an almost everywhere pointwise limit.

For general sequences of nonnegative functions, we have the following immediate consequence.

Lemma A.1.36 (Fatou's lemma). *Let $(X, \mathcal{A}, \mu)$ be a measure space. For any sequence $(f_n)_{n=1}^{\infty}$ of nonnegative measurable functions,*

$$\int_X \liminf_{n\to\infty} f_n\,d\mu \leq \liminf_{n\to\infty} \int_X f_n\,d\mu.$$

Proof. For every $x \in X$ and $n \in \mathbb{N}$, define $g_n(x) = \inf\{f_i(x) : 1 \leq i \leq n\}$ and apply the monotone convergence theorem to the sequence $(g_n)_{n=1}^{\infty}$. For more detail, see Theorem 1.28 in Rudin [58]. □

The following lemma is another application of the monotone convergence theorem. It offers another way of integrating a nonnegative function.

Lemma A.1.37. *Let $(X, \mathcal{A}, \mu)$ be a measure space. Let f be a nonnegative measurable function and $A \in \mathcal{A}$. Then*

$$\int_A f\,d\mu = \int_0^\infty \mu(\{x \in A : f(x) > r\})\,dr.$$

Proof. Suppose that $f = \mathbb{1}_B$ for some $B \in \mathcal{A}$. Then

$$\begin{aligned}\int_0^\infty \mu(\{x \in A : f(x) > r\})\,dr &= \int_0^1 \mu(\{x \in A : \mathbb{1}_B(x) > r\})\,dr \\ &= \int_0^1 \mu(A \cap B)\,dr \\ &= \mu(A \cap B) \\ &= \int_X \mathbb{1}_{A\cap B}\,d\mu = \int_X \mathbb{1}_A \cdot \mathbb{1}_B\,d\mu = \int_A f\,d\mu.\end{aligned}$$

So the equality holds for characteristic functions. We leave it to the reader to show that the equality prevails for all nonnegative measurable simple functions. If f is a general nonnegative measurable function, then by Theorem A.1.17 there exists an increasing sequence $(s_n)_{n=1}^\infty$ of nonnegative measurable simple functions such that $\lim_{n\to\infty} s_n(x) = f(x)$ for every $x \in X$. For every $r \geq 0$, let

$$\tilde{f}(r) = \mu(\{x \in A : f(x) > r\}).$$

This function is obviously nonnegative and decreasing. Hence, by Theorem A.1.15 it is Borel measurable since $\tilde{f}^{-1}((t,\infty])$ is an interval for all $t \in [0,\infty]$ and the sets $\{(t,\infty]\}_{t\in[0,\infty]}$ generate the Borel σ-algebra of $[0,\infty]$. Fix momentarily $r \geq 0$. Since $s_n \nearrow f$, the sets $(\{x \in A : s_n(x) > r\})_{n=1}^\infty$ form an ascending sequence such that

$$\bigcup_{n=1}^\infty \{x \in A : s_n(x) > r\} = \{x \in A : f(x) > r\}.$$

Then by Lemma A.1.19(d),

$$\tilde{f}(r) = \mu\left(\bigcup_{n=1}^\infty \{x \in A : s_n(x) > r\}\right) = \lim_{n\to\infty} \mu(\{x \in A : s_n(x) > r\}) = \lim_{n\to\infty} \tilde{s}_n(r).$$

Observe that $(\tilde{s}_n)_{n=1}^\infty$ is an increasing sequence of nonnegative decreasing functions. So $(\tilde{s}_n)_{n=1}^\infty$ is a sequence of nonnegative Borel measurable functions, which increases pointwise to $\tilde{f}$. It follows from the monotone convergence theorem (Theorem A.1.35) that

$$\begin{aligned}\int_0^\infty \mu(\{x \in A : f(x) > r\})\,dr &= \int_0^\infty \tilde{f}(r)\,dr = \int_0^\infty \lim_{n\to\infty} \tilde{s}_n(r)\,dr \\ &= \lim_{n\to\infty} \int_0^\infty \tilde{s}_n(r)\,dr \\ &= \lim_{n\to\infty} \int_0^\infty \mu(\{x \in A : s_n(x) > r\})\,dr \\ &= \lim_{n\to\infty} \int_A s_n\,d\mu = \int_A \lim_{n\to\infty} s_n\,d\mu \\ &= \int_A f\,d\mu. \end{aligned}$$ □

Pointwise convergence of a sequence of integrable functions does not guarantee convergence in L^1 (see Exercise 8.5.14). However, under one relatively weak additional assumption, this becomes true. The second fundamental theorem of convergence applies to sequences of functions which have an almost everywhere pointwise limit and are dominated (i. e., uniformly bounded) almost everywhere by an integrable function.

Theorem A.1.38 (Lebesgue's dominated convergence theorem). *If a sequence of measurable functions $(f_n)_{n=1}^\infty$ on a measure space $(X, \mathcal{A}, \mu)$ converges pointwise μ-a. e. to a function f and if there exists $g \in L^1(\mu)$ such that $|f_n(x)| \le g(x)$ for all $n \in \mathbb{N}$ and μ-a. e. $x \in X$, then $f \in L^1(\mu)$ and*

$$\lim_{n\to\infty} \|f_n - f\|_1 = 0 \quad \text{and} \quad \lim_{n\to\infty} \int_X f_n\,d\mu = \int_X f\,d\mu.$$

Proof. Apply Fatou's lemma to $2g - |f_n - f| \ge 0$. See Theorem 1.34 in Rudin [58] for more detail. □

Note that

$$\lim_{n\to\infty} \|f_n - f\|_1 = 0 \implies \lim_{n\to\infty} \|f_n\|_1 = \|f\|_1$$

since $\big|\|f_n\|_1 - \|f\|_1\big| \le \|f_n - f\|_1$ and

$$\lim_{n\to\infty} \|f_n - f\|_1 = 0 \implies \lim_{n\to\infty} \int_X f_n\,d\mu = \int_X f\,d\mu$$

by applying Lemma A.1.34(c) to $f_n - f$. The opposite implications do not hold in general. Nevertheless, the following lemma states that any sequence of integrable functions

$(f_n)_{n=1}^\infty$ that converges pointwise almost everywhere to an integrable function f will also converge to that function in L^1 if and only if their L^1 norms converge to the L^1 norm of f.

Lemma A.1.39 (Scheffé's lemma). *Let $(X,\mathcal{A},\mu)$ be a measure space. If a sequence $(f_n)_{n=1}^\infty$ of functions in $L^1(\mu)$ converges pointwise μ-a. e. to a function $f \in L^1(\mu)$, then*

$$\lim_{n\to\infty} \|f_n - f\|_1 = 0 \iff \lim_{n\to\infty} \|f_n\|_1 = \|f\|_1.$$

In particular, if $f_n \geq 0$ μ-a. e. for all $n \in \mathbb{N}$, then $f \geq 0$ μ-a. e. and

$$\lim_{n\to\infty} \|f_n - f\|_1 = 0 \iff \lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu.$$

Proof. The direct implication $\Rightarrow$ is trivial. For the converse implication, assume that $\lim_{n\to\infty} \|f_n\|_1 = \|f\|_1$. Suppose first that $f_n \geq 0$ for all $n \in \mathbb{N}$. Then $f \geq 0$ and hence our assumption reduces to $\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu$. Let $\ell_n = \min\{f, f_n\}$ and $u_n = \max\{f, f_n\}$. Then both $(\ell_n)_{n=1}^\infty$ and $(u_n)_{n=1}^\infty$ converge pointwise μ-a. e. to f. Also, $|\ell_n| = \ell_n \leq f$ for all n, so Lebesgue's dominated convergence theorem asserts that

$$\lim_{n\to\infty} \int_X \ell_n \, d\mu = \int_X f \, d\mu.$$

Observing that $u_n = f + f_n - \ell_n$, we also get that

$$\lim_{n\to\infty} \int_X u_n \, d\mu = \int_X f \, d\mu + \lim_{n\to\infty} \int_X f_n \, d\mu - \lim_{n\to\infty} \int_X \ell_n \, d\mu = \int_X f \, d\mu.$$

Thus

$$\lim_{n\to\infty} \|f_n - f\|_1 = \lim_{n\to\infty} \int_X |f_n - f| \, d\mu = \lim_{n\to\infty} \left(\int_X u_n \, d\mu - \int_X \ell_n \, d\mu \right) = 0.$$

So the implication $\Leftarrow$ holds for nonnegative functions. As $g = g_+ - g_-$ and $((f_n)_+)_{n=1}^\infty$ and $((f_n)_-)_{n=1}^\infty$ are sequences of functions in $L^1(\mu)$ converging pointwise μ-a. e. to $f_+ \in L^1(\mu)$ and $f_- \in L^1(\mu)$, respectively, it is easy to see that the general case follows from the case for nonnegative functions. $\square$

If g is a L^1 function on a general measure space $(X,\mathcal{A},\mu)$, then the sequence of nonnegative measurable functions $(g_M)_{M=1}^\infty$, where $g_M = |g| \cdot \mathbb{1}_{\{|g|\geq M\}}$, decreases to 0 pointwise and is dominated by $|g|$. Therefore, the monotone convergence theorem (or, alternatively, Lebesgue's dominated convergence theorem) affirms that

$$\lim_{M\to\infty} \int_{\{|g|\geq M\}} |g| \, d\mu = 0.$$

This suggests introducing the following concept.

Definition A.1.40. Let $(X, \mathcal{A}, \mu)$ be a measure space. A sequence of measurable functions $(f_n)_{n=1}^{\infty}$ is *uniformly integrable* if

$$\lim_{M\to\infty} \sup_{n\in\mathbb{N}} \int_{\{|f_n|\geq M\}} |f_n| \, d\mu = 0.$$

On finite measure spaces, there exists a generalization of Lebesgue's dominated convergence theorem (Theorem A.1.38).

Theorem A.1.41. *Let $(X, \mathcal{A}, \mu)$ be a finite measure space and $(f_n)_{n=1}^{\infty}$ a sequence of measurable functions that converges pointwise μ-a. e. to a function f.*

(a) *If $(f_n)_{n=1}^{\infty}$ is uniformly integrable, then $f_n \in L^1(\mu)$ for all $n \in \mathbb{N}$ and $f \in L^1(\mu)$. Moreover,*

$$\lim_{n\to\infty} \|f_n - f\|_1 = 0 \quad \textit{and} \quad \lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu.$$

(b) *If $f, f_n \in L^1(\mu)$ and $f_n \geq 0$ μ-a. e. for all $n \in \mathbb{N}$, then $\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu$ implies that $(f_n)_{n=1}^{\infty}$ is uniformly integrable.*

Proof. See Theorem 16.14 in Billingsley [7]. □

Corollary A.1.42. *Let $(X, \mathcal{A}, \mu)$ be a finite measure space and $(f_n)_{n=1}^{\infty}$ a sequence of integrable functions that converges pointwise μ-a. e. to an integrable function f. Then the following conditions are equivalent:*

(a) *The sequence $(f_n)_{n=1}^{\infty}$ is uniformly integrable.*
(b) $\lim_{n\to\infty} \|f_n - f\|_1 = 0$.
(c) $\lim_{n\to\infty} \|f_n\|_1 = \|f\|_1$.

Proof. Part (a) of Theorem A.1.41 yields (a)⇒(b). That (b)⇒(c) follows from $\big|\|f_n\|_1 - \|f\|_1\big| \leq \|f_n - f\|_1$. Finally, replacing f_n by $|f_n - f|$ and f by 0 in part (b) of Theorem A.1.41 gives (c)⇒(a). □

Obviously, any sequence of measurable functions that converges uniformly on an entire space does converge pointwise. It is well known that the converse is not true in general, and it is thus natural to ask whether, in some way, a pointwise convergent sequence converges "almost" uniformly.

Definition A.1.43. Let $(X, \mathcal{A}, \mu)$ be a measure space. A sequence $(f_n)_{n=1}^{\infty}$ of measurable functions on X is said to *converge μ-almost uniformly* to a function f if for every $\varepsilon > 0$ there exists $Y \in \mathcal{A}$ such that $\mu(Y) < \varepsilon$ and $(f_n)_{n=1}^{\infty}$ converges uniformly to f on $X \setminus Y$.

It is clear that almost uniform convergence implies almost everywhere pointwise convergence. The converse is not true in general but these two types of convergence are one and the same on any finite measure space.

Theorem A.1.44 (Egorov's theorem). *Let $(X, \mathcal{A}, \mu)$ be a finite measure space. A sequence $(f_n)_{n=1}^{\infty}$ of measurable functions on X converges pointwise μ-almost everywhere to a limit function f if and only if that sequence converges μ-almost uniformly to f.*

Proof. See Chapter 3, Exercise 16 in Rudin [58]. □

The reader ought to convince themself that this result does not generally hold on infinite spaces.

Convergence in measure is another interesting type of convergence.

Definition A.1.45. Let $(X, \mathcal{A}, \mu)$ be a measure space. A sequence $(f_n)_{n=1}^{\infty}$ of measurable functions *converges in measure* to a measurable function f provided that for each $\varepsilon > 0$,

$$\lim_{n\to\infty} \mu(\{x \in X : |f_n(x) - f(x)| > \varepsilon\}) = 0.$$

Lemma A.1.46. *Let $(X, \mathcal{A}, \mu)$ be a measure space. If a sequence $(f_n)_{n=1}^{\infty}$ of measurable functions converges in $L^1(\mu)$ to a measurable function f, then $(f_n)_{n=1}^{\infty}$ converges in measure to f.*

Proof. Let $\varepsilon > 0$. Then

$$\mu(\{x \in X : |f_n(x) - f(x)| > \varepsilon\}) \le \int_{\{x\in X : |f_n(x)-f(x)|>\varepsilon\}} \frac{|f_n - f|}{\varepsilon}\, d\mu \le \frac{1}{\varepsilon}\|f_n - f\|_1.$$

Taking the limit of both sides as $n \to \infty$ completes the proof. □

When the measure is finite, there is a close relationship between pointwise convergence and convergence in measure.

Theorem A.1.47. *Let $(X, \mathcal{A}, \mu)$ be a finite measure space and $(f_n)_{n=1}^{\infty}$ a sequence of measurable functions.*

(a) *If $(f_n)_{n=1}^{\infty}$ converges pointwise μ-a. e. to a function f, then $(f_n)_{n=1}^{\infty}$ converges in measure to f.*
(b) *If $(f_n)_{n=1}^{\infty}$ converges in measure to a function f, then there exists a subsequence $(f_{n_k})_{k=1}^{\infty}$ which converges pointwise μ-a. e. to f.*
(c) *$(f_n)_{n=1}^{\infty}$ converges in measure to a function f if and only if each subsequence $(f_{n_k})_{k=1}^{\infty}$ admits a further subsequence $(f_{n_{k_l}})_{l=1}^{\infty}$ that converges pointwise μ-a. e. to f.*

Proof. See Theorem 20.5 in Billingsley [7]. □

The previous two results reveal that, on a finite measure space, a sequence of integrable functions that converges in L^1 to an integrable function admits a subsequence which converges pointwise almost everywhere to that function. In general, the sequence itself might not converge pointwise almost everywhere (see Exercise 8.5.17).

In some sense, the following result is a form of convergence theorem. It asserts that Borel measurable functions can be approximated by continuous functions on "arbitrarily large" portions of their domain.

Theorem A.1.48. *Let $(X, \mathcal{B}(X), \mu)$ be a finite Borel measure space and let $f : X \to \overline{\mathbb{R}}$ be a Borel measurable function. Given any $\varepsilon > 0$, for every $B \in \mathcal{B}(X)$ there is a closed set E with $\mu(B \setminus E) < \varepsilon$ such that $f|_E$ is continuous. If B is locally compact, then the set E can be chosen to be compact and then there is a continuous function $f_\varepsilon : X \to \overline{\mathbb{R}}$ with compact support that coincides with f on E and such that $\sup_{x \in X} |f_\varepsilon(x)| \le \sup_{x \in X} |f(x)|$.*

A.1.7 Mutual singularity, absolute continuity and equivalence of measures

We now leave aside convergence of sequences of functions and recall the definitions of mutually singular, absolutely continuous, and equivalent measures.

Definition A.1.49. Let $(X, \mathcal{A})$ be a measurable space, and μ and ν be two measures on $(X, \mathcal{A})$.

(a) The measures μ and ν are said to be *mutually singular*, denoted by $\mu \perp \nu$, if there exist disjoint sets $X_\mu, X_\nu \in \mathcal{A}$ such that $\mu(X \setminus X_\mu) = 0 = \nu(X \setminus X_\nu)$.

(b) The measure μ is said to be *absolutely continuous with respect to* ν, denoted $\mu \ll \nu$, if $\nu(A) = 0 \implies \mu(A) = 0$.

(c) The measures μ and ν are said to be *equivalent* if $\mu \ll \nu$ and $\nu \ll \mu$.

The Radon–Nikodym theorem provides a characterization of absolute continuity. Though it is valid for σ-finite measures, the following version for finite measures is sufficient for our purposes.

Theorem A.1.50 (Radon–Nikodym theorem). *Let $(X, \mathcal{A})$ be a measurable space and let μ and ν be two finite measures on $(X, \mathcal{A})$. Then the following statements are equivalent:*

(a) $\mu \ll \nu$.

(b) *For every $\varepsilon > 0$ there exists $\delta > 0$ such that $\nu(A) < \delta \implies \mu(A) < \varepsilon$.*

(c) *There exists a ν-a. e. unique function $f \in L^1(\nu)$ such that $f \ge 0$ and*

$$\mu(A) = \int_A f \, d\nu, \quad \forall A \in \mathcal{A}.$$

Proof. See relation (32.4) and Theorem 32.2 in Billingsley [7]. □

Remark A.1.51. The function f is often denoted by $\frac{d\mu}{d\nu}$ and called the *Radon–Nikodym derivative* of μ with respect to ν.

Per Lemma A.1.34(f), two integrable functions are equal almost everywhere if and only if their integrals are equal over every measurable set. When the measure is finite, we can restrict our attention to any generating π-system.

Corollary A.1.52. *Let $(X, \mathcal{A}, \nu)$ be a finite measure space and suppose that $\mathcal{A} = \sigma(\mathcal{P})$ for some π-system $\mathcal{P}$. Let $f, g \in L^1(\nu)$. Then*

$$f = g \ \ \nu\text{-a.e.} \quad \Longleftrightarrow \quad \int_P f\,d\nu = \int_P g\,d\nu, \quad \forall P \in \mathcal{P}.$$

Proof. The direct implication $\Rightarrow$ is obvious. So let us assume that $\int_P f\,d\nu = \int_P g\,d\nu$ for all $P \in \mathcal{P}$. The measures $\mu_f(A) := \int_A f\,d\nu$ and $\mu_g(A) := \int_A g\,d\nu$ are equal on the π-system $\mathcal{P}$. According to Lemma A.1.26, this implies that $\mu_f = \mu_g$. It follows from the uniqueness part of the Radon–Nikodym theorem that $f = g$ ν-almost everywhere. □

A.1.8 The space $C(X)$, its dual $C(X)^*$ and the subspace $M(X)$

Another important result is Riesz representation theorem. Before stating it, we first establish some notation. Let X be a compact metrizable space. Let $C(X)$ be the set of all continuous real-valued functions on X. This set becomes a normed vector space when endowed with the supremum norm

$$\|f\|_\infty := \sup\{|f(x)| : x \in X\}. \tag{A.3}$$

This norm defines a metric on $C(X)$ in the usual way:

$$d_\infty(f, g) := \|f - g\|_\infty = \sup\{|f(x) - g(x)| : x \in X\}.$$

The topology induced by the metric d_∞ on $C(X)$ is called the *topology of uniform convergence* on X. Indeed, $\lim_{n\to\infty} d_\infty(f_n, f) = 0$ if and only if the sequence $(f_n)_{n=1}^\infty$ converges to f uniformly on X. It is not hard to see that $C(X)$ is a separable Banach space (i. e., a separable and complete normed vector space).

Let $C(X)^*$ denote the *dual space* of $C(X)$, that is,

$$C(X)^* := \{F : C(X) \to \mathbb{R} \mid F \text{ is continuous and linear}\}.$$

Recall that a real-valued function F defined on $C(X)$ is called a *functional* on $C(X)$. It is well known that a linear functional F is continuous if and only if it is *bounded*, that is, if and only if its operator norm $\|F\|$ is finite, where

$$\|F\| := \sup\{|F(f)| : f \in C(X) \text{ and } \|f\|_\infty \le 1\}. \tag{A.4}$$

So $C(X)^*$ can also be described as the normed vector space of all bounded linear functionals on $C(X)$. The operator norm defines a metric on $C(X)^*$ in the usual manner:

$$d(F, G) := \|F - G\|.$$

The topology induced by the metric d on $C(X)^*$ is called the *operator norm topology*, or strong topology, on $C(X)^*$.

It is not difficult to see that $C(X)^*$ is a separable Banach space. Furthermore, a linear functional F is said to be *normalized* if $F(1) = 1$ and is called *positive* if $F(f) \geq 0$ whenever $f \geq 0$.

Finally, we denote the set of all Borel probability measures on X by $M(X)$. This set is clearly convex and can be characterized as follows.

Theorem A.1.53 (Riesz representation theorem). *Let X be a compact metrizable space, and let F be a normalized and positive linear functional on $C(X)$. Then there exists a unique $\mu \in M(X)$ such that*

$$F(f) = \int_X f\, d\mu, \quad \forall f \in C(X). \tag{A.5}$$

Conversely, any $\mu \in M(X)$ defines a normalized positive linear functional on $C(X)$ via formula (A.5). This linear functional is bounded.

Proof. The converse statement is straightforward to check. For the other direction, see Theorem 2.14 in Rudin [58]. □

It immediately follows from Riesz representation theorem that every Borel probability measure on a compact metrizable space is uniquely determined by the way it integrates continuous functions on that space.

Corollary A.1.54. *If μ and ν are two Borel probability measures on a compact metrizable space X, then*

$$\mu = \nu \quad \Longleftrightarrow \quad \int_X f\, d\mu = \int_X f\, d\nu, \quad \forall f \in C(X).$$

Let us now discuss the weak* *topology* on the set $M(X)$. Recall that if Z is a set and $(Z_\alpha)_{\alpha \in A}$ is a family of topological spaces, then the weak topology induced on Z by a collection of maps $\{\psi_\alpha : Z \to Z_\alpha \mid \alpha \in A\}$ is the smallest topology on Z that makes each ψ_α continuous. Evidently, the sets $\psi_\alpha^{-1}(U_\alpha)$, for U_α open in Z_α, constitute a subbase for the weak topology. The weak* topology on $M(X)$ is the weak topology induced by $C(X)$ on its dual space $C(X)^*$, where measures in $M(X)$ and normalized positive linear functionals in $C(X)^*$ are identified via the Riesz representation theorem. Note that $M(X)$ is metrizable, although $C(X)^*$ with the weak* topology usually is not. Indeed, both $C(X)$ and its subspace $C(X, [0, 1])$ of continuous functions on X taking values in $[0, 1]$, are separable since X is a compact metrizable space. Then for any dense subset $\{f_n\}_{n=1}^{\infty}$ of $C(X, [0, 1])$, a metric on $M(X)$ is

$$d(\mu, \nu) = \sum_{n=1}^{\infty} \frac{1}{2^n} \left| \int_X f_n\, d\mu - \int_X f_n\, d\nu \right|.$$

In this book, we will denote the convergence of a sequence of measures $(\mu_n)_{n=1}^{\infty}$ to a measure μ in the weak* topology of $M(X)$ by $\mu_n \xrightarrow{*} \mu$.

Remark A.1.55. Note that this notion is often presented as "weak convergence" of measures. This can be slightly confusing at first sight, but it helps to bear in mind that, as we have seen above, the weak* topology is just one instance of a weak topology.

The following theorem gives several equivalent characterizations of weak* convergence of Borel probability measures.

Theorem A.1.56 (Portmanteau theorem). *Let $(\mu_n)_{n=1}^{\infty}$ and μ be Borel probability measures on a compact metrizable space X. The following statements are equivalent:*
(a) $\mu_n \xrightarrow{*} \mu$.
(b) *For all continuous functions $f : X \to \mathbb{R}$,*

$$\lim_{n\to\infty} \int_X f\, d\mu_n = \int_X f\, d\mu.$$

(c) *For all closed sets $F \subseteq X$,*

$$\limsup_{n\to\infty} \mu_n(F) \le \mu(F).$$

(d) *For all open sets $G \subseteq X$,*

$$\liminf_{n\to\infty} \mu_n(G) \le \mu(G).$$

(e) *For all sets $A \in \mathcal{B}(X)$ such that $\mu(\partial A) = 0$,*

$$\lim_{n\to\infty} \mu_n(A) = \mu(A).$$

Proof. See Theorem 2.1 in Billingsley [8]. □

For us, the most important result concerning weak* convergence of measures is that the set $M(X)$ of all Borel probability measures on a compact metrizable space X is a compact and convex set in the weak* topology.

In order to establish this, we need to remember Banach–Alaoglu's theorem. In this theorem, note that the boundedness and closedness are with respect to the operator norm on the dual space while the compactness is with respect to the weak* topology on the dual space.

Theorem A.1.57 (Banach–Alaoglu's theorem). *The closed unit ball in the dual space B^* of a Banach space B is compact in the* weak* *topology on B^*. Furthermore, every closed, bounded subset of B^* is compact in the* weak* *topology on B^*.*

Proof. See Theorem V.4.2 and Corollary V.4.3 in Dunford and Schwartz [20]. □

Theorem A.1.58. *Let X be a compact metrizable space. The set $M(X)$ is compact and convex in the* weak* *topology of $C(X)^*$.*

Proof. The set $M(X)$ is closed with respect to the operator norm topology on $C(X)^*$. Indeed, suppose that $(\mu_n)_{n=1}^{\infty}$ is a sequence in $M(X)$ which converges to a $F \in C(X)^*$ in the operator norm topology of $C(X)^*$. In other words, suppose that $\lim_{n\to\infty} \|\mu_n - F\| = 0$. By definition of the operator norm (see (A.4)) and thanks to the linearity of F, this implies that

$$F(f) = \lim_{n\to\infty} \int_X f \, d\mu_n, \quad \forall f \in C(X).$$

In particular, F is normalized (since $F(1) = 1$) and positive (as $F(f) \geq 0$ for all $f \geq 0$). By Riesz representation theorem (Theorem A.1.53), there is $\mu \in M(X)$ that represents F. So $F \in M(X)$, and thus $M(X)$ is closed in the operator norm topology on $C(X)^*$.

The set $M(X)$ is also bounded in that topology. Indeed, if $\mu \in M(X)$ then

$$\|\mu\| \leq \sup\left\{ \int_X |f| \, d\mu : f \in C(X), \|f\|_\infty \leq 1 \right\} = 1.$$

Since X is a compact metrizable space, the space $C(X)$ is a Banach space, as earlier mentioned. We can then infer from Banach–Alaoglu's theorem that the set $M(X)$ is compact in the weak* topology.

The convexity of $M(X)$ is obvious. Indeed, if $\mu, \nu \in M(X)$ so is any convex combination $m = \alpha\mu + (1-\alpha)\nu$, where $\alpha \in [0,1]$. □

A.1.9 Expected values and conditional expectation functions

The mean or expected value of a function over a set is a straightforward generalisation of the mean value of a real-valued function defined on an interval of the real line.

Definition A.1.59. Let $(X, \mathcal{A}, \mu)$ be a probability space and let $\varphi \in L^1(\mu)$. The *mean* or *expected value* $E(\varphi|A)$ of the function φ over the set $A \in \mathcal{A}$ is defined to be

$$E(\varphi|A) := \begin{cases} \frac{1}{\mu(A)} \int_A \varphi \, d\mu & \text{if } \mu(A) > 0 \\ 0 & \text{if } \mu(A) = 0. \end{cases}$$

Given that μ is a probability measure, the expected value of φ over the entire space is simply given by

$$E(\varphi) := E(\varphi|X) = \int_X \varphi \, d\mu =: \mu(\varphi).$$

Our next goal is to give the definition of the conditional expectation of a function with respect to a σ-algebra. Let $(X, \mathcal{A}, \mu)$ be a probability space and $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Let also $\varphi \in L^1(X, \mathcal{A}, \mu)$. Notice that $\varphi : X \to \mathbb{R}$ is not necessarily measurable if X is endowed with the sub-σ-algebra $\mathcal{B}$ instead of the σ-algebra $\mathcal{A}$. In short, we say that φ is $\mathcal{A}$-measurable but not necessarily $\mathcal{B}$-measurable. We aim to find a function $E(\varphi|\mathcal{B}) \in L^1(X, \mathcal{B}, \mu)$ such that

$$\int_B E(\varphi|\mathcal{B})\, d\mu = \int_B \varphi\, d\mu, \quad \forall B \in \mathcal{B}. \tag{A.6}$$

This condition means that the function $E(\varphi|\mathcal{B})$ has the same expected value as φ on every measurable set belonging to the sub-σ-algebra $\mathcal{B}$. Accordingly, $E(\varphi|\mathcal{B})$ is called the *conditional expectation of φ with respect to $\mathcal{B}$*.

We now demonstrate the existence and μ-a. e. uniqueness of the conditional expectation. Let us begin with the existence of that function. Suppose first that the function φ is nonnegative. If $\varphi = 0$ μ-a. e., then simply set $E(\varphi|\mathcal{B}) = 0$. If $\varphi \neq 0$ μ-a. e. then the set function $\nu(A) := \int_A \varphi\, d\mu$ defines a finite measure on $(X, \mathcal{A})$ which is absolutely continuous with respect to μ. The restriction of ν to $\mathcal{B}$ also determines a finite measure on $(X, \mathcal{B})$ which is absolutely continuous with respect to the restriction of μ to $\mathcal{B}$. So by the Radon–Nikodym theorem (Theorem A.1.50), there exists a μ-a. e. unique nonnegative function $\widehat{\varphi} \in L^1(X, \mathcal{B}, \mu)$ such that $\nu(B) = \int_B \widehat{\varphi}\, d\mu$ for every $B \in \mathcal{B}$. Then

$$\int_B \widehat{\varphi}\, d\mu = \nu(B) = \int_B \varphi\, d\mu, \quad \forall B \in \mathcal{B}.$$

The point here is that although it may look as if we have not really achieved anything, we have actually gained that $\widehat{\varphi}$ is $\mathcal{B}$-measurable, whereas φ may not be. Therefore, $\widehat{\varphi}$ is the sought-after conditional expectation $E(\varphi|\mathcal{B})$ of φ with respect to $\mathcal{B}$.

If φ takes both negative and positive values, write $\varphi = \varphi_+ - \varphi_-$, where $\varphi_+(x) := \max\{\varphi(x), 0\}$ is the positive part of φ and $\varphi_-(x) := \max\{-\varphi(x), 0\}$ is the negative part of φ. Then define the conditional expectation linearly, that is, set

$$E(\varphi|\mathcal{B}) := E(\varphi_+|\mathcal{B}) - E(\varphi_-|\mathcal{B}).$$

This proves the existence of the conditional expectation function. Its μ-a. e. uniqueness follows from its defining property (A.6) and Lemma A.1.34(f).

The conditional expectation exhibits several natural properties. We mention a few of them in the next proposition.

Proposition A.1.60. *Let $(X, \mathcal{A}, \mu)$ be a probability space, let $\mathcal{B}$ and $\mathcal{C}$ denote sub-σ-algebras of $\mathcal{A}$ and let $\varphi \in L^1(X, \mathcal{A}, \mu)$.*
(a) *If $\varphi \geq 0$ μ-a. e., then $E(\varphi|\mathcal{B}) \geq 0$ μ-a. e.*
(b) *If $\varphi_1 \geq \varphi_2$ μ-a. e., then $E(\varphi_1|\mathcal{B}) \geq E(\varphi_2|\mathcal{B})$ μ-a. e.*
(c) $|E(\varphi|\mathcal{B})| \leq E(|\varphi|\,|\,\mathcal{B})$.

(d) *The functional* $E(\cdot|\mathcal{B})$ *is linear, i. e. for any* $c_1, c_2 \in \mathbb{R}$ *and* $\varphi_1, \varphi_2 \in L^1(X, \mathcal{A}, \mu)$,

$$E(c_1\varphi_1 + c_2\varphi_2 \big| \mathcal{B}) = c_1 E(\varphi_1|\mathcal{B}) + c_2 E(\varphi_2|\mathcal{B}).$$

(e) *If* φ *is already* $\mathcal{B}$*-measurable, then* $E(\varphi|\mathcal{B}) = \varphi$*. In particular, we have that* $E(E(\varphi|\mathcal{B})\big|\mathcal{B}) = E(\varphi|\mathcal{B})$*. Also, if* $\varphi = c \in \mathbb{R}$ *is a constant function, then* $E(\varphi|\mathcal{B}) = \varphi = c$.

(f) *If* $\mathcal{C} \subseteq \mathcal{B}$*, then* $E(\varphi|\mathcal{C}) = E\big(E(\varphi|\mathcal{B})\,\big|\,\mathcal{C}\big)$.

Proof. This is left as an exercise to the reader. □

We will now determine the conditional expectation of an arbitrary integrable function φ with respect to various sub-σ-algebras of particular interest.

Example A.1.61. Let $(X, \mathcal{A}, \mu)$ be a probability space. The family $\mathcal{N}$ of all measurable sets that are either of null or of full measure constitutes a sub-σ-algebra of $\mathcal{A}$. Let $\varphi \in L^1(X, \mathcal{A}, \mu)$. Then the function $E(\varphi|\mathcal{N})$ has to belong to $L^1(X, \mathcal{N}, \mu)$ and must satisfy condition (A.6). In particular, $E(\varphi|\mathcal{N})$ must be $\mathcal{N}$-measurable. This means that for each Borel subset R of $\mathbb{R}$, the function $E(\varphi|\mathcal{N})$ must be such that $E(\varphi|\mathcal{N})^{-1}(R) \in \mathcal{N}$. Among others, for every $t \in \mathbb{R}$ we must have $E(\varphi|\mathcal{N})^{-1}(\{t\}) \in \mathcal{N}$; in other words, for each $t \in \mathbb{R}$ the set $E(\varphi|\mathcal{N})^{-1}(\{t\})$ must be of measure zero or of measure one. Also bear in mind that

$$X = E(\varphi|\mathcal{N})^{-1}(\mathbb{R}) = \bigcup_{t\in\mathbb{R}} E(\varphi|\mathcal{N})^{-1}(\{t\}).$$

Since the above union consists of mutually disjoint sets of measure zero and one, it follows that only one of these sets can be of measure one. In other words, there exists a unique $t \in \mathbb{R}$ such that $E(\varphi|\mathcal{N})^{-1}(\{t\}) = A$ for some $A \in \mathcal{A}$ with $\mu(A) = 1$. Because the function $E(\varphi|\mathcal{N})$ is unique up to a set of measure zero, we may assume without loss of generality that $A = X$. Hence, $E(\varphi|\mathcal{N})$ is a constant function. More specifically, its value is

$$E(\varphi|\mathcal{N}) = \int_X E(\varphi|\mathcal{N})\, d\mu = \int_X \varphi\, d\mu.$$

Example A.1.62. Let $(X, \mathcal{A})$ be a measurable space and let $\alpha = \{A_n\}_{n=1}^{\infty}$ be a countable measurable partition of X. That is, each $A_n \in \mathcal{A}$, $A_i \cap A_j = \emptyset$ for all $i \neq j$ and $X = \bigcup_{n=1}^{\infty} A_n$. The sub-$\sigma$-algebra of $\mathcal{A}$ generated by α is the family of all sets which can be written as a union of elements of α, that is,

$$\sigma(\alpha) = \left\{A \subseteq X : A = \bigcup_{j\in J} A_j \text{ for some } J \subseteq \mathbb{N}\right\}.$$

When α is finite, so is $\sigma(\alpha)$. When α is countably infinite, $\sigma(\alpha)$ is uncountable. Let μ be a probability measure on $(X, \mathcal{A})$. Let $\varphi \in L^1(X, \mathcal{A}, \mu)$ and set $\mathcal{B} = \sigma(\alpha)$. Then the conditional expectation $E(\varphi|\mathcal{B}) : X \to \mathbb{R}$ has to be a $L^1(X, \mathcal{B}, \mu)$ function that satisfies condition (A.6). In particular, $E(\varphi|\mathcal{B})$ must be $\mathcal{B}$-measurable. Thus, for any $t \in \mathbb{R}$ we must have $E(\varphi|\mathcal{B})^{-1}(\{t\}) \in \mathcal{B}$, that is, the set $E(\varphi|\mathcal{B})^{-1}(\{t\})$ must be a union of elements of α. This means that the conditional expectation function $E(\varphi|\mathcal{B})$ is constant on each element of α. Let $A_n \in \alpha$. If $\mu(A_n) = 0$ then $E(\varphi|\mathcal{B})|_{A_n} = 0$. Otherwise,

$$E(\varphi|\mathcal{B})|_{A_n} = \frac{1}{\mu(A_n)} \int_{A_n} E(\varphi|\mathcal{B})\, d\mu = \frac{1}{\mu(A_n)} \int_{A_n} \varphi\, d\mu = E(\varphi|A_n).$$

In summary, the conditional expectation $E(\varphi|\mathcal{B})$ of a function φ with respect to a sub-σ-algebra generated by a countable measurable partition is constant on each element of that partition. More precisely, on any given element of the partition, $E(\varphi|\mathcal{B})$ is equal to the mean value of φ on that element.

The next result is a special case of a theorem originally due to Doob and called the martingale convergence theorem. But, first, let us define the martingale itself.

Definition A.1.63. Let $(X, \mathcal{A}, \mu)$ be a probability space. Let $(\mathcal{A}_n)_{n=1}^{\infty}$ be a sequence of sub-σ-algebras of $\mathcal{A}$. Let also $(\varphi_n : X \to \mathbb{R})_{n=1}^{\infty}$ be a sequence of $\mathcal{A}$-measurable functions. The sequence $((\varphi_n, \mathcal{A}_n))_{n=1}^{\infty}$ is called a *martingale* if the following conditions are satisfied:
(a) $(\mathcal{A}_n)_{n=1}^{\infty}$ is an ascending sequence, that is, $\mathcal{A}_n \subseteq \mathcal{A}_{n+1}$ for all $n \in \mathbb{N}$.
(b) φ_n is $\mathcal{A}_n$-measurable for all $n \in \mathbb{N}$.
(c) $\varphi_n \in L^1(\mu)$ for all $n \in \mathbb{N}$.
(d) $E(\varphi_{n+1}|\mathcal{A}_n) = \varphi_n$ μ-a. e. for all $n \in \mathbb{N}$.

Theorem A.1.64 (Martingale convergence theorem). *Let $(X, \mathcal{A}, \mu)$ be a probability space. If $((\varphi_n, \mathcal{A}_n))_{n=1}^{\infty}$ is a martingale such that*

$$\sup_{n \in \mathbb{N}} \|\varphi_n\|_1 < \infty,$$

then there exists $\widehat{\varphi} \in L^1(X, \mathcal{A}, \mu)$ such that

$$\lim_{n\to\infty} \varphi_n(x) = \widehat{\varphi}(x) \quad \textit{for } \mu\textit{-a. e. } x \in X \quad \textit{and} \quad \|\widehat{\varphi}\|_1 \le \sup_{n\in\mathbb{N}} \|\varphi_n\|_1.$$

Proof. See Theorem 35.5 in Billingsley [7]. □

One natural martingale is formed by the conditional expectations of a function with respect to an ascending sequence of sub-σ-algebras.

Example A.1.65. Let $(X, \mathcal{A}, \mu)$ be a probability space and let $(\mathcal{A}_n)_{n=1}^{\infty}$ be an ascending sequence of sub-σ-algebras of $\mathcal{A}$. For any $\varphi \in L^1(X, \mathcal{A}, \mu)$, the sequence $\{(E(\varphi|\mathcal{A}_n), \mathcal{A}_n)\}_{n=1}^{\infty}$

is a martingale. Indeed, set $\varphi_n = E(\varphi|\mathcal{A}_n)$ for all $n \in \mathbb{N}$. Condition (a) in Definition A.1.63 is automatically fulfilled. Conditions (b) and (c) follow from the very definition of the conditional expectation function. Regarding condition (d), a straightforward application of Proposition A.1.60(f) gives

$$E(\varphi_{n+1}|\mathcal{A}_n) = E\big(E(\varphi|\mathcal{A}_{n+1}) \mid \mathcal{A}_n\big) = E(\varphi|\mathcal{A}_n) = \varphi_n \ \mu\text{-a.e.}, \quad \forall n \in \mathbb{N}.$$

So $\{(E(\varphi|\mathcal{A}_n), \mathcal{A}_n)\}_{n=1}^{\infty}$ is a martingale. Using Proposition A.1.60(c), note that

$$\sup_{n\in\mathbb{N}} \|\varphi_n\|_1 = \sup_{n\in\mathbb{N}} \int_X \big|E(\varphi|\mathcal{A}_n)\big|\, d\mu \le \sup_{n\in\mathbb{N}} \int_X E\big(|\varphi| \mid \mathcal{A}_n\big)\, d\mu = \int_X |\varphi|\, d\mu = \|\varphi\|_1 < \infty.$$

According to Theorem A.1.64, there thus exists $\widehat{\varphi} \in L^1(X, \mathcal{A}, \mu)$ such that

$$\lim_{n\to\infty} E(\varphi|\mathcal{A}_n)(x) = \widehat{\varphi}(x) \quad \text{for } \mu\text{-a.e. } x \in X \quad \text{and} \quad \|\widehat{\varphi}\|_1 \le \|\varphi\|_1.$$

What is $\widehat{\varphi}$? This is the question we will answer in Theorem A.1.67.

Beforehand, we establish the uniform integrability of this martingale (see Definition A.1.40).

Lemma A.1.66. *Let $(X, \mathcal{A}, \mu)$ be a probability space and let $(\mathcal{A}_n)_{n=1}^{\infty}$ be a sequence of sub-σ-algebras of $\mathcal{A}$. For any $\varphi \in L^1(X, \mathcal{A}, \mu)$, the sequence $(E(\varphi|\mathcal{A}_n))_{n=1}^{\infty}$ is uniformly integrable.*

Proof. Without loss of generality, we may assume that $\varphi \ge 0$. Let $\varepsilon > 0$. Since $\nu(A) = \int_A \varphi\, d\mu$ is absolutely continuous with respect to μ, it follows from the Radon–Nikodym theorem (Theorem A.1.50) that there exists $\delta > 0$ such that

$$A \in \mathcal{A},\ \mu(A) < \delta \quad \Longrightarrow \quad \int_A \varphi\, d\mu < \varepsilon. \tag{A.7}$$

Set $M > \int_X \varphi\, d\mu/\delta$. For each $n \in \mathbb{N}$, let

$$X_n(M) = \big\{x \in X : E(\varphi|\mathcal{A}_n)(x) \ge M\big\}.$$

Observe that $X_n(M) \in \mathcal{A}_n$ since $E(\varphi|\mathcal{A}_n)$ is $\mathcal{A}_n$-measurable. Therefore,

$$\mu\big(X_n(M)\big) \le \frac{1}{M} \int_{X_n(M)} E(\varphi|\mathcal{A}_n)\, d\mu = \frac{1}{M} \int_{X_n(M)} \varphi\, d\mu \le \frac{1}{M} \int_X \varphi\, d\mu < \delta$$

for all $n \in \mathbb{N}$. Consequently, by (A.7),

$$\int_{X_n(M)} E(\varphi|\mathcal{A}_n)\, d\mu = \int_{X_n(M)} \varphi\, d\mu < \varepsilon$$

for all $n \in \mathbb{N}$. Thus

$$\sup_{n\in\mathbb{N}} \int\limits_{\{E(\varphi|\mathcal{A}_n)\geq M\}} E(\varphi|\mathcal{A}_n)\, d\mu \leq \varepsilon.$$

Since this holds for all large enough M's and since $\varepsilon > 0$ is arbitrary, we have

$$\lim_{M\to\infty} \sup_{n\in\mathbb{N}} \int\limits_{\{E(\varphi|\mathcal{A}_n)\geq M\}} E(\varphi|\mathcal{A}_n)\, d\mu = 0,$$

that is, the sequence $(E(\varphi|\mathcal{A}_n))_{n=1}^{\infty}$ is uniformly integrable. □

Theorem A.1.67 (Martingale convergence theorem for conditional expectations). *Let $(X, \mathcal{A}, \mu)$ be a probability space and $\varphi \in L^1(X, \mathcal{A}, \mu)$. Let $(\mathcal{A}_n)_{n=1}^{\infty}$ be an ascending sequence of sub-σ-algebras of $\mathcal{A}$ and*

$$\mathcal{A}_\infty := \sigma\left(\bigcup_{n=1}^{\infty} \mathcal{A}_n\right).$$

Then

$$\lim_{n\to\infty} \big\| E(\varphi|\mathcal{A}_n) - E(\varphi|\mathcal{A}_\infty) \big\|_1 = 0 \quad \text{and} \quad \lim_{n\to\infty} E(\varphi|\mathcal{A}_n) = E(\varphi|\mathcal{A}_\infty) \ \ \mu\text{-a. e. on } X.$$

Proof. Let $\varphi_n = E(\varphi|\mathcal{A}_n)$. In Example A.1.65 and Lemma A.1.66, we have seen that $((\varphi_n, \mathcal{A}_n))_{n=1}^{\infty}$ is a uniformly integrable martingale such that

$$\lim_{n\to\infty} \varphi_n = \widehat{\varphi} \quad \mu\text{-a. e. on } X$$

for some $\widehat{\varphi} \in L^1(X, \mathcal{A}, \mu)$. For all $n \in \mathbb{N}$ the function φ_n is $\mathcal{A}_\infty$-measurable since it is $\mathcal{A}_n$-measurable and $\mathcal{A}_n \subseteq \mathcal{A}_\infty$. Thus $\widehat{\varphi}$ is $\mathcal{A}_\infty$-measurable, too. Moreover, it follows from Theorem A.1.41 that

$$\lim_{n\to\infty} \|\varphi_n - \widehat{\varphi}\|_1 = 0 \quad \text{and} \quad \lim_{n\to\infty} \int\limits_A \varphi_n\, d\mu = \int\limits_A \widehat{\varphi}\, d\mu, \quad \forall A \in \mathcal{A}.$$

Therefore, it just remains to show that $\widehat{\varphi} = E(\varphi|\mathcal{A}_\infty)$.

Let $k \in \mathbb{N}$ and $A \in \mathcal{A}_k$. If $n \geq k$, then $A \in \mathcal{A}_n \subseteq \mathcal{A}_\infty$, and thus

$$\int\limits_A \varphi_n\, d\mu = \int\limits_A E(\varphi|\mathcal{A}_n)\, d\mu = \int\limits_A \varphi\, d\mu = \int\limits_A E(\varphi|\mathcal{A}_\infty)\, d\mu.$$

Letting $n \to \infty$ yields

$$\int\limits_A \widehat{\varphi}\, d\mu = \int\limits_A E(\varphi|\mathcal{A}_\infty)\, d\mu, \quad \forall A \in \mathcal{A}_k.$$

Since k is arbitrary,

$$\int_B \widehat{\varphi}\, d\mu = \int_B E(\varphi|\mathcal{A}_\infty)\, d\mu, \quad \forall B \in \bigcup_{k=1}^{\infty} \mathcal{A}_k.$$

Since $\bigcup_{k=1}^{\infty} \mathcal{A}_k$ is a π-system generating $\mathcal{A}_\infty$ and since both $\widehat{\varphi}$ and $E(\varphi|\mathcal{A}_\infty)$ are $\mathcal{A}_\infty$-measurable, Corollary A.1.52 affirms that $\widehat{\varphi} = E(\varphi|\mathcal{A}_\infty)$ μ-a. e. □

There is also a counterpart of this theorem for descending sequences of σ-algebras.

Theorem A.1.68 (Reversed martingale convergence theorem for conditional expectations). *Let $(X, \mathcal{A}, \mu)$ be a probability space and $\varphi \in L^1(X, \mathcal{A}, \mu)$. If $(\mathcal{A}_n)_{n=1}^{\infty}$ is a descending sequence of sub-σ-algebras of $\mathcal{A}$, then*

$$\lim_{n\to\infty} \left\| E(\varphi|\mathcal{A}_n) - E\left(\varphi \,\middle|\, \bigcap_{n=1}^{\infty} \mathcal{A}_n\right) \right\|_1 = 0 \quad \textit{and} \quad \lim_{n\to\infty} E(\varphi|\mathcal{A}_n) = E\left(\varphi \,\middle|\, \bigcap_{n=1}^{\infty} \mathcal{A}_n\right) \mu\textit{-a. e.}$$

Proof. See Theorem 35.9 in Billingsley [7]. □

Theorems A.1.67/A.1.68 are especially useful for the calculation of the conditional expectation of a function with respect to a sub-σ-algebra generated by an uncountable measurable partition which can be approached by an ascending/descending sequence of sub-σ-algebras generated by countable measurable partitions. See Exercise 8.5.22.

They can also be used to approximate a measurable set by one from a generating sequence of sub-σ-algebras.

Corollary A.1.69. *Let $(X, \mathcal{A}, \mu)$ be a probability space. Let $(\mathcal{A}_n)_{n=1}^{\infty}$ be an ascending sequence of sub-σ-algebras of $\mathcal{A}$ and set $\mathcal{A}_\infty = \sigma(\bigcup_{n=1}^{\infty} \mathcal{A}_n)$. Let $B \in \mathcal{A}_\infty$. For every $\varepsilon > 0$, there exists $A \in \bigcup_{n=1}^{\infty} \mathcal{A}_n$ such that $\mu(A \triangle B) < \varepsilon$.*

Proof. Let $B \in \mathcal{A}_\infty$. It ensues from Theorem A.1.67 that

$$\lim_{n\to\infty} E(\mathbb{1}_B|\mathcal{A}_n) = E(\mathbb{1}_B|\mathcal{A}_\infty) = \mathbb{1}_B \quad \mu\text{-a. e. on } X.$$

By Theorem A.1.47, we deduce that

$$\lim_{n\to\infty} \mu\left(\left\{x \in X : \left|E(\mathbb{1}_B|\mathcal{A}_n)(x) - \mathbb{1}_B(x)\right| \geq \frac{1}{8}\right\}\right) = 0.$$

For every $n \in \mathbb{N}$, let

$$B_n := \left\{x \in X : \left|E(\mathbb{1}_B|\mathcal{A}_n)(x) - \mathbb{1}_B(x)\right| \geq \frac{1}{8}\right\}.$$

Then there exists $N = N(\varepsilon) \in \mathbb{N}$ such that

$$\mu(B_n) \leq \varepsilon/2, \quad \forall\, n \geq N.$$

For every $n \in \mathbb{N}$, let

$$A_n := \left\{x \in X : \left|E(\mathbb{1}_B|\mathcal{A}_n)(x) - 1\right| \le \frac{1}{4}\right\} \in \mathcal{A}_n.$$

On one hand,

$$x \in B \setminus A_n \implies \left|E(\mathbb{1}_B|\mathcal{A}_n)(x) - \mathbb{1}_B(x)\right| > \frac{1}{4} \implies x \in B_n.$$

This means that

$$B \setminus A_n \subseteq B_n.$$

On the other hand,

$$x \in A_n \setminus B \implies \left|E(\mathbb{1}_B|\mathcal{A}_n)(x) - \mathbb{1}_B(x)\right| = \left|E(\mathbb{1}_B|\mathcal{A}_n)(x)\right| \ge \frac{3}{4} \implies x \in B_n.$$

This means that

$$A_n \setminus B \subseteq B_n.$$

Therefore,

$$\mu(A_n \triangle B) = \mu(A_n \setminus B) + \mu(B \setminus A_n) \le 2\mu(B_n) \le \varepsilon, \quad \forall n \ge N.$$

Since $A_n \in \mathcal{A}_n$, we have found some $A \in \bigcup_{n=1}^{\infty} \mathcal{A}_n$ with $\mu(A \triangle B) < \varepsilon$. □

We will now give a proof of Lemma A.1.32 in the case where the algebra is countable.

Proof. Let $\mathcal{A} = \{A_n\}_{n=1}^{\infty}$ be a countable algebra on a set X and let μ be a probability measure on $(X, \sigma(\mathcal{A}))$. Set $\mathcal{A}'_n = \{A_k\}_{k=1}^{n}$. Since $\mathcal{A}'_n$ is a finite set, the algebra $\mathcal{A}(\mathcal{A}'_n)$ it generates is also finite, and thus $\sigma(\mathcal{A}'_n) = \mathcal{A}(\mathcal{A}'_n)$. Moreover, since $\mathcal{A}$ is an algebra, $\sigma(\mathcal{A}'_n) = \mathcal{A}(\mathcal{A}'_n) \subseteq \mathcal{A}$. It follows that

$$\sigma(\mathcal{A}) = \sigma\left(\bigcup_{n=1}^{\infty} \sigma(\mathcal{A}'_n)\right).$$

Let $B \in \sigma(\mathcal{A})$ and $\varepsilon > 0$. Since $(\sigma(\mathcal{A}'_n))_{n=1}^{\infty}$ is an ascending sequence of sub-σ-algebras of $\sigma(\mathcal{A})$ such that $\sigma(\mathcal{A}) = \sigma(\bigcup_{n=1}^{\infty} \sigma(\mathcal{A}'_n))$, it ensues from Corollary A.1.69 that there exists $A \in \bigcup_{n=1}^{\infty} \sigma(\mathcal{A}'_n)$ with $\mu(A \triangle B) < \varepsilon$. Since $\bigcup_{n=1}^{\infty} \sigma(\mathcal{A}'_n) \subseteq \mathcal{A}$, we have found some $A \in \mathcal{A}$ such that $\mu(A \triangle B) < \varepsilon$. □

Finally, we introduce the concept of conditional measure and relate it to the concept of expected value.

Definition A.1.70. Let $(X, \mathcal{A}, \mu)$ be a probability space and let $B \in \mathcal{A}$ be such that $\mu(B) > 0$. The set function $\mu_B : \mathcal{A} \to [0, 1]$ defined by setting

$$\mu_B(A) := \frac{\mu(A \cap B)}{\mu(B)}, \quad \forall A \in \mathcal{A}$$

is a probability measure on $(X, \mathcal{A})$ called the *conditional measure of* μ *on* B.

Note that for every $\varphi \in L^1(X, \mathcal{A}, \mu)$,

$$\int_X \varphi \, d\mu_B = \int_B \varphi \, d\mu_B + \int_{X \setminus B} \varphi \, d\mu_B = \frac{1}{\mu(B)} \int_B \varphi \, d\mu + 0 = E(\varphi | B).$$

A.2 Analysis

Theorem A.2.1 (Inverse function theorem). *Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and $x_0 \in \mathcal{X}$. If F is a C^1 (i. e., continuously differentiable) function on some neighborhood of x_0 such that $F'(x_0)$ is invertible, then there exists an open neighborhood U of x_0 such that $F(U)$ is open in $\mathcal{Y}$ and F bijectively maps U to $F(U)$. Furthermore, the inverse function of $F|_U$, mapping $F(U)$ to (U), is continuously differentiable.*

Bibliography

[1] Aaronson, J., An Introduction to Infinite Ergodic Theory. Mathematical Surveys and Monographs, vol. **50**. American Mathematical Society (1997).

[2] Adler, R. L., Konheim, A. G., and McAndrew, M. H., Topological entropy. Trans. Am. Math. Soc. **114**, 309–319 (1965).

[3] Alsedá, L., Libre, J., and Misiurewicz, M., Combinatorial Dynamics and Entropy in Dimension One, 2nd edn. World Scientific (2000).

[4] Anosov, D., Geodesic flows on Riemannian manifolds with negative curvature. Proceedings of the Steklov Institute of Mathematics, vol. **90**. AMS, Providence, RI (1967).

[5] Anosov, D., Sinai, Y., Some smooth ergodic systems. Russ. Math. Surv. **22**(5), 103–167 (1967).

[6] Avila, A., Forni, G., Weak mixing for interval exchange transformations and translation flows. Ann. Math. **165**, 637–664 (2007).

[7] Billingsley, P., Probability and Measure, 3rd edn. John Wiley and Sons, New York (1995).

[8] Billingsley, P., Convergence of Probability Measures, 2nd edn. John Wiley and Sons, New York (1999).

[9] Birkhoff, G. D., Proof of the ergodic theorem. Proc. Natl. Acad. Sci. USA **17**, 656–660 (1931).

[10] Bowen, R., Entropy for group endomorphisms and homogeneous maps. Trans. Am. Math. Soc. **153**, 401–414 (1971).

[11] Bowen, R., Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms. Lect. Notes in Math., vol. **470**. Springer, (1975).

[12] Brin, M., Katok, A., On Local Entropy. In: Geometric Dynamics. Lecture Notes in Mathematics, vol. **1007**, pp. 30–38. Springer, Berlin-New York (1983).

[13] Brin, M., Stuck, G., Introduction to Dynamical Systems. Cambridge University Press, Cambridge (2002).

[14] Bryant, B. F., On expansive homeomorphisms. Pac. J. Math. **10**, 1163–1167 (1960).

[15] Conway, J. B., A Course in Functional Analysis, 2nd edn. Springer, New York (1990).

[16] Cornfeld, I. P., Fomin, S. V., and Sinai, Y. G., Ergodic Theory. Springer, New York (1982).

[17] Coven, E. M., Reddy, W. L., Positively expansive maps of compact manifolds. Lect. Notes in Math., vol. **819** (1980).

[18] Denker, M., Urbański, M., Absolutely continuous invariant measures for expansive rational maps with rationally indifferent periodic points. Forum Math. **3**, 561–579 (1991).

[19] Devaney, R. L., An Introduction to Chaotic Dynamical Systems, 2nd edn. Westview Press, Boulder, Colorado (2003).

[20] Dunford, N., Schwartz, J. T., Linear Operators, Part I: General Theory. John Wiley and Sons, New York (1964).

[21] Einsiedler, M., Ward, T., Ergodic Theory with a view towards Number Theory. Graduate Texts in Mathematics, vol. **259**. Springer, London (2011).

[22] Epstein, D., Shub, M., Expanding endomorphisms of flat manifolds. Topology **7**(2), 139–141 (1968).

[23] Frink, A. H., Distance functions and the metrization problem. Bull. Am. Math. Soc. **43**, 133–142 (1937).

[24] Goetz, A., Piecewise isometries – an emerging area of dynamical systems. http://math.sfsu.edu/goetz/Research/graz/graz.pdf.

[25] Gottschalk, W. H., Hedlund, G. A., Topological Dynamics. Amer. Math. Soc. Colloquium Publications, vol. **36**. Providence (1955).

[26] Hadamard, J., Les surfaces à courbures opposées et leur lignes géodésiques. J. Math. Pures Appl. **4**, 27–73 (1898).

[27] Halmos, P. R., Measure Theory. Van Nostrand, New York (1950).

[28] Hatcher, A., Algebraic Topology. Cambridge University Press, Cambridge (2002).

https://doi.org/10.1515/9783110702682-014

[29] Herrlich, H., Axiom of Choice. Lect. Notes in Math., vol. **1876**. Springer, Berlin Heidelberg (2006).

[30] Hirsch, M. W., Differential Topology. Graduate Texts in Mathematics, vol. **33**. Springer, New York (1976).

[31] Katok, A., Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. Publ. Math. IHÉS **51**, 137–173 (1980).

[32] Katok, A., Interval exchange transformations and some special flows are not mixing. Isr. J. Math. **35**(4), 301–310 (1980).

[33] Katok, A., Hasselblatt, B., Introduction to the Modern Theory of Dynamical Systems. Cambridge University Press, Cambridge (1996).

[34] Keane, M., Interval exchange transformations. Math. Z. **141**, 25–31 (1975).

[35] Kechris, A., Classical Descriptive Set Theory. Graduate Texts in Mathematics. Springer, New York (1995).

[36] Kesseböhmer, M., Stratmann, B. O., Fractal analysis for sets of non-differentiability of Minkowski's question mark function. J. Number Theory **128**, 2663–2686 (2008).

[37] Keynes, H. B., Robertson, J. B., Generators for topological entropy and expansiveness. Math. Syst. Theory **3**, 51–59 (1969).

[38] Kingman, J. F. C., Taylor, S. J., Introduction to Measure and Probability. Cambridge University Press, London (1966).

[39] Kitchens, B., Symbolic Dynamics: One-Sided, Two-Sided and Countable State Markov Shifts. Universitext. Springer, Berlin Heidelberg (1998).

[40] Kotus, J., Urbański, M., Existence of invariant measures for transcendental subexpanding functions. Math. Z. **243**, 25–36 (2003).

[41] Kotus, J., Urbański, M., Ergodic Theory, Geometric Measure Theory, Conformal Measures and the Dynamics of Elliptic Functions, preprint. http://arxiv.org/pdf/2007.13235.pdf.

[42] Krzyżewski, K., Szlenk, W., On invariant measures for expanding differentiable mappings. Stud. Math. **33**, 83–92 (1969).

[43] Lind, D. A., Marcus, B., An Introduction to Symbolic Dynamics and Coding. Cambridge University Press, Cambridge (1995).

[44] Martens, M., The existence of σ-finite invariant measures, Applications to real one-dimensional dynamics. Math. ArXiv, http://front.math.ucdavis.edu/math.DS/9201300.

[45] Masur, H., Interval exchange transformations and measured foliations. Ann. Math. **115**(1), 169–200 (1982).

[46] Mattila, P., Geometry of Sets and Measures in Euclidean Spaces. Cambridge University Press, Cambridge (1995).

[47] Mauldin, R. D., Urbański, M., Graph Directed Markov Systems: Geometry and Dynamics of Limit Sets. Cambridge Tracts in Mathematics, vol. **148**. Cambridge University Press, Cambridge (2003).

[48] Minkowski, H., Geometrie der Zahlen. Gesammelte Abhandlungen, vol. **2** (1911). Reprinted by Chelsea, New York, 43–52 (1967).

[49] Misiurewicz, M., A short proof of the variational principle for a $\mathbb{Z}_+^N$ action on a compact space. Astérisque **40**, 147–187 (1976).

[50] Misiurewicz, M., Szlenk, W., Entropy of piecewise monotone mappings. Stud. Math. **67**(1), 45–63 (1980).

[51] Morse, M., Hedlund, G. A., Symbolic Dynamics. Am. J. Math. **60**, 815–816 (1938).

[52] von Neumann, J., Proof of the quasi-ergodic hypothesis. Proc. Natl. Acad. Sci. USA **18**(1), 70–82 (1932).

[53] Nicol, M., Petersen, K., Ergodic Theory: Basic Examples and Constructions. Encyclopedia of Complexity and Systems Science. Springer (2009).

[54] Oxtoby, J. C., Measure and Category. Springer, New York (1980).

[55] Parthasarathy, K. R., Probability Measures on Metric Spaces. Academic Press, New York (1967).
[56] Pesin, Y. B., Dimension Theory in Dynamical Systems. Chicago Lectures in Mathematics. University of Chicago Press, Chicago (1997).
[57] Przytycki, F., Urbański, M., Conformal Fractals: Ergodic Theory Methods. London Mathematical Society Lecture Note Series, vol. **371**. Cambridge University Press (2010).
[58] Rudin, W., Real and Complex Analysis, 3rd edn. McGraw-Hill, Singapore (1987).
[59] Ruelle, D., Statistical Mechanics. Rigorous Results. Benjamin, New York (1969).
[60] Ruelle, D., Statistical mechanics on a compact set with $\mathbb{Z}^\nu$ action satisfying expansiveness and specification. Bull. Am. Math. Soc. **78**, 988–991 (1972). Transactions AMS **185**, 237–251 (1973).
[61] Ruelle, D., Thermodynamic formalism. Encyclopedia of Math. and Appl., vol. **5**. Addison-Wesley, Reading Mass (1976).
[62] Salem, R., On some singular monotonic functions which are strictly increasing. Trans. Am. Math. Soc. **53**(3), 427–439 (1943).
[63] Sears, M., Expansiveness on locally compact spaces. Math. Syst. Theory **7**(4), 377–382 (1973).
[64] Shannon, C., A Mathematical Theory of Communication. Bell Syst. Tech. J. **27**, 379–423 & 623–656 (1948).
[65] Shannon, C., The Mathematical Theory of Communication. University of Illinois Press, Urbana, Illinois (1949).
[66] Shub, M., Endomorphisms of compact differentiable manifolds. Am. J. Math. **91**(1), 175–199 (1969).
[67] Sinai, Ya. G., On the Notion of Entropy of a Dynamical System. Dokl. Russ. Acad. Sci. **124**, 768–771 (1959).
[68] Sinai, Ya. G., Gibbs Measures in Ergodic Theory. Russ. Math. Surv. **27**, 21–70 (1972).
[69] Stroppel, M., Locally Compact Groups. European Mathematical Society (2006).
[70] Sumi, H., Urbański, M., Measures and dimensions of Julia sets of semi-hyperbolic rational semigroups. Discrete Contin. Dyn. Syst., Ser. A **30**, 313–363 (2011).
[71] Szlenk, W., An Introduction to the Theory of Smooth Dynamical Systems. PWN/Wiley, Warszawa/New York (1984).
[72] Taylor, J. C., An Introduction to Measure and Probability. Springer, New York (1997).
[73] Utz, W. R., Unstable homeomorphisms. Proc. Am. Math. Soc. **1**, 769–774 (1950).
[74] Veech, W. A., Gauss measures for transformations on the space of interval exchange maps. Ann. Math. **115**(1), 201–242 (1982).
[75] Walters, P., A Variational Principle for the Pressure of Continuous Transformations. Am. J. Math. **97**, 937–971 (1975).
[76] Walters, P., An Introduction to Ergodic Theory. Graduate Text in Mathematics. Springer (2000).
[77] Willard, S., General Topology. Addison-Wesley, Mass. (1970).
[78] Williams, R. F., A note on unstable homeomorphisms. Proc. Am. Math. Soc. **6**, 308–309 (1955).
[79] Yomdin, Y., Volume growth and entropy. Isr. J. Math. **57**, 285–300 (1987).
[80] Zinsmeister, M., Thermodynamic Formalism and Holomorphic Dynamical Systems. SMF/AMS Texts and Monographs, vol. **2**. American Mathematical Society (1999).

Index

De Gruyter Expositions in Mathematics

Volume 68
Yuri A. Bahturin
Identical Relations in Lie Algebras, 2021
ISBN 978-3-11-056557-7, e-ISBN 978-3-11-056665-9

Volume 34
Marek Jarnicki, Peter Pflug
Extension of Holomorphic Functions, 2nd Edition, 2020
ISBN 978-3-11-062766-4, e-ISBN 978-3-11-063027-5

Volume 67
Alcides Lins Neto, Bruno Scárdua
Complex Algebraic Foliations, 2020
ISBN 978-3-11-060107-7, e-ISBN 978-3-11-060205-0

Volume 66
Anke Kalauch, Onno van Gaans
Pre-Riesz Spaces, 2018
ISBN 978-3-11-047539-5, e-ISBN 978-3-11-047629-3

Volume 65
Yakov Berkovich, Zvonimir Janko
Groups of Prime Power Order. Volume 6, 2018
ISBN 978-3-11-053097-1, e-ISBN 978-3-11-053314-9

Volume 64
Yakov Berkovich, Lev S. Kazarin, Emmanuel M. Zhmud’
Characters of Finite Groups. Volume 2, 2019
ISBN 978-3-11-022408-5, e-ISBN 978-3-11-022409-2

Volume 63
Yakov Berkovich, Lev S. Kazarin, Emmanuel M. Zhmud’
Characters of Finite Groups. Volume 1, 2017
ISBN 978-3-11-022406-1, e-ISBN 978-3-11-022407-8

Volume 62
Yakov Berkovich, Zvonimir Janko
Groups of Prime Power Order. Volume 5, 2016
ISBN 978-3-11-029534-4, e-ISBN 978-3-11-029535-1

www.degruyter.com