Selected Topics

on

ANTENNA SYNTHESIS

The Unity of

Mathematical Formalisms

and

Technical Sense

Yuri Choni

Selected Topics on Antenna Synthesis

Selected Topics on Antenna Synthesis:

The Unity of Mathematical Formalisms and Technical Sense

By

Yuri Choni

Cambridge
Scholars
Publishing

Selected Topics on Antenna Synthesis:
The Unity of Mathematical Formalisms and Technical Sense

By Yuri Choni

This book first published 2021

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2021 by Yuri Choni

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-5275-6072-4
ISBN (13): 978-1-5275-6072-7

To the memory of my supervisor and colleague Vadim Nikolaevich Dymsky, who encouraged me to believe in the harmony of electromagnetic laws and strive for the unity of mathematical formalism and its physical sense.

"Students should not learn thoughts—they should learn to think"
Immanuel Kant (1724–1804)

TABLE OF CONTENTS

ILLUSTRATIONS

Figures

Tables

PREFACE

I firmly believe that the core quality of any technician that makes them successful in the profession is a special style of mind, an engineering style of thinking. The features of this book stemmed from the fact that the author set himself the goal of not only providing knowledge on the designated issue, but also, perhaps mainly, honing the engineering style of thinking. Antenna synthesis provides an excellent opportunity to achieve this goal, since it differs from many other technical disciplines in the organic unity of complex mathematical abstractions and physical manifestations, which is a vital aspect of the engineering style of thinking. The photo in Fig. 3-3 can serve as a visual confirmation of this fact. It shows a setup that implements such a mathematical abstraction as an adjoint operator in the form of an electromagnetic field. Amazing! Isn't it?

The mentioned features are as follows.

- First, the emphasis is not on the results of the synthesis, but on methodological aspects: how to formulate and solve the corresponding problems. At the same time, physical meaning and compliance with practical needs are the focus of attention. For example, in the synthesis of a given amplitude antenna radiation pattern (ARP), the deviation of ARPs in the form, taking into account random errors of reproduction of the nominal solution, need to be considered.

- Second, it is best to teach creativity with your own ideas and developments. That is why the book is not an overview of the works in the field of antenna synthesis, so its bibliography contains only 49 references. The subject matter of the book is what the author did.

- Third, Chapter 1 is devoted to the basics of functional analysis, without which it is hardly possible to understand the essence of antenna synthesis. This material is of broader interest, as is the Introduction, which discusses what an engineering style of thinking is.

Who is this book for? Naturally, it is primarily for those who are busy with or interested in the development of antennas, since antenna synthesis is the theoretical basis of such activities.

In addition, due to its focus on the formation of technical thinking, it is useful to students, graduate students and young engineers of many specialties and first radio communications.

As for the style of the book, at the end of the chapters there are training tasks and Mathcad programs in the Appendices, which allow the interested reader to think over or conduct their own research on the issues posed.

In some places, the author's comments interrupt the text. Smaller fonts and an enlarged left margin highlight them so you can skip them.

Graphic visualization of everything is one of the facets of technical thinking. It is therefore not surprising that the book has so many figures.

Acknowledgments

The author considers it his pleasant duty to express his gratitude to his colleagues for many years of communication and cooperation, thanks to which his professional philosophy has become what it is.

INTRODUCTION:
ENGINEERING STYLE OF THINKING

The field of activity associated with techniques and technologies is extremely diverse and extensive. Nevertheless, some universal psychological qualities largely predetermine the ability to assimilate the relevant knowledge, to develop necessary skills, and success in the profession. A special mindset, which could be called an engineering way of thinking, is the most important, I think.

The list that indexes different facets of thinking (thinking styles, methods and ways of thinking) consists of more than a hundred items. Not surprisingly, there are several scales used by psychologists to classify styles of thinking in accordance with their penchant for philosophy, sociology and management [1, 2], or psychology [3].

Along with Kirton's adaptive vs. innovative cognitive styles [3], V. Gulenko defined four styles of thinking from another point of view [4]: causal, dialectic-algorithmic, holographic and vortex. Note that psychologists understand style of thinking merely as universal cognitive techniques without its relation to the subject of a thought. The engineering style of thinking, as I understand it, is much narrower and refers to the techniques of comprehending the world of technique and technologies. It relates closely to the criteria of competence, which is a very important measure of purpose and quality of training in the high school [5, 6].

Of what the engineering style of thinking consists

The following mental qualities and skills form the engineering style of thinking:

Firstly, curiosity and an overwhelming desire to understand the essence of any natural phenomena and the laws of the technical world. Of course, human curiosity does not restrict itself by the field of technique. Once during the spring I thought, "The spring sap will start moving. I wonder how the sap reaches the foliage at an altitude of 15 or 20 meters. What makes it rise? What pushes it or pulls it?" My reader, try and answer this. If you think of pressure, it should be of unbelievable values of 1.5 atm. Think of something else.

Secondly, a strong predilection for the quantitative analysis of physical processes and things. Three skills are of importance here: 1) to select measurable characteristics and parameters of the essential manifestations; 2) to ignore irrelevant factors; 3) to identify quantitative relationships and laws that govern the world of technology.

Thirdly, a proper arsenal of basic mathematical methods and techniques needed to solve a wide range of tasks in the relevant technical area.

Fourthly, the ability to perceive the physical nature of analyzed processes, hidden in mathematical formalism. This not only gives a meaningful interpretation of the obtained results, but also can often speed up calculations through the use of heuristic assumptions and reasonable simplifications. Besides, very often it happens that among strict mathematical solutions, just a few comply with the terms of technical feasibility or other practical limitations, and we have to separate them from the formal solutions. To do this we must understand the physical nature of the processes and quantitative laws reflected by mathematical expressions.

In Fig. 1, the diagram illustrates the role of thinking in an engineer's activities. Just like any chart of a socio-psychological nature, it does not require detailed comments. We note only that there are two main functions of engineering thinking, as a summand of engineer competencies. It is the capability to plunge easily from the real physical world into the virtual world of mathematical descriptions (mathematical formalism), and then just as easily make the reverse transition to the physical reality, to the interpretation and understanding of the formal results.
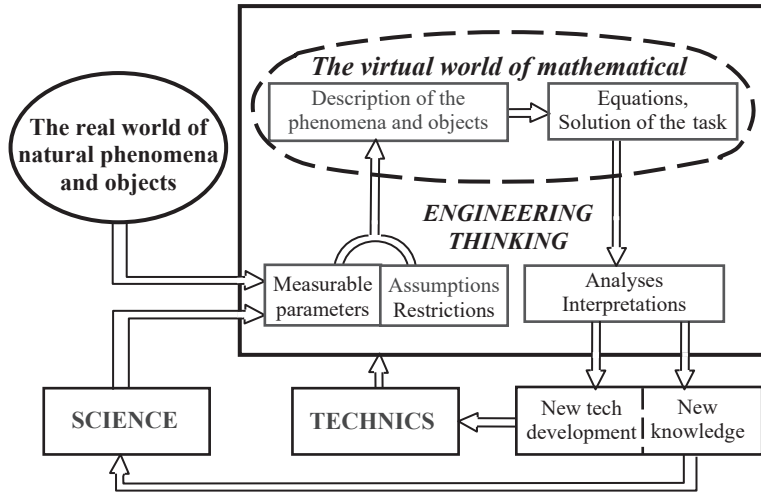
Figure 1: Activities of an engineer

It is impossible and hardly necessary to devise several (say, ten) specialized lessons, the result of which the engineering way of thinking would have been acquired by all the students. This is the aim of the whole process of technical training in its entirety.

Education, in general, is a product of the interaction of three subjects: the Authority's manager, the teacher and the student (Fig. 2).
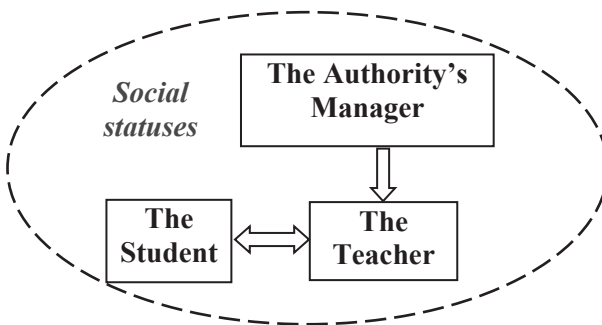
Figure 2: Three subjects of the educational process

Certainly, educational standards and curriculums (partly the product of officials' efforts) are necessary and important. However, they have a very distant relation towards the formation of an engineering style of thinking.

How clever and ambitious (or not) school graduates are, those who go to universities is largely dependent on social prestige of the profession they have chosen. While prestige is a moral issue, in relation to this topic it has a very clear quantitative measure: the income that an average employee receives in the profession.

Fig. 2 presents the subjects of educational processes along with social conditions that, catalyst-like, "do not participate in the reaction, but can greatly speed up or slow down it." Always and in all circumstances, the professionalism of teachers is the most essential thing concerning the quality of education. Over many years of teaching at the Electronics and Telecommunications Institute (courses "Antennas and microwave devices," "Electrodynamics" and others) pedagogical methods were worked out, that can be hopefully useful at teaching any technical discipline.

Psychological issues

1. It is very helpful to keep in mind that terms, definitions, physical and functional characteristics are as vital for any technical discipline, as the structural elements (bricks, beams, logs, etc.) for the building. That is why the question "What is it?" is pronounced so often in exams. Draw the students' attention to a pleasant and inspiring fact that each term and each definition contains a small (easily exhaustible) amount of information. No matter how paradoxical it may sound, any student can and should know as clearly as professors or academicians what the spectrum of a signal, a band pass filter, the electric field vector, the antenna gain etc. are.

2. A good student or a qualified engineer can make the right judgment whether he knew what was discussed, or not. So, sometimes he answers, "I do not know." An unsuccessful student or engineer, as a rule, has no certainty on this matter. Typically, this is due to unsystematic learning—attempts to grasp the current topic without mastering previous often-simpler themes.

A good piece of advice to such people is the following. Develop the habit, when faced with difficulties, of asking yourself two questions: a) whether the terms of the studied text are clear, b) whether what it is about is clear (not in detail, but in general). If necessary, leaf through the textbook again. In critical situations, seek someone's help. In addition, of course, train your inner voice: "I know it," or "I don't understand it." Such a strategy will not only result in successful learning but also will reduce the time consumption.

3. For so many students and engineers, a formula is a fetish: it is impossible to derive it and to find it in a textbook is the only way. There is a need to convince them that not even all the books together contain the answers to all situations. A strategy that is much more practical is to understand the general principles and laws and remember a small number of the corresponding formulas and relations. This will make it possible to get a formula for a particular case, without wasting time on an unsuccessful search.

4. Some do not know how to control the calculations, find errors and assess the acceptability of results. There are standard techniques that help to cope with these limitations:

First, dimensional analysis is elementary and surprisingly effective. If the obtained formula results in calculating the sine of 3 kg or the logarithm of 1 mm or something like that, do not do any calculations and look for errors. If this formula is from a textbook, it is likely you do not correctly perceive the notation of the variables.

Second, it is useful to check formulas at special values of their variables, for which you know what it should be: in case of a short circuit or discontinuity, at zero or infinitely large value, etc.

5. A very useful habit is to not rush into solving a new task, but try to understand it as a whole, find out its general structure or its features in order to pave the best way to a solution. I call this tactic a "shell approach."

Nature and its description

A feature of the engineering profession as the brainchild of Applied Physics and Applied Mathematics is that it deals with something not completely known. For example, electrodynamics, based on paradigm of the macroscopic, explores notion of electromagnetic field (EMF). It elucidates the relevant lows that create the solid foundations for developing antennas, high-frequency devices, radio systems, and many other applications. Although a comprehensive answer to the question "What is EMF itself?" still awaited. Perception of the real complex and multifaceted world of nature through its manifestations is the fundamental principle of engineering thinking. Regularities of these manifestations established as strict quantitative relationships pave the way for the use of physical phenomena in technology.

The transition from the world of reality to the world of quantitative descriptions of its manifestation begins with an introduction of significant measurable characteristics of the considered phenomenon, structure or

device. To the engineer, it is useful to bear in mind that the choice of these characteristics is not strictly predetermined. They are largely the result of professionals' conventions.

For example, apart from in our perception, the electric field vector, the "famous" vector $\mathbf{E}$, does not exist. What does exist is an ability of EMF to exert a force $\mathbf{F}$ on a particle of matter, which is arranged in the given point in space and has a charge q. Coulomb's law has proven useful in characterizing this property of EMF by vector $\mathbf{E}=\mathbf{F}$/q. Similarly, Ampere's law regarding the force interaction of electric currents had led to an agreement to characterize the magnetic field associated with the currents by vector $\mathbf{H}$. The pair of vectors $\mathbf{E}$ and $\mathbf{H}$ so fully characterize the properties of EMF, that specialists perceive these vectors as EMF itself.

Such identification we use very often. For instance, we say, "transmit energy," although the energy does not exist as a material something. After all, it is a property of a matter, or a body, or a field, and apart from its carrier does not exist. The same situation concerns information: it is customary to speak of it, as if it is something tangible and exists in itself. We must understand the conventionality of such expressions used for brevity.

Creativity and ingenuity

To exercise the creativity and ingenuity inherent to engineering thinking, unpretentious crafty tasks are useful. For the sake of concreteness, here are two examples of such tasks, which do not require the knowledge beyond the high school physics. I hope that they will please inquisitive readers, who, I hope, will try to answer the tasks before reading the paragraphs "***If you still have not coped with the task**, read the answer*."

Cryo contr reflector (D = 0.5 m)   Main reflector (D = 12 m, F = 2.8 m)

External heat shield
Internal heat shield I
Internal heat shield II
Deployable truss

Flat mirror
Cryo container
D = 3 m, h = 1 m

Radio & laser TM antennas

Service module
D = 3.5 m, h = 1.5 m
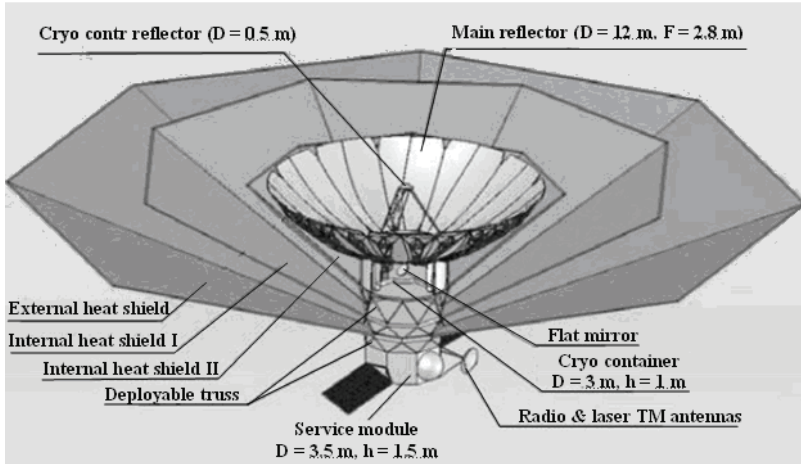
Figure 3: Layout of the 12 m space telescope (a picture from [7])

The first of them I thought up under the impression of the Russia-European project named "Millimetron." The plan is to launch a large space telescope for exploring the universe [7]. Fig. 3 depicts its overall view, which is so grandiose that I cannot resist showing it. Among others, there are two amazing moments. 1) To achieve the desired angular resolution and needed mirror efficiency the overall surface accuracy should be 10 microns for the whole 12 meter dish surface! 2) To detect the relict radiation (verifying the theory of primary blast is one of the many aims of the project), the required sensitivity is so high that no super receiver can achieve it. The only option is photon counters [8]!

The last fact provoked the idea of the following **task A**.

Task A

Conditions:

Let us assume that a signal of a feeble cell phone's transmitter of the power $P = 5$ mW, is radiated uniformly in all directions of a full solid angle. The carrier frequency of the signal is $\nu = 1$ GHz.

What to do:

Estimate the distance R from the cellular phone, which can serve as a boundary between two areas: one where the phone field exists in the form of a continuous electromagnetic wave (EMW) and the other where it become a discrete flux of photons (Fig. 4).
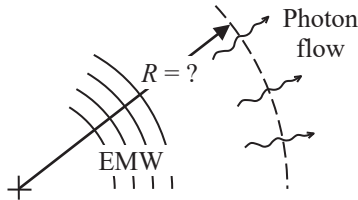
Figure 4: The subject of Task A

Some tips:

1) Note that EMF can be something continuous in space (for example, EMW) only under the condition that the density of photons—elementary particles of the field—is extremely high.
2) Photon energy E is determined by Planck's relation $E = h\nu$, where $h = 6.626 \; 10^{-34}$ [J s] is the Planck constant and ν is the frequency.
3) EMF propagates in free space in the form of a spherical wave, therefore the density of flaw of electromagnetic energy subsidizes along the distance r in the proportion of $1/r^2$.

Discussion:

Of course, this task has no exact single-valued answer, because the condition that determines the boundary between the electromagnetic wave and the "photon discrete flux" is uncertain—there is no distinct determination. In fact, in this lies the usefulness of the task—a reader must make his own independent decision based on common sense. It is very typical for engineers to need to make a decision in the situation of essential uncertainty.

If you still have not coped with the task, *read the answer:*

As to the task, assume that we have to consider EMF as a discrete flux of photons if less then $N_{ph} = 10^5$ photons fly through a surface area dS of 1 mm^2 during a time span dt of 1c. Then the obvious equation $P \, dt \, dS / (4\pi R^2) = N_{ph} \, h \, \nu$ takes place, which results in the estimation $R \approx 77.5$ km.

Task B

Conditions:

A tire of a rear wheel of a sports car became flat (Fig. 5a). There was no spare wheel in the car. A car garage was 5 km away along a highway. The driver became very distressed, since the tire would inevitably come into disrepair while driving or towing to the garage. Fortunately, he was a good engineer before becoming a professional driver. Therefore, clever thoughts came to his mind (Fig. 5b). He closed his eyes, moved his lips, shouted "Eureka!" and pressed the accelerator pedal down.

a)

b)   c)
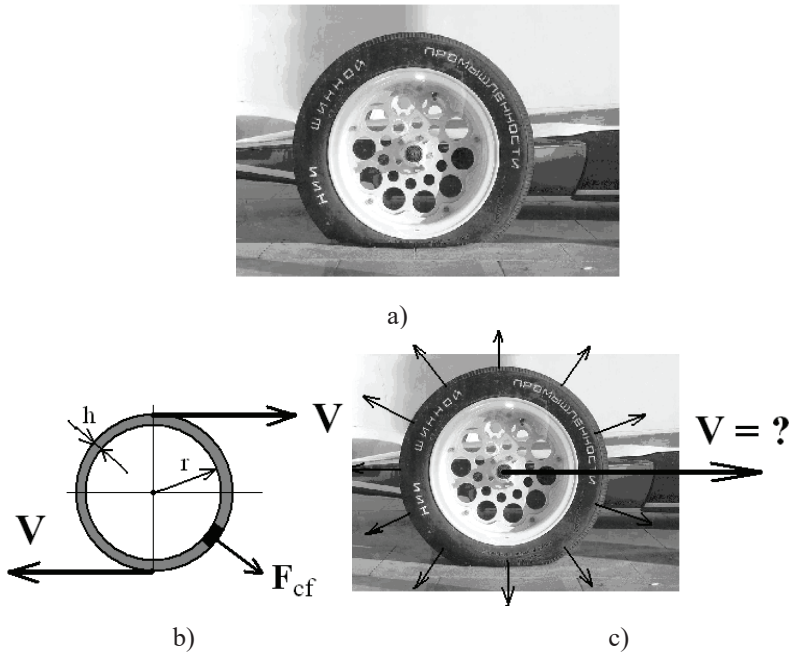
Figure 5: A flat tire in three situations: a) a standing car; b) the driver's thinking; c) a moving car

What to do:

Estimate the speed V at which the driver had to drive in order to save the tire (Fig. 5c).

Some tips:
1) Note that the centrifugal force can essentially substitute the force of air pressure expanding the tire in the normal situation.
2) Neglecting the sidewalls, we can consider a tire as a section of cylindrical tube (Fig. 5b).

Discussion:

This task, like the previous one, has no exact quantitative answer. However, we deal with assessing. Therefore, physical dependences and common sense fight uncertainty and pave the way to a meaningful estimation. Indeed, the pressure of a pumped tire P_0 [Pa] gives rise to the force $F_0 = P_0 \Delta S$ [N] on area ΔS of the tread surface. On the other hand, the centrifugal force F_{cf} introduced by the rotation of a tire due to moving the car with a speed V is $F_{cf} = \rho \, (h \, \Delta S) \, V^2 / r$ where ρ, h and r are the density, the depth and the radius of the tire, respectively. The desired estimation results from equation $F_{cf} = F_0$.

If you still have not coped with the task, *read the answer:*

Let us assume that the pressure P_0 is about 2 atm, the wheel radius r is 25 cm, tire thickness h is 25mm, and the tire density ρ exceeds the density of water due to steel cords at $1.1 \; 10^3$ kg/m³. After transforming all the values to the uniform units, the estimation $V \approx 154.5$ km/h comes in force.

ABBREVIATIONS

AA	antenna array
AAA	adaptive antenna array
APhD	amplitude and phase distribution
ARP	antenna radiation pattern
CCPW	converging cluster of plane wavs
EMF	electromagnetic field
EMW	electromagnetic wave
PhF	phase front
LPhC	local phase center
LPF	low-pass filter
MMRA	modified minimal residual algorithm
MRA	minimal residual algorithm
PhC	phase center
PEMW	plane electromagnetic wave
RMSD	root-mean-square deviation
WLM	Woodward-Lawson method
WV	weighting vector

$k = 2\pi/\lambda$	wave number in free space
[...]	dimension of physical value

$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$ is the so-called Kronecker delta

$\text{sinc}(x) = \sin(x) / x$ is the so-called sinc-function

CHAPTER ONE

SOME FUNDAMENTALS
OF FUNCTIONAL ANALYSIS

1.1. Introduction

To master the topics of the following chapters with understanding, the knowledge of some fundamentals of functional analysis is necessary. The corresponding issues, as a rule, lie on the periphery of such university courses as "Special sections of mathematics" or "Methods of computational mathematics." My teaching experience shows that many students do not know that. However, if you are an advanced reader, you may skip this chapter. Nevertheless, in my opinion, Section 1.3 is worth studying or, at least, reading. It outlines the ideology and effective algorithms used to analyze, synthesize and design not only antennas but also many other radio and electronics devices.

I got acquainted with functional analysis [9, 10], a fascinating branch of mathematics, during my postgraduate studies, when I was working on a dissertation devoted to the problems of antenna synthesis. Since then, I have never ceased to admire how exquisitely and elegantly mathematicians are able to comprehend the fundamental structural concepts of the real world and to convey them in a generalized form, naturally, to the world of mathematical abstractions. It is impossible not to admire the effectiveness of methods and algorithmic tools generated from these abstractions to solve various tasks. In this there is something very much in common with the engineering style of thinking, more precisely its facet, connected with the transition from reality to quantitative descriptions reflecting the main manifestations of physical phenomena or a technical object.

Therefore, I hope that this chapter will provide the reader with ideas about universal means of obtaining the best approximation to the desired functional characteristics of the devices designed. At the same time, it will contribute to the formation of a certain philosophical framework that

serves as reliable support and a compass in dealing with complex technical problems.

Of course, only the basic issues of functional analysis will be discussed here without in-depth nuances, but with an emphasis on aspects that are important for the synthesis of antennas.

1.2. Metric and space of functions

Let us think about "What is space?" Try to define it not from the position of its physical essence (whether it is emptiness, whether a special kind of matter called ether), but in a purely ordinary sense as a universal container (without boundaries) for all that exists. By the way, any material object, existing in space, has spatial properties itself. Back to the question, what sense is behind our intuitive feel of space, the space in which we live? The question sounds very simple, but it is not easy to find a meaningful answer for it.

Mathematics, with exceptional elegance, answers this question: "the space X is the set of points $\{x\}$, to each pair of which there is associated a number, called the distance $\rho(x_i, x_j)$ or the metric[1] of space." The distance/metric, more precisely the rule of its definition, must satisfy the following axioms: for arbitrary points x_i, x_j, x_k in X

- the distance is a positive real value, i.e. $\rho(x_i, x_j) \geq 0$;
- if $x_i = x_j$, then $\rho(x_i, x_j) = 0$, and vice versa (!) (the axiom of identity);
- $\rho(x_i, x_j) = \rho(x_j, x_i)$ (the axiom of symmetry);
- $\rho(x_i, x_j) \leq \rho(x_i, x_k) + \rho(x_k, x_j)$ (the axiom of triangle inequality).

The first axiom (the distance cannot be negative) follows from the remaining axioms. Therefore, strictly speaking, it can be omitted, and mathematicians do this [10]. However, for us engineers, it is worth mentioning this property explicitly because it expresses an important fragment of intuitive human notions of distance in space. In some mathematical books, it is also present.

It is obvious that these axioms are not the invention of mathematicians. It is the formalization of our intuitive perception of space and distance in it. In particular, the triangle inequality corresponds to the fact that the path from x_i to x_j is shorter or at least not longer than the path first traversed from x_i to x_k, and then from x_k to x_j. After intuition transformed into clear mathematical relations, huge possibilities arise for the use of the concept

---

[1] If the elements $\{x\}$ of the set X are functions or objects of nature other than ordinary (Euclidean) points, then the term "distance" is not linguistically right. Therefore, the term "metric" is preferable, although both have the same meaning.

of distance as a quantitative measure of the closeness of elements of any nature, in particular functions.

Consider the set $\mathcal{F}$ of functions $f(x)$, referring to a particular function $f_i(x)$ and treating it as an abstract point of $\mathcal{F}$. If we define a rule according to which the quantity $\rho(f_i, f_j)$ satisfying the above axioms is determined (computed), then the concept of the distance between points in the set $\mathcal{F}$ (the distance between functions!) is introduced, and the set turns into a functional space—a space of functions. Since linguistically the phrase "distance between two functions" sounds strange, it is customary to use the more abstract term "metric" as applied to the functional space.

Among the vast variety of possible metrics and the functional spaces generated by them, two metrics are most often used. The first is a metric defined by the equality:

$$\rho_{C_2}(f_1(x), f_2(x)) = \sqrt{\int_a^b \left| f_1(x) - f_2(x) \right|^2} \tag{1-1}$$

which means the root-mean-square deviation (RMSD) of the functions. The relevant space is commonly denoted as C_2.

The second is a metric defined as follows:

$$\rho_L(f_1(x), f_2(x)) = \max_{a \le x \le b} \left| f_1(x) - f_2(x) \right| \tag{1-2}$$

that measures the value of the maximum deviation of the functions in the interval (a, b) where they exist. This metric name is "linear," and L usually denotes the functional space that it generates.

Figure 1-1 illustrates the difference in how the functions approach the same desired function $f_0(x)$ if the C_2 or L metric is used. Taking into account that the metric Ex. 1-1 stands for RMSD, we can see that the value $\rho_{C_2}(f_1(x), f_0(x))$ is remarkably bigger than $\rho_{C_2}(f_2(x), f_0(x))$, therefore, the function $f_2(x)$ approaches $f_0(x)$ better than $f_1(x)$. The metric Ex. 1-2 controls local deviations and therefore, contrary to the previous situation, the value $\rho_L(f_1(x), f_0(x))$ is smaller than $\rho_L(f_2(x), f_0(x))$, and the function $f_1(x)$ approaches to $f_0(x)$ nearer than the function $f_2(x)$.
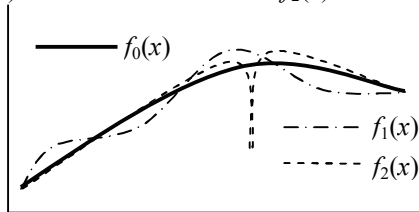
Figure 1-1: Functions $f_1(x)$ and $f_2(x)$ approaching the function $f_0(x)$

1.3. Hilbert space of functions

1.3.1. Introductory note

Another, perhaps even more impressive, example of the extremely sophisticated generalization of the notion of our Euclidean space to abstract mathematical spaces is the notion of orthogonality, in particular, the orthogonality of functions.

In common Euclidean space, the space in which we live, each point x can be determined by the radius vector $\mathbf{x}$—a line segment directed from the coordinate system's origin to the point x. The angle between the vectors and, in particular, their perpendicularity represents an important property of Euclidean space. To represent an arbitrary vector $\mathbf{x}$ as the sum of the basis vectors $\{\xi_n\}$: $\mathbf{x} = \sum_n a_n \xi_n$ it is extremely convenient to use a system of mutually perpendicular unit vectors $\{\xi_n\}$ forming the so-called orthonormal basis. The fact is that in this case each of the coefficients a_n of the expansion is determined independently of the others, as the projection of the vector $\mathbf{x}$ onto the basis vector ξ_n: $a_n = (\mathbf{x}, \xi_n)$. If the basis vectors are not perpendicular to each other, then the procedure for finding the coefficients a_n becomes much more complicated and, as a result, requires solving a system of algebraic equations.

Naturally, for function spaces the expansion of an arbitrary function $f(x)$ over the system of basis functions $\{g_n(x)\}$ of $f(x) = \sum_n a_n g_n(x)$ plays the same important role as for the vector space. It is tempting that the basis functions have the same property of mutual "perpendicularity." However, how do you define this property for functions? Can we introduce the notion of "angle" between functions and the property of their "perpendicularity"? Is it possible to introduce something analogous to the notion of "angle" between functions or the property of their "perpendicularity"? This problem is definitely more complicated than the notion of the distance between functions. Indeed, perceiving distance as a quantitative measure of the deviation of functions from each other, we can easily guess we should use the standard deviation as the distance between them. To introduce the concept of "perpendicularity" of functions, a much more sophisticated idea is required. To solve this problem, mathematicians generalize the notion of an inner product.

1.3.2. The cosine of the "angle" between functions. Orthogonality of functions

Each student knows that the scalar (or inner) product of two Euclidean vectors is the product of their lengths/sizes by the cosine of the angle α between them: $(\xi_1, \xi_2) = |\xi_1||\xi_2|\cos(\alpha)$, where ξ_1 and ξ_2 are any vectors. From the geometric meaning of a inner product, a number of properties follow, the main of which are as follows. **1.** If one of the vectors is a sum of vectors, then the inner product is equal to the sum of the inner products with each summand (property of linearity). **2.** If the vectors ξ_1 and ξ_2 are the same ($\xi_1 = \xi_2 = \xi$), then the inner product is not a negative quantity equal to the square of the length of the vector: $(\xi, \xi) = |\xi|^2$ (the property of positive definiteness), and the equality $(\xi, \xi) = 0$ is satisfied only in the case of the zero vector $\xi \equiv \varnothing$. **3.** In addition to that, it is clear that $|(\xi_1, \xi_2)| \leq |\xi_1||\xi_2|$, since $\cos(\alpha) \leq 1$. The perpendicularity of the non-zero vectors ξ_1 and ξ_2 leads to the fact that their inner product is zero, since $\cos(\alpha) = 0$ for $\alpha = 90°$.

Reserving these semantic properties while defining a notion of inner product for elements of an arbitrary nature, in particular for the space $\mathcal{F}$ of complex-valued functions, mathematics defines it as follows. This is a complex number associating to each pair of functions $f_1 = f_1(x)$ and $f_2 = f_2(x)$ and denoted by (f_1, f_2), for which the following conditions hold:

- for any three elements f_1, f_2 and f of $\mathcal{F}$ and any numbers α and β, the equality $(\alpha f_1 + \beta f_2, f) = \alpha (f_1, f) + \beta (f_2, f)$ is true (the linearity of the inner product with respect to the first argument);
- for any elements f_1 and f_2 the following equality holds $(f_1, f_2) = (f_2, f_1)^*$, where the asterisk * denotes complex conjugation;
- for any element f we have $(f, f) \geq 0$, where $(f, f) = 0$ only for the zero-element $f \equiv \varnothing$ (positive definiteness of the inner product).

The quantity $\|f\| = \sqrt{(f, f)}$, called the norm of the element f (in particular, a function), plays the same role as the length of the vector. The norm of the difference of two elements satisfies all the axioms of the metric and thereby generates the corresponding space in which the distance is as follows $\rho(f_1, f_2) = \|f_1 - f_2\| = \sqrt{(f_1 - f_2, f_1 - f_2)}$. The well-known and important Cauchy–Bunyakovsky–Schwarz inequality states that for all elements f_1 and f_2 of a Hilbert space it is true that

$$|(f_1, f_2)| \leq \|f_1\| \|f_2\|. \tag{1-3}$$

The same inequality for Euclidean vectors means triviality: the cosine of the angle between the vectors cannot be greater than one.

Non-zero functions f_1 and f_2, whose inner product is equal to zero $|(f_1, f_2)| = 0$, for linguistic reasons are called not "perpendicular," but orthogonal. It is clear that these terms are so absolute synonyms that, even with respect to vectors, the term "orthogonal" often replaces the term "perpendicular." For example, they call a system, which unit vectors are mutually perpendicular, exclusively as an <u>orthogonal</u> coordinate system.

There are different ways to define the inner product. For functions $f(x)$ that are square integrable on the interval (a, b), the following expression

$$(f_1(x), f_2(x)) = \int_a^b f_1(x) f_2^*(x) \, dx \qquad (1\text{-}4)$$

or

$$(f_1(x), f_2(x)) = \int_a^b f_1(x) f_2^*(x) \, \rho(x) \, dx \qquad (1\text{-}4')$$

most often defines it. Here, $\rho(x)$ is a real positive function reflecting the weight or importance of the different parts of the interval (a, b). It is apparent that the metric generated by the inner product Ex. 1-3, i.e.

$$\rho(f_1(x), f_2(x)) = \sqrt{(f_1(x) - f_2(x), f_1(x) - f_2(x))}, \qquad (1\text{-}5)$$

is equal to the metric Ex. 1-1 because of trivial $f(x) f^*(x) = |f(x)|^2$.

In the case of a vector space, Cartesian, cylindrical and spherical coordinate systems are of great interest, since their unit vectors install a set of mutually perpendicular vectors $\{\xi_j\}$ ($j = 1, \ldots, N$) that satisfy the condition $(\xi_i, \xi_j) = \delta_{ij}$ where $\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$ is so-called Kronecker delta. In this case, for an arbitrary vector $\mathbf{A}$, the coefficients a_i of its decomposition $\mathbf{A} = \sum a_i \xi_i$ in the basis vectors $\{\xi_i\}$ have the explicit expression and therefore are calculated very simply: $a_i = (\mathbf{A}, \xi_i)$.

In the case of an N-dimensional functional space, the orthonormal basis formed by the set N mutually orthogonal normalized functions[2] $\{g_j(x)\}$ plays the same role. If Ex. 1-4 defines the inner product in the space under consideration and if for any pair of indices i and j the functions $\{g_j(x)\}$ satisfy the condition $\int_a^b g_i(x) g_j^*(x) \, dx = \delta_{ij}$, then any arbitrary function $f(x)$ in this space has the following representation

$$f(x) = \sum_j a_j g_j(x) = \sum_j (f(x), g_j(x)) g_j(x), \qquad (1\text{-}6)$$

the coefficients a_j of which are defined explicitly as $a_j = \int_a^b f(x) g_j^*(x) \, dx$.

---

[2] That is, functions satisfying the analogical condition $(g_i(x), g_j(x)) = \delta_{ij}$.

If the basis functions $\{g_j(x)\}$ are mutually orthogonal, but differ in norms, then Ex. 1-6 turns into the slightly different formula

$$f(x) = \sum_j \frac{(f(x), g_j(x))}{\| g_j(x) \|^2} g_j(x) \,. \tag{1-7}$$

The space of functions with the notion of an inner product introduced for its elements and the metric generated by it is a so-called Hilbert space. The Hilbert space possesses all the properties of the Euclidean space of vectors. Therefore, it is very useful and meaningful to use vectors as a graphical visualization of functions in Hilbert space.

1.3.3. Schmidt orthogonalization procedure

An arbitrary set of N linearly independent functions $\{f_n(x)\}$ can be transformed to an equivalent system of N mutually orthogonal functions $\left\{ f_n^{\perp}(x) \right\}$. Equivalence is understood in the sense that any function $f(x)$ that is a series $f(x) = \sum_n a_n f_n(x)$ can be represented by the series $f(x) = \sum_n b_n f_n^{\perp}(x)$. The corresponding procedure is the so-called Gram-Schmidt process. That is because from any linearly independent functions $f_1(x)$ and $f_2(x)$ we can construct a function $f_2^{\perp}(x) = f_2(x) - \alpha_{21} f_1(x)$, which is orthogonal to the function $f_1(x)$. Indeed, from the condition of orthogonality $(f_2(x) - \alpha_{21} f_1(x), f_1(x)) = (f_2(x), f_1(x)) - \alpha_{21} \|f_1(x)\|^2 = 0$ immediately follows $\alpha_{21} = (f_2(x), f_1(x)) / \|f_1(x)\|^2$.

The Gram-Schmidt process consists of the following recursions. The first function remains unchanged: $f_1^{\perp}(x) = f_1(x)$. The second function is orthogonalizing as described by $f_2^{\perp}(x) = f_2(x) - \alpha_{21} f_1(x)$. The third function is orthogonalizing with respect to both functions $f_1^{\perp}(x)$ and $f_2^{\perp}(x)$. At step K, a function $f_K^{\perp}(x)$ is formed from the function $f_K(x)$

$$f_K^{\perp}(x) = f_K(x) - \sum_{j=1}^{K-1} \frac{(f_K(x), f_j^{\perp}(x))}{\left\| f_j^{\perp}(x) \right\|^2} f_j^{\perp}(x) \,, \tag{1-8}$$

which is orthogonal to all the previously constructed functions $\{f_j^{\perp}(x)\}$. At each step or after completion of them, the normalization, which means replacing $f_j^{\perp}(x)$ by $f_j^{\perp}(x) / \|f_j^{\perp}(x)\|$, can produce an orthonormal basis of the space generated by the initial set of functions. Obtaining an orthonormal basis greatly simplifies further calculations.

1.3.4. Projection onto a subspace versus minimization
of a root-mean-square deviation

For simplicity and, most importantly, for getting a visual image that reflects the meaning of the process, it is expedient to start this subsection with the case of a vector space. Let $\mathbf{F}_0$ be some vector of an N-dimensional space $\mathcal{F}$, and let $\{\mathbf{f}_k\}$ ($k = 1,...K$) be a set of K ($K < N$) certain vectors in the same space $\mathcal{F}$ and, besides, K is less than N. How do we construct a linear combination of these vectors $\mathbf{F} = \Sigma\, a_k\mathbf{f}_k$, providing the best approximation to the vector $\mathbf{F}_0$ in the sense of the minimum deviation $\varepsilon^2 = \|\,\mathbf{F}_0 - \mathbf{F}\,\|^2$?

There are two logically different ways to obtain the result answering the question. It is useful to know both.

Approach A.
Let us write

$$\varepsilon^2 = \left(\mathbf{F}_0 - \sum_k a_k\mathbf{f}_k\,,\, \mathbf{F}_0 - \sum_k a_k\mathbf{f}_k\right) = \left\|\mathbf{F}_0\right\|^2 - \left(\mathbf{F}_0,\ \sum_k a_k\mathbf{f}_k\right) - \left(\sum_k a_k\mathbf{f}_k, \mathbf{F}_0\right) + \left(\sum_k a_k\mathbf{f}_k, \sum_k a_k\mathbf{f}_k\right) \tag{1-9}$$

where ($k,\ j = 1,...K$), and consider this expression as a function of K variables α_k. It is clear that the condition of equality to zero of all K partial derivatives determines the desired extrema $\partial\varepsilon^2 / \partial a_k = -2(\mathbf{F}_0, \mathbf{f}_k) + 2(\mathbf{f}_k, \sum_{j=1}^{K} a_j\mathbf{f}_j) = 0$. These equations form a system of K algebraic equations for K unknown variables, the matrix form of which is $\langle\mathbf{L}\rangle\ \boldsymbol{\alpha} = \mathbf{b}$, where $\langle\mathbf{L}\rangle$ is the square matrix of the coefficients $L_{jk} = (\mathbf{f}_k, \mathbf{f}_j)$, and the columnar vectors $\boldsymbol{\alpha}$ and $\mathbf{b}$ are formed by coefficients $\{\alpha_k\}$ and $\{b_k = (\mathbf{F}_0, \mathbf{f}_k)\}$, respectively.

Approach B.
Imagine that the vector $\mathbf{F}_0$ of our three-dimensional space represents an N-dimensional vector $\mathbf{F}_0$ belonging to the space $\mathcal{F}$, and the plane containing the vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ is the K-dimensional subspace $\mathcal{F}_K$, as shown below in Fig. 1-2. Correspondingly, the required vector $\mathbf{F} = \Sigma\, a_k\,\mathbf{f}_k$ is a certain vector of this plane. Obviously, the deviation vector $\mathbf{F}^\perp = \mathbf{F}_0 - \mathbf{F}$ has the minimum length if the vector $\mathbf{F}$ coincides the projection of the vector $\mathbf{F}_0$ onto the "plane" $\mathcal{F}_K$. In other words, the deviation vector $\mathbf{F}^\perp$ must be perpendicular to any vector $\mathbf{f}_k$ of the set $\mathcal{F}_K$.
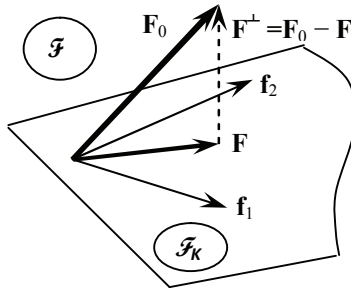
Figure 1-2: Projection onto subspace $\mathcal{F}_K$ of a vector space $\mathcal{F}$

Transforming the text above in terms of a mathematical relation, we easily gain the vector equation

$$\left(\mathbf{F}_0 - \sum_{k=1}^{K} a_k \mathbf{f}_k, \mathbf{f}_j\right) = (\mathbf{F}_0, \mathbf{f}_j) - \sum_{j=1}^{K} a_k (\mathbf{f}_k, \mathbf{f}_j) = 0 \qquad (1\text{-}10)$$

that, of course, coincides with the Ex. 1-9 obtained earlier.

After defining the notion of an inner product of functions, there is no problem in minimizing the mean square deviation of the linear combination $F(x) = \sum_{k=1}^{K} a_k f_k(x)$ of a given set of functions $\{f_k(x)\}$ from the required function $F_0(x)$. We can easily do this in the above manner. Moreover, in the case of a functional space, we can use a meaningful graphical image (Fig. 1-2), representing functions in the form of vectors, since the inner product transfers the main characteristics of a vector space to a functional one. Applying approach B to this problem, we obtain the required coefficients $\{a_k\}$ of the expansion as a solution of the system of equations

$$<\mathbf{L}>\mathbf{a} = \mathbf{p}, \qquad (1\text{-}11)$$

where $<\mathbf{L}>$ is a square matrix of order K, the coefficients of which are the mutual inner products of the functions $\{f_k(x)\}$, namely $L_{kj} = (f_j(x), f_k(x))$; $\mathbf{p}$ is a column vector of the coefficients $\{p_k = (F_0(x), f_k(x))\}$; and $\mathbf{a}$ is a column vector of the though-for coefficients $\{a_k\}$.

1.4. Adjoint operator. Selfadjoint operator and its spectrum

In mathematics, the operator is the rule for mapping elements x from X to elements y from Y. Commonly, the denote A is in use for the operator. It is said that the operator A maps X onto Y or into Y depending on

whether the set of its values coincides with the whole set Y or with its subset. It is clear that "operator" is a generalization of the notion of function to the case when the argument is not a numerical variable x, but an element of a different nature, for example, a function. If the set Y is a numerical axis, then we have an operator called a functional. Thus, an operator, a function, and a functional are terms for denoting certain types of mappings of one set on or into another.

If the mentioned sets X and Y are Hilbert spaces $\mathcal{G}$ and $\mathcal{F}$ of the functions $g(y)$ and $f(x)$ respectively, then mathematicians introduce the notion of the operator A* adjoint to the operator A. The operator A* is the adjoint to A, if the equality

$$(Ag(x), f(y)) = (g(x), A^* f(y))$$ (1-12)

is true for arbitrary elements $g \in \mathcal{G}$ and $f \in \mathcal{F}$. At first glance, the Ex. 1-12 is a pure formalism, the utility of which, perhaps, consists only in the fact that the operator A* acts in the opposite direction: mapping $\mathcal{G}$ onto $\mathcal{F}$ (Fig. 1-3). The remarkable features of the adjoint operator are not obvious. Some of them are discussed below, and others in Chapters 2 and 3 devoted to the use of the adjoint operator for solving some problems of antenna synthesis.
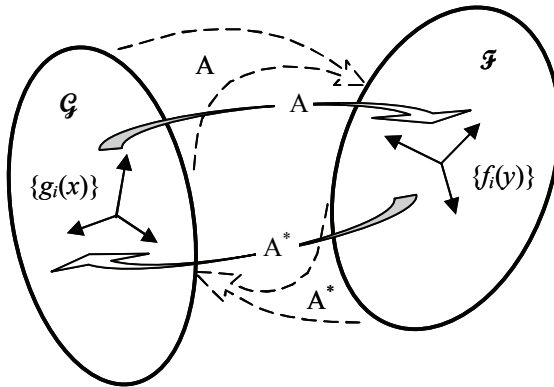
Figure 1-3: Functional spaces $\mathcal{G}$ and $\mathcal{F}$, operators A and A*, biorthogonal bases $\{g_i(x)\}$ and $\{f_i(y)\}$

Being an adherent or, perhaps, even a slave of the engineering style of thinking, which has an eagerness to identify cause-effect relationships and to form on this basis a systemic frame of knowledge, I have long been looking for an explanation of how the adjoint operator correlates with other mathematical concepts. I cannot insist that I am right, but it seems to me that mathematicians, introducing the concept of the adjoint operator,

had the purpose to transfer the notion of the transposed matrix to a functional space.

Indeed, the classical scheme for solving the SLAE $<A>$ $\mathbf{x} = \mathbf{y}$, where $<A>$ is an M-by-N matrix, $\mathbf{x}$ and $\mathbf{y}$ are column-vectors of size M and $N<M$, respectively, is as follows. The left-multiplication of both parts of the original equation by transposed matrix $<A>^T$ leads to SLAE $<B>\mathbf{x} = <A>^T\mathbf{y}$ with square matrix $<B> = <A>^T <A>$ of order N that can be easily solved. In the case of complex-valued matrices and vectors, the Hermitian matrix, which is a transposed matrix with complex conjugate elements, is in use. In the books on matrix algebra, I did not notice any mention of the fact that the transposed (or Hermitian) matrices satisfy the equation $(<A>\mathbf{x}, \mathbf{y}) = (\mathbf{x}, <A>^T\mathbf{y})$ for any vectors $\mathbf{x}$ and $\mathbf{y}$ of dimension N and M, respectively. Most likely, this is because the proof of this assertion is elementary. It suffices to write the equation in the form of sums. In Section 1.5, there is a task inviting the reader to do so.

Returning to the Hilbert space of functions, let us mention the following properties used in the next chapters.

An operator L mapping $\mathcal{G}$ onto itself is said to be selfadjoint if for any pair of elements $g_1(x)$ and $g_2(x)$ the equality $(Lg_1(x), g_2(x)) = (g_1(x), Lg_2(x))$ is true. In mathematics, it is proved that the eigenfunctions $\{g_i(x)\}$ of a selfadjoint operator L, i.e. such functions for which the equality

$$L\, g_i(x) = \lambda_i\, g_i(x) \tag{1-13}$$

holds, form an orthogonal basis of the space in which this operator is defined. The numbers $\{\lambda_i\}$ are eigenvalues. The set of eigenvalues $\{\lambda_i\}$ and eigenfunctions $\{g_i(x)\}$ forms the operator spectrum. It plays a great role, allowing not only to justify various computational algorithms, but also to understand both the properties of the operator L itself and, more importantly, the features of the inverse operator L^{-1}. This refers to the following.

Let $\{g_i(x)\}$ be an orthonormal basis consisting of the normalized eigenfunctions of L. Let us expand an arbitrary function $g(x)$ in the series

$$g(x) = \sum_i a_i\, g_i(x)\,. \tag{1-14}$$

Thanks to the orthonormality of the basis $\{g_i(x)\}$, the coefficients $\{a_i\}$ are equal to the inner products $a_i = (g(x), g_i(x))$ that lead to easy calculations. The operator L transforms the function $g(x)$ into the function $\beta(x)$

$$\beta(x) = Lg(x). \tag{1-15}$$

Taking into account Ex. 1-14, we have

$$\beta(x) = \sum_i a_i\, L g_i(x) \; = \sum_i \lambda_i a_i\, g_i(x) \; = \sum_i b_i\, g_i(x)\,. \tag{1-16}$$

Thus, the action of the operator L results in the transformation of the original coefficients $\{a_i\}$ into the coefficients $\{b_i = \lambda_i\, a_i\}$. In other words, the operator L is similar to a filter that weakens those components of the function $g(x)$, to which small eigenvalues λ_i correspond.

Now let us consider Ex. 1-15 as an equation for the unknown function $g(x)$ and the given function $\beta(x)$. To construct the inverse operator L^{-1}, let us take into account the equality Ex. 1-16

$$g(x) \;=\mathrm{L}^{-1}\beta(x) \;= \sum_i \frac{b}{\lambda_i} g_i(x)\;. \tag{1-17}$$

Thus, the inverse operator "emphasizes or rises" the components corresponding to small eigenvalues. This leads to follows: if the spectrum of the operator L contains very small eigenvalues and the corresponding components are present in the given function $\beta(x)$, then the norm of the solution $g(x) = \mathrm{L}^{-1}g(x)$ can increase catastrophically, which is an attribute of solution instability.

It is obvious that the operator $\mathrm{L} = \mathrm{A}^*\,\mathrm{A}$, formed by multiplication (sequential application) of the operators A^* and A, is a selfadjoint operator. If $\{g_i(x)\}$ is an orthogonal basis composed of eigenfunctions of the operator L, then it is not difficult to verify that the functions $\{f_i(y) = \mathrm{A}g_i(x)\}$ form an orthogonal basis of the space $\mathcal{F}$. The bases $\{f_i(y)\}$ and $\{g_i(x)\}$, which are mutually transformed into each other by the operators A* and A respectively, are called biorthogonal and play an important role in studying the properties of these operators. In Fig. 1-3, triples of mutually perpendicular vectors depict these bases and figured arrows depict the actions of the operators A and A^*.

1.5. Training tasks

Task 1A

Transform a set of six power functions $\{x^n\}$ $(n = 0\ldots5)$ given on the interval $(-1 < x < 1)$ into orthonormalize basis $\{g_n(x)\}$. Imagine that you have only a sheet of paper and a pencil, and a head on your shoulders, of course.

A hint: following the "shell approach"[3] principle or a clue that standard Schmidt orthogonalization procedure will give you at the very first step, you can guess how to simplify the problem.

---

[3] Introduction/Psychological issues/point 5.

Task 1B

Make sure that the operators $A\,g(x) = \int_a^b g(x)\,\psi(x,y)\,dx = \alpha(y)$ and

$A^*f(y) = \int_c^d f(y)\,\psi^*(x,y)\,dx = \beta(x)$ mapping complex functions defined

on the intervals $(a,\ b)$ and $(c.\ d)$, respectively, are mutually adjoint operators, i.e. satisfy the condition Ex. 1-12.

Task 1C

Assume that a selfadjoint operator L has two different orthogonal eigenfunctions, say, $g_k(x)$ and $g_m(x)$, to which the same eigenvalue corresponds, i.e. $\lambda_k\ =\ \lambda_m\ =\ \lambda$. Check whether the combination $g(x) = a\,g_k(x) + b\,g_m(x)$ is an eigenfunction of the operator L?

Task 1D

> *A preliminary remark*: If in the spectrum of some operator L there are *n* equal eigenvalues λ then the multiplicity of the eigenvalue λ is equal to *n*.

In the situation resembling the previous task, somebody shows that each of five functions $\{f_i(x)\}$ is an eigenfunction of the operator L and all of them correspond to the same eigenvalue λ. John says, "This means that the multiplicity of the eigenvalue λ is equal to five." Sam replies, "You are mistaken. It seems to me that the multiplicity of the eigenvalue λ is equal to three, if not less."

> What would you do to settle the dispute?
> *A hint*: note that a set of functions may be linearly dependent.

CHAPTER TWO

ANTENNA SYNTHESIS
FOR THE PREDETERMINED
COMPLEX RADIATION PATTERN

2.1. Introduction

For the book, this chapter is central because it is devoted to the original and fertile idea of V.N. Dymsky: to use the adjoint operator for solving the problems of antenna synthesis. Just as a streamlet gives life to a stream that, supported by other sources turns into a river, this idea in combination with heuristic methods gave birth to a series of algorithms, theoretical and practical results. Chapter 2 combines the mathematical formalism of functional analyses with profound physical interpretation. On this basis, some original notions emerge, such as the power radiated in the desired antenna pattern, the pattern's relief deviation, coefficient of error sensitivity, purely radiating distributions, etc. The adjoint operator converts the desired directional pattern into a distribution possessing some useful properties. It results in some convenient computational algorithms for the step-by-step optimization of a solution.

An attractive feature of the issues discussed in the chapter is that, with reference to the antenna technique, "dry" mathematical abstractions turn into something that has a clear physical nature. We can reproduce them on the physical level and "perceive" them with the help of a measuring setup. This chapter trains a thoughtful reader with a quality important for engineers: the ability to see, to "feel" the physical meaning in formulas, mathematical transforms and their consequences. In the Introduction[4], this quality is a mentioned ability to return with ease from the world of mathematical descriptions to the world of technical realities.

I think that here it would be appropriate to say a few words about the relationship between physics and mathematics, which reveals itself sometimes in verbal swordplay between physicists and mathematicians. In

---

[4] See the diagram in Fig. 1-3.

the year 1963, I witnessed a scene of this kind. At a seminar on the diffraction of electromagnetic waves intended for young specialists, the eminent Soviet physicists and mathematicians: academician Fock V.A., corresponding members of the academy Vainshtein L.A. and Kupradze V.D., Professors Ufimzev Ya.P., Kurochkin A.P., Boldyrev Yu.Ya. and others delivered lectures.

In one of his lectures—I quote from memory, perhaps not verbatim, but exactly in meaning—Lev Albertovich Vainshtein said:

> A rigorous formulation of the problem results in a differential equation of the hyperbolic type. However, the value of the coefficient for the last term on the left side of the equation is rather small. Therefore, we omit it and obtain an equation of parabolic type, which is much easier to solve. Without waiting for the mathematicians prove that such a replacement is permissible, we, the physicists, have received and continue to receive many new practically important results.

Hearing the last phrase, Victor D. Kupradze (mathematician, President of the Academy of Sciences of the former Georgian Soviet Socialist Republic) jumped up from his seat, and with a slight Georgian accent, very emotionally but politely retorted:

> Lev Albertovich, you must consider the age of the listeners. These are young people, and your last words can cause them the false impression that mathematics is engaged in something known to the devil that has nothing to do with practice. No, I insist that the intention and purpose of mathematics is to study the laws of the real world and nothing else than this.

Certainly, V. Kupradze was right: mathematical abstractions and formalities always correspond to reality, although often indirectly and remotely. By the way, sometimes physicists also resort to abstract concepts, the material realization of which does not exist. Vivid examples from the field of electrodynamics are so-called <u>magnetic</u> charges and currents. In the real world (strictly speaking, in the investigated region of the universe), no such field sources have been found. From my point of view, the concept of magnetic sources emerged not least for aesthetic reasons[5] (the third factor that stimulates technological progress): from the belief in the harmony of the basic laws of nature, which should reveal

---

[5] Another thing is that the formalism of magnetic field sources leads to simplification of calculations for some real sources, for instance, by replacing a small loop of electric current with a magnetic dipole or a slot in the screen with a strip of surface magnetic current on the screen.

itself in the symmetry of Maxwell's equations (compare the two versions of them given in the Tab. 2-1). The left-hand expressions correspond to the equations containing only real electric sources of the field (ρ^e and $\boldsymbol{\delta}^e$). The right-hand expressions involve fictitious magnetic sources (ρ^m and $\boldsymbol{\delta}^m$) as well. Are they not cuter? By the way, it seems to be quite possible that Maxwell guessed to insert the summand $\partial(\varepsilon\mathbf{E})/\partial t$ into the first equation, also believing in the harmony of the laws of nature. At least, at that time there was no evidence or even a hint of the existence of this term.

Table 2-1. Two forms of Maxwell equations

Just electric sources		Electric and magnetic sources
$\begin{cases} \operatorname{rot}\mathbf{H} = \boldsymbol{\delta}^e + \partial(\varepsilon\mathbf{E})/\partial t \\ \operatorname{rot}\mathbf{E} = -\partial(\mu\mathbf{H})/\partial t \\ \operatorname{div}\varepsilon\mathbf{E} = \rho^e \\ \operatorname{div}\mu\mathbf{H} = 0 \end{cases}$	from $\Longrightarrow$ to	$\begin{cases} \operatorname{rot}\mathbf{H} = \boldsymbol{\delta}^e + \partial(\varepsilon\mathbf{E})/\partial t \\ \operatorname{rot}\mathbf{E} = -\boldsymbol{\delta}^m - \partial(\mu\mathbf{H})/\partial t \\ \operatorname{div}\varepsilon\mathbf{E} = \rho^e \\ \operatorname{div}\mu\mathbf{H} = -\rho^m \end{cases}$

The content of this and next chapters represents a striking example of a close and organic unity of very formal mathematical abstractions and physical sense. In this capacity, the chapter is useful for a wide range of readers as a whetstone for sharpening the engineering style of thinking.

2.2. Operators describing antenna systems

2.2.1. Direct, inverse and quasi-inverse operators

In this chapter, we will consider the most general, fundamental issues of antenna synthesis. Therefore, assume that the synthesized antenna is formed by the field sources (electric currents, for certainty) distributed into a region v nearby some physical body w (Fig. 2-1). Let $\boldsymbol{\rho}$ denote the radius vector of points of the region v in a Cartesian coordinate system[6] xyz and $\mathbf{r}^o$ denote the radial unit vector of direction to the points on the spherical surface Ω where we control the antenna radiation field. The area Ω may be

---

[6] Theoretically, the location of the origin does not matter. However, the desire to simplify formulas and calculations prompts us to coordinate its location with the antenna system geometry.

arbitrary: the whole sphere or a solid angle of any shape, or a plane, or a set of M directions $\{ \mathbf{r}_m^o \}$ $(m = 1...M)$.
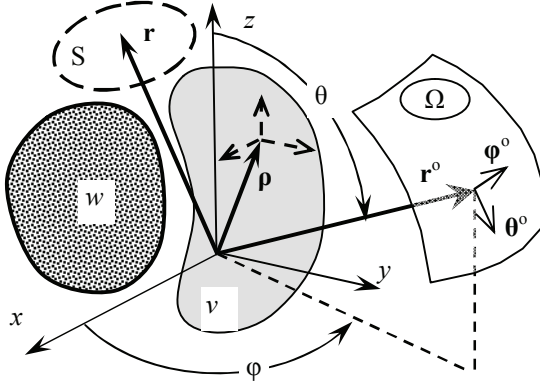
Figure 2-1: Geometric reference: v – antenna region; w – physical body; Ω – far field zone; S – near-field area of possible interest

The following expression describes the dependence of the complex-valued vector radiation pattern $\mathbf{F}(\mathbf{r}^o)$ on the current distribution $\mathbf{I}(\boldsymbol{\rho})$ in the most compact and general form due to the use of the tensor formalism

$$\mathbf{F}(\mathbf{r}^o) = \int_v \langle \mathrm{E} \rangle \mathbf{I}(\boldsymbol{\rho})\, dv . \tag{2-1}$$

It is obvious that ARP has two components $\mathbf{F}(\mathbf{r}^o) = F_\theta(\mathbf{r}^o)\, \boldsymbol{\theta}^o + F_\varphi(\mathbf{r}^o)\, \boldsymbol{\varphi}^o$, and the vector function $\mathbf{I}(\boldsymbol{\rho})$, describing the amplitude, phase and polarization of the sources, consists of three components in the general case. To shorten the subsequent expressions, let us use notation ξ symbolizing a set of unit vectors $\{\mathbf{x}^o, \mathbf{y}^o, \mathbf{z}^o\}$—a triplet of vectors depicted by the dashed line in Fig. 2-1—to represent the polarization of the current. This leads to the following compact expression

$$\mathbf{I}(\boldsymbol{\rho}) = \sum_\xi I_\xi(\boldsymbol{\rho})\, \xi^o \tag{2-2}$$

instead of $\mathbf{I}(\boldsymbol{\rho}) = I_x(\boldsymbol{\rho})\, \mathbf{x}^o + I_y(\boldsymbol{\rho})\, \mathbf{y}^o + I_z(\boldsymbol{\rho})\, \mathbf{z}^o$. In addition, if a cylindrical or spherical coordinate system is suitable for describing the polarization of the sources with regard to the geometry of the antenna, the Ex. 2-2 remains intact since we can agree that ξ and ξ^o symbolize, respectively, the axes and the unit vectors of the corresponding coordinate system.

It is clear that the tensor $\langle \mathrm{E} \rangle$ —in principle, it is sufficient to consider it as a matrix whose entries are functions—should transform the current distribution $\mathbf{I}(\boldsymbol{\rho})$ into its far field, of course, up to without a certain

multiplier[7]. It consists of six entries $\{E_{\theta\xi}(\boldsymbol{\rho}, \mathbf{r}^{\text{o}}), E_{\varphi\xi}(\boldsymbol{\rho}, \mathbf{r}^{\text{o}})\}$, each of which determines θ or φ components of the electric field corresponding to the unit value of the relevant component ($I_\xi = 1$) of the current $\mathbf{I}$ located at the point $\boldsymbol{\rho}$. The tensor

$$\langle\mathrm{E}\rangle = \begin{vmatrix} 0 & 0 & 0 \\ E_{\theta\xi_1}(\boldsymbol{\rho},\mathbf{r}^{\text{o}}) & E_{\theta\xi_2}(\boldsymbol{\rho},\mathbf{r}^{\text{o}}) & E_{\theta\xi_3}(\boldsymbol{\rho},\mathbf{r}^{\text{o}}) \\ E_{\varphi\xi_1}(\boldsymbol{\rho},\mathbf{r}^{\text{o}}) & E_{\varphi\xi_2}(\boldsymbol{\rho},\mathbf{r}^{\text{o}}) & E_{\varphi\xi_3}(\boldsymbol{\rho},\mathbf{r}^{\text{o}}) \end{vmatrix}$$

absorbs—or masks—all the computational difficulties associated with the diffraction of the EMF on the body w, except for the simplest bodies[8], for which analytical solutions exist.

It is useful to see what the universal Ex. 2-1 looks like in specific situations. First, if the antenna sources have fixed polarization, which remains unchanged in the synthesis process, then the required excitation distribution is worth determining as a scalar complex function of $I(\boldsymbol{\rho})$, and Ex. 1-1 takes the following form

$$\mathbf{F}(\mathbf{r}^{\text{o}}) = \int_v I(\boldsymbol{\rho})\,\mathbf{f}(\boldsymbol{\rho},\mathbf{r}^{\text{o}})\,dv \tag{2-3}$$

where $\mathbf{f}(\boldsymbol{\rho}, \mathbf{r}^{\text{o}})$ is the individual radiation pattern of the source located at the point $\boldsymbol{\rho}$.

Secondly, suppose that there is no body, so an electric current of arbitrary polarization (Ex. 2-2) exists in the region v in free space. Taking into account the fact that the radiation pattern of the differential volume dv at the origin with the unit current $\mathbf{I} = \boldsymbol{\xi}^{\text{o}}$ in it is equal to the vector product $[\mathbf{r}^{\text{o}}\,[\boldsymbol{\xi}^{\text{o}}\,\mathbf{r}^{\text{o}}]]\,dv$, we can write

$$\mathbf{F}(\mathbf{r}^{\text{o}}) = \int_v \exp(jk\,\boldsymbol{\rho}\,\mathbf{r}^{\text{o}})\sum_\xi I_\xi(\boldsymbol{\rho})\,[\mathbf{r}^{\text{o}}[\boldsymbol{\xi}^{\text{o}}\mathbf{r}^{\text{o}}]]\,dv\,. \tag{2-4}$$

It is clear that the well-known phase term $\exp(jk\,\boldsymbol{\rho}\,\mathbf{r}^{\text{o}})$ is due to the difference in the path of the rays from the origin and from the point $\boldsymbol{\rho}$ to the observation point in the direction $\mathbf{r}^{\text{o}}$. In this term, the multiplication of two vectors ($\boldsymbol{\rho}$ by $\mathbf{r}^{\text{o}}$) means their inner product.

Ex. 2-1 is so universal that it also includes antenna arrays, since the volume measure[9] dv can be the set of δ-functions $\{\delta(\boldsymbol{\rho} - \boldsymbol{\rho}_n)\}$ and, of course, in this case the integral in Ex. 2-1 turns into the summation

$$\mathbf{F}(\mathbf{r}^{\text{o}}) = \sum_{n=1}^N \langle\mathrm{E}\rangle\mathbf{I}(\boldsymbol{\rho}_n)\,. \tag{2-5}$$

---

[7] By the way, the dimension of this multiplier is meter/volt.
[8] For example, a metal body in the form of a plane, a cylinder, wedge, sphere or ellipsoid.
[9] In mathematics, this type of integration is the Stieltjes integral [11].

A more common expression is relevant to the case of an antenna array of sources having predetermined polarization:

$$\mathbf{F}(\mathbf{r}^o) = \sum_{n=1}^{N} I_n \mathbf{f}_n(\mathbf{r}^o) \exp(jk\, \boldsymbol{\rho}_n\, \mathbf{r}^o) . \tag{2-6}$$

Here, I_n and $\mathbf{f}_n(\mathbf{r}^o)$ are, respectively, the excitation coefficient and the individual vector pattern of the nth element if it locates at the origin of the global coordinate system. The meaning of the phase term $\exp(jk\, \boldsymbol{\rho}_n\, \mathbf{r}^o)$ is the same as above.

We will consider functions $\mathbf{I}(\boldsymbol{\rho})$ and $\mathbf{F}(\mathbf{r}^o)$ as elements of Hilbert spaces $\mathcal{I}$ and $\mathcal{F}$, respectively, with the inner products defined as follows

$$
\begin{aligned}
\left(\mathbf{I}_1(\boldsymbol{\rho}), \mathbf{I}_2(\boldsymbol{\rho}) \right) &= \int_v \mathbf{I}_1(\boldsymbol{\rho})\, \mathbf{I}_2^*(\boldsymbol{\rho})\, dv \\
\left(\mathbf{F}_1(\mathbf{r}^o), \mathbf{F}_2(\mathbf{r}^o) \right) &= \int_\Omega \mathbf{F}_1(\mathbf{r}^o)\, \mathbf{F}_2^*(\mathbf{r}^o)\, d\Omega
\end{aligned}
\tag{2-7}
$$

Here, as it was above, the asterisk * denotes complex conjugation; multiplied vectors ($\mathbf{I}_1$ and $\mathbf{I}_2$, $\mathbf{F}_1$ and $\mathbf{F}_2$) mean their inner product. The metrics generated—in accordance with Ex. 1-4—by inner products Ex. 2-7 mean standard deviation. The square of norms usually associates with the antenna excitation power $P_{\mathbf{I}} = \|\mathbf{I}(\boldsymbol{\rho})\|^2$ and radiation power $P_\Sigma = \|\mathbf{F}(\mathbf{r}^o)\|^2$, respectively.

Ex. 2-1 determines a linear integral operator—we will denote it by U— that maps the space $\mathcal{I}$ of excitation functions into the subspace $\mathcal{F}_v \subset \mathcal{F}$ of the radiation patterns that are realizable by sources distributed within the given region v. Thus, the simplest expression

U $\mathbf{I}(\boldsymbol{\rho})$ = $\mathbf{F}(\mathbf{r}^o)$ (2-8)

replaces initial Ex. 2-1 along with all its variations, for instance, Ex. 2-3– 2-6, as well as more detailed ones. We call the operator U, which describes what the ARP corresponds to the distribution $\mathbf{I}(\boldsymbol{\rho})$, a direct operator, bearing in mind that it acts from antenna input to its "output".

The task of synthesizing antennas is to solve Ex. 2-8 with reference to the desired ARP $\mathbf{F}_0(\mathbf{r}^o)$ (the opposite direction from antenna "output" to its input). Since in the case of an unrealizable FRP $\mathbf{F}_0(\mathbf{r}^o)$ there is no such excitation distribution $\mathbf{I}(\boldsymbol{\rho})$ that would satisfy Ex. 2-8, then the inverse operator—in the strict sense of the word—does not exist. The realistic aim of antenna synthesis is to construct a quasi-inverse operator U_σ^{-1} that converts an arbitrary desired pattern $\mathbf{F}_0(\mathbf{r}^o)$ into an excitation distribution $\mathbf{I}_\sigma(\boldsymbol{\rho}) = U_\sigma^{-1} \mathbf{F}_0(\mathbf{r}^o)$, which provides the best—in the certain sense— approach to the desired pattern: $U\, \mathbf{I}_\sigma(\boldsymbol{\rho}) \approx \mathbf{F}_0(\mathbf{r}^o)$.

2.2.2. Adjoint operator, purely radiating and non-radiating distributions

The operator that is adjoint to the operator U—we denote it by V—has some very attractive features, due to which it is useful to use it to construct a quasi-inverse operator. In his PhD thesis (Kazan Aviation Institute, 1964, in Russian), Dymsky V.N. showed that the following integral transformation determines the adjoint operator V

$$\mathbf{V}\mathbf{F}(\mathbf{r}^{o}) = \oint_{\Omega} \langle E \rangle^{*} \mathbf{F}(\mathbf{r}^{o}) \, d\Omega, \tag{2-9}$$

where asterisk * at a tensor denotes a Hermitian conjugation that is a combination of matrix transpose and complex conjugation.

One of these features relates to the following. In antennas of some configuration, there may be specific current distributions that do not radiate[10]. They form a corresponding subspace $\mathcal{J}_Q \subset \mathcal{J}$ of so-called reactive distributions. The orthogonal complement of the subspace $\mathcal{J}_Q$ to the space $\mathcal{J}$, which we denote by $\mathcal{J}_\Sigma$, corresponds to the semantic name "subspace of purely radiating distributions." It turns out that the adjoint operator V maps the space $\mathcal{F}$ of arbitrary patterns $\mathbf{F}(\mathbf{r}°)$ onto the subspace $\mathcal{J}_\Sigma$. Moreover, for any purely radiating distribution $\mathbf{I}(\boldsymbol{\rho}) \subset \mathcal{J}_\Sigma$ such a pattern $\mathbf{F}(\mathbf{r}°)$ exists that the equation $\mathbf{I}(\boldsymbol{\rho}) = \mathrm{V}\,\mathbf{F}(\mathbf{r}°)$ is true.

The proof of the first statement follows easily from the Cauchy–Bunyakovsky–Schwarz inequality (Ex. 1-3) and the adjoint operator definition (Ex. 1-10). Indeed, assume that the operator V maps a certain ARP $\mathbf{F}_0(\mathbf{r}^o)$ to a current distribution $\mathbf{p}(\boldsymbol{\rho})$, i.e. $\mathbf{p}(\boldsymbol{\rho}) = \mathrm{V}\mathbf{F}_0(\mathbf{r}^o)$. Like a catalyst, this ARP does not affect the result of our reasoning, but it takes part in the process itself. Suppose that there exists a distribution $\mathbf{I}(\boldsymbol{\rho})$ different from the distribution $\mathbf{p}(\boldsymbol{\rho})$, such that the ARPs of both coincide: $\mathrm{U}\mathbf{I}(\boldsymbol{\rho}) = \mathrm{U}\mathbf{p}(\boldsymbol{\rho}) = \mathbf{F}(\mathbf{r}^o)$. The following elementary manipulations that use Ex. 1-10 and the assumption just made, is true:

$$\left\|\mathbf{p}(\boldsymbol{\rho})\right\|^2 = \big(\mathbf{p}(\boldsymbol{\rho}), \mathbf{p}(\boldsymbol{\rho})\big) = \big(\mathrm{V}\mathbf{F}_0(\mathbf{r}^o), \mathbf{p}(\boldsymbol{\rho})\big) = \big(\mathbf{F}_0(\mathbf{r}^o), \mathrm{U}\mathbf{p}(\boldsymbol{\rho})\big) =$$
$$\big(\mathbf{F}_0(\mathbf{r}^o), \mathrm{U}\mathbf{I}(\boldsymbol{\rho})\big) = \big(\mathrm{V}\mathbf{F}_0(\mathbf{r}^o), \mathbf{I}(\boldsymbol{\rho})\big) = \big(\mathbf{p}(\boldsymbol{\rho}), \mathbf{I}(\boldsymbol{\rho})\big) \tag{2-10}$$

Because of the Cauchy–Bunyakovsky–Schwarz inequality, we have a strict[11] inequality $|(\mathbf{p}(\boldsymbol{\rho}), \mathbf{I}(\boldsymbol{\rho}))| < \|\mathbf{p}(\boldsymbol{\rho})\| \, \|\mathbf{I}(\boldsymbol{\rho})\|$. Given Ex. 2-10 and replacing the left hand-side of the considered inequality with $\|\mathbf{p}(\boldsymbol{\rho})\|^2$, we

---

[10] The radial-polarized current $\mathbf{I}(\boldsymbol{\rho}) = I(\boldsymbol{\rho}) \, \boldsymbol{\rho}^o$ distributed central-symmetrically within a sphere is an obvious example. Some less trivial examples are below.

[11] Since functions $\mathbf{p}(\boldsymbol{\rho})$ and $\mathbf{I}(\boldsymbol{\rho})$ are different, as we assumed.

get $\|\mathbf{p}(\boldsymbol{\rho})\| < \|\mathbf{I}(\boldsymbol{\rho})\|$. In other words, among all distributions $\mathbf{I}(\boldsymbol{\rho})$, ARPs of which $\mathbf{F}(\mathbf{r}^o)$ are identical, the distribution $\mathbf{p}(\boldsymbol{\rho})$, yielded by the operator V, has the smallest norm.

This clearly indicates that the distribution $\mathbf{p}(\boldsymbol{\rho})$ is a purely radiating distribution. Indeed, an arbitrary distribution consists of two terms $\mathbf{I}(\boldsymbol{\rho}) = \mathbf{I}_\Sigma(\boldsymbol{\rho}) + \mathbf{I}_Q(\boldsymbol{\rho})$:   (2-11)
purely radiating and non-radiating, respectively. Since the terms are orthogonal, the equality $\|\mathbf{I}(\boldsymbol{\rho})\|^2 = \|\mathbf{I}_\Sigma(\boldsymbol{\rho})\|^2 + \|\mathbf{I}_Q(\boldsymbol{\rho})\|^2$ is in force. Consequently, the non-radiating term that does not change ARP simply increases the distribution norm. Therefore, the minimal distribution norm corresponds to the absence of this term.

To see how pure radiating and non-radiating current distributions look, consider the two-dimensional case (Fig. 2-2): the electric current $\mathbf{I}(\boldsymbol{\rho})$—which may be uniformed along z-axis—is distributed in a region of radius $R = 1.25\ \lambda$ in the cross plane H of an infinite ideally conducting wedge, the outer angle of which is $\alpha_0 = 300°$. The ARP generated by the current is controlled in the same plane H and, of course, has only a φ-polarization $\mathbf{F}(\varphi) = F_\varphi(\varphi)\ \boldsymbol{\varphi}^o$.
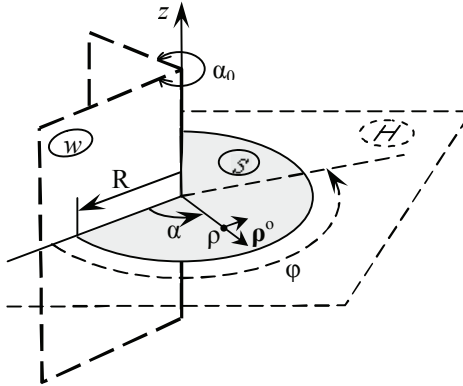
Figure 2-2: Geometric reference: w is metallic wedge of outer angle α_0; S is the region of current sources; ρ and α are the polar coordinates of points within the region S; H is the plane where antenna sources locate and ARP $F_\varphi(\varphi)$ is under control

Perhaps the idea of how to find these distributions is no less important than the observation of their appearance. Electrical currents of both ρ- and α- polarizations produce a radiation field of the required polarization. Assume that the current distribution is
$$\mathbf{I}(\rho,\alpha) = I_\rho(\rho)\ \sin(v_n\alpha)\ \boldsymbol{\rho}^o + I_\alpha(\rho)\ \cos(v_n\alpha)\ \boldsymbol{\alpha}^o. \qquad (2\text{-}12)$$

Here, $\nu_n = n\pi/\alpha_0$ is the nth angular harmonic multiplier. Taking into consideration the well-known solution for diffraction on a wedge [12, 13], we can up to a constant multiplier, write the pattern corresponding to this current[12] as following

$$F_\varphi(\varphi) \; \propto \; \cos(\nu_n\varphi)\int_{\rho=0}^{R}\Big\{I_\rho(\rho)\;\big(J_{\nu_n-1}(k\rho)+J_{\nu_n+1}(k\rho)\big)+ \\ I_\alpha(\rho)\;\big(J_{\nu_n-1}(k\rho)-J_{\nu_n+1}(k\rho)\big)\Big\}\,\rho\,d\rho \qquad (2\text{-}13)$$

Here, $J_\nu(x)$ denotes a Bessel function of the first kind for order ν.

Ex. 2-13 denotes operator U for the task under consideration. Let the desired ARP is $\mathbf{F}_0(\varphi) = \cos(\nu_n\varphi)\,\boldsymbol{\varphi}^{\mathrm{o}}$. Taking into account that each summand in Ex. 2-13 corresponds to the relevant terms of the current Ex. 2-12, we have

$$\mathbf{p}(\boldsymbol{\rho}) \; = \; V\mathbf{F}_0(\varphi) \; \propto \; \big(J_{\nu_n-1}(k\rho)+J_{\nu_n+1}(k\rho)\big)\sin(\nu_n\alpha)\,\boldsymbol{\rho}^{\mathrm{o}} + \\ \big(J_{\nu_n-1}(k\rho)-J_{\nu_n+1}(k\rho)\big)\cos(\nu_n\alpha)\,\boldsymbol{\alpha}^{\mathrm{o}} \qquad (2\text{-}14)$$

Fig. 2-3 shows halves of three current distributions $\mathbf{I}(\rho,\alpha)$. The shades of gray, from the maximum (black) to zero (white), represent the amplitude reliefs and the short arrows indicate directions of the currents. Light gray lines are a coordinate grid with $\lambda/4$ spacing. The purely radiating distribution given by Ex. 2-14 is shown in Fig. 2-3a. It is clear that its ARP is $\mathbf{F}(\varphi) = \cos(\nu_n\varphi)\,\boldsymbol{\varphi}^{\mathrm{o}}$.
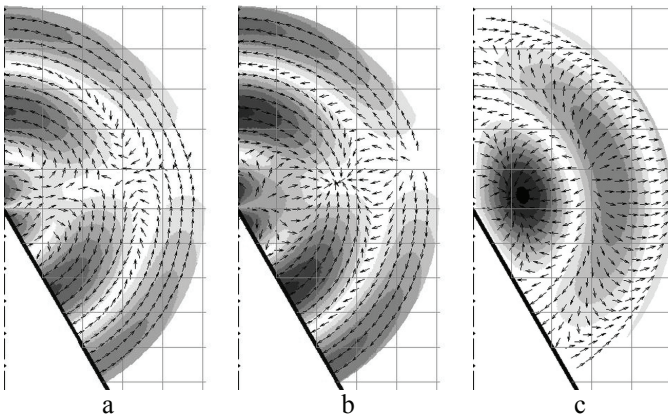
Figure 2-3: Current distributions: a) a pure radiating distribution forming the pattern $\mathbf{F}_0(\varphi) = \cos(\nu_n\varphi)\,\boldsymbol{\varphi}^{\mathrm{o}}$; b) and c) − non-radiating distributions

---

[12] Except for the number $n = 0$

Perhaps the easiest way to construct a non-radiating distribution is to use the fact that the radiation of ρ- and α- components of the current $\mathbf{p}(\boldsymbol{\rho})$ is of the same pattern $\mathbf{F}(\varphi) = \cos(v_n\varphi)\ \boldsymbol{\varphi}^o$. Consequently, they will compensate each other's field if the amplitudes of their fields are equal, and the phases differ by π. This idea results in the formula

$$\mathbf{I}_{Q1}(\boldsymbol{\rho}) = \mathrm{V}\mathbf{F}_0(\varphi) \propto \left(J_{v_n-1}(k\rho) + J_{v_n+1}(k\rho) \right) \sin(v_n\alpha)\,\boldsymbol{\rho}^o - \sqrt{\lambda_\rho / \lambda_\alpha} \left(J_{v_n-1}(k\rho) - J_{v_n+1}(k\rho) \right) \cos(v_n\alpha)\,\boldsymbol{\alpha}^o \quad . \tag{2-15}$$

Here, λ_ρ is equal to $\int_0^R | J_{v_n-1}(k\rho) + J_{v_n+1}(k\rho)|^2\ \rho\, d\rho$ and λ_α is equal to $\int_0^R | J_{v_n-1}(k\rho) - J_{v_n+1}(k\rho)|^2\ \rho\, d\rho$. Fig. 2-3b shows the non-radiating distribution $\mathbf{I}_{Q1}(\boldsymbol{\rho})$.

A more formal and at the same time a more fertile way of constructing non-radiation distributions can consist of two stages. First, it is necessary to choose some distribution $\mathbf{I}(\boldsymbol{\rho})$ that forms the same ARP as the distribution $\mathbf{p}(\boldsymbol{\rho})$. In our case, $\mathbf{F}(\varphi) = \cos(v_n\varphi)\ \boldsymbol{\varphi}^o$ and, therefore, $\mathbf{I}(\rho,\alpha)$, given by Ex. 2-12 with arbitrary functions $I_\rho(\rho)$ and $I_\alpha(\rho)$, is suitable. In accordance with Ex. 2-11, it consists of two terms: $\mathbf{I}_\Sigma(\boldsymbol{\rho})$, which in the situation under consideration is equal to $\mathbf{p}(\boldsymbol{\rho})$, given by Ex. 2-14, and $\mathbf{I}_Q(\boldsymbol{\rho})$. At the second stage, it is necessary to select the non-radiating term $\mathbf{I}_Q(\boldsymbol{\rho})$ from the chosen distribution $\mathbf{I}(\boldsymbol{\rho})$. This goal is easily achievable via the Schmidt orthogonalization procedure (see Ex. 1-6). As a result, we have

$$\mathbf{I}_Q(\boldsymbol{\rho}) = \mathbf{I}(\boldsymbol{\rho}) - \frac{(\mathbf{I}(\boldsymbol{\rho}), \mathbf{p}(\boldsymbol{\rho}))}{\|\mathbf{p}(\boldsymbol{\rho})\|^2}\,\mathbf{p}(\boldsymbol{\rho}) . \tag{2-16}$$

Fig. 2-3c shows the non-radiating distribution $\mathbf{I}_{Q2}(\boldsymbol{\rho})$, obtained with the help of the above expression from the distribution $\mathbf{I}(\boldsymbol{\rho})$, in which both functions $I_\rho(\rho)$ and $I_\alpha(\rho)$ are equal to ρ: $I_\rho(\rho) = I_\alpha(\rho) = \rho$.

Varying the functions $I_\rho(\rho)$ and $I_\alpha(\rho)$ in an arbitrary way, we can obtain as many different non-radiating distributions as we like. Is it not surprising that for the considered space $\mathbf{I}$ of distributions generated by combinations $I_\rho(\rho)\sin(v_n\alpha)\ \boldsymbol{\rho}^o + I_\alpha(\rho)\cos(v_n\alpha)\ \boldsymbol{\alpha}^o$, for every nth angular harmonic, there is only one purely radiating and an unlimited number of non-radiating distributions? This means that there are an unlimited number of distributions $\mathbf{I}(\boldsymbol{\rho})$ generating the same ARP $\mathbf{F}(\varphi) = \cos(v_n\varphi)\ \boldsymbol{\varphi}^o$.

It is difficult not to wonder how the above-mentioned is consistent with the electromagnetic uniqueness theorem. The theorem asserts that sources located on the surface can create any external electromagnetic field. This means that there is no need for internal sources to obtain any

radiation pattern. Thus, the electromagnetic uniqueness theorem has nothing to do with sources distributed continuously inside a closed surface and, therefore, it does not say anything about them. Moreover, the theorem hints at the fact that non-radiating distributions can exist. A thoughtful reader can take pleasure in deciphering a hint and obtaining evidence.

Of course, the non-radiation problem is much more theoretical than practical, not least because antennas formed by a volumetric current hardly exist as a real construction. An inhomogeneous dielectric body illuminated by a feed makes the rarest exception. In addition to this, in widely used antenna arrays, which are discrete structures, non-radiating distributions do not exist in principle. Although in some very extravagant cases they can exist. For example, if the set of directions inaptly chosen to control the ARP so that its directions coincide with the nulls of ARP for some distribution in the antenna array, then this distribution turns out to be a non-radiating distribution. In addition, if the distance between the array elements is small enough and their number is large, then almost non-radiating distributions exist. Indeed, the radiation of neighbor elements excited out of phase is insignificant. Summarizing the issue, we can state that even in practical cases it is worthwhile being aware of the possible existence of non-radiating distributions.

2.3. Operators involving in antenna synthesis and their spectrums

2.3.1. Selfadjoint and quasi-inverse operators

Suppose that the desired ARP $\mathbf{F}_0(\mathbf{r}^o)$ is realizable—belonging to the subspace $\mathcal{F}_v$—and the equation

$$U \, \mathbf{I}(\boldsymbol{\rho}) = \mathbf{F}_0(\mathbf{r}^o) \tag{2-17}$$

has an exact solution $\mathbf{I}(\boldsymbol{\rho})$. Multiplying this equation by the adjoint operator V, we obtain the equation

$$L \, \mathbf{I}(\boldsymbol{\rho}) = \mathbf{p}_0(\boldsymbol{\rho}) \tag{2-18}$$

with the selfadjoint operator $L = VU$ and a purely radiating distribution $\mathbf{p}_0(\boldsymbol{\rho})$ on the right-hand side to which the operator V maps the ARP $\mathbf{F}_0(\mathbf{r}^o)$: $\mathbf{p}_0(\boldsymbol{\rho}) = V \, \mathbf{F}_0(\mathbf{r}^o)$. Using the inverse operator L^{-1}, we can determine the exact solution as follows:

$$\mathbf{I}(\boldsymbol{\rho}) = L^{-1} V \, \mathbf{F}_0(\mathbf{r}^o). \tag{2-19}$$

Suppose now that the desired ARP does not belong to the subspace $\mathcal{F}_v$. This means that it contains the term $\mathbf{F}_0^{\perp}(\mathbf{r}^o)$, which is orthogonal to the subspace $\mathcal{F}_v$, that is, $\mathbf{F}_0(\mathbf{r}^o) = \mathbf{F}_{0v}(\mathbf{r}^o) + \mathbf{F}_0^{\perp}(\mathbf{r}^o)$. Thanks to the useful features of the adjoint operator V, it "ignores" the term $\mathbf{F}_0^{\perp}(\mathbf{r}^o)$, and

therefore the solution Ex. 2-19 reproduces only the orthogonal projection of an arbitrary desired pattern $\mathbf{F}_0(\mathbf{r}^o)$ onto the subspace $\boldsymbol{\mathcal{F}}_v$: $U\,\mathbf{I}(\boldsymbol{\rho}) = \mathbf{F}_{0v}(\mathbf{r}^o)$. In other words, the solution Ex 2-19 provides the best mean square approximation of the desired pattern and thereby defines the quasi-inverse operator $U_\sigma^{-1} = L^{-1}V$.

The spectrum of a selfadjoint operator L, which is the set of its eigenfunctions $\{\mathbf{g}_n(\boldsymbol{\rho})\}$ and eigenvalues λ_n, determines the basic properties of both itself and the inverse operator. Therefore, it would be useful and interesting to get the mentioned spectrum. Unfortunately, in the general case, this requires laborious calculations, and only in some particular cases is it possible to obtain it with the help of some sophisticated analytical reasoning.

Denote the set of normalized eigenfunctions of the operator L by $\{\mathbf{g}_n(\boldsymbol{\rho})\}$. This means that $L\mathbf{g}_n(\boldsymbol{\rho}) = \lambda_n\,\mathbf{g}_n(\boldsymbol{\rho})$ and $(\mathbf{g}_n(\boldsymbol{\rho}), \mathbf{g}_m(\boldsymbol{\rho})) = \delta_{nm}$[13]. The set $\{\mathbf{g}_n(\boldsymbol{\rho})\}$ composes an orthonormal basis of the subspace $\boldsymbol{\mathcal{I}}_\Sigma$ of purely radiating distributions.

It is easy to see that, due to the structure of the operator L, the patterns corresponding to the distributions $\{\mathbf{g}_n(\boldsymbol{\rho})\}$ are mutually orthogonal. Indeed, $(U\mathbf{g}_n(\boldsymbol{\rho}), U\mathbf{g}_m(\boldsymbol{\rho})) = (VU\mathbf{g}_n(\boldsymbol{\rho}), \mathbf{g}_m(\boldsymbol{\rho})) = (L\mathbf{g}_n(\boldsymbol{\rho}), \mathbf{g}_m(\boldsymbol{\rho})) = \lambda_n\,(\mathbf{g}_n(\boldsymbol{\rho}), \mathbf{g}_m(\boldsymbol{\rho})) = \lambda_n\,\delta_{nm}$. Thus, the radiation patterns

$$\mathbf{G}_n(\mathbf{r}^o) \;=\; U\,\mathbf{g}_n(\boldsymbol{\rho})\,/\,\sqrt{\lambda_n} \qquad\qquad (2\text{-}20)$$

compose the orthonormal basis of the subspace $\boldsymbol{\mathcal{F}}_v$. It is just as elementary, as above, to verify that the adjoint operator V performs a symmetrical inverse transformation

$$\mathbf{g}_n(\boldsymbol{\rho}) \;=\; V\,\mathbf{G}_n(\mathbf{r}^o)\,/\,\sqrt{\lambda_n}\;. \qquad\qquad (2\text{-}21)$$

If the bases $\mathbf{g}_n(\boldsymbol{\rho})$ and $\{\mathbf{G}_n(\mathbf{r}^o)\}$ are known, then the following explicit expression defines the quasi-inverse operator U_σ^{-1}

$$\mathbf{I}(\boldsymbol{\rho}) \;= U_\sigma^{-1}\mathbf{F}_0(\mathbf{r}^o) = \sum_n \frac{(\mathbf{F}_0(\mathbf{r}^o), \mathbf{G}_n(\mathbf{r}^o))}{\sqrt{\lambda_n}}\,\mathbf{g}_n(\boldsymbol{\rho})\,. \qquad (2\text{-}22)$$

Ex. 2-22 is not very important for practical implementations but it plays a large methodological role clearly indicating that the norm of the solution $\mathbf{I}(\boldsymbol{\rho})$ can raise catastrophically if the desired pattern $\mathbf{F}_0(\mathbf{r}^o) = \sum_n a_n\,\mathbf{G}_n(\mathbf{r}^o)$ contains terms corresponding to very small eigenvalues.

Fig. 2-4 illustrates the basic properties exhibited by operators U and V: the first maps the entire space $\boldsymbol{\mathcal{I}}$ of arbitrary distributions on the subspace $\boldsymbol{\mathcal{F}}_v$ of the patterns that are inherent to the antenna, and the second maps the

---

[13] Here, too, the notation δ_{nm} means Kronecker delta.

entire space $\mathcal{F}$ of arbitrary patterns on the subspace $\mathcal{I}_\Sigma$ of purely radiating distributions. They also convert the bases $\{g_n(\rho)\}$ and $\{G_n(r^o)\}$ into each other.
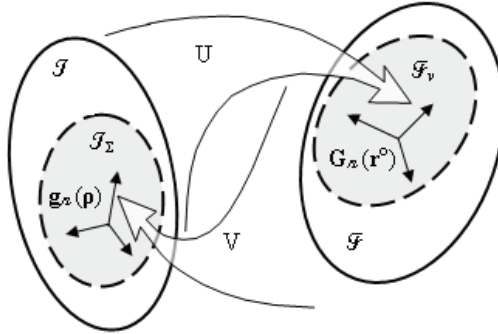
Figure 2-4: Mappings performing by operators U and V

It is easy to verify that in subspaces $\mathcal{I}_\Sigma$ and $\mathcal{F}_v$ there exist a unique pair of orthogonal bases that transform into each other with the help of operators U and V, as in Ex. 2-21 and Ex. 2-22. In some cases, this avoids the hard problem of solving the corresponding equation (Ex. 1-11) in order to obtain the eigenfunctions of the operator L. Indeed, we can simply choose some orthogonal basis in the subspace $\mathcal{F}_v$ or $\mathcal{I}_\Sigma$, and then, after using the operator V or U, respectively, check whether the resulting set of functions is orthogonal—we have the thought-for bases—or not—the attempt was unsuccessful.

The aforementioned choice is a guess, in fact, and there is no universal rule on how to do it. However, in the case of antennas of some regular geometry, including antenna arrays, the right choice may seem obvious— with caution in the non-use, or unrealizable ARP or non-radial distributions, respectively.

2.3.2. Examples of spectrums: $\{g_n\}$, $\{G_n\}$, $\{\lambda_n\}$ – for some structures of sources

For simplicity's sake, let us limit ourselves to 2D situations. The following expressions [14] will be very useful to achieve the objective of this section:

$$e^{j x \cos \psi} = \sum_{m=0}^{\infty} (1 + \delta_{0m}) j^m J_m(x) \cos(m\psi) , \tag{2-23}$$

$$\cos\psi\, e^{jx\cos\psi} = \sum_{m=0}^{\infty}(2-\delta_{0m})j^{m+1}[J_{m-1}(x)-J_{m+1}(x)]\cos(m\psi) \quad (2\text{-}24)$$

where $J_m(x)$ is Bessel function of order m. We will consider two cases: the electric current distributed on the plane of the section inside (or on the boundary) of the circular cylinder around the edge of an ideally conducting wedge (Fig. 2-2) or similarly distributed current in free space. For both cases the coordinate systems shown in Fig. 2-2—$\{z, \rho, \alpha\}$ for sources and $\{\theta, \varphi\}$ for a far-zone field—will be in use.

Example 1: a circle of z-polarized current in free space

Let the radius of the cycle be ρ_0, and the current is $I_z(\alpha) = \exp(\pm j\, n\, \alpha)$. For $\theta = 90°$, as it is, the obvious equality $[\mathbf{r}^o[\mathbf{r}^o\ \mathbf{z}^o]] = \mathbf{\theta}^o$ is valid, and Ex. 2-4 for the straight operator U turns to be as follows:

$$F_\theta(\varphi) = \int_0^{2\pi} e^{jk\rho_0\,\cos(\varphi-\alpha)}\, e^{\pm jn\alpha} d\alpha \quad (2\text{-}25)$$

because the distribution $\mathbf{I}(\rho)$ consists only of the z-component $\mathbf{I}(\rho_0, \alpha) = \exp(\pm j\, n\alpha)\, \mathbf{z}^o$ and all the entries of the tensor $<\mathbf{E}>$ are equal to zero, except for the single entry $E_{\theta z}$, which is $\exp(jk\rho_0\cos(\varphi - \alpha))$. For simplicity, taking into account the further normalization of ARP, we omitted the factor ρ_0 in Ex. 2-25.

Substituting Ex. 2-23 (with $\psi = \varphi - \alpha$) in Ex. 2-25 and performing the integration it is easy to obtain the following closed expression

$$F_\theta(\varphi) = j^n J_n(k\rho_0)\, e^{\pm jn\varphi}. \quad (2\text{-}26)$$

In this particular case, the Hermitian conjugated tensor $<\mathbf{E}>^*$ has only one non-zero entry $E_{z\theta} = \exp(-jk\rho_0\cos(\varphi - \alpha))$, therefore the adjoint operator V defined by Ex. 2-9, looks like this

$$p_z(\alpha) = V F_\theta(\varphi) = \int_0^{2\pi} e^{-jk\rho_0\,\cos(\varphi-\alpha)}\, F_\theta(\varphi)\, d\varphi. \quad (2\text{-}27)$$

Ex. 2-27 and Ex. 2-25 have a similar structure; therefore, if $F_\theta(\varphi) = \exp(\pm j\, n\varphi)$, then we have

$$p_z(\alpha) = j^n J_{\mp n}(k\rho_0)\, e^{\pm jn\alpha}. \quad (2\text{-}28)$$

Since the functions $\exp(\pm nx)\}$ are mutually orthogonal in the interval $0 < x < 2\pi$, we have—after normalization and taking into account the equality $|J_{-n}(x)| = |J_n(x)|$—the sought-for bases and eigenvalues as follows:

$$\{g_n(\alpha) = e^{\pm jn\alpha}/\sqrt{2\pi\rho_0}\}, \quad \{G_n(\varphi) = e^{\pm jn\varphi}/\sqrt{2\pi}\},$$
$$\{\lambda_n = |J_n(k\rho_0)|\}. \quad (2\text{-}29)$$

Try to answer the question of whether there is any non-radiating distribution in the case under consideration or not. For a thoughtful reader who wants to make their own decision before reading the hint, the text of the hint is "disguised" in italics below.

If the radius ρ_0 of the sources' circle is equal to a root of Bessel function, i.e. $J_n(k\rho_0) = 0$, then $\lambda_n = 0$, which means that the current distribution $g_n(\alpha)$ does not radiate. The next question is how many non-radiating distributions can exist? It seems to me that there is only one. Although a rigorous answer should be sought somewhere in the very depth of the theory of Bessel functions concerning the problem of the multiplicity of the root of the equation $J_n(k\rho_0) = 0$, with a fixed value of $k\rho_0$ and variable integer n, if any.

If the value of $k\rho_0$ does not coincide with some root of any Bessel function $J_n(x)$ then, theoretically, non-radiating distributions do not exist. However, Bessel functions $J_n(x)$ converges to zero if $n \gg x$. This means that the corresponding distributions, practically, do not radiate which accords to the physical sense: the size of the antenna puts a limit to the sharpness of its ARP. In other words, the number of effectively radiating angular harmonics is in some proportion with the antenna size.

A rough approximation to this proportion is the following inequality $n < 1.2\,\rho_0$. Unfortunately, even in this interval the set of eigenvalues $\{\lambda_n\}$ given by Ex. 2-29 can contain and, for a relatively large radius, inevitably contains very small eigenvalues corresponding to extremely weakly radiating distributions.

Example 2: a circular loop with a α-polarized current in free space

Using the above notations with the current $I_\alpha(\alpha) = \exp(\pm j\,n\,\alpha)$ and obvious equality $[\mathbf{r}^o[\mathbf{r}^o\,\boldsymbol{\alpha}^o]] = \cos(\varphi-\alpha)\,\boldsymbol{\varphi}^o$, we have Ex. 2-4 for the direct operator U, which turns into the next

$$F_\alpha(\varphi) = \int_0^{2\pi} \cos(\varphi-\alpha)\, e^{jk\rho_0\,\cos(\varphi-\alpha)}\, e^{\pm jn\alpha} d\alpha \ . \tag{2-30}$$

It is because the distribution $\mathbf{I}(\boldsymbol{\rho})$ consists merely of the α-component $\mathbf{I}(\rho_0,\alpha) = \exp(\pm j\,n\alpha)\,\boldsymbol{\alpha}^o$ and all the entries of the tensor $<E>$ are equal to zero, except for the only entry $E_{\varphi\alpha}$, which is $\cos(\varphi-\alpha)\exp(jk\rho_0\cos(\varphi-\alpha))$.

In this situation, the above line of reasoning results in the following expressions:

$$\{g_n(\alpha) = e^{\pm jn\alpha} / \sqrt{2\pi\rho_0}\}, \quad \{G_n(\varphi) = e^{\pm jn\varphi} / \sqrt{2\pi}\},$$
$$\{\lambda_n = |J_{n-1}(k\rho_0) - J_{n+1}(k\rho_0)|\}. \tag{2-31}$$

As the derivation $J_n'(x) = dJ_n(x)/dx$ is equal to $[J_{n-1}(x) - J_{n+1}(x)]/2$ then the existence of non-radiating distributions depends on the condition $J_n'(k\rho_0) = 0$, that is completely analogical to the previous condition $J_n(k\rho_0) = 0$. Therefore, all the above reasoning is true including about the extremely weakly radiating distributions.

Example 3: a circular loop with a ρ-polarized current in free space

Repeating the above logic in relation to the current $I_\rho(\alpha) = \exp(\pm j\, n\, \alpha)$ and $[\mathbf{r}^\circ[\mathbf{r}^\circ\ \boldsymbol{\rho}^\circ]] = \sin(\varphi - \alpha)\ \boldsymbol{\varphi}^\circ$, we will gain the following final expressions:

$$\{g_n(\alpha) = e^{\pm j n \alpha} / \sqrt{2\pi \rho_0}\}, \quad \{G_n(\varphi) = j\, e^{\pm j n \varphi} / \sqrt{2\pi}\},$$

$$\{\lambda_n = |J_{n-1}(k\rho_0) + J_{n+1}(k\rho_0)|\}. \tag{2-32}$$

As for the non-radiating distributions, the situation is the same, as the previous one, plus the following addition: irrespective of the radius ρ_0, the uniform distribution ($n = 0$) is non-radiating. This is clear and physically —the z-component of the fields of any pair of currents symmetrically located relative to the observation point, cancel each other—and formally: Ex. 2-31 yields zero for $n = 0$, since $J_{-1}(x) = -J_1(x)$.

Example 4: a circle of Huygens elements in free space

Independent of whether z- or α- polarization is the case, the individual pattern of the sources in the plane of circle is the same with regard to z- or φ- field component, respectively: $f(\alpha, \varphi) = 1 + \cos(\varphi - \alpha)$. Clearly, Ex. 2-3 for $I(\alpha) = \exp(\pm j\, \alpha)$ transforms to the sum of Ex. 2-25 and Ex. 2-30

$$F_{z|\varphi}(\varphi) = \int_0^{2\pi} [1 + \cos(\varphi - \alpha)] e^{j k \rho_0\ \cos(\varphi - \alpha)}\ e^{\pm j n \alpha} d\alpha \tag{2-33}$$

Here the notation "$z|\varphi$" means or z- or φ- polarization.

In this case, taking into account Ex. 2-29, Ex. 2-31, and being cautious with the imaginary unit in Ex. 2-31, we obtain

$$\{g_n(\alpha) = e^{\pm j n \alpha} / \sqrt{2\pi \rho_0}\}, \quad \{G_n(\varphi) = j\, e^{\pm j n \varphi} / \sqrt{2\pi}\},$$

$$\{\lambda_n = |J_n(k\rho_0) + j J_n'(k\rho_0)|\}. \tag{2-34}$$

The advantageous, in practical terms, feature of this case is that, unlike the previous three cases, there are no small eigenvalues $\{\lambda_n\}$ in the interval $0 \le n < 1.2\ \rho_0$. This is due to the fact that if $J_n'(k\rho_0) = 0$, then the value of $|J_n(k\rho_0)|$ has local maxima, and, conversely, if $J_n(k\rho_0) = 0$, then the value of $|J_n'(k\rho_0)|$ is close to local maxima. Therefore, all the angular harmonics belonging to the mentioned interval are well-radiating distributions.

Let us compare the analyzed structures. Fig. 2-5 demonstrates the above-mentioned and very important features of the spectrums $\{\lambda_n\}$ inherent in field sources of analyzed configurations.
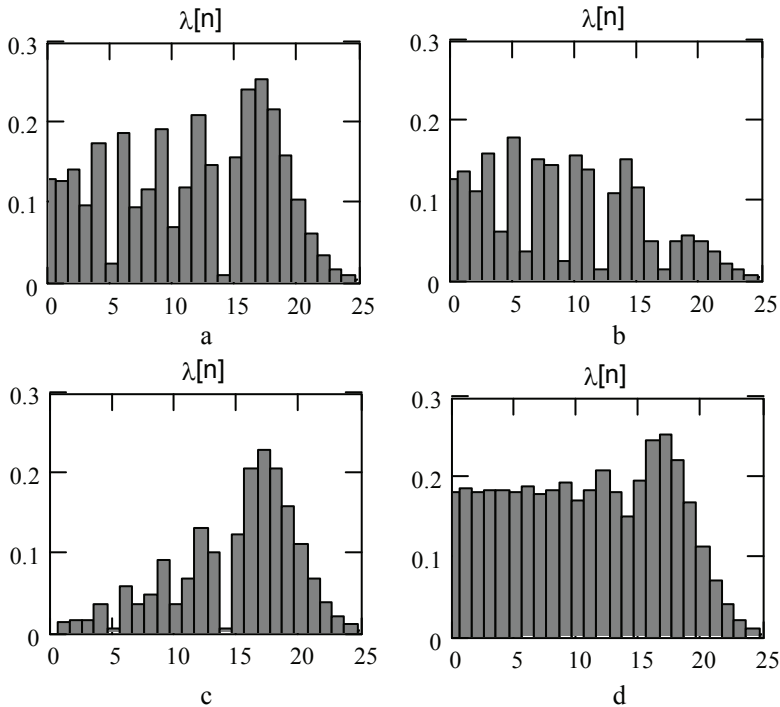
Figure 2-5: Eigenvalues $\{\lambda_n\}$ inherent in sources continuously distributed along a circle of 3 λ radius: a) z-polarized electric current; b) α-polarized electric current; c) ρ-polarized electric current; d) Huygens elements

The presented computational data are consistent with understandable regularities. First, in all the cases, the eigenvalues λ_n for $n > 25$—that is, more than 1.3 $k\rho_0$—are sufficiently small to be of interest. Secondly, the current of ρ- polarization is hardly suitable for emission in the plane of its location, because so many of its initial harmonics radiate poorly. "Example 3" is present mainly for the sake of theoretical completeness. However, it is necessary when synthesizing the current distribution with non-predetermined polarization—similar to Ex. 2-14. Thirdly, Huygens sources have the flattest dependence $\lambda(n)$ in the interval $n < k\rho_0$, which means they can radiate each of the corresponding harmonics equally effectively.

Physically, it is clear that in the first three cases, the small eigenvalues correspond to those harmonics at which a resonance occurs when, due to isotropic or quasi-isotropic radiation, the fields emitted by sources and

opposite sources compensate each other. Those resonances are impossible in the case of Huygens sources due to their directed—cardioid—radiation pattern.

Example 5: a circular array of z-polarized currents in free space

Let there be N sources, which are z-polarized currents, located uniformly along the circle of radius ρ_0 in free space. Their angular coordinates are $\alpha_k = 2\pi k / N$, where $k = 0 \ldots N{-}1$.

Bearing in mind the same idea of using current distributions and ARPs of the harmonica type, we cannot start at trying-out the set of APRs $\{G_n(\varphi) = \exp(\pm j n \varphi)\}$, as it was convenient and successful above. The point is that being a discrete structure the antenna array cannot produce those patterns precisely; consequently, they do not belong to subspace $\mathcal{F}_v$ of the patterns inherent in the antenna array.

Let's try to start in the opposite way: to check whether the set of harmonics $\{g_n(\alpha_k) = \exp(j n \alpha_k)\}$ with $n = 0 \ldots N{-}1$ makes the orthogonal basis of the subspace $\mathcal{J}_\Sigma$ of purely radiating distributions. For simplicity's sake, we postpone normalization to the final stage. Since in any antenna array, non-radiating distributions cannot be, the subspace $\mathcal{J}_\Sigma$ coincides with the total space $\mathcal{J}$. After and if the verification is successful, it will be necessary to find out whether the diagrams $\{G_n(\varphi) = U\, g_n(\alpha)\}$ are mutually orthogonal.

Advice. My dear reader, try to guess, please, why do we ignore the set $\{g_n(\alpha_k) = \exp(-j n \alpha_k)\}$? It is useful—not least for training the engineering style of thinking—to reveal at the end of your studies a striking similarity to the well-known stroboscopic effect: here—discreteness in space, there—discreteness in time.

In the case under consideration, the following formula defines the inner product of any pair of distributions $g_n(\alpha)$ and $g_m(\alpha)$ in antenna array

$$(g_n(\alpha),\, g_m(\alpha)) = \sum\nolimits_{k=0}^{N-1} \exp(j(n-m)\,\alpha_k)\,. \qquad (2\text{-}35)$$

If $n \neq m$ that is the situation of our interest, Ex. 2-35 actually contains the sums of cosine and sine. In the outstanding mathematical reference book[14] Gradshteyn I.S. and Ryzhik I.M. "*Table of Integrals, Series, and Products.*" (Eighth edition, 2014), one can find two formulas of trigonometric summation that we have combined into following two-story expression

$$\sum_{k=0}^{n} {\sin \atop \cos}(kx) = {\sin \atop \cos}(nx/2)\ \sin((n+1)x/2)\ \operatorname{cosec}(x/2)\,. \qquad (2\text{-}36)$$

---

[14] Available at: https://www.sciencedirect.com/book/9780123849335/table-of-integrals-series-and-products.

To apply Ex. 2-36 to Ex. 2-35, the following substitutions are necessary: $n = N - 1$, $x = 2\pi(n - m)/N$, which leads to the equality $(n + 1)x/2 = \pi(n - m)$. This means that the middle term on the right-hand side of Ex. 2-36 is equal to zero for any n and $m \neq n$. If m is equal to n, then it is obvious that Ex. 2-35 yields the value of N. By the way, Ex. 2-36 gives the same[15]. Consequently, the set of N mutually orthogonal functions $\{g_n(\alpha)\}$ does form an orthogonal basis of the N-dimensional space $\mathcal{J}_\Sigma$.

As for the set of ARPs $\{G_n(\varphi) = U\, g_n(\alpha)\}$, in the case of the antenna array Ex. 2-23 converts to the sum

$$G_n(\varphi) = \sum_{k=0}^{N-1} e^{j n \alpha_k}\, e^{j k \rho_0 \, \cos(\varphi - \alpha_k)} \,. \tag{2-37}$$

The substitution Ex. 2-23 into Ex. 2-37 yields $1+\delta_{0m}$:

$$G_n(\varphi) = \sum_{m=0}^{\infty} (1+\delta_{0m})\, j^m J_m(k\rho_0) \sum_{k=0}^{N-1} e^{j n \alpha_k} \cos m(\varphi - \alpha_k). \tag{2-38}$$

In fact, Ex. 2-38 is a Fourier series. A possible way to find out whether functions $\{G_n(\varphi)\}$ are mutually orthogonal involves analyzing the coefficients of the series. The substitution of the trivial $\exp(j\, n\, \alpha_k) = \cos(n\, \alpha_k) + j \sin(n\, \alpha_k)$ transforms the last summation in Ex. 2-38 to the following: $\sum_k \exp(j\, n\, \alpha_k) \cos(\varphi - \alpha_k) = [cc(n, m) + j\, sc(n, m)] \cos(m\, \varphi) + j\, [ss(n, m) - j\, cs(n, m)] \sin(m\, \varphi)$. Here the notations

$$cc(n, m) = \sum_{k=0}^{N-1} \cos(n\alpha_k) \cos(m\alpha_k),$$
$$ss(n, m) = \sum_{k=0}^{N-1} \sin(n\alpha_k) \sin(m\alpha_k), \tag{2-39}$$

$$sc(n, m) = \sum_{k=0}^{N-1} \sin(n\alpha_k) \cos(m\alpha_k),$$
$$cs(n, m) = \sum_{k=0}^{N-1} \cos(n\alpha_k) \sin(m\alpha_k) \tag{2-40}$$

are in use. First, it is easy to see, that in view of Ex. 2-36 both $sc(n, m)$ and $cs(n, m)$ are equal to zero, regardless of the values of numbers n and m. This leads to the following entry for Ex. 2-38

$$G_n(\varphi) = \sum_{m=0}^{\infty} \zeta_m J_m(k\rho_0)[cc(n,m)\cos(m\varphi) + j\, ss(n,m)\sin(m\varphi)]. \tag{2-41}$$

Here, ζ_m denotes $(1+\delta_{0m})\, j^m$.

Secondly, rearranging Ex. 2-39 to the summation of $\cos(k\,\pi(n \pm m)/N)$ makes it possible to use Ex. 2-36 and obtain a resultant expression, the terms of which, unfortunately, need a cumbersome analysis with regard for the various combinations of the values of the numbers N, n and m and their parity as well. In order not to make mistakes during this analysis, it is very

---

[15] Clearly, $\cos(\pi(N-1)(n-m)/N)|_{m=n}$ is equal to 1 and the discovering uncertainty $\sin(\pi(n-m))\operatorname{cosec}(\pi(n-m)/N)|_{m=n}$ yields N.

useful to calculate Ex. 2-39 directly for a series of combinations of numbers N, n, m.

Advice: People who are committed to the engineering style of thinking strive to visualize everything they are dealing with. Such a habit helps to reveal and understand the basic regularities or any features that are much more difficult to identify from formulas. In the case under consideration, we can easily draw—e.g. in MathCad—the dependencies $cc(n, m)$ and $ss(n, m)$ as functions of m with a set of combinations of parameters N and n, and then to draw the tables summarizing the results as shown below.

In Fig. 2-6, empty cells mean that both sums $cc(n, m)$ and $ss(n, m)$ are equal to zero. Cells crossed diagonally contain symbols representing sums—$cc(n, m)$ above and $ss(n, m)$ below the diagonal—normalized to the maximum equal to 2, which is equal to N. Symbols "1", "0", "+" and "−" denote the values 1, 0, 0.5 and −0.5, respectively.
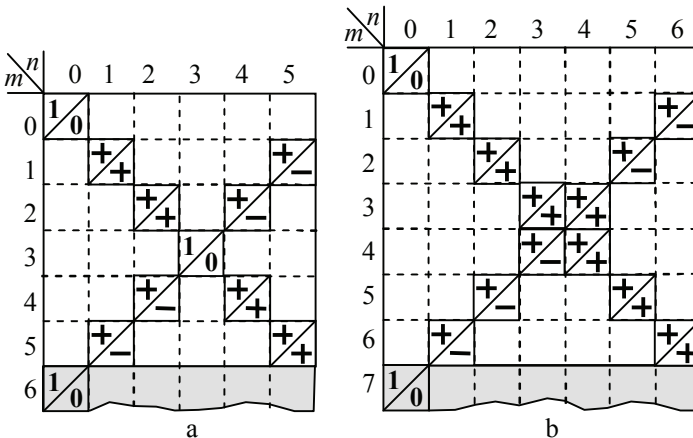
Figures 2-6: Typical patterns of the sums $cc(n, m)$ and $ss(n, m)$: a) the case of an even number N; b) the case of for an odd number N

Obviously, the number of sources N limits the number of harmonics ($0 \leq n \leq N - 1$), and both trigonometric sums $cc(n, m)$ and $ss(n, m)$ in reference to the number m have a period equal to N. Therefore, it is enough to draw the initial ($0 \leq m \leq N - 1$) part of the tables. Depending on the parity of N, the sum patterns are slightly different, retaining a cross-like structure. Fig. 2-6a corresponds to $N = 6$ and Fig. 2-6b – to $N = 7$.

Taking into account Ex. 2-41 and meaning of symbols "+" and "−", and omitting the factor $N/2$ for simplicity, we see that ⊞ and ⊟ denote

$$(\cos(m\varphi) + j\sin(m\varphi)) = \mathrm{e}^{jm\varphi} \qquad \text{and} \qquad (\cos(m\varphi) - j\sin(m\varphi)) = \mathrm{e}^{-jm\varphi},$$

respectively. Then ARP $G_n(\varphi) = \mathrm{U}g_n(\alpha_k))$ given by Ex. 2-41, seems to be as follows

$$G_n(\varphi) = N\sum_{p=0}^{\infty} j^{pN+n} J_{pN+n}(k\rho_0)\, \mathrm{e}^{j(pN+n)\varphi} \; +$$

$$N\sum_{p=1}^{\infty} j^{pN-n} J_{pN-n}(k\rho_0)\, \mathrm{e}^{-j(pN-n)\varphi} \; . \tag{2-42}$$

Perhaps not from Ex. 2-42 but with reference to Fig. 2-6 it is clear that $(G_n(\varphi),\, G_s(\varphi))|_{n \neq s} = 0$. Indeed, if $s \neq N - n$ then the functions $G_n(\varphi)$ and $G_s(\varphi)$ do not contain harmonics with the same indexes m which is an explicit indication that the functions are orthogonal. Otherwise, the functions $G_n(\varphi)$ and $G_s(\varphi)$ have terms with the same indices m, but they correspond to the opposite phase variations $\mathrm{e}^{jm\varphi}$ and $\mathrm{e}^{-jm\varphi}$, and therefore they are orthogonal in the interval $0 \leq \varphi \leq 2\pi$.

Thus, the functions given by Ex. 2-42 are mutually orthogonal, and after normalization, we obtain the following final expressions:

$$g_n(\alpha_k) = \frac{\mathrm{e}^{jn\alpha_k}}{\sqrt{N}}, \quad \lambda_n = \sqrt{\sum_{p=0}^{\infty}\left|J_{pN+n}(k\rho_0)\right|^2 + \sum_{p=1}^{\infty}\left|J_{pN-n}(k\rho_0)\right|^2}$$

$$G_n(\varphi) = \frac{1}{\lambda_n\sqrt{2\pi N}}\sum_{p=0}^{\infty} j^{pN+n} J_{pN+n}(k\rho_0)\, \mathrm{e}^{j(pN+n)\varphi} \; + \tag{2-43}$$

$$\frac{1}{\lambda_n\sqrt{2\pi N}}\sum_{p=1}^{\infty} j^{pN-n} J_{pN-n}(k\rho_0)\, \mathrm{e}^{-j(pN-n)\varphi}$$

Since Bessel functions $J_m(x)$ rapidly decrease for $m > 1.2\, x$, it is practically acceptable to limit the summations in Ex. 2-43 by the condition $p < [3.6\, k\rho_0/N]$ or even by $p < [2.4\, k\rho_0/N]$. Here, square brackets mean a value truncated to an integer.

Example 6: z-polarized current distributed continuously in free space between two circles of radius ρ_1 and ρ_2.

Unlike Example 1, the current distribution is now a function of two variables $I_z(\rho, \alpha)$. As for the dependencies on α, it remains tempting to choose harmonics. However, what functions should we try for dependencies on ρ? Intuition is silent.

On the other hand, it is clear that Fourier series can represent any ARP that is a smooth function in interval $0 \leq \varphi < 2\pi$. So why not try the basis $\{G_n(\varphi) = \mathrm{e}^{\pm jn\varphi}/\sqrt{2\pi}\}$? The adjoint operator V maps a function $G_n(\varphi)$ to

$$g_n(\rho,\alpha) = \int_0^{2\pi} e^{jk\rho\cos(\varphi-\alpha)} \, G_n(\varphi) \, d\varphi = j^n 2\pi \, J_n(k\rho) \, e^{\pm j n\alpha} \qquad (2\text{-}44)$$

It is clear that inner product of any pair $g_n(\rho,\ \alpha)$ and $g_m(\rho,\ \alpha)$ $(g_n(\rho,\alpha), g_m(\rho,\alpha)) = \int_{\rho_1}^{\rho_2} \int_0^{2\pi} g_n(\rho,\alpha) \, g_m^*(\rho,\alpha) \, \rho \, d\rho \, d\alpha$ is equal to zero if $n \neq m$, i.e. functions $\{g_n(\rho,\ \alpha)\}$ are mutually orthogonal. After normalization, the following formulas arise

$$g_n(\rho,\alpha) = j^n \frac{J_n(k\rho)}{\sqrt{\lambda_n}} e^{\pm j n\alpha}, \quad G_n(\varphi) = \frac{1}{\sqrt{2\pi}} e^{\pm j n\varphi}$$

$$\lambda_n = 2\pi \int_{\rho=\rho_1}^{\rho_2} |J_n(k\rho)|^2 \rho \, d\rho \qquad \qquad (2\text{-}45)$$

As for non-radiation distributions, the situation is similar to that described in Section 2.2.2 above. The distribution $I_z(\rho,\alpha) = I(\rho) \, e^{j n\alpha}$ with an arbitrary function $I(\rho)$ produces the same ARP $F_\theta \sim e^{j n\varphi}$ and only one of them— $g_n(\rho,\ \alpha)$ given by Ex. 2-45—is the purely radiating distribution. Any others contain a non-radiating term belonging to the subspace $\boldsymbol{J}_Q$. We can select this term using Schmidt orthogonalization procedure (Ex. 1-6) with reference to the distribution $g_n(\rho,\alpha)$.

Other Examples. In some other cases we can obtain the spectrums $\{g_n(\rho,\ \alpha)\}$, $\{G_n(\varphi)\}$, $\{\lambda_n\}$ using the above arguments. Tab. 2-1 is a directory for a number of 2D situations.

2.4. Set of parameters evaluating the quality of antenna synthesis solutions

The most common characteristics that evaluate the quality of obtained excitation distribution are those that feature the preciseness, energetic effectiveness and stability of the solution. Of course, there are other important characteristics, such as directivity, beam width, side-lobe level, bandwidth, and so on. Nevertheless, we can note that the directivity factor and the side-lobe level somehow correlate with the accuracy of the solution—if the desired pattern is a δ-function, of course—and bandwidth is directly dependent on solution stability.

1. Concept of the power radiated into the desired pattern

It is always possible and very useful to represent any real antenna pattern $\mathbf{F}(\mathbf{r}^o)$ as a sum of two orthogonal terms $\mathbf{F}(\mathbf{r}^o) = \alpha \, \mathbf{F}_0(\mathbf{r}^o) + \mathbf{F}^\perp(\mathbf{r}^o)$, in which the first term differs from the desired pattern $\mathbf{F}_0(\mathbf{r}^o)$ only by some scalar factor α, as is shown in Fig. 2-7.
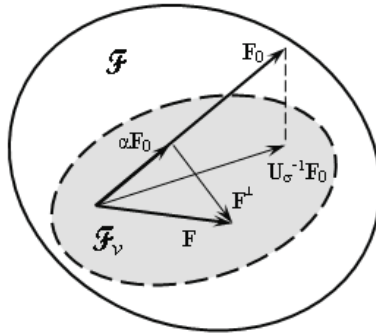
Figure 2-7: Orthogonal decomposition of the ARP $\mathbf{F}(\mathbf{r}°)$

According to this decomposition, the power $P_\Sigma = \|\mathbf{F}_0(\mathbf{r}°)\|^2$ radiated by the antenna is of two terms $P_\Sigma = P_{0\Sigma} + P^\perp$, the first of which V.N. Dymsky called "the power radiated into the desired pattern," and following him, we will use this meaningful term. The second term is an energy measure of the deviation of the actual pattern $\mathbf{F}(\mathbf{r}°)$ from the desired $\mathbf{F}_0(\mathbf{r}°)$ one. Since according to Ex. 1-6 the factor α is equal to $(\mathbf{F}(\mathbf{r}°), \mathbf{F}_0(\mathbf{r}°)) / \|\mathbf{F}_0(\mathbf{r}°)\|^2$ then the power radiated into the desired pattern is as follows

$$P_{0\Sigma} = |(\mathbf{F}(\mathbf{r}°), \mathbf{F}_0(\mathbf{r}°))|^2 / \|\mathbf{F}_0(\mathbf{r}°)\|^2. \tag{2-46}$$

2. Mean square deviation of an antenna pattern's shape

The value of the normalized standard deviation an antenna pattern $\mathbf{F}(\mathbf{r}°)$ from the desired one

$$\sigma = \| \mathbf{F}(\mathbf{r}°) - \mathbf{F}_0(\mathbf{r}°) \| / \| \mathbf{F}_0(\mathbf{r}°) \| \tag{2-47}$$

is a common characteristic evaluating the accuracy of the solution. Note that the value of σ depends on the scale factor. Suppose we have two patterns $\mathbf{F}_1(\mathbf{r}°)$ and $\mathbf{F}_2(\mathbf{r}°)$, which differ only in some scale factor μ: $\mathbf{F}_2(\mathbf{r}°) = \mu\, \mathbf{F}_1(\mathbf{r}°)$. In a practical sense, these patterns are equal, but standard deviation σ estimates their accuracy as not equal to each other.

It would be reasonable to slightly modify the characteristic σ so that it depends only on the shape of the ARP (on its angular function) regardless of the scale factor. A simple idea of how to achieve the goal is to calculate the standard deviation, but for a pre-scaled ARP as follows: $\underline{\sigma} = \| \mathbf{F}_0(\mathbf{r}°) - \gamma_{opt}\mathbf{F}(\mathbf{r}°) \| / \| \mathbf{F}_0(\mathbf{r}°) \|$. The search for the optimal scale factor leads to the following variation task: $\min_\gamma \| \mathbf{F}_0(\mathbf{r}°) - \gamma \mathbf{F}(\mathbf{r}°) \|^2$. The condition $d\|\ldots\|^2/d\gamma = 0$ results in the following optimal complex value $\gamma_{opt} = ((\mathbf{F}_0(\mathbf{r}°), \mathbf{F}(\mathbf{r}°))/\|\mathbf{F}(\mathbf{r}°)\|^2$. Finally, the following formula for the

normalized mean square deviation of antenna pattern's shape is in force:

$$\underline{\sigma}^2 = 1 - |\,(\mathbf{F}_0(\mathbf{r}°), \mathbf{F}(\mathbf{r}°)\,|^2 \big/ \|\,\mathbf{F}_0(\mathbf{r}°)\,\|^2 \|\,\mathbf{F}(\mathbf{r}°)\,\|^2 \,. \tag{2-48}$$

It is interesting to note that variables $\underline{\sigma}^2$, P_Σ, $P_{0\Sigma}$, and $P^\perp$ are in a very close and meaningful dependence: $\underline{\sigma}^2 = 1 - P_{0\Sigma}/P_\Sigma = P^\perp/P_\Sigma$.

Of course, the values of both measures—the standard deviation of σ and $\underline{\sigma}$—are equal for the solution obtained by the MRSD algorithm, since the scaling of that solution is automatically optimal. However, if the solution arises because of optimization under additional constraints, for example, with restriction of its norm (so-called regularization) or during various iterative procedures, then $\underline{\sigma}$ more adequately estimates the accuracy of the solution.

Commentary with reference to $\underline{\sigma}^2$

Honestly, I do not understand why such a rational and useful thing as the deviation of the shapes of radiation patterns published many years ago—in my dissertation [15] and some articles [16-18]—remains unnoticed. Of course, I did not expect something like applause or the widespread use of this thing. Nevertheless, even being a modest person, I expected that the idea would be worth recognizing by the relevant scholars. Sometimes, when discussing such situations in a conversation "from heart to heart," I tell the students: "It's not my fault; it's a prank of humanity that it ignores something worthwhile."

By the way, mathematicians use the meaningful notation {$\boldsymbol{\rho}°$, $\boldsymbol{\varphi}°$, $\mathbf{z}°$} and {$\mathbf{r}°$, $\boldsymbol{\theta}°$, $\boldsymbol{\varphi}°$} for a standard basis set of vectors for cylindrical and spherical coordinate systems, respectively. However, why for the Cartesian system do they use the meaningless notation {$\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$} instead of the logical {$\mathbf{x}°$, $\mathbf{y}°$, $\mathbf{z}°$}? Perhaps this is so because humanity values the vagueness of the humanitarian style of thinking above a clearness of the engineering style. Nevertheless, following Dymsky, I prefer to use the notation {$\mathbf{x}°$, $\mathbf{y}°$, $\mathbf{z}°$}, and the students more than approve of it, they like it.

3. The solution's sensitivity to random errors

An important feature of the solution—as an amplitude and phase distribution (APhD) in antenna—is the relation $Q = \|\mathbf{I}(\boldsymbol{\rho})\|^2/\|\mathbf{F}(\mathbf{r}°)\|^2$, often called the reactivity of APhD and interpreted as the ratio between the total excitation power (with regard to the reactive component) and the radiated power. In my opinion, the interpretation is not strict but physically justified, because it somehow reflects the reactivity of the antenna.

However, the parameter Q has another meaning simultaneously[16]. With a typical statistics of random errors $\delta\mathbf{I}(\boldsymbol{\rho})$, which inevitably accompany the reproduction of the nominal distribution $\mathbf{I}(\boldsymbol{\rho})$, the relative dispersion $\delta^2 = D(\delta\mathbf{F}(\mathbf{r}^o)) / \|\mathbf{F}(\mathbf{r}^o)\|^2$ of random deviations $\delta\mathbf{F}(\mathbf{r}^o)$ of ARP, caused by the distribution errors $\delta\mathbf{I}(\boldsymbol{\rho})$, is equal to the following: $\delta^2 = \varepsilon^2\, Q/\, Q_0$. Here, $\delta^2 = D(\delta\mathbf{I}(\boldsymbol{\rho})) / \|\mathbf{I}(\boldsymbol{\rho})\|^2$ is the relative dispersion of errors $\delta\mathbf{I}(\boldsymbol{\rho})$, and Q_0 is a value that depends only on the geometry of the antenna.

Thus, $Q = \|\mathbf{I}(\boldsymbol{\rho})\|^2/\|\mathbf{F}(\mathbf{r}^o)\|^2$ is **solution's sensitivity to random errors** of its reproduction.

2.5. Uniqueness and stability of antenna synthesis solution

2.5.1. Introductory remarks

As for the uniqueness of antenna synthesis solutions, the situation is simple. If there are two[17] solutions $\mathbf{I}_1(\boldsymbol{\rho})$ and $\mathbf{I}_2(\boldsymbol{\rho})$ to which the same ARP corresponds, then the distribution $\mathbf{I}(\boldsymbol{\rho}) = \mathbf{I}_1(\boldsymbol{\rho}) - \mathbf{I}_2(\boldsymbol{\rho})$ does not radiate. The converse is true as well: if there are non-radiating distributions in the antenna, the solution of the antenna synthesis cannot be unique. Thus, non-uniqueness of the antenna synthesis solution is the other side of the existence of non-radiating distributions.

It is not difficult to cope with this problem, imposing an additional condition on the sought-for distribution. For example, the mentioned condition may be as following: $\|\mathbf{I}(\boldsymbol{\rho})\|^2 \leq P_{\max}$ or equal to it the condition $\|\mathbf{I}(\boldsymbol{\rho})\|^2 \leq Q_0\, \|\mathrm{U}\,\mathbf{I}(\boldsymbol{\rho})\|^2$. The choice of certain values for $P_{\max}$ or Q_0 is a vague task, for the solution of which there is hardly anything else than a method of trial and error.

The other way to cope with the problem of non-uniqueness is to get the best approximation to the desired ARP $\mathbf{F}_0(\mathbf{r}^o)$, seeking for a solution into subspace purely radiating distributions $\mathcal{J}_\Sigma$. In fact, the quasi-inverse operator $\mathrm{U}_\sigma^{-1} = \mathrm{L}^{-1}\mathrm{V}$, given by Ex. 2-22, determines this precisely, thanks to the peculiarities of the adjoint operator V that maps an arbitrary desired pattern $\mathbf{F}_0(\mathbf{r}^o)$ to the subspace $\mathcal{J}_\Sigma$.

The problem of the stability of the solution is much broader. Even in the case of antenna arrays in which non-radiating distributions do not exist and the solution of any synthesis task is unique, its stability can be unacceptably poor, since some eigenvalues can be very small. Moreover, this is so in an infinite-dimensional situation, when the continuously

---

[16] See Section 2.4.
[17] Especially if there are several of them.

distributed sources are along a certain curve, on a surface or inside a volume. Mathematicians have proved [19] that eigenvalues of any selfadjoint operator are positive and zero can be the only point of their convergence. This means that in the case of continuously distributed sources, the operator L has infinitesimal eigenvalues causing an infinite increase in the solution norm (Ex. 2-22 shows this explicitly) and its instability. There are two ways to enhance the stability of the solution: regularization (stems from formal reasoning) and optimization (stems from physical reasons).

2.5.2. Regularization

A brief remark about an ill-posed problem

From a definition given by Jacques Hadamard, a well-posed problem should have the following three properties [20]: 1) a solution exists; 2) the solution is unique; 3) the solution's behavior changes continuously with the initial conditions. If any of the conditions do not satisfy, the problem becomes an ill-posed one.

The inverse problems, to which the synthesis of antennas relates, are often ill-posed, mainly because of the third condition. Even if a problem is well-posed, it can still be ill-conditioned, meaning that a small deviation in the right side of the equation can result in much larger alteration in the solution, as it is with antenna arrays.

In the case of continuously distributed sources, the antenna synthesis equation $U\mathbf{I}(\boldsymbol{\rho}) = \mathbf{F}_0(\mathbf{r}^\circ)$ is certainly an ill-posed mathematical problem, since infinitesimal changes $\delta\mathbf{F}_0(\mathbf{r}^\circ)$ in the right side of the equation can cause finite changes $\delta\mathbf{I}(\boldsymbol{\rho})$ in its solution. In such situations, Tikhonov regularization is a commonly used method [21, 22]. The general idea of regularization is to impose an additional condition onto the sought solution. Very often this condition is the inequality constraining the norm of the solution: $\|\mathbf{I}(\boldsymbol{\rho})\|^2 \le P_{max}$.

Using the method of Lagrange multipliers to solve the optimization task that is to minimize $\|\mathbf{F}_0(\mathbf{r}^\circ) - U\,\mathbf{I}(\boldsymbol{\rho})\|^2$ while satisfying the condition $\|\mathbf{I}(\boldsymbol{\rho})\|^2 \le P_{max}$, we obtain the following equation instead of Ex. 2-18

$$L\,\mathbf{I}(\boldsymbol{\rho}) + \mu\,\mathbf{I}(\boldsymbol{\rho}) = (L + \mu\,E)\,\mathbf{I}(\boldsymbol{\rho}) = \mathbf{p}(\boldsymbol{\rho}), \qquad (2\text{-}49)$$

where E is the identity operator—a unit matrix in the finite dimensional case—and μ denotes the Lagrange multiplier whose value must be such that the imposed condition is met.

It is clear that the eigenfunctions $\{\mathbf{g}_n(\boldsymbol{\rho})\}$ of operator L serve simultaneously the eigenfunctions $\{\mathbf{g}_n(\boldsymbol{\rho})\}$ of operator $(L + \mu\,E)$ with eigenvalues $\{\lambda_n + \mu\}$. Indeed, a series of equalities $(L + \mu\,E)\,\mathbf{g}_n(\boldsymbol{\rho}) =$

$L\mathbf{g}_n(\boldsymbol{\rho}) + \mu\, E\mathbf{g}_n(\boldsymbol{\rho}) = \lambda_n\, \mathbf{g}_n(\boldsymbol{\rho}) + \mu\, \mathbf{g}_n(\boldsymbol{\rho}) = (\lambda_n + \mu)\, \mathbf{g}_n(\boldsymbol{\rho})$ is obviously true. Thus, the solution Ex. 2-22 transforms into regularized distribution, which is

$$\mathbf{I}_\mu(\boldsymbol{\rho}) = \sum_n \frac{(\mathbf{F}_0(\mathbf{r}°), \mathbf{G}_n(\mathbf{r}°))}{\lambda_n + \mu} \sqrt{\lambda_n}\, \mathbf{g}_n(\boldsymbol{\rho})\,. \tag{2-50}$$

Consequently, by virtue of Ex. 2-20, the ARP $\mathbf{F}_\mu(\mathbf{r}°) = U\, \mathbf{I}_\mu(\boldsymbol{\rho})$, corresponding to this solution, turns into the following expansion in functions $\{\mathbf{G}_n(\mathbf{r}°)\}$

$$\mathbf{F}_\mu(\mathbf{r}°) = \sum_n \frac{(\mathbf{F}_0(\mathbf{r}°), \mathbf{G}_n(\mathbf{r}°))}{\lambda_n + \mu} \lambda_n\, \mathbf{G}_n(\mathbf{r}°) \tag{2-51}$$

If $\mu = 0$ then Ex. 2-51 yields $\mathbf{F}_\mu(\mathbf{r}°)|_{\mu=0} = \sum_n (\mathbf{F}_0(\mathbf{r}°), \mathbf{G}_n(\mathbf{r}°))\, \mathbf{G}_n(\mathbf{r}°)$, which coincides with the orthogonal projection $\mathbf{F}_\sigma(\mathbf{r}°)$ of the desired ARP $\mathbf{F}_0(\mathbf{r}°)$ onto the subspace $\mathscr{F}_\nu$: $\mathbf{F}_\sigma(\mathbf{r}°) = \sum_n (\mathbf{F}_0(\mathbf{r}°), \mathbf{G}_n(\mathbf{r}°))\, \mathbf{G}_n(\mathbf{r}°)$. In other words, the pattern $\mathbf{F}_\sigma(\mathbf{r}°)$ is the best (in sense of the RMSD) approximation, which the considered antenna can provide to the desired pattern $\mathbf{F}_0(\mathbf{r}°)$.

To study the quality of regularized solution we will analyze how its preciseness and reactivity depend on the value of Lagrange multiplier μ. So as not to be dependent on the orthogonal component $\mathbf{F}_o^\perp(\mathbf{r}°)$ of the desired ARP, calculating RMSD σ as a deviation from the best approach $\mathbf{F}_\sigma(\mathbf{r}°)$ is worthwhile. By the way, this will also be useful because its range of values is clear in advance: between zero, if the solution is the most accurate, and a unit, if the solution itself is null. Thus, the following formula will be in use

$$\sigma^2 = \| \mathbf{F}_\sigma(\mathbf{r}°) - \mathbf{F}_\mu(\mathbf{r}°) \|^2 \,/\, \| \mathbf{F}_\sigma(\mathbf{r}°) \|^2 =$$
$$\sum_n | a_n |^2 [1 - \lambda_n^2 / (\lambda_n + \mu)^2] \,\Big/\, \sum_n | a_n |^2\,. \tag{2-52}$$

Here, $a_n = (\mathbf{F}_0(\mathbf{r}°), \mathbf{G}_n(\mathbf{r}°))$ are the expansion coefficients of the pattern $\mathbf{F}_\sigma(\mathbf{r}°)$ in the orthonormal basis $\{\mathbf{G}_n(\mathbf{r}°)\}$.

To study the regularization effectiveness in detail let as consider synthesizing a circular—cylindrical, more precisely—antenna of Huygens elements in free space. Assume that the radius of circle ρ_0 is equal to 1.5 λ and the desired ARP is a δ-function in direction $\varphi = 0°$.

Several simplifying comments before proceeding to the calculations

First, in the case under consideration, theoretically, there are an infinite number of terms in any of the series Ex. 2-50 ÷ 2-52. However, from computational constraints, not to mention practical reasons, this number $N_{\max}$ must be finite. Say, one hundred or even dozens can be satisfactory.

Secondly, taking into account the parity of the task, which leads to the equality of the coefficients for the harmonics corresponding values $+n$ and $-n$, it is advantageous to combine these harmonics by going over to the

cosine decompositions. This will halve the number of the series' terms and transform Ex. 2-34 to the followings:

$$\{g_n(\alpha) = \cos(n\,\alpha)\,/\,\sqrt{\varepsilon_{0n}\,\pi\rho_0}\}, \quad \{G_n(\varphi) = \cos(n\,\varphi)\,/\,\sqrt{\varepsilon_{0n}\,\pi}\},$$

$$\{\lambda_n = |J_n(k\rho_0) + jJ_n'(k\rho_0)|\}, \qquad \varepsilon_{0n} = \begin{cases} 2 & \text{if } n = 0 \\ 1 & \text{if } n \neq 0 \end{cases}. \tag{2-53}$$

To calculate the derivation $J_n'(k\rho_0)$, the equation $J_n'(k\rho_0) = [J_{n-1}(k\rho_0) - J_{n+1}(k\rho_0)]/2$ is usually used, known in the Bessel functions theory [23].

Thirdly, in accordance with the above, the projection $F_\sigma(\varphi)$ of the δ-function onto the space $\mathcal{F}_v$ substitutes the desired ARP. Taking into account Ex. 2-53 and the fact that $a_n = (\delta(\varphi - 0), G_n(\varphi)) = \dfrac{1}{\sqrt{\varepsilon_{0n}\,\pi}}$ we can write the following formulas for the desired ARP $F_\sigma(\varphi)$, the regularized distribution $I_\mu(\alpha)$, the reproduced ARP $F_\mu(\varphi)$ and their characters σ and Q:

$$F_\sigma(\varphi) = \sum_{n=0}^{N_{\max}} \frac{\cos(n\varphi)}{\varepsilon_{0n}}, \quad \lambda_n^2 = J_n^2(k\rho_0) + \frac{[J_{n-1}(k\rho_0) - J_{n+1}(k\rho_0)]^2}{4},$$

$$I_\mu(\alpha) = \sum_{n=0}^{N_{\max}} \frac{\sqrt{\lambda_n}}{\lambda_n + \mu} \frac{\cos(n\alpha)}{\varepsilon_{0n}}, \qquad F_\mu(\varphi) = \sum_{n=0}^{N_{\max}} \frac{\lambda_n}{\lambda_n + \mu} \frac{\cos(n\varphi)}{\varepsilon_{0n}}, \qquad (2\text{-}54)$$

$$\sigma^2 = \frac{\displaystyle\sum_{n=0}^{N_{\max}} \frac{\mu^2}{\varepsilon_{0n}(\lambda_n + \mu)^2}}{N_{\max} + 0.5} = \begin{cases} 0, & \text{if } \mu = 0 \\ 1, & \text{if } \mu \to \infty \end{cases}, \quad Q = \frac{\displaystyle\sum_{n=0}^{N_{\max}} \frac{\lambda_n}{\varepsilon_{0n}(\lambda_n + \mu)^2}}{\displaystyle\sum_{n=0}^{N_{\max}} \frac{\lambda_n^2}{\varepsilon_{0n}(\lambda_n + \mu)^2}}.$$

Finally, the reactivity of eigenfunctions $\{g_n(\alpha)\}$ is obviously inversely proportional to the eigenvalues: $Q_n = \|g_n(\alpha)\|^2 / \|Ug_n(\alpha)\|^2 = 1/\lambda_n$. Thus, the minimal and maximal reactivity are inherent in the eigenfunctions corresponding to $\lambda_{\max}$ and $\lambda_{\min}$, respectively. The reactivity of any other distribution $\mathbf{I}(\boldsymbol{\rho}) = \sum_n a_n \mathbf{g}_n(\boldsymbol{\rho})$ is between those limits: $Q_{\min} = 1/\lambda_{\max}$ and $Q_{\max} = 1/\lambda_{\min}$. In the case under consideration, the limits are $Q_{\min} = 3.92$ and $Q_{\max} = 200.52$.

Calculations

In view of the structure of the spectrum $\{\lambda_n\}$ shown in Fig. 2-5d, in order not to complicate displaying the calculation results in detail, we will restrict the number of terms in the expansions by $N_{\max} = 25$. Fig. 2-8a shows dependences of accuracy σ or $\underline{\sigma}$ and reactivity Q—the crucial characteristics—of solution $I_\mu(\alpha)$ obtained while the Lagrange multiplier μ varies over a wider range than usual.
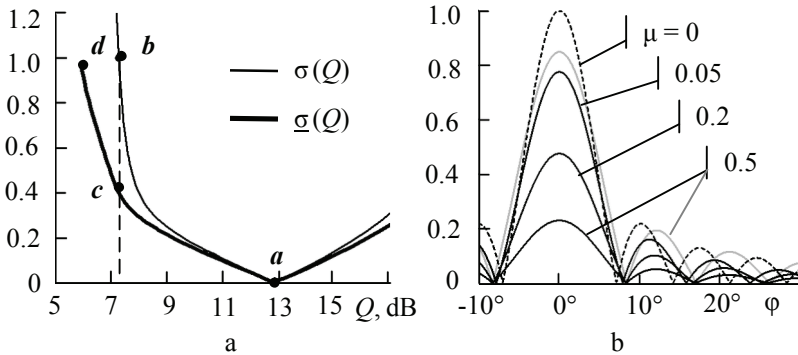
Figure 2-8: Regularization of the antenna synthesis solution: a) the dependences of standard deviations σ and σ̲ on the reactivity Q in decibel; b) ARPs $|F_\mu(\varphi)|$ corresponding to four values of the Lagrange multiplier μ

The common range of μ variation is from 0 (the point a: $\sigma|_{\mu=0} = 0$; $Q|_{\mu=0} = 19.25$) to ∞ (the point b: $\sigma|_{\mu=\infty} = 1$; $Q|_{\mu=\infty} = 5.34$). The exact solution $I_\mu(\alpha)|_{\mu=0}$ (point a) can be unacceptable due to high reactivity, whereas the solution $I_\mu(\alpha)|_{\mu=\infty}$ seems entirely useless with regard to approaching the desired ARP (point b with $\sigma = 1(!)$). However, this follows from Ex. 2-54 that if the value of μ is significantly higher than λ_{max}, then the relations $I_\mu(\alpha) \approx \dfrac{1}{\mu} \sum\limits_{n=0}^{N_{max}} \sqrt{\lambda_n} \dfrac{\cos(n\,\alpha)}{\varepsilon_{0n}} = \dfrac{1}{\mu} p_0(\alpha)$ are in force. Here, $p_0(\alpha) = V\,F_\sigma(\varphi)$, as we shall see in the next section, is a fairly good solution. The problem behind the value of $\sigma = 1$ for the point b is that the solution $I_\mu(\alpha)|_{\mu\to\infty}$ becomes vanishingly small due to scale factor $1/\mu$.

In Fig. 2-8b, there are four patterns $F_\mu(\varphi)$ the first of which corresponds to the value $\mu = 0$, i.e. is exactly the pre-assigned ARP $F_\sigma(\varphi)$. The dashed line highlights it. When the multiplier μ increases to the level near the value of the minor eigenvalues, the contribution of the "high-frequency" terms corresponding to large values of n and small values of λ_n, decreases in accordance with the factor $\lambda_n/(\lambda_n + \mu)$. This leads to some expansion of the main lobe. After μ becomes larger than λ_{max}, the shape of the pattern $F_\mu(\varphi)$ practically does not alter. Only its level reduces inversely proportional to the value of μ: $F_\mu(\varphi) \approx U p_0(\varphi) / \mu$. To make this pictorially clear, in Fig. 2-8b the curve in the light gray color displays this pattern after optimal scaling in accordance with the aforementioned factor $((F_0(\varphi), F_\mu(\varphi))/\|F_\mu(\varphi)\|^2$. In doing so, the square deviation $\underline{\sigma}^2$ becomes equal to 0.16 that is not so bad.

At the same variation of μ from 0 to ∞, the curve of dependency $\underline{\sigma}(Q)$ starts at the point **a** (Fig. 2-8b) and ends at the point **c** ($\underline{\sigma}|_{\mu=\infty} = 0.401$; $Q|_{\mu=\infty} = 5.34$). Note that the solution $I_\mu(\alpha)$ remains unchanged, but the value of the corresponding deviation $\underline{\sigma}$, determined by Ex. 2-48, is significantly smaller than σ, due to optimal scaling of the patterns $F_\mu(\varphi)$.

The range of reactivity inherent to the considered system of sources—from $Q_{min} = 3.92$ to $Q_{max} = 200.52$—is wider than the range $5.34 \div 19.25$ to which the reactivity of regularized solutions $I_\mu(\alpha)$ belongs. It is interesting how the dependences of $\sigma(Q)$ and $\underline{\sigma}(Q)$ behave in the entire range of Q. To find this out, we need to solve the following constrained optimization problem: to minimize $\|F_\sigma(r^o) - U\,I(\alpha)\|^2$ while satisfying the condition $\|I(\alpha)\|^2 / \|U\,I(\alpha)\|^2 = Q$.

Rewriting the condition in the form $Q\,\|U\,I(\alpha)\|^2 - \|I(\alpha)\|^2 = 0$, and using the method of Lagrange multipliers reduces the optimization problem with the restriction on reactivity Q to the same equation given by Ex. 2-49. The difference is that the μ factor can now be negative.

If the value of μ varies from zero to $-\lambda_{min}$, then the solution reactivity rises from the value $Q|_{\mu=0} = 19.25$ to the maximum possible value of $Q_{max} = 1/\lambda_{min} = 200.52$, while the accuracy deteriorates. Obviously, these solutions are of no practical interest. Therefore, in Fig. 2-8a, corresponding portions of the curves $\sigma(Q)$ and $\underline{\sigma}(Q)$ to the right of the point **a** are present on the truncated interval only for the sake of universality.

If the value of μ varies from $-\infty$ to $-\lambda_{max}$, then the reactivity of the solution decreases from the value of $Q|_{\mu=\infty} = 5.34$ to the minimum value of $Q_{min} = 3.92$, while the accuracy deteriorates. The portion of the curve $\underline{\sigma}(Q)$ between points **c** and **d** ($\underline{\sigma}|_{\mu=-\lambda_{max}} = 0.98$; $Q|_{\mu=-\lambda_{max}} = 3.92$) displays the quality of the corresponding solutions $I_\mu(\alpha)$. Since low reactivity is a very attractive factor, a solution with $Q < 5.34$—to the left, but, of course, near the point **c**—can be of interest.

The values of the variables σ and $\underline{\sigma}$ for the regularized solutions $I_\mu(\alpha)$ differ significantly for the most important type of low-reactive distributions. It is clear that the variable $\underline{\sigma}$ gives a much more accurate evaluation of the solution quality, and the variable σ unreasonably provokes us to use more reactive distributions.

2.5.3. Solution optimization with consideration for random errors of its subsequent implementation

Regularization, as a method for solving ill-posed or ill-conditioned problems, constructs such an algorithm that ensures the solution's stability

according to Hadamard. In practice, this means that small changes on the right-hand side of the equation could result in rather small, or at least not drastic, alterations in the solution. Regularization perfectly fits inverse tasks arising in many fields of technology, in particular, for any kind of indirect measurements.

Bearing in mind the antenna technique, we can refer to retrieving the antenna current distribution $\mathbf{I}_a(\boldsymbol{\rho})$ from the measured[18] ARP $\mathbf{F}_a(\mathbf{r}^o)$, which reduces to the solution of the following equation U $\mathbf{I}_a(\boldsymbol{\rho}) = \mathbf{F}_a(\mathbf{r}^o)$. Since instrumental errors $\delta\mathbf{F}_a(\mathbf{r}^o)$ distort the measurement data $\mathbf{F}_a(\mathbf{r}^o)$, it is vital to restrict the sensitivity of the retrieval algorithm towards random distortions $\delta\mathbf{F}_a(\mathbf{r}^o)$. The regularization paves the way to do so.

The situation with antenna synthesis, which deals with the same equation U $\mathbf{I}_\sigma(\boldsymbol{\rho}) = \mathbf{F}_0(\mathbf{r}^o)$, see Ex. 2-17, is markedly different from the above. The right part of this equation is the desired optimal ARP $\mathbf{F}_0(\mathbf{r}^o)$, uniquely determined by the requirements for the operation of the radio system to which this antenna serves. Suppose that Ivan and Michael do not know each other. They develop their own radio system and asked Diego—the mathematician—to solve the relevant Ex. 2-17. It so happened that the optimal patterns for both systems turned out to be very close.

Obviously, Ivan does not care whether the solution given to Michael will be close or substantially different from the solution that he received from Diego. Just like Michael is indifferent to the solution for Ivan. Maybe Diego would be not pleased obtaining essentially different solutions for very close tasks. Being a mathematician and following Hadamard, he desires that the behavior of the solution changes continuously with the initial conditions, and perhaps he used regularization.

If so, what reasoning did he use to choose the value of the Lagrange multiplier μ? With the indirect measurements mentioned above, when we know the level of instrumental errors changing the right side of the equation, even then the choice of the value of μ is such a vague task that after all, we have to make a heuristic decision based on the previous experience.

By the way, it looks absurd, but if the desired ARP $\mathbf{F}_0(\mathbf{r}^o)$ coincides with the eigenfunction $\mathbf{G}_m(\mathbf{r}^o)$, which corresponds to the minimal eigenvalue $\lambda_m = \lambda_{\min}$ and, accordingly, to the worst distribution $\mathbf{g}_m(\boldsymbol{\rho})$, then, according to the Hadamard concept of sustainability, the problem of antenna synthesis acquires the highest stability. Indeed, let the dimension of the task is limited by N (say, we deal with an antenna array of N elements).

---

[18] The measurements can take place in the Fraunhofer zone that allows the use of a compact chamber.

Assume that distortion $\delta\mathbf{F}(\mathbf{r}^{o}) = \sum_{n}\alpha_{n}\mathbf{G}_{n}(\mathbf{r}^{o})$ of the right side of the synthesis equation Ex. 2-8 is a random function, which coefficients $\{\alpha_{n}\}$ are statistically independent with zero mean and equal variances $\sigma^{2} = \varepsilon^{2}/N$. Then, as a consequence of Ex. 2-22, the variance $\overline{\|\delta\mathbf{I}(\boldsymbol{\rho})\|}^{2}$ of distortions of the solution $\delta\mathbf{I}(\boldsymbol{\rho})$ is $\varepsilon^{2} \sum_{n}(1/\lambda_{n})$. The solution $\mathbf{I}(\boldsymbol{\rho}) = U^{-1} \mathbf{G}_{m}(\mathbf{r}^{o})$ is, obviously, equal to $\mathbf{g}_{m}(\boldsymbol{\rho})/\sqrt{\lambda_{m}}$. Therefore, the relative level of the solution distortions is $\overline{\|\delta\mathbf{I}(\boldsymbol{\rho})\|}^{2} /\| \mathbf{I}(\boldsymbol{\rho}) \|^{2} = \lambda_{m} \varepsilon^{2} \sum_{n}(1/\lambda_{n})$. This equality means that the smaller the value of λ_{m}, the higher the stability of the solution Ex. 2-8 (in the sense of the Hadamard concept). The worse the desired pattern $\mathbf{F}_{0}(\mathbf{r}^{o})$, the better. Is not it discouraging?

As for Ivan and Michael, they do have an interest in the stability of the solution of equation Ex. 2-17, however, this stability is in the opposite direction: from left to right. This is because at synthesizing antenna, the source of deviations is not the deformation of the desired ARP $\mathbf{F}_{0}(\mathbf{r}^{o})$, but distortion the nominal current distribution $\mathbf{I}_{\varepsilon}(\boldsymbol{\rho})$ due to unavoidable random errors $\delta\mathbf{I}(\boldsymbol{\rho})$ accompanying its implementation.

In these circumstances, the current ARP formed by the antenna is $\mathbf{F}(\mathbf{r}^{o}) = U (\mathbf{I}_{\varepsilon}(\boldsymbol{\rho}) + \delta\mathbf{I}(\boldsymbol{\rho})) = \mathbf{F}_{\varepsilon}(\mathbf{r}^{o}) + \delta\mathbf{F}(\mathbf{r}^{o})$, where $\mathbf{F}_{\varepsilon}(\mathbf{r}^{o}) = U \mathbf{I}_{\varepsilon}(\boldsymbol{\rho})$ is the nominal ARP and $\delta\mathbf{F}(\mathbf{r}^{o}) = U \delta\mathbf{I}(\boldsymbol{\rho})$ is a random deviation from it due to errors $\delta\mathbf{I}(\boldsymbol{\rho})$. Therefore, the standard deviation σ determined by Ex. 2-47, is a random variable.

Clearly, the level of errors $\delta\mathbf{I}(\boldsymbol{\rho})$ is not an absolute value, but a relative value. In other words, the variance of errors $\overline{\| \delta\mathbf{I}(\boldsymbol{\rho}) \|}^{2}$ —here and further, the bar over the variable denotes the mathematical expectation—is proportional to the square of the nominal distribution norm as follows

$$\xi^{2} = \overline{\| \delta\mathbf{I}(\boldsymbol{\rho}) \|}^{2} = \varepsilon^{2} \| \mathbf{I}_{\varepsilon}(\boldsymbol{\rho}) \|^{2} . \tag{2-55}$$

Here, ε is a percentage accuracy of realization.

It is of great practical interest to find such a nominal solution $\mathbf{I}_{\varepsilon}(\boldsymbol{\rho})$, which provides the minimum mean value of the standard deviation σ ARP, while the errors $\delta\mathbf{I}(\boldsymbol{\rho})$ have a certain level ε. Below I discuss the approach and the results very close to those obtained by Shifrin Ya.S. and Kornienko L.G., who dealt with this problem in the framework of the statistical theory of antennas [24]. The fact is that we all conducted our studies independently and contemporaneously.

Let us write the random function $\delta\mathbf{I}(\boldsymbol{\rho})$ in the form of the sum $\sum_{n}\alpha_{n} \mathbf{g}_{n}(\boldsymbol{\rho})$ where coefficients $\{\alpha_{n}\}$ correspond to a widely used statistical model hat embodies following assumptions: the coefficients are

statistically independent complex values with zero means and equal variances of their real and imaginary parts. In particular, this means that in the case of the antenna array, the errors $\{\delta\mathbf{I}_n(\boldsymbol{\rho})\}$ of the excitation coefficients of its elements are also, as usual, independent complex values with equal variances and zero means. It is clear that because of Ex. 2-55 the variance of any random coefficient α_n is

$$\overline{|\alpha_n|^2} = \overline{\|\delta\mathbf{I}(\boldsymbol{\rho})\|^2} / N = \varepsilon^2 \|\mathbf{I}_\varepsilon(\boldsymbol{\rho})\|^2 / N \ . \tag{2-56}$$

With an eye on the mean square deviation (MSD) σ^2 (Ex. 2-47), let us write the obvious equation $\|\mathbf{F}_0(\mathbf{r}^o)-\mathbf{F}(\mathbf{r}^o)\|^2 = \|\mathbf{F}_0(\mathbf{r}^o)-\mathbf{F}_\varepsilon(\mathbf{r}^o)\|^2 - 2\operatorname{Re}[(\mathbf{F}_0(\mathbf{r}^o)-\mathbf{F}_\varepsilon(\mathbf{r}^o), \delta\mathbf{F}(\mathbf{r}^o))] + \|\delta\mathbf{F}(\mathbf{r}^o)\|^2$. It is clear that the average value of the middle term on the right-hand side of the equation is equal to zero. Taking into account the equalities $\delta\mathbf{F}(\mathbf{r}^o) = \mathrm{U}\,\delta\mathbf{I}(\boldsymbol{\rho}) = \sum_n \alpha_n \sqrt{\lambda_n}\,\mathbf{G}_n(\mathbf{r}^o)$

and Ex. 2-55, we obtain the following result for a variance $\overline{\sigma^2}$ of the variable σ:

$$\overline{\sigma^2} = \sigma_\varepsilon^2 + \varepsilon^2 \frac{\|\mathbf{I}_\varepsilon(\boldsymbol{\rho})\|^2}{\|\mathbf{F}_0(\mathbf{r}^o)\|^2} \lambda_{\mathrm{avr}} \ . \tag{2-57}$$

Here, $\sigma_\varepsilon = \|\mathbf{F}_0(\mathbf{r}^o)-\mathbf{F}_\varepsilon(\mathbf{r}^o)\| / \|\mathbf{F}_0(\mathbf{r}^o)\|$ is the root-mean-square deviation of the realized ARP $\mathbf{F}_\varepsilon(\mathbf{r}^o)$ from the desired ARP $\mathbf{F}_0(\mathbf{r}^o)$ for the nominal distribution $\mathbf{I}_\varepsilon(\boldsymbol{\rho})$ and $\lambda_{\mathrm{avr}} = \sum_n \lambda_n / N$ is the average value of the eigenvalues $\{\lambda_n\}$.

Ex. 2-57 explicitly asserts that the nominal distribution $\mathbf{I}_\varepsilon(\boldsymbol{\rho})$, which minimize the value $\overline{\sigma^2}$, coincides with the relevant solution of the regularization problem (Ex. 2-50). Indeed, having a certain norm, it must minimize the value of σ_ε, i.e. belong to the set of solutions that Ex 2-50 defines. The difference is that instead of a rather vague restriction on the norm $\|\mathbf{I}_\varepsilon(\boldsymbol{\rho})\|$, now with regard to the specified level of errors ε^2, the condition of minimization of the value $\overline{\sigma^2}$ uniquely defines the exact value of regularization multiplier μ. What do you think about this difference? Is it significant or trifling?

To clarify the situation, Fig. 2-9a shows the dependencies $\overline{\sigma^2}(Q)$ calculated using Ex. 2-57 and the regularized distribution Ex. 2-50 as the nominal solution $\mathbf{I}_\varepsilon(\boldsymbol{\rho})$. At that, we considered the same situation as in the Section 2.4.2 (see the paragraph "Calculations" in it), but with a larger number of angular harmonics $N = 28$ for obtaining more expressive dependencies.
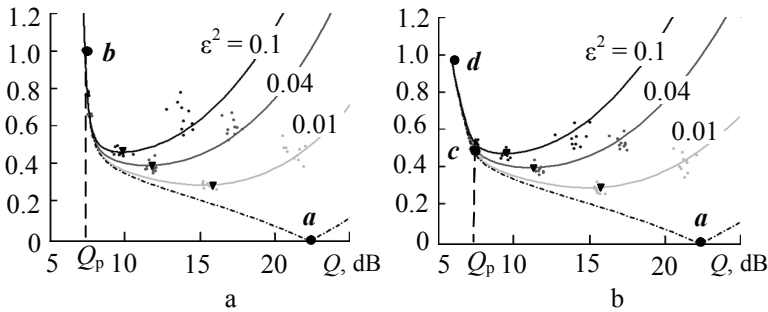
Figure 2-9: The standard deviation of the σ and σ̲ variables (a and b, respectively) as a function of the reactivity of the antenna synthesis solution for three levels of accuracy: black, dark gray and light gray, respectively, to a value of ε2 equal to 0.1, 0.04 and 0.01

The value of the Lagrange multiplier μ varies in the same range as in the case of Fig. 7 and the reactivity Q of the regularized solution Ex. 2-50 is the abscissa of the chart. The dash-dotted line shows the dependence of the root-mean-square deviation σ of ARP for the nominal solution $\mathbf{I}_\varepsilon(\boldsymbol{\rho})$ (as if ε^2 is equal to 0). This dependence corresponds to that in Fig. 2-8a, and the points a, b mark the same situations: $\mu = 0$, $\mu = \infty$, respectively. The reactivity $Q_p = 7.28$ dB corresponds to the point b and characterizes a very attractive solution discussed in the next chapter.

In Fig. 8a, three solid lines of different shade shows dependencies $\overline{\sigma^2}(Q)$ for the values of error variance $\varepsilon^2 = \|\delta\mathbf{I}(\boldsymbol{\rho})\|^2/\|\mathbf{I}(\boldsymbol{\rho})\|^2$ equal to 0.1, 0.04 and 0.01, respectively. The small triangles on them mark the optimal solutions minimizing the mean value $\overline{\sigma^2}$ if the implementation accuracy corresponds to a certain level (given by ε^2).

Of course, the mathematically average value of σ is a decisive parameter. However, it is interesting to see the features according to some cases of reproducing the nominal solution while errors of a certain level are inevitable. To satisfy this desire, we simulated ten random realizations $\delta\mathbf{I}(\boldsymbol{\rho})$ having a normal distribution of the coefficients $\{a_n\}$ with zero mean and a given variance (Ex. 2-56). In Fig. 2-9a, near each of the three curves, you can see three clusters of ten points that represent the results for ten random implementations of the following solutions: the optimal solution (central cluster), the solution with a higher reactivity (cluster on the right),

and the solution with less reactivity (barely noticeable[19] left cluster). Fig 2-8a demonstrates that the results of direct statistical simulations are in a good agreement with the dependencies predicted by Ex. 2-57.

Let us carry out a similar analysis of the statistical characteristics of the variable $\underline{\sigma}$ determined by Ex. 2-48. Denoting $\xi^2 = \varepsilon^2 \, \|\mathbf{I}(\boldsymbol{\rho})\|^2/N$ and $P_0 = \sum_n \lambda_n \, |a_n|^2$ we can easily obtain the following formulas $\overline{|(\mathbf{F}_0(\mathbf{r}^\circ),\ \mathbf{F}(\mathbf{r}^\circ)+\delta\mathbf{F}(\mathbf{r}^\circ))|^2} = |(\mathbf{F}_0(\mathbf{r}^\circ),\ \mathbf{F}(\mathbf{r}^\circ))|^2 + \xi^2 \, P_0$ and $\overline{\|\mathbf{F}(\mathbf{r}^\circ)+\delta\mathbf{F}(\mathbf{r}^\circ)\|^2} = \overline{(\mathbf{F}(\mathbf{r}^\circ)+\delta\mathbf{F}(\mathbf{r}^\circ),\mathbf{F}(\mathbf{r}^\circ)+\delta\mathbf{F}(\mathbf{r}^\circ))} = \|\mathbf{F}(\mathbf{r}^\circ)\|^2 + \xi^2 \sum_n \lambda_n$,

which lead to the expression $\overline{\sigma^2} \cong 1 - \dfrac{|(\mathbf{F}_0(\mathbf{r}^\circ),\ \mathbf{F}(\mathbf{r}^\circ))|^2 + \xi^2 P_0}{\|\mathbf{F}_0(\mathbf{r}^\circ)\|^2 \, (\|\mathbf{F}(\mathbf{r}^\circ)\|^2 + \xi^2 \sum_n \lambda_n)}$.

Finally, assuming that the value of ξ^2 is rather small, we can write

$$\overline{\sigma^2} \cong 1 - \frac{|(\mathbf{F}_0(\mathbf{r}^\circ),\mathbf{F}(\mathbf{r}^\circ))|^2}{\|\mathbf{F}_0(\mathbf{r}^\circ)\|^2 \|\mathbf{F}(\mathbf{r}^\circ)\|^2}(1-\varepsilon^2 Q\,\lambda_{\mathrm{avr}}) - \varepsilon^2 Q \frac{P_0}{N\,\|\mathbf{F}_0(\mathbf{r}^\circ)\|^2}. \qquad (2\text{-}58)$$

Fig. 2-9b shows results of calculations by formula Ex. 2-58 with data similar to those shown in Fig. 2-9a. Points *a*, *c* and *d* on the dash-dotted curve correspond to the same points in Fig. 2-8b. It is clear that the dependences in Fig. 2-9a and Fig. 2-9b are very similar because parameters σ and $\underline{\sigma}$ are inherently close. However, in the range of moderate values of reactivity Q (about Q_{p}) that is most important for practice, they are clearly different. It is worth being aware that the value of the parameter σ (Ex. 2-47) "suffers" from a decrease in the norm of the solution (due to the large value of μ) and unjustifiably encourages us to use solutions with a higher reactivity. Therefore, using the parameter $\underline{\sigma}$ (Ex. 2-48) for statistical optimization is more reliable, since it is free from this disadvantage.

The discussed problem closely relates to the so-called super directivity problem. The point is this. As the number of eigenfunctions increases—for an antenna array, it coincides with the number of antenna elements—the accuracy of approaching a desired ARP $F_0(\theta)$ increases even with other conditions being the same.

Consider a linear antenna array. Let $F_0(\theta)$ be the δ-function in the direction normal to the array. The best practical solution is an in-phase distribution with equal amplitudes. The narrowest main beam corresponds to it. Nevertheless, with a large number N, the solution of the synthesis problem

---

[19] The less reactivity the closer the points of the cluster are located. This is why each left cluster looks like a slightly expanded spot.

gives a complex amplitude-phase distribution, which corresponds to a narrower beam. That is why such a solution is called super directed. From the above, it is clear that this solution is extremely sensitive to random errors of implementation. The example shown in Appendix A of the Mathcad program "Super-Directionality and Errors" allows you to explore the problem. I believe that the reader who decides not to miss this opportunity will get a certain pleasure from a visual acquaintance with an interesting phenomenon.

2.6. Local amendments of the antenna radiation pattern

The mean square approach to the desired ARP $\mathbf{F}(\mathbf{r}°)$ that solution Ex. 2-19 provides does not always meet the requirements in full. Most often, this refers to the level of side lobes. What can we do to improve the solution?

Of course, we can define the inner product underlying the synthesis task and its solution Ex. 2-19, using Ex. 1-4′, and increase the value of the weight function $\rho(\mathbf{r}°)$ for those areas where side lobes caused our concern. However, this way is rather vague. The strategy is clear, but it is uncertain a priori how much the increase in the values of the function $\rho(\mathbf{r}°)$ will affect the side lobes. The only possibility on this path is a tedious process of trial and error.

Another way to solve a problem stems from a sensible and practical idea that everyone can easily generate. In order to change the ARP in a certain narrow region, leaving it almost intact outside the region, it is enough to add a beam-like ARP with a proper complex multiplier or a set of those ARPs, if necessary.

For simplicity's sake let us consider the task in scalar formulation. For antenna of arbitrary geometry, we can get a beam $g_m(\mathbf{r}°)$ pointing in a certain direction $\mathbf{r}_m^{\circ}$ by synthesizing the desired radiation pattern, which is the delta function $\mathrm{F}_0(\mathbf{r}°) = \delta(\mathbf{r}° - \mathbf{r}_m^{\circ})$. At that, the adjoint operator V gives a very attractive approximate solution $p(\boldsymbol{\rho}) = \mathrm{V}\,\mathrm{F}_0(\mathbf{r}°) = f^*(\boldsymbol{\rho}, \mathbf{r}_m^{\circ})$ that completely accords to a reasonable physical principle: in order to get a maximum at the direction $\mathbf{r}_m^{\circ}$, the excitation distribution $p(\boldsymbol{\rho})$ should provide the in-phase addition of the fields of all antenna elements at this direction. Thus, the following formula represents a wanted beam:

$$g_m(\mathbf{r}°) = \mathrm{U}\,p(\boldsymbol{\rho}) = \begin{cases} \int_v f^*(\boldsymbol{\rho}, \mathbf{r}_m^{\circ}) f(\boldsymbol{\rho}, \mathbf{r}°)\, dv \\ \sum_n f_n^*(\mathbf{r}_m^{\circ}) f_n(\mathbf{r}°) \end{cases}. \tag{2-59}$$

Here v is the area of antenna sources—a line or a surface, for instance—$f_n(\mathbf{r}°)$ is ARP of the n-th element of the antenna array.

Assume the ARP that needs to be improved is $F(\mathbf{r}°)$ and the maxima of its side lobes are at a set of directions $\{\mathbf{r}_m°\}$ ($m = 1\ldots M$). Then M conditions are obvious:

$$\left| F(\mathbf{r}_{m'}°) + \sum_{m=1}^{M} a_m g_m(\mathbf{r}_{m'}°) \right| = \beta , \qquad (2\text{-}60)$$

where m' is the same index as m, $\{a_m\}$ are sought-for coefficients, and β is the limitation value, form a system of M second-degree equations—after squaring, naturally. To work around the difficulties connected with solving the system Ex. 2-60, we can rearrange conditions in the following manner

$$\sum_{m=1}^{M} a_m g_m(\mathbf{r}_{m'}°) = -\left(|F(\mathbf{r}_{m'}°)| - \beta \right) F(\mathbf{r}_{m'}°) / |F(\mathbf{r}_{m'}°)| . \qquad (2\text{-}61)$$

The idea behind Ex. 2-61 is simple: to get the sum of the beams, the amplitude of which in the $\{\mathbf{r}_{m'}°\}$ directions is equal to the needed difference $|F(\mathbf{r}_{m'}°)| - \beta$, and the phase is opposite to the phase $F(\mathbf{r}_{m'}°)$. Ex. 2-61 is an M order system of linear equations for M unknowns $\{a_m\}$.

Fig. 2-10 shows what the above procedure results in for a linear source $L = 7.5\ \lambda$ and initial ARP $F(\theta) = \text{sinc}((\pi L /\lambda)\sin(\theta))$.[20] In this case, beams are $g_m(\theta) = \text{sinc}[(\pi L /\lambda)(\sin(\theta) - \sin(\theta_m))]$. We decided to reduce five side lobes locating in a set $\{\theta_m\} = \{-27.8°, -19.5°, -11.5°, 11.5°, 19.5°\}$—why not—to the level $\beta = 0.0562 = -25$ dB. The above starting directions $\{\theta_m\}$ after minor corrections turned to be $\{-24°, -17°, -11.5°, 11.5°, 17°\}$.
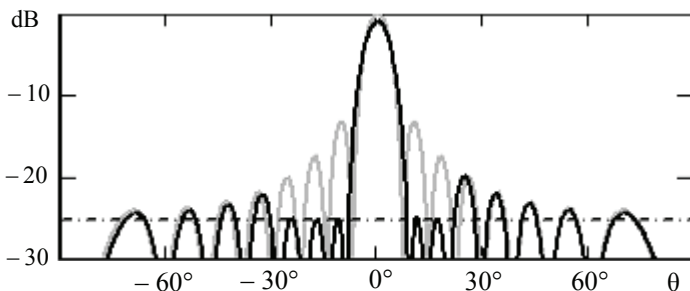
Figure 2-10: Reducing five side lobes of a linear 20λ-source to the level –25 dB: gray and black lines show initial and amended ARPs, respectively

---

[20] The so-called sinc-function is $\text{sinc}(x) = \sin(x) / x$.

Fig. 2-11 shows analogical results for a circle antenna array of twenty elements, the individual ARP of which are radial oriented cardioids $f_n(\theta) = (1 + \cos(\theta - \alpha_n)) \exp(ja \cos(\theta - \alpha_n))$. The spacing between elements is equal to 0.45 λ, so radius $a = 0.225\, \lambda\, N\, /\pi$.
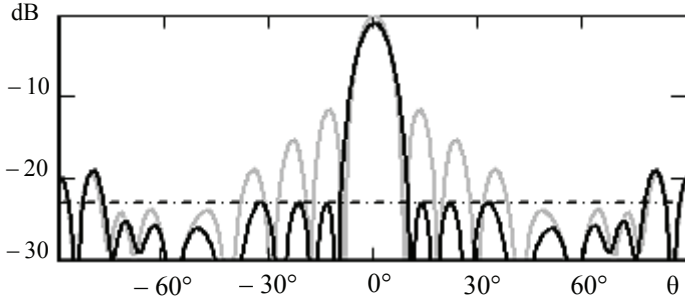
Figure 2-11: Reducing six side lobes of a circle array of twenty elements with spacing 0.45λ to the level β = –23dB: gray and black lines show initial and amended ARPs, respectively

Ex 2-59 determines the beams $\{ g_m(\theta) = \sum_n f_n^*(\theta_m) f_n(\theta) \}$. The initial ARP was the beam at the direction $\theta_m = 0$. The level of restriction for six side lobes was equal to –23 dB. The same as above, a starting set of directions $\{\theta_m\}$ was equal to positions of the initial ARP maxima, and got minor corrections in course of calculations.

The interested reader can use the programs "Reducing the Level of a Set of Side Lobes of Antenna Array" from Appendixes B and C to explore the problem and understand why the correction is useful.

2.7. Training tasks

Task 2A

Try to prove that for any purely radiating distribution $\mathbf{I}(\boldsymbol{\rho}) \subset \mathcal{J}_\Sigma$ there is such a pattern $\mathbf{F}(\mathbf{r}°)$ that the adjoint operator V maps into $\mathbf{I}(\boldsymbol{\rho})$. In other words, this means that the operator V maps the space $\mathcal{F}$ not into subspace $\mathcal{J}_\Sigma$, but on it—as a whole.

A hint: Use proof by contradiction assuming that $\mathbf{I}(\boldsymbol{\rho})$ is orthogonal to subspace, to which the operator V maps the space F. This assumption equal to the equation $(\mathbf{I}(\boldsymbol{\rho}), \mathbf{V}\mathbf{F}(\mathbf{r}°)) = 0$, from which you can logically come to the needed statement.

Task 2B

Let distributions $\mathbf{I}_1(\boldsymbol{\rho})$ and $\mathbf{I}_2(\boldsymbol{\rho})$ differ, but their ARPs be the same $\mathrm{U}\mathbf{I}_1(\boldsymbol{\rho}) = \mathrm{U}\mathbf{I}_2(\boldsymbol{\rho})$. Certainly, $\Delta\mathbf{I}(\boldsymbol{\rho}) = \mathbf{I}_1(\boldsymbol{\rho}) - \mathbf{I}_2(\boldsymbol{\rho})$ is a non-radiating

distribution. Assume that functions $\mathbf{I}_1^{\perp}(\boldsymbol{\rho})$ and $\mathbf{I}_2^{\perp}(\boldsymbol{\rho})$ are obtained as a result of orthogonalization the initial functions towards the $\Delta\mathbf{I}(\boldsymbol{\rho})$. The question is "Are the functions $\mathbf{I}_1^{\perp}(\boldsymbol{\rho})$ and $\mathbf{I}_2^{\perp}(\boldsymbol{\rho})$ purely radiating or not?"

Task 2C
Prove that the pair of orthogonal bases $\{\mathbf{g}_n(\boldsymbol{\rho})\}$ and $\{\mathbf{G}_n(\boldsymbol{\rho})\}$ in the subspaces $\mathcal{J}_\Sigma$ and $\mathcal{F}_v$ are the only pair of orthogonal bases mutually interconnected by the operators U and V, as shown in Ex. 2-20 and 2-21.
A hint: Assume that all the eigenvalues $\{\lambda_n\}$ are different.

Task 2D
Consider a pure 2D situation when a current $\mathbf{I}(\alpha) = I(\alpha)\,\boldsymbol{\alpha}^{\circ}$ flows in free space along a circular cylinder of radius R. Are there any non-radiating distributions $I(\alpha)$ there?

Task 2E
Are there any non-radiating distributions $\{I_n\}$ in the above 2D situation with cosine currents $\{\mathbf{I}_n(\alpha) = I_n \cos(\alpha)\,\boldsymbol{\alpha}^{\circ}\}$ on a set of N circular cylinders of radiuses $\{R_n\}$ $n = 1,\dots N$? If yes, how many linear independent non-radiating distributions exist?

Task 2F
Why, under a constrained minimization of $\underline{\sigma}$, are the conditions $\|\mathbf{I}(\boldsymbol{\rho})\| < I_0$ and $Q < Q_0$ equivalent to each other, while at minimization of σ they are not equivalent?

Task 2G
The bases $\{\mathbf{G}_n(\mathbf{r}^{\circ})\}$ and $\{\mathbf{g}_n(\boldsymbol{\rho})\}$ correspond to an antenna array of N elements, and eigenvalues are in descending order. The required ARP is $F_0(\mathbf{r}^{\circ}) = \sum_n a_n G_n(\mathbf{r}^{\circ})$. A regularized solution to the problem is equal to $\sum_n a_n g_n(\mathbf{r}^{\circ})\sqrt{\lambda_n}\,/(\lambda_n + \mu)$. Write the formulas for σ^2 and $\underline{\sigma}^2$ for two values of the regularization parameter: $\mu = -\lambda_1$ and $\mu = -\lambda_N$.

Task 2H
Prove that Ex. 2-61 is equal to Ex. 2-60.

Task 2I
To obtain the results shown in Fig. 2-10 the set of directions $\{\theta_m\} = \{-27.8^{\circ}, -19.5^{\circ}, -11.5^{\circ}, 11.5^{\circ}, 19.5^{\circ}\}$ corresponding to maxima of the initial ARP, we have changed to $\{\theta_m\} = \{-24^{\circ}, -17^{\circ}, -11.5^{\circ}, 11.5^{\circ}, 17^{\circ}\}$. Why we do so? What would happen if we did not make the changes?

CHAPTER THREE

THE METHODOLOGY OF THE ADJOINT OPERATOR

3.1. Introduction

Dear reader, before you start looking through this chapter, I have to confess that it is my pleasure to explain anything associated with the adjoint operator, which is relevant to antenna synthesis. There are two good reasons why I like it.

First, being a devoted admirer of the engineering style of thinking, I used to seek a deep physical interpretation for any mathematical formalities, much more for such a sophisticated one as the adjoint operator. Of course, often there is only a heuristic way to take a guess about the physical nature hidden behind mathematical formalities. It is especially nice if this happens, as it did with the adjoint operator in my case.

Second, the adjoint operator inherent several very attractive and useful features are inherent to the adjoint operator in addition to mentioned one, which is mapping any arbitrary desired ARP $\mathbf{F}_0(\mathbf{r}^o)$ into the subspace I_Σ of purely radiating distributions.

3.2. Adjoint operator applied to the synthesis of antennas. Energetically optimal solution

The adjoint operator V depends on the direct operator U through the condition Ex. 1-10, which means equality for the following inner products (see Ex. 2-7) in spaces $\mathcal{F}$ and $\mathcal{J}$:

$$(\mathrm{U}\mathbf{I}(\boldsymbol{\rho}), \mathbf{F}(\mathbf{r}^o)) = (\mathbf{I}(\boldsymbol{\rho}), \mathrm{V}\mathbf{F}(\mathbf{r}^o)). \tag{3-1}$$

Thanks to tensor symbolism, Ex. 2-1 and Ex. 2-9 define the operators U and V in the most compact and universal forms. However, their compactness makes it difficult to perceive some physical aspects behind

them. Therefore, let us consider specific antennas, describing them by commonly used clearer formulas.

Assume that the antenna's sources continually locating in area v have a fixed orientation/polarization. Therefore, a scalar complex function $I(\boldsymbol{\rho})$ describes the excitation distribution (amplitude and phase dependencies). Let the radiation pattern of a source located at the point $\boldsymbol{\rho}$ correspond to a vector function $\mathbf{f}(\boldsymbol{\rho}, \mathbf{r}^o)$. Then the following pair of equations define the operators U and V

$$\mathbf{F}(\mathbf{r}^o) = \mathrm{U}\, I(\boldsymbol{\rho}) = \int_v I(\boldsymbol{\rho})\, \mathbf{f}(\boldsymbol{\rho}, \mathbf{r}^o)\, dv$$

$$p(\boldsymbol{\rho}) = \mathrm{V}\mathbf{F}(\mathbf{r}^o) = \oint_\Omega \mathbf{F}(\mathbf{r}^o)\, \mathbf{f}^*(\boldsymbol{\rho}, \mathbf{r}^o)\, d\Omega \qquad (3\text{-}2)$$

The top line of Ex. 3-2 is actually a copy of Ex. 2-3. In the integral on the second line, the multiplication of vector functions means their inner product, and the asterisk symbolizes the complex conjugate value. The function $p(\boldsymbol{\rho})$ is an antenna distribution to which the adjoint operator V maps an arbitrary radiation pattern $\mathbf{F}(\mathbf{r}^o)$. It is not difficult to make sure that the formula in Ex. 3-2 for adjoint operator V satisfies the condition Ex. 3-1.

For an antenna array consisting of N elements of fixed polarization, the expressions for the operators U and V are obviously very similar

$$\mathbf{F}(\mathbf{r}^o) = \mathrm{U}\, I(\boldsymbol{\rho}) = \sum_n I(\boldsymbol{\rho}_n)\, \mathbf{f}_n(\mathbf{r}^o)$$

$$p(\boldsymbol{\rho}_n) = \mathrm{V}\mathbf{F}(\mathbf{r}^o) = \oint_\Omega \mathbf{F}(\mathbf{r}^o)\, \mathbf{f}_n^*(\mathbf{r}^o)\, d\Omega \qquad (3\text{-}3)$$

If the polarization of antenna sources (say electrical current) may be arbitrary then generally speaking the distribution $\mathbf{I}(\boldsymbol{\rho})$ must be a vector function having three components in every point $\boldsymbol{\rho}$. We will use the shorten record mentioned in Section 2.2.1 $\mathbf{I}(\boldsymbol{\rho}) = \sum_\xi I_\xi(\boldsymbol{\rho})\, \boldsymbol{\xi}^o$ in which ξ and $\boldsymbol{\xi}^o$ denote three axes of some coordinate system and its vectors, respectively. Therefore, the summation actually is $\sum_{m=1}^3 I_{\xi_m}(\boldsymbol{\rho})\, \boldsymbol{\xi}_m^o$. Using nation ξ, we can transform Ex. 3-2 to the following expressions

$$\mathbf{F}(\mathbf{r}^o) = \mathrm{U}\mathbf{I}(\boldsymbol{\rho}) = \sum_\xi \int_v I_\xi(\boldsymbol{\rho})\, \mathbf{f}_\xi(\boldsymbol{\rho}, \mathbf{r}^o)\, dv$$

$$\mathbf{p}(\boldsymbol{\rho}) = \mathrm{V}\mathbf{F}(\mathbf{r}^o) = \sum_\xi \left(\oint_\Omega \mathbf{F}(\mathbf{r}^o)\, \mathbf{f}_\xi^*(\boldsymbol{\rho}, \mathbf{r}^o)\, d\Omega \right) \boldsymbol{\xi}^o \qquad (3\text{-}4)$$

Here, $\mathbf{f}_\xi(\boldsymbol{\rho}, \mathbf{r}^o)$ denotes the radiation pattern of the sources located at the point $\boldsymbol{\rho}$ and polarized along the vector $\boldsymbol{\xi}°$.

At last, in case of antenna array with elements of variable polarization, the following expressions define the direct and adjoint operators

$$\mathbf{F}(\mathbf{r}^o) = U\,\mathbf{I}(\boldsymbol{\rho}) = \sum_n \sum_\xi I_\xi(\boldsymbol{\rho}_n)\,\mathbf{f}_\xi(\boldsymbol{\rho}_n, \mathbf{r}^o)$$
$$\mathbf{p}(\boldsymbol{\rho}_n) = V\mathbf{F}(\mathbf{r}^o) = \sum_\xi \left(\oint_\Omega \mathbf{F}(\mathbf{r}^o)\,\mathbf{f}_\xi^*(\boldsymbol{\rho}_n, \mathbf{r}^o)\,d\Omega \right) \boldsymbol{\xi}^o . \tag{3-5}$$

The adjoint operator V, in addition to the attractive ability to map any—desired for example—radiation pattern into a purely radiating distribution, disclosed in Section 2.2.2, has other very useful properties.

First, the well-known Cauchy–Bunyakovsky–Schwarz inequality states that the inequality $|(f_1(x), f_2(x))| \leq \|f_1(x)\|\,\|f_2(x)\|$ is true for any two functions $f_1(x)$ and $f_2(x)$ in a Hilbert space. It turns into equality, only if the functions $f_1(x)$ and $f_2(x)$ are equal to each other or differ by a constant factor. Let us compose the ratio

$$\frac{|(\mathbf{F}_0(\mathbf{r}^o), U\mathbf{I}(\boldsymbol{\rho}))|^2}{\|\mathbf{F}_0(\mathbf{r}^o)\|^2 \|\mathbf{I}(\boldsymbol{\rho})\|^2} = P_{0\Sigma} / P_\mathbf{I} , \tag{3-6}$$

where $P_{0\Sigma}$, determined by Ex. 2-46, is the power radiated by distribution $\mathbf{I}(\boldsymbol{\rho})$ into the desired pattern $\mathbf{F}_0(\mathbf{r}^o)$, and the square of the norm $P_\mathbf{I} = \|\mathbf{I}(\boldsymbol{\rho})\|^2$ is, as commonly interpreted, the power of the excitation $\mathbf{I}(\boldsymbol{\rho})$.

From the Cauchy–Bunyakovsky–Schwarz inequality accounting for the basic property of the adjoint operator $(\mathbf{F}_0(\mathbf{r}^o), U\mathbf{I}(\boldsymbol{\rho})) = (V\mathbf{F}_0(\mathbf{r}^o), \mathbf{I}(\boldsymbol{\rho}))$, it follows that the considered ratio $P_{0\Sigma} / P_\mathbf{I}$ reaches the maximal value if the distribution $\mathbf{I}(\boldsymbol{\rho})$ is equal to $V\mathbf{F}_0(\mathbf{r}^o)$. In other words, the distribution $\mathbf{I}(\boldsymbol{\rho}) = V\mathbf{F}_0(\mathbf{r}^o)$ is optimal in the following sense: among all distributions with a restricted norm $\|\mathbf{I}(\boldsymbol{\rho})\|^2 \leq$ const it radiates into the desired pattern $\mathbf{F}_0(\mathbf{r}^o)$ the maximum power $P_{0\Sigma}$. With this in mind, Dymsky called this distribution an energetically optimal solution to the synthesis problem and introduced a special notation $\mathbf{p}_0(\boldsymbol{\rho}) = V\mathbf{F}_0(\mathbf{r}^o)$, to which we will follow.

Very interesting and pleasant to those who are fond of the adjoint operator is the fact that it, the adjoint operator, provides a very useful and widely used in practice distributions. Let us find out these facts.

Situation A. If a desired pattern $\mathbf{F}_0(\mathbf{r}^o) = \delta(\mathbf{r}^o - \mathbf{r}_m^o)\boldsymbol{\zeta}^o$ is a delta function in some direction $\mathbf{r}_m^o$ with polarization given by the Jones vector $\boldsymbol{\zeta}^{o}$ [21], then, taking into account Ex. 3-4, we have

$$\mathbf{p}_0(\boldsymbol{\rho}) = \sum_\xi \oint_{4\pi} \mathbf{f}_\xi^*(\boldsymbol{\rho}, \mathbf{r}^o)\,\delta(\mathbf{r}^o - \mathbf{r}_m^o)\,\boldsymbol{\zeta}^o d\Omega = \sum_\xi (\mathbf{f}_\xi^*(\boldsymbol{\rho}, \mathbf{r}_m^o)\,\boldsymbol{\zeta}^o)\,\boldsymbol{\xi}^o . \tag{3-7}$$

---

[21] In the most general case, the Jones vector $\boldsymbol{\xi}^o$ consists of θ- and φ- components reproducing, elliptical polarization.

Here, ξ symbolizes three axes of a coordinate system used to describe vector structure of the distribution $\mathbf{I}(\boldsymbol{\rho})$ or $\mathbf{p}_0(\boldsymbol{\rho})$; $\{\boldsymbol{\xi}^o\}$ is the base set of coordinate vectors; $\mathbf{f}_\xi(\boldsymbol{\rho}, \mathbf{r}^o)$ is the vector function describing the radiation pattern of a source that has ξ-polarization and locates in the point $\boldsymbol{\rho}$. Ex. 3-7 shows that distribution $\mathbf{p}_0(\boldsymbol{\rho}) = V\,\delta(\mathbf{r}^o - \mathbf{r}_m^o)\,\boldsymbol{\zeta}^o$ corresponds to the widely used beamforming principle. That is, the excitation of sources must be proportional to the complex conjugate values of their individual radiation patterns in the given direction $\mathbf{r}_m^o$, taking into account the prescribed polarization $\boldsymbol{\zeta}^o$ of the desired ARP, as well.

In the very common and simplest situation, the orientation of antenna sources is fixed and the polarization of their patterns is similar. Then Ex. 3-6 simplifies to the well-known scalar equality

$$p_0(\boldsymbol{\rho}) = f^*(\boldsymbol{\rho},\ \mathbf{r}_m^o), \tag{3-8}$$

that expressively represents the beamforming principle.

Situation B. Another fact confirming the significance of the adjoint operator is as follows. The ARP $F(\theta)$ of a $2a$ long linear antenna composed of isotropic sources—the array factor, in other words—corresponds to the obvious formula [25] that in terms of the commonly used variables z and θ is as follows:

$$F(\theta) = \int_{-a}^{a} I(z)\,e^{jkz\cos\theta}\ dz\,. \tag{3-9}$$

Defining, as usual, the abstract angular variable $u = k\cos\theta$, we can rewrite Ex. 3-9 as a Fourier transform $F(u) = \int_{-a}^{a} I(z)\,e^{juz}\ dz$. Therefore, the inverse Fourier transform $I(z) = \int_{-\infty}^{\infty} F(u)\,e^{-jzu}\ du$ defines the distribution $I(z)$ corresponding to the ARP $F(u)$. Here the integral along with the interval $|u| \le k$ of real angles $-1 \le \sin\theta \le 1$ includes the interval $|u| > k$ of imaginary angles $|\sin\theta| > 1$.

Clearly, the square of the norm $\|F(u)\|^2$, corresponding to the first interval, is the radiated power P_Σ and the square of the norm $\|F(u)\|^2$, corresponding to the second interval, corresponds—as it is believed—to the reactive power P_Q. The desired ARP $F_0(u)$ is usually not realizable by an antenna of a given length $2a$ and, therefore, has no analytic continuation into the interval $|u| > k$ of imaginary angles. In addition, we always strive to minimize antenna reactivity. So it is quite reasonable to set $F_0(u) = 0$ for the intervals $|u| > k$. This leads to the following expression

$$I_0(z) = \int_{-k}^{k} F_0(u) e^{-jzu} \, du = k \int_{-\pi/2}^{\pi/2} F_0(\theta) e^{-jkz\cos\theta} \sin\theta \, d\theta \, . \qquad (3\text{-}10)$$

Of course, the function Ex. 3-10 exists along the entire z-axis, since the antenna must be infinite in order to approach an unrealizable ARP. A common and successful method for creating an effective real antenna is to implement the excitation $I_0(z)$ given by Ex. 3-10 along the actual antenna length $-a \leq z \leq a$ [25]. For example, this means that if the desired ARP $F_0(\theta) = \delta(\theta - \theta_0)$ is a δ-function in the θ_0 direction, then the distribution $I_0(z) = e^{-jkz\cos\theta_0}$ corresponds to a very practical solution, which is the uniform amplitude and linear phase distribution. In the direction θ_0, it provides the in-phase summation of waves emitted by the antenna sources, which is in good agreement with common sense if we strive to maximize the directivity in this direction.

Another well-known consequence of Ex. 3-9 is that if the desired ARP $F_0(\theta)$ is a sector function $F_0(\theta) = \begin{cases} 1 & \text{if } |\theta| \leq \Delta\theta \\ 0 & \text{if } |\theta| > \Delta\theta \end{cases}$, then a very practical

distribution in the antenna is the truncation of function $\mathrm{sinc}(z)$, that is

$$I_0(z) = \begin{cases} \dfrac{\sin(kz\sin(\Delta\theta))}{kz\sin(\Delta\theta)} & |z| \leq a \\ 0 & |z| > a \end{cases} \, .$$

It is easy to see that in the case under consideration, replacing $\mathbf{r}^0$ by θ, $\boldsymbol{\rho}$ by z, $f(\boldsymbol{\rho}, \mathbf{r}^0)$ by $\exp(j\,k\,z\cos\theta)$ and $d\Omega$ by $2\pi\sin\theta$, we convert Ex. 3-2 for operators U and V to the following expressions

$$F(\theta) = \mathrm{U}\, I(z) = \int_{-a}^{a} I(z) e^{jkz\sin\theta} \, dx$$
$$p_0(z) = \mathrm{V} F_0(\theta) = 2\pi \int_{-\pi/2}^{\pi/2} F_0(\theta) e^{-jkz\sin\theta} \cos\theta \, d\theta \qquad (3\text{-}11)$$

Thus, the adjoint operator V in this case is nothing but the inverse Fourier transform Ex. 3-10, commonly used to synthesize a linear antenna.

Second, let assume that the desired ARP is realizable, i.e. has an exact decomposition into a basis set $\{\mathbf{G}_n(\mathbf{r}^\circ)\}$

$$\mathbf{F}_0(\mathbf{r}^\circ) = \sum_n a_n \, \mathbf{G}_n(\mathbf{r}^\circ) \, . \qquad (3\text{-}12)$$

In fact, the distribution $\mathbf{p}_0(\boldsymbol{\rho})$ is an approximate solution of the synthesis problem $\mathrm{U}\mathbf{I}(\boldsymbol{\rho}) = \mathbf{F}_0(\mathbf{r}^\circ)$, to which the following series correspond

$$\mathbf{p}_0(\boldsymbol{\rho}) = \sum_n a_n \sqrt{\lambda_n}\, \mathbf{g}_n(\boldsymbol{\rho}), \qquad \mathrm{U}\mathbf{p}_0(\boldsymbol{\rho}) = \sum_n a_n \lambda_n \, \mathbf{G}_n(\mathbf{r}^\circ) \, . \qquad (3\text{-}13)$$

Comparison with the series for the exact solution Ex. 3-12 reveals that the terms of the series Ex. 3-13 differ by the factors λ_n. In what does this result? Of course, it leads to some deterioration in accuracy. However, on

the other hand, it leads to a decrease in reactivity, i.e. to better energy efficiency of the antenna. The reactivity of the exact solution $\mathbf{I}(\boldsymbol{\rho})$ and

solution $\mathbf{p}_0(\boldsymbol{\rho})$ are $Q_\mathrm{I} = \dfrac{\|\mathbf{I}(\boldsymbol{\rho})\|^2}{\|\mathrm{U}\mathbf{I}(\boldsymbol{\rho})\|} = \dfrac{\sum_n |a_n|^2 / \lambda_n}{\sum_n |a_n|^2}$ and

$Q_\mathrm{p} = \dfrac{\|\mathbf{p}_0(\boldsymbol{\rho})\|^2}{\|\mathrm{U}\mathbf{p}_0(\boldsymbol{\rho})\|} = \dfrac{\sum_n |a_n|^2 \lambda_n}{\sum_n |a_n|^2 \lambda_n^2}$, respectively. The very important and

interesting thing is, that such characteristics of the solution $\mathbf{p}_0(\boldsymbol{\rho})$ as the values of the reactivity Q_p and the square deviation $\underline{\sigma}^2$ belong to the curve $\underline{\sigma}^2(Q)$ of their optimal combination. In Fig. 2-7a and Fig. 2-8b, there are points marked by c that are related to the solution $\mathbf{p}_0(\boldsymbol{\rho})$. This attractive quality of the adjoint operator V inspires the idea of constructing a sequential refinement algorithm, based on using the operator V at each step of recursions. Section 3.3 is devoted to this algorithm.

Third, the most impressive feature of operator V is that it is defined at any space point $\boldsymbol{\rho}$ regardless of whether there is an antenna element or not. It turns out that the function $\mathbf{p}_0(\boldsymbol{\rho})$ refers to the electromagnetic field created by a converging stream—or cluster in a discrete version—of the incident plane waves, whose complex amplitudes and polarization depend on the required ARP $\mathbf{F}_0(\mathbf{r}^\mathrm{o})$. Therefore, we can calculate or even measure (!) the function $\mathbf{p}_0(\boldsymbol{\rho})$ and use it at the design stage. It turns out that this is especially useful for choosing the location of the elements of the antenna array. The corresponding procedure is under consideration in Section 3.4.

3.3. Recursive procedure applying the adjoint operator to enhance the solution

3.3.1. Introductory note

As justified above, the distribution $\mathbf{p}_0(\boldsymbol{\rho})$ to which adjoint operator V maps a prescribed ARP $\mathbf{F}_0(\mathbf{r}^\mathrm{o})$, is a rather good solution of the antenna synthesis problem that provides an attractive compromise between its accuracy and reactivity. However, in certain situations, accuracy can be of most importance. Then it seems promising to develop a recursive procedure, at each step of which the adjoint operator V gives a consequent amendment used in some way to correct the solution. Due to the energy efficiency of the operator V, this way of improving the solution accuracy will lead to a reasonably moderate increase in its reactivity. One can continue recursion to achieve acceptable accuracy of the solution or the limit to its reactivity.

At the simplest iterations $\mathbf{I}_{k+1}(\boldsymbol{\rho}) = \mathbf{I}_k(\boldsymbol{\rho}) + V\mathbf{F}_k^\perp(\mathbf{r}^\circ)$, where $\mathbf{F}_k^\perp(\mathbf{r}^\circ) = \mathbf{F}_0(\mathbf{r}^\circ) - U\mathbf{I}_k(\boldsymbol{\rho})$. This means that the amendment to the previous solution $\mathbf{I}_k(\boldsymbol{\rho})$ corresponds exactly to what the V-operator maps the remainder $\mathbf{F}_k^\perp(\mathbf{r}^\circ)$ of the required ARP from the previous step. Obviously, the convergence of such iterations leaves much to be desired. Say, if $\mathbf{F}_0(\mathbf{r}^\circ) = \mathbf{G}_N(\mathbf{r}^\circ)$, then at the kth iteration we have $\sigma_k^2 = \|\mathbf{F}_k^\perp(\mathbf{r}^\circ)\|^2 = (1 - \lambda_N)^{2k}$. This means a very slow convergence if the eigenvalue λ_N is sufficiently small.

Therefore, let us rethink the significantly better and more popular algorithm, the so-called minimal residual algorithm (MRA) [26], which corresponds to the following iterations

$$\mathbf{I}_{k+1}(\boldsymbol{\rho}) = \mathbf{I}_k(\boldsymbol{\rho}) + \gamma_k \mathbf{p}_k(\boldsymbol{\rho}) \tag{3-14}$$

with $\mathbf{p}_k(\boldsymbol{\rho}) = V\mathbf{F}_k^\perp(\mathbf{r}^\circ)$ and $\gamma_k = (\mathbf{F}_k^\perp(\mathbf{r}^\circ), U\mathbf{p}_k(\boldsymbol{\rho}))/\|U\mathbf{p}_k(\boldsymbol{\rho})\|^2$. The factor γ_k is similar to that in the Schmidt orthogonalization procedure—see Section 1.3.3. Consequently, in step $k + 1$, it minimizes the deviation between the amending ARP $\gamma_k U\mathbf{p}_k(\boldsymbol{\rho})$ and the remainder $\mathbf{F}_k^\perp(\mathbf{r}^\circ) = \mathbf{F}_0(\mathbf{r}^\circ) - U\mathbf{I}_k(\boldsymbol{\rho})$ left after the previous step. Due to this, in the above situation, when the desired ARP $\mathbf{F}_0(\mathbf{r}^\circ)$ coincides with some eigenfunction $\mathbf{G}_n(\mathbf{r}^\circ)$, we get the exact solution in the very first iteration. In any case, the factor γ_k improves solution by optimal scaling the amendment in each step, and consequently it speeds the convergence of the iterations.

In general, as it clearly follows from Ex. 3-12 and Ex. 3-13, the accuracy of the solution $\mathbf{p}_0(\boldsymbol{\rho}) = V\mathbf{F}_0(\mathbf{r}^\circ)$ depends strongly on the operator L = VU spectrum $\{\lambda_n\}$ and the spectrum $\{a_n\}$ of the desired ARP, that is the set of coefficients in the decomposition of a function $\mathbf{F}_0(\mathbf{r}^\circ)$ into a basis set $\{\mathbf{G}_n(\mathbf{r}^\circ)\}$.

A good idea would be to find the worst combination first, study the situation in order to understand the problem, and then develop an efficient algorithm to overcome it. There are several ways to find the worst combination of the spectrums $\{\lambda_n\}$ and $\{a_n\}$, which in the first step corresponds to the maximum value of the square deviation

$$\sigma^2 = \|\mathbf{F}_0(\mathbf{r}^\circ) - \gamma_0 U\mathbf{p}_0(\boldsymbol{\rho})\|^2 / \|\mathbf{F}_0(\mathbf{r}^\circ)\|^2. \tag{3-15}$$

One of these ways stems from the engineering style of thinking and is surprisingly easy and enlightening.

First, let us turn Ex. 3-15 into an explicit expression with respect to spectrums $\{\lambda_n\}$ and $\{a_n\}$. The optimal scaling up by the factor γ_0 turns σ^2

into the normalized mean square deviation of antenna pattern's form $\underline{\sigma}^2$ given by Ex. 2-48, which is now

$$\underline{\sigma}^2 = 1 - \frac{|(\mathbf{F}_0(\mathbf{r}^\circ), U\mathbf{p}_0(\boldsymbol{\rho}))|^2}{\|\mathbf{F}_0(\mathbf{r}^\circ)\|^2 \|U\mathbf{p}_0(\boldsymbol{\rho})\|^2} . \tag{3-16}$$

One   can   obtain   this   result   directly   plugging $\gamma_0 = (\mathbf{F}_0(\mathbf{r}^\circ), U\mathbf{p}_0(\boldsymbol{\rho})) / \|U\mathbf{p}_0(\boldsymbol{\rho})\|^2$ into Ex. 3-15 and making necessary simplifications. In view of Ex. 3-12 and Ex. 3-13, the final expression is, as follows

$$\underline{\sigma}^2 = 1 - \frac{|\sum_{n=1}^{N} \lambda_n |a_n|^2|^2}{\sum_{n=1}^{N} |a_n|^2 \sum_{n=1}^{N} \lambda_n^2 |a_n|^2} . \tag{3-17}$$

Suppose that we synthesize a certain antenna, and corresponding spectrum $\{\lambda_n\}$ is fixed in descending order with the maximum value $\lambda_1 = \lambda_{max}$ and the minimum value $\lambda_N = \lambda_{min}$. Bearing in mind the search for the extreme spectrum $\{a_n\}$ and taking into account Ex. 3-17, we denote new unknowns $\{x_n = |a_n|^2\}$ simplifying Ex. 3-17 to the form

$$\underline{\sigma}^2 = 1 - \frac{|\sum_{n=1}^{N} \lambda_n x_n|^2}{\sum_{n=1}^{N} x_n \sum_{n=1}^{N} \lambda_n^2 x_n} . \tag{3-18}$$

There are at least three ways to solve the problem.

3.3.2. The good students' way: constrained optimization method

The need to solve some constrained optimization problem often arises in any technical field. Therefore, if not knowledge, then at least awareness of the relevant methods is important for any engineer, and not just for antenna designers. That is why I put forward this perhaps slightly formal but universal and elegant method in the first place.

Let us start. It is clear[22] that the value of Ex. 3-18 does not depend on the scaling factor for the sought-for coefficients $\{x_n\}$. Therefore, without loss of generality, we can set $\sum_{n=1}^{N} x_n = 1$ and seek for a maximum of Ex. 3-18 by solving the following constrained optimization problem

$$\min f(x_1, x_2 ... x_N) = |\sum_{n=1}^{N} \lambda_n x_n|^2 \Big/ \sum_{n=1}^{N} \lambda_n^2 x_n$$

subject to constrain: $g(x_1, x_2 ... x_N) = \sum_{n=1}^{N} x_n - 1 = 0$ \qquad (3-19)

---

[22] Read the arguments in the paragraph above Ex. 2-48.

The Lagrange multipliers method is a common and effective way to solve the problem [27]. Nowadays, searching for a "constrained optimization problem" on the Internet, any good student can easily find relevant chapters of a book or lectures on this topic[23] with comprehensive examples and detailed instructions for using the method.

The Lagrangian function for the minimization problem is $L(\{x_n\}, \mu) = f(\{x_n\}) - \mu\, g(\{x_n\})$. A thought-for solution must turn the values of Lagrangian's partial derivatives to zero with respect to any of the variables $\{x_n\}$ and μ. As a result, after elementary simplifications in view of the obvious $\sum_{n=1}^{N} \lambda_n^2 x_n > 0$, we can write the following system of equations

$$\begin{cases} \left(2\lambda_n \sum_k \lambda_k^2 x_k - \lambda_n^2 \sum_k \lambda_k x_k\right)\sum_k \lambda_k x_k - \mu\left(\sum_k \lambda_k^2 x_k\right)^2 = 0 \\ \sum_k x_k - 1 = 0 \end{cases} \quad (3\text{-}20)$$

Note that changing the variables $\{x_n\}$ alters two values only, say $a = \sum_k \lambda_k x_k$ and $b = \sum_k \lambda_k^2 x_k$ figuring in the first line of Ex. 3-20 as multipliers. This line actually defines N equations for $n = 1...N$, but only two of them can be satisfied. Denote the corresponding indices by i and j. Therefore, variables with other indices should not be present, i.e. $x_n = 0$ for $n \neq i$ and $n \neq j$, and system Ex. 3-20 turns to three equations

$$\begin{cases} 2\lambda_i ab - \lambda_i^2 a^2 - \mu b^2 = 0 \\ 2\lambda_j ab - \lambda_j^2 a^2 - \mu b^2 = 0, \\ x_i + x_j - 1 = 0 \end{cases} \quad (3\text{-}21)$$

where $a = \lambda_i x_i + \lambda_j x_j$ and $b = \lambda_i^2 x_i + \lambda_j^2 x_j$. Subtracting the second equation from the first to eliminate the variable μ and substituting the third equation $(x_j = 1 - x_i)$ there, after simple but tedious manipulations, we get the solution $\{x_i = \lambda_j / (\lambda_i + \lambda_j), x_j = \lambda_i / (\lambda_i + \lambda_j)\}$.

It is a typical situation regarding constrained optimization. Initially, we get some swarm of solutions from which we have to separate suitable ones or one. In our case, it is about uncertainty with respect to indices i and j. Looking to Ex. 3-17 and bearing in mind that there are only two terms in each sum, we can easily guess that the less λ_i differs from λ_j, the closer the subtracted term approaches unity. Consequently, the right solution is

$$\{x_1 = \lambda_{min}/(\lambda_{max}+\lambda_{min}),\ x_N = \lambda_{max}/(\lambda_{max}+\lambda_{min})\}, \quad (3\text{-}22)$$

---

[23] For instance, to read "Constrained Optimization: Step by Step", one can use URL: https://www3.nd.edu/~jstiver/FIN360/Constrained%20Optimization.pdf

if eigenvalues have a descending order, i.e. $\lambda_1 = \lambda_{max}$ and $\lambda_N = \lambda_{min}$. As for the spectrum $\{a_n\}$ corresponding to the worst situation for the solution $\mathbf{p}_0(\boldsymbol{\rho}) = V \mathbf{F}_0(\mathbf{r}^o)$, which interests us now, it looks like this: $\left\{\sqrt{\lambda_{min}}, 0...0, \sqrt{\lambda_{max}}\right\}$.

3.3.3. The average students' way: a partial derivatives method

Imagine you are an average student. Of course, an average student likes simplifications—readily uses Ex. 3-18 instead of Ex. 3-17—and knows that at the extrema of the function $f(x_1, x_2 ... x_N) = 1 - \left(\sum_n \lambda_n x_n\right)^2 / \left(\sum_n x_n \sum_n \lambda_n^2 x_n\right)$, its partial derivatives must be equal to zero. This fact leads to a system of N equations $\partial f/\partial x_n = 0$. He is lazy enough to write long expressions, and smart enough to set up the following notations $\mathbf{x} = \{x_n\}$, $u(\mathbf{x}) = \sum_n \lambda_n x_n$, $v(\mathbf{x}) = \sum_n x_n$, $w(\mathbf{x}) = \sum_n \lambda_n^2 x_n$, which shrinks the formula to $f(\mathbf{x}) = 1 - u^2(\mathbf{x})/[v(\mathbf{x}) w(\mathbf{x})]$. He understands that $\partial u(\mathbf{x})/\partial x_n = \lambda_n$, $\partial v(\mathbf{x})/\partial x_n = 1$, and $\partial w(\mathbf{x})/\partial x_n = \lambda_n^2$. As a result, nth equation of the system is as follows $[2 \lambda_n u(\mathbf{x}) v(\mathbf{x}) w(\mathbf{x}) - u^2(\mathbf{x}) (w(\mathbf{x}) + \lambda_n^2 v(\mathbf{x}))] / [v(\mathbf{x}) w(\mathbf{x})]^2 = 0$. It is clear that for any solution $\mathbf{x}_0$—except for the zero vector—the values of all functions $u(\mathbf{x}_0) = a$, $v(\mathbf{x}_0) = b$, and $w(\mathbf{x}_0) = c$ are greater than zero. Therefore, after multiplying by $[v(\mathbf{x}_0) w(\mathbf{x}_0)]^2$ and dividing by $u(\mathbf{x}_0)$, the last equation takes a very simple form

$$2\lambda_n b c - a (c + \lambda_n^2 b) = 0. \tag{3-23}$$

A sought-for solution $\mathbf{x}_0$, defining the values a, b, and c, must satisfy a linear system of order N represented by Ex. 3-20. Even if these values are independent variables—the student thinks—the above system had no solution if $N > 3$. This means that the function f has no extrema inside the boundary determined by the obvious constraints $x_n = |a_n|^2 > 0$ and gets maximal and/or minimal values at some boundary points only. Consequently, at least one of the variables $\{x_n\}$ must be equal to zero. Deleting this variable from their set and repeating the same reasoning for the function of the remaining variables, they discover that the function f gets the maximal value at the boundary point at which all the variables are zero, with the exception of perhaps three of them, say $n = 1$, $n = 2$. $n = 3$. In other words, a sought-for solution has to be $\mathbf{x}_0 = \{x_1, x_2, x_3, 0...0\}$. "Great, I love simplifications," the student rejoices.

Now let us deal with the first two equations: $2\lambda_1 b c - a (c + \lambda_1^2 b) = 0$ and $2\lambda_2 b c - a (c + \lambda_2^2 b) = 0$. Subtracting the second from the first gives

$2 (\lambda_1 - \lambda_2) b c - (\lambda_1^2 - \lambda_2^2) a b = 0.$ (3-24)

From Ex. 3-24 it is strikingly clear that the variable b does not matter, i.e. may have any value. Fine, suppose that $b = 1$. Since only two variables—a, and c—remain, the third equation is superfluous and, consequently, $x_3 = 0$. The assumption $b = 1$ means $x_2 = 1 - x_1$, $a = \lambda_1 x_1 + \lambda_2 (1 - x_1)$ and $c = \lambda_1^2 x_1 + \lambda_2^2 (1 - x_1)$. This turns Ex. 3-24 into equation of one variable x_1: $2\lambda_1^2 x_1 + 2\lambda_2^2 (1 - x_1) = (\lambda_1 + \lambda_2) (\lambda_1 x_1 + \lambda_2 (1 - x_1))$. Having performed the necessary substitutions and permutations—with errors and subsequent corrections—the student finally obtained the result $\{x_1 = \lambda_2/(\lambda_1 + \lambda_2), x_2 = \lambda_1/(\lambda_1 + \lambda_2)\}$, which satisfies the equation Ex. 3-24.

3.3.4. The creative students' way: a method of engineering style

Dear reader, the aim of this section is to show how effective and elegant it would be solving the above problem under "the guidance" of an engineering style of thinking. First, it is worth trying to find out some structural features of the problem, without going into details. I used to call it the shell principle: do not start with an in-depth investigation; try to understand a frame of the problem as a whole, as if looking at it from above. Let us go.

What is the score of the problem? The spectrum $\{\lambda_n\}$ is a given. We search for a spectrum $\{a_n\}$ to which the ARP $\mathbf{F}(\mathbf{r}^o) = UV \mathbf{F}_0(\mathbf{r}^o) = \sum_n a_n \lambda_n \mathbf{G}_n(\mathbf{r}°)$ makes the worst approach to the desired ARP $\mathbf{F}_0(\mathbf{r}^o) = \sum_n a_n \mathbf{G}_n(\mathbf{r}°)$. A mean square deviation of the forms—given by Ex. 3-18—serves a measure of the patterns deviation. Then the thought-for spectrum $\{a_n\}$ must contain at least two non-zero components, because the ARPs $\mathbf{F}_0(\mathbf{r}^o)$ and $\mathbf{F}(\mathbf{r}^o)$ have the same form—$\underline{\sigma} = 0$—otherwise. It is also clear from Ex. 3-18 that the mutual ratio of the variables $\{x_n = |a_n|^2\}$ matters, not a scale factor.

Someone can quickly decide that the desired worst situation arises when using the components of the spectrum $\{\lambda_n\}$ with the greatest difference, namely λ_{max} and λ_{min}. Accordingly, he will think about solution $\mathbf{x}_0 = \{x_1, 0\ldots0, x_N\}$, deciding what the ratio x_1/x_N should be.

For an unhurried person who does not like to "jump" to a decision and prefers to approach it gradually, the following reasoning may be appropriate. First, it is clear that the spectrum $\mathbf{x}_0$ must have at least two components. Second, x_1 and x_N must be present in order for the maximum deviation between $\mathbf{F}_0(\mathbf{r}^o)$ and $\mathbf{F}(\mathbf{r}^o)$ to be possible. Assume that $x_2 > 0$. This will appear as if instead of λ_1 and λ_2 there is a resulting eigenvalues λ, which lies between λ_1 and λ_2—depending on the ratio x_2/x_1. Thus, although

we are interested in increasing the ARP deviation, it will decrease, because the difference $\lambda - \lambda_2$ becomes smaller than $\lambda_1 - \lambda_2$. The situation is the same with any other element x_n. Accordingly, the solution $\mathbf{x}_0$ should be $\{x_1, 0\ldots0, x_N\}$.

It is very interesting that we can find the desired ratio between x_1 and x_N from an engineering point of view, avoiding the tedious analysis of Ex. 3-18 partial derivatives. Indeed, what does the engineer do in the first place, trying to solve the problem? He "depicts a problem" to perceive its core. Let us go this way.

The commonly used vector representation for functions leads to the image in Fig. 3-1a. There, vectors $\mathbf{G}_1$, $\mathbf{G}_N$, $\mathbf{F}_0$, and $\mathbf{F}$ depict the base functions $\mathbf{G}_1(\mathbf{r}^o)$, $\mathbf{G}_N(\mathbf{r}^o)$, the desired ARP $\mathbf{F}_0(\mathbf{r}^o)$ and ARP $\mathbf{F}(\mathbf{r}^o) = \text{UV}\,\mathbf{F}_0(\mathbf{r}^o)$, respectively. In what way does the objective function $\underline{\sigma}^2$ "peep out" from the picture? It is clear that the ARP scale factor relates to the length of the corresponding vector, and the ARP form determines its orientation. Thus, $\underline{\sigma}^2$ is in some—maybe nonlinear—proportion with the angle $\varphi_0 - \varphi$. Moreover, given that $|(\mathbf{F}_0, \mathbf{F})| = |\mathbf{F}_0|\,|\mathbf{F}|\cos(\varphi_0 - \varphi)$, we can see that Ex. 3-16, in this case, turns to $\underline{\sigma}^2 = \sin^2(\varphi_0 - \varphi)$.
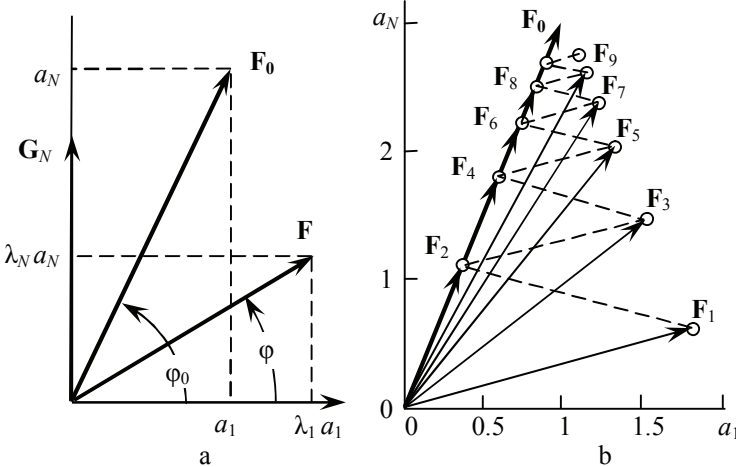
Figure 3-1: Graphical representation of iterations of the minimal residual algorithm: a) searching the worst situation in the first step; b) stepwise convergence of ARPs $\mathbf{F}_k$ to the exact solution $\mathbf{F}_0$.

Given the eigenvalues λ_1 and λ_N, we have to find such a vector $\mathbf{F}_0$—in fact, the coefficients a_1 and a_N—from which the vector $\mathbf{F}$ deviates as much as possible. Denoting $\xi = a_N / a_1$ and $\alpha = \lambda_N / \lambda_1$ we can write $\text{tg}(\varphi_0) = \xi$ and $\text{tg}(\varphi) = \alpha\,\xi$. To simplify the task it is worth searching for maxima not of the angle $(\varphi_0 - \varphi)$ itself or its sine, but for maxima of $\text{tg}(\varphi_0 - \varphi)$. Then the

task turns into maximizing the function $f(\xi) = \xi / (1 + \alpha\,\xi^2)$ with the elementary result $\xi = 1/\sqrt{\alpha}$. In the above notation, this means $x_N/x_1 = |a_N|^2/|a_1|^2 = \lambda_1/\lambda_N$.

Figure 3-1b shows the results in nine initial steps of the MRA with spectrums $\{\lambda_1 = \lambda_{max} = 1, \lambda_N = \lambda_{min} = 1/9\}$ and $\{a_1 = 1, a_N = 3\}$. Here, $\mathbf{F}_k$ is ARP obtained in step k. The dashed line traces the process of stepwise convergence ARP $\mathbf{F}_k$ to the prescribed ARP $\mathbf{F}_0$. It is striking that in the second step—as it is in any even step k—the ARP $\mathbf{F}_k$ actually coincides with $\mathbf{F}_0$, except for some scale factor. However, following the MRA course, $\mathbf{F}_k$ will approach $\mathbf{F}_0$ after many steps of the zigzag route.

3.3.5. Modified minimal residual algorithm

Based on the above, in order to accelerate the convergence of the MRA in the worst situation, a modified minimal residual algorithm (MMRA) had emerged [28]. It includes optimal scaling not only for the amendment in the step but also for the whole solution at the end of this step. As a result, iterations Ex 3-14 turn to the following

$$\mathbf{F}_k^{\perp}(\mathbf{r}^\circ) = \mathbf{F}_0(\mathbf{r}^\circ) - \mathrm{U}\mathbf{I}_k(\boldsymbol{\rho})$$

$$\mathbf{p}_k(\boldsymbol{\rho}) = \mathrm{V}\mathbf{F}_k^{\perp}(\mathbf{r}^\circ), \quad \gamma_k = (\mathbf{F}_k^{\perp}(\mathbf{r}^\circ), \mathrm{U}\,\mathbf{p}_k(\boldsymbol{\rho})) / \parallel \mathrm{U}\,\mathbf{p}_k(\boldsymbol{\rho}) \parallel^2$$

$$\mathbf{J}_k(\boldsymbol{\rho}) = \mathbf{I}_k(\boldsymbol{\rho}) + \gamma_k \mathrm{V}\mathbf{F}_k^{\perp}(\mathbf{r}^\circ), \quad \beta_k = (\mathbf{F}_0(\mathbf{r}^\circ), \mathrm{U}\,\mathbf{J}_k(\boldsymbol{\rho})) / \parallel \mathrm{U}\,\mathbf{J}_k(\boldsymbol{\rho}) \parallel^2 \qquad (3\text{-}25)$$

$$\mathbf{I}_{k+1}(\boldsymbol{\rho}) = \beta_k \mathbf{J}_k(\boldsymbol{\rho})$$

Of course, the benefit of implementing this algorithm depends on spectrums $\{\lambda_n\}$ and $\{a_n\}$. As for antenna syntheses, the solution reactivity Q is of no less importance than a convergence speed. Fig. 3-2 displays the numerical results for two intentionally selected situations.

The first one corresponds to the rather typical spectrum $\{\lambda_n = \cos^2(\pi\,n/(2N+2))\}$ with dimension $N = 12$. The deliberate assumption of the desired ARP spectrum was of inverse proportion $\{a_n = \lambda_N/\lambda_n\}$ to get a hard situation, in which the modified algorithm shows an impressive speed advantage—in step $k = 27$, for example, σ is –30 dB instead of –10 dB—and noticeable decrees in solution reactivity as well.

The second situation is less extravagant with a spectrum[24] $\{\lambda_n = |J_n(a)|\}$ corresponding to an antenna array of $N = 18$ isotropic elements evenly located on a circle with a radius of electrical length $ka = 2.2\,\pi$. The chart in the right top corner of Fig. 3-2b shows this spectrum. The used uniform spectrum $\{|a_n| = 1\}$ corresponds to the desired ARP, which is a beam in

---

[24] See Ex 2-29.

some direction having the maximum directivity. From Fig. 3-2b, we can see that the results for both algorithms are very close to the optimal combination of $\underline{\sigma}$ and Q—the dashed line—but MMRA achieves –30 dB accuracy at step $k = 26$, while MRA achieves it at the 50th step.
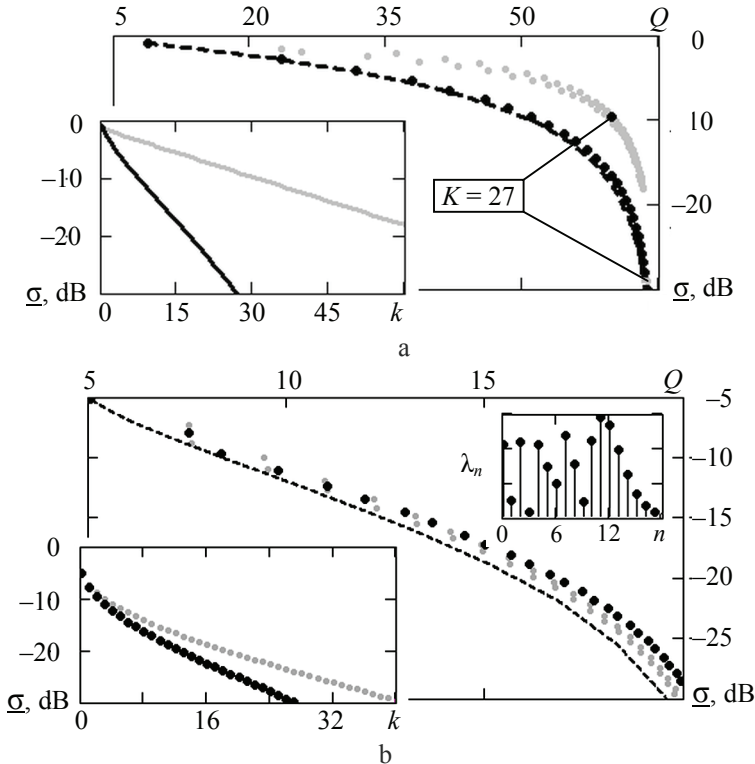
Figure 3-2: Comparison of algorithms: black dots – MMRA; gray dots – MRA; dashed line – optimal dependency $Q(\underline{\sigma})$. The main chart represents the reactivity Q_k and the root square deviation $\underline{\sigma}_k$ of solutions generated by both algorithms. The charts in bottom left corner show $\underline{\sigma}_k$ for 60 (a) or 40 (b) steps

Of course, not only the number of steps is important, but also the duration of the calculations. MMRA requires additional multiplications that increase time-consuming. The thorough comparison has to consider this. However, this goes beyond the scope of this book, which focuses on antenna synthesis methodology, ideas and methods, and, to a lesser extent, results.

3.4. Configuring the antenna array geometry with regard to a desired radiation pattern

3.4.1. Sequential selection of antenna elements

Let us start with a trivial thing. If $\mathbf{f}(\boldsymbol{\rho}, \mathbf{r}^0)$ is the vector pattern of the antenna source located at some point $\boldsymbol{\rho}$, and $\mathbf{A}(\mathbf{r}^0)$ denotes the polarization and the complex amplitude of the plane wave incident from the direction $\mathbf{r}^0$, then the dot product $\mathbf{A}(\mathbf{r}^o)\,\mathbf{f}(\boldsymbol{\rho}, \mathbf{r}^0)$ determines the signal $s(\boldsymbol{\rho})$ that the wave induces on the antenna element. Suppose now that the incident waves come from all directions, forming a convergent cluster of plane waves (CCPW). In this situation, the following expression defines the complex amplitude of the received signal

$$s(\boldsymbol{\rho}) = \oint_{\Omega} \mathbf{A}(\mathbf{r}^o)\,\mathbf{f}(\boldsymbol{\rho}, \mathbf{r}^o)\,d\Omega\,. \tag{3-26}$$

Comparing Ex. 3-26 with the second line of Ex. 3-2, one can easily guess that if $\mathbf{A}(\mathbf{r}^o) = \mathbf{F}_0^*(\mathbf{r}^o)$ then the signal $s(\boldsymbol{\rho})$ is equal to $p_0^*(\boldsymbol{\rho}) = [\mathrm{V}\,\mathbf{F}_0(\mathbf{r}^o)]^*$. This godsend makes it possible to develop such a technical setup that is actually the adjoint operator V in the flesh. Indeed, it suffices to place the proper number M of radiating sources onto the surrounding sphere of a large radius in free space, excite them in accordance with the function $\mathbf{F}_0^*(\mathbf{r}^o)$ conjugate to the required ARP and measure the signals from the antenna elements. If the antenna does not exist yet, the following way is appropriate. A probe equivalent to the antenna element is moving along the aperture of the future antenna or into a set of points $\{\boldsymbol{\rho}_n\}$ in the case of an array antenna. The conjugate of the measured signals $s(\boldsymbol{\rho})$ is the sought-for APhD $p_0(\boldsymbol{\rho})$.

In Fig 3-3a, the photo shows "a device for approximate solving antenna synthesis problem" [USSR Patent 1297006, 1971], which realizes the above idea for two-dimensional situations. It consists of six panels with four half-wave dipoles in each, arranged into a semicircle with a radius of 10 m[25]. Uniformly excited dipoles correspond to a sector ARP $\mathbf{F}_0^*(\mathbf{r}^o)$, which is a sector of 180° with a phase center in the center of the circle.

---

[25] Because of overall dimensions, the only room able to hold it was the assembly hall of the faculty. In the background, you can see the stage with the curtain down and the blackboard.
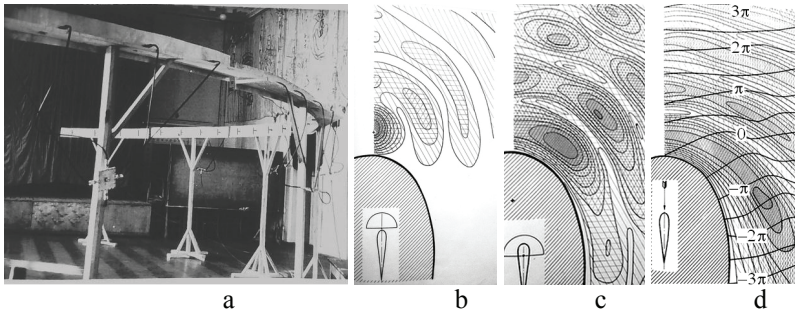
Figure 3-3: The measurement setup (a) and experimental results for the desired ARP as follows: 180°-sector pattern with a phase center at a distance of λ/4 in front of the body (b) and inside the body (c); δ-function at the axis of symmetry (d)

To the side of the photo there are three pictures of the measurement results. They show how the signal received by an electric dipole depends on its location in the area near the metal body of a certain aerodynamic profile. In Fig 3-3d, the picture shows not only the amplitude relief, as well as others, but also phase fronts with 90°-discrete. It is clear that a single emitting dipole was in use to create an incident plane wave for the case of $\mathbf{F}_0(\mathbf{r}^o) = \delta(\mathbf{r}^o - 0°)$ related to Fig. 3-3d.

In the past century, as applied to a physical body of a complex form, obtaining a solution $\mathbf{p}_0(\boldsymbol{\rho}) = V\mathbf{F}_0(\mathbf{r}^o)$ from an experiment might be of some practical interest since the necessary calculations were impossible. At present, it remains useful mainly as a curious idea to develop the capacity for creativity and innovation, since the feasibility of implementing electrodynamics modeling has increased tremendously.

The adjoint operator V has two attractive features. The first one, which the pictures in Fig. 3-3 give a clear hint at, is that the operator maps a desired ARP into some field $\mathbf{p}_0(\boldsymbol{\rho}) = V\mathbf{F}_0(\mathbf{r}^o)$, determined—existing, we can say—at any point $\boldsymbol{\rho}$, regardless of whether there is an antenna element there or not. The second one is as follows. It is easy to show[26] that in the case of a single source located at a point $\boldsymbol{\rho}$, the ratio $P_{\Sigma 0}/P_I$, that determines the source's energetic efficiency in the sense of Ex. 3-6 is proportional to $|\mathbf{p}_0(\boldsymbol{\rho})|^2$. So relief $|\mathbf{p}_0(\boldsymbol{\rho})|^2$ describes how the energy efficiency of a source with respect to the ARP $\mathbf{F}_0(\mathbf{r}^o)$ depends on its location $\boldsymbol{\rho}$.

Bearing in mind the desired pattern $\mathbf{F}_0(\mathbf{r}^o)$, it seems to be reasonable to choose the locations of array antenna's sources in the following manner. The first source is to be placed at point $\boldsymbol{\rho}_1$ of the maxima of the field

---

[26] Explanations for problem 2 in section 3-5 are helpful.

$|\mathbf{V}\,\mathbf{F}_0(\mathbf{r}^\circ)|$. Let's assume the pattern $\mathbf{F}_1(\mathbf{r}^\circ)$ is the best mean square approximation to $\mathbf{F}_0(\mathbf{r}^\circ)$ at this stage, and $\mathbf{F}_1^{\perp}(\mathbf{r}^\circ) = \mathbf{F}_0(\mathbf{r}^\circ) - \mathbf{F}_1(\mathbf{r}^\circ)$ is the difference between them which should be regarded as a desired pattern at the next step. Therefore, it is worth placing the second element at point $\boldsymbol{\rho}_2$ of the maxima of the field $|\mathbf{V}\mathbf{F}_1^{\perp}(\mathbf{r}^\circ)|$. Let's assume the pattern $\mathbf{F}_2(\mathbf{r}^\circ)$ is the best mean square approximation to $\mathbf{F}_0(\mathbf{r}^\circ)$ with the help of two chosen sources, and $\mathbf{F}_2^{\perp}(\mathbf{r}^\circ) = \mathbf{F}_0(\mathbf{r}^\circ) - \mathbf{F}_1(\mathbf{r}^\circ)$ is the difference between them, and so on.

Of course, this is a heuristic but advisable procedure also because, in addition to the positive features mentioned above, at any kth step, the value of the relief $|\mathbf{V}\mathbf{F}_{k-1}^{\perp}(\mathbf{r}^\circ)|$ equal to zero at all points $\{\boldsymbol{\rho}_i\}$ where the previous sources ($i = 1\ldots k-1$) are. Consequently, each new source cannot be too close to the others, and the antenna array designed in this way will have a reasonable spacing. This prevents the occurrence of high reactive antenna arrays.

3.4.2. Comparison results

To assess the benefits of applying the procedure described above, and to demonstrate suitable soft computing techniques, consider the following formal two-dimensional task—see Fig. 3-4a. We need to place seven infinite current filaments parallel to the edge of an ideally conducting wedge of outer angle $\alpha_0 = 270°$, bearing in mind the desired ARP $F_0(\varphi)$, which is a 180°-sector angled $\varphi_1 = 15°$ from the face of the wedge. A circle of radius 1.75 λ limits a sources' region.
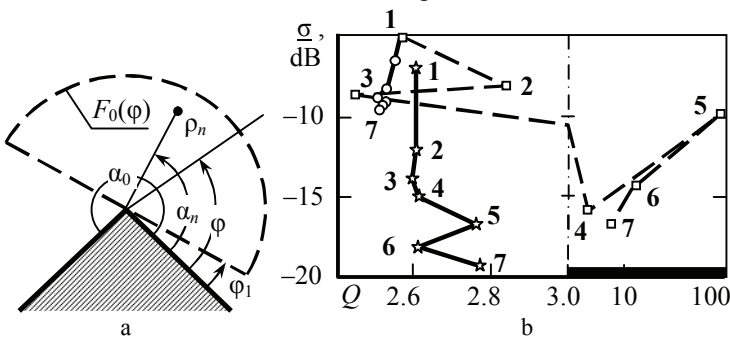
Figure 3-4: a) geometry of the task; b) MSD $\underline{\sigma}_N$ in decibel and reactive ratio Q_N for three types of arrays build of $N = 1\ldots7$ elements. Stars, circles and squares depict a type obtained using the operator V, a "flying geese" and a circle, respectively

Using Sommerfeld's strict solution of the problem of the diffraction of electromagnetic waves on an ideally conducting wedge [12] and omitting non-essential—as it is for any ARP—multipliers, we can write the ARP of a source located at the point (ρ, α) as follows

$$f(\rho, \alpha, \varphi) = \sum\nolimits_{m=1}^{\infty} \mathrm{e}^{-j\nu_m \pi/2} \, J_{\nu_m}(k\rho) \, \sin(\nu_m \alpha) \, \sin(\nu_m \varphi) \,. \qquad (3\text{-}27)$$

Here, $\nu_m = m\pi/\alpha_0$, k is the wave number, and $J_{\nu_m}(x)$ is the Bessel function of non-integer order ν_m. Using notation $f_m(\rho, \alpha)$ for coefficients of the series above converts Ex. 3-27 in a more compact form

$$f(\rho, \alpha, \varphi) \cong \sum\nolimits_{m=1}^{\infty} f_m(\rho, \alpha) \, \sin(\nu_m \varphi),$$

$$\text{where } f_m(\rho, \alpha) = \mathrm{e}^{-j\nu_m \pi/2} \, J_{\nu_m}(k\rho) \, \sin(\nu_m \alpha). \qquad (3\text{-}27')$$

On the interval $(0...\alpha_0)$ functions $\{G_m(\varphi) = \sin(\nu_m \varphi)\}$, each having square of the norm $\|G_m(\varphi)\|^2$ equal to $\alpha_0/2$, form an orthogonal basis. Therefore, the following series represents the desired ARP

$$F_0(\varphi) = \sum\nolimits_{m=1}^{\infty} a_m \, \sin(m\nu\varphi) \,, \qquad (3\text{-}28)$$

under the obvious condition—see Ex. 1-7—that the coefficients $\{a_m\}$ are as follows $a_m = (F_0(\varphi), G_m(\varphi))/\|G_m(\varphi)\|^2 = 2\left[\cos(m\nu\varphi_1 - \cos(m\nu\varphi_2)\right]/m$.

Let us consider Nth step of the described above procedure of consequent positioning the antenna elements. There are $N-1$ sources located already at points $\{(\rho_n, \alpha_n)\}$ $(n = 1...N-1)$ that produce some ARP $F_{N-1}(\varphi)$ and leave for the current step N the residual pattern $F_{N-1}^{\perp}(\varphi) = F_0(\varphi) - F_{N-1}(\varphi)$ as desired ARP. Assume that $F_{N-1}^{\perp}(\varphi) = \sum\nolimits_{m=1}^{\infty} b_m \, \sin(\nu_m \varphi)$.

First, we have to explore the amplitude relief of the function $VF_{N-1}^{\perp}(\varphi)$, which is

$$p_N(\rho, \alpha) = \left| \left(F_{N-1}^{\perp}(\varphi), f(\rho, \alpha, \varphi) \right) \right| = \frac{\alpha_0}{2} \left| \sum\nolimits_{m=1}^{\infty} b_m f_m^*(\rho, \alpha) \right|, \qquad (3\text{-}29)$$

second, to find the point (ρ_N, α_N) of its maximum for the Nth sources, third, to resolve the system of equations, corresponding to Ex. 1-11. Here, this system is $<\mathbf{L}> \mathbf{u}_N = \mathbf{v}_N$, where the matrix entries L_{nj} $(n, j = 1...N)$ are the dot products of elements' ARP $(f(\rho_j, \alpha_j, \varphi), f(\rho_n, \alpha_n, \varphi))$, and $\mathbf{v}_N$ is a column vector of dot products $(F_0(\varphi), f(\rho_n, \alpha_n, \varphi))$. The column vector $\mathbf{u}_N$ of sought-fore coefficients $\{u_n\}$ defines the current—in Nth step—antenna's pattern $F_N(\varphi) = \sum\nolimits_{n=1}^{N} u_n \, f(\rho_n, \alpha_n, \varphi)$, which provides the best mean square approximation to the ARP $F_0(\varphi)$. The new residual

$F_N^\perp(\varphi) = F_0(\varphi) - F_N(\varphi)$ will give rise to the next step. In the initial step $k = 1$, as said above, the ARP $F_{k-1}^\perp(\varphi)$, which should determine the choice of the first element, is the desired ARP $F_0(\varphi)$.

The data in the first column of Tab. 3-1 correspond to an antenna array synthesized by arranging its elements as described above. There are relieves $|p_N(\rho, \alpha)|$ for four odd steps—$N = \{1, 3\ 5, 7\}$. The shades of gray from white to black represent the amplitudes from zero to the maximum value, which, due to normalization, was always equal to one. A grid of lines with an interval of $\lambda / 4$ makes it possible to judge about the real position of any point around the wedge. The thin and bold dashed lines represent, respectively, the desired ARP $|F_0(\varphi)|$ and the ARP reproduced by the array antenna of N elements. In the current step, a new element lays in the dark area of the relief, so the white circle depicts it. In contrary, black circles depict the previously chosen elements.

The other two columns correspond to some regular geometric structures of antenna arrays, just to compare and evaluate the profitability of configuring the antenna array in a special way using the operator V. To make the comparison fair, the "wings" of the "flying geese" arrays were parallel to the wedge faces and offset from them by a distance of $\lambda / 4$. The spacing of the array elements was $\lambda / 2$, as usual in practice.

As for circular arrays, their geometry, as it seems, corresponds to a reasonable principle. The elements are distributed uniformly on a circle with a spacing of $\lambda / 2$. In addition, the outer elements are at a distance of $\lambda / 4$ from the wedge faces. This leads to a simple dependence between the radius R of the circle and the number N of elements $R = 0.5\ N\ \lambda/\ \alpha_0$.
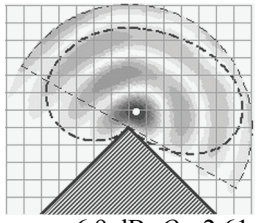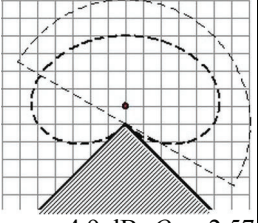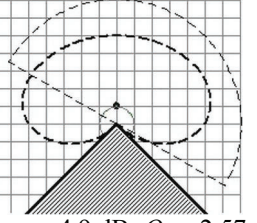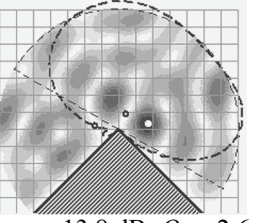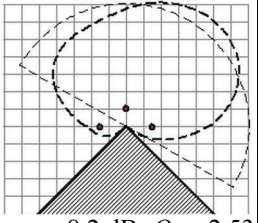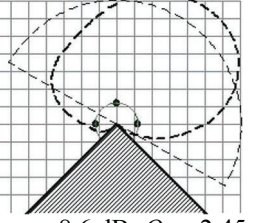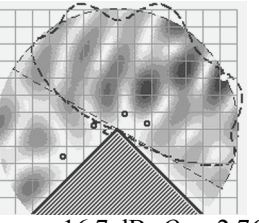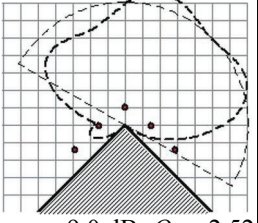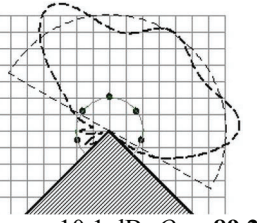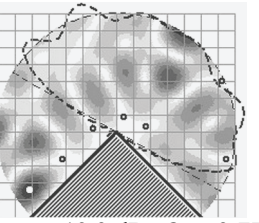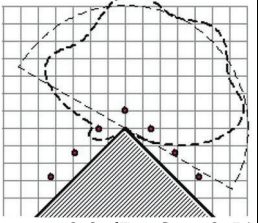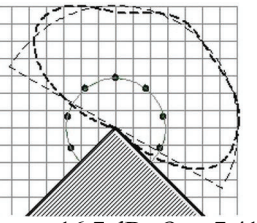
Under each picture in Tab. 3-1, there are data for the most important characteristics of the antenna, i.e. MSD σ_N, valuating how the realizable ARP $F_N(\varphi)$ approximates the required ARP $F_0(\varphi)$, and the reactivity Q of the antenna. For clarity, these data plotted in Fig. 3-4b in the form of lines marked with stars, circles and squares for the three compared cases, respectively. Moreover, there are all seven points—for $N = 1\ldots7$—on each line. The abscissa in Fig. 3-4b is composed of two segments: linear for $2.4 < Q < 3$ and a logarithmic for $3 < Q < 100$, highlighted by a thick abscissa line.

The data in Tab 3-1 and dots $\{\sigma_N, Q_N\}$ in Fig. 3-4b show that synthesized array, which has difficult to predict geometry, is the best in terms of both MSD and low reactivity. The circular arrays for $N = 5$ turns out to be very bad due to the high reactivity $Q_5 = 89.2$. It is a consequence of the unfortunate value of the radius R of the circle for a particular number $N = 5$ of sources.

To avoid the restrictions caused by the voluntary choice of regular geometry for antenna array, we conducted a statistical study. Fig. 3-5 shows its results. There is a plot with 1,500 dots, each of which represents values σ_7 and Q_7 for an antenna array of seven sources randomly placed in the following manner.

The radius of 1.75 λ, as before, limited the area for the sources. A complete set of source locations was a grid of evenly distributed 30,937 points, separated by a distance of 0.02 λ. A random number generator selected a point for the first source. In the next steps of the process, a random point supplemented the existing array only if the distance between it and each of the previous ones was greater than 0.2 λ. This reduces the probability of impractical geometry arising when the sources are too close to each other. Otherwise, a poorly radiating distribution, to which high reactivity corresponds, would appear more often.

Table 3-1: Geometry of arrays, the reproduced ARP and the quality characteristics σ$_N$ and Q_N for the odd numbers N antenna elements from one to seven

Antenna types:		
Synthesized array	"Flying geese" array	Circular array
σ$_1$ = –6.9 dB; Q_1=2.61	σ$_1$ = –4.9 dB; Q_1 = 2.57	σ$_1$ = –4.9 dB; Q_1 = 2.57
σ$_3$ = –13.8 dB; Q_3 = 2.6	σ$_3$ = –8.2 dB; Q_3 = 2.53	σ$_3$ = –8.6 dB; Q_3 = 2.45
σ$_5$ = –16.7 dB; Q_5 = 2.76	σ$_5$ = –9.0 dB; Q_5 = 2.52	σ$_5$ = –10.1 dB; Q_5 = **89.2**
σ$_7$ = **–19.2 dB**; Q_7 = **2.77**	σ$_7$ = –9.3 dB; Q_7 = 2.51	σ$_7$ = –16.7 dB; Q_7 = 7.41

Figure 3-5: MSD σ_7 and reactive ratio Q_7 for 1500 randomly generated arrays of seven elements

The biggest dot in Fig. 3-5 corresponds to the best of the 1,500 arrays. Its quality corresponds to the values $\sigma_7 = -16.5$ dB and $Q_7 = 7.78$—8.9 in decibel. The synthesized array of seven elements demonstrates better results for both $\sigma_7 = -19.2$ dB and $Q_7 = 2.77$—see highlighted in bold the data in the bottom left cell of Tab. 3-1.

A priori, it is clear that most of the randomly generated arrays will be of poor quality, and good ones will rarely appear. If $\sigma_7 < -15$ dB is an acceptable value then, in accordance with Fig. 3-5, the probability of the appearance of an acceptable array is about 0.0045.

However, we have to bear in mind that the above-promoted method of combining the antenna elements is quite reasonable, but no one has proved its optimality. Therefore, in principle, some random array can surpass the synthesized one. Nevertheless, the best results we obtained after a large number of repetitions of a random choice were σ_7 around -17.2 dB and Q_7 in the range from 5 to 15.

3.5. Training tasks

Task 3A

Taking into account that $\gamma_0 = (\mathbf{F}_k^{\perp}(\mathbf{r}^\circ), \mathbf{U}\mathbf{p}_k(\boldsymbol{\rho}))/\parallel \mathbf{U}\mathbf{p}_k(\boldsymbol{\rho}) \parallel^2$, show that $\sigma^2 = \parallel \mathbf{F}_0(\mathbf{r}^\circ) - \gamma_0 \mathbf{U}\mathbf{p}_0(\boldsymbol{\rho}) \parallel^2 / \parallel \mathbf{F}_0(\mathbf{r}^\circ) \mid$ coincides with the normalized mean square deviation of antenna pattern's form $\underline{\sigma}^2$ given by Ex. 2-48 after substitution $\mathbf{F}(\mathbf{r}^\circ) = \mathbf{U}\mathbf{p}_0(\boldsymbol{\rho})$.

Task 3B

Show that for a single source located at a point $\boldsymbol{\rho}$, the ratio $P_{\Sigma 0}/P_{\mathbf{I}}$ is equal to $|\mathbf{p}_0(\boldsymbol{\rho})|^2$. Here, $\mathbf{p}_0(\boldsymbol{\rho}) = V\mathbf{F}_0(\mathbf{r}^o)$, $P_{\Sigma 0}$ is the power radiated into desired ARP—see Ex. 2-46—and $P_{\mathbf{I}}$ is the power of excitation that turns to $|\mathbf{p}_0(\boldsymbol{\rho})|^2$ in the considered case.

Task 3C

The setup for measuring the APhD $\mathbf{p}(\boldsymbol{\rho}) = V\mathbf{F}_0(\mathbf{r}^o)$, shown in Fig. 3-3, needs a device tuning the excitation distribution in accordance with the function $\mathbf{F}_0{}^*(\mathbf{r}^o)$. Unfortunately, you have a tuning device that is actually a multi-channel receiver which cannot work in transmission mode. Is it possible to reorganize the system so that it uses the given multi-channel receiver as a tuning device? If yes then prove it.

Task 3D

Is it possible to modify the measurement setup shown in Fig. 3-3 to get rid of phase shifters that adjust arg(F0*(ro)), and confine ourselves to using power dividers only?

Task 3E

Make sure that the formula in Ex. 3-2 for adjoint operator V satisfies the condition Ex. 3-1.

Chapter Four

Phase Pattern of the Antenna as an Additional Factor in Optimization

4.1. Introduction

Not only for reading this chapter but in principle it is important to keep in mind that ARP $F(\mathbf{r}^{\circ})$, describing the far field of an antenna at a certain frequency, cannot be anything but a complex function of the direction represented by the radial unit vector $\mathbf{r}^{\circ}$. In calculations, it usually consists of the real and imaginary parts $F(\mathbf{r}^{\circ}) = F'(\mathbf{r}^{\circ}) + jF''(\mathbf{r}^{\circ})$ but from a physical point of view, it is more meaningful to present it as a combination of the amplitude $A(\mathbf{r}^{\circ})$ and phase $\psi(\mathbf{r}^{\circ})$ patterns $F(\mathbf{r}^{\circ}) = A(\mathbf{r}^{\circ})e^{j\psi(\mathbf{r}^{\circ})}$. Even if ARP $F(\mathbf{r}^{\circ})$ is a real function $F'(\mathbf{r}^{\circ})$ or $A(\mathbf{r}^{\circ})$—if we allow(!?) amplitude to be negative—it does not mean a disappearing phase pattern at all. This means only that the antenna radiates, or we want it to create, an ideally spherical wave with a phase center exactly at the origin.

In practice, the phase patterns of the overwhelming majority of antennas do not play any role and can be arbitrary. The exceptions are feeds of parabolic antennas or lenses that must emit a spherical wave, or antennas of some rare navigation systems for which the first variation of phase $\psi(\varphi) = \varphi$ is of great importance. By the way, on-board satellite navigation antennas for both GPS and GLONASS systems must meet extremely stringent requirements for the stability of their phase diagrams. However, these antennas are unique in terms of their number.

It would seem that the design of the antenna for the required amplitude pattern should be the mainstream problem in the theory of synthesis. However, the situation is different. As far as I can judge, there are two reasons for this. First, the magnitude of a series of complex functions nonlinearly depends on the coefficients of the series. Therefore, the minimization of the deviation of the antenna's amplitude pattern $|F(\mathbf{r}^{\circ})|$ from the required amplitude pattern $|F_0(\mathbf{r}^{\circ})|$—even in the sense of least squares—refers to the issues of non-convex programming [29] with all its difficulties, not to mention the problem of non-uniqueness of the solution.

Second, antennas of widely used geometry, such as linear, circular, planar array or aperture, usually have an ARP with the phase center at their physical center.

> Among the many works that the Internet responds with to the request "non-convex programming", I would like to recommend the presentation of the lecture "Non-convex optimization"[27], which suites for those who intend to get an idea of the problem, avoiding deep immersion in it.

As mentioned above, the mainstream research concerns the problems of synthesis in the classical formulation for the desired ARP, given as a complex function. This is the case even if sometimes they look like real functions. In fact, as stated above, these are just situations that correspond to the phase center at the origin. The least squares method for approximating the required complex ARP $F_0(\mathbf{r}^\circ)$ is well developed and provides analytical solutions based on the orthogonal projection scheme described in Section 1.3.4. The method is applicable to very general structures of antennas, for example, an antenna array of any geometry consisting of not-identical sources, as well as sources distributed continuously over some surface.

It is obvious that the solutions of the synthesis problems depend both on the prescribed phase pattern $\psi_0(\mathbf{r}^\circ)$ and on the desired amplitude pattern $|F_0(\mathbf{r}^\circ)|$. If a phase pattern $\psi_0(\mathbf{r}^\circ)$ does not matter, why not use it as an additional factor to enhance the sought-for solution? This chapter is about how to actualize this opportunity in several situations for relevant purposes.

4.2. Synthesis of the antenna array for a desired amplitude pattern

Assume that the array antenna consists of N elements with individual ARPs $\{f_n(\mathbf{r}^\circ)\}$ $(n = 1…N)$ and the desired amplitude ARP is $A_0(\mathbf{r}^\circ)$. To use the least squares method, let us prescribe the desired ARP $F_0(\mathbf{r}^\circ) = A_0(\mathbf{r}^\circ)\,e^{j\psi_0(\mathbf{r}^\circ)}$, where $\psi_0(\mathbf{r}^\circ)$ is an arbitrarily selected function, to which we assign the role of a desired phase pattern. In particular, it can be $\psi_0(\mathbf{r}^\circ) = $ const, say, zero.

By making the N x N matrix <**L**>, which entries $L_{nk} = (f_k(\mathbf{r}^\circ), f_n(\mathbf{r}^\circ))$ are equal to inner products of individual ARP of the antenna elements $(n, k = 1…N)$, and solving the following system of equations

$$<\mathbf{L}>\mathbf{W} = \mathbf{p}, \tag{4-1}$$

---

[27] URL: http://www.cs.cornell.edu/courses/cs6787/2017fa/Lecture7.pdf

where **p** is the vector of N entries $p_n = (F_0(\mathbf{r}^\circ), f_n(\mathbf{r}^\circ))$, we harvest the vector **W** of N excitation coefficients $\{W_n\}$, which provides a minimum to the squares deviation of realized ARP $F(\mathbf{r}^\circ) = \sum_n W_n f_n(\mathbf{r}^\circ)$ from the desired one $F_0(\mathbf{r}^\circ)$.

Let us denote the amplitude and phase patterns of the realized ARP $F(\mathbf{r}^\circ)$ by $A_1(\mathbf{r}^\circ)$ and $\psi_1(\mathbf{r}^\circ)$, respectively. Otherwise, this means $F(\mathbf{r}^\circ) = A_1(\mathbf{r}^\circ)e^{j\psi_1(\mathbf{r}^\circ)}$. Denote the square deviations of the complex and amplitude patterns as $\sigma_1^2 = \|F_0(\mathbf{r}^\circ) - F(\mathbf{r}^\circ)\|^2$ and $\Delta_1^2 = \|A_0(\mathbf{r}^\circ) - A_1(\mathbf{r}^\circ)\|^2$, respectively.

Since the square norm $\|...\|^2$ means integration $\int_\Omega |...|^2 \, d\Omega$, where Ω is a manifold of the variable $\mathbf{r}^\circ$—simply put, the area of a sphere or part of a circle of a unit radius in which ARPs exist—then the inequality $\sigma_1^2 \geq \Delta_1^2$ holds true in any case. For further reasoning, it is a very important fact that inequality turns into equality only if phase patterns $\psi_0(\mathbf{r}^\circ)$ and $\psi_1(\mathbf{r}^\circ)$ coincide.

If the phase pattern $\psi_1(\mathbf{r}^\circ)$ differs from our choice of $\psi_0(\mathbf{r}^\circ)$, then let us respect the antenna's "response" and use it as a new prescribed phase pattern in combination with a fixed amplitude pattern. This leads to a new desired pattern $F_0(\mathbf{r}^\circ) = A_0(\mathbf{r}^\circ)e^{j\psi_1(\mathbf{r}^\circ)}$ and, consequently, to a new vector **p**. Solving the updated system of equations given by Ex. 4-1, we get a new vector **W** of excitation coefficients, which produces a new ARP $F(\mathbf{r}^\circ) = A_2(\mathbf{r}^\circ)e^{j\psi_2(\mathbf{r}^\circ)}$. How does this affect the square of the deviation at the second step σ_2^2? Let us evaluate its value.

In the manifold of vectors **W**, there is the previous vector $\mathbf{W}_1$. If it appeared at the current stage, then σ_2^2 would be equal to Δ_1^2. Indeed, the ARP $F_2(\mathbf{r}^\circ)$ would coincide with $F_1(\mathbf{r}^\circ) = A_1(\mathbf{r}^\circ)e^{j\psi_1(\mathbf{r}^\circ)}$. Since the current desired pattern is $A_0(\mathbf{r}^\circ)e^{j\psi_1(\mathbf{r}^\circ)}$, the following chain of equalities would be true $\sigma_2^2 = \left\|A_0(\mathbf{r}^\circ)e^{j\psi_1(\mathbf{r}^\circ)} - A_1(\mathbf{r}^\circ)e^{j\psi_1(\mathbf{r}^\circ)}\right\|^2 = \left\|A_0(\mathbf{r}^\circ) - A_1(\mathbf{r}^\circ)\right\|^2 = \Delta_1^2$. However, the least squares method generates—by solving the matrix equation Ex. 4-1—a different vector **W**, which is the best in sense of square deviation σ^2. Consequently, σ_2^2 is less than Δ_1^2.

Continuing the above process of successive corrections of the prescribed phase pattern, we get a number of gradually improving solutions. The following infinite chain of inequalities characterizes their accuracy

$$\sigma_1^2 > \Delta_1^2 > \sigma_2^2 > \Delta_2^2 > \ldots > \sigma_k^2 > \Delta_k^2 > \sigma_{k+1}^2 > \Delta_{k+1}^2 > \ldots \qquad (4\text{-}2)$$

Ex. 4-2 clearly shows that reducing the deviation of complex APR goes in parallel with bettering the amplitude pattern, which is of the highest concern to us.

From Ex. 4-2, it seems almost obvious that the sequence of the phase patterns $\{\psi_k(\mathbf{r}^\circ)\}$ arising during the above process converges to a certain phase pattern $\psi^{st}(\mathbf{r}^\circ)$. A particularity of this pattern is that if we combine it with the desired amplitude pattern $A_0(\mathbf{r}^\circ)$—prescribing the complex pattern $F_0(\mathbf{r}^\circ)$—then a phase pattern of the realizable ARP $F(\mathbf{r}^\circ)$ is exactly the phase pattern $\psi^{st}(\mathbf{r}^\circ)$. In terms of Chapter 2—see the last paragraph of Section 2.2.1—the following concise expression represents this statement

$$F(\mathbf{r}^\circ) = U_\sigma^{-1}\, A_0(\mathbf{r}^\circ)\, e^{j\underline{\psi^{st}(\mathbf{r}^\circ)}} = A^{st}(\mathbf{r}^\circ)\, e^{j\underline{\psi^{st}(\mathbf{r}^\circ)}}\,. \tag{4-3}$$

Here, the lines under the phase patterns are to emphasize the mentioned fact. Since the quasi-inverse operator U_σ^{-1} maps the phase pattern $\psi^{st}(\mathbf{r}^\circ)$ into itself—in combination with a fixed amplitude diagram, of course—it is logical to call it a stationary phase pattern. The superscript used comes from this name.

In my thesis [16], I proved several theorems regarding the synthesis of the desired amplitude patterns. My article [30] includes the three most valuable ones.

Consider the functional

$$\Delta^2(\mathbf{W}) = \left\| A_0(\mathbf{r}^\circ) - \left| \sum_n W_n f_n(\mathbf{r}^\circ) \right| \right\|^2, \tag{4-3}$$

which serves a measure of mean square deviation for two amplitude patterns: the desired one $A_0(\mathbf{r}^\circ)$ and the realizable one corresponding to the excitation vector $\mathbf{W}$. Vector $\mathbf{W}$ can be arbitrary.

First, there is at least one stationary phase diagram for any antenna and an arbitrary desired amplitude pattern. Second, any point of the local minimum of the functional $\Delta^2(\mathbf{W})$ corresponds to a certain stationary phase pattern. Third, strictly speaking, the converse is not true, since the functional $\Delta^2(\mathbf{W})$ may have saddle points [31] which also correspond to stationary patterns, but rather specific ones. Their meaningful name is unsustainable stationary phase patterns.

Indeed, suppose that the above calculations came very close to a saddle point moving towards its minimum. Due to the finite accuracy of the calculations, a current point $\mathbf{W}$ can easily "slide" to the area where the saddle shows itself as having a maximum. Then the process will move

away—perhaps rather slowly at the beginning—from the saddle point and come to the end at some genuine local minimum[28].

Unfortunately, unlike the functional $\sigma^2(\mathbf{W})$, minimization of which is the main issue of Chapter 3, the functional $\Delta^2(\mathbf{W})$ is non-convex. This means that it may have several, perhaps a number of local or even global, minima. We can easily construct an appropriate formal example. Obviously, if an array antenna consists of N elements, the individual amplitude patterns of which are the same, say, $a_o(\mathbf{r}^\circ)$, and the required amplitude pattern $A_o(\mathbf{r}^\circ)$ is equal to $a_o(\mathbf{r}^\circ)$, then the functional $\Delta^2(\mathbf{W})$ has N global minima $\Delta^2 = 0$, each of them corresponds to excitation of the single element of the array. Thus, in this situation, a set of stationary phase patterns has at least N phase patterns that are the phase patterns of the antenna elements.

I tried to find out how many stationary phase patterns an array antenna can have. At first, it seemed me that no more than the number of elements. However, the simplest example had destroyed that hope. Let the two above-mentioned elements be in points $x_1 = -d$ and $x_2 = d$, where spacing d is less than $\lambda/2$, and $A_o(\varphi)$ is $a_o(\varphi)$, as before. Then, in addition to the two stationary phase patterns, which are $\psi^{st}(\varphi) = \exp(\pm j\,k\,d\cos(\varphi))$, the third one $\psi^{st}(\varphi) = 0$ appears, as is easy to see. The question also remains open whether there is a combination of the array antenna and the desired pattern $A_o(\mathbf{r}^\circ)$, to which the only stationary phase pattern corresponds.

To demonstrate how the process of an iterative approach to a desired amplitude pattern works, consider the array antenna shown in Fig. 4-1.
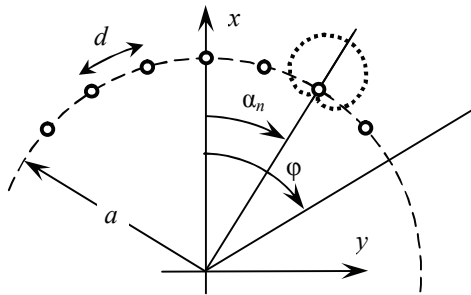

Figure 4-1: Geometry of an array antenna of elements with cardioid ARP

---

[28] This is a rare situation when we have to thank the finite preciseness of the calculus.

On a circular arc of radius $R_a = 2\lambda$, there are seven sources distributed symmetrically versus the x-axis with a spacing $d = 0.55\lambda$. Therefore, their angular coordinates are $\alpha_n = (n - (N-1)/2)\, d/R_a$, where $N = 7$ and $n = 0\ldots N-1$. Striving for a nontrivial situation, let us assume that the required amplitude of ARP is a non-symmetrical sector function $A_0(\varphi) = 1$ for the interval ($\varphi_1 = -45° < \varphi < \varphi_2 = 75°$) and $A_0(\varphi) = 0$ outside of it.

A complete set of computational formulas is as follows. Elements' ARPs are $f_n(\varphi) = (1 + \cos(\varphi - \alpha_n))\, e^{jka\cos(\varphi - \alpha_n)}$. In Ex. 4-1, the entries of the matrix **L**, which are mutual inner products of the functions $f_n(\varphi)$, are $L_{n,k} = \int_0^{2\pi} f_k(\varphi) f_n^*(\varphi)\, d\varphi$. The column vector **p** depends on the phase pattern $\psi_m(\varphi)$ at the current step m, and its entries are $p_n = \int_{\varphi_1}^{\varphi_2} e^{j\psi_m(\varphi)} f_n^*(\varphi)\, d\varphi$.

It is surprising and interesting that the entries $L_{n,k}$ turn out to be real values. Their explicit expression follows from a very useful series—for analytical investigation in antenna analysis and synthesis—that is $e^{jka\cos(\varphi)} = \sum_{n=0}^{\infty} (2 - \delta_{0n})\, j^n J_n(ka) \cos(n\varphi)$, where δ_{0n} is a Kronecker symbol. Some tips on how to derive the expression are contained in the training tasks 4-D and 4-E, aiming towards such a valuable facet of engineering style of thinking as the ability to perform analytical studies. Of course, modern numerical modeling tools reduce the need for analytical research. However, only analytical studies can reveal universal properties or regularities of the problem under interest.

Fig. 4-2 – Fig. 4.5 demonstrate how the considered recursions work depending on initial phase patterns $\psi_0(\varphi)$. The dash-point line depicts the desired amplitude pattern $A_0(\varphi)$. The bold dashed line and the bold line correspond to the ARP $F_1(\varphi)$, respectively, in the first step $m = 1$ and in the final step $m = 80$ of the procedure. The thin gray lines show ARP $F_m(\varphi)$ for an intentionally selected sequence of steps $m = 2, 3, 6, 11, 20, 40$. The idea of choice is clear: to take into account that the rate of convergence slows down dramatically when the process approaches a point of a local minimum.

Amplitude and phase patterns refer to the left and right axis, respectively. Phase patterns are of interest—and, by the way, well determined—only on the interval ($\varphi_1 < \varphi < \varphi_2$). In figures captions, there are values Δ_m^2 of the square deviation of the amplitude patterns $A_m(\varphi) = |F_m(\varphi)|$ from the desired one.
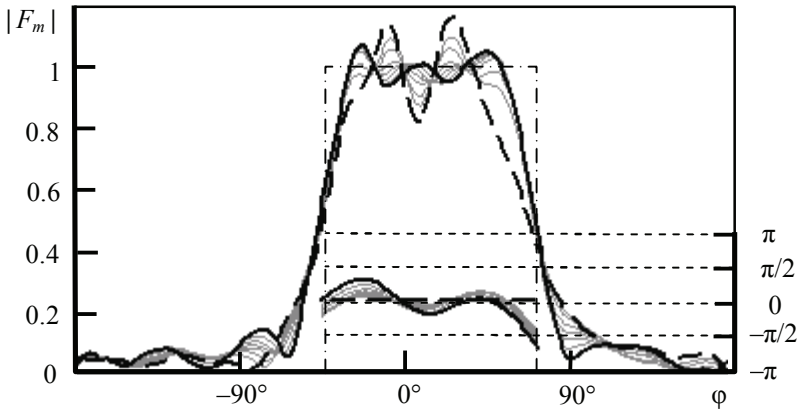
Figure 4-2: Synthesizing the sector amplitude pattern $A_o(\varphi)$. The initial phase pattern $\psi_0(\varphi) = 0$ corresponds to phase center at the origin. Deviations Δ_m^2 are $\Delta_1^2 = 0.14$; $\Delta_2^2 = 0.11$; $\Delta_3^2 = 0.10$; $\Delta_6^2 = 0.097$; $\Delta_{11}^2 = 0.095$; $\Delta_{20}^2 = 0.092$; $\Delta_{40}^2 = 0.089$; $\Delta_{80}^2 = 0.086$
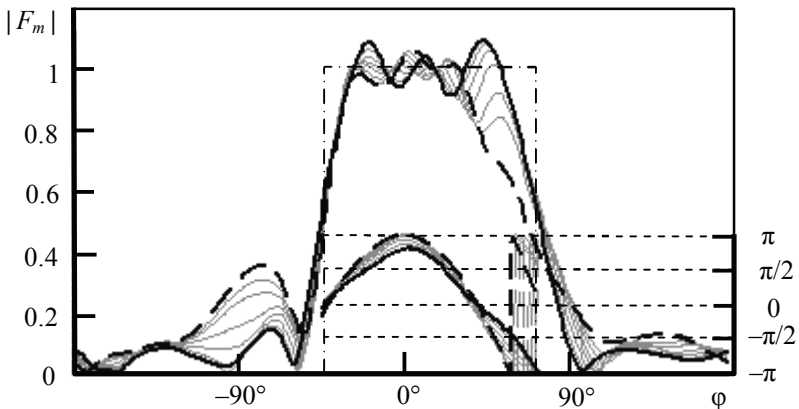
Figure 4-3: Synthesizing the sector amplitude pattern $A_o(\varphi)$. The initial phase pattern $\psi_0(\varphi) = k\,a\,\cos(\varphi)$ corresponds to phase center at the central element of the array. Deviations Δ_m^2 are $\Delta_1^2 = 0.21$; $\Delta_2^2 = 0.16$; $\Delta_3^2 = 0.14$; $\Delta_6^2 = 0.12$; $\Delta_{11}^2 = 0.105$; $\Delta_{20}^2 = 0.096$; $\Delta_{40}^2 = 0.091$; $\Delta_{80}^2 = 0.089$

The choice of $\psi_0(\varphi) = 0$ represented by Fig. 4-2, is a formal but rather typical one for those who do not take into account that this means gratifying the desire to have a phase center of ARP at the origin, located quite far outside the area occupied by the antenna sources. It seems to be

better to place the phase center close to the antenna area, say, at the point $(a, 0)$ of the central source. Fig. 4-2 shows the corresponding results that do not meet our expectations for a better approximation. Before drawing any conclusions, let us try two more situations with a phase center to the left and right side of the area.
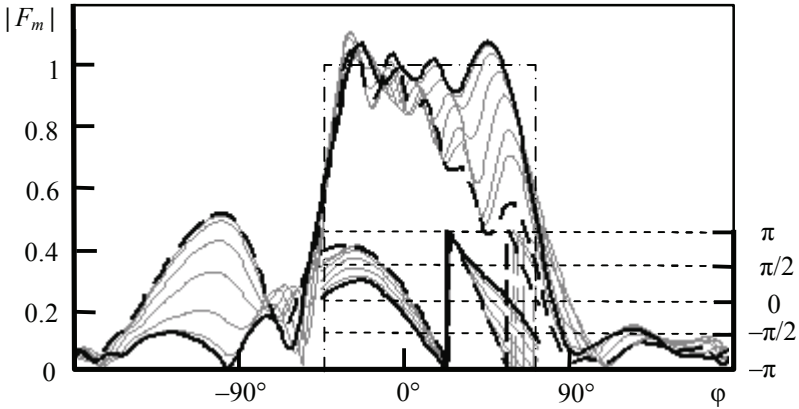
Figure 4-4: Synthesizing the sector amplitude pattern $A_0(\varphi)$. The initial phase pattern $\psi_0(\varphi) = 0.8\,ka\,\cos(\varphi) - 0.5\,ka\,\sin(\varphi)$ corresponds to phase center by the left side element of the array. Deviations Δ_m^2 are $\Delta_1^2 = 0.39$; $\Delta_2^2 = 0.34$; $\Delta_3^2 = 0.31$; $\Delta_6^2 = 0.25$; $\Delta_{11}^2 = 0.18$; $\Delta_{20}^2 = 0.13$; $\Delta_{40}^2 = 0.11$; $\Delta_{80}^2 = 0.10$
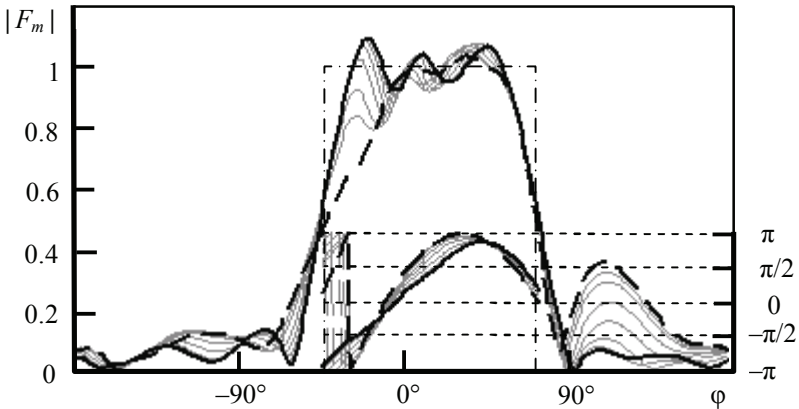
Figure 4-5: Synthesizing the sector amplitude pattern $A_0(\varphi)$. The initial phase pattern $\psi_0(\varphi) = 0.8\,ka\,\cos(\varphi) + 0.5\,ka\,\sin(\varphi)$ corresponds to phase center by the right side element of the array. Deviations Δ_m^2 are $\Delta_1^2 = 0.20$; $\Delta_2^2 = 0.16$; $\Delta_3^2 = 0.14$; $\Delta_6^2 = 0.12$; $\Delta_{11}^2 = 0.102$; $\Delta_{20}^2 = 0.089$; $\Delta_{40}^2 = 0.083$; $\Delta_{80}^2 = 0.083$

What do we see? First, all four stationary phase patterns differ from each other and are rather close to the initial ones. The most amazing thing is that in the situation shown in Fig. 4-2 it is as if the local phase center was moving inside a rather small area at the origin. This means that the arc antenna array can create a field, whose phase front in a limited range of angles corresponds mainly to an expanding cylinder. This explains the seeming paradox.

Second, in each situation the process comes to almost the same value of deviation Δ^2 of about 0.086. This makes us think that the relief of the functional $\Delta^2(\mathbf{W})$ resembles a hilly landscape with dents of the same depth. By the way, I found that the phase pattern $\psi_0(\varphi) = 0.4\ ka\ \cos(\varphi) - 0.03\ ka\ \sin(\varphi)$, corresponding to the phase center at the point $(0.4\ a, -0.03\ a)$, almost exactly is stationary. At least, the deviation Δ^2 value was equal to 0.088 from the first step to the 80th.

It is quite predictable that starting with the function $\psi_0(\varphi)$ corresponding to the phase center far away from the antenna area, the process converges very quickly from an initial value of 0.6 or more to a value of about 0.09 with a phase pattern similar to one of the above.

Since the subject of the book is mainly ideas and methodology for solving problems of synthesis and not results in themselves, I provide interested readers with a program in order they to try to investigate functional $\Delta^2(\mathbf{W})$ on its local and global minima.

4.3. Modified Woodward-Lawson method

The idea of the Woodward-Lawson method (WLM), meaningfully called the method of superposition of beams, is very sound and at the same time rather simple. Obviously, the ARP of a linear antenna having a length L and consisting of isotropic sources with the excitation distribution $I(x) = I_0 \exp(-j\,\beta\,x)$, corresponds—after scaling—to the formula

$$F(u) = \int_{-L/2}^{L/2} \exp\left(j(kx\sin(\theta) - \beta x)\right) dx = \frac{\sin(u - u_\beta)}{u - u_\beta} = \mathrm{sinc}(u - u_\beta), \qquad (4\text{-}4)$$

where $u = L\pi\sin(\theta)/\lambda,\quad u_\beta = \beta L/2.$

As usual, θ is a real angle from the antenna normal and the variable u is a so-called generalized angular variable uniquely associated with θ in the interval $\pm\pi L/\lambda$.

Ex. 4–4 defines a beam of a typical form $\mathrm{sinc}(u - u_\beta)$ that has the equal to unity maximum at the point $u = u_\beta$ and zeros at all the points $u_n = u_\beta \pm n\pi$, where n is an integer. The WLM uses this particularity of the sinc-function to synthesize the desired ARP $F_0(\theta)$. It deals with a set of

beams $\{g_n(u) = \mathrm{sinc}(u - u_0 - n\pi)\}$ to approximate the function $F_0(u)^{[29]}$ by the series

$$F(u) = \sum_n A_0(u_0 + n\pi)\, g_n(u) = \sum_n F_0(u_0 + n\pi)\, \mathrm{sinc}(u - u_0 - n\pi),$$

$$I(x) = \sum_n F_0(u_0 + n\pi)\, \mathrm{e}^{-j(u_0 + n\pi)x}. \tag{4-5}$$

Here index n alters in the interval satisfying the inequality $-N\pi < u_0 + n\pi < N\pi$, were N is a truncated value of L/λ, and coefficients a_n are equal to the values of the desired ARP at the directions $\{u_0 + n\pi\}$.

It is worth noting that WLM is attractive not only by extreme simplicity of obtaining the solution Ex. 4-5. The reactivity $Q = \|I(x)\|^2 / \|F(\theta)\|^2$ of the generated solution $I(x)$ is rather low by virtue of the fact that the functions $\{f_n(u)\}$ form an almost orthogonal basis due to a proper spacing for sinc-functions, which is π.

An attentive reader can note that the traditional expressions for WLM do not contain the term u_0, as if the value of u_0 in Ex. 4-5 is zero. What is the meaning of the term u_0? If $u_0 = 0$, then the maxima of the functions $f_n(u)$ are located at the points $u_n = n\pi$ that are multiples of π. Alteration of the term u_0 results in shifting the entire set of functions $\{f_n(u)\}$ along the u-axis by the distance u_0 without changing their mutual arrangement. Suppose that the required ARP $F_0(u)$ has a sharp maximum at some point u_0. It is obvious that among the set $\{f_n(u)\}$ it is useful to have a function whose maximum is at this point precisely. Ex. 4-5 provides this opportunity, not excluding the traditional situation.

Due to mentioned particularities of the base functions $\{f_n(u)\}$ the values of both ARPs—$F(u)$ specified by Ex. 4-5, and $F_0(u)$—coincide in the set of points $\{u_n = u_0 + n\pi\}$, which means coincidence of functions $F(\theta)$ and $F_0(\theta)$ in the set of corresponding angles $\{\theta_n\}$.

The proposed by us [32] modification of the WLM includes much more significant thing than a mere nothing u_0. The idea stems from a rather instructive point of view inherent in engineering style of thinking. I mean the permanent attention to and disclosure of the physical meaning of mathematical formalities.

It is important to bear in mind that the desired ARP in Ex. 4-5 is actually a complex function. Even if we set its values $\{F_0(u_n)\}$ as the real values, this—as above—just means that we require the ARP $F_0(u)$ having a phase center at the origin. However, we are omly interested in a certain amplitude pattern $A_0(u)$, while the phase pattern does no matter and can be arbitrary.

---

[29] This function is the desired ARP, explicit expression of which appears after replacing u by $L\pi \sin(\theta)/\lambda$.

The solution given by Ex. 4-5 has a very interesting feature. Due to the peculiarities of the used basis functions $\{g_n(u) = \text{sinc}(u - u_0 - n\pi)\}$, the equalities $F(u_n) = a_n = F_0(u_n)$ are true. Consequently, the equations $A(u_n) = |F(u_n)| = |F_0(u_n)| = A_0(u_n)$ are true regardless of the phase terms $\{\psi_n)\}$ in the set $\{F_0(u_n) = A_0(u_n) \exp(j\,\psi_n)\}$. Why not use this extra degree of freedom to improve the amplitude ARP $A(u) = |F(u)|$ in reference to the desired one $A_0(u)$?

It is clear that finding the optimal set of phase values $\{\psi(u_n)\}$ as a solution of the task $\min\limits_{\{\psi_n\}} \| A_0(u) - |\sum_n A_0(u_n)e^{j\psi_n}g_n(u)| \|^2$ is a pretty hard problem of non-convex optimization, the main challenge of which is imposed by the presence of a set of local minima. A common method of solving such a problem is stochastic gradient descent, for which the better the starting point—that is the initial set $\{\psi_n\}$—the better. Of course, we can force a computer to test many random sets $\{\psi_n\}$ for this, but it is very thrilling to try to help—or even compete—the computer with human shrewdness.

Let us start thinking. In any case, we can equate to zero one of the phases of a set. It is logical to do this for the "main", say, the nth element of maximum amplitude $|a_n|$. Changing the phase ψ_{n+1} of the next term will noticeably affect how the function $|F(u)|$ behaves in the interval $u_0 + n\pi < u < u_0 + (n+1)\pi$ between the nth and $(n+1)$th reference points, whereas the conditions $|F(u_n)| = A_0(u_n)$ and $|F(u_{n+1})| = A_0(u_{n+1})$ remain intact. Therefore, there is a certain value ψ_{n+1} that provides the best approximation to the required function $A_0(u)$ in the considered interval. For instance, we can try to fulfill the equality $|F(u_k)| = A_0(u_k)$, where $u_k = (u_n + u_{n+1})/2$ is the midpoint of the interval. On the same reason we will find the value ψ_{n+2} taking into account the next interval $u_0 + (n+1)\pi < u < u_0 + (n+2)\pi$, etc. In the same way, we can move down from the nth element.

> **Remark**: it is worth emphasizing that such a method of optimizing the phase set uses the fact that each sinc-function $g_n(u)$ acts efficiently in its rather narrow local range because outside the interval $\pm\pi$ its values are relatively small. Of course, the above-mentioned iterative procedure easily turns into a computational program, but it is quite fascinating to execute it manually. In Appendix D, there is a corresponding Mathcad program called "modified WLM" to provide interested readers with the pleasure of beating our results shown below. Certainly, it is possible.

To demonstrate the above way of enhancing the WLM's solutions we considered the following task: to synthesize the cosecant pattern:

$A_0(\theta) = \sin(\theta_0)/\sin(\theta)$ if $\theta_0 \leq \theta \leq \pi/2$, and $A_0(\theta) = 0$ otherwise, with a ten-wavelength line source. The angle of a cosecant maximum $\theta_0 = 6°$. In terms of the generalized angle $u = 10 \pi \sin(\theta)$ we have $u_0 = 3.284$, $A_0(u) = u_0/u$ and $a_n = u_0/(u_0 + (n - 10) \pi)$ if $n \geq 10$, and $a_n = 0$ otherwise. Thus, Ex. 4-5 turns to be

$$F(u) = \sum_{n=10}^{20} A_0(u_0 + n\pi) e^{j \psi_n} \operatorname{sinc}(u - u_0 - n\pi) . \tag{4-6}$$

In several minutes, we came to the following set of phases $\{\psi_{10}, \ldots \psi_{20}\} = \{0°, 88°, 39°, 95°, 50°, 89°, 44°, 103°\}$. Fig. 4-6 shows the gained results. The x-axis is actually the variable u divided by $L\pi/\lambda$. What positive changes related to the modified WLM can we see? First of all, in addition to the reference points $\{u_n = u_0 + n \pi\}$ where ARP $|F(u_n)|$ and $A_0(u_n)$ are the same—marked by circles—new points of coincidence appear between them. Second, the ripple to the right of u_0 has become noticeably less.

Figure 4-6: Synthesizing cosecant ARP: gray line **1** is the desired amplitude pattern $A_0(\theta)$; bold dashed gray line **2** and bold solid black line **3** are ARP derived from classical WLM and modified WLM respectively

Remark: in [33], the authors proposed an iterative algorithm for the improvement of an original ARP of a linear source by adding new beams where local amendments in the current ARP are desirable. In particular, they demonstrate, as we do, a cosecant ARP of a ten-wavelength line source produced in accordance with WLM and after improving. They did not change the phase pattern, and the results obtained are less impressive.

Unfortunately, the side lobes to the left of u_0 have increased. This may be acceptable if the cosecant ARP illuminates the ground and side lobes

look up, where there are no sources of intense interference or noise. However, it would be better to rectify the situation here. Section 2.6 gives one of the possible approaches to the local amending ARP. However, we can think about less strict limitations on the side lobes, just wanting to lower them. Then the following way can match these intentions.

Since the side lobes have maxima between the zeros located at $\{u_n = u_0 + (n - 10)\,\pi\}$ $(n = 0..9)$, so it is worth making a pi/2-shift for all these ten functions in order to align the positions of their maxima with the maxima of the side lobes. In result, we obtain functions $\{\tilde{g}_n(u) = \mathrm{sinc}(u - u_n + \pi/2)\}$ $(n = 0,..9)$. In Fig. 4-6, thin solid and dashed lines depict two shifted functions $\tilde{g}_9$ and $\tilde{g}_8$, respectively. For clarity, we have lowered them by 5 dB.

Adding these functions with weights $\{b_n\}$ to the Ex. 4-6 gives $\tilde{F}(u) = F(u) + \sum_{n=0}^{9} b_n\,\tilde{g}_n(u)$. There are several ways to determine the coefficients $\{b_n\}$, including the heuristic one. To reduce values $\{F(\tilde{u}_n)\}$ of side lobes' maxima it seems to be reasonable to set $b_n = -\mu\,F(\tilde{u}_n)$, where μ is a real coefficient of value $0 < \mu < 1$.

$$\tilde{F}(u) = F(u) - \mu \sum_{n=0}^{9} F(\tilde{u}_n)\,\tilde{g}_n(u) . \tag{4-7}$$

This rule provides subtraction of the "beams" $\tilde{g}_n(u)$ from the corresponding side lobe, and μ defines the degree of suppression. The bold black line in Fig. 4-7 shows the amplitude ARP according to Ex. 4-7 with $\mu = 0.64$.
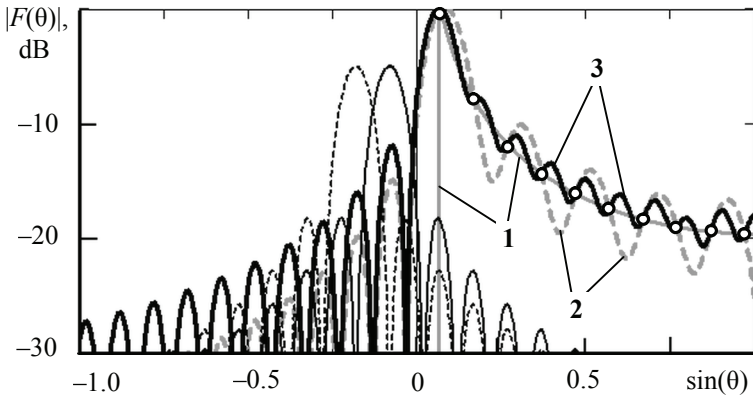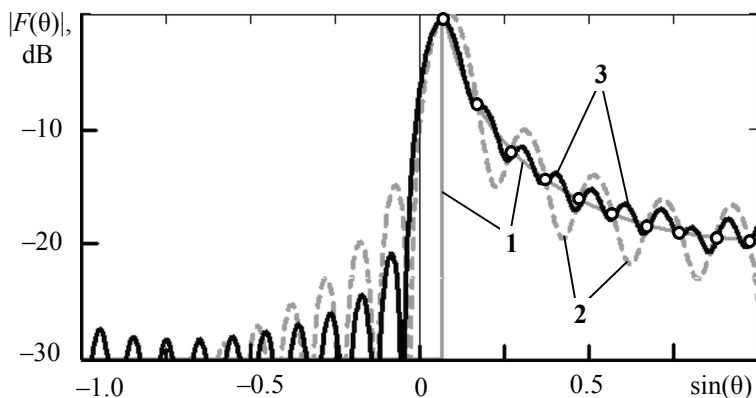

Figure 4-7: Synthesizing cosecant ARP: gray line **1** is the desired amplitude pattern $A_0(\theta)$; bold dashed gray line **2** and bold solid black line **3** are, respectively, ARP derived from classical WLM and modified WLM plus side lobes suppression

The level of side lobes became even lower than that of the Woodward-Lawson pattern represented by the dashed gray line **2** in Fig. 4-7, though the starting ascent slightly deteriorated. A reader can use the mentioned "modified WLM" program from Appendix D to see the influence of μ on the solution and find a compromise between the side-lobe level and the steepness of the ascent that he prefers.

Of course, the techniques of lowering the level of side lobes to a prescribed value—see Section 2.5—are applicable here. However, the purpose of the book is mainly to discuss the various approaches to the problem of synthesis, rather than to demonstrate the best achievable results. The interested reader can combine the programs "modified WLM" and "Reducing the Level of a Set of Side Lobes" to restrict, for instance, three side lobes to the level –25 dB or five of them to –30 dB.

4.4. Antenna array synthesis with attenuation of a near-field in the given area

In some cases, the structure of the antenna near-field must meet additional requirements. For example, to facilitate electromagnetic compatibility, it is important to minimize the field intensity in the area S provided for the protected device. Similarly, in situations where the physical body may be present or absent near the antenna, in order to make ARP fixed, it is important that the near-field in the corresponding area S be as weak as possible.

All antenna synthesis techniques presented in Chapter 2 will be applicable to these situations if they use the extended ARP concept, which includes the near-field of the antenna. It would be natural to call the combination of ARP and near-field in the area S as "generalized ARP." Let us use the following notation $\mathbf{F}(\mathbf{r}°, \mathbf{r}) = \{\mathbf{F}(\mathbf{r}°), \mu \mathbf{E}(\mathbf{r})\}$ for it. Here $\mathbf{r}$ denotes a radius vector of the points in S (Fig. 2-1), $\mathbf{E}(\mathbf{r})$ is the antenna near-field, and μ is a scaling factor, the role of which will be clear—from below. Now the inner product is

$$\left(\mathbf{F}_1(\mathbf{r}°, \mathbf{r}), \mathbf{F}_2(\mathbf{r}°, \mathbf{r})\right) = \oint_\Omega \mathbf{F}_1(\mathbf{r}°)\mathbf{F}_2^*(\mathbf{r}°)\,d\Omega + \mu^2 \oint_S \mathbf{E}_1(\mathbf{r})\mathbf{E}_2^*(\mathbf{r})\,dS . \qquad (4\text{-}8)$$

For simplicity, let us confine ourselves to the case of an antenna array. We denote the generalized individual pattern of the n-th element as $\mathbf{f}_n(\mathbf{r}°, \mathbf{r}) = \{\mathbf{f}_n(\mathbf{r}°), \mu\, \mathbf{e}_n(\mathbf{r})\}$ and the excitation complex amplitude as I_n. Then ARP and near-field are $\mathbf{F}(\mathbf{r}°) = \sum_n I_n \mathbf{f}_n(\mathbf{r}°)$ and $\mathbf{E}(\mathbf{r}) = \sum_n I_n \mathbf{e}_n(\mathbf{r})$, respectively. With regard to Ex. 4-8 the square

deviation of generalized ARP $\mathbf{F}(\mathbf{r}^\circ, \mathbf{r})$ from a desired one $\mathbf{F}_0(\mathbf{r}^\circ, \mathbf{r})$ is as follows

$$\left\| \mathbf{F}(\mathbf{r}^\circ, \mathbf{r}) - \mathbf{F}_0(\mathbf{r}^\circ, \mathbf{r}) \right\|^2 = \oint_\Omega \left| \mathbf{F}(\mathbf{r}^\circ) - \mathbf{F}_0(\mathbf{r}^\circ) \right|^2 d\Omega +$$
$$\mu^2 \oint_S \left| \mathbf{E}(\mathbf{r}) - \mathbf{E}_0(\mathbf{r}) \right|^2 dS \qquad (4\text{-}9)$$

In the above-mentioned case, ideally, there should be $\mathbf{E}_0(\mathbf{r}) = \mathbf{0}$ (no field in S). Then the right-hand side of Ex. 4-9 turns to be $\sigma^2 + \mu^2\,\mathbb{C}$, where σ^2 and $\mathbb{C}$ denote, respectively, square deviation ARPs and near-field energy in S.

It is very clear that the scaling factor μ plays two roles. First, its dimension balances the dimensions of the addends—this may be a formal nothing but is important in meaning. Second, it sets the proportion between the requirements of the accuracy of the ARP and attenuation of the near-field: the larger the value of μ, the stricter limitation to the field intensity.

Finding out whether it is possible to eliminate the influence of a physical body—metallic as a rule—located in front of the antenna aperture, we obtained very interesting, even intriguing results. Definitely, the above approach is suitable for this study, because if the field in S is absent, then the current in the body brought in here does not appear, and the body does not affect the ARP. It turns out that the antenna phase pattern again plays a crucial role in this.

For simplicity, we have considered a 2D situation: a linear array of N sources, which are parallel infinite electric current filaments. The desired radiation pattern was a sector function from 30° to 150° for the angle θ measured from the x-axis of the array. The area S was a circle of radius λ centered on the sector symmetry line at a distance of 4 λ from the origin. Therefore, it certainly eclipsed the aperture of the antenna array.

To balance the summands in objective function Ex. 4-8, we used the value $\mu = 1/\,\mathbb{C}_1$ where $\mathbb{C}_1$ is the value of the energy in S when a single source located at the origin has excitation coefficient $W_1 = 1$. We synthesized the antenna array in two ways.

<u>Strategy **A**</u> is the one described in Section 3.4. The required ARP $F_0(\theta)$ is a real function, which means there is a desire to have a phase center at the origin. The number of sources N increases consistently from 1 to 25, and at each step, the adjoint operator V determines the place for the current successive source. Fig 4-8 demonstrates the corresponding results.

<u>Strategy **B**</u> is similar to the previous one with regard to the consistent increasing of sources and the use of the operator V for placing a sequential source. However, it differs in that at each stage it optimizes the phase diagram, as described in Section 4.2. Fig. 4-9 shows the results obtained here.

In all the charts below, the shades of black—the darker the more intense—present relief of the antenna near-field in the region ($-7\lambda < x < 12\lambda$, $0 < y < 13\lambda$). The spacing of grid lines is λ. The dots under the x-axis represent the positions of the sources. Under each chart, there are δ^2- and $\in$-values characterizing the quality of synthesized array of N sources: δ^2 is the square deviation of the amplitude pattern $A_N(\theta)$ at the current step N from the required $A_0(\theta)$; $\in$ is the normalized value in decibels relative to the value $\in 1$ of a single source at the origin.

(a) $N = 5$, $\delta^2 = 0.188$, $\in = -2.26$ dB   (b) $N = 10$, $\delta^2 = 0.155$, $\in = -3.14$ dB

(c) $N = 15$, $\delta^2 = 0.143$, $\in = -3.27$ dB   (d) $N = 25$, $\delta^2 = 0.135$, $\in = -3.33$ dB
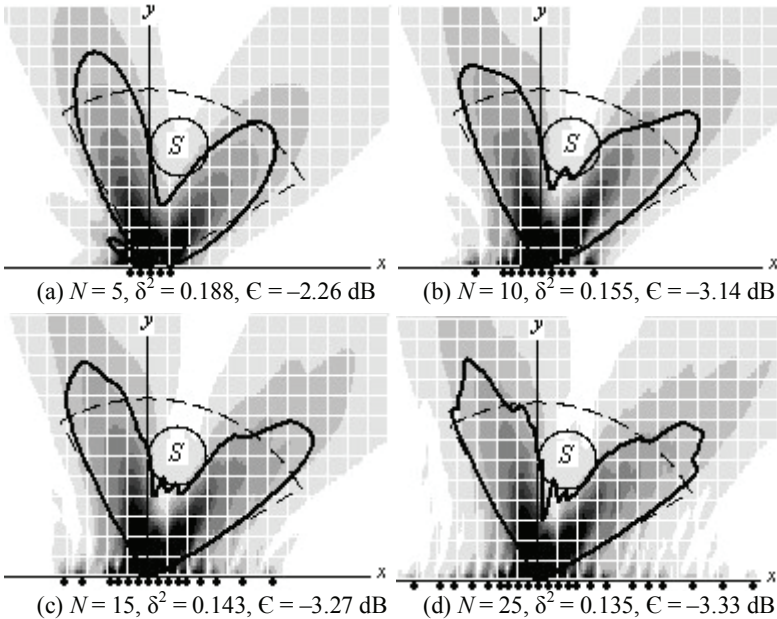
Figure 4-8: Linear array of N sources synthesized to form a sector ARP with phase center at the origin and limited field energy in the area S (**Strategy A**)

Data presented in Fig. 4-8 clearly show that the suppression of the near-field in the area S leads to a deep decreasing ARP in the directions down-shadowed by that area. This seems quite understandable from common sense. Moreover, these data indicate that this negative effect persists almost independently of the number of N sources.

The situation becomes significantly different if the antenna phase pattern is not predetermined, but changes to optimize the objective function Ex. 4-9. Fig. 4-9 shows the results of the corresponding calculations. They confirm

the encouraging fact that, in principle, it is possible to achieve at the same time both a good approximation to the desired ARP and a very weak field in the area S. Obviously, the number N of sources—the more, the better—plays a crucial role in achieving this.

From the standpoint of physics, the situation may be as follows. The left-hand side of the antenna array radiates mainly to the left of the area S, the right-hand side—mainly to the right of the area S. In other words, the phase pattern corresponds to the local phase center, which moves along a complex trajectory along—or, most likely, around—the antenna array. Such a phase structure causes leakage of the field to the shadow region of the area S.

(a) $N = 5$, $\delta^2 = 0.143$, $\mathcal{E} = -5.48$ dB   (b) $N = 10$, $\delta^2 = 0.07$, $\mathcal{E} = -10.3$ dB

(c) $N = 15$, $\delta^2 = 0.034$, $\mathcal{E} = -12.3$ dB   (d) $N = 25$, $\delta^2 = 0.013$, $\mathcal{E} = -16.9$ dB
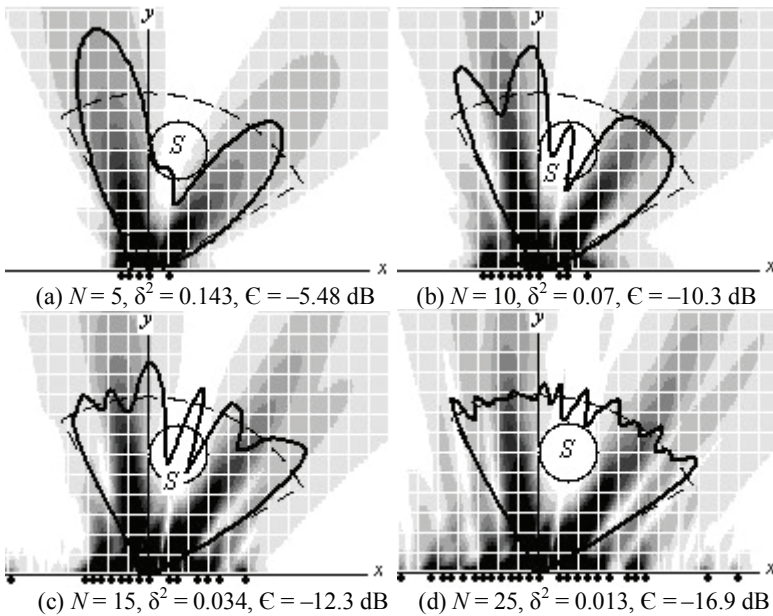
Figure 4-9: Linear array of N elements synthesized to form a sector amplitude pattern with limited field energy in the area S (**Strategy B**)

Fig. 4-10 shows in detail the dependences $\delta^2(N)$ and $\mathcal{E}(N)$ for both methods, and two values of the scale factor μ: $\mu_1 = 1/\mathcal{E}_1$ and $\mu_2 = 3.3\mu_1$.
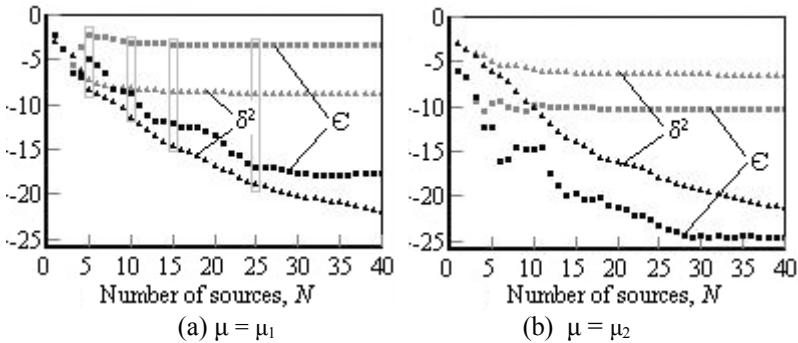
Figure 4-10: Dependences $\delta^2(N)$ and $\mathfrak{C}(N)$ corresponding to: **Strategy A** – gray color and **Strategy B** – black color

The light gray boxes in Fig. 4-10(a) highlight a set of situations that coincide with those depicted in Fig. 4-8. In general, the dependences in Fig. 4-10 are trivial: an increase in the coefficient μ leads to a significant decrease in the value of $\mathfrak{C}$ accompanied by some deterioration of the approach to the desired ARP. However, in the case of Strategy **A**—with fixed phase pattern—the deterioration is very serious, while for Strategy **B** it is much less so. Another striking difference between the strategies is that increasing the number of sources within Strategy **A** has a very limited effect, while for Strategy **B** this is a powerful factor.

Let us think about why the optimization of the phase pattern gives such a wonderful effect of improving ARP and weakening the field in S at the same time. If we were psychologists, our emotional and reasonable answer could be this: "The antenna likes a gentle developer who listens and takes into account its response to his requests, and is ready to please him."

It seems that changes in the phase structure make the near-field flow around the obstacle *S*, and this is to some extent true: as noted above, the field fills the shaded area. However, this happens rather gradually than abruptly.

4.5. Local phase center of antenna far-field and its 3D hodograph

4.5.1. Introductory notes

Any antenna radiates a spherical wave, so each component of its far field has the following structure: $[A(\theta,\varphi)\exp(j\psi(\theta,\varphi))]\exp(-jkr)\,/\,r$, where

$A(\theta,\varphi)$ and $\psi(\theta,\varphi)$ are amplitude and phase patterns, respectively, and the last .factor describes spherical divergence of the field.

A phase center (PhC) of an antenna is such a point that, if it is taken as an origin, the phase of a given field component over the surface of the radiation sphere is constant: $\psi(\theta,\varphi) = \psi_0$. Strictly speaking, a real antenna hardly has a PhC regarding a solid angle of 4π sr. Therefore, for practice, it is vital to define the concept of a local phase center (LPhC), which is the center of a sphere reproducing a rather small piece of the phase front (PhF) of a radiation field near the remote observation point.

The location of LPhC depends on the observation direction. As a result, the LPhC draws a curve in a two-dimensional situation or a surface in a general situation. Often their configuration is unpredictable. For brevity, we will call them the hodograph 2D or 3D, respectively.

Since the LPhC hodograph is completely dependent on the phase pattern $\psi(\theta, \varphi)$, this is actually just another way of graphical representation of the function $\psi(\theta, \varphi)$. However, this way is more meaningful in the applications[30], when the location and biases of LPhC are extremely important functional characteristics of the antenna.

Unfortunately, not many works are devoted to methodological issues of LPhC definition and analysis of its hodographs.

4.5.2. Calculus techniques

Let the function $\psi(\theta, \varphi)$ be the phase pattern of an antenna. In the far field region, the equiphase surface $r(\theta, \varphi)$, i.e. PhF, accords to an obvious formula

$$r(\theta, \varphi) = R_0 + \psi(\theta, \varphi)/k. \tag{4-10}$$

A piece of sphere, the center of which (x_c, y_c, z_c) is LPhC corresponding to the direction (θ, φ), should approximate a small area dS of the surface $r(\theta, \varphi)$ at a given point $r(\theta, \varphi)$. Fig. 4-11 visualizes the problem.

---

[30] The feeds for lens or parabolic reflectors must have a PhC at the focus point. PhC becomes of great interest in relation to satellite positioning systems, the high accuracy of which is due to phase measurements at the carrier frequency.
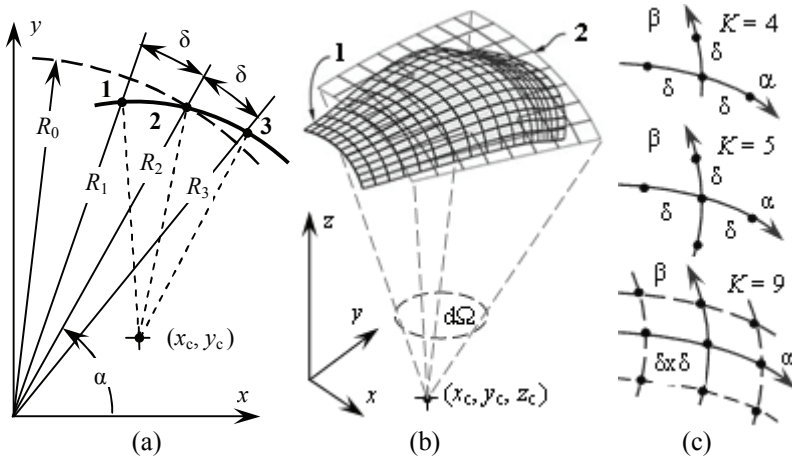
Figure 4-11: Calculating LPhC location (x_c, y_c, z_c): (a) 2D situation; (b) 3D situation, **1** - a piece of the PhF surface, **2** - approximating sphere; (c) three typical sets of four, five and nine reference points

Clearly, there is some degree of uncertainty in the very concept of LPhC. First, in general, at the observation point $r(\theta, \varphi)$, the principal curvatures of the phase front can differ. Second, the result of approximation depends on the techniques used. Third, the distance R_0 and the angle span $d\Omega$ of the area dS—the matters of choice—to some extent affect the result of the calculation.

Using inverse trigonometric functions to calculate the phase pattern $\psi(\theta, \varphi)$ results in a saw-tooth shape of the phase pattern and spurious jumps of the LPC. If available, phase unwrapping functions[31] cope with the problem. Otherwise, standard inverse trigonometric functions supplemented by comparing the new value with the previous one and changing—if necessary—the memorized number of additional 2π-periods will do.

2D Hodograph.

If it is clear that ARP has a principal plane, say the *xy*-plane, in which LPhC hodograph lies, hen 2D situation is the case (Fig. 4-11a). Let define three points **1**, **2**, **3** on a PhF line, which correspond to the set of angles $\{\alpha + \delta, \alpha, \alpha - \delta\}$ with a relatively small angular spacing δ. Their location points and distances from the origin are (x_1, y_1), (x_2, y_2), (x_3, y_3) and $R_1, R_2,$

---

[31] Matlab and Octave have such a function called **unwrap()**.

R_3, respectively. The LPhC coordinates (x_c, y_c) must satisfy two obvious equations:

$$\begin{cases} (x_2 - x_c)^2 + (y_2 - y_c)^2 = (x_1 - x_c)^2 + (y_1 - y_c)^2 \\ (x_3 - x_c)^2 + (y_3 - y_c)^2 = (x_1 - x_c)^2 + (y_1 - y_c)^2 \end{cases}, \quad (4\text{-}11)$$

the solutions of which are

$$\begin{aligned} x_c &= (d_2(y_3 - y_1) - d_3(y_2 - y_1))/D \\ y_c &= (d_3(x_2 - x_1) - d_2(x_3 - x_1))/D \end{aligned}. \quad (4\text{-}12)$$

Here, the following denotes $d_2 = (R_2^2 - R_1^2)/2$, $d_3 = (R_3^2 - R_1^2)/2$, and $D = (x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)$ are in use.

When the angle α varies, the coordinates (x_c, y_c) of the LPhC change, drawing a curve $\{x_c(\alpha), y_c(\alpha)\}$ that is a 2D hodograph of LPhC.

3D Hodograph

Let there are K points on the PhF surface within a rather small solid angle $d\Omega$ (Fig. 4-11b). That is $\mathbf{R}_i = R_i\, \mathbf{r}_i$, where $\mathbf{r}_i$ is the radial unit vector pointed to the ith direction (θ_i, φ_i), and $R_i = R_0 + \psi(\theta_i, \varphi_i)/k$ is the distance to the ith point of the PhF, given by Ex. 4-10; $i = 1...K$ indexes the points. Let $\mathbf{d}_s = x_s\, \mathbf{x}^o + y_s\, \mathbf{y}^o + z_s\, \mathbf{z}^o$ be the radius vector to the center of the approximating spherical surface of radius R_s. Taking into account the remoteness of the PhF reference points, we can use the following expression for the distance $\rho_i = |\mathbf{R}_i - \mathbf{d}_s|$: $\rho_i = R_i - [(x_c \cos\varphi_i + y_c \sin\varphi_i)\sin\theta_i + z_c \cos\theta_i]$. The quadratic approximation error σ^2 is a function of unknown variables: $\sigma^2(x_c, y_c, z_c, R_s) = \sum_i^K (R_s - \rho_i)^2$. Equating its derivatives to zero results in a system of four linear algebraic equations written below in a matrix form

$$\begin{vmatrix} \sum_i cs_i^2 & \sum_i cs_i ss_i & \sum_i cs_i c_i & \sum_i cs_i \\ \sum_i cs_i ss_i & \sum_i ss_i^2 & \sum_i ss_i c_i & \sum_i ss_i \\ \sum_i cs_i c_i & \sum_i ss_i c_i & \sum_i c_i^2 & \sum_i c_i \\ \sum_i cs_i & \sum_i ss_i & \sum_i c_i & K \end{vmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ R_s \end{bmatrix} = \begin{bmatrix} \sum_i R_i cs_i \\ \sum_i R_i ss_i \\ \sum_i R_i c_i \\ \sum_i R_i \end{bmatrix}, \quad (4\text{-}13)$$

where $cs_i = \cos\varphi_i \sin\theta_i$, $ss_i = \sin\varphi_i \sin\theta_i$, $c_i = \cos\theta_i$, and $R_i = R_0 + \psi(\theta_i, \varphi_i)/k$.

Clearly, the number K of points cannot be less than four, and the angular spacing δ between them must be small enough, but not necessarily equal. Intuition and expediency restrict potential arbitrariness. In particular, δ must be of tenths of the angular resolution of the antenna. Besides, as the

reference points are located within a small solid angle dΩ, there is no reason to use a large number of them. Fig. 4-11(c) shows three most practical sets of the reference points $K = \{4, 5, 9\}$. In our calculations, the spacing δ of both angles was the same and varied from 0.5° to 5°.

4.5.3. Phase patterns and 2D hodographs related to Section 4.4

Perhaps it does not matter much, but it is interesting to see how the LPhC hodographs change when a specific phase pattern arises in the course of its optimization towards the amplitude ARP. In [34], there are some examples regarding the synthesis of linear arrays of five or fifteen sources reproducing, respectively, sector or cosecant amplitude patterns.

Let us look at the features of phase patterns inherent in solutions for constrained optimization that are the matter of Section 4.4. Fig. 4-12 replicates the ARP from Fig. 4-8(c) and Fig. 4-9(c) in Cartesian coordinates and additionally shows phase patterns. It is quite expected that the phase pattern $\psi(\theta)$ in Fig. 4-12(a) deviates little from zero within the interval from 30° to 120°, whereas in Fig. 4-12(b) the same is changing much more. Due to this difference, the respective amplitude pattern is much closer to the desired sector.

If, as it was in Section 4.4, the antenna array consists of parallel infinite filaments of electric current, then a two-dimensional situation takes place, and LPhC certainly moves in the *x-y* cross-section drawing 2D hodograph $(x_c(\theta), y_c(\theta))$. Fig. 4-13 shows hodographs that relate to the phase patterns represented in Fig. 4-12.

The hodograph graphical image is a set of small dots corresponding to the angles $\{\theta_i\}$ within the mentioned interval with a spacing of 1°. Between the dots, there are small tip arrows indicated directions of LPhC movement. The grid of gray lines with λ-spacing allows us to estimate the location of hodograph points. Small squares in the *x*-axis point out the sources' positions. On some of the hodographs there are labels with values of corresponding angles $\{\theta_i\}$.
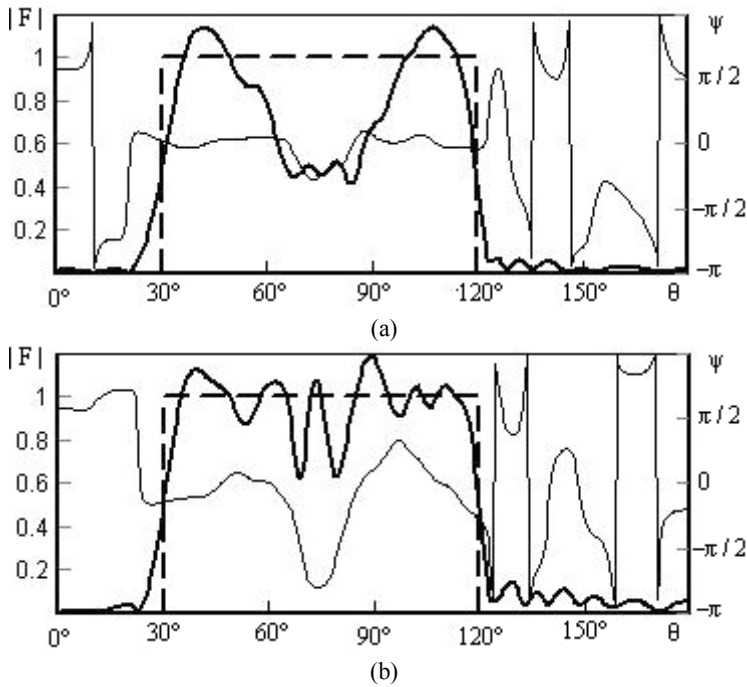
Figure 4-12: Amplitude and phase patterns related to the linear antenna array of fifteen sources ($N = 15$) synthesized for producing a sector ARP: (a) matches accords **Strategy A** with phase center at the origin; (b) accords **Strategy B** with optimization of the phase pattern. Bold solid and dotted lines depict amplitude patterns $|F(\theta)|$ and $|F_0(\theta)|$, respectively; thin line relating to the right axis depicts phase patterns

(a)   (b)
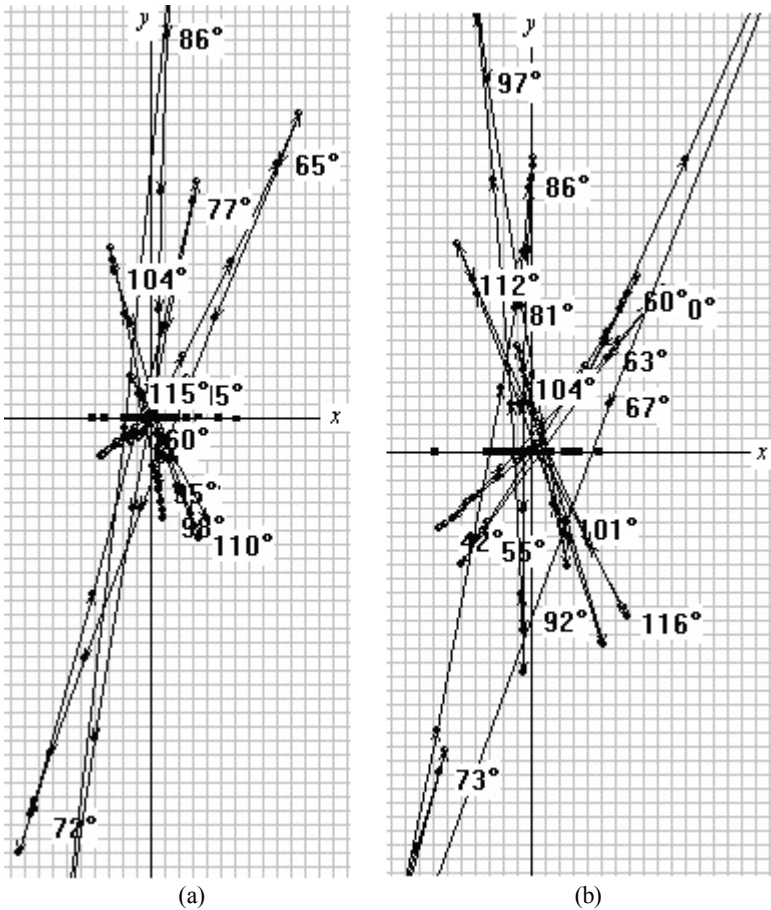
Figure 4-13: 2D hodographs of the phase patterns $\psi(\theta)$ shown in Fig. 4-12

To clearly understand the interrelation between the phase pattern $\psi(\theta)$ and its hodograph, it is important to keep in mind that the first and second derivatives of the function $\psi(\theta)$ result in shifting LPhC transversally and along the θ-line, respectively. We see that in the case of **Strategy B** the movements of LPhC are more intense everywhere and in particular at interval $90° < \theta < 120°$.

4.5.4. Competition: 3D hodograph versus 2D hodograph

To render a final verdict on which of the two hodographs is better, let us consider a concrete example. Suppose, N elements lie uniformly on a circle of radius $a = \lambda / 2$ in the *x-y* plane. Their angular coordinates, measured from the *x*-axis, are $\alpha_n = 2\,\pi\,(n-1)/N$. Suppose that the individual patterns are axially symmetric with respect to the direction α_n, and have the shape of rotational cardioids $f_n(\theta,\ \varphi) = 1 + \sin(\theta)\,\cos(\varphi - \alpha_n)$. The excitation distribution U_n is in-phase and uniform. Thus, the array pattern is

$$F(\theta,\varphi)\ = \sum_{n=1}^{N}[1 + \sin\theta\,\cos(\varphi - \alpha_n)]\exp(j\,k\,a\sin\theta\,\cos(\varphi - \alpha_n))\,. \quad (4\text{-}14)$$

It is clear that Ex. 4-14 defines ARP that is symmetric with respect to the *x-y* plane, and LPhC lies in this plane—$\theta = \pi/2$—for any direction φ in it. Therefore, it is tempting to determine the 2D hodograph defining LPhC coordinates $\{x_c(\varphi_i), y_c(\varphi_i)\}$ by formulas Ex. 4-12 with three points $\{\varphi_i - \delta,\ \varphi_i,\ \varphi_i + \delta\}$. Moreover, it is interesting and easy to find out that if, in addition to the points mentioned, the fourth one $(\varphi_i, \theta_i = \pi/2 - \delta)$ or $(\varphi_i, \theta_i = \pi/2 + \delta)$ is in use, then Ex. 4-13 defines a 3D hodograph corresponding to the possible displacement of LPhC from the *x-y* plane, while its x_c and y_c coordinates remain unchanged.

When debugging the program, we varied the distance R_0 and the angular step δ in the ranges of $10^2\,\lambda \div 10^4\,\lambda$ and $0.5° \div 5°$, respectively. The constancy of the obtained results confirmed the reliability of both algorithms leading to Ex. 4-12 and Ex. 4-13. The final calculations were carried out at $R_0 = 10^3\,\lambda$ and $\delta = 2.5°$.

Fig. 4-14 shows 2D hodographs corresponding to the phase patterns in the *x-y* plane of the above circular array. The gray lines form a coordinate grid with a mesh density of $\lambda/8$.
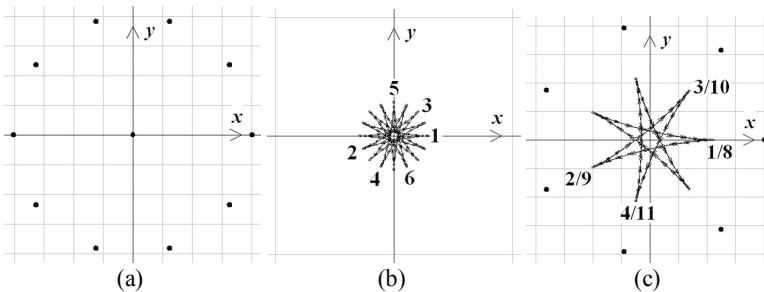

Figure 4-14: 2D hodographs of a circular array of N elements: (a) $N = 10$; (b) $N = 8$; (c) $N = 7$

The dependence of 2D hodographs on the number of array elements N is completely predictable. The number $N \geq 10$ is large enough for the phase pattern to become flat $\psi(\varphi) \approx$ const, and the center of curvature identified with LPhC to lay at the origin. As N decreases, the fluctuations in the phase pattern $\psi(\varphi)$ increase, and, therefore, the alternating oscillations of the second derivative $\psi''(\varphi)$ increase. Due to that, the radius of curvature decreases for $\psi''(\varphi) > 0$ and increases for $\psi''(\varphi) < 0$, and LPhC moves along a star-like line. Each hodograph has several numbered teeth, which begin with "**1**" for the angle $\varphi = 0°$, and continue for increasing angles φ. In the case $N = 8$, the "asterisk" of the hodograph is so small that the shown image region is only $\lambda/8 \times \lambda/8$. It is clear that the number of 2D hodographs' teeth is 2 N, since the function $\psi''(\varphi)$ is N-periodic. At first glance, in the case of $N = 7$, this rule is incorrect for an uncertain reason. However, there are actually fourteen overlapping teeth, and in Fig. 4-14(c), the slashes separate their double numbers.

Fig. 4-15 shows the isometric projections of the 3D hodographs for the same situations as above. They consist of "ribs" ($\varphi =$ const) and "hoops" ($\theta =$ const) in 5° increments for both angles.
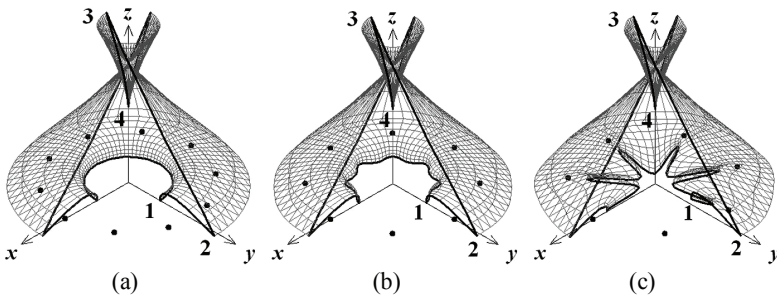
Figure 4-15: 3D hodographs of a circular array of N elements: (a) $N = 10$; (b) $N = 8$; (c) $N = 7$

The results obtained from Ex. 4-13 with five and nine reference points—see Fig. 4.11c, $K = 5$ and $K = 9$—are too close, so the case $K = 5$ is not presented for brevity. Taking into account the symmetry of the patterns with respect to the x-y plane, we show a 3D hodograph in the half-space $z > 0$. The bold dots represent the antenna element locations in isometric projections. Remembering that the radius of the antenna array is equal to $\lambda/2$, one can judge the extents of the hodograph surfaces.

For the sake of shape observation, in Fig. 4-15 there are 3D hodograph fractions corresponding to the solid angle of the observation directions

$90° \leq \theta < 180°$ and $90° \leq \varphi < 360°$. Therefore, the "starting" point "**1**" ($\theta = 90°$, $\varphi = 90°$) lies on the *y*-axis.

As the number of elements $N = 10$ is sufficiently large for a radius of $\lambda/2$, the considered array is equal to a continuous in-phase excited circle. Bearing in mind that the array factor of the circular antenna of isotropic elements is a real Bessel function $J_0(ka \sin\theta)$, which means that the antenna has an exact PhC located at the origin, someone may suspect (like the author suspected) that the represented 3D hodograph for the case of $K = 10$ is not true. However, given the cardioid patterns, another expression appears $F(\theta,\varphi) = \int_0^{2\pi}(1 + \sin\theta\cos(\varphi - \alpha))\,e^{jka\sin\theta\cos(\varphi - \alpha)}\,d\alpha$ resulting in the phase pattern $\psi(\theta)|_{\varphi = 90°} = \text{arctg}\left(\dfrac{\sin\theta\,J_1(ka\sin\theta)}{J_0(ka\sin\theta)}\right)$. In accordance with Ex. 4-13, it generates the "rib" shown in bold in Fig. 4-15(a).

By the way, the location of the point "**1**" has a clear physical sense. The frontal part of the circle, the elements of which have the patterns oriented in the positive direction of the *y*-axis, makes the main contribution to the field at the direction ($\theta = 90°$, $\varphi = 90°$). This is why LPhC shifts from the origin to a large part of the radius. Due to the axial symmetry of the pattern, the "rib" moves in a circle as φ changes.

It is clear that in the region near the *x-y* plane ($\theta \sim 90°$), when the number N of elements decreases, the fluctuations increase both in the patterns and, consequently, in the hodographs. Moreover, the LPhC motions caused by changing φ at $\theta = 90°$ do not repeat the corresponding 2D hodograph. In the polar region ($\theta \approx 180°$), the hodographs shown demonstrate the shape stability. This is due to the fact that the pattern of a circle array defined by Ex. 4-14 contains the factor "$a \sin\theta$", which, as it were, decreases the radius a and, consequently, a smaller number of N elements makes the array to be equal to a continuous circle.

4.6. Communication system's adaptive antenna arrays with the phase pattern steering circuit

4.6.1. Introduction

An adaptive antenna array (AAA) is an effective tool to increase the resistance of radio telecommunication systems to interference. In receive mode, it automatically reduces the ARP magnitude in the direction of interference sources to suppress their signals. AAA consists of antenna array (AA) of several antenna elements whose weighted signals form the

output signal. Fig. 4-16 shows the relevant block diagram where the control unit, commonly called a processor, produces the current signals to adjust the weighting coefficients $\{W_n\}$ that compose a weighting vector (WV) $\mathbf{W}$. It implements the appropriate control algorithm, usually using digital signal processing.
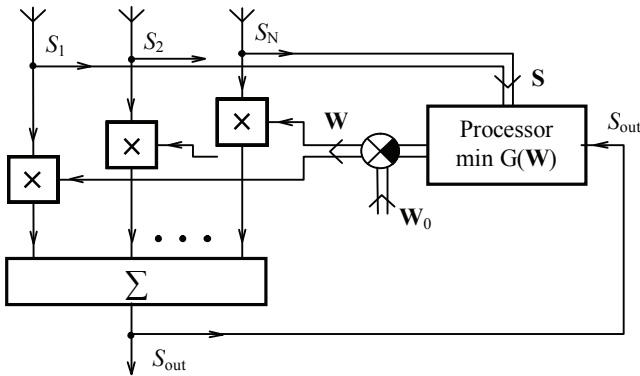
Figure 4-16: Block diagram of an Applebaum adaptive antenna array

The basic properties of AAA depend primarily on this algorithm. Let us try to determine the structure of the processor in the same way as to synthesize the antenna. This means that the starting point is the objective function. It is crucial that the objective function $G(\mathbf{W})$ minimized by the algorithm—the optimization criteria, in other words—would be adequate to the requirements for the functioning of the system. What are the requirements for any communication system?

Obviously, in the absence of interference, the processor must form such a WV $\mathbf{W}_0$, which yields the optimal ARP $F_0(\mathbf{r}°)$, covering the service area. When interference occurs, it is due to change WV in order to form ARP with dips in the directions of interference sources and, at the same time, to preserve the service area as much as possible.

Option 1. It turns out to be [35] that the classical Howells-Applebaum scheme [36] minimizes the objective function $G_1(\mathbf{W}) = P_{out} + \mu \, |\mathbf{W} - \mathbf{W}_0|^2$, where P_{out} is the output power of interference[32]. The true sense of this function is that decreases of P_{out} take place under restrictions for the

---

[32] For simplicity, we neglect receiver noise, since it is a minor factor in relation to strong interference.

deviation of WV $\mathbf{W}$ from the original $\mathbf{W}_0$. It is evident that the coefficient μ changes the ratio between the achieved decrease in P_{out} and the value of the deviation $\|\mathbf{W} - \mathbf{W}_0\|^2$, which assesses the deterioration in coverage of the required area, albeit indirectly.

The very practical AAA type implements a real-time gradient descent procedure, corresponding to a simple and comprehensive idea: to send a signal proportional to the gradient component at the integrator input. With tips given in the Task 4G, one can easily obtain the following expression for the gradient of the objective function $G_1(\mathbf{W})$

$$\Delta G_1(\mathbf{W}) = <\mathbf{R}>\mathbf{W} + \mu \,|\,\mathbf{W} - \mathbf{W}_0\,|^2 . \qquad (4\text{-}15)$$

Here, $<\mathbf{R}>$ is a covariance matrix of interference signals $\{S_n\}$ at the AA outputs. The first term in Ex. 4-15, which we will denote as the vector $\mathbf{g}$, is in fact a vector of covariance of the output signal S_{out} with signals $\{S_n\}$, i.e. $\mathbf{g} = <\mathbf{R}>\mathbf{W} = \{\overline{S_{out}S_n^*}\}$. The upper line means averaging over an interval greater than the interference correlation time, which is the reciprocal of the receiver bandwidth. Fig. 4-17 shows a block diagram of a processor that adjusts WV according to the gradient descent procedure, which realizes the rule $d\mathbf{W}/dt \sim -\Delta G_1(\mathbf{W})$. To understand its operation, it is useful to take into account that a low-pass filter (LPF) with a large time constant is almost equivalent to an integrator.
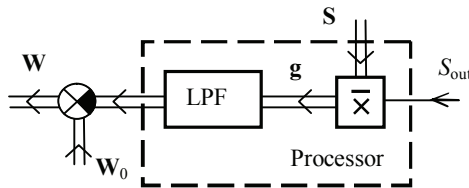
Figure 4-17: Block diagram of a processor minimizing objective function $G_1(\mathbf{W})$

Option 2. Thinking about improving AAA functionality, we can easily guess that it is better to control directly the deviation of the current ARP $F(\mathbf{r}^\circ) = \sum W_n f_n(\mathbf{r}^\circ)$ from the initial optimal ARP $F_0(\mathbf{r}^\circ)$ using the objective function $G_2(\mathbf{W}) = P_{out} + \mu \, \| F(\mathbf{r}^\circ) - F_0(\mathbf{r}^\circ)\|^2$. This leads to the following expression for the gradient of the function $G_2(\mathbf{W})$:

$$\Delta G_2(\mathbf{W}) = <\mathbf{R}>\mathbf{W} + \mu <\mathbf{Z}>(\mathbf{W} - \mathbf{W}_0) . \qquad (4\text{-}16)$$

Here $<\mathbf{Z}>$ is a square matrix of the so-called mutual resistances of AA elements $Z_{nk} = \left(f_n(\mathbf{r}^\circ), f_k(\mathbf{r}^\circ)\right) = \oint_\Omega f_n(\mathbf{r}^\circ) f_k^*(\mathbf{r}^\circ)\, d\Omega$.

The similarity of the gradients in Ex. 4-15 and Ex. 4-16 results in a likeness of the relevant block diagram of the AAA processors that realize the gradient descent algorithm. Fig. 4-18 relates to the objective function $G_2(\mathbf{W})$.
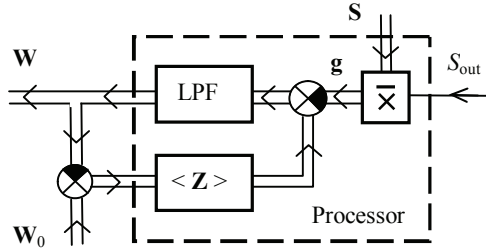
Figure 4-18: Block diagram of a processor minimizing objective function $G_2(\mathbf{W})$

Option 3. Clearly, coverage of the service area depends on the current amplitude pattern $|F(\mathbf{r}°)|$ only. The phase pattern of ARP $F(\mathbf{r}°)$ does not matter. Consequently, a constraint that is more adequate to the communication system could be $\| A_0(\mathbf{r}°) - |F(\mathbf{r}°)| \|^2$ that results in the following objective function

$$\Delta G_3(\mathbf{W}) = <\mathbf{R}> \mathbf{W} + \mu \| A_0(\mathbf{r}°) - | F(\mathbf{r}°) | \|^2 . \qquad (4\text{-}17)$$

Here $A_0(\mathbf{r}°) = |F_0(\mathbf{r}°)|$ is the initial optimal amplitude pattern corresponding to WV $\mathbf{W}_0$. Taking into account the issue of Section 4.2, we can propose the structure shown in Fig. 4-19, which realizes in real time the above iteration method of best mean square approach to the desired amplitude pattern $A_0(\mathbf{r}°)$.
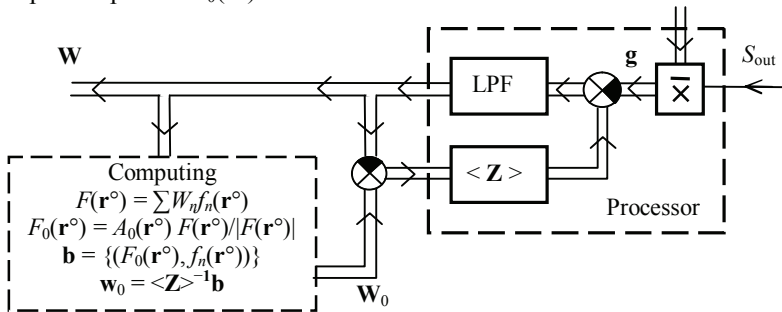

Figure 4-19: Block diagram of the processor minimizing objective function $G_3(\mathbf{W})$

Certainly, the temp, in which the computing unit renews the current desired WV $\mathbf{W}_0$, must be in less than the temp of changing WV $\mathbf{W}$ by the processor. Say, its period times must exceed the time constant of the LPF.

As the aim of this Section is to teach, how the methodology of antenna synthesis concerning optimization of a phase pattern can help in improving AAA functionality we do not demonstrate in these examples. The interested reader can find them in [35]. Although, with regard to assessing the efficiency of the transformation of the first objective function into the following two, from the general judgments, the following is clear.

First, with regard to assessing the efficiency of converting the first objective function into the next two, it is clear from general judgments that, in a real situation, the functions $G_2(\mathbf{W})$ and $G_1(\mathbf{W})$ are almost equal, since for AA with a very usual spacing around $\lambda/2$ the matrix $<\mathbf{Z}>$ is close to a diagonal one. Therefore, the norms $\|F(\mathbf{r}°) - F_0(\mathbf{r}°)\|$ and $\|\mathbf{W}(\mathbf{r}°) - \mathbf{W}_0(\mathbf{r}°)\|$ are practically proportional to each other.

Second, the objective functions $G_3(\mathbf{W})$ and $G_1(\mathbf{W})$ can be significantly different, subject to certain situations. In [35], there are some relevant examples.

4.7. Training tasks

Task 4A
Prove that the chain of inequalities Ex. 4-2 is true in the general situation. Find out the condition under which the strict inequality sign ">" needs to be replaced with "≥" or even "=."

Task 4B
Prove that any local minima of the functional $\Delta^2(\mathbf{W})$ correspond to a certain stationary phase pattern.

Task 4C
Try to construct the simplest situation, say $N = 2$, to which the number of stationary phase patterns is greater than N.

Task 4D
Using the decomposition $\exp(jx\cos\varphi) = \sum_{n=0}^{\infty} j^n(2-\delta_{0n})J_n(x)\cos(n\varphi)$, make sure that in the case of a circular antenna of isotropic sources and ARP controlled in the plane of the circle, we have: $\{g_n(\alpha) = \cos(n\alpha)\}$, $\{G_n(\varphi) = \cos(n\varphi)\}$, $\{\lambda_n = J_n(ka)\}$, where a is the radius of the circle.

Hint. Take into account the fact that if the operator U maps an orthogonal set of functions $\{u_n(\alpha)\}$ into an orthogonal set of functions $\{w_n(\alpha)\}$ then these sets are the biorthogonal bases $\{g_n(\alpha)\}$ and $\{G_n(\alpha)\}$, respectively.

Task 4E

Using the decomposition from the task 4D show that the individual pattern of a cardioid source with the polar coordinates (a, α) is

$$f(\alpha,\varphi) = \sum_{n=0}^{\infty} j^n \left[(2-\delta_{0n})J_n(ka) + j\left(J_{n-1}(ka) - J_{n+1}(ka)\right) \right] \cos(n(\varphi-\alpha)) .$$

Task 4F

Write up the formula for $10\,\lambda$ linear sources distribution $I(x)$, which ARP accords to Ex. 4-7.

Task 4G

Taking into account the obvious $P_{\text{out}} = \overline{|S_{\text{out}}|^2}$ and $S_{\text{out}} = \sum_n W_n S_n$, show that gradient $\Delta P_{\text{out}}(\mathbf{W}) = (\partial P_{\text{out}}/\partial W_n' + j\, \partial P_{\text{out}}/\partial W_n'')^{\text{T}} = <\mathbf{R}> \mathbf{W}$.

CHAPTER FIVE

USING ANTENNA SYNTHESIS TECHNIQUES FOR ANTENNA PATTERN RETRIEVAL FROM DATA TAKEN IN NON-IDEAL ENVIRONMENTS

5.1. Introductory notes

As mentioned, antenna synthesis, at its core, refers to the inverse problems of mathematics. Similar problems arise in other fields of engineering. Therefore, the techniques developed for the synthesis of antennas are of wider interest.

In acoustics, for example, one of the ways to suppress the radiation field of a membrane or a metal shell excited by external sources is to use an active vibration compensation system. The system consists of several special vibrators, a number of sensors, and a control circuit that generates the excitation signals for the vibrators necessary to compensate for—to suppress—the external field of vibration. For several years, the author has been working on this problem in cooperation with the Acoustic Institute of the USSR Academy of Sciences [37−39].

The most dangerous vibrations occur at resonant frequencies. Therefore, the problem of active vibration compensation is in many ways analogous to the problem of antenna array synthesis. The signals $\{S_m\}$ from the sensors, more precisely, their current complex amplitudes, taken with reverse phases, play the role of the required ARP $F_0(\mathbf{r}_m)$. The matrix of transfer coefficients from each vibrator to each sensor—which depends on the physical properties and geometry of the body—is the matrix of the direct operator U. By the way, the adjoint operator V, used as shown above in Section 3.4, can "advise" where to place the vibrators if their location is not predetermined.

In some ways, however, the problem of compensating the field of vibration is simpler than the synthesis of an antenna. First, it does not include the task of optimizing the phase pattern of the generated vibration field, because it must be exactly the inverse of the phase pattern of the current vibration. Second, there are no problems with side lobes or local

amendments to the solution. For this reason, we have limited ourselves to mentioning this problem with the above references and devoted this chapter to much more enlightening issues related to specific ARP measurement procedures.

Modern vector network analyzers allow taking ampliphasometric measurements in automatic mode with subsequent digital processing of data. This enables us to use in practice sophisticated procedures for the retrieval of ARP from measurements taken under conditions far from ideal. For example, the illuminating field is significantly different from a plane wave due to the short distance to the radiating antenna and/or due to reflections from the walls of the measuring chamber.

The pioneer works in this field were theoretical rather than practical [40–42]. Since then, many studies on various aspects of primary measurements and data processing algorithms have been carried out, for example. The recently approved doctoral dissertations [43, 44] confirm the relevance of research in this field until nowadays. Among the measurement methods that include the subsequent reconstruction of ARP, there are three groups [43]. First, planar parallel scanning the antenna near-field and transformation of obtained data to the far field. Second, measuring the antenna field in the Fresnel zone, and transforming it to the far field. Third, a probe of the illuminating field within the boundary of the test area by a sensor with precisely known ARP, and using the obtained data to retrieve the true ARP of measurement of the antenna under test. The methodology of the CCPW proposed by us [45, 46] belongs to the third group. So as not to send English-speaking readers to these publications in Russian, the following paragraph briefly repeats the methodological background of the CCPW techniques.

In the practice of antenna measurements, compact anechoic chambers are widely used. Sometimes, a special device, the so-called collimator, serves as an irradiator that reproduces a plane wave, more precisely, a piece of a plane wave within the test area. Retrieving ARP from measurement data taken in non-ideal conditions allows the weakening of the requirements both to the absorbers coating the camera walls and to the collimator up to using a regular feeding horn. This is an additional reason for the interest in retrieving ARP.

5.2. Methodology of a converging cluster of plane waves at a retrieving antenna pattern

The possibility of using CCPW for the approximation of a real field in an area free of sources arises from two circumstances [46]. **First**, the following integral transformation gives the Green's function of a free space [47] and, therefore, an arbitrary field in any region outside the source

$$G(x,y,z) = \frac{1}{8\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\exp(\pm\sqrt{\beta_x^2 + \beta_y^2 - k^2}\, z - j\beta_x x - j\beta_y y)}{\sqrt{\beta_x^2 + \beta_y^2 - k^2}} \, d\beta_x \, d\beta_y \, . \tag{5-1}$$

From the physical point of view, Ex. 5-1 is a continuous solid (two-dimensional) cluster of plane electromagnetic waves propagating from a source located at the coordinate origin. This cluster includes both <u>uniform waves</u>—for $\beta_x^2 + \beta_x^2 > k^2$—and <u>evanescent waves</u>—for $\beta_x^2 + \beta_x^2 < k^2$. Since the inhomogeneities capable of supporting evanescent waves are quite far away from the antenna under test, therefore the contribution of those waves to the field in the antenna test area is negligible. It is clear that the illuminating field is a converging cluster of waves, since its sources are outside the test zone, and because of the above reasons, it contains only <u>uniform</u> plane waves.

Second, since ARP determines the response of the antenna to the incident uniform plane wave, the output signal of the antenna illuminated by CCPW corresponds to a convolution type integral whose core is ARP.

For simplicity's sake, let us focus on the two-dimensional scalar situation, when the ARP $F_a(\varphi)$ is measured in a plane and the field illuminating it is a CCPW in the same plane. In Fig. 5-1, a gray arrow with a transverse segment, symbolizing the phase front, depicts a separate plane wave of CCPW. A triplet of them artificially shows the cluster itself.

*Assume that the source of the illuminating cylindrical wave is located at point **D** relatively close to the antenna under test (AUT). At the first stage, it is necessary to determine the complex amplitudes $A(\varphi)$ of cluster's waves—called the spatial spectrum of the cluster—reproducing the illuminating field. For this, the probe, whose ARP $f_0(\psi)$ is precisely known, rotates along a circle of radius R_0.*
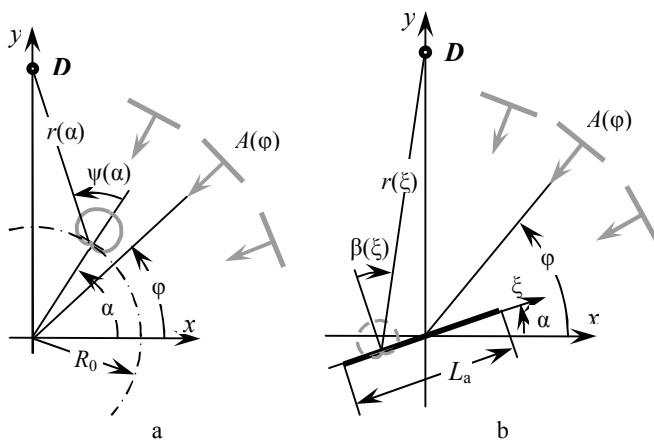
Figure 5-1: Reference geometries: (a) stage 1 − probing CCPW; (b) stage 2 − measuring with AUT

It is obvious that the signal $V(\alpha)$ at the probe output, where α is the angular position of the probe, is the sum of the signals from each plane wave. Thus, the following convolution integral describes dependency between the functions $V(\alpha)$ and $A(\varphi)$

$$V(\alpha) = \int_0^{2\pi} F_{\mathrm{pr}}(\varphi - \alpha)\, A(\varphi)\, d\varphi \,. \tag{5-2}$$

Here, $F_{\mathrm{pr}}(\varphi - \alpha) = f_0(\varphi - \alpha) \exp(jkR_0 \cos(\varphi - \alpha))$ is the ARP of the probe displaced from the origin by radius R_0. After the measurements at step 1, the function $V(\alpha)$ is known, and the solution of the equation Ex. 5-2 defines the spatial spectrum $A(\varphi)$.

It is not hard to see that Ex. 5-2 coincides with the equation for the synthesis of a circular antenna of radius R_0, the sources of which have an individual ARP $f_0(\psi)$. In that case, functions $V(\alpha)$ and $A(\varphi)$ are the desired ARP is the sought-for distribution, respectively.

At the second stage, we rotate AUT—say, the linear antenna of length L_{a} (Fig. 5-1(b)), measure the signal $U(\alpha)$ at its output, and reconstruct the ARP $F_{\mathrm{a}}(\varphi)$ by solving the following convolution equation

$$U_{\mathrm{a}}(\alpha) = \int_0^{2\pi} A(\varphi) F_{\mathrm{a}}(\varphi - \alpha)\, d\varphi \,, \tag{5-3}$$

the kernel $A(\varphi)$ of which we know from the previous stage.

It is clear that the surface S of probe rotation should cover the test area: $R_0 > L_{\mathrm{a}}/2$. Indeed, the coincidence of the tangential components of the electric field of the illuminating wave and that of the CCPW—as a

consequence of Ex. 5-2—allows us to assert[33] that in the case under consideration the CCPW field inside S coincides with the illuminating field. At the same time, as the calculations show, the field outside S has an amazing and hardly predictable structure, which has nothing to do with the real field [46]. By the way, in the same situation, if the electric and magnetic surface currents, in accordance with the Lava equivalence, produce an illumination field inside S, then they give zero fields outside S.

Likely, it would be very practical to use a spherical wave source, for example, a small horn as an illuminating antenna. In this case, strictly speaking, the uniqueness theorem forces us to go over to a closed surface S and to a continuous CCPW. Nevertheless, the calculations show that if we deal with the measurement of the antenna of a small transverse size, then the approximation of the illuminating field by a plane CCPW provides acceptable results.

In the absence of an analytic solution, it is necessary to reduce Ex. 5-2 to a system of linear algebraic equations (SLAE). The simplest way, which is hardly inferior to the Galerkin-Ritz method [48], is to replace the continuous CCPW with a discrete cluster of N plane waves with the angle spacing $\Delta = 2\pi/N$ between them, and perform measurements in quite a large number $M > N$ of points with spacing $\delta = 2\pi/M$. Then each of the equations Ex. 5-2 and Ex. 5-3 transforms into a system of M equations. If ARP of the probe is a cardioid, and the antenna pattern is sought-for as a series $F_{\mathrm{a}}(\psi) = \sum_{k=1}^{K} a_k\, g_k(\psi)$ with using K basis functions $\{g_k(\psi)\}$, then two following SLAE appear with N unknowns $\{A_n\}$ and K unknowns $\{a_k\}$, respectively

$$V(\alpha_m) = \sum_{n=0}^{N-1} A_n \left(1 + \cos(\varphi_n - \alpha_m)\right) \exp(jkR_0 \cos(\varphi_n - \alpha_m)), \qquad (5\text{-}4)$$

$$U_{\mathrm{a}}(\alpha_m) = \sum_{k=1}^{K} \left(a_k \sum_{n=0}^{N-1} A_n g_k(\varphi_n - \alpha_m) \right) \qquad (5\text{-}5)$$

where $\varphi_n = (n-1)\,\Delta$ and $\alpha_m = (m-1)\,\delta$.

5.3. Calculation results

At present, when studying any problem, computer simulation can and usually does successfully substitute the physical experiment. In our case, we need to simulate the "measurements", that is, to determine the appropriate formulas for the signals $V(\alpha_m)$ and $U(\alpha_m)$ received at position α_m.

---

[33] On the grounds of the uniqueness of the Maxwell's equations solution.

If the probe's ARP is a cardioid $f_0(\psi) = 1 + \cos\psi$, then the following expression is true

$$V(\alpha_m) = \left((1 + \cos\psi(\alpha_m)) / \sqrt{r(\alpha_m)}\right) \exp(-jk\,r(\alpha_m)),\qquad(5\text{-}6)$$

where $r(\alpha_m) = \sqrt{D^2 + R_0^{\,2} - 2DR_0\sin\alpha_m}$ is a distance from the source at D to the probe—see Fig. 5-1(a))—and $\cos(\psi(\alpha_m))$ is obviously equal to $(D\cos\alpha_m - R_0)/r(\alpha_m)$.

Let assume that AUT is a linear antenna of length L_a formed by in-phase and uniformly excited sources with cardioid individual patterns (Fig. 5-1(b)). Then the following expression defines the signal $U(\alpha_m)$ received by the rotating antenna

$$U(\alpha_m) = \int_{-L_a/2}^{L_a/2} \frac{(1 + \cos(\beta))}{\sqrt{r(\alpha_m,\xi)}} \exp(-jk\,r(\alpha_m,\xi))\ d\xi,\qquad(5\text{-}7)$$

where the integration variable ξ is a coordinate along the antenna, $r(\alpha_m,\xi) = \sqrt{\xi^2 + D^2 - 2D\xi\sin(\alpha_m)}$ is a distance from the integration point ξ to the point D; β is the angle between the normal $\mathbf{n}°$ to the antenna and the direction to the source in the point D. With some work—just write $\mathbf{n} = -\sin(\alpha_m)\mathbf{x}° + \cos(\alpha_m)\mathbf{y}°$ and $\mathbf{r}(\alpha_m,\ \xi) = -\xi\cos(\alpha_m)\,\mathbf{x}° + (D -\xi\sin(\alpha_m))\,\mathbf{y}°$ and use the dot product—it can be shown that cosine of the angle β corresponds to the following formula $\cos(\beta) = D\cos(\alpha_m)/r(\alpha_m,\ \xi)$.

The results presented below relate to the following geometry: the test area of radius $R_0 = 8\ \lambda$ locates at the distance of $35\ \lambda$ from the source D. AUT is the aforementioned linear antenna with a length of $L_a = 15\ \lambda$.

Fig. 5-2 (a) shows the spatial spectrum $\{A_n\}$, gained by solving the equation Ex. 5-4. Bold bars refer to the left axis and depict amplitudes $|A_n|$. The black dots connected by a gray line refer to the right axis and display the argument of the complex values A_n.
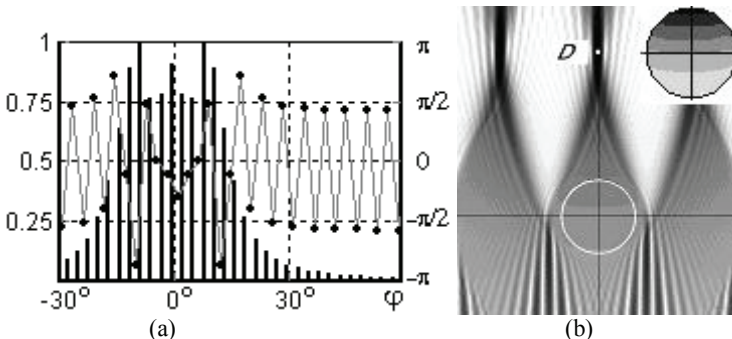

Figure 5-2: Stage 1: (a) spatial spectrum $\{A_n\}$; (b) field of CCPW

Because of the geometrical symmetry about the y-axis, the spectrum $\{A_n\}$ is an even function with respect to the angle $\varphi = 90°$. For obvious reasons, its main part is in the interval $60° < \varphi_n < 120°$. The CCPW obtained sets up the electric field $E(x, y) = \sum_n A_n \exp(k(x \cos \varphi_n + y \sin \varphi_n))$. The plot in Fig. 5-2(b) shows the relief of the amplitude of this field in shades of gray from white to black, the more, the darker. Dimensions of the plot are 60 λ along the x-axis and 70 λ along the y-axis. It is clear that the field maximum is at point $\boldsymbol{D}$, where the source is located. A white circle outlines the test area of radius R_0 = 8 λ. Although difficult, we can see that inside it the field corresponds to a cylindrical wave emanating from $\boldsymbol{D}$. In the upper right corner of the plot, there is a more detailed picture of the field in this area after normalizing to its maximal value.

Fig. 5-3 illustrates what happens in the second stage. The distance D = 35 λ is too short, and the phase distribution of the illuminating field along the x-axis is very close to a quadratic dependence with a phase shift of 286° at the ends $x = \pm 15$ λ. That is why the signal $U(\alpha)$—a gray curve— resembles the ARP of a linear antenna with a similar quadratic phase distribution, which leads to the expansion and bifurcation of the beam.
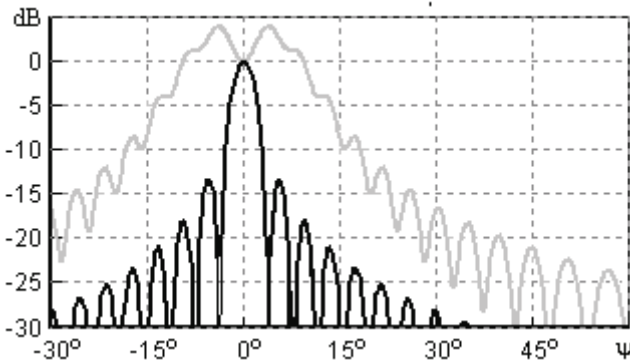
Figure 5-3: Stage 2: ―――― signal $U(\alpha)$; ―――― retrieved ARP $F_a(\psi)$

The retrieval ARP—a black curve in Fig. 5-3—coincides with the true ARP, which is $F_a(\theta) = 0.5 (1 + \cos\theta) \operatorname{sinc}(0.5 kL_a \sin\theta)$. Here θ is the angular coordinate measured from the normal to the antenna, and function $\operatorname{sinc}(x)$ is $\sin(x)/x$. Of course, several factors limit the retrieval precession. The main factors are the measurement accuracy and algorithm's parameters

such as numbers N, M, K that influent onto conditioning SLAE Ex. 5-4 and 5-5, and on methodological accuracy as well. The interested reader will find the relating issues in [46]. Antenna synthesis techniques can alleviate some of the above problems.

5.4. An improvement of retrieval techniques

Let us look at Ex. 5-4 from the point of antenna synthesis. Bearing in mind its equation $\sum_n U_n f(x_n, \psi) = F_0(\psi)$ let us use the notation $v(\varphi_n, \alpha)$ for the function $(1 + \cos (\varphi_n - \alpha)) \exp(jk\, R_0 \cos (\varphi_n - \alpha))$. It has a clear meaning of a signal induced in a rotating probe by a plane wave incident from the direction of φ_n. Then Ex. 5-4 turns to be $\sum_n A_n v(\varphi_n, \alpha) = V(\alpha)$. It differs from the antenna synthesis equation only in the notations: A_n, φ_n, and α are equal to U_n, x_n, and ψ, respectively; signals $v(\varphi_n, \alpha)$ and measured signal $V(\alpha)$ play the role of an individual pattern of antenna array's element and the required ARP, respectively. In these terms, CCPW is a specific antenna array whose elements are plane waves, and the signals at the output of the probe, depending on the angle α, substitute antenna patterns. Clearly, "the required ARP" is a complex function, the phase pattern of which is important.

Formally, we have to use a very small angular spacing $\Delta = 2\pi/N$, converting Ex. 5-2 to SLAE Ex. 5-4 by substituting the integration for the summation. However, by doing this, very soon we find that the so-called curse of dimension imposes a hard limit on the maximum value of N. Thus, in the above situation, Mathcad "refuses" to solve Ex. 5-4, if $N > 130$. On the other hand, if $N < 110$, the number of terms in series Ex. 5-4 is not sufficient to get a good approximation for the measured signal $V(\alpha)$. Thus, with reference to value N, we have to balance between a poor approximation and an ill-conditioned system of equations.

It is clear that Tikhonov's regularization can help overcome the problem of poor conditionality. However, why not try to deal with the problem from the other—less formal and typical to the engineering style of thinking—point of view. I mean the perception of Ex. 5-4 as the task of approximation of the measured function $V(\alpha)$ with CCPW, which is a set of directions $\{\varphi_n\}$ and a spatial spectrum $\{A_n\}$. As mentioned above, this task is completely analogous to the synthesis of an antenna array. Thinking this way, we immediately ask ourselves the questions: "Why should the waves be distributed evenly? In the end, it seems useful to increase the density of the waves in the direction where the real source is. Why not use the idea of sequential choosing their directions $\{\varphi_n\}$ as described in Section 3.4.1?"

Though I am sure that the idea is promising, I have not yet verified the idea by appropriate calculations and invite interested readers to do it on their own, possibly at the same time as me.

Numerical simulation of the investigated procedure for the ARP retrieval from the results of measurements taken in non-ideal conditions demonstrates the stability of the results as well as the possibility of achieving high accuracy, which requires a rational choice of the number of CCPW waves. Evidently, this number depends on the wavelength of the working area. The methodology may be of interest for the measurement setup with a collimator, allowing you to increase the size of the working area with respect to the collimator size. Evaluation of the effectiveness of this approach can be the subject of further research.

APPENDICES

A: Program "Super Directivity and Errors"

Notations:

L − length of the linear antenna array [in wavelengths!]

N − number of elements

ε − relative error of implementation of the nominal solution

d − spacing [in radians]

$\{x_n\}$ − coordinates of the antenna elements [in radians]

<A> − matrix of SLAE for synthesizing the nominal distribution

{B} − the unit vector of the right side of SLAE

{U} − the nominal distribution

F0(θ, U) − ARP according to the nominal distribution

F0_(θ) − ARP according to the in-phase uniform distribution

δ − standard deviation of the errors

F1(θ, V1) − F5(θ, V5) − ARPs of five implementations of the nominal distribution

|U| − the norm of the nominal distribution

Recommendations

1. Rewrite the program code shown on the next page in the Mathcad environment. Where necessary, use imaginary units.

2. Format the X-Y Plot with seven traces as shown below.

Legend Label	Symbol	Line	Color	Type	Weight
trace 1	none	solid	grn	lines	1
trace 2	none	solid	grn	lines	1
trace 3	none	solid	grn	lines	1
trace 4	none	solid	grn	lines	1
trace 5	none	solid	grn	lines	1
trace 6	none	solid	blk	lines	2
trace 7	none	dash	blk	lines	1

| trace 1 | none ▾ | solid ▾ | grn ▾ | lines ▾ | 1 ▾ |

3. Run the program; change the values of ε, *N*, and L. Make sure that when N exceeds a certain threshold—while L is constant—the ARP distortions increase catastrophically.

Program code

$$L := 1.5 \quad N := 8 \quad \varepsilon := 7 \times 10^{-3} \quad n := 0 .. N - 1 \quad m := 0 .. N - 1$$

$$dx := \frac{2\pi \cdot L}{N - 1} \quad x_n := dx \cdot n \quad A_{n,m} := J0[dx \cdot (n - m)] \quad B_n := 1$$

$$U := \text{lsolve}(A, B) \quad \theta := 0 .. 180 \quad Fa(\theta, U) := \left| \sum_n U_n \cdot e^{i x_n \cdot \cos(\deg \cdot \theta)} \right|$$

$$\delta := \varepsilon \cdot |U|$$

$$V1 := U + \text{rnorm}(N, 0, \delta) + i \cdot \text{rnorm}(N, 0, \delta) \qquad // \text{ V1 - V5 are}$$
$$V2 := U + \text{rnorm}(N, 0, \delta) + i \cdot \text{rnorm}(N, 0, \delta) \qquad // \text{ five ARP with}$$
$$V3 := U + \text{rnorm}(N, 0, \delta) + i \cdot \text{rnorm}(N, 0, \delta) \qquad // \text{ random errors:}$$
$$V4 := U + \text{rnorm}(N, 0, \delta) + i \cdot \text{rnorm}(N, 0, \delta) \qquad // \text{ in-phase and}$$
$$V5 := U + \text{rnorm}(N, 0, \delta) + i \cdot \text{rnorm}(N, 0, \delta) \qquad // \text{ quadrature ones}$$

// The sets {mF0} - {mF5} are the corresponding normalized ARPs

$$mF0_\theta := Fa(\theta, U) \qquad F0(\theta) := \frac{Fa(\theta, U)}{\max(mF0)}$$

$$mF1_\theta := Fa(\theta, V1) \qquad F1(\theta) := \frac{Fa(\theta, V1)}{\max(mF1)}$$

$$mF2_\theta := Fa(\theta, V2) \qquad F2(\theta) := \frac{Fa(\theta, V2)}{\max(mF2)}$$

$$mF3_\theta := Fa(\theta, V3) \qquad F3(\theta) := \frac{Fa(\theta, V3)}{\max(mF3)}$$

$$mF4_\theta := Fa(\theta, V4) \qquad F4(\theta) := \frac{Fa(\theta, V4)}{\max(mF4)}$$

$$mF5_\theta := Fa(\theta, V5) \qquad F5(\theta) := \frac{Fa(\theta, V5)}{\max(mF5)}$$

$$u(\theta) := \pi \cdot L \cdot \cos\left(\pi \frac{\theta}{180}\right) \qquad F0\_(\theta) := \left| \frac{\sin(u(\theta))}{u(\theta)} \right|$$

$|U| = 57.797$

F1(θ)
F2(θ)
F3(θ)
F4(θ)
F5(θ)
F0(θ)
F0_(θ)

θ
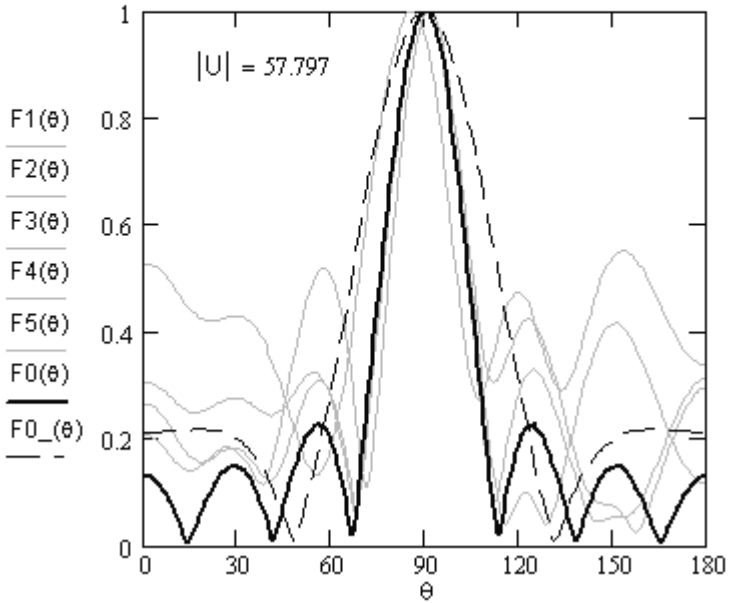
B: Program "Reducing the Level of a Set of Side Lobes of a Linear Antenna Array of Cardioid Sources"

Notations:

d − spacing [in radians!]
N − number of elements
$\psi 0$ − main beam direction [in radians]
K − the number of side lobes that are subject to decrease
β − the level of restriction
$f(n, \theta)$ − ARP of n*th* element
$\{I0_n\}$ − initial excitation distribution to produce the main beam
$F0(\theta)$ − initial normalized ARP
Ψ − set of K angles [in radians] to which restrictions apply
$beam(k, \theta)$ − a function defining k*th* normalized beam
$\{A_{k, m}\}$ − matrix of the SLAE for the restrictions
g − the right side vector of the SLAE
$F1(\theta)$ − ARP after local amendments

Recommendations

4. Rewrite the program code shown in the next page in the Mathcad environment. Where necessary, use imaginary units and complex conjugates (looks like a line under a variable or function).

5. Format the X-Y Plot with the number of grids on X-axis equal to $5*(X_{max} − X_{min})$ and tree traces as shown below.

Legend Label	Symbol	Line	Color	Type	Weight
trace 1	none	solid	grn	lines	2
trace 2	none	solid	blk	lines	2
trace 3	none	dadot	blk	lines	1

| trace 1 | none ▼ | solid ▼ | grn ▼ | lines ▼ | 2 ▼ |

6. Run the program; change the restriction value β and, if necessary, adjust the set Ψ; change the values $\psi 0$, N, K and the set Ψ. For starters, you can keep in mind the following combinations:
   N = 10, K = 6, $\psi 0$ = 0, Ψ = (-0.65 -0.45 -0.28 0.28 0.45 0.68),
   N = 15, K = 6, $\psi 0$ = 0, Ψ = (-0.42 -0.3 -0.2 0.2 0.3 0.42).

7. To more accurately install the Ψ array, watching the side lobes of the ARP, expand the graph and increase the number of grids along the X-axis.

8. In the end, be free to change all the parameters and improve the program itself, as you like.

Program code

$$N := 10 \quad d := \pi \quad \psi0 := 0.2 \quad K := 6 \quad \beta := 0.056 \quad n := 0..N$$

$$f(n,x) := (1 + \cos(x)) \cdot e^{i \cdot n \cdot d \cdot \sin(x)}$$

$$I0_n := \overline{f(n,\psi0)} \qquad F\_0(\theta) := \sum_n I0_n f(n,\theta)$$

$$F0(\theta) := \frac{F\_0(\theta)}{F\_0(\psi0)} \qquad k := 0..K-1 \quad m := 0..K-1$$

$$\Psi := \begin{pmatrix} -0.42 & -0.25 & -0.08 & 0.5 & 0.7 & 0.95 \end{pmatrix}^T$$

$$\text{beam}(k,\theta) := \frac{\left(\sum_n \overline{f(n,\Psi_k)}\, f(n,\theta) \right)}{\left| \sum_n \overline{f(n,\Psi_k)}\, f(n,\Psi_k) \right|} \qquad g_k := F0(\Psi_k) \cdot \left(\frac{\beta}{\left| F0(\Psi_k) \right|} - 1 \right)$$

$$A_{k,m} := \text{beam}(m,\Psi_k)$$

$$dI := A^{-1} g$$

$$F1(\theta) := F0(\theta) + \sum_k dI_k \cdot \text{beam}(k,\theta) \qquad dB(x) := 20 \cdot \log(|x|)$$

$$y(x) := dB(\beta) \qquad y0(x) := dB(F0(x)) \qquad y1(x) := dB(F1(x))$$

C: Program "Reducing the Level of a Set of Side Lobes of a Circular Antenna Array of Cardioid Sources"

Notations:

d − spacing [in wavelength!] between elements

N − number of elements

a=d N − radius of the circle [in radians!]

$\{\alpha_n\}$ − angle coordinates of antenna elements

$\psi 0$ − main beam direction [in radians]

K − the number of side lobes that are subject to decrease

β − the level of restriction

f(n, φ) − ARP of n*th* element

$\{I0_n\}$ − initial excitation distribution to produce the main beam

F0(φ) − initial normalized ARP

Ψ − set of K angles [in radians] to which restrictions apply

beam(k, φ) − a function defining k*th* normalized beam

$\{A_{k,m}\}$ − matrix of the SLAE for the restrictions

g − the right side vector of the SLAE

F1(φ) − ARP after local amendments

Recommendations

1. Rewrite the program code shown in the next page in the Mathcad environment. Where necessary, use imaginary units and complex conjugates (looks like a line under a variable or function).

2. Format the X-Y Plot with the number of grids on X-axis equal to fifteen and three traces as shown below.

Legend Label	Symbol	Line	Color	Type	Weight
trace 1	none	solid	grn	lines	2
trace 2	none	solid	blk	lines	2
trace 3	none	dadot	blk	lines	1

trace 1	none ▼	solid ▼	grn ▼	lines ▼	2 ▼

3. Run the program; change the restriction value β and, if necessary, adjust the set Ψ; change the values ψ0, N, K and the set Ψ. Be free to change all the parameters and improve the program itself, as you like.

4. To more accurately install the Ψ array, watching the side lobes of the ARP, expand the graph and increase the number of grids along the X-axis.

Program code

$$d := 0.45 \quad N := 20 \quad a := d \cdot N \quad \psi 0 := 0 \quad n := 0 .. N - 1 \quad \alpha_n := 2 \cdot n \cdot \frac{\pi}{N}$$

$$f(n, \phi) := (1 + \cos(\phi - \alpha_n)) \cdot e^{i \cdot a \cdot \cos(\phi - \alpha_n)} \qquad \beta := 0.07$$

$$K := 6 \qquad \Psi := (-1.168 \quad -0.78 \quad -0.47 \quad 0.47 \quad 0.78 \quad 1.168)^T$$

$$I0_n := \overline{f(n, \psi 0)} \qquad F\_0(x) := \sum_n I0_n f(n, x) \qquad F0(\phi) := \frac{F\_0(\phi)}{F\_0(\psi 0)}$$

$$\text{beam}(k, x) := \frac{\left(\sum_n \overline{f(n, \Psi_k)} f(n, x) \right)}{\left| \sum_n \overline{f(n, \Psi_k)} f(n, \Psi_k) \right|} \qquad \begin{array}{l} k := 0 .. K - 1 \\ \\ m := 0 .. K - 1 \end{array} \quad A_{k, m} := \text{beam}(m, \Psi_k)$$

$$g_k := F0(\Psi_k) \cdot \left(\frac{\beta}{|F0(\Psi_k)|} - 1 \right) \qquad dI := A^{-1} g$$

$$F1(x) := F0(x) + \sum_k dI_k \cdot \text{beam}(k, x) \qquad dB(x) := 20 \cdot \log(|x|)$$

$$y(x) := dB(\beta) \qquad y0(x) := dB(F0(x)) \qquad y1(x) := dB(F1(x))$$
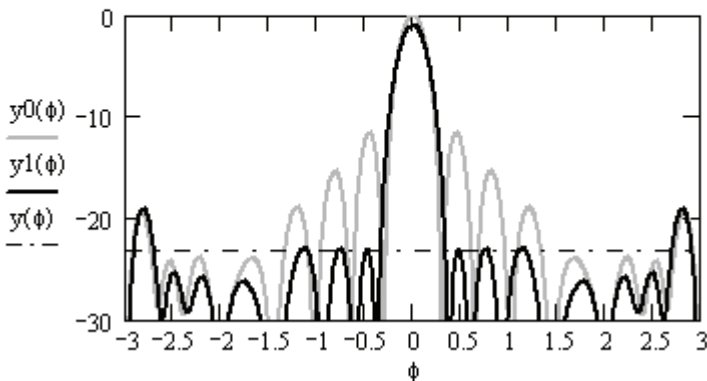
D: Program "Modified Woodward-Lawson Method" for Synthesizing Linear Source with Cosecant Radiation Pattern

Notations:

L − the length of the source [in wavelength]

N − the number of π-sections of abstract angle coordinate $u = 0.5\,kL\,\sin\theta$

δ − the direction of the cosecant maximum [in radians!]

N1 − the number of right-hand beams to form a cosecant descent

N2 − the number of left-hand side lobes of ARP

A(u) − cosecant ARP

{u1} − a set of directions for the right-hand beams

{u2} − a set of directions for the left-hand beams

f1(n1,u) − the right-hand beam with index number n1

f2(n2,u) − the left-hand beam with index number n2

F0(u) − ARP corresponding to the classical WLM

{β} − a set of phases for the right-hand beams [in radians!]

{w1} − a set of excitation coefficients for the right-hand beams

F1(u) − ARP corresponding to the modified WLM

μ − to what proportion should the side lobes be reduced

{w2} − a set of excitation coefficients for the left-hand beams

F2(u) − ARP corresponding to the modified WLM plus reducing the level of lobe labels

Recommendations

1. Rewrite the program code shown in the next page in the Mathcad environment. Where necessary, use imaginary unit *i*.

2. Format the X-Y Plots with traces as shown below, respectively.

Legend	Symbol	Line	Color	Type	Weight
trace 1	none	solid	grn	lines	2
trace 2	none	dash	blk	lines	2
trace 3	none	solid	blk	lines	2
trace 4	none	solid	blk	lines	1
trace 5	none	dash	blk	lines	1

| trace 1 | none ▾ | solid ▾ | grn ▾ | lines ▾ | 2 ▾ |

Legend	Symbol	Line	Color	Type	Weight
trace 1	none	solid	grn	lines	2
trace 2	none	dash	blk	lines	2
trace 3	none	solid	blk	lines	2

| trace 1 | none ▾ | solid ▾ | grn ▾ | lines ▾ | 2 ▾ |

3. Run the program; alter the phase column {β}, change the values μ. Be free to change all the parameters and improve the program itself, as you like.

Program code

$$L := 6.5 \quad N := \text{round}(\pi L) \qquad N = 20$$

$$\delta := 6 \cdot \text{deg} \qquad N1 := \text{round}(0.5N) \quad N2 := N - 1 - N1$$

$$n2 := 0 .. N2 \quad n1 := 0 .. N1 - 1$$

$$A(u) := \text{if}\left(u > \frac{\pi}{2} - 0.01 + \delta, \frac{-\sin(\delta)}{\cos(u)}, 0.0315\right)$$

The right hand rays are arranged as usual, but with a shift δ
from π/2. The column {u1} is of their angles.
The {u2} column is for left hand beams, numbered from
right to left (from p2 to 0).

$$u1_{n1} := 0.5\pi + \delta + n1 \cdot \frac{\pi}{N} \qquad u2_{n2} := \frac{\pi}{2} - \frac{1.5\pi}{N} + \left(\delta - n2 \cdot \frac{\pi}{N}\right)$$

$$f1(n1, u) := \frac{\sin\left[N \cdot \left(u - u1_{n1}\right)\right]}{N \cdot \left(u - u1_{n1}\right)} \quad f2(n2, u) := \frac{\sin\left[N \cdot \left(u - u2_{n2}\right)\right]}{N \cdot \left(u - u2_{n2}\right)}$$

$$F0(u) := \sum_{n1} A\left(u1_{n1}\right) \cdot f1(n1, u)$$

In the column {b}, set the optimal phases [in radians (!)]
for the right hand beams at your discretion.
If you have increased the number H, do not forget
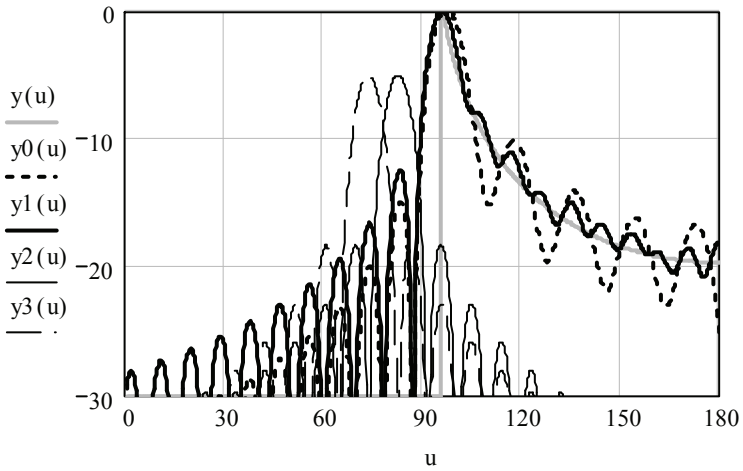to write additional elements of the column {b}.

$$ß := (0 \quad 1.43 \quad 0.8 \quad 1.43 \quad 0.8 \quad 1.47 \quad 0.99 \quad 1.457 \quad 0.77 \quad 1.95)^T$$

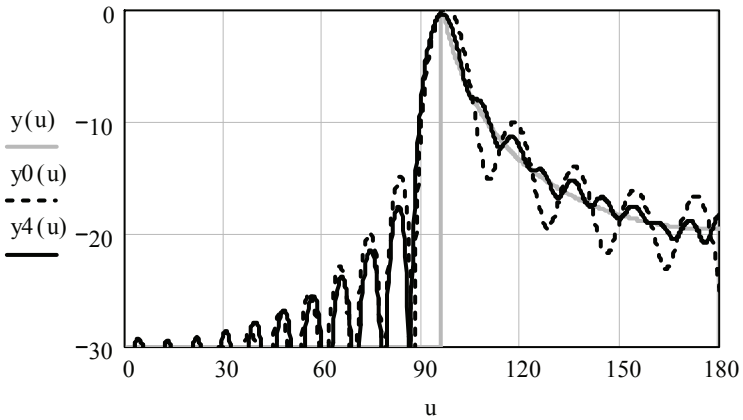$$w1_{n1} := A\left(u1_{n1}\right) \cdot e^{i \cdot ß_{n1}} \qquad F1(u) := \sum_{n1} w1_{n1} \cdot f1(n1, u)$$

$$\mu := 0.44$$

$$w2_{n2} := -\mu \cdot F1\left(u2_{n2}\right) \qquad F2(u) := F1(u) + \sum_{n2} w2_{n2} \cdot f2(n2, u)$$

$$dB(x) := 20 \cdot \log(|x|) \qquad y(x) := dB(A(x \cdot deg))$$

$$y0(x) := dB(F0(x \cdot deg)) \qquad y1(x) := dB(F1(x \cdot deg))$$

$$y2(x) := dB(f2(0, x \cdot deg)) - 5 \qquad y3(x) := dB(f2(1, x \cdot deg)) - 5$$

$$y4(x) := dB(F2(x \cdot deg))$$

BIBLIOGRAPHY

1. Davis B. *Inventions of Teaching: a Genealogy*. (Routledge, London, 2004), 344.
2. Kirton, M. "Adaptors and Innovators: a Description and Measure." *Journal of Applied Psychology*. 61 (1976): 622−29.
3. Stum J. "Kirton's Adaption-Innovation Theory: Managing Cognitive Styles in Times of Diversity and Change." *Emerging Leadership Journeys*, 1 (2009): 66−78.
4. Gulenko V.V. "Forms of Thinking." *Socionics, Mentality and Psychology of Individuals*. 43 (2002), (in Ukrainian).
5. Krasnova V.I. "Implementation of the Competency Approach in the Educational Process of High School." *Kazan Pedagogical Journal*. 3 (2009): 10-14, (in Russian).
6. Zhuk O.L. *Teaching Students: Competence Approach*. (RIHS, Minsk, 2009), 336 (in Russian).
7. "Spektr-M Observatory Takes Shape." Accessed July 09, 2019. http://www.russianspaceweb.com/spektr_m.html
8. Karasik, B.S., and Sergeev, A.V. "THz Hot-Electron Photons Counter." *IEEE Transactions on Applied Superconductivity*, 15 (2005): 618−623.
9. Vulih B.Z. *Introduction to functional analysis* (2nd ed.). (Science, Moscow, 1967): 416.
10. Kolmogorov, A.N., Fomin, S.V. *Elements of the Theory of Functions and Functional Analysis*. Dover Publications; Dover Books on Mathematics edition (February 16, 1999): 128.
11. Jeffreys, H. and Jeffreys, B. S. "Integration: Riemann, Stieltjes." In *Methods of Mathematical Physics, 3rd ed*. 26-36 Cambridge, England: Cambridge University Press, 1988.
12. Ufimtsev, P.Ya. "Edge Wave Diffraction Theory." In *Introduction to the Physical Theory of Diffraction* (Binom, Moscow, 2012): 372 (in Russian).
13. Hönl H., Maue A.W., Westpfahl K. *Theorie der Beugung. Handbuch der physic.* (Springer-Verlag OHG, Berlin, Göttingen, Heidelberg, 1961), 218.
14. Yanke E., Emde. F., Lesh F. *Special Functions: Formulas, Graphs, Tables* (Nauka, Moscow, 1964): 344 (in Russian).
15. Choni Yu.I. "Synthesis of Emitters Located near the Metal Body." PhD thesis. (Kazan Aviation Institute, 1968) (in Russian).

16. Choni Yu.I., Morozov G.A. "Optimization of Solutions for Antenna Synthesis Problems with Regard to Random Errors of their Implementation." *KAI Proceedings*. 164 (1974): 108−111 (in Russian).

17. Dymski, V. N.; Choni, Yu. I. "An approximate solution of problems of antenna synthesis allowing experimental simulation." *Radiophysics and Quantum Electronics*, 13 (September, 1970): 1069−1075. doi: 10.1007/BF01032776

18. Choni Yu.I. "Adjoint Operator Method and its Aspects in Regard to Antenna Synthesis." 86-91. In *IX Proceedings of the 2013 IX International Conference on Antenna Theory and Techniques*, (Odessa, Ukraine, September 2013) doi: 10.1109/ICATT.2013.6650690

19. Hook, Stephen M. "Inequalities for Eigenvalues of Selfadjoint Operators." *Transactions of the American Mathematical Society,* 318 (March, 1990): 237−259.

20. Blanchet L., Faye G., "Hadamard regularization", *Journal of Mathematical Physics*, 11 (2000): 7675–7714. doi:10.1063/1.1308506

21. Tikhonov, A.N., Arsenin V.Y. *Solution of Ill-posed Problems*. (Washington: Winston & Sons, 1977): 258.

22. Golub, G. H., Hansen P. Ch., O'Leary D. P. "Tikhonov Regularization and Total Least Squares," *SIAM Journal on Matrix Analysis and Applications*, 1 (1999), pp. 185-194. doi.org/10.1137/S0895479897326432

23. Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* (Cambridge, England: Cambridge University Press, 1966): 816.

24. Kornienko, L.G., Shifrin, Ya.S. "Statistical synthesis of antennas." 275−297 In *Problems of Antenna Engineering*, ed. L.D. Bakhrakh, D.I. Voskresensky. Moscow: Radio and Communication, 1989, 368 (in Russian).

25. Balanis, Constantine A. *Antenna Theory: Analysis and Design*, *4th Edition* (New Jersey, John Wiley & Sons, 2016): 1072

26. Krasnoselsky M.A., Vainikko G.M., Zabreiko PL. et al. *Approximate solution of operator equations*. (Nauka, Moscow, 1969): 456.

27. Bertsekas, D.P. *Constrained Optimization and Lagrange Multiplier Methods (Computer Science & Applied Mathematics)* (Academic Press, Cambridge, 1982), 412, doi: org/10.1016/C2013-0-10366-2

28. Choni, L.V., Choni, Yu.I. "A Modified Minimum-residual Algorithm." *Computational Mathematics and Mathematical Physics*. 36 (1996): 139−144.

29. Boyd, S., Vandenberghe, L. *Convex optimization*. (Cambridge University Press, New York, 2009), 730.

30. Choni, Yu.I. "Synthesis of an antenna according to a given amplitude radiation pattern." *Radio Engineering and Electronic Physics*. 16 (1971): 770−778.

31. Buck, R.C. *Advanced Calculus*. (Waveland Press, Long Grove, 2003), 160.

32. Choni, Yu.I., Tsupikov A.E. "Improving the Accuracy of Reproduction of a Desired Pattern in the Synthesis of Linear Source or Array Antenna in the Framework of the Beamforming Method." *Antennas*, 123 (2007): 34−36 (in Russian).

33. Raju, G.R.L.V.N. Srinivasa, Raju, G.S.N. "Synthesis of Cosecant and Square Patterns for EMC Applications." *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, 9 (July - August 2014): 01-06.

34. Choni, Yu.I. "Hodograph of Antenna's Local Phase Center: Computation and Analysis." *IEEE Transactions on Antennas and Propagation*, 63 (June 2015): 2819−2823, doi: 10.1109/ TAP.2015.2417894.

35. Choni, Yu.I., Hassan, A. "Optimal Adaptive Antenna Arrays for Asynchronous Communication System." *IEEE Transactions on Antennas and Propagation.* 60 (2012): 3071− 3076.

36. Monzingo, R.A., Miller, T.W. *Introduction to Adaptive Arrays*. (New York, John Wiley & Sons, 1980), 555.

37. Gavrilov AM, Dymsky V.N., Chabdarov Sh.M "Design of exciter system based on a given energy spectrum of random vibration." *Soviet Aeronaut.* 18 (1975): 29−33.

38. Vyalyshev A.I., Gavrilov A.M., Lyubashevsky G.S., Tartakovsky B.D., Choni Yu.I. "Synthesis of Systems for Compensation of Vibration and Sound Fields." *Soviet Physics. Acoustics.* 23 (1977): 136−139.

39. Anfinogentov V.I., Lyubashevsky G.S., Tartakovsky B.D., Choni Yu.I. "On the Optimal Distribution of the Excitation Potentials in the Synthesis of Compensation Systems of Vibro-Acoustic Fields." *Akusticheskij Zhurnal*. 29 (1983): 728-232.

40. Bennet J.C. and Griziotis, A. "Removal of environmental effects from antenna radiation patterns by deconvolution processing." *Proceedings of the IEE Conference*, 1 (1983): 224–228.

41. Bakhrakh, L.D., Kremenetsky, S.D., Kurochkin, A.P., Usin, V.A., Shifrin, Ya.S. *Methods for measuring parameters of radiating systems in near-field*. (Science, Leningrad, USSR, 1985), 272 (in Russian).

42. Plokhikh, S.A., Sazonov, D.M., Scherbakov, V.I. "Reconstruction of Antenna Patterns sing the Reference Antenna Method from Measuring Amplitude and Phases in the Near-Field", *Izvestiya Vuzov, Radioelectronics*, 2 (1987): 59−64, (in Russian).

43. Zhao, W. "Retrieval of Free Space Radiation Patterns through Measured Data in a Non-Anechoic Environment." PhD dissertation, (Syracuse University, USA), 2013.

44. Krivosheev, Yu.V. "Measurement of Antenna's Characteristics in the Fresnel Zone on a Sparse Grid of Angles," Doctoral dissertation, (Moscow, RF, "MPEU"), 2014 (in Russian).

45. Choni, Yu.I., Pirozhenko, S.A. "Retrieval of the Antenna Pattern from the Data Obtained in Non-Ideal Environments," *Izvestiya Vuzov, Radioelectronics*, 2 (1992): 43−50 (in Russian).

46. Choni, Yu.I., Abuhadma, L.K.T., Danilov, I.Yu. "Antenna Pattern Retrieval from Measurements in Non-Ideal Anechoic Chamber." 376−382 In *2018 Systems of Signals Generating and Processing in the Field of on Board Communications*. Moscow, (March 2018) doi: 10.1109/SOSG.2018.8350647.

47. Markov, G.T., Petrov, B.M., Grudinskaya, G.P. *Electrodynamics and propagation of radio waves* (Sovetskoe radio, Moscow, 1979), 376.

48. R. L. Taylor, "Ritz & Galerkin: The road to the finite element method." *Bulletin of International Association for Computational Mechanics*, 12 (2002): 2–5.