UNIVERSITY OF TRENTO - ITALY

PH.D. IN MATHEMATICS

XVIII CYCLE

Riccardo Aramini

On some open problems in the implementation of the linear sampling method

Supervisor

Prof. Michele Piana

Contents

Preface

Electromagnetic scattering is a physical phenomenon in which an electromagnetic incident wave is scattered by an obstacle or an inhomogeneity and the total field at any point in space can be expressed as the sum of the original incident field and the scattered one. The direct electromagnetic scattering problem consists in determining the scattered field, once the geometrical and physical properties of the scatterer, as well as the incident field, are known. Among the various corresponding inverse electromagnetic scattering problems we can conceive, we are particularly interested in the following one: to get information on the support of the scatterer, once the incident wave and the far-field pattern, i.e. the scattered wave considered at large distances from the scatterer, are known.

In correspondence with the recent development of several new techniques in remote sensing and non-invasive investigation, in the last years inverse electromagnetic scattering problems have increasingly drawn the attention of scientific community, in particular with regard to the following applications[1]:

1. *medical diagnostics and therapy:* for example, in using microwaves to detect bone marrow cancer (leukaemia) or breast cancer, as well as to make hyperthermia treatment;

2. *non-destructive testing:* for example, in looking for small cracks inside metallic or plastic structures;

3. *mine removal,* in the case in which one wants to recover the location of mines in a minefield from aerial measurements of the wave scattered by such mines when reached by a known electromagnetic field sent by a scout plane flying over them;

4. *radar,* when not only the presence and the number of some moving objects are to be detected, but also some information about their dimensions and shapes is needed.

In general, there are two main difficulties making inverse scattering problems hard to solve:

[1]For an interesting review of some applications and methods in inverse electromagnetic (and acoustic) scattering, we refer to [2].

(a) they are ill-posed (in the sense of Hadamard [40]);

(b) they are non-linear.

In order to briefly discuss these two points, we can observe what follows.

(a') Any reliable approach to the solution of an ill-posed problem has to face, at some stage, questions of uniqueness and stability; in particular, it is well-known that any numerical implementation of a method for solving an ill-posed (or an ill-conditioned) problem needs to involve, at some step, a regularization procedure in order to damp the wild oscillations that, without regularization, would completely blur the solution owing to the presence of noise in the measured data and an actually uncontrolled error propagation from the data to the solution themselves. However, such pathologies can be satisfactorily cured, at least for linear problems, by regularization theory: some basic concepts and techniques of the latter are, in fact, introduced in chapter 1.

(b') Unfortunately, the genuine non-linearity of inverse scattering problems in general prevents one from using the powerful tools of regularization theory holding in the linear case. Traditional approaches for solving such problems are substantially of two kinds:

(i') non-linear optimization schemes, which provide an iterative (and, often, very accurate) reconstruction of the scatterer starting from an initial guess about its geometrical properties;

(ii') weak scattering approximation methods, which allow one to linearize the problem by means of suitable approximations, such as physical optics (holding when the wavelength is much smaller than the linear dimensions of the scatterer and the latter is a conductor) or Born approximation (holding when the wavelength is larger than the linear dimensions of the scatterer and the latter is a penetrable object having a low contrast with respect to the background medium).

Both kinds of approach suffer from some heavy drawbacks. As regards (i'), we point out that the difficulty in implementing iterative optimization algorithms is twofold: first, they require long (and, sometimes, extremely long) reconstruction times; second, they need to be quite accurately initialized, while, on the other hand, there are several applications, e.g. in medical imaging, in which the a priori available information about the geometrical properties of the scatterer does not allow for an initial guess that is accurate enough (if any at all). As regards (ii'), we point out that weak scattering approximation methods typically require a priori knowledge of the physical conditions under which the scattering phenomenon has generated the measured far-field pattern: more precisely, one should know whether the object has been penetrated by the incident wave or not, and, in the latter case, what kind of boundary condition the total field satisfies on the boundary of

the object itself. Moreover, situations occur in which no weak scattering approximation is possible: a typical example is microwave tomography in medical imaging applications.

In order to overcome the above mentioned drawbacks, in 1996 a new approach to inverse scattering problem solving has been proposed by Colton and Kirsch [26] (with two important additional contributions by Colton, Piana, Potthast in 1997 [29] and by Cakoni, Colton, Haddar in 2002 [17]): it is the by now quite famous *linear sampling method*, which is essentially a computational procedure providing a visualization of the support of the scatterer. Roughly speaking, it works as follows: the $N \times N$ measurements of the far-field pattern of the scattered wave at N observation angles and for N incident fixed-frequency fields are put in a $N \times N$ matrix, called *far-field matrix*, and a finite grid of sampling points in a spatial region containing the scatterer is chosen; then, for each grid point z_l, a linear algebraic system is written, whose left-hand side coefficients are the elements of the far-field matrix, while the right-hand side is a known vector with N components depending on z_l; moreover, for each z_l, a regularized solution of the above linear system is determined and its Euclidean norm is computed. Finally, the shape of the scatterer can be detected by the set of grid points in which such a norm (playing the role of indicator function) is mostly large. Of course, several different indicator functions are possible, since a visualization of the scatterer profile can also be obtained by mapping the values of a suitable monotonically increasing or decreasing function I of the norm itself. However, for the mathematical and technical aspects concerning the linear sampling method in its traditional formulation, we directly refer, as regards this PhD thesis, to sections 2.4 and 2.5 of chapter 2 (which is entirely devoted to introducing the direct and, even more, inverse scattering problem we are interested in).

Here we would rather point out the main features of the linear sampling method and explain the reasons for its usefulness:

- as indicated by the name itself, the linear sampling method is actually linear: more precisely, the indicator function is ultimately obtained by solving a finite number of ill-conditioned linear systems (one for each grid point). This means that the implementation is computationally simple and the numerical instability typical of noisy inverse problems can be easily handled by using the classical tools of regularization theory for ill-conditioned linear systems;

- the above linearity does not derive from any sort of approximation based on particular physical conditions; in other terms, no approximation concerning the wavelength or the physical properties of the scatterer is needed;

- very little a priori information on the scatterer is required: more precisely, it is not necessary to know a priori the number of connected pieces forming the scatterer, nor

whether they are penetrable by the wave or not, and, in the latter case, which kind of (possibly mixed) boundary conditions the total field satisfies (piecewise) on the boundary of the scatterer. It suffices only to know that the latter is actually inside the chosen grid of sampling points;

- the implementation of the linear sampling method is computationally fast: more precisely, the reconstruction of a two-dimensional scatterer from real data requires a few minutes, while for complex three-dimensional objects not more than a couple of hours is typically necessary.

On the other hand, the linear sampling method has also some drawbacks. The main one is that, in the case of scattering from a penetrable and inhomogeneous object, it can only provide a visualization of the support of the scatterer, but no information at all about the point values of the index of refraction; however, one should remember that, even in a purely theoretical context, it is possible to prove that in several situations (e.g. anisotropic objects) only the support of the scatterer is uniquely determined and not the point values of the index of refraction. Another drawback is that scatterers are, in general, not accurately recovered as regards their possible concavities, which tend to be "convexified", as pointed out in [22].

Finally, the linear sampling method still presents some open problems from three points of view:

(a) its mathematical foundation;

(b) its numerical implementation;

(c) the quantitative assessment of its performances.

The original results obtained by working at this PhD thesis mainly[2] concern the previous points (b) and (c) and are illustrated in chapter 3, according to the approach proposed in [3] and adding some further details or applications. More precisely, we try to discuss and face the following four problems:

(i) is there a criterion suggesting how to choose the parameters of an "optimal" grid containing the scatterer (i.e. number of points and sampling distance)?

(ii) is it possible to give a characterization of the indicator function in terms of its physical meaning or analytical properties?

[2] As far as we know, also the two theorems 1.7.6 and 1.7.7, as well as the blended regularization presented in subsection 1.8.4, can be considered original results.

(iii) which is the spatial resolution power achievable by means of the linear sampling method?

(iiii) once the visualization map, i.e. the indicator function, is available, which general criterion can suggest the thresholding level for its values? Or, in other terms, what can be considered "large" or "small" for the indicator function (respectively depending on the increasing or decreasing monotonicity of the function I)?

To this end, in section 3.1 we present a new (no-sampling) implementation of the linear sampling method in which the set of the angle-discretized far-field equations for all sampling points is replaced by a single functional equation formulated in a Hilbert space defined as a direct sum of L^2 spaces; this removes the problem of choicing a grid of sampling points and allows one to determine, by means of a unique regularization process, a regularized solution of the above functional equation, in such a way that the regularization parameter does not depend any longer on the sampling point and an analytical representation for any indicator function is therefore possible without any sampling in the space. Then, for sake of simplicity, in section 3.2 we choose a particular indicator function whose analytical expression allows one to show that it is band-limited and, consequently, to obtain (in section 3.3) some theoretical information about the spatial resolution achievable by the method. Moreover, in the same framework of our no-sampling implementation, in section 3.4 we discuss the possibility of using a different family of indicator functions (with no apparent gain in visualization accuracy), while in section 3.5 we outline the technique of deformable contour models in order to face the problem stated in the previous point (iiii).

Finally, this PhD thesis ends with two appendices: appendix A, which collects in few pages a good number of definitions, notations, theorems and properties which we often need to use and recall (mainly in chapter 2), and appendix B, which contains all the figures[3] illustrating chapters 2 and 3 (chapter 1 has no figures): this should avoid an excess of fragmentation in the written text and make it more easily readable.

Throughout the text, the black square, i.e. the symbol ■, indicates the end of the proof of a theorem, while the empty square, i.e. the symbol □, indicates the end of a remark or of an example.

Acknowledgements

During the last three years, working under the guide of my advisor professor Michele Piana and with my colleague doctor Massimo Brignone has been a pleasure and an enriching experience for me, not only from a strictly professional or scientific viewpoint, but also by virtue of their valued friendship. I deeply hope that it will be possible to continue such a collaboration during the next years or decades.

[3] All of them have been realized by doctor Massimo Brignone, who is kindly acknowledged.

I also thank several professors and PhD students of the University of Trento, for their helpfulness, kindness and friendship: all of them have contributed to enable me to spend a good time in Trento.

Finally, I wish to express my gratitude to my friends and, above all, to my parents, who, although not directly involved in my scientific activity, have often helped me with their (not only psychological) support.

Trento, March 13th, 2007

CHAPTER 1

Inverse problems and regularization

In this chapter we introduce the concept of *inverse problem* and we explain that, in general, an inverse problem is ill-posed (or ill-conditioned), this implying, in particular, that its solution is either non-existing, or not unique or completely blurred by noise and, consequently, devoid of any physical meaning.

Then we show that such pathologies can be cured by regularization theory, which allows one to define a new concept of solution of an inverse problem: it is the so-called *regularized solution*, which represents a generally satisfying compromise between accuracy in reproducing the noisy data and stability with respect to noisy perturbations of the data themselves. Although a certain number of different regularization methods is known in literature, we mainly focus on Tikhonov's one, since it is just the one we shall use in chapters 2 and 3 to implement the linear sampling method.

1.1. Direct and inverse problems

From a strictly mathematical point of view, the concept of *inverse problem* is quite ambiguous, in the sense that it would be only possible to speak about a couple of reciprocally inverse problems, as suggested by a well-known statement of J. B. Keller [43]: "We call two problems *inverses* of one another if the formulation of each involves all or part of the solution of the other". However, the same author goes on: "Often, for historical reasons, one of the two problems has been studied extensively for some time, while the other has never been studied and is not so well understood. In such cases, the former is called the *direct problem*, while the latter is the *inverse problem*".

Mathematical physics is rich of such problems: their peculiarity is a sort of duality, by which the data of one problem are all or part of the unknowns of the other and conversely, so that it may be asked by virtue of which criterion a direct problem can be distinguished by an inverse

one. The fact is that from a physical point of view the situation is quite different, since the two problems are not on the same level: the direct one starts from known causes to compute unknown effects, i.e. it is oriented along a cause-effect sequence, while the corresponding inverse one works backwards, since it consists in computing the unknown causes of given effects. Generally speaking, direct problems have always been considered by physicists more fundamental than inverse ones, and consequently they also have been more studied.

Thus, the historical reasons mentioned by Keller are basically physical reasons, since only physical laws can establish what are the causes and what are the effects, and provide the equations relating the effects to the causes. Let us see some examples about this.

First of all, if we consider Newtonian mechanics, we know that its second law relates force (cause) to acceleration (effect) and, consequently, to trajectory. So a direct problem is, for instance, the computation of trajectories of particles from the forces acting upon them (and their initial conditions), while the corresponding inverse problem is the determination of the forces from the knowledge of the trajectories. From this point of view, Newton succeeded in solving the first inverse problem when he drew the explicit form of the gravitation force from the Kepler laws describing the trajectories of the planets.

However, with regard to the application of modern methods in inverse problem solving, other examples are more suitable. In scattering and diffraction theory, the aim of the direct problem is to calculate the scattered (or diffracted) waves starting from the knowledge of the sources and the obstacles, while the inverse problem consists in determining the obstacles when the data are the sources and the scattered (or diffracted) waves. This kind of inverse problems is very important in non-destructive evaluation (e.g. medical imaging), which consists in sounding an object by means of an appropriate radiation source.

Another typical example of direct problem can be found in wave-propagation theory, when, starting from the knowledge of a given source, one has to compute the field radiated by it (for instance, the radiation pattern of a given antenna); obviously, the corresponding inverse problem consists in determining the source from the knowledge of the radiated field (in the previous example, the aim is to compute the current distribution in the antenna, given the radiation pattern).

We can also consider potential theory: a direct problem is computing the potential generated by a known distribution of masses (or charges), while the corresponding inverse problem consists in determining the mass (or charge) distribution, given the potential generated by it.

Another field rich of this kind of problems is instrumental physics, i.e. the physics of instruments such as electronic devices, imaging systems and so on. In these cases, the direct problem consists in determining the output of the instrument (e.g. the image) from the knowledge of the input (e.g. the object) and the characteristics of the instrument (impulse response function, etc.), while the inverse problem is the computation of the input from the knowledge of the instrument and of its output.

We have already said that a direct problem is oriented along a cause-effect sequence; now we want to point out that it is also directed towards a loss of information, in the sense that its solution is the result of a transition from a physical quantity with a certain information content to another physical quantity with a smaller information content. This is a common feature for most direct problems and we shall investigate it more precisely in the next section. Here we only observe that in general the solution of a direct problem is much smoother than the data: for instance, the image yielded by a bandlimited system is smoother than the corresponding object (if the object involves out of band frequencies), the wave scattered by a rough obstacle can be smooth, and so on. An interesting example of this property can be found in [9], p. 3-4.

As a consequence, a conceptual difficulty common to most inverse problems arises, since their solution requires a transformation which should involve a gain of information. This difficulty is referred to as *ill-posedness* and we shall consider it in the next section.

Finally, although there exists a certain number of mathematically interesting nonlinear inverse problems, in the following we shall consider only linear inverse problems. The reason is threefold:

1) linear problems, eventually deriving from the linearization of nonlinear ones, are currently the most important for the applications;

2) well-known mathematical methods and efficient numerical algorithms for the computation of their solutions are already at disposal;

3) we are mainly interested in implementing the *linear sampling method* (abbr. LSM, to which section 2.5 is dedicated) for the solution of inverse scattering problems.

1.2. Well-posed and ill-posed problems

In the previous section we have mentioned ill-posedness as a typical feature of inverse problems: our aim is now to give some definitions and comments in order to be more precise in handling this important property.

First of all, let us recall the basic concept of *well-posed problem*, introduced for the first time in 1902 by the French mathematician Jacques Hadamard: he gave a definition of such a concept in a paper on boundary-value problems for partial differential equations and their physical interpretation [39]. In this first formulation, a problem was called *well-posed* if its solution was unique and existed for arbitrary data. However, in a subsequent treatise [40], written in 1923, Hadamard pointed up the requirement of continuous dependence of the solution on the data, since a solution that varies very much for small changes of the data cannot be considered a solution from a physical point of view: in fact, physical data are always affected by errors and an uncontrolled propagation of them in the solution makes the latter physically meaningless.

Definition 1.2.1. *A problem is well-posed (in the sense of Hadamard) if it satisfies the following three properties:*

1. *the solution is unique;*
2. *the solution exists for arbitrary data;*
3. *the solution depends continuously on the data.*

Definition 1.2.2. *A problem is ill-posed if at least one of the previous three properties is not verified.*

Therefore a problem is ill-posed if its solution is not unique or[1] does not exist for arbitrary data or[2] does not depend continuously on the data.

Hadamard was convinced that problems deriving from physics had always to be well-posed; this point of view was heavily conditioned by the physics of the nineteenth century. The mathematical requirements of existence, uniqueness and continuity of the solution correspond to the "philosophical" ideal of a unique, complete and stable determination of the physical events. Consequently, ill-posed problems were considered for a long time (up to the late 1960s) as mathematical pathologies, devoid of real interest in the context of applied mathematics, and were not seriously studied.

Anyway, the subsequent discovery of the ill-posedness of most inverse problems has fully changed this point of view: with the development of the inverse problems theory and its application to many areas of applied sciences, ill-posedness has become a crucial point for their solution, both in functional and numerical analysis.

For instance, the following impressive example of ill-posed problem, due to Hadamard himself [40], has been considered for many years of merely mathematical interest, but in 1977 it turned out that the basic inverse problem of electrocardiography [25], i.e. the reconstruction of the epicardial potential from body surface maps, can be formulated just as a Cauchy problem for an elliptic equation, i.e. a generalization of the Laplace equation.

Example 1.2.1. Let us consider the Laplace equation in two dimensions

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{1.1}$$

and the associated Cauchy problem characterized by the data

$$u(x,0) = \frac{1}{n}\cos(nx), \qquad \frac{\partial u}{\partial y}(x,0) = 0. \tag{1.2}$$

[1] Here *or* has the same meaning of the Latin *vel*.
[2] Idem.

Then the unique solution of this Cauchy problem is

$$u(x,y) = \frac{1}{n}\cos(nx)\cosh(ny). \tag{1.3}$$

If the same Laplace equation (1.1) is considered with the Cauchy data

$$u(x,0) = 0, \qquad \frac{\partial u}{\partial y}(x,0) = 0, \tag{1.4}$$

the unique solution of the new Cauchy problem is $u(x,y) = 0$. Now, for sufficiently large n the distance (measured, e.g., by means of the supremum norm) between the Cauchy data (1.2) and (1.4) can be made arbitrarily small for any x, while at any given finite distance from the x-axis the solution (1.3) grows to infinity. This is a classical example showing the effects produced by a non-continuous dependence of the solution on the data. It is interesting to observe that if the oscillating function (1.2) represents the experimental errors on the data, then, by linearity, the error propagation from the data to the solution is described by the function (1.3), and its effect is so heavy that the solution corresponding to these real data is physically meaningless. Besides, it can be shown that the solution doesn't exist for any data, but only for data endowed with some specific analyticity properties. □

1.3. Formulation of a linear inverse problem

The observations and comments in the previous sections suggest the following general statement: a direct problem, i.e. a problem oriented along a cause-effect sequence, is well-posed, while the corresponding inverse problem, which implies a reversal of the cause-effect sequence, is ill-posed. However, this statement is not completely meaningful unless we can yield an appropriate mathematical environment for the description of direct and inverse problems. For sake of perspicuity, we shall often use terms and expressions that are frequently used in imaging problems: obviously, this is not necessary, but it can help in some cases by virtue of the intuitive meaning of this terminology.

The first point is to define the direct problem: its solution provides a linear operator A, whose domain is the linear space X of the *objects* to be imaged, which correspond to suitable functions with certain properties, and whose range is in the linear space Y of the *images*, which correspond to appropriate functions describing, in the inverse problem, the measurable data. Naturally, X is called *object space* and Y *image space*; in this context, they represent functional spaces that are typically taken as Hilbert spaces. There are good reasons for this choice: first of all, we need a *distance*, in order to know whether two objects (or images) are close or not, so that our spaces have to be *metric*. Secondly, a scalar product turns out to be particularly appropriate in the case of discrete data: indeed, in such a case the operator A maps a function

into a vector and, if A is continuous[3], each component of this vector can be represented, by virtue of Riesz representation theorem[4], just in terms of a useful scalar product. Finally, the space L^2 of the measurable functions such that $\int |f(x)|^2\,dx < \infty$ contains the finite energy signals, which are the only ones to be physically achievable.

The second point is to observe that the space Y cannot coincide with the range $\mathcal{R}(A)$ of the operator A, but can only strictly contain it. In fact, recalling the obvious decomposition

$$Y = \overline{\mathcal{R}(A)} \oplus \mathcal{R}(A)^{\perp}, \tag{1.5}$$

it's easy to realize that $\mathcal{R}(A)$ is the set of all the *noise-free images*: the direct problem being well-posed, the operator A associates a unique image to each object. As already observed in section 1.1, this image may be rather smooth, since its information content is smaller than the one of the corresponding object. However, a *measured image* is certainly a *noisy image*, since it corresponds to a noise-free image corrupted by the noise affecting the measurement process: as a consequence, the smoothness property mentioned above may not be satisfied and, in any case, the measured image may not belong to $\mathcal{R}(A)$.

So the third point is to define properly the Hilbert space Y, in such a way that it contains both the noise-free and the noisy images.

Summing up, the solution of the direct problem defines a linear continuous operator $A : X \to Y$ between two Hilbert spaces: X is the space of the objects, Y the space of all the possible images, noise-free and noisy ones.

Definition 1.3.1. *Let X and Y be two normed vector spaces (not necessarily complete); we shall denote with $\mathcal{B}(X,Y)$ the space of the linear continuous (i.e. bounded) operators from X to Y.*

Remark 1.3.1. We recall that the operatorial norm, defined in any one of the following equivalent ways:

1. $\|A\| := \sup_{\|x\|_X \neq 0} \dfrac{\|Ax\|_Y}{\|x\|_X}$;
2. $\|A\| := \sup_{\|x\|_X = 1} \|Ax\|_Y$;
3. $\|A\| := \sup_{\|x\|_X \leq 1} \|Ax\|_Y$;
4. $\|A\| := \inf\{C \in \mathbb{R} \mid \|Ax\|_Y \leq C\|x\|_X \ \forall x \in X\}$,

makes $\mathcal{B}(X,Y)$ a normed vector space, which is also complete if Y is complete in turn. *In the following, unless otherwise specified, we shall always consider the particular case in which X and Y are Hilbert spaces.* □

[3] As we shall always admit, insofar as the direct problem is well-posed.

[4] See theorem A.1.3 in appendix A.

By virtue of the previous mathematical setting, now we can explain more precisely the loss of information that characterizes the solution of a direct problem. First of all, it may obviously happen that two, or even more, different objects have exactly the same image. Since the operator is linear, this fact corresponds to the existence of objects (called *invisible*) whose image is zero. In other terms, given any object in the space X, if an invisible object is added to it, we obtain a different object which has exactly the same image of the previous one. Secondly, it may happen that two very distant objects have very close images; in other words, there may exist very large sets of different objects that are mapped by the operator A into very small sets of images.

Thus, if we now consider the inverse problem

$$g = Af, \tag{1.6}$$

i.e. the problem of determining the object f corresponding to a given image g, it is easy to realize that its ill-posedness is strictly related to the loss of information that affects the solution of the direct one. Indeed, if the image g corresponds to two (or more) different objects, the solution of the inverse problem is not unique (in this case, $\mathcal{N}(A) \neq \{0\}$, being $\mathcal{N}(A)$ the kernel of A). If g is a noisy image, which doesn't belong to $\mathcal{R}(A)$, then the solution to the inverse problem doesn't exist (in this case, $D(A^{-1}) \neq Y$, being $D(A^{-1})$ the domain of the inverse operator A^{-1}). Finally, if we have two neighbouring images g_1, g_2 such that the corresponding objects f_1, f_2 are very distant, then the solution of the inverse problem doesn't depend continuously on the data (in this case, the operator A^{-1} is not continuous). Obviously, all three cases may occur in the same inverse problem. So we can understand why the requirements of the following definition, which is simply the reformulation of definition 1.2.1 for a linear inverse problem, are not, in general, fulfilled:

Definition 1.3.2. *A linear inverse problem is well-posed if the three following properties hold:*

1. $\mathcal{N}(A) = \{0\}$;
2. $D(A^{-1}) = Y$;
3. $A^{-1} : Y \to X$ *is continuous.*

However, it is worthwhile noticing, also for future purpose, that if the first two requirements are satisfied, also the third one is: this is a consequence of the two following theorems.

Theorem 1.3.1. [Open mapping theorem]. *Let X and Y be Banach spaces; if $A \in \mathcal{B}(X, Y)$ is surjective, then the image by A of an open set in X is an open set in Y, i.e. the map A is open.*

Proof. This is a fundamental theorem in functional analysis. For a proof, see, for instance, [57]. ■

Theorem 1.3.2. *Let X and Y be Banach spaces; if $A \in \mathcal{B}(X,Y)$ is bijective, then A^{-1} is continuous.*

Proof. In particular, A is continuous and surjective: by theorem 1.3.1, it maps open sets into open sets. Furthermore, A is injective, so that A^{-1} exists, $D(A^{-1}) = Y$ and the inverse image by A^{-1} of an open set in X is an open set in Y, i.e. A^{-1} is continuous. ■

It is also interesting to observe that neither the mere continuity of the inverse operator A^{-1}, which is trivially verified in a finite-dimensional context[5], would be sufficient to assure the stability of the solution: in other terms, well-posedness is not a sufficient condition for the stability of the solution of a linear inverse problem. Indeed, let us assume that A^{-1} is well-defined and continuous. Then, with reference to equation (1.6), if δg is a small variation of the datum and δf is the corresponding variation of the solution, the continuity of A^{-1} implies

$$\|\delta f\|_X \leq \|A^{-1}\|\|\delta g\|_Y, \tag{1.7}$$

where $\|\cdot\|_X$, $\|\cdot\|_Y$ now denote respectively the norms in the Hilbert spaces X and Y induced by the scalar products in X and in Y themselves.

On the other hand, the continuity of A implies

$$\|f\|_X \geq \frac{\|g\|_Y}{\|A\|}, \tag{1.8}$$

so that

$$\frac{\|\delta f\|_X}{\|f\|_X} \leq \|A\|\|A^{-1}\|\frac{\|\delta g\|_Y}{\|g\|_Y}. \tag{1.9}$$

The real positive number

$$C(A) := \|A\|\|A^{-1}\| \tag{1.10}$$

is said *condition number* and provides an estimate of the instability of the problem. Since we have

$$\|g\|_Y = \|Af\|_Y \leq \|A\|\|f\|_X = \|A\|\|A^{-1}g\|_X \leq \|A\|\|A^{-1}\|\|g\|_Y, \tag{1.11}$$

it is always

$$C(A) \geq 1. \tag{1.12}$$

Hence, if $C(A) \gg 1$ (in some LSM-applications[6], for instance, it is up to the order of 10^{10}), a small variation δg on the datum can produce an enormous variation δf on the solution: in such a case, the inverse problem is said *ill-conditioned*. In other terms, relation (1.9) shows

[5] This typically happens when the original continuous problem is discretized.

[6] See chapter 2.

that the presence of an error, however small, on the datum of an ill-conditioned problem can make its solution extremely unstable[7].

So the next point is: how to cure ill-posedness (or ill-conditioning)?

1.4. Facing ill-posedness

First of all, we could observe that the property of non-continuous dependence of the solution on the data is strictly verified only for ill-posed problems formulated in infinite-dimensional spaces; in practice, one has to deal with discrete data and with discrete, finite-dimensional problems. Now, the discrete version of a continuous linear inverse problem is a linear algebraic system, apparently an easy mathematical problem: there exist a lot of methods that yield a numerical solution to it. However, this problem is obtained by discretizing a problem with very bad mathematical properties: indeed, its mathematical solution simply doesn't work, in the sense that it is physically meaningless. If we now remember the last part of the previous section, we can easily imagine what happens.

We already know that in the continuous case small oscillating data can produce large oscillating solutions. In any inverse problem, data are always affected by noise, which can be considered as a small randomly oscillating function. Thus, the solution method amplifies the noise generating a large and wildly oscillating function which fully hides the physically meaningful solution corresponding to the exact, i.e. noise-free data. This property is still true for the discrete version of the continuous ill-posed problem, since the corresponding linear algebraic system is *ill-conditioned*: even if the solution exists and is unique, it is completely corrupted by a minimum error on the data.

Summing up, whereas on the one hand we can compute a unique solution of our algebraic system, on the other hand this solution is meaningless; the physically meaningful solution we are seeking is not a solution of the problem but only an approximate solution, in the sense that, when mapped by the matrix representing the discretized version of the operator A, it reproduces the data not exactly, but only within the experimental errors. Anyway, if we search for approximate solutions, they turn out to form a very large set, which contains completely different functions, as a consequence of the loss of information in the direct problem. Thus our problem is: how can we choose the good ones?

We can now state the so-called *golden rule* for solving ill-posed inverse problems: *look for approximate solutions satisfying additional constraints coming from the physics of the problem.* Let us clarify this statement by means of the mathematical model introduced just above.

The set of the approximate solutions that reproduce (within a certain amount of error) the

[7]We point out that inequality (1.9) cannot be improved since, in some cases, equality holds true: see [9], p. 82-83.

same data function is the set of the objects whose images are close to the measured one. The set of such objects is too large, due to the loss of information in the direct problem. Thus we need some additional information, also called *a priori* or *prior* information, in order to compensate this loss. This information is additional in the sense that it cannot be retrieved from the image or from the properties of the mapping A that describes the direct problem, but represents some expected physical properties of the object. Its role is to reduce the set of the objects that are compatible with the measured image, or also to distinguish meaningful objects from spurious ones, generated by overwhelming propagation of the noise affecting the image.

Let us see some simple examples of additional information.

1. The object cannot be too large: this implies a constraint in the form of an upper bound on the object itself, or its intensity, or its energy, etc.

2. The object is smooth, so that, for example, its derivatives must be smaller than a certain quantity.

3. The object is known to be non-negative.

4. The object must be different from zero only inside a given bounded region.

Furthermore, a quite different kind of additional information may be represented by statistical properties of the objects. In this case, the objects to be restored are assumed to be realizations of a random process with known probability distribution (this can be a way of expressing our previous experience in object restoring). Although a complete knowledge of the probability distribution is not always at disposal, also a partial knowledge of statistical properties of the object (for instance, the expectation values or covariance matrices) may be useful.

Thus, the principle of the *regularization methods* is to use the additional information explicitly, from the very beginning, to construct families of approximate solutions, that is of objects compatible with the measured image. These methods are now one of the most effective tools for the solution of inverse problems.

1.5. Generalized inverse operators

Given an ill-posed linear inverse problem, a first step towards the objective determination of an approximate solution consists of looking for functions minimizing in some sense the distance between their image by the operator A and the datum. More precisely, we introduce the least-squares problem associated to the linear inverse one (1.6), defined as the problem of determining $f \in X$ such that

$$\|Af - g\|_Y = \text{minimum}, \tag{1.13}$$

where X and Y are Hilbert spaces.

Definition 1.5.1. *A solution of the least-squares problem (1.13) is said a normal solution or pseudosolution.*

The characterization of pseudosolutions is given by the following theorem.

Theorem 1.5.1. *Given $A \in \mathcal{B}(X,Y)$, let P denote the linear projection onto $\overline{\mathcal{R}(A)}$, closure of the range of A, and let g be a generic element of Y. Then the following properties of $u \in X$ are equivalent:*

(i) $Au = Pg$;

(ii) $\|Au - g\|_Y \leq \|Af - g\|_Y \quad \forall f \in X$;

(iii) $A^*Au = A^*g$,

where A^* denotes the adjoint operator of A, defined, as usual, by the condition[8] $(Af, g)_Y = (f, A^*g)_X \ \forall f \in X, \forall g \in Y$.

Proof. (i) $\Rightarrow$ (ii): by virtue of the decomposition $Y = \overline{\mathcal{R}(A)} \oplus \mathcal{R}(A)^\perp$, one has that $Pg - g \in \mathcal{R}(A)^\perp$; besides $Af - Pg \in \overline{\mathcal{R}(A)}$, so that

$$\|Af - g\|_Y^2 = \|Af - Pg\|_Y^2 + \|Pg - g\|_Y^2. \tag{1.14}$$

Then, by using hypothesis (i), we have

$$\|Af - g\|_Y^2 = \|Af - Pg\|_Y^2 + \|Au - g\|_Y^2 \geq \|Au - g\|_Y^2 \qquad \forall f \in X, \tag{1.15}$$

i.e. (ii).
(ii) $\Rightarrow$ (iii): since $Pg \in \overline{\mathcal{R}(A)}$ and $\overline{\mathcal{R}(A)}$ is closed, there exists a sequence $\{f_n\}_{n=1}^\infty$ such that $Pg = \lim_{n\to\infty} Af_n$, i.e. $\lim_{n\to\infty} \|Af_n - Pg\|_Y = 0$. If we now assume that $u \in X$ is such that (ii) holds, by virtue of the continuity of the norm we have

$$\|Au - g\|_Y^2 = \|Au - Pg\|_Y^2 + \lim_{n\to\infty} \|Af_n - g\|_Y^2 \geq \|Au - Pg\|_Y^2 + \|Au - g\|_Y^2. \tag{1.16}$$

It follows that $Au - Pg = 0$, then $Au - g = Pg - g$ and $A^*Au - A^*g = A^*(Pg - g)$. But $Pg - g \in \mathcal{R}(A)^\perp = \mathcal{N}(A^*)$, so that (iii) is true.
(iii) $\Rightarrow$ (i): if (iii) holds, then $Au - g \in \mathcal{N}(A^*) = \mathcal{R}(A)^\perp$ and (i) follows. ■

[8] In the following, we shall denote with $(\ ,\ )_X$ the scalar product in a Hilbert space X. We choose "right component conjugation", i.e. $(x_1, ax_2)_X = \bar{a}(x_1, x_2)_X \ \forall a \in \mathbb{C}$ and $\forall x_1, x_2 \in X$, where we obviously denote with $\bar{a}$ the complex conjugate of $a \in \mathbb{C}$.

If $g \in Y$, the set of the pseudosolutions associated to g will be denoted with S_g. Clearly, by virtue of condition (i) of theorem 1.5.1, it follows that S_g is empty if and only if $\mathcal{R}(A)$ is not closed and $g \in \overline{\mathcal{R}(A)} \setminus \mathcal{R}(A)$. In other terms, it holds

$$S_g \neq \emptyset \Leftrightarrow g \in \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}. \tag{1.17}$$

Let us now suppose that S_g is not empty and let u_0 be a generic pseudosolution: it's trivial to show that

$$S_g = \{u = u_0 + \varphi, \quad A\varphi = 0\}, \tag{1.18}$$

or, in a shorter form,

$$S_g = u_0 + \mathcal{N}(A). \tag{1.19}$$

Then, it is easy to realize that the set S_g is convex and closed in the Hilbert space X; therefore, for a well-known theorem, there exists a unique pseudosolution with minimal norm. Hence, we can introduce the following definition.

Definition 1.5.2. *Given the inverse problem $g = Af$, if the set S_g of its pseudosolutions is not empty, the unique element of S_g having minimal norm will be called* generalized solution *of the problem and it will be denoted with $f^{\dagger}$. Moreover, the operator $A^{\dagger} : D(A^{\dagger}) \to X$, defined by the condition*

$$A^{\dagger}g = f^{\dagger} \qquad \forall g \in D(A^{\dagger}) \subset Y, \tag{1.20}$$

will be called generalized inverse operator.

Recalling the double implication (1.17), we can immediately realize that

$$D(A^{\dagger}) = \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}. \tag{1.21}$$

It is worthwhile observing that the condition of minimal norm in the definition of generalized solution may have a quite natural physical meaning. Indeed, for example, the L^2-norm of a signal is a measure of its energy and minimizing the energy of a signal is a typical way to reduce its instability[9].

Of course, the concepts of generalized solution and of generalized inverse operator are fundamental in studying linear inverse problems. In order to describe them in a better way, we give the two following theorems.

Theorem 1.5.2. *The generalized solution $f^{\dagger}$ is the unique pseudosolution belonging to $\mathcal{N}(A)^{\perp}$.*

[9]However it may happen that minimum conditions other than the L^2 minimization more realistically fulfil the problem requirements. A possible generalization of the definition of generalized solution is given by replacing the minimum norm condition with the more general constraint $\|Cf\|_Z =$ minimum, where C is a closed linear operator with range in the Hilbert space Z. The closedness condition is due to the fact that the use of C-generalized inverse operators is particularly useful in the case of closed differential operators.

Proof. Since $f^\dagger \in X$, we have the obvious decomposition

$$f^\dagger = u_1 + u_2, \tag{1.22}$$

with $u_1 \in \mathcal{N}(A)^\perp$ and $u_2 \in \mathcal{N}(A)$. We immediately have

$$u_1 = f^\dagger - u_2, \tag{1.23}$$

and, by remembering representation (1.18) and recalling that $f^\dagger$ is, in particular, a pseudosolution, we get that u_1 is a pseudosolution too. Furthermore, we can write

$$\|f^\dagger\|_X^2 = \|u_1 + u_2\|_X^2 = \|u_1\|_X^2 + \|u_2\|_X^2 \geq \|u_1\|_X^2. \tag{1.24}$$

Since by definition $f^\dagger$ is the unique pseudosolution of minimal norm, the relation (1.24) can hold only if $u_2 = 0$. This implies that $f^\dagger = u_1$ and so $f^\dagger \in \mathcal{N}(A)^\perp$. Finally, if we substitute $f^\dagger$ to u_0 in the decomposition (1.19), we find that $f^\dagger$ is the *unique* pseudosolution in $\mathcal{N}(A)^\perp$. ■

Theorem 1.5.3. *The generalized inverse operator $A^\dagger$, defined by relation (1.20), is linear.*

Proof. Let g_1 and g_2 be two elements of the domain $D(A^\dagger)$ of the generalized inverse operator $A^\dagger$. Then, remembering condition (i) of theorem 1.5.1 and relation (1.20), we have immediately

$$AA^\dagger g_1 = Pg_1, \qquad AA^\dagger g_2 = Pg_2. \tag{1.25}$$

Since Pg_1, $Pg_2 \in \mathcal{R}(A)$, by linearity of A and P we have that also $P(g_1 + g_2) = Pg_1 + Pg_2 \in \mathcal{R}(A)$. Thus, not only $g_1 + g_2 \in D(A^\dagger)$ and $AA^\dagger(g_1 + g_2) = P(g_1 + g_2)$, but, recalling the two relations (1.25), we also get

$$AA^\dagger g_1 + AA^\dagger g_2 = AA^\dagger(g_1 + g_2), \tag{1.26}$$

and then, by linearity of A, we have that $A^\dagger g_1 + A^\dagger g_2 - A^\dagger(g_1 + g_2) \in \mathcal{N}(A)$. But $A^\dagger g_1$, $A^\dagger g_2$ and $A^\dagger(g_1 + g_2)$ are the generalized solution respectively corresponding to the data g_1, g_2 and $g_1 + g_2$, so that $A^\dagger g_1, A^\dagger g_2, A^\dagger(g_1 + g_2) \in \mathcal{N}(A)^\perp$. Since the latter is a linear subspace of X, it follows that $A^\dagger g_1 + A^\dagger g_2 - A^\dagger(g_1 + g_2) \in \mathcal{N}(A)^\perp$. Summing up, we have found that

$$A^\dagger g_1 + A^\dagger g_2 - A^\dagger(g_1 + g_2) \in \mathcal{N}(A) \cap \mathcal{N}(A)^\perp. \tag{1.27}$$

But obviously $\mathcal{N}(A) \cap \mathcal{N}(A)^\perp = \{0\}$, so that

$$A^\dagger g_1 + A^\dagger g_2 = A^\dagger(g_1 + g_2). \tag{1.28}$$

In a fully analogous way it can be shown that $A^\dagger(ag) = aA^\dagger g \ \ \forall a \in \mathbb{C}$. ■

We can also establish a relation between the range of the generalized inverse operator and the range of the adjoint one: this is the aim of the following theorem.

Theorem 1.5.4. *Given $A \in \mathcal{B}(X,Y)$, it holds $\mathcal{R}(A^*) \subset \mathcal{R}(A^\dagger)$. Furthermore, if $\mathcal{R}(A)$ is closed, then $\mathcal{R}(A^*) = \mathcal{R}(A^\dagger)$.*

Proof. If $u \in \mathcal{R}(A^*)$, then $u \in \overline{\mathcal{R}(A^*)} = \mathcal{N}(A)^\perp$. If we define the element $g := Au$, then u is the generalized solution corresponding to g, since it is trivially pseudosolution and has no components in $\mathcal{N}(A)$. Hence $u = A^\dagger g$ and so $u \in \mathcal{R}(A^\dagger)$; summing up, $\mathcal{R}(A^*) \subset \mathcal{R}(A^\dagger)$.
Furthermore, let us now suppose that $\mathcal{R}(A)$ is closed; then, it suffices to prove the other inclusion $\mathcal{R}(A^\dagger) \subset \mathcal{R}(A^*)$. If $u \in \mathcal{R}(A^\dagger)$, it is also $u \in \mathcal{N}(A)^\perp$ by virtue of theorem 1.5.2. Now, if $\mathcal{R}(A)$ is closed, also $\mathcal{R}(A^*)$ is closed[10] and then $\mathcal{N}(A)^\perp = \mathcal{R}(A^*)$. It follows that $u \in \mathcal{R}(A^*)$ and, finally, $\mathcal{R}(A^\dagger) \subset \mathcal{R}(A^*)$. ■

Since we have introduced the generalized solution and the generalized inverse operator, we are in a position to formulate a new inverse problem that consists in determining the solution of two subsequent minimum problems, described by the two equations

$$\|Af - g\|_Y = \text{minimum} \tag{1.29}$$

and

$$\|f\|_X = \text{minimum}. \tag{1.30}$$

Such a problem is well-posed if and only if, $\forall g \in Y$, the generalized solution exists unique and the generalized inverse operator is continuous. There is an entire class of operators in $\mathcal{B}(X,Y)$ for which the well-posedness of the problem (1.29), (1.30) is ensured. Indeed, if $\mathcal{R}(A)$ is closed, the space of the data can be decomposed in the form $Y = \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$. It follows that $D(A^\dagger) = Y$ and then, $\forall g \in Y$, the set S_g of pseudosolutions is not empty. Hence, the existence and uniqueness of the generalized solution is a direct consequence of its definition, while the continuity of the generalized inverse operator is guaranteed by the following lemma and the subsequent theorem.

Lemma 1.5.5. *If $A \in \mathcal{B}(X,Y)$ and $\mathcal{R}(A)$ is closed, then $\exists\, m > 0$ such that*

$$\|Af\|_Y \geq m\|f\|_X \qquad \forall f \in \mathcal{N}(A)^\perp. \tag{1.31}$$

Proof. If A is the null operator $A = 0$, then $\mathcal{N}(A)^\perp = \{0\}$ and, as a consequence, inequality (1.31) is trivially verified by choosing an arbitrary $m \in \mathbb{R}$.

If $A \neq 0$, i.e. $\mathcal{N}(A)^\perp \neq \{0\}$, then the restriction $A' : \mathcal{N}(A)^\perp \to \mathcal{R}(A)$ of A is a bijective and continuous operator between two Hilbert spaces and then, by theorem 1.3.2, $(A')^{-1}$ is continuous. Hence, $\forall f \in \mathcal{N}(A)^\perp$, we have

$$\|f\|_X = \|(A')^{-1}Af\|_X \leq \|(A')^{-1}\|\|Af\|_Y, \tag{1.32}$$

[10] The proof of this statement is not immediate and can be found, e.g., in [11], p. 72.

so that the thesis holds with $m := \|(A')^{-1}\|^{-1}$. ■

Theorem 1.5.6. *Let $A \in \mathcal{B}(X,Y)$: then $A^\dagger$ is continuous $\Leftrightarrow$ $\mathcal{R}(A)$ is closed.*

Proof. "⇒": by absurd, if $\mathcal{R}(A)$ were not closed, then $D(A^\dagger)$ would be dense in Y, being $D(A^\dagger) = \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$. Clearly, $\forall g \in D(A^\dagger)$ it holds:

$$AA^\dagger g = Pg. \tag{1.33}$$

Since $A^\dagger$ is linear, continuous and densely defined in Y, it can be extended to a linear continuous operator $\hat{A}^\dagger$ defined on all Y. Hence, if $g \in D(A^\dagger)$, equation (1.33) holds; if $g \in Y \setminus D(A^\dagger)$, let $\{g_n\}_{n=0}^\infty$ be a sequence in $D(A^\dagger)$ such that $\lim_{n\to\infty} g_n = g$, i.e. $\lim_{n\to\infty} \|g_n - g\|_Y = 0$. Since $\{g_n\}_{n=0}^\infty \subset D(A^\dagger)$, by (1.33) we have $AA^\dagger g_n = Pg_n$ $\forall n \in \mathbb{N}$ or, equivalently,

$$A\hat{A}^\dagger g_n = Pg_n \qquad \forall n \in \mathbb{N}; \tag{1.34}$$

then, taking the limit as $n \to \infty$ of both members of the previous equation and using the continuity of A, $\hat{A}^\dagger$ and P, we finally get

$$A\hat{A}^\dagger g = Pg. \tag{1.35}$$

Summing up, we have found that

$$A\hat{A}^\dagger g = Pg \qquad \forall g \in Y. \tag{1.36}$$

But this means that equation $Af = g$ has a pseudosolution, precisely $\hat{A}^\dagger g$, for each datum $g \in Y$; on the other hand, it is always possible to choose $g \in \overline{\mathcal{R}(A)} \setminus \mathcal{R}(A)$, so that, for such a g, no pseudosolution of $Af = g$ exists (see the double implication (1.17)). Hence, we have got a contradiction.
"⇐": since we have in this case

$$AA^\dagger g = Pg \qquad \forall g \in Y, \tag{1.37}$$

we can write

$$\|g\|_Y \geq \|Pg\|_Y = \|AA^\dagger g\|_Y \geq m\|A^\dagger g\|_X \qquad \forall g \in Y, \tag{1.38}$$

where the last inequality holds by virtue of lemma 1.5.5. Finally, we get

$$\|A^\dagger g\|_X \leq \frac{1}{m}\|g\|_Y, \tag{1.39}$$

i.e. $A^\dagger$ is continuous. ■

Theorem 1.5.6 states that if $\mathcal{R}(A)$ is closed, the problem of determining the generalized solution is well-posed, while if $\mathcal{R}(A)$ is not closed, the determination of $f^\dagger$ is surely ill-posed.

1.5.1. The case of compact operators

Among the operators whose range, in general, is not closed, there are some extremely meaningful classes that are frequently met in the applications: an important example is the one of compact operators (see, e.g., [57]). Firstly we recall some basic definitions and theorems about them.

Definition 1.5.3. *Let K be a subset of a topological space; K is said* compact *if and only if for any (not necessarily countable) family of open sets $\{U_i\}_{i\in I}$ such that $\cup_{i\in I}A_i \supset K$, there exists a finite subset $J \subset I$ such that $\cup_{i\in J}A_i \supset K$.*

The previous definition can be summarized saying that K is called *compact* if and only if from any one of its open coverings it is always possible to extract a finite subcovering.

Definition 1.5.4. *A subset of a topological space is said to be* relatively compact *if its closure is compact.*

Remark 1.5.1. If K is a compact subset of a normed vector space X (not necessarily complete), then it is bounded. Indeed, in a normed vector space we can consider a generic open ball of centre x_0 and radius r, defined as:

$$B(x_0, r) := \{x \in X \mid \|x - x_0\|_X < r\}. \tag{1.40}$$

Then, fixed $r > 0$, for each point $x_i \in K$ we can consider the ball $B_i := B(x_i, r)$. Clearly $\{B_i\}_{i\in I}$ is an open covering of K; since K is compact, we can extract a finite subcovering, i.e. there exists a finite subset $J \subset I$ such that $K \subset \cup_{i\in J}B_i$. If we now consider the maximum R of the distances between the centres of the balls, i.e.

$$R := \max_{i,j\in J} \|x_i - x_j\|_X, \tag{1.41}$$

and arbitrarily choose an x_{i_0} among the centres x_i, $i \in J$, we easily realize that $\cup_{i\in J}B_i \subset B(x_{i_0}, R + r)$; hence $K \subset B(x_{i_0}, R + r)$, and this means that K is bounded. □

Definition 1.5.5. *Let $A : X \to Y$ be a linear operator between the normed vector spaces (not necessarily complete) X and Y; A is said* compact *if it maps bounded sets onto relatively compact sets.*

Remark 1.5.2. If a linear operator is compact, then it is bounded. Indeed, if, according to notation (1.40), we denote with $\overline{B(0,1)}$ the closure of the open ball in X with centre in $0 \in X$ and radius 1, we have that, by definition of compact operator, the set $\overline{A\left(\overline{B(0,1)}\right)}$ is compact in Y: hence, by virtue of remark 1.5.1, such a set is also bounded. In particular, this implies that there exists $C > 0$ such that

$$\|Ax\|_Y \leq C \qquad \forall x \in X \ \text{ such that } \ \|x\|_X = 1; \tag{1.42}$$

it follows that

$$\sup_{\|x\|_X=1} \|Ax\|_Y \leq C, \tag{1.43}$$

i.e. A is bounded. □

We now state two theorems: the first one, remembering theorem 1.5.6, implies that if the inverse problem we are interested in is modelled by a linear compact operator, then, in general, it is ill-posed; the second one is very useful in the computational implementations, in which also continuous problems need to be discretized in finite-dimensional spaces.

Theorem 1.5.7. *Let X and Y be Banach spaces; if $A \in \mathcal{B}(X,Y)$ is compact and its range is closed, then the range dimension is finite (and the operator is said* of finite range*).*

Proof. The operator $A : X \to \mathcal{R}(A)$ is linear, continuous and surjective between two Banach spaces. Then, by virtue of the *open mapping theorem* 1.3.1, given any $g \in \mathcal{R}(A)$, $g = Af$, the image by A of the unitary open sphere of centre f is an open set in $\mathcal{R}(A)$ containing g. Since A is compact, the closure of such an open set (which is still contained in $\mathcal{R}(A)$) is compact; thus, each element of $\mathcal{R}(A)$ has a compact neighbourhood in $\mathcal{R}(A)$. This means that $\mathcal{R}(A)$ is locally compact and then, being also a normed vector space, it is of finite dimension by virtue of a theorem by Riesz [49]. ■

Theorem 1.5.8. *Let X and Y be normed vector spaces (not necessarily complete); if $A \in \mathcal{B}(X,Y)$ is of finite range, then it is compact.*

Proof. If $V \subset X$ is bounded, i.e. there exists $C < \infty$ such that $\|x\|_X \leq C \ \forall x \in V$, then $\|Ax\|_Y \leq \|A\| \, \|x\|_X \leq C\|A\| \ \forall x \in V$, i.e. $A(V)$ is bounded. Moreover, since $\dim\mathcal{R}(A) = n < \infty$, $\mathcal{R}(A)$ is closed; hence $\overline{A(V)}$ is a closed and bounded subset of the normed vector space of finite dimension $\mathcal{R}(A)$, which, as such, is always homeomorphic to an $\mathbb{R}^n$ (in our case, we obviously have $n = \dim\mathcal{R}(A)$). By virtue of Heine-Borel's theorem, $\overline{A(V)}$ is compact; hence, A is compact. ■

Also for future purpose, we are now going to show that if X, Y are Hilbert spaces and $A : X \to Y$ is compact, the generalized solution, if it exists, can be written explicitly in terms of the datum g and the *singular system* of the operator A.

Let us briefly recall that if A is compact, then the operators A^*A and AA^* are compact, self-adjoint and positive[11]. They also have the same positive eigenvalues with the same multiplicity. Let σ_k^2 be these eigenvalues, ordered in a decreasing sequence ($\sigma_0^2 \geq \sigma_1^2 \geq \sigma_2^2 \geq \dots$): except in

[11] We remind that a linear operator $T : X \to X$, with X a Hilbert space, is called *positive* if $(x, Tx)_X \geq 0$ $\forall x \in X$; T is called *strictly positive* if $(x, Tx)_X > 0$ $\forall x \in X$ with $x \neq 0$. It is possible to prove that if a linear and continuous operator is positive, then it is also self-adjoint.

degenerate cases where they are in finite number, the sequence $\{\sigma_k^2\}_{k=0}^{\infty}$ tends to zero when n tends to infinity. It can be shown that it is always possible to find vector sets $\{u_k\}_{k=0}^{\infty} \subset X$ and $\{v_k\}_{k=0}^{\infty} \subset Y$ such that:

1. denoting with σ_k the positive square root of σ_k^2, it holds:

$$Au_k = \sigma_k v_k, \qquad A^* v_k = \sigma_k u_k; \tag{1.44}$$

2. the set $\{u_k\}_{k=0}^{\infty} \subset X$ form an orthonormal basis in $\mathcal{N}(A)^{\perp}$.

Furthermore, it trivially turns out that $\{u_k\}_{k=0}^{\infty} \subset X$ are all the eigenvectors (in $\mathcal{N}(A^*A)^{\perp} = \mathcal{N}(A)^{\perp}$) of the operator A^*A, while $\{v_k\}_{k=0}^{\infty} \subset Y$ are all the eigenvectors (in $\mathcal{N}(AA^*)^{\perp} = \mathcal{N}(A^*)^{\perp}$) of AA^* and form an orthonormal basis in $\overline{\mathcal{R}(A)}$.

Let us now observe that any $f \in X$ can be univocally decomposed as

$$f = f_1 + f_2, \qquad \text{with} \quad f_1 \in \mathcal{N}(A)^{\perp}, \ f_2 \in \mathcal{N}(A), \tag{1.45}$$

so that $Af = Af_1$. Since it obviously holds

$$f_1 = \sum_{k=0}^{\infty} (f, u_k)_X \, u_k, \tag{1.46}$$

by means of the continuity and linearity of A and of the first of relations (1.44), we get

$$Af_1 = A\left(\lim_{n\to\infty} \sum_{k=0}^{n} (f, u_k)_X \, u_k \right) = \lim_{n\to\infty} \sum_{k=0}^{n} (f, u_k)_X Au_k = \sum_{k=0}^{\infty} (f, u_k)_X \, \sigma_k v_k. \tag{1.47}$$

Summing up, we have found the following relation

$$Af = \sum_{k=0}^{\infty} \sigma_k (f, u_k)_X \, v_k \qquad \forall f \in X, \tag{1.48}$$

and in an analogous way we can get one for A^*, i.e.

$$A^* g = \sum_{k=0}^{\infty} \sigma_k (g, v_k)_Y \, u_k \qquad \forall g \in Y. \tag{1.49}$$

Definition 1.5.6. *The positive numbers σ_k and the vectors u_k, v_k, for $k \in \mathbb{N}$, are respectively called the* singular values *and the* singular vectors (functions) *of the compact operator A. The set of triples $\{\sigma_k, u_k, v_k\}_{k=0}^{\infty}$ is said the* singular system *of A; the representation (1.48) [(1.49)] is named the* singular representation *of A [A^*].*

Remark 1.5.3. The singular representations (1.48) and (1.49) easily imply that $\|A\| = \sigma_0$ and $\|A^*\| = \sigma_0$ respectively. Let us prove this result for A (the argument for A^* is obviously the same). To this end, let $f \in X$ be such that $\|f\|_X = 1$: then, by virtue of (1.48), we have

$$\|Af\|_Y = \left\| \sum_{k=0}^{\infty} \sigma_k (f, u_k)_X \, v_k \right\|_Y = \sqrt{\sum_{k=0}^{\infty} \sigma_k^2 \left| (f, u_k)_X \right|^2} \leq \sigma_0 \sqrt{\sum_{k=0}^{\infty} \left| (f, u_k)_X \right|^2} \leq \sigma_0, \tag{1.50}$$

where equality can hold (it suffices to take, e.g., $f = u_0$). Hence, remembering definition[12] $\|A\| := \sup_{\|f\|_X = 1} \|Af\|_Y$, from inequality (1.50) we immediately get $\|A\| = \sigma_0$. □

Let us now consider the so-called *Euler equation*, which is exactly the third condition that characterizes pseudosolutions in theorem (1.5.1), i.e.

$$A^* A f = A^* g. \tag{1.51}$$

By inserting in both sides of the previous equation the singular representations (1.48) and (1.49), we get

$$\sigma_j^2 (f, u_j)_X = \sigma_j (g, v_j)_Y \qquad \forall j \in \mathbb{N}, \tag{1.52}$$

that is

$$(f, u_j)_X = \frac{1}{\sigma_j} (g, v_j)_Y \qquad \forall j \in \mathbb{N}. \tag{1.53}$$

It easily follows that necessary and sufficient condition for the existence of pseudosolutions, i.e. solutions of equation (1.51), is that

$$\sum_{k=0}^{\infty} \frac{1}{\sigma_k^2} \left| (g, v_k)_Y \right|^2 < \infty, \tag{1.54}$$

which is called *Picard's condition* and is basically posed on the datum g. If the Picard's condition holds, it is immediate to realize that the following series

$$\sum_{k=0}^{\infty} \frac{1}{\sigma_k} (g, v_k)_Y \, u_k \tag{1.55}$$

converges to a pseudosolution, even better to the generalized solution $f^\dagger$, since all the vectors u_k are in $\mathcal{N}(A)^\perp$. Summing up, we have found the explicit representation

$$A^\dagger g = f^\dagger = \sum_{k=0}^{\infty} \frac{1}{\sigma_k} (g, v_k)_Y \, u_k. \tag{1.56}$$

[12] See remark 1.3.1, definition No 2.

Coming back to the case of a generic operator $A \in \mathcal{B}(X, Y)$ (also not compact), we conclude this section 1.5 noticing that, analogously to what observed in section 1.3 about inequality (1.9), the fact that $\mathcal{R}(A)$ is closed does not ensure the stability of the generalized solution. Indeed, it is possible to prove [10] the following inequality:

$$\frac{\|\delta f^\dagger\|_X}{\|f^\dagger\|_X} \leq C(A)\frac{\|\delta g\|_Y}{\|g\|_Y}, \tag{1.57}$$

where, this time, the condition number is given by

$$C(A) = \|A\|\|A^\dagger\|. \tag{1.58}$$

It can be shown [8] that, also in this case, it is always $C(A) \geq 1$. Obviously, if $C(A) \gg 1$ and the datum is affected by error, the generalized solution is numerically unstable. Hence, looking for the generalized solution of an inverse problem, instead of the solution itself, does not free us from the necessity of using *regularization algorithms*. Of course, this does not mean that the concept of generalized solution is useless; on the contrary, it is fundamental just in handling the regularization algorithms themselves.

1.6. Regularization theory: a general formulation

Given a linear inverse problem, if the generalized inverse operator is not continuous or the problem is characterized by a very large condition number, the knowledge of the generalized solution, if it exists, is nearly useless from the point of view of applications to real data. In such cases, indeed, as a consequence of any measurement operation, there is always an error on the datum; this error propagates on the generalized solution and makes it numerically unstable and then physically meaningless. In such a situation, some methods yielding stable estimates of the generalized solutions are needed. In scientific literature there exist various algorithms of this kind and their description is rigorously developed in the ambit of regularization theory (see, e.g., [33], [37], [6], [67], [52]). Here we shall only give some basic definitions and some hints about the fundamental techniques we are going to employ in the following chapters.

In general terms, *regularization* is the approximation of an ill-posed problem by a one-parameter (usually denoted with α) family of neighbouring well-posed problems. We now want to motivate the future definition of a *regularization operator* and of a *regularization method* by making some considerations about the fact that our aim is to approximate the generalized solution $f^\dagger = A^\dagger g$ of the usual (and exact) linear inverse problem

$$Af = g \tag{1.59}$$

in the most general case in which the error or the noise affect not only the exact datum g, but also the exact operator A, i.e. there may be also *modelling errors*, so that we don't know

neither g nor A, but only some approximations of them. In practice, we shall have to deal with a noisy version of the previous problem, i.e.

$$A_h f = g_\delta, \tag{1.60}$$

where g_δ represents a perturbed version of the datum and A_h is an approximate version of the operator; we shall work in a *deterministic* framework, i.e. we shall assume to know a priori the following noise bounds:

$$\|g_\delta - g\|_Y \leq \delta \tag{1.61}$$

and

$$\|A_h - A\| \leq h. \tag{1.62}$$

In the following, for notational convenience, we shall sometimes denote with $\eta := (\delta, h)$ the two noise levels together.

Remark 1.6.1. Also for future purpose, it is useful to observe that, in any case, for each exact datum g, its noisy version g_δ can be represented as

$$g_\delta = g + w_\delta, \tag{1.63}$$

where w_δ is the *noise function* and is such that $\|w_\delta\|_Y \leq \delta$.

Analogously, for each exact operator A, its noisy version A_h can be represented as

$$A_h = A + N_h \tag{1.64}$$

where N_h is the *noise operator* and is such that $\|N_h\| \leq h$.

Expressions (1.63) and (1.64) may not be explicitly known, but it is always possible to assume that they exist. In fact they are quite general and do not mean that the noise is necessarily additive, since w_δ and N_h may respectively depend on g and A. □

Remark 1.6.2. If the exact equation (1.59) falls within the mathematical model of a physical phenomenon, then one has necessarily that $g \in \mathcal{R}(A)$ (otherwise the model would be inconsistent, giving rise to an impossible equation). In such a case, since $\mathcal{R}(A) \subset D(A^\dagger)$, we have that $g = Af^\dagger$ and consequently expression (1.63) can be rewritten as

$$g_\delta = Af^\dagger + w_\delta. \tag{1.65}$$

However, in some situations, the exact equation is or may be impossible, i.e. one may have $g \notin \mathcal{R}(A)$: as we shall see, this is, in general, just the case of the linear sampling method, whose basic equation, although physically interpretable as a focusing condition [22], does not actually model any physical phenomenon. Obviously, in such a circumstance representation (1.65) may not hold. □

Now, we observe that, in general, the solution of equation (1.60) does not exist, since the random components of the noise may carry the datum outside the range of A_h (or of A, in the exact operator case $h = 0$). Furthermore, in general, the kernel of A_h is not empty and there is not continuous dependence of the solution on the data. In this ill-posed case, since $A_h^\dagger$ (or $A^\dagger$ if $h = 0$) is in general unbounded, $A_h^\dagger g_\delta$ (or $A^\dagger g_\delta$) is certainly not a good approximation of $A^\dagger g$ even if it exists (which will, in general, also not be the case, since $D(A_h^\dagger)$ (or $D(A^\dagger)$) is in general a proper subset of Y). Consequently, we have to look for some approximation, say f_α^η, of $f^\dagger$ which does, on the one hand, depend continuously on the noisy data g_δ and on the approximate operators A_h, so that it can be computed in a stable way, and has, on the other hand, the property that as the noise levels δ and h decrease to zero and the *regularization parameter* α is chosen appropriately (whatever this means), then f_α^η tends to $f^\dagger$. The construction of f_α^η will in general involve the operator A_h (or A if $h = 0$); although this seems to be a trivial remark, there are situations where this is not necessarily the case: if $\|g_\delta\|_Y \leq \delta$, i.e. if the noise level is larger than or equal to the signal, one might be best off just to take $f_\alpha^\eta := 0$, independently of the operator A_h (or A), since in such a situation the noisy datum contains no information at all anyway. But except in this case, the operator A_h (or A) certainly has to play some role in the construction of f_α^η. So it is convenient to split any regularization method into two logical steps:

1. we regularize the (in general, unbounded) generalized inverse operator $A^\dagger$ on $D(A^\dagger)$ by replacing it with a one-parameter-depending family $\{R_\alpha^{(h)}\}_{\alpha>0}$ of continuous operators[13] defined on all Y. At this level, the regularization parameter α is completely free; furthermore, in the construction of each $R_\alpha^{(h)}$, the operator A_h (or A if $h = 0$) plays an important role, while the datum g_δ has no one. In this sense, we are considering (1.59) as a collection of equations, one for each $g \in D(A^\dagger)$;

2. then, we consider the particular equation we have to deal with, i.e. a certain $g \in D(A^\dagger)$, and as approximation of its generalized solution $f^\dagger$ we take $f_{\alpha^*}^\eta := R_{\alpha^*}^{(h)}(g_\delta)$, where we are now regarding α as a function α^* of δ, h, g_δ, A_h: such a function $\alpha^*(\delta, h, g_\delta, A_h)$ will be called *parameter choice rule* and two requirements for it are that if the noise levels δ, h tend to zero, then the *regularized solution* $f_{\alpha^*}^\eta$ tends to $f^\dagger$ and α^* itself tends to zero. We point out that $f_{\alpha^*}^\eta$ should be computable in a stable way (at least in principle, since $R_\alpha^{(h)}$ is assumed to be continuous[14] $\forall \alpha > 0$).

[13] The apex (h) in notation $\{R_\alpha^{(h)}\}_{\alpha>0}$ points out that each operator $R_\alpha^{(h)}$ is constructed, in general, by means of an explicit use of the noisy operator A_h. In the following, if $h = 0$, we shall often write $\{R_\alpha\}_{\alpha>0}$ or R_α instead of, respectively, $\{R_\alpha^{(0)}\}_{\alpha>0}$ or $R_\alpha^{(0)}$, except in some specific cases, in which such a shorthand may be misleading.

[14] Obviously, a priori there might be a problem of ill-conditioning; however, if a regularization method were ill-conditioned, there would be no sense in using it.

Summing up, we shall define regularization operators for the whole collection of equations (1.59), with $g \in D(A^\dagger)$, but parameter choice rules for a specific equation out of this collection. Both together then form a regularization method for solving one specific equation. These considerations lead to the following definition.

Definition 1.6.1. *Given $A \in \mathcal{B}(X,Y)$ and $\alpha_0 \in (0,+\infty]$, for every $\alpha \in (0,\alpha_0)$ let*

$$R_\alpha^{(h)} : Y \to X \tag{1.66}$$

be a continuous (not necessarily linear) operator, constructed by means of a noisy version $A_h \in \mathcal{B}(X,Y)$ of the exact operator A. The family $\{R_\alpha^{(h)}\}_{\alpha>0}$ is called a regularization *or a* regularization operator *(for $A^\dagger$) if, for each $g \in D(A^\dagger)$, there exists a* parameter choice rule

$$\begin{aligned} \alpha^* : \mathbb{R}^+ \times \mathbb{R}^+ \times Y \times \mathcal{B}(X,Y) &\to (0,\alpha_0) \\ (\delta, h, g_\delta, A_h) &\mapsto \alpha^*(\delta, h, g_\delta, A_h) \end{aligned} \tag{1.67}$$

such that the two following conditions hold:

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta, A_h} \left\{ \left\| R^{(h)}_{\alpha^*(\delta,h,g_\delta,A_h)} g_\delta - A^\dagger g \right\|_X \;\middle|\; \|g_\delta - g\|_Y \le \delta;\ \|A_h - A\| \le h \right\} = 0 \tag{1.68}$$

and

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta, A_h} \{\alpha^*(\delta,h,g_\delta,A_h) \mid \|g_\delta - g\|_Y \le \delta;\ \|A_h - A\| \le h\} = 0. \tag{1.69}$$

Finally, when considering a certain $g \in D(A^\dagger)$ and a specific parameter choice α^*, the pair $(R_{\alpha^*}^{(h)}, \alpha^*)$ is called a (convergent) regularization method *or* regularization algorithm *for solving $Af = g$ if conditions (1.68) and (1.69) hold for that g and for that α^*.*

Thus, a regularization method consists of a regularization operator and a parameter choice rule which together form a convergent algorithm in the sense that, if the regularization parameter is chosen according to that rule, then the regularized solution converges (in the norm) to the generalized one as the noise levels tend to zero; this is assured for any collection of noisy data and noisy operators compatible with their respective noise levels, then it is a "worst case" concept of convergence.

Remark 1.6.3. The parameter choice rule $\alpha^* = \alpha^*(\delta, h, g_\delta, A_h)$ depends (so far) explicitly on the noise levels δ, h, on the noisy datum g_δ and on the perturbed operator A_h. However, it is useful to remember that we defined it for every specific $g \in D(A^\dagger)$, so that α^* depends also on the exact datum g. Since g is not known, this dependence can only be on some qualitative a priori knowledge about g like smoothness properties. Analogously, it is evident from condition (1.68) that α^* (qualitatively) depends also on the exact operator A (or at least on its generalized inverse $A^\dagger$) which, in general, is unknown too. We might have denoted these implicit dependencies by writing $\alpha^* = \alpha^*_{g,A}(\delta, h, g_\delta, A_h)$, but we have avoided it not to make too heavy our notations. □

Remark 1.6.4. Let us note that in definition 1.6.1 we did not require the regularization operator $\{R_\alpha^{(h)}\}_{\alpha>0}$ to be a family of *linear operators*. If the $R_\alpha^{(h)}$ are linear, then we call the corresponding method a *linear regularization method*, and the family $\{R_\alpha^{(h)}\}_{\alpha>0}$ a *linear regularization operator*. However, it also makes sense to consider nonlinear regularization methods for solving linear problems, like the method of conjugate gradient. □

Remark 1.6.5. Since each operator $R_\alpha^{(h)}$ is in general constructed by using explicitly the noisy operator A_h (or A if $h = 0$), it is not useless to specify that when we write (as in (1.68)) $R^{(h)}_{\alpha^*(\delta,h,g_\delta,A_h)}$, we intend that the operator A_h appearing as an argument of α^* is exactly the same employed to form $R_\alpha^{(h)}$ for any $\alpha > 0$. □

Remark 1.6.6. Although the regularization parameter α is typically a real positive number, it may also be a natural number N: this happens when dealing with iterative regularization algorithms. In such a case, the previous definition needs trivial changes: for example, condition (1.69) becomes:

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta,A_h} \{N^*(\delta,h,g_\delta,A_h) \mid \|g_\delta - g\|_Y \le \delta;\ \|A_h - A\| \le h\} = \infty. \tag{1.70}$$

In subsection 1.7.1 we shall see TSVD as an example of iterative regularization algorithm. □

Definition 1.6.2. *Let $\alpha^* = \alpha^*(\delta, h, g_\delta, A_h)$ be a parameter choice rule. If there is dependence neither on g_δ , nor on A_h, but only on δ and h, then we denote such a rule with $\alpha^*(\delta, h)$ and we call it an* a priori parameter choice rule. *Otherwise, we call α^* an* a posteriori parameter choice rule.

Thus, an a priori parameter choice rule depends only on the noise levels, not on the actual data or the perturbed operator and, consequently, not on results obtained during the actual computation, like the so-called *residuals*, defined as $\left\|A_h R^{(h)}_{\alpha^*} g_\delta - g_\delta\right\|_Y$; such a rule may be devised before the actual calculation, whence the name "a priori parameter choice rule".

One could also think of parameter choice rules that depend *only* on g_δ, A_h and not on the noise levels δ or h. However, the following theorem due to Bakushinskii [5] shows that, for an ill-posed problem, such rules cannot be part of a regularization algorithm satisfying definition 1.6.1.

Theorem 1.6.1. *Given $A \in \mathcal{B}(X, Y)$, let us assume that there exists a regularization $\{R_\alpha^{(h)}\}_{\alpha>0}$ for $A^\dagger$ with a parameter choice rule α^* which depends on g_δ, A_h and not on δ or h, such that the regularization method $(R^{(h)}_{\alpha^*}, \alpha^*)$ is convergent for every $g \in D(A^\dagger)$. Then $A^\dagger$ is continuous (and $D(A^\dagger) = Y$).*

Proof. If α^* is independent of δ and h, i.e. $\alpha^* = \alpha^*(g_\delta, A_h)$, then it follows from (1.68) that, for each $g \in D(A^\dagger)$, we have

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta,A_h} \left\{ \left\| R^{(h)}_{\alpha^*(g_\delta,A_h)} g_\delta - A^\dagger g \right\|_X \mid \|g_\delta - g\|_Y \le \delta;\ \|A_h - A\| \le h \right\} = 0. \tag{1.71}$$

Hence, we easily get

$$\left\| R^{(0)}_{\alpha^*(g,A)} g - A^\dagger g \right\|_X = 0 \qquad \forall g \in D(A^\dagger); \tag{1.72}$$

indeed, it suffices to observe that the left-hand side of the previous equation (1.72) is completely independent[15] of δ and h, so that, if by absurd it were $s := \left\| R^{(0)}_{\alpha^*(g,A)} g - A^\dagger g \right\|_X > 0$ for some $g \in D(A^\dagger)$, also the supremum internal to the limit in (1.71) would be strictly positive and not less than $s > 0$ for every δ and h; thus the limit itself should be greater or equal to s, against (1.71). Hence, by virtue of (1.72) and (1.71), for any sequence $\{g_n\}_{n=0}^{\infty}$ in $D(A^\dagger)$ which converges to a g in $D(A^\dagger)$ as $n \to \infty$, we have

$$A^\dagger g_n = R^{(0)}_{\alpha^*(g_n,A)} g_n \to A^\dagger g \qquad \text{as } n \to \infty, \tag{1.73}$$

so that $A^\dagger$ is continuous on $D(A^\dagger)$; but then, theorem 1.5.6 and relation (1.21) imply that $D(A^\dagger) = Y$. ■

Thus, if $A^\dagger$ is unbounded, no *error-free* parameter choice rule can yield a convergent regularization method. However, this is an asymptotic result, so it does not imply that error-free parameter choice rules cannot behave well for finite noise levels δ, h.

We can now ask the following questions:

1. How can one construct regularization operators?
2. How can one construct parameter choice rules that give rise to convergent regularization methods?
3. How can these steps be performed in some "optimal" way?

With regard to the third point, we shall not deal with it: for a treatment, see, for example, [33].

On the other hand, the following two theorems 1.6.2 and 1.6.3 give a preliminary answer to the first and second question in the particular, but very important, case in which the operator A is known exactly (i.e. $h = 0$).

Theorem 1.6.2. *If the operator A is known exactly and if there exists a family $\{R_\alpha\}_{\alpha>0}$ of continuous (possibly non-linear) operators such that*

$$\lim_{\alpha \to 0^+} \left\| R_\alpha g - A^\dagger g \right\|_X = 0 \qquad \forall g \in D(A^\dagger), \tag{1.74}$$

the family $\{R_\alpha\}_{\alpha>0}$ itself is a regularization for $A^\dagger$ and there exists, for every $g \in D(A^\dagger)$, an a priori parameter choice rule $\alpha^* = \alpha^*(\delta)$ such that (R_{α^*}, α^*) is a convergent regularization method for solving $Af = g$.

[15] Remember remark 1.6.5 and footnote 13.

Proof. Let $g \in D(A^\dagger)$ be arbitrary, but fixed. By assumption, there exists an increasing monotonic function $\sigma : \mathbb{R}^+ \to \mathbb{R}^+$, with $\lim_{\varepsilon \to 0^+} \sigma(\varepsilon) = 0$, such that, for every $\varepsilon > 0$, it holds

$$\left\| R_{\sigma(\varepsilon)} g - A^\dagger g \right\|_X \leq \frac{\varepsilon}{2}. \tag{1.75}$$

Furthermore, since each $R_{\sigma(\varepsilon)}$ is continuous, for every $\varepsilon > 0$ there exists a $\rho(\varepsilon)$ such that, if $\|z - g\|_Y \leq \rho(\varepsilon)$, then

$$\left\| R_{\sigma(\varepsilon)} z - R_{\sigma(\varepsilon)} g \right\|_X \leq \frac{\varepsilon}{2}. \tag{1.76}$$

The previous property (1.76) enables us to consider $\rho(\varepsilon)$ as a function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ (which we can assume, without loss of generality, to be strictly increasing monotonic and continuous) endowed with the property that $\lim_{\varepsilon \to 0^+} \rho(\varepsilon) = 0$. Hence, the inverse function ρ^{-1} exists on the range of ρ, is strictly increasing monotonic and continuous, and has the property that $\lim_{\delta \to 0^+} \rho^{-1}(\delta) = 0$. We can extend ρ^{-1} to all of $\mathbb{R}^+$ and thus define

$$\begin{aligned} \alpha^* : \mathbb{R}^+ &\to \mathbb{R}^+ \\ \delta &\mapsto \sigma(\rho^{-1}(\delta)). \end{aligned} \tag{1.77}$$

The function α^* is increasing monotonic and has the property that $\lim_{\delta \to 0^+} \alpha^*(\delta) = 0$. Furthermore, for each $\varepsilon > 0$, there is a $\delta > 0$, namely $\delta := \rho(\varepsilon)$, such that, if $\|g_\delta - g\|_Y \leq \delta$, then

$$\left\| R_{\alpha^*(\delta)} g_\delta - A^\dagger g \right\|_X \leq \left\| R_{\alpha^*(\delta)} g_\delta - R_{\alpha^*(\delta)} g \right\|_X + \left\| R_{\alpha^*(\delta)} g - A^\dagger g \right\|_X \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \tag{1.78}$$

having remembered relations (1.75), (1.76) and observed that $\alpha^*(\delta) = \sigma(\varepsilon)$. Thus, for the method (R_{α^*}, α^*), conditions (1.68) and (1.69) trivially hold and the function $\alpha^* = \alpha^*(\delta)$ defines an a priori parameter choice rule. ■

In the previous theorem 1.6.2 we have seen that, if (1.74) holds, then there exists an a priori parameter choice rule $\alpha^* = \alpha^*(\delta)$ such that (R_{α^*}, α^*) is a convergent regularization method. Such parameter choice rules, in the particular but very frequent case of linear regularization, can be characterized as follows.

Theorem 1.6.3. *Let $\{R_\alpha\}_{\alpha > 0}$ be a family of linear and continuous operators such that*

$$\lim_{\alpha \to 0^+} \left\| R_\alpha g - A^\dagger g \right\|_X = 0 \qquad \forall g \in D(A^\dagger) \tag{1.79}$$

and, for each $g \in D(A^\dagger)$, let $\alpha^* : \mathbb{R}^+ \to (0, \alpha_0)$, $\alpha^* = \alpha^*(\delta)$ be an a priori parameter choice rule. Then (R_{α^*}, α^*) is a convergent regularization method if and only if

$$\lim_{\delta \to 0^+} \alpha^*(\delta) = 0 \tag{1.80}$$

and

$$\lim_{\delta \to 0^+} \delta \left\| R_{\alpha^*(\delta)} \right\| = 0. \tag{1.81}$$

Proof. First of all, since $\alpha^* = \alpha^*(\delta)$ is assumed to be an a priori parameter choice rule, (1.80) holds by definition in any case.
"$\Leftarrow$": let us assume that (1.80) and (1.81) hold. For each $g \in D(A^\dagger)$ and for any $g_\delta \in Y$ such that $\|g_\delta - g\|_Y \leq \delta$, we can easily get

$$\begin{aligned}\left\|R_{\alpha^*(\delta)}g_\delta - A^\dagger g\right\|_X &\leq \left\|R_{\alpha^*(\delta)}g - A^\dagger g\right\|_X + \left\|R_{\alpha^*(\delta)}g_\delta - R_{\alpha^*(\delta)}g\right\|_X \leq \\ &\leq \left\|R_{\alpha^*(\delta)}g - A^\dagger g\right\|_X + \left\|R_{\alpha^*(\delta)}\right\|\delta, \end{aligned} \tag{1.82}$$

where the last member does not depend on g_δ. Hence, we get

$$\sup_{g_\delta}\left\{\left\|R_{\alpha^*(\delta)}g_\delta - A^\dagger g\right\|_X \;|\; \|g_\delta - g\| \leq \delta\right\} \leq \left\|R_{\alpha^*(\delta)}g - A^\dagger g\right\|_X + \delta\left\|R_{\alpha^*(\delta)}\right\|. \tag{1.83}$$

By virtue of (1.79), (1.80) and (1.81), inequality (1.83) implies relation (1.68).
"$\Rightarrow$": let us assume, by absurd, that (1.81) does not hold (we have already noticed that (1.80) holds by definition). Then, there exist an $\varepsilon > 0$ and a sequence $\{\delta_n(\varepsilon)\}_{n=0}^\infty \equiv \{\delta_n\}_{n=0}^\infty$ such that $\lim_{n\to\infty}\delta_n = 0$ and $\delta_n\|R_{\alpha^*(\delta_n)}\| > \varepsilon \;\; \forall n \in \mathbb{N}$. Hence, there is a sequence $\{z_n\}_{n=0}^\infty$ in Y, with $\|z_n\|_Y = 1 \; \forall n \in \mathbb{N}$, such that $\delta_n\|R_{\alpha^*(\delta_n)}z_n\|_X \geq \varepsilon/2 \;\; \forall n \in \mathbb{N}$. Thus, for any $g \in D(A^\dagger)$ and for any $g_n \in Y$ of the form $g_n := g + \delta_n z_n$, so that $\|g_n - g\| \leq \delta_n$, let us consider the element of X given by:

$$R_{\alpha^*(\delta_n)}g_n - A^\dagger g = (R_{\alpha^*(\delta_n)}g - A^\dagger g) + \delta_n R_{\alpha^*(\delta_n)}z_n. \tag{1.84}$$

It immediately follows that

$$\left\|R_{\alpha^*(\delta_n)}g_n - A^\dagger g\right\|_X \geq \left|\left\|(R_{\alpha^*(\delta_n)}g - A^\dagger g)\right\|_X - \delta_n\left\|R_{\alpha^*(\delta_n)}z_n\right\|_X\right| \qquad \forall n \in \mathbb{N}. \tag{1.85}$$

The second term in the right-hand side of previous inequality, as stated before, remains not less than $\varepsilon/2$, while the first one tends to zero as $n \to \infty$ by virtue of (1.80) and (1.79). Hence, it is impossible that

$$\lim_{n\to\infty}\sup_{g_n}\left\{\left\|R_{\alpha^*(\delta_n)}g_n - A^\dagger g\right\|_X \;|\; \|g_n - g\| \leq \delta_n\right\} = 0; \tag{1.86}$$

this clearly implies that condition (1.68) cannot be verified, so that (R_{α^*}, α^*) cannot be a convergent regularization method, against the hypothesis. ■

Remark 1.6.7. Relation (1.82) represents a basic inequality in linear regularization theory and deserves a comment. First of all, let us rewrite it, considering, for a moment, α as a free parameter:

$$\left\|R_\alpha g_\delta - A^\dagger g\right\|_X \leq \left\|R_\alpha g - A^\dagger g\right\|_X + \|R_\alpha\|\,\delta. \tag{1.87}$$

The first term at the right-hand side represents the approximation error due to the use of R_α instead of the generalized inverse operator; by virtue of equation (1.79), it tends to zero when

$\alpha \to 0^+$. On the other hand, the second term is an estimate of the error on the regularized solution $R_\alpha g_\delta$ due to the presence of noise on the datum and it grows up to a very large number or to infinity[16] when $\alpha \to 0^+$, since the bounded operators R_α are, as $\alpha \to 0^+$, more and more accurate approximations of the operator $A^\dagger$, which has, in general, very big norm or is unbounded.

Therefore it is necessary to find a compromise between approximation and error magnification. If we assume that the two terms at the right-hand side of inequality (1.87) are monotonic functions of α (this condition is satisfied by all the regularizing algorithms used in practice) and, more precisely, that the first one is an increasing function, while the second one a decreasing function of α, it turns out that there exists a unique value of α which minimizes the (g_δ-independent) right-hand side of inequality (1.87) and represents the optimal compromise between accuracy and stability: we shall denote such a value of α with $\alpha_{\text{opt}}(\delta)$, since it is optimal and depends, in general, on the noise level δ affecting the datum g_δ.

Knowing that such an optimal value of α exists does not imply, however, that it is easily determinable. On the contrary, estimating $\alpha_{\text{opt}}(\delta)$ is, in general, one of the hardest tasks in regularization theory: the main reason for this difficulty is that the calculation of $\alpha_{\text{opt}}(\delta)$ requires the knowledge of $f^\dagger$, and this is impossible if we know only the noisy datum g_δ and not the exact one, i.e. g. On the other hand, whenever an algorithm depending on a parameter is employed, one has to give a rule in order to fix a suitable value of the parameter itself. Hence, we shall have to be satisfied with a criterion verifying some basic properties and by means of which we can choose a suitable α (depending in general both on δ and g_δ, and denoted with $\alpha^*(\delta, g_\delta)$) which is (hopefully) not too far from the optimal value $\alpha_{\text{opt}}(\delta)$. As already established, we call such a criterion a *parameter choice rule.* □

As regards a possible converse of theorem 1.6.2, we can say that it actually holds in the following sense: if $(R^{(h)}_{\alpha^*}, \alpha^*)$ is a convergent regularization method[17], then from relation (1.68) we easily get

$$\lim_{(\delta,h)\to(0^+,0^+)} \left\| R^{(0)}_{\alpha^*(\delta,g,h,A)} g - A^\dagger g \right\|_X = 0 \qquad \forall g \in D(A^\dagger). \tag{1.88}$$

If α^* is continuous in δ and h, this implies that

$$\lim_{\sigma\to 0^+} \left\| R^{(0)}_\sigma g - A^\dagger g \right\|_X = 0 \qquad \forall g \in D(A^\dagger); \tag{1.89}$$

otherwise, this holds only over the set of σ-values which are in the range of $\alpha^*(\cdot, \cdot, g, A)$. Summing up, when evaluated for $h = 0$, regularizations are, substantially, pointwise approximations of the generalized inverse $A^\dagger$ in $D(A^\dagger)$.

[16] For more precise statements about $\|R_\alpha\|$, see theorem 1.6.6 in the following.

[17] Here we can admit again that the regularization operator depends on A_h; as regards limit (1.88), remember remark 1.6.5.

We are now going to show that such pointwise approximations of $A^\dagger$ in $D(A^\dagger)$ cannot be, in general, approximations in the operator norm. In order to do this, we recall some theorems that are well-known in functional analysis and concern the so-called *Principle of Uniform Boundedness*. We shall not demonstrate them; for their proofs, see, for example, [32], part I.

Theorem 1.6.4. *Let X, Y be Banach spaces, and let $\{T_n\}_{n=0}^{\infty}$ be a sequence of bounded linear operators on X to Y. Then the limit $Tx = \lim_{n\to\infty} T_n x$ exists for every x in X if and only if*

1. *the limit Tx exists for every x in a fundamental set*[18]*, and*
2. *for each x in X the supremum $\sup_n \|T_n x\|_Y < \infty$.*

When the limit Tx exists for each x in X, the operator T is linear, bounded, and

$$\|T\| \leq \liminf_{n\to\infty} \|T_n\| \leq \sup_n \|T_n\| < \infty. \tag{1.90}$$

Theorem 1.6.5. *Let X, Y be Banach spaces, and $\{T_\beta\}_{\beta\in I}$ a family (not necessarily countable) of bounded linear operators on X to Y. Then the following statements are equivalent:*

1. $\sup_{\beta\in I} \|T_\beta\| < \infty$;
2. $\sup_{\beta\in I} \|T_\beta\, x\|_Y < \infty \qquad \forall x \in X.$

Now, let us suppose that the regularization $\{R_\alpha^{(0)}\}_{\alpha>0} \equiv \{R_\alpha\}_{\alpha>0}$ is linear and uniformly bounded, i.e. $\sup_{\alpha\in(0,\alpha_0)} \|R_\alpha\| < \infty$; by theorem 1.6.5, this is equivalent to

$$\sup_{\alpha\in(0,\alpha_0)} \|R_\alpha g\|_X < \infty \qquad \forall g \in Y. \tag{1.91}$$

It turns out that if $\mathcal{R}(A)$ is non-closed, it is impossible that $\lim_{\alpha\to 0^+} \left\|R_\alpha - A^\dagger\right\| = 0$ (regarding R_α and $A^\dagger$ as defined on the normed vector space $D(A^\dagger)$). Indeed, if, by absurd, it were so, condition (1.74) would hold too, since convergence in operatorial norm implies strong convergence; then, in particular, it should exist a sequence $\{\alpha_n\}_{n=0}^{\infty} \subset \mathbb{R}^+$ such that $\lim_{n\to\infty} \alpha_n = 0^+$ and

$$\lim_{n\to\infty} \left\|R_{\alpha_n} g - A^\dagger g\right\|_X = 0 \qquad \forall g \in D(A^\dagger). \tag{1.92}$$

But $D(A^\dagger)$ is fundamental in Y (it is even dense), thus, recalling (1.91), which clearly implies

$$\sup_{\alpha_n\in(0,\alpha_0)} \|R_{\alpha_n} g\|_X < \infty \qquad \forall g \in Y, \tag{1.93}$$

[18] Let us briefly recall the concept of *fundamental set*. First of all, the subspace spanned by a set B in a linear space X will be denoted by $sp(B)$ and its closure by $\overline{sp(B)}$. Then, if $\overline{sp(B)} = X$, the set B is called *fundamental*.

and applying theorem 1.6.4, it would exist a linear bounded operator $\hat{A}^\dagger : Y \to X$ that extends $A^\dagger$. This is a contradiction, since, by theorem 1.5.6, $A^\dagger$ is not bounded owing to the non-closedness of $\mathcal{R}(A)$.

By means of analogous arguments, we can now state the theorem cited above in footnote 16.

Theorem 1.6.6. *If condition (1.74) holds, regularization $\{R_\alpha\}_{\alpha>0}$ is linear and $\mathcal{R}(A)$ is non-closed, then it holds*

$$\lim_{\alpha\to 0^+} \|R_\alpha\| = \infty. \tag{1.94}$$

Proof. Indeed, let us suppose, by absurd, the contrary: then, it would exist a sequence $\{\alpha_n\}_{n=0}^{\infty} \subset \mathbb{R}^+$ such that $\lim_{n\to\infty} \alpha_n = 0^+$ and $\sup_{n\in\mathbb{N}} \|R_{\alpha_n}\| < \infty$, i.e., recalling theorem 1.6.5, such that (1.93) holds. On the other hand, by virtue of (1.74), also (1.92) holds. Hence, since (1.93) and (1.92) hold together again, we can proceed and obtain the same contradiction as before. ■

Moreover, if (1.74) and (1.94) hold for the linear regularization $\{R_\alpha\}_{\alpha>0}$, then, by applying theorem 1.6.5, it easily follows that

$$\exists\, g \in Y \mid \lim_{\alpha\to 0^+} \|R_\alpha g\|_X = \infty. \tag{1.95}$$

Such a g cannot obviously belong to $D(A^\dagger)$, since (1.74) holds; hence, the elements g satisfying (1.95) belong to a set $W \subset Y \setminus D(A^\dagger)$. It turns out that under a (reasonable) additional condition, the set W is exactly the complement of $D(A^\dagger)$ in Y; in fact, the following theorem holds.

Theorem 1.6.7. *Let $\{R_\alpha\}_{\alpha>0}$ be a linear regularization. Then*

$$\lim_{\sigma\to 0^+} \left\|R_\sigma g - A^\dagger g\right\|_X = 0 \qquad \forall g \in D(A^\dagger). \tag{1.96}$$

Moreover, if

$$\sup_{\alpha\in(0,\alpha_0)} \|AR_\alpha\| < \infty, \tag{1.97}$$

then

$$\lim_{\sigma\to 0^+} \|R_\sigma g\|_X = \infty \qquad \forall g \in Y \setminus D(A^\dagger). \tag{1.98}$$

Here, limits (1.96) and (1.98) are to be understood as explained about limit (1.89).

Proof. Equation (1.96) follows exactly as limit (1.89): this means (in the sense described there) that $R_\sigma \to A^\dagger$ pointwise on $D(A^\dagger)$, so that $AR_\sigma \to AA^\dagger$ pointwise on the dense set $D(A^\dagger)$; moreover, we have already seen (cf. relation (1.33)) that $AA^\dagger = P|_{D(A^\dagger)}$, where P is the usual orthogonal projector P onto $\overline{\mathcal{R}(A)}$. Since, by assumption, $\|AR_\alpha\|$ is uniformly bounded, it

follows, from theorems 1.6.5 and 1.6.4, that it is possible to extend continuously the operator $AA^\dagger$ to a continuous operator defined on all Y, i.e. P itself, such that $AR_\sigma \to P$ pointwise on all Y.

Now, for a certain $g \in Y$, let us define, in general, $\sigma_n := \alpha^*(\delta_n, h_n, g, A)$ and assume that there is a sequence $\{\sigma_n\}_{n=0}^\infty \subset \mathbb{R}^+$, with $\lim_{n\to\infty} \sigma_n = 0^+$, such that the set $\{R_{\sigma_n} g\}_{n=0}^\infty$ is bounded; then [13] there exists a subsequence $\{R_{\sigma_{n(k)}} g\}_{k=0}^\infty$ which converges weakly to an $x \in X$ as $k \to \infty$. Since A is also weakly sequentially continuous[19], it holds $AR_{\sigma_{n(k)}} g \rightharpoonup Ax$. On the other hand, we have $AR_{\sigma_{n(k)}} g \to Pg$, so that $Ax = Pg$. Hence, by theorem 1.5.1 and definition 1.5.2, $g \in D(A^\dagger)$. Thus, if $g \notin D(A^\dagger)$, no bounded sequence $\{\|R_{\sigma_n} g\|_X\}_{n=0}^\infty$ can exist; hence (1.98) holds. ■

1.7. Regularization algorithms

1.7.1. TSVD (case of exact operator)

Let us consider the case of a *compact* operator A, which we assume to know exactly: if we remember representation (1.56) for the generalized solution and observe that, instead of knowing the exact datum g (which is supposed belonging to $D(A^\dagger)$), we only have at disposal a noisy version g_δ of it, we would be tempted to write

$$f^\dagger = \sum_{k=0}^{\infty} \frac{1}{\sigma_k} (g_\delta, v_k)_Y \, u_k. \tag{1.99}$$

Nevertheless, as previously stated, when g_δ does not satisfy the Picard's condition (1.54) (which typically happens, owing to the stochastic nature of the noise affecting the measured data), this expansion has only a formal meaning. However, from equation (1.99) it is rather natural to introduce the one-parameter family of functions in X given by

$$f_N := \sum_{k=0}^{N} \frac{1}{\sigma_k} (g_\delta, v_k)_Y \, u_k, \qquad N < \infty. \tag{1.100}$$

If we now insert (1.63) into (1.100), observe that $(g, v_k)_Y = (Pg, v_k)_Y \ \forall k = 0, \ldots, N$ (since $v_k \in \mathcal{R}(A) \ \forall k \in \mathbb{N}$, see (1.44)), remember that $Pg = Af^\dagger$ (see condition (i) of theorem 1.5.1), use the definition of adjoint operator and employ the second of relations (1.44), we obtain

$$f_N = \sum_{k=0}^{N} (f^\dagger, u_k)_X \, u_k + \sum_{k=0}^{N} \frac{1}{\sigma_k} (w_\delta, v_k)_Y \, u_k. \tag{1.101}$$

[19] See, for example, [13], p. 58.

It is clear that for increasing N the first term at the right-hand side of equation (1.101) converges to the generalized solution, while the second term, in general, grows up (or, in case, blows up) inducing an overwhelming effect of the noise on the solution, due to the behaviour of the singular values at high k (i.e. to the fact that $\{\sigma_k\}_{k=0}^{\infty}$ is a (in general, not strictly) decreasing monotonic sequence that vanishes as $k \to \infty$). However, the next theorem and the following remark show that it is possible to choose a finite and suitable $N^* = N^*(\delta)$ in such a way that f_{N^*} is an acceptable approximation of $f^\dagger$.

Theorem 1.7.1. *The one-parameter family $\{R_N\}_{N\geq 0}$ (with $N \in \mathbb{N}$) defined by*

$$R_N g = \sum_{k=0}^{N} \frac{1}{\sigma_k}(g, v_k)_Y \, u_k \qquad \forall g \in Y \tag{1.102}$$

defines a linear regularization for $A^\dagger$.

Proof. First of all, it is obvious that each R_N is linear and continuous. Moreover, as already observed, $(g, v_k)_Y = (Pg, v_k)_Y \ \forall g \in Y$ and $\forall k \in \mathbb{N}$, so that $R_N g = R_N P g$ and, in particular, if $g \in D(A^\dagger)$, $R_N g = R_N A f^\dagger$. Then we easily get

$$R_N g = \sum_{k=0}^{N} \frac{1}{\sigma_k}(g, v_k)_Y \, u_k = \sum_{k=0}^{N} \frac{1}{\sigma_k}(Af^\dagger, v_k)_Y \, u_k = \sum_{k=0}^{N}(f^\dagger, u_k)_X \, u_k \qquad \forall g \in D(A^\dagger). \tag{1.103}$$

Thus, for $N \to \infty$ the regularized solution $R_N g$ tends to the generalized solution, i.e.

$$\lim_{N\to\infty} \left\| R_N g - f^\dagger \right\|_X = 0 \qquad \forall g \in D(A^\dagger). \tag{1.104}$$

By virtue of theorem 1.6.2, this implies that the family $\{R_N\}_{N\geq 0}$ is a regularization for $A^\dagger$. ■

Remark 1.7.1. Furthermore, it follows from the same theorem 1.6.2 that there exists, for every $g \in D(A^\dagger)$, an a priori parameter choice rule $N^* = N^*(\delta)$ such that (R_{N^*}, N^*) is a convergent regularization method for solving $Af = g$. Now, it is easy to realize that

$$\|R_N\| = \frac{1}{\sigma_N} \qquad \forall N \geq 0. \tag{1.105}$$

In fact, for each $h \in Y$, we can write

$$\|R_N h\|_X^2 = \left\| \sum_{k=0}^{N} \frac{1}{\sigma_k}(h, v_k)_Y \, u_k \right\|_X^2 = \sum_{k=0}^{N} \frac{|(h, v_k)_Y|^2}{\sigma_k^2} \leq \sum_{k=0}^{N} \frac{|(h, v_k)_Y|^2}{\sigma_N^2} \leq \frac{\|h\|_Y^2}{\sigma_N^2} \tag{1.106}$$

and equality holds, e.g., for $h = v_N$. Thus we immediately get

$$\frac{\|R_N h\|_X}{\|h\|_Y} \leq \frac{1}{\sigma_N} \qquad \forall h \in Y \setminus \{0\} \tag{1.107}$$

and then, since equality can hold,

$$\|R_N\| := \sup_{0\neq h\in Y} \frac{\|R_N h\|_X}{\|h\|_Y} = \frac{1}{\sigma_N}, \tag{1.108}$$

i.e. (1.105). So we can apply theorem 1.6.3 to conclude that, if $N^* = N^*(\delta)$ is an a priori parameter choice rule, then (R_{N^*}, N^*) is a convergent regularization method if and only if

$$\lim_{\delta\to 0^+} N^*(\delta) = \infty, \qquad \lim_{\delta\to 0^+} \frac{\delta}{\sigma_{N^*(\delta)}} = 0. \tag{1.109}$$

It might be worthwhile observing that the two conditions (1.109) are not contradictory: for example, if we define the a priori parameter choice rule $N^* = N^*(\delta)$ according to the prescription $\sigma^2_{N^*+1} \leq \delta \leq \sigma^2_{N^*}$, it is easy to see that both relations (1.109) are verified. □

Remark 1.7.2. The regularization algorithm of theorem 1.7.1 is called, for obvious reasons, *Truncated Singular Value Decomposition* (abbr. TSVD) and it is perhaps the easiest regularization algorithm. It often provides coarse reconstructions, but it may be helpful when a fast estimate of the solution of the inverse problem is needed. □

1.7.2. Tikhonov's method

Case of exact operator

Tikhonov's method is, historically, the first algorithm rigorously described in regularization theory [64] [65]. The first step to define such a method consists in considering the one-parameter family of minimum problems

$$\|Af_\alpha - g\|^2_Y + \alpha\|f_\alpha\|^2_X = \text{minimum}, \tag{1.110}$$

where $g \in Y$ is the generic (exact or noisy) datum of the problem and α is a real positive number. For convenience, let us define the following functional of f:

$$\Phi_\alpha[f; g] := \|Af - g\|^2_Y + \alpha\|f\|^2_X. \tag{1.111}$$

Roughly speaking, the central idea of this method is the following. If we consider an element f that makes the so-called *residual* $\|Af - g\|_Y$ too little, f itself is too close to the generalized solution $f^\dagger$ (if it exists), so it turns out to be substantially unstable, i.e. f may oscillate too wildly for small variations of g (if it is actually a noisy datum g_δ), and consequently it is not a good candidate to be a regularized solution. Hence, minimization of the functional (1.111) is a compromise between accuracy and stability, i.e. between the two opposite needs to keep small both the residual and the "penalty term" $\|f\|_X$. In other words, the first term in the functional (1.111), when small enough, guarantees that f is "nearly" a least squares solution (i.e., when

mapped by A, it reproduces with sufficient accuracy the datum g, which is, in general, noisy), while the second term, when small enough, tends to damp out wild instabilities in f itself.

In the following, we are going to give some theorems in order to rigorously describe Tikhonov's method.

Theorem 1.7.2. *For each α the minimum problem (1.110) is equivalent to the so-called* Euler equation *of the functional $\Phi_\alpha[f;g]$, i.e.*

$$(A^*A + \alpha I) f_\alpha = A^* g. \tag{1.112}$$

Proof. Indeed, f_α is a solution of the minimum problem (1.110) if and only if, for all complex numbers t and for all elements h of the Hilbert space X, we have

$$\|Af_\alpha - g\|_Y^2 + \alpha \|f_\alpha\|_X^2 \leq \|A(f_\alpha + th) - g\|_Y^2 + \alpha \|f_\alpha + th\|_X^2. \tag{1.113}$$

By writing the norms as scalar products one easily obtains

$$\begin{aligned} &|t|^2(\|Ah\|_Y^2 + \alpha\|h\|_X^2) + \\ &+t\{(Ah, Af_\alpha - g)_Y + \alpha(h, f_\alpha)_X\} + \\ &+\bar{t}\{(Af_\alpha - g, Ah)_Y + \alpha(f_\alpha, h)_X\} \geq 0 \qquad \forall t \in \mathbb{C},\ \forall h \in X, \end{aligned} \tag{1.114}$$

i.e.

$$|t|^2(\|Ah\|_Y^2 + \alpha\|h\|_X^2) + 2\mathrm{Re}\{t[(Ah, Af_\alpha - g)_Y + \alpha(h, f_\alpha)_X]\} \geq 0 \qquad \forall t \in \mathbb{C},\ \forall h \in X. \tag{1.115}$$

Since the term quadratic in $|t|$ is non-negative, it is not difficult to see that condition (1.115) can be satisfied if and only if the term linear in t is zero. Using the definition of adjoint operator, this condition can be written as

$$(h, A^*Af_\alpha + \alpha f_\alpha - A^*g)_X = 0 \qquad \forall h \in X, \tag{1.116}$$

whence the Euler equation (1.112) follows. ■

The previous theorem states that f_α is a minimum point of $\Phi_\alpha[f;g]$ if and only if it is a solution of the Euler equation (1.112). The next theorem states that the latter has always a unique solution, which belongs to $\mathcal{N}(A)^\perp$.

Theorem 1.7.3. *If $\alpha > 0$, equation (1.112) has a unique solution, denoted with f_α, for any $g \in Y$; furthermore $f_\alpha \in \mathcal{N}(A)^\perp$.*

Proof. The operator at the left-hand side of equation (1.112) is strictly positive, as follows from the inequality

$$(A^*Af + \alpha f, f)_X = \|Af\|_Y^2 + \alpha\|f\|_X^2 \geq \alpha\|f\|_X^2 \qquad \forall f \in X. \tag{1.117}$$

Then, by applying the Cauchy-Schwarz inequality to the scalar product in the first member, we obtain

$$\|(A^*A + \alpha I)f\|_X \geq \alpha\|f\|_X \qquad \forall f \in X. \tag{1.118}$$

The previous inequality (1.118) has the following implications:

1. the equation $(A^*A+\alpha I)f = 0$ has the unique solution $f = 0$, i.e. the solution of equation (1.112) is unique;

2. the inverse operator $(A^*A + \alpha I)^{-1} : D((A^*A + \alpha I)^{-1}) \subset X \to X$ is bounded (and its norm is bounded by α^{-1}).

We now recall that for any linear and continuous operator $T : X \to X$, with X a Hilbert space, it holds $\mathcal{N}(T^*) = \mathcal{R}(T)^\perp$; if we now observe that $(A^*A + \alpha I)$ is self-adjoint and that its kernel is the null space by virtue of the previous point 1, we get

$$\{0\} = \mathcal{N}(A^*A + \alpha I) = \mathcal{R}(A^*A + \alpha I)^\perp, \tag{1.119}$$

so that

$$\overline{\mathcal{R}(A^*A + \alpha I)} = X, \tag{1.120}$$

i.e. $D((A^*A+\alpha I)^{-1}) = \mathcal{R}(A^*A+\alpha I)$ is dense in X. It follows that the operator $(A^*A+\alpha I)^{-1}$ can be univocally extended to a continuous operator (which we shall denote in the same way) defined on all X. Hence equation (1.112) has a (unique) solution $\forall g \in X$; this solution can be written in the form:

$$f_\alpha = (A^*A + \alpha I)^{-1}A^*g. \tag{1.121}$$

Finally, in general we can obviously write

$$f_\alpha = f_{1,\alpha} + f_{2,\alpha}, \qquad \text{with } f_{1,\alpha} \in \mathcal{N}(A)^\perp,\ f_{2,\alpha} \in \mathcal{N}(A). \tag{1.122}$$

Of course, we have

$$\|Af_\alpha - g\|_Y = \|Af_{1,\alpha} - g\|_Y; \tag{1.123}$$

if, by absurd, it were $f_{2,\alpha} \neq 0$, it would hold

$$\|f_\alpha\|_X > \|f_{1,\alpha}\|_X, \tag{1.124}$$

so that we would get

$$\Phi_\alpha[f_\alpha; g] > \Phi_\alpha[f_{1,\alpha}; g] \tag{1.125}$$

and this is absurd, since f_α is the minimum point of the functional $\Phi_\alpha[f; g]$. This concludes the proof. ■

Remark 1.7.3. The last point of the previous theorem, i.e. the fact that $f_\alpha \in \mathcal{N}(A)^\perp$, can also be proved by considering expression (1.121) and showing that the range of the operator

$$R_\alpha := (A^*A + \alpha I)^{-1}A^* \tag{1.126}$$

is contained in $\mathcal{N}(A)^\perp$. For future purpose, we want to follow also this second way. Let us start with the obvious identity

$$(A^*A + \alpha I)A^* = A^*(AA^* + \alpha I). \tag{1.127}$$

From the proof of the previous theorem 1.7.3, we already know that $A^*A + \alpha I$ has a bounded inverse $(A^*A+\alpha I)^{-1}$ that can be thought as defined on all X. With exactly the same arguments, one proves that also $AA^* + \alpha I$ has a bounded inverse $(AA^* + \alpha I)^{-1}$ that can be thought as defined on all Y: indeed, both the operators $A^*A + \alpha I$ and $AA^* + \alpha I$ are continuous, strictly positive and (therefore) self-adjoint. Hence, by multiplying both the members of identity (1.127) on the left by $(A^*A + \alpha I)^{-1}$ and on the right by $(AA^* + \alpha I)^{-1}$, and remembering definition (1.126), we get

$$R_\alpha = A^*(AA^* + \alpha I)^{-1}. \tag{1.128}$$

This implies that

$$\mathcal{R}(R_\alpha) \subset \mathcal{N}(A)^\perp, \tag{1.129}$$

since it obviously holds $\mathcal{R}(R_\alpha) \subset \mathcal{R}(A^*) \subset \overline{\mathcal{R}(A^*)} = \mathcal{N}(A)^\perp$. □

We can now enunciate and prove the most important theorem of this subsection. There are at least two quite different proofs of it; since they are equally interesting and instructive, we give them both.

Theorem 1.7.4. *The one-parameter family of operators $\{R_\alpha\}_{\alpha>0}$ defined by (1.126), i.e.*

$$R_\alpha := (A^*A + \alpha I)^{-1}A^*, \tag{1.130}$$

defines a linear regularization for $A^\dagger$.

Proof No 1. It is obvious that each R_α is linear and continuous, since it is a composition of two operators, i.e. $(A^*A + \alpha I)^{-1}$ and A^*, that are endowed with these two properties[20]. Furthermore, remembering that $\mathcal{N}(A^*) = \mathcal{R}(A)^\perp$ and definition (1.130), we immediately realize that

$$R_\alpha g = R_\alpha P g \qquad \forall g \in Y, \tag{1.131}$$

where P denotes, as usual, the orthogonal projection onto $\overline{\mathcal{R}(A)}$. But when $g \in Y$ is such that $Pg \in \mathcal{R}(A)$, we have that $Pg = Af^\dagger$, so that

$$R_\alpha g = R_\alpha A f^\dagger \qquad \forall g \in D(A^\dagger). \tag{1.132}$$

[20] About the continuity of $(A^*A + \alpha I)^{-1}$, see the proof of theorem 1.7.3.

It follows that hypothesis (1.74) of theorem 1.6.2, which we now want to prove, can be rewritten in our case as

$$\lim_{\alpha\to 0^+} \left\| R_\alpha A f^\dagger - f^\dagger \right\|_X = 0. \tag{1.133}$$

In order to prove condition (1.133), we shall use arguments based on the spectral theory of linear, continuous and self-adjoint operators [11], [51], [33]. If we denote with dE_λ the spectral measure defined by the spectral family associated to the self-adjoint and positive operator A^*A, the following integral representation holds[21]:

$$\left\| R_\alpha A f^\dagger - f^\dagger \right\|_X = \left\| \int_0^{\|A\|^2} \frac{\alpha}{\lambda+\alpha} dE_\lambda f^\dagger \right\|_X . \tag{1.134}$$

Let us consider, at first, the following limit:

$$\lim_{\alpha\to 0^+} \int_0^{\|A\|^2} \frac{\alpha}{\lambda+\alpha} dE_\lambda f^\dagger; \tag{1.135}$$

for each $\alpha > 0$, the function $\frac{\alpha}{\lambda+\alpha}$ of λ is integrable with respect to the spectral measure over $[0, \|A\|^2]$ and is bounded by 1, integrable over the same interval. Then we can apply the dominated convergence theorem and carry the limit inside the integral. Since it is obviously:

$$\lim_{\alpha\to 0^+} \frac{\alpha}{\lambda+\alpha} = \begin{cases} 1 & \text{if } \lambda = 0 \\ 0 & \text{if } \lambda \in (0, \|A\|^2] \end{cases} \tag{1.136}$$

we get

$$\lim_{\alpha\to 0^+} \int_0^{\|A\|^2} \frac{\alpha}{\lambda+\alpha} dE_\lambda f^\dagger = E_0 f^\dagger, \tag{1.137}$$

where E_0 is the projection onto $\mathcal{N}(A^*A) = \mathcal{N}(A)$, so that $E_0 f^\dagger = 0$. Finally, by means of the continuity of $\|\cdot\|_X$, we have

$$0 = \left\| \lim_{\alpha\to 0^+} \int_0^{\|A\|^2} \frac{\alpha}{\lambda+\alpha} dE_\lambda f^\dagger \right\|_X = \lim_{\alpha\to 0^+} \left\| \int_0^{\|A\|^2} \frac{\alpha}{\lambda+\alpha} dE_\lambda f^\dagger \right\|_X , \tag{1.138}$$

so that, by recalling equation (1.134), we finally get relation (1.133). By virtue of theorem 1.6.2, this implies that the family $\{R_\alpha\}_{\alpha>0}$ is a (linear) regularization for $A^\dagger$. ■

[21] We incidentally observe that representation (1.134) implies that $\left\| R_\alpha A f^\dagger - f^\dagger \right\|_X$ is an increasing function of α.

Proof No 2. First of all, we proceed exactly as in the previous proof until relation (1.133), which we are going to prove. Then we observe that, by virtue of theorems 1.7.2 and 1.7.3, it holds

$$R_\alpha A f^\dagger = \operatorname{argmin} \Phi_\alpha \left[f; A f^\dagger \right], \tag{1.139}$$

i.e.

$$\Phi_\alpha \left[R_\alpha A f^\dagger; A f^\dagger \right] = \text{minimum}. \tag{1.140}$$

Moreover, we define:

$$f_\alpha := R_\alpha A f^\dagger. \tag{1.141}$$

Now, let $\{\alpha_n\}_{n=0}^\infty$ be any sequence such that $\alpha_n > 0 \ \forall n \in \mathbb{N}$ and $\lim_{n\to\infty} \alpha_n = 0$. It is easy to realize that the following inequalities or equalities hold:

$$\alpha_n \left\| f_{\alpha_n} \right\|_X^2 \leq \Phi_{\alpha_n} \left[f_{\alpha_n}; A f^\dagger \right] \leq \Phi_{\alpha_n} \left[f^\dagger; A f^\dagger \right] = \alpha_n \left\| f^\dagger \right\|_X^2. \tag{1.142}$$

Summing up, the relations (1.142) imply that

$$\left\| f_{\alpha_n} \right\|_X \leq \left\| f^\dagger \right\|_X \qquad \forall n \in \mathbb{N}, \tag{1.143}$$

i.e. the sequence $\{f_{\alpha_n}\}_{n=0}^\infty$ is bounded in the Hilbert space X, and therefore [13] it has a subsequence $\{f_{\alpha_{n(k)}}\}_{k=0}^\infty$ that is weakly convergent to an element $f^* \in X$, i.e.

$$f_{\alpha_{n(k)}} \rightharpoonup f^*. \tag{1.144}$$

Furthermore, it is easy to realize that the functional $F : X \to \mathbb{R}$ defined by $F(x) := \|x\|_X$ is continuous and convex; therefore [7] it is also weakly lower semicontinuous, so that the first of the following inequalities holds (the other ones are trivial):

$$\left\| f^* \right\|_X \leq \liminf_{k\to\infty} \left\| f_{\alpha_{n(k)}} \right\|_X \leq \limsup_{k\to\infty} \left\| f_{\alpha_{n(k)}} \right\|_X \leq \left\| f^\dagger \right\|_X. \tag{1.145}$$

Since also the functional $G : X \to \mathbb{R}$ defined by $G(x) := \left\| Ax - Ax^\dagger \right\|_Y$ is continuous and convex, it is weakly lower semicontinuous too, so that

$$\left\| A f^* - A f^\dagger \right\|_Y \leq \liminf_{k\to\infty} \left\| A f_{\alpha_{n(k)}} - A f^\dagger \right\|_Y \leq \limsup_{k\to\infty} \left\| A f_{\alpha_{n(k)}} - A f^\dagger \right\|_Y. \tag{1.146}$$

Now, it is not difficult to see that the following inequalities or equalities hold:

$$\left\| A f_{\alpha_{n(k)}} - A f^\dagger \right\|_Y^2 \leq \Phi_{\alpha_{n(k)}} \left[f_{\alpha_{n(k)}}; A f^\dagger \right] \leq \Phi_{\alpha_{n(k)}} \left[f^\dagger; A f^\dagger \right] = \alpha_{n(k)} \left\| f^\dagger \right\|_X^2, \tag{1.147}$$

whence we immediately get

$$\limsup_{k\to\infty} \left\| A f_{\alpha_{n(k)}} - A f^\dagger \right\|_Y = 0. \tag{1.148}$$

Substituting this result into (1.146), we find that $Af^* = Af^\dagger$; but the generalized solution $f^\dagger$ is the unique minimum norm solution and, on the other hand, we have found (see (1.145)) $\|f^*\|_X \leq \left\|f^\dagger\right\|_X$. Hence, we have that $f^* = f^\dagger$ and so, again from (1.145), we immediately get that

$$\lim_{k\to\infty} \left\|f_{\alpha_{n(k)}}\right\|_X = \left\|f^\dagger\right\|_X . \tag{1.149}$$

Summing up, the two relations (1.144) (with f^* replaced by $f^\dagger$) and (1.149) respectively say that the subsequence $\{f_{\alpha_{n(k)}}\}_{k=0}^\infty$ converges weakly to $f^\dagger$ and that the subsequence $\{\|f_{\alpha_{n(k)}}\|_X\}_{k=0}^\infty$ converges to $\left\|f^\dagger\right\|_X$. For a well-known theorem[22] [7], this implies that $\{f_{\alpha_{n(k)}}\}_{k=0}^\infty$ converges strongly to $f^\dagger$, i.e.

$$\lim_{k\to\infty} \left\|f_{\alpha_{n(k)}} - f^\dagger\right\|_X = 0. \tag{1.150}$$

Then we have found that, for each sequence $\{\alpha_n\}_{n=0}^\infty$ such that $\alpha_n > 0\ \forall n \in \mathbb{N}$ and $\lim_{n\to\infty} \alpha_n = 0$, there exists a subsequence $\{\alpha_{n(k)}\}_{k=0}^\infty$ such that relation (1.150) holds: it is not difficult to see that this imply

$$\lim_{\alpha\to 0^+} \left\|f_\alpha - f^\dagger\right\|_X = 0. \tag{1.151}$$

Indeed, let us suppose, by absurd, that (1.151) is not true; then, it would exist an $\varepsilon > 0$ such that for any right neighbourhood $U_{0,n}^+ \subset \mathbb{R}^+$ of 0 (with $0 \notin U_{0,n}^+$) there exists at least one value $\alpha_n(\varepsilon) \equiv \alpha_n \in U_{0,n}^+$ of α such that

$$\left\|f_{\alpha_n} - f^\dagger\right\|_X > \varepsilon. \tag{1.152}$$

Then, let us consider a family of right neighbourhoods $\left\{U_{0,n}^+\right\}_{n=0}^\infty$ such that $U_{0,0}^+ \supset U_{0,1}^+ \supset U_{0,2}^+ \ldots$ and $\operatorname{mis} U_{0,n}^+ < 1/n$; for each $U_{0,n}^+$, let us choose arbitrarily an $\alpha_n > 0$ such that (1.152) holds: thus we obtain a sequence $\{\alpha_n\}_{n=0}^\infty$ (such that $\lim_{n\to\infty} \alpha_n = 0$) and the corresponding sequence $\{f_{\alpha_n}\}_{n=0}^\infty$ in X. Hence, by virtue of the previous arguments, it is possible to extract a subsequence $\left\{f_{\alpha_{n(k)}}\right\}_{k=0}^\infty$ such that (1.150) holds; on the other hand, by construction of the subsequence itself, relation (1.152) holds too, i.e.

$$\left\|f_{\alpha_{n(k)}} - f^\dagger\right\|_X > \varepsilon \qquad \forall k \in \mathbb{N}. \tag{1.153}$$

This is a contradiction, therefore relation (1.151) must hold. If we remember definition (1.141), we have found

$$\lim_{\alpha\to 0^+} \left\|R_\alpha A f^\dagger - f^\dagger\right\|_X = 0, \tag{1.154}$$

i.e. exactly relation (1.133). Summing up, we have proved, by means of theorem 1.6.2, that $\{R_\alpha\}_{\alpha>0}$ is a (linear) regularization for $A^\dagger$. ■

[22] It is the so-called *H-property*, by virtue of which weak convergence and norm convergence imply strong convergence. A Hilbert space has the H-property.

Remark 1.7.4. It is worthwhile observing that, by virtue of theorem 1.6.2 itself, there exists, for each $g \in D(A^\dagger)$, an a priori parameter choice rule $\alpha^* = \alpha^*(\delta)$ such that (R_{α^*}, α^*) is a convergent regularization method for solving $Af = g$. On the other hand, it is possible to show that

$$\|R_\alpha\| \leq \frac{1}{\sqrt{\alpha}} \qquad \forall \alpha > 0. \tag{1.155}$$

Indeed, by means of relation (1.128), we can write, for any $g \in Y$ and any $\alpha > 0$,

$$\begin{aligned} \|R_\alpha g\|_X^2 &= (A^*(AA^* + \alpha I)^{-1} g, A^*(AA^* + \alpha I)^{-1} g) = \\ &= (AA^*(AA^* + \alpha I)^{-1} g, (AA^* + \alpha I)^{-1} g); \end{aligned} \tag{1.156}$$

by applying the Cauchy-Schwarz inequality to the last term, one has

$$\|R_\alpha g\|_X^2 \leq \|AA^*(AA^* + \alpha I)^{-1}\| \|(AA^* + \alpha I)^{-1}\| \|g\|_Y^2. \tag{1.157}$$

Now, if we denote with $\Lambda(AA^*)$ the spectrum of AA^*, we get

$$\|AA^*(AA^* + \alpha I)^{-1}\| = \sup_{\lambda \in \Lambda(AA^*)} \frac{\lambda}{\lambda + \alpha} \leq 1 \tag{1.158}$$

and

$$\|(AA^* + \alpha I)^{-1}\| = \sup_{\lambda \in \Lambda(AA^*)} \frac{1}{\lambda + \alpha} \leq \frac{1}{\alpha}, \tag{1.159}$$

so that relation (1.155) follows.

Hence, we can apply theorem 1.6.3 to conclude that, if $\alpha^* = \alpha^*(\delta)$ is an a priori parameter choice rule, then a sufficient condition for (R_{α^*}, α^*) to be a convergent regularization method is that the two following (and clearly not contradictory) relations hold:

$$\lim_{\delta \to 0^+} \alpha^*(\delta) = 0, \qquad \lim_{\delta \to 0^+} \frac{\delta}{\sqrt{\alpha^*(\delta)}} = 0. \tag{1.160}$$

If we cling only to our previous considerations, we must conclude that, since inequality (1.155) gives only a bound for $\|R_\alpha\|$, but not an exact value, the second of conditions (1.160) is not necessary (while the first, obviously, is). Actually, it can be shown that it is also necessary (see [6], p. 82). □

Remark 1.7.5. Inequality (1.155) can be improved: indeed, it is possible to demonstrate that

$$\|R_\alpha\| \leq \frac{1}{2\sqrt{\alpha}} \qquad \forall \alpha > 0. \tag{1.161}$$

For a proof, one can use, e.g., inequality (2.48) at p.45 in [33]. □

Remark 1.7.6. If $g \in \mathcal{R}(A)$, the result obtained in the previous remark 1.7.4, i.e. the sufficiency of conditions (1.160) for (R_{α^*}, α^*) to be a convergent regularization method, can be obtained in a completely different way, by following a reasoning that is very similar to the one of previous proof No 2. We shall employ this technique in the more general case of a noisy operator A_h (see theorem 1.7.5), so that also our current situation, in which $h = 0$, will be covered. □

Remark 1.7.7. Let us now consider the particular case in which the operator A is compact; let $\{\sigma_k, u_k, v_k\}_{k=0}^{\infty}$ be, as usual, its singular system: then, for each $\alpha > 0$ the solution $f_\alpha \in \mathcal{N}(A)^\perp$ of the minimum problem (1.110) can be expanded as

$$f_\alpha = \sum_{k=0}^{\infty} (f_\alpha, u_k)_X \, u_k. \tag{1.162}$$

If we now recall the singular representations (1.48), (1.49) and substitute them, together with (1.162), in the Euler equation (1.112), we straightforwardly obtain

$$f_\alpha = \sum_{k=0}^{\infty} \frac{\sigma_k}{\sigma_k^2 + \alpha} (g, v_k)_Y \, u_k, \tag{1.163}$$

where $g \in Y$ is the generic (exact or noisy) datum of the problem. A comparison with (1.56) clearly shows the stabilization: errors in $(g, v_k)_Y$ are not propagated into the result with the increasing factors $\frac{1}{\sigma_k}$, but only with the factors $\frac{\sigma_k}{\sigma_k^2+\alpha}$, which remain bounded, till vanishing as $k \to 0$. □

Case of noisy operator

Till now, we have seen that, given the linear inverse problem

$$Af = g_\delta, \tag{1.164}$$

with exact operator A and noisy datum g_δ (being $\|g_\delta - g\|_Y \leq \delta$), its regularized solution, according to Tikhonov's method, is given by

$$f_\alpha^\delta := \operatorname{argmin} \Phi_\alpha^\delta[f; g_\delta], \tag{1.165}$$

where

$$\Phi_\alpha^\delta[f; g_\delta] := \|Af - g_\delta\|_Y^2 + \alpha \|f\|_X^2. \tag{1.166}$$

It follows that

$$f_\alpha^\delta = R_\alpha g_\delta, \tag{1.167}$$

where

$$R_\alpha := (A^*A + \alpha I)^{-1} A^*. \tag{1.168}$$

Now, if we consider the linear inverse problem

$$A_h f = g_\delta, \tag{1.169}$$

with perturbed operator A_h and noisy datum g_δ satisfying, as usual, the conditions

$$\|g_\delta - g\|_Y \leq \delta, \tag{1.170}$$

$$\|A_h - A\| \leq h, \tag{1.171}$$

the most natural generalization we can think of is obviously to consider as smoothing functional the following one[23]:

$$\Phi_\alpha^\eta[f; g_\delta] := \|A_h f - g_\delta\|_Y^2 + \alpha \|f\|_X^2 \tag{1.172}$$

and, consequently, to define the new regularized solution as

$$f_\alpha^\eta := \operatorname{argmin} \Phi_\alpha^\eta[f; g_\delta]. \tag{1.173}$$

Clearly, since theorems 1.7.2, 1.7.3 and remark 1.7.3 hold by virtue of the mere continuity of A, they keep on holding if we substitute everywhere A_h to A: in particular, the Euler equation of the functional $\Phi_\alpha^\eta[f; g_\delta]$ is obtained by replacing A with A_h (and f_α with f_α^η) in equation (1.112), i.e.:

$$(A_h^* A_h + \alpha I)\, f_\alpha^\eta = A_h^*\, g_\delta. \tag{1.174}$$

Hence, we take as regularization operator the family $\{R_\alpha^{(h)}\}_{\alpha>0}$, defined by

$$R_\alpha^{(h)} := (A_h^* A_h + \alpha I)^{-1} A_h^*, \tag{1.175}$$

so that the regularized solution turns out to be

$$f_\alpha^\eta = R_\alpha^{(h)} g_\delta. \tag{1.176}$$

Of course, by means of the same arguments used in remark 1.7.7, we can show that, if A_h is compact and $\{\sigma_k^h, u_k^h, v_k^h\}_{k=0}^\infty$ is its singular system, then the following representation for f_α^η holds:

$$f_\alpha^\eta = \sum_{k=0}^{\infty} \frac{\sigma_k^h}{(\sigma_k^h)^2 + \alpha} (g_\delta, v_k^h)_Y\, u_k^h. \tag{1.177}$$

However, theorem 1.7.4 cannot be trivially generalized, since $R_\alpha^{(h)}$ depends on A_h and then theorem 1.6.2 cannot be applied. Hence, in order to show that $\{R_\alpha^{(h)}\}_{\alpha>0}$ is a regularization, we shall have to consider from the very beginning a certain class of parameter choice rules (or a specific one, as it will be the case for the discrepancy method in section 1.8) and to deal directly with definition 1.6.1. As an illustration of this fact, we can give the following theorem.

[23] We recall that we denote with $\eta := (\delta, h)$ the two noise levels together.

Theorem 1.7.5. *Let, as usual, $A_h f = g_\delta$ be the noisy version of the exact linear inverse problem $Af = g$, with $g \in \mathcal{R}(A)$, and let relations (1.170), (1.171) hold. If $\{R_\alpha^{(h)}\}_{\alpha>0}$ is the one-parameter family of operators defined by (1.175), i.e.*

$$R_\alpha^{(h)} := (A_h^* A_h + \alpha I)^{-1} A_h^*, \tag{1.178}$$

and if $\alpha^* = \alpha^*(\delta, h)$ is an a priori parameter choice rule such that[24]

$$\lim_{(\delta,h)\to(0^+,0^+)} \alpha^*(\delta, h) = 0, \qquad \lim_{(\delta,h)\to(0^+,0^+)} \frac{(h+\delta)^2}{\alpha^*(\delta, h)} = 0, \tag{1.179}$$

then the pair $(R_{\alpha^*}^{(h)}, \alpha^*)$ is a convergent regularization method for solving the exact equation $Af = g$.

Proof. Let $\{(\delta_n, h_n)\}_{n=0}^\infty$ be any sequence such that $\delta_n > 0$, $h_n > 0$ $\forall n \in \mathbb{N}$ and, additionally, $\lim_{n\to\infty}(\delta_n, h_n) = (0, 0)$; for notational convenience, we shall denote this sequence with $\{\eta_n\}_{n=0}^\infty$ and we shall write $\eta_n > 0$ $\forall n \in \mathbb{N}$ and $\lim_{n\to\infty} \eta_n = 0$. Then, let us consider, for each η_n, the corresponding $\alpha_n^* := \alpha^*(\delta_n, h_n) > 0$ and let us choose arbitrarily $g_{\delta_n} \in Y$ and $A_{h_n} \in \mathcal{B}(X, Y)$ such that, respectively, $\|g_{\delta_n} - g\|_Y \leq \delta_n$ and $\|A_{h_n} - A\| \leq h_n$. Moreover, let us consider the correspondent regularized solution, defined as

$$f_{\alpha_n^*}^{\eta_n} := \mathrm{argmin}\Phi_{\alpha_n^*}^{\eta_n}[f; g_{\delta_n}]. \tag{1.180}$$

It is not difficult to see that the following equalities or inequalities hold:

$$\begin{aligned}
\alpha_n^* \left\| f_{\alpha_n^*}^{\eta_n} \right\|_X^2 \leq \Phi_{\alpha_n^*}^{\eta_n}\left[f_{\alpha_n^*}^{\eta_n}; g_{\delta_n}\right] \leq \Phi_{\alpha_n^*}^{\eta_n}\left[f^\dagger; g_{\delta_n}\right] = \left\| A_h f^\dagger - g_{\delta_n} \right\|_Y^2 + \alpha_n^* \left\| f^\dagger \right\|_X^2 \leq \\
\leq \left(\left\| A_h f^\dagger - A f^\dagger \right\|_Y + \left\| A f^\dagger - g_{\delta_n} \right\|_Y \right)^2 + \alpha_n^* \left\| f^\dagger \right\|_X^2 \leq \\
\leq \left(h_n \left\| f^\dagger \right\|_X + \delta_n \right)^2 + \alpha_n^* \left\| f^\dagger \right\|_X^2,
\end{aligned} \tag{1.181}$$

where, in the last passage, we have observed that $Af^\dagger = g$ (since, by hypothesis, $g \in \mathcal{R}(A)$) and remembered conditions (1.170), (1.171). From (1.181) it follows immediately that

$$\left\| f_{\alpha_n^*}^{\eta_n} \right\|_X^2 \leq \frac{\left(h_n \left\| f^\dagger \right\|_X + \delta_n \right)^2}{\alpha_n^*} + \left\| f^\dagger \right\|_X^2. \tag{1.182}$$

The second of conditions (1.179) implies, in particular, that there exists a constant $C > 0$ such that

$$\frac{\left(h_n \left\| f^\dagger \right\|_X + \delta_n \right)^2}{\alpha_n^*} \leq C \qquad \forall n \in \mathbb{N}. \tag{1.183}$$

Relations (1.182) and (1.183) together imply that the sequence $\left\{ f_{\alpha_n^*}^{\eta_n} \right\}_{n=0}^\infty$ is bounded in the Hilbert space X, and therefore [13] it has a subsequence $\left\{ f_{\alpha_{n(k)}^*}^{\eta_{n(k)}} \right\}_{k=0}^\infty$ (which we shall denote,

[24] The first of the two relations (1.179) is obviously a trivial rewriting of condition (1.69) for α^* to be an a priori parameter choice rule. Hence, the only task of the following proof is to demonstrate condition (1.68).

for notational convenience, with $\left\{f_{\alpha_k^*}^{\eta_k}\right\}_{k=0}^{\infty}$) that it is weakly convergent to an element $f^* \in X$, i.e.

$$f_{\alpha_k^*}^{\eta_k} \rightharpoonup f^*. \tag{1.184}$$

Hence, using the weak lower semicontinuity of the norm $\|\cdot\|_X$ [7], the second of conditions (1.179) and inequality (1.182), we can easily get:

$$\|f^*\|_X \leq \liminf_{k\to\infty} \left\|f_{\alpha_k^*}^{\eta_k}\right\|_X \leq \limsup_{k\to\infty} \left\|f_{\alpha_k^*}^{\eta_k}\right\|_X \leq \left\|f^\dagger\right\|_X. \tag{1.185}$$

Since also the functional $x \mapsto \left\|Ax - Ax^\dagger\right\|_Y$ is weakly lower semicontinuous (see proof No 2 of theorem 1.7.4), we get

$$\left\|Af^* - Af^\dagger\right\|_Y \leq \liminf_{k\to\infty} \left\|Af_{\alpha_k^*}^{\eta_k} - Af^\dagger\right\|_Y \leq \limsup_{k\to\infty} \left\|Af_{\alpha_k^*}^{\eta_k} - Af^\dagger\right\|_Y. \tag{1.186}$$

Using, in particular, the triangle inequality, the fact that $Af^\dagger = g$ and relations (1.181), (1.182), (1.183), we have the following inequalities:

$$\begin{aligned}\left\|Af_{\alpha_k^*}^{\eta_k} - Af^\dagger\right\|_Y &\leq \left\|Af_{\alpha_k^*}^{\eta_k} - A_{h_k}f_{\alpha_k^*}^{\eta_k}\right\|_Y + \left\|A_{h_k}f_{\alpha_k^*}^{\eta_k} - g_{\delta_k}\right\|_Y + \|g_{\delta_k} - g\|_Y \leq \\ &\leq h_k \left\|f_{\alpha_k^*}^{\eta_k}\right\|_X + \left(\Phi_{\alpha_k^*}^{\eta_k}\left[f_{\alpha_k^*}^{\eta_k}; g_{\delta_k}\right]\right)^{1/2} + \delta_k \leq \\ &\leq h_k \left(C + \left\|f^\dagger\right\|_X^2\right)^{1/2} + \left(\left(h_k \left\|f^\dagger\right\|_X + \delta_k\right)^2 + \alpha_k^* \left\|f^\dagger\right\|_X^2\right)^{1/2} + \delta_k,\end{aligned}$$

whence we immediately get

$$\limsup_{k\to\infty} \left\|Af_{\alpha_k^*}^{\eta_k} - Af^\dagger\right\|_Y = 0. \tag{1.187}$$

Substituting this result into (1.186), we find that $Af^* = Af^\dagger$; but the generalized solution is the unique minimum norm solution and, on the other hand, we have found (see (1.185)) $\|f^*\|_X \leq \left\|f^\dagger\right\|_X$. Hence, we have that $f^* = f^\dagger$ and so, again from (1.185), we immediately get that

$$\lim_{k\to\infty} \left\|f_{\alpha_k^*}^{\eta_k}\right\|_X = \left\|f^\dagger\right\|_X. \tag{1.188}$$

Summing up, by virtue of the H-property (see footnote 22), the two relations (1.184) (with f^* replaced by $f^\dagger$) and (1.188) imply that

$$\lim_{k\to\infty} \left\|f_{\alpha_k^*}^{\eta_k} - f^\dagger\right\|_X = 0. \tag{1.189}$$

Now, it is not difficult to see that equation (1.189) implies our thesis, i.e., according to condition (1.68) and remembering (1.176),

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta, A_h} \left\{\left\|f_{\alpha^*(\delta,h)}^{\eta} - f^\dagger\right\|_X \mid \|g_\delta - g\|_Y \leq \delta;\ \|A_h - A\| \leq h\right\} = 0. \tag{1.190}$$

In fact, let us suppose, by absurd, that (1.190) is not true; then, there exist an $\varepsilon > 0$ and a sequence[25] $\{\eta_n(\varepsilon)\}_{n=0}^{\infty} \equiv \{\eta_n\}_{n=0}^{\infty}$, with $\eta_n > 0\ \forall n \in \mathbb{N}$, such that $\lim_{n\to\infty} \eta_n = 0$ and

$$\sup_{g_{\delta_n}, A_{h_n}} \left\{ \left\| f_{\alpha_n^*}^{\eta_n} - f^\dagger \right\|_X \mid \|g_{\delta_n} - g\|_Y \le \delta_n;\ \|A_{h_n} - A\| \le h_n \right\} > \varepsilon \quad \forall n \in \mathbb{N}. \tag{1.191}$$

Inequality (1.191) clearly implies that $\forall n \in \mathbb{N}$ there exist $\tilde{g}_{\delta_n} \in Y$ and $\tilde{A}_{h_n} \in \mathcal{B}(X,Y)$, with $\|\tilde{g}_{\delta_n} - g\|_Y \le \delta_n$ and $\|\tilde{A}_{h_n} - A\| \le h_n$, such that the corresponding regularized solution

$$\tilde{f}_{\alpha_n^*}^{\eta_n} := \left(\tilde{A}_{h_n}^* \tilde{A}_{h_n} + \alpha_n^* I \right)^{-1} \tilde{A}_{h_n}^* \tilde{g}_{\delta_n} \tag{1.192}$$

satisfies the inequality $\left\| \tilde{f}_{\alpha_n^*}^{\eta_n} - f^\dagger \right\|_X \ge \frac{\varepsilon}{2}$; in other terms, there exists a sequence $\left\{ \tilde{f}_{\alpha_n^*}^{\eta_n} \right\}_{n=0}^{\infty}$ such that

$$\left\| \tilde{f}_{\alpha_n^*}^{\eta_n} - f^\dagger \right\|_X \ge \frac{\varepsilon}{2} \qquad \forall n \in \mathbb{N}. \tag{1.193}$$

On the other hand, as we have seen above, from the sequence $\left\{ \tilde{f}_{\alpha_n^*}^{\eta_n} \right\}_{n=0}^{\infty}$ itself we can extract a subsequence, say $\left\{ \tilde{f}_{\alpha_k^*}^{\eta_k} \right\}_{k=0}^{\infty}$, that verifies relation (1.189), i.e. such that

$$\lim_{k\to\infty} \left\| \tilde{f}_{\alpha_k^*}^{\eta_k} - f^\dagger \right\|_X = 0. \tag{1.194}$$

Relations (1.193) and (1.194) are obviously contradictory, so equation (1.190) must hold. This concludes the proof. ■

Remark 1.7.8. It is easy to see that the arguments employed in proving the previous theorem 1.7.5 keep on holding in the more general case of an a posteriori parameter choice rule, provided that we substitute hypotheses (1.179) with the following ones:

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta, A_h} \{ \alpha^*(\delta, h, g_\delta, A_h) \mid \|g_\delta - g\|_Y \le \delta;\ \|A_h - A\| \le h \} = 0, \tag{1.195}$$

and

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta, A_h} \left\{ \frac{(h+\delta)^2}{\alpha^*(\delta, h, g_\delta, A_h)} \mid \|g_\delta - g\|_Y \le \delta;\ \|A_h - A\| \le h \right\} = 0. \tag{1.196}$$

In fact, let us remember that if, on the one hand, at the very beginning of the previous proof we have arbitrarily chosen, for each η_n, a noisy datum g_{δ_n} and a noisy operator A_{h_n} such that, respectively, $\|g_{\delta_n} - g\|_Y \le \delta_n$ and $\|A_{h_n} - A\| \le h_n$, on the other hand each $\alpha_n^* := \alpha^*(\delta_n, h_n)$ did not depend on g_{δ_n} or A_{h_n}. Hence, the role of the uniform convergence with respect to parameters g_δ and A_h in limits (1.195), (1.196) is clearly to allow us to use again the very same arguments of the previous proof with new values of $\alpha_n^* := \alpha^*(\delta_n, h_n, g_{\delta_n}, A_{h_n})$, which now depend also on g_{δ_n} and A_{h_n}. □

[25] See also the last part of proof No 2 of theorem 1.7.4

Remark 1.7.9. As regards the previous theorem 1.7.5, how can we proceed in the general case, in which $g \in D(A^\dagger)$, but not necessarily $g \in \mathcal{R}(A)$? The most natural approach one can think of is the following: we already know, from theorem 1.5.1, that $f^\dagger$ is the generalized solution of $Af = g$ if and only if $f^\dagger$ is the minimum norm solution of the Euler equation $A^*Af = A^*g$. In other terms, if we construct the new exact problem

$$Bf = u \tag{1.197}$$

by defining the new exact operator and the new exact datum respectively as

$$B := A^*A, \qquad u := A^*g, \tag{1.198}$$

it follows immediately that the generalized solution $f^\dagger$ of problem (1.197) exists if and, mostly important, *only if* $u \in \mathcal{R}(B)$; moreover, the generalized solution of (1.197) exists if and only if the one of problem $Af = g$ exists, and they are, of course, the same. Summing up, when dealing with problem (1.197), we have no more the drawback that the generalized solution may exist without the exact datum belonging to the range of the exact operator. However, what now remains to do is not completely straightforward. In fact, if we decide to start from the new exact problem (1.197), we shall actually have to deal with its noisy version

$$B_H f = u_\Delta, \tag{1.199}$$

having defined

$$B_H := A_h^* A_h, \qquad u_\Delta := A_h^* g_\delta \tag{1.200}$$

and having denoted with H and Δ the new noise levels on the new noisy operator B_H and the new noisy datum u_Δ respectively. Hence, we could clearly restate theorem 1.7.5 replacing everywhere A, g, h, δ, A_h, g_δ, $R_\alpha^{(h)}$ respectively with B, u, H, Δ, B_H, u_Δ, $\hat{R}_\alpha^{(h)}$, having obviously defined the latter as

$$\hat{R}_\alpha^{(h)} := (B_H^* B_H + \alpha I)^{-1} B_H^*. \tag{1.201}$$

At this level, however, such a theorem would be completely useless in any application, as far as we know only the "physical" values δ and h of the noise levels, and not the new "mathematical" ones Δ and H. On the other hand, since such a theorem holds when considering Δ and H as free quantities that can tend to zero in whatever manner, even more so it will keep on holding when Δ and H tend to zero according to a certain law, endowed with suitable properties. More precisely, such a theorem remains true if we can give an estimate of Δ, H in terms of δ, h (and, in case, of $\|g_\delta\|_Y$, $\|A_h\|$), i.e., as we shall see very soon, $\Delta(\delta, h, \|g_\delta\|_Y, \|A_h\|)$ and $H(h, \|A_h\|)$, such that the three following conditions hold:

$$\Delta(\delta, h, \|g_\delta\|_Y, \|A_h\|) > 0, \quad H(h, \|A_h\|) > 0 \qquad \forall \delta, h > 0; \quad \forall \|g_\delta\|_Y, \|A_h\| \geq 0; \tag{1.202}$$

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta, A_h} \{\Delta(\delta, h, \|g_\delta\|_Y, \|A_h\|) \mid \|g_\delta - g\|_Y \leq \delta; \ \|A_h - A\| \leq h\} = 0^+ \tag{1.203}$$

and

$$\lim_{h\to 0^+} \sup_{A_h} \{H(h, \|A_h\|) \mid \|A_h - A\| \leq h\} = 0^+. \tag{1.204}$$

About the uniform convergence with respect to the parameters g_δ and A_h, see the previous remark 1.7.8.

Now, it is not so difficult to see that all three previous conditions can be satisfied. Firstly, let us recall that [32] if $T : X \to Y$ is a linear continuous operator between two Hilbert spaces X and Y, then it holds

$$\|T\| = \|T^*\|; \tag{1.205}$$

furthermore, if $S : X \to Y$ is another linear continuous operator between the same spaces and $a, b \in \mathbb{C}$, then it also holds:

$$(aS + bT)^* = \bar{a}S^* + \bar{b}T^*, \tag{1.206}$$

where $\bar{a}$, $\bar{b}$ obviously denote the complex conjugate of a, b respectively. Moreover, if we remember that

$$|\, \|A_h\| - \|A\| \,| \leq \|A_h - A\|\,, \tag{1.207}$$

from inequality (1.171) we immediately have

$$|\, \|A_h\| - \|A\| \,| \leq h; \tag{1.208}$$

hence, recalling definitions (1.198), (1.200), representations (1.63), (1.64), properties (1.205), (1.206) and inequality (1.208), we get

$$\begin{aligned} \|u_\Delta - u\|_X &= \|A_h^* g_\delta - A^* g\|_X = \|(A^* + N_h^*)(g + w_\delta) - A^* g\|_X = \\ &= \|N_h^* g_\delta + A^* w_\delta\|_X \leq h\|g_\delta\|_Y + \|A\|\delta \leq h\|g_\delta\|_Y + (\|A_h\| + h)\delta. \end{aligned} \tag{1.209}$$

This allows us to define

$$\Delta = \Delta(\delta, h, \|g_\delta\|_Y, \|A_h\|) := h\|g_\delta\|_Y + (\|A_h\| + h)\delta. \tag{1.210}$$

Analogously, one gets

$$\begin{aligned} \|B_h - B\| &= \|(A + N_h)^*(A + N_h) - A^*A\| = \|A^*N_h + N_h^*A + N_h^*N_h\| \leq \\ &\leq \|A^*\|\|N_h\| + \|N_h^*\|\|A\| + \|N_h\|^2 \leq 2h\|A\| + h^2 \leq 2h(\|A_h\| + h) + h^2, \end{aligned} \tag{1.211}$$

so that we can define

$$H = H(h, \|A_h\|) := 2h\|A_h\| + 3h^2. \tag{1.212}$$

From definitions (1.210), (1.212) it follows immediately that condition (1.202) is satisfied. Moreover, recalling again inequalities (1.170) e (1.171), we easily get

$$\sup_{g_\delta, A_h} \{\Delta(\delta, h, \|g_\delta\|_Y, \|A_h\|) \mid \|g_\delta - g\|_Y \leq \delta;\ \|A_h - A\| \leq h\} \leq h(\|g\|_Y + \delta) + (\|A\| + 2h)\delta \tag{1.213}$$

and

$$\sup_{A_h} \{H(h, \|A_h\|) \mid \|A_h - A\| \leq h\} \leq 2h(\|A\| + h) + 3h^2, \tag{1.214}$$

so that conditions (1.203) and (1.204) are immediately verified. Hence, if we observe that

$$[H(h, \|A_h\|) + \Delta(\delta, h, \|g_\delta\|_Y, \|A_h\|)]^2 = \left[2h\|A_h\| + 3h^2 + h\|g_\delta\|_Y + (\|A_h\| + h)\delta\right]^2, \tag{1.215}$$

we can finally restate theorem 1.7.5 as follows*:*

Theorem 1.7.6. *Let, as usual, $A_h f = g_\delta$ be the noisy version of the exact linear inverse problem $Af = g$, with $g \in D(A^\dagger)$, and let relations (1.170), (1.171) hold. If $\{R_\alpha^{(h)}\}_{\alpha>0}$ is the one-parameter family of operators defined by*

$$R_\alpha^{(h)} := \left((A_h^* A_h)^2 + \alpha I\right)^{-1} A_h^* A_h A_h^*, \tag{1.216}$$

and if $\alpha^* = \alpha^*(\delta, h, g_\delta, A_h)$ is an a posteriori parameter choice rule such that

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta, A_h} \{\alpha^*(\delta, h, g_\delta, A_h) \mid \|g_\delta - g\|_Y \leq \delta;\ \|A_h - A\| \leq h\} = 0, \tag{1.217}$$

and

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta, A_h} \left\{ \frac{\left[2h\|A_h\| + 3h^2 + h\|g_\delta\|_Y + (\|A_h\| + h)\delta\right]^2}{\alpha^*(\delta, h, g_\delta, A_h)} \ \middle|\ \|g_\delta - g\|_Y \leq \delta;\ \|A_h - A\| \leq h \right\} = 0, \tag{1.218}$$

then the pair $(R_{\alpha^*}^{(h)}, \alpha^*)$ is a convergent regularization method for solving the exact equation $Af = g$. □

A remarkable property of Tikhonov regularization is stated in the following theorem.

Theorem 1.7.7. *Let, as usual, $A_h f = g_\delta$ be the noisy version of the exact linear inverse problem $Af = g$, with $g \in D(A^\dagger)$, and let (cf. relations (1.175), (1.176))*

$$f_\alpha^\eta := (A_h^* A_h + \alpha I)^{-1} A_h^* g_\delta \tag{1.219}$$

be the Tikhonov regularized solution of the noisy problem $A_h f = g_\delta$. Furthermore, let $\alpha^* = \alpha^*(\delta, h, g_\delta, A_h)$ be any (a posteriori or, in case, a priori) parameter choice rule such that relation (1.68) holds, i.e.

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta, A_h} \left\{ \left\| f^\eta_{\alpha^*(\delta,h,g_\delta,A_h)} - f^\dagger \right\|_X \mid \|g_\delta - g\|_Y \leq \delta;\ \|A_h - A\| \leq h \right\} = 0, \tag{1.220}$$

where $f^\dagger$ is, as usual, the generalized solution of the exact problem $Af = g$. Then, if $f^\dagger \neq 0$, relation (1.69) holds too, i.e.

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta, A_h} \{\alpha^*(\delta, h, g_\delta, A_h) \mid \|g_\delta - g\|_Y \leq \delta;\ \|A_h - A\| \leq h\} = 0. \tag{1.221}$$

Proof. For any $\alpha > 0$, the following inequality

$$\left| \|f_\alpha^\eta\|_X - \left\|f^\dagger\right\|_X \right| \leq \left\|f_\alpha^\eta - f^\dagger\right\|_X \tag{1.222}$$

is clearly true. Then, from relations (1.220) and (1.222) we immediately get:

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta, A_h} \left\{ \left| \left\| f^\eta_{\alpha^*(\delta,h,g_\delta,A_h)} \right\|_X - \left\|f^\dagger\right\|_X \right| \ | \ \|g_\delta - g\|_Y \leq \delta; \ \|A_h - A\| \leq h \right\} = 0, \tag{1.223}$$

so that, for (δ, h) small enough and for any g_δ and A_h such that, respectively, $\|g_\delta - g\|_Y \leq \delta$ and $\|A_h - A\| \leq h$, it surely holds

$$\left\| f^\eta_{\alpha^*(\delta,h,g_\delta,A_h)} \right\|_X \geq \frac{\left\|f^\dagger\right\|_X}{2} > 0, \tag{1.224}$$

being $f^\dagger \neq 0$ by hypothesis. Moreover, from relation (1.219), holding for a generic $\alpha > 0$, we easily obtain

$$A_h^* \left(A_h f_\alpha^\eta - g_\delta\right) = -\alpha f_\alpha^\eta \tag{1.225}$$

and then

$$\left\|A_h^* \left(A_h f_\alpha^\eta - g_\delta\right)\right\|_X = \alpha \left\|f_\alpha^\eta\right\|_X \qquad \forall \alpha > 0. \tag{1.226}$$

Thus, if we replace the generic α with $\alpha^*(\delta, h, g_\delta, A_h)$ in relation (1.226), we get

$$\left\| A_h^* \left(A_h f^\eta_{\alpha^*(\delta,h,g_\delta,A_h)} - g_\delta \right) \right\|_X = \alpha^*(\delta, h, g_\delta, A_h) \left\| f^\eta_{\alpha^*(\delta,h,g_\delta,A_h)} \right\|_X . \tag{1.227}$$

Now, let us assume, for a moment, that the following limit holds (we shall give a proof of it soon below):

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta, A_h} \left\{ \left\| A_h^* \left(A_h f^\eta_{\alpha^*(\delta,h,g_\delta,A_h)} - g_\delta \right) \right\|_X \ | \ \|g_\delta - g\|_Y \leq \delta; \ \|A_h - A\| \leq h \right\} = 0; \tag{1.228}$$

hence, by virtue of relations (1.228) and (1.227), we have that

$$\lim_{(\delta,h)\to(0^+,0^+)} \sup_{g_\delta, A_h} \left\{ \alpha^*(\delta, h, g_\delta, A_h) \left\| f^\eta_{\alpha^*(\delta,h,g_\delta,A_h)} \right\|_X \ | \ \|g_\delta - g\|_Y \leq \delta; \ \|A_h - A\| \leq h \right\} = 0. \tag{1.229}$$

Now, let us suppose, by absurd, that our thesis (1.221) is not true: then, there exist an $\varepsilon > 0$ and a sequence $\{\eta_n(\varepsilon)\}_{n=0}^\infty \equiv \{\eta_n\}_{n=0}^\infty := \{(\delta_n, h_n)\}_{n=0}^\infty$, with $\eta_n > 0 \ \forall n \in \mathbb{N}$, such that $\lim_{n\to\infty} \eta_n = 0$ and

$$\sup_{g_{\delta_n}, A_{h_n}} \{\alpha^*(\delta_n, h_n, g_{\delta_n}, A_{h_n}) \mid \|g_{\delta_n} - g\|_Y \leq \delta_n; \ \|A_{h_n} - A\| \leq h_n\} > \varepsilon \qquad \forall n \in \mathbb{N}. \tag{1.230}$$

Then, by virtue of relations (1.224) and (1.230), there exists $N \in \mathbb{N}$ such that

$$\begin{aligned} \sup_{g_{\delta_n}, A_{h_n}} & \left\{ \alpha^*(\delta_n, h_n, g_{\delta_n}, A_{h_n}) \left\| f^\eta_{\alpha^*(\delta_n,h_n,g_{\delta_n},A_{h_n})} \right\|_X \ | \ \|g_{\delta_n} - g\|_Y \leq \delta_n; \ \|A_{h_n} - A\| \leq h_n \right\} \geq \\ & \geq \frac{\left\|f^\dagger\right\|_X}{2} \sup_{g_{\delta_n}, A_{h_n}} \{\alpha^*(\delta_n, h_n, g_{\delta_n}, A_{h_n}) \mid \|g_{\delta_n} - g\|_Y \leq \delta_n; \ \|A_{h_n} - A\| \leq h_n\} > \\ & > \frac{\left\|f^\dagger\right\|_X}{2} \varepsilon > 0 \qquad \forall n \geq N. \end{aligned} \tag{1.231}$$

Inequality (1.231) and limit (1.229) are clearly contradictory: hence, summing up, if we prove limit (1.228), our demonstration is complete.

To this end, we firstly observe that, for any $\alpha > 0$, the following chain of equalities or inequalities holds:

$$\begin{aligned}
&\left\|A_h^*\left(A_h f_\alpha^\eta - g_\delta\right) - A^*\left(Af^\dagger - g\right)\right\|_X = \left\|A_h^* A_h f_\alpha^\eta - A^* A f^\dagger + A^* g - A_h^* g_\delta\right\|_X \leq \\
&\leq \left\|A_h^* A_h f_\alpha^\eta - A_h^* A_h f^\dagger\right\|_X + \left\|A_h^* A_h f^\dagger - A^* A f^\dagger\right\|_X + \left\|A^* g - A_h^* g\right\|_X + \left\|A_h^* g - A_h^* g_\delta\right\|_X \leq \\
&\leq \left\|A_h^* A_h\right\| \left\|f_\alpha^\eta - f^\dagger\right\|_X + \left\|A_h^* A_h - A^* A\right\| \left\|f^\dagger\right\|_X + \left\|A^* - A_h^*\right\| \left\|g\right\|_Y + \left\|A_h^*\right\| \left\|g - g_\delta\right\|_Y .
\end{aligned} \tag{1.232}$$

Now, remembering relations (1.205), (1.206), (1.171) and (1.208), we easily obtain:

$$\left\|A_h^*\right\| = \left\|A_h\right\| \leq \left\|A\right\| + h, \tag{1.233}$$

$$\left\|A^* - A_h^*\right\| = \left\|A - A_h\right\| \leq h, \tag{1.234}$$

$$\begin{aligned}
\left\|A_h^* A_h - A^* A\right\| &\leq \left\|A_h^* A_h - A_h^* A\right\| + \left\|A_h^* A - A^* A\right\| \leq \left\|A_h^*\right\| \left\|A_h - A\right\| + \left\|A_h^* - A^*\right\| \left\|A\right\| \leq \\
&\leq \left\|A_h - A\right\| \left(\left\|A_h\right\| + \left\|A\right\|\right) \leq h\left(2\left\|A\right\| + h\right),
\end{aligned} \tag{1.235}$$

$$\left\|A_h^* A_h\right\| \leq \left\|A_h\right\|^2 \leq \left(\left\|A\right\| + h\right)^2 ; \tag{1.236}$$

hence, recalling also relation (1.170), the last inequality in (1.232) implies that:

$$\begin{aligned}
&\left\|A_h^*\left(A_h f_\alpha^\eta - g_\delta\right) - A^*\left(Af^\dagger - g\right)\right\|_X \leq \\
&\leq \left(\left\|A\right\| + h\right)^2 \left\|f_\alpha^\eta - f^\dagger\right\|_X + h\left(2\left\|A\right\| + h\right) \left\|f^\dagger\right\|_X + h\left\|g\right\|_Y + \left(\left\|A\right\| + h\right)\delta.
\end{aligned} \tag{1.237}$$

Now we observe that $\left(Af^\dagger - g\right) \in \mathcal{R}(A)^\perp = \mathcal{N}(A^*)$, and then

$$A^*\left(Af^\dagger - g\right) = 0; \tag{1.238}$$

substituting this result into (1.237), we immediately find:

$$\left\|A_h^*\left(A_h f_\alpha^\eta - g_\delta\right)\right\|_X \leq \left(\left\|A\right\| + h\right)^2 \left\|f_\alpha^\eta - f^\dagger\right\|_X + h\left(2\left\|A\right\| + h\right) \left\|f^\dagger\right\|_X + h\left\|g\right\|_Y + \left(\left\|A\right\| + h\right)\delta. \tag{1.239}$$

If we now remember hypothesis (1.220), we realize that inequality (1.239) implies limit (1.228). This concludes the proof. ■

1.8. The generalized discrepancy principle

1.8.1. Preliminary considerations

At first, we consider the case in which the exact operator A is known; soon afterwards, we shall also treat the case of noisy operator A_h.

Given the noisy linear inverse problem $Af = g_\delta$, whatever concept of solution we may have (proper, generalized or regularized), such a solution, when mapped by A, is expected to be able to reproduce, in a certain measure, the noisy datum g_δ.

More precisely, a proper solution, if it exists (in general, not unique), reproduces the datum exactly, i.e. $\|Af - g_\delta\|_Y = 0$, while the (noisy) generalized solution, if it exists, is the minimum norm element in X that minimizes the distance in Y between the datum g_δ and its possible "reconstructions" as images of the operator A, i.e.

$$f_\delta^\dagger = \operatorname{argmin}\|Af - g_\delta\|_Y \qquad (\text{with } \left\|f_\delta^\dagger\right\|_X = \text{minimum}). \tag{1.240}$$

Of course, if proper solutions exist, the (noisy) generalized solution is a proper one too. However, we have already observed that the generalized solution is completely corrupted by noise and then it turns out to be physically meaningless (see inequalities (1.9), (1.57) and their respective comments soon below).

Hence, when dealing with the regularized solution f_α^δ, what kind of requirement can we reasonably conceive about the quantity $\left\|Af_\alpha^\delta - g_\delta\right\|_Y$ (which is called *residual* or *discrepancy*), in such a way that f_α^δ is a stable and reliable approximation of the generalized solution $f^\dagger$ of the exact problem? The so-called *discrepancy principle* (due to Morozov [53]) yields a possible answer to this question: in fact, it is an algorithm that gives rise to a specific a posteriori parameter choice rule[26].

Roughly speaking, the central idea of the discrepancy principle is the following. We want to solve the exact problem $Af = g$, but, instead of g, we have only its noisy version g_δ and know that $\|g_\delta - g\|_Y \leq \delta$; hence, it does not make sense to look for an approximate solution f_α^δ with a discrepancy $\left\|Af_\alpha^\delta - g_\delta\right\|_Y < \delta$, since a residual in the order of δ is the best we should ask for: actually, since the datum g_δ is *noisy*, there is no sense in trying to reproduce it *exactly* by means of Af_α^δ. In other terms, when we merely write the expression of the discrepancy, i.e. $\left\|Af_\alpha^\delta - g_\delta\right\|_Y$, we necessarily commit an error bounded by δ, i.e. the discrepancy rises with a "default" error not greater than δ, so that requiring $\left\|Af_\alpha^\delta - g_\delta\right\|_Y < \delta$ would imply that the information contained in the discrepancy itself may be, in general, completely covered by noise.

Among the possible and technically different versions of this principle, we cite here the simplest one, that consists in choosing a value α^* (depending on δ and g_δ) of α such that

$$\left\|Af_{\alpha^*}^\delta - g_\delta\right\|_Y = \delta. \tag{1.241}$$

Obviously, some results and theorems are needed in order to show that such a choice is "well-posed" (i.e. not ambiguous or impossible) and, when considered together with a certain regularization operator $\{R_\alpha\}_{\alpha>0}$, gives rise to a convergent regularizing algorithm (R_{α^*}, α^*). We

[26] Strictly speaking, this is not completely true, since, as we shall see in the following (cf., in particular, remark 1.8.4), there are particular cases (i.e. when $f^\dagger = 0$ or the noise is somehow too large) in which no value of the regularization parameter is actually selected.

are going to see all this in the more general case of noisy operator (only for Tikhonov's method, although the discrepancy principle can be usefully applied in many other kinds of regularization).

In the case of perturbed operator A_h, such that $\|A_h - A\| \leq h$, we firstly observe that whenever $A_h x$ is computed for any $x \in X$, an error bounded by $h\|x\|_X$ is committed, since obviously

$$\|A_h x - Ax\|_Y \leq \|A_h - A\|\|x\|_X \leq h\|x\|_X. \tag{1.242}$$

Now, it trivially holds

$$\|A_h f_\alpha^\eta - g_\delta\|_Y \leq \|A_h f_\alpha^\eta\|_Y + \|g_\delta\|_Y \tag{1.243}$$

and the error that affects the first term at the right-hand side is bounded by $h\|f_\alpha^\eta\|_X$, while for the second term the bound on the error is δ. Hence, when the discrepancy $\|A_h f_\alpha^\eta - g_\delta\|_Y$ is computed, it does not make sense to look for an approximate solution f_α^η such that

$$\|A_h f_\alpha^\eta - g_\delta\|_Y < \delta + h\|f_\alpha^\eta\|_X, \tag{1.244}$$

since the discrepancy itself is affected by a "default" error bounded by $\delta + h\|f_\alpha^\eta\|_X$; the best we can ask for is to choose a value α^* (depending on δ, h, g_δ, A_h) of α such that in (1.244) equality holds:

$$\|A_h f_{\alpha^*}^\eta - g_\delta\|_Y = \delta + h\,\|f_{\alpha^*}^\eta\|_X\,. \tag{1.245}$$

This recipe is the core of the simplest form of the so-called *generalized discrepancy principle*; however, we shall need two more sophisticated versions of it, one[27] of which provides for the fact that the exact datum $g \in Y$ may not belong, in general, to the range of the exact operator A. In order to explain such a principle more carefully, we need to consider the following auxiliary functions of the regularization parameter α:

$$\gamma_\eta(\alpha) := \|f_\alpha^\eta\|_X^2\,; \tag{1.246}$$

$$\beta_\eta(\alpha) := \|A_h f_\alpha^\eta - g_\delta\|_Y^2\,. \tag{1.247}$$

These functions verify a lot of properties; here we recall only the ones we need for our purposes. For the proofs and a more general treatment, see [67].

Lemma 1.8.1. *As functions of α, it turns out that $\gamma_\eta(\alpha)$, $\beta_\eta(\alpha)$ have the following properties:*

1. *they are continuous in $(0, +\infty)$;*
2. *$\gamma_\eta(\alpha)$ is monotonically nonincreasing, $\beta_\eta(\alpha)$ is monotonically nondecreasing in $(0, +\infty)$;*
3. *if $\alpha_0 > 0$ is such that $f_{\alpha_0}^\eta \neq 0$, then $\gamma_\eta(\alpha)$ is strictly decreasing in $(0, \alpha_0)$, while $\beta_\eta(\alpha)$ is strictly increasing in $(0, \alpha_0)$;*

[27] See the next subsection 1.8.2.

4. $\lim_{\alpha\to+\infty} \gamma_\eta(\alpha) = 0$;

5. $\lim_{\alpha\to+\infty} \beta_\eta(\alpha) = \|g_\delta\|_Y^2$;

6. $\lim_{\alpha\to 0^+} \beta_\eta(\alpha) = \left[\inf_{f\in X} \|A_h f - g_\delta\|_Y\right]^2$.

Definition 1.8.1. *The following quantities:*

$$\mu := \inf_{f\in X} \|Af - g\|_Y, \tag{1.248}$$

$$\mu_\eta(g_\delta, A_h) := \inf_{f\in X} \|A_h f - g_\delta\|_Y, \tag{1.249}$$

$$\hat{\mu}_\eta(g_\delta, A_h) := \inf_{f\in X} (\delta + h\|f\|_X + \|A_h f - g_\delta\|_Y) \tag{1.250}$$

are called, respectively,

1. incompatibility measure *of the exact problem* $Af = g$;

2. (simple) incompatibility measure *of the noisy problem* $A_h f = g_\delta$;

3. modified incompatibility measure *of the noisy problem* $A_h f = g_\delta$.

Remark 1.8.1. It immediately follows from definitions (1.249) and (1.250) that, if $\delta > 0$, it holds:

$$\mu_\eta(g_\delta, A_h) < \hat{\mu}_\eta(g_\delta, A_h), \tag{1.251}$$

obviously $\forall g_\delta \in Y$ and $\forall A_h \in \mathcal{B}(X, Y)$ such that, respectively, $\|g_\delta - g\|_Y \leq \delta$ and $\|A_h - A\| \leq h$. On the other hand, if $\delta = 0$ (and, consequently, $g_\delta = g_0 = g$), we can only have a weak inequality, i.e.

$$\mu_\eta(g_0, A_h) \leq \hat{\mu}_\eta(g_0, A_h); \tag{1.252}$$

indeed, in (1.252) equality can hold: to this end, it suffices that $g_0 \in \mathcal{R}(A_h)^\perp$, so that the generalized solution $f_\eta^\dagger$ of the noisy problem $A_h f = g_0$ exists and is zero. Such a remark plays a role in the statement No 3 of theorem 1.8.3 in the following. Of course, the weak inequality

$$\mu_\eta(g_\delta, A_h) \leq \hat{\mu}_\eta(g_\delta, A_h) \tag{1.253}$$

holds in any case. □

In the following, we shall always admit that the generalized solution $f^\dagger$ of the exact problem exists (otherwise, clinging to our previous introductory treatment of the theory of regularization, the study of the convergence itself of any regularization algorithm would be meaningless): this implies that

$$\mu = \left\|Af^\dagger - g\right\|_Y. \tag{1.254}$$

1.8.2. The incompatible case

We are now going to illustrate the generalized discrepancy principle in the most general (i.e. *incompatible*[28]) case, in which the exact datum $g \in Y$ may not belong to the range $\mathcal{R}(A)$ of the exact operator A. We begin by recalling the following lemma (for a proof, see [67]).

Lemma 1.8.2. *The following relations hold:*

$$\hat{\mu}_\eta(g_\delta, A_h) \geq \mu; \tag{1.255}$$

$$\lim_{\eta \to 0^+} \sup_{g_\delta, A_h} \{\hat{\mu}_\eta(g_\delta, A_h) \mid \|g_\delta - g\|_Y \leq \delta;\ \|A_h - A\| \leq h\} = \mu. \tag{1.256}$$

However, we point out that, in general, $\hat{\mu}_\eta(g_\delta, A_h)$ may not be computed exactly, but rather with error $\kappa_1 \geq 0$, which is supposed to match with the noise η, in the sense that $\kappa_1 = \kappa_1(\eta) \to 0$ as $\eta \to 0^+$ (for example, $\kappa_1(\eta) \equiv \kappa_1(\delta, h) := \delta + h$). We shall denote with $\hat{\mu}_\eta^{\kappa_1}(g_\delta, A_h)$ the approximate estimate of $\hat{\mu}_\eta(g_\delta, A_h)$ and assume that

$$\hat{\mu}_\eta(g_\delta, A_h) \leq \hat{\mu}_\eta^{\kappa_1}(g_\delta, A_h) \leq \hat{\mu}_\eta(g_\delta, A_h) + \kappa_1. \tag{1.257}$$

We easily observe that, if $\kappa_1(\eta) \to 0$ as $\eta \to 0^+$, from relations (1.256) and (1.257) it immediately follows:

$$\lim_{\eta \to 0^+} \sup_{g_\delta, A_h} \left\{\hat{\mu}_\eta^{\kappa_1}(g_\delta, A_h) \mid \|g_\delta - g\|_Y \leq \delta;\ \|A_h - A\| \leq h\right\} = \mu. \tag{1.258}$$

Definition 1.8.2. *The function of α defined as*

$$\hat{\rho}_\eta^{\kappa_1}(\alpha) := \|A_h f_\alpha^\eta - g_\delta\|_Y^2 - \left(\delta + h\|f_\alpha^\eta\|_X + \hat{\mu}_\eta^{\kappa_1}(g_\delta, A_h)\right)^2, \tag{1.259}$$

i.e., recalling definitions (1.246) and (1.247),

$$\hat{\rho}_\eta^{\kappa_1}(\alpha) := \beta_\eta(\alpha) - \left(\delta + h\sqrt{\gamma_\eta(\alpha)} + \hat{\mu}_\eta^{\kappa_1}(g_\delta, A_h)\right)^2, \tag{1.260}$$

is called generalized discrepancy (for the incompatible case).

Theorem 1.8.3. *The generalized discrepancy $\hat{\rho}_\eta^{\kappa_1}(\alpha)$ has the following properties:*

1. *$\hat{\rho}_\eta^{\kappa_1}(\alpha)$ is continuous and monotonically nondecreasing in $(0, +\infty)$;*
2. *$\exists\, \hat{\rho}_{\eta,\infty}^{\kappa_1} := \lim_{\alpha \to +\infty} \hat{\rho}_\eta^{\kappa_1}(\alpha) = \|g_\delta\|_Y^2 - \left(\delta + \hat{\mu}_\eta^{\kappa_1}(g_\delta, A_h)\right)^2$;*
3. *$\exists\, \hat{\rho}_{\eta,0}^{\kappa_1} := \lim_{\alpha \to 0^+} \hat{\rho}_\eta^{\kappa_1}(\alpha) \leq [\mu_\eta(g_\delta, A_h)]^2 - \left(\delta + \hat{\mu}_\eta^{\kappa_1}(g_\delta, A_h)\right)^2 \begin{cases} < 0 & \text{if } \delta > 0 \\ \leq 0 & \text{in any case} \end{cases}$;*

[28] We point out that, in the current framework, the incompatibility is possible, not necessary.

4. if there exists (in general, not unique) an $\alpha_1^* \in (0,+\infty)$ such that $\hat{\rho}_\eta^{\kappa_1}(\alpha_1^*) = 0$, the corresponding regularized solution $f_{\alpha_1^*}^\eta$ is uniquely defined.

Proof. Properties No 1, 2 and 3 follow straightforwardly from the previous definitions 1.8.1, 1.8.2, remark 1.8.1 and lemma 1.8.1.

With regard to property No 4, we observe that since $\hat{\rho}_\eta^{\kappa_1}(\alpha)$ is not, in general, strictly monotonic, α_1^* can be defined in a non-unique manner. More precisely, the set of roots of equation $\hat{\rho}_\eta^{\kappa_1}(\alpha) = 0$ fills, in general, a certain interval $[\alpha_{1,m}^*, \alpha_{1,M}^*]$. Hence, if $\hat{\rho}_\eta^{\kappa_1}(\alpha)$ is zero on such an interval, the (not strict) monotonicity of $\beta_\eta(\alpha)$ and $\gamma_\eta(\alpha)$ implies that these functions are constant on the same interval. Now, let us consider $\alpha_{1,a}^*$, $\alpha_{1,b}^* \in [\alpha_{1,m}^*, \alpha_{1,M}^*]$ (with $\alpha_{1,a}^* \neq \alpha_{1,b}^*$) and the corresponding regularized solutions:

$$f_{\alpha_{1,a}^*}^\eta = \operatorname{argmin}\left[\|A_h f - g_\delta\|_Y^2 + \alpha_{1,a}^* \|f\|_X^2\right], \tag{1.261}$$

$$f_{\alpha_{1,b}^*}^\eta = \operatorname{argmin}\left[\|A_h f - g_\delta\|_Y^2 + \alpha_{1,b}^* \|f\|_X^2\right]. \tag{1.262}$$

Since it holds

$$\beta_\eta(\alpha_{1,a}^*) = \beta_\eta(\alpha_{1,b}^*), \quad \text{i.e.} \quad \left\|A_h f_{\alpha_{1,a}^*}^\eta - g_\delta\right\|_Y^2 = \left\|A_h f_{\alpha_{1,b}^*}^\eta - g_\delta\right\|_Y^2, \tag{1.263}$$

and

$$\gamma_\eta(\alpha_{1,a}^*) = \gamma_\eta(\alpha_{1,b}^*), \quad \text{i.e.} \quad \left\|f_{\alpha_{1,a}^*}^\eta\right\|_X^2 = \left\|f_{\alpha_{1,b}^*}^\eta\right\|_X^2, \tag{1.264}$$

it follows immediately that

$$\left\|A_h f_{\alpha_{1,a}^*}^\eta - g_\delta\right\|_Y^2 + \alpha_{1,a}^* \left\|f_{\alpha_{1,a}^*}^\eta\right\|_X^2 = \left\|A_h f_{\alpha_{1,b}^*}^\eta - g_\delta\right\|_Y^2 + \alpha_{1,a}^* \left\|f_{\alpha_{1,b}^*}^\eta\right\|_X^2, \tag{1.265}$$

i.e. both $f_{\alpha_{1,a}^*}^\eta$ and $f_{\alpha_{1,b}^*}^\eta$ minimize the functional $\|A_h f - g_\delta\|_Y^2 + \alpha_{1,a}^* \|f\|_X^2$; but we already know that the minimum point of such a functional is unique[29]: hence, $f_{\alpha_{1,a}^*}^\eta = f_{\alpha_{1,b}^*}^\eta$. ■

Remark 1.8.2. Since the function $\hat{\rho}_\eta^{\kappa_1}(\alpha)$ is continuous and monotonically nondecreasing in $(0,+\infty)$, a sufficient condition for the existence of an $\alpha_1^* = \alpha_1^*(\eta, g_\delta, A_h) \in (0,+\infty)$ such that $\hat{\rho}_\eta^{\kappa_1}(\alpha_1^*(\eta, g_\delta, A_h)) = 0$ is that

$$\hat{\rho}_{\eta,\infty}^{\kappa_1} > 0 \qquad \text{and} \qquad \hat{\rho}_{\eta,0}^{\kappa_1} < 0. \tag{1.266}$$

Actually, the following theorem shows that the first of conditions (1.266) implies the second one[30], as well as, even more, the strict monotonicity of $\hat{\rho}_\eta^{\kappa_1}(\alpha)$ and, consequently, the uniqueness of the solution to the equation $\hat{\rho}_\eta^{\kappa_1}(\alpha) = 0$. □

[29] Cf. theorems 1.7.2 and 1.7.3, which, as already observed soon below definition (1.173), hold also in the case of noisy operators.

[30] Provided that $\delta^2 + h^2 \neq 0$.

Theorem 1.8.4. *Given the noisy version $A_h f = g_\delta$ of the exact problem $Af = g$ (with, as usual, $\|g_\delta - g\|_Y \leq \delta$, $\|A_h - A\| \leq h$), let us assume that $\hat{\rho}^{\kappa_1}_{\eta,\infty} > 0$, i.e., recalling statement No 2 in theorem 1.8.3, that*

$$\|g_\delta\|_Y > \delta + \hat{\mu}^{\kappa_1}_\eta(g_\delta, A_h). \tag{1.267}$$

Then the generalized discrepancy function $\hat{\rho}^{\kappa_1}_\eta(\alpha)$ is strictly increasing monotonic in $(0, +\infty)$ and, if $h > 0 \vee \delta > 0$, it holds $\hat{\rho}^{\kappa_1}_{\eta,0} < 0$.

Proof. First of all, we have already shown, without using condition (1.267) (see statement No 3 of theorem 1.8.3), that if $\delta > 0$, then $\hat{\rho}^{\kappa_1}_{\eta,0} < 0$; however, for sake of completeness (see relation (1.275) at the end of the current proof), we have repeated such a statement here. From now on, we shall make no assumptions about δ.

Starting from hypothesis (1.267) and recalling relations (1.257), (1.253) as well as definition (1.249), we can write the following chain of equalities or inequalities:

$$\|g_\delta\|_Y > \delta + \hat{\mu}^{\kappa_1}_\eta(g_\delta, A_h) \geq \delta + \hat{\mu}_\eta(g_\delta, A_h) \geq \delta + \mu_\eta(g_\delta, A_h) = \delta + \inf_{f \in X} \|A_h f - g_\delta\|_Y , \tag{1.268}$$

i.e.

$$\|g_\delta\|_Y > \delta + \inf_{f \in X} \|A_h f - g_\delta\|_Y . \tag{1.269}$$

The previous inequality implies that $g_\delta \notin \mathcal{R}(A_h)^\perp$; indeed, if, by absurd, it were $g_\delta \in \mathcal{R}(A_h)^\perp$, then the generalized solution $f^\dagger_\eta$ of the noisy problem $A_h f = g_\delta$ would exist and would be zero: then we would have $\inf_{f \in X} \|A_h f - g_\delta\|_Y = \|g_\delta\|_Y$, which, substituted into (1.269), gives the following absurd:

$$\|g_\delta\|_Y > \delta + \|g_\delta\|_Y. \tag{1.270}$$

Hence, if we denote with P_h the orthogonal projection onto $\overline{\mathcal{R}(A_h)}$, we have that

$$P_h\, g_\delta \neq 0. \tag{1.271}$$

This implies, in particular, that there exists no value of the regularization parameter α such that the corresponding regularized solution is zero. Indeed, let us suppose, by absurd, that there exists a value $\tilde{\alpha}$ of α such that $f^\eta_{\tilde{\alpha}} = 0$: then, recalling the Euler equation (1.174), we would get $A^*_h\, g_\delta = 0$, i.e.

$$g_\delta \in \mathcal{N}(A^*_h) = \mathcal{R}(A_h)^\perp, \tag{1.272}$$

which is in contradiction with (1.271). Hence, since it holds

$$f^\eta_\alpha \neq 0 \qquad \forall \alpha > 0, \tag{1.273}$$

we have, by virtue of statement No 3 in lemma 1.8.1, that $\gamma_\eta(\alpha)$ is strictly decreasing monotonic in $(0, +\infty)$, while $\beta_\eta(\alpha)$ is strictly increasing monotonic in $(0, +\infty)$: remembering definition (1.260), this suffices to conclude the strictly increasing monotonicity of $\hat{\rho}^{\kappa_1}_\eta(\alpha)$.

Moreover, definition (1.246) (i.e. $\gamma_\eta(\alpha) := \|f_\alpha^\eta\|_X^2$) and inequality (1.273) together imply that $\gamma_\eta(\alpha) > 0 \ \forall \alpha > 0$; hence, by virtue of the strictly decreasing monotonicity of $\gamma_\eta(\alpha)$ in $(0, +\infty)$, it holds:

$$\exists\, \gamma_0 := \lim_{\alpha \to 0^+} \gamma_\eta(\alpha), \tag{1.274}$$

with $\gamma_0 \in \mathbb{R}^+$ or $\gamma_0 = +\infty$.

Finally, recalling statements No 6 of lemma 1.8.1 and No 3 of theorem 1.8.3, as well as definitions (1.249), (1.260), we get

$$\exists\, \hat\rho_{\eta,0}^{\kappa_1} := \lim_{\alpha \to 0^+} \hat\rho_\eta^{\kappa_1}(\alpha) = [\mu_\eta(g_\delta, A_h)]^2 - \left(\delta + h\sqrt{\gamma_0} + \hat\mu_\eta^{\kappa_1}(g_\delta, A_h)\right)^2 \begin{cases} < 0 & \text{if } \delta > 0 \vee h > 0\,; \\ \leq 0 & \text{if } \delta = 0 \wedge h = 0\,. \end{cases} \tag{1.275}$$

This concludes the proof. ■

Remark 1.8.3. Obviously, the case $\delta = h = 0$ requires, in principle, no regularization: indeed, in such a case the exact problem is not only a theoretical reference, but it is actually at disposal also for computational purposes, so that we can directly determine its generalized solution $f^\dagger$, which we always admit to be existing, as pointed out just before relation (1.254). As regards the other possible cases, in which $\delta > 0 \vee h > 0$ and, consequently, regularization is needed, we can summarize remark 1.8.2 and theorem 1.8.4 saying that condition (1.267) is sufficient for the existence and the uniqueness of the zero of the generalized discrepancy function $\hat\rho_\eta^{\kappa_1}(\alpha)$. □

We can now state the *generalized discrepancy principle* (for the incompatible case) as follows. Given the noisy version

$$A_h f = g_\delta \tag{1.276}$$

of the exact problem $Af = g$ (with, as usual, $\|g_\delta - g\|_Y \leq \delta$, $\|A_h - A\| \leq h$),

1. if it holds

$$\|g_\delta\|_Y \leq \delta + \hat\mu_\eta^{\kappa_1}(g_\delta, A_h), \tag{1.277}$$

let $f^\eta = 0$ be the selected approximation of the generalized solution $f^\dagger$ of the exact problem;

2. if it holds

$$\|g_\delta\|_Y > \delta + \hat\mu_\eta^{\kappa_1}(g_\delta, A_h), \tag{1.278}$$

there exists a unique $\alpha_1^*(\eta, g_\delta, A_h) > 0$ such that $\hat\rho_\eta^{\kappa_1}(\alpha_1^*(\eta, g_\delta, A_h)) = 0$ and then we take $f^\eta_{\alpha_1^*(\eta, g_\delta, A_h)}$ as approximation of $f^\dagger$.

Now, let us put:

$$f^\eta_{[\alpha_1^*(\eta, g_\delta, A_h)]} := \begin{cases} f^\eta = 0 & \text{if } \|g_\delta\|_Y \leq \delta + \hat\mu_\eta^{\kappa_1}(g_\delta, A_h) \\ f^\eta_{\alpha_1^*(\eta, g_\delta, A_h)} & \text{if } \|g_\delta\|_Y > \delta + \hat\mu_\eta^{\kappa_1}(g_\delta, A_h). \end{cases} \tag{1.279}$$

Theorem 1.8.5. *The generalized discrepancy principle (for the incompatible case) is a regularizing algorithm*[31] *for solving $Af = g$, that is, remembering definition 1.6.1, the following limits hold:*

$$\lim_{\eta\to 0^+} \sup_{g_\delta, A_h} \left\{ \left\| f^\eta_{[\alpha_1^*(\eta, g_\delta, A_h)]} - f^\dagger \right\|_X \mid \|g_\delta - g\|_Y \le \delta;\ \|A_h - A\| \le h;\ \hat\rho_\eta^{\kappa_1}(\alpha_1^*(\eta, g_\delta, A_h)) = 0 \right\} = 0, \tag{1.280}$$

$$\lim_{\eta\to 0^+} \sup_{g_\delta, A_h} \left\{ \alpha_1^*(\eta, g_\delta, A_h) \mid \|g_\delta - g\|_Y \le \delta;\ \|A_h - A\| \le h;\ \hat\rho_\eta^{\kappa_1}(\alpha_1^*(\eta, g_\delta, A_h)) = 0 \right\} = 0. \tag{1.281}$$

Proof. If $f^\dagger = 0$, then, using the triangle inequality and remembering relations (1.254), (1.255), (1.257) and (1.170), we get:

$$\|g_\delta\|_Y = \left\| Af^\dagger - g + g - g_\delta \right\|_Y \le \mu + \delta \le \hat\mu_\eta^{\kappa_1}(u_\delta, A_h) + \delta; \tag{1.282}$$

hence inequality (1.277) holds and consequently we take $f^\eta = 0$ as our approximation of $f^\dagger$, which is zero too, and nothing else needs to be proved.

If $f^\dagger \neq 0$, then, recalling equality (1.254) and the uniqueness of the generalized solution, we have

$$\mu = \left\| Af^\dagger - g \right\|_Y < \|A(0) - g\|_Y = \|g\|_Y, \tag{1.283}$$

i.e.

$$\mu < \|g\|_Y. \tag{1.284}$$

On the other hand, since $|\|g_\delta\|_Y - \|g\|_Y| \le \|g_\delta - g\|_Y \le \delta$, it holds:

$$\lim_{\delta\to 0} \inf_{g_\delta} \{\|g_\delta\|_Y \mid \|g_\delta - g\|_Y \le \delta\} = \|g\|_Y, \tag{1.285}$$

while, remembering relation (1.258), we immediately get

$$\lim_{\eta\to 0^+} \sup_{g_\delta, A_h} \left\{ \left(\delta + \hat\mu_\eta^{\kappa_1}(g_\delta, A_h)\right) \mid \|g_\delta - g\|_Y \le \delta;\ \|A_h - A\| \le h \right\} = \mu. \tag{1.286}$$

Hence, taking into account relations (1.284), (1.285) and (1.286), we easily see that condition (1.278) holds, at least for sufficiently small η; then, for vanishing η, we can actually take $f^\eta_{\alpha_1^*(\eta, g_\delta, A_h)}$ as approximation of $f^\dagger$ and prove limit (1.280) in the case $f^\eta_{[\alpha_1^*(\eta, g_\delta, A_h)]} = f^\eta_{\alpha_1^*(\eta, g_\delta, A_h)}$. Moreover, by virtue of theorem 1.7.7, the proof of (1.280) in such a case implies that limit (1.281) holds too.

Now, let

$$\mathcal{F} := \{(\delta_n, h_n, g_{\delta_n}, A_{h_n}) \mid n \in \mathbb{N}\} \tag{1.287}$$

be any set of 4-uples $(\delta_n, h_n, g_{\delta_n}, A_{h_n})$ such that

[31] To this purpose, see also the following remark 1.8.4.

1. $\{(\delta_n, h_n)\}_{n=0}^{\infty}$ is a sequence such that $\delta_n > 0$, $h_n > 0$ $\forall n \in \mathbb{N}$ and $\lim_{n\to\infty}(\delta_n, h_n) = (0,0)$; for notational convenience, we shall denote this sequence with $\{\eta_n\}_{n=0}^{\infty}$ and we shall write $\eta_n > 0$ $\forall n \in \mathbb{N}$ and $\lim_{n\to\infty} \eta_n = 0$;

2. $g_{\delta_n} \in Y$ $\forall n \in \mathbb{N}$ and
$$\|g_{\delta_n} - g\|_Y \leq \delta_n \qquad \forall n \in \mathbb{N}; \tag{1.288}$$

3. $A_{h_n} \in \mathcal{B}(X,Y)$ $\forall n \in \mathbb{N}$ and
$$\|A_{h_n} - A\| \leq h_n \qquad \forall n \in \mathbb{N}; \tag{1.289}$$

4. $\forall n \in \mathbb{N}$, $\eta_n > 0$ is taken small enough, in such a way that $\forall n \in \mathbb{N}$ there exists a (unique) $\alpha^*_{1,n} := \alpha^*_1(\eta_n, g_{\delta_n}, A_{h_n}) > 0$ satisfying $\hat{\rho}^{\kappa_1}_{\eta_n}(\alpha^*_{1,n}) = 0$.

Hence, $\forall\, (\delta_n, h_n, g_{\delta_n}, A_{h_n}) \in \mathcal{F}$, we have a (unique) $\alpha^*_{1,n}$ satisfying $\hat{\rho}^{\kappa_1}_{\eta_n}(\alpha^*_{1,n}) = 0$ and, consequently, we can consider the corresponding regularized solution, defined as

$$f^{\eta_n}_{\alpha^*_{1,n}} := \mathrm{argmin}\Phi^{\eta_n}_{\alpha^*_{1,n}}[f; g_{\delta_n}]. \tag{1.290}$$

By virtue of (1.290) and recalling definition (1.172), we get

$$\left\|A_{h_n} f^{\eta_n}_{\alpha^*_{1,n}} - g_{\delta_n}\right\|^2_Y + \alpha^*_{1,n}\left\|f^{\eta_n}_{\alpha^*_{1,n}}\right\|^2_X \leq \left\|A_{h_n} f^\dagger - g_{\delta_n}\right\|^2_Y + \alpha^*_{1,n}\left\|f^\dagger\right\|^2_X. \tag{1.291}$$

Moreover, since $\alpha^*_{1,n}$ is such that $\hat{\rho}^{\kappa_1}_{\eta_n}(\alpha^*_{1,n}) = 0$, remembering definition 1.8.2 we have

$$\left\|A_{h_n} f^{\eta_n}_{\alpha^*_{1,n}} - g_{\delta_n}\right\|^2_Y = \left(\delta_n + h_n\left\|f^{\eta_n}_{\alpha^*_{1,n}}\right\|_X + \hat{\mu}^{\kappa_1}_{\eta_n}(g_{\delta_n}, A_{h_n})\right)^2, \tag{1.292}$$

while from the triangle inequality, together with the usual error bounds (1.288), (1.289) and relation (1.254), it follows that

$$\begin{aligned}\left(\left\|A_{h_n} f^\dagger - g_{\delta_n}\right\|_Y\right)^2 &= \left(\left\|(A_{h_n} - A)f^\dagger + g - g_{\delta_n} + Af^\dagger - g\right\|_Y\right)^2 \leq \\ &\leq \left(\left\|(A_{h_n} - A)f^\dagger\right\|_Y + \|g - g_{\delta_n}\|_Y + \left\|Af^\dagger - g\right\|_Y\right)^2 \leq \\ &\leq \left(h_n\left\|f^\dagger\right\|_X + \delta_n + \mu\right)^2. \end{aligned} \tag{1.293}$$

Substituting (1.292) and (1.293) into (1.291), we get

$$\left(\delta_n + h_n\left\|f^{\eta_n}_{\alpha^*_{1,n}}\right\|_X + \hat{\mu}^{\kappa_1}_{\eta_n}(g_{\delta_n}, A_{h_n})\right)^2 + \alpha^*_{1,n}\left\|f^{\eta_n}_{\alpha^*_{1,n}}\right\|^2_X \leq \left(h_n\left\|f^\dagger\right\|_X + \delta_n + \mu\right)^2 + \alpha^*_{1,n}\left\|f^\dagger\right\|^2_X. \tag{1.294}$$

Now, as we have already seen in the last of inequalities (1.282), it holds

$$\mu \leq \hat{\mu}^{\kappa_1}_{\eta}(u_{\delta_n}, A_{h_n}); \tag{1.295}$$

then, from relation (1.294), we immediately get (changing the order of some terms):

$$\left(\delta_n + \mu + h_n \left\| f^{\eta_n}_{\alpha^*_{1,n}} \right\|_X\right)^2 + \alpha^*_{1,n} \left\| f^{\eta_n}_{\alpha^*_{1,n}} \right\|^2_X \leq \left(\delta_n + \mu + h_n \left\| f^\dagger \right\|_X\right)^2 + \alpha^*_{1,n} \left\| f^\dagger \right\|^2_X . \tag{1.296}$$

Now, the real function $\psi : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ defined as $\psi(t) := (\delta_n + \mu + h_n t)^2 + \alpha^*_{1,n} t^2$ is easily seen to be strictly increasing monotonic; then, from (1.296) (regarding t as $\|f\|_X$, with $f \in X$), we have

$$\left\| f^{\eta_n}_{\alpha^*_{1,n}} \right\|_X \leq \left\| f^\dagger \right\|_X \qquad \forall n \in \mathbb{N}. \tag{1.297}$$

Relation (1.297) means that the sequence $\left\{ f^{\eta_n}_{\alpha^*_{1,n}} \right\}_{n=0}^{\infty}$ is bounded in the Hilbert space X, and therefore [13] it has a subsequence $\left\{ f^{\eta_{n(k)}}_{\alpha^*_{1,n(k)}} \right\}_{k=0}^{\infty}$ (which we shall denote, for notational convenience, with $\left\{ f^{\eta_k}_{\alpha^*_{1,k}} \right\}_{k=0}^{\infty}$) that it is weakly convergent to an element $f^* \in X$, i.e.

$$f^{\eta_k}_{\alpha^*_{1,k}} \rightharpoonup f^*. \tag{1.298}$$

Furthermore, using the weak lower semicontinuity of the norm $\|\cdot\|_X$ [7] and inequality (1.297), we can easily get:

$$\left\| f^* \right\|_X \leq \liminf_{k\to\infty} \left\| f^{\eta_k}_{\alpha^*_{1,k}} \right\|_X \leq \limsup_{k\to\infty} \left\| f^{\eta_k}_{\alpha^*_{1,k}} \right\|_X \leq \left\| f^\dagger \right\|_X . \tag{1.299}$$

Since also the functional $x \mapsto \|Ax - g\|_Y$ is weakly lower semicontinuous (see proof No 2 of theorem 1.7.4), we get

$$\left\| Af^* - g \right\|_Y \leq \liminf_{k\to\infty} \left\| Af^{\eta_k}_{\alpha^*_{1,k}} - g \right\|_Y \leq \limsup_{k\to\infty} \left\| Af^{\eta_k}_{\alpha^*_{1,k}} - g \right\|_Y . \tag{1.300}$$

Using, in particular, the triangle inequality and relations (1.288), (1.289), (1.292), (1.297), we have the following inequalities:

$$\begin{aligned}
\left\| Af^{\eta_k}_{\alpha^*_{1,k}} - g \right\|_Y &\leq \left\| Af^{\eta_k}_{\alpha^*_{1,k}} - A_{h_k} f^{\eta_k}_{\alpha^*_{1,k}} \right\|_Y + \left\| A_{h_k} f^{\eta_k}_{\alpha^*_{1,k}} - g_{\delta_k} \right\|_Y + \left\| g_{\delta_k} - g \right\|_Y \leq \\
&\leq h_k \left\| f^{\eta_k}_{\alpha^*_{1,k}} \right\|_X + \left(\delta_k + h_k \left\| f^{\eta_k}_{\alpha^*_{1,k}} \right\|_X + \hat{\mu}^{\kappa_1}_{\eta_k}(g_{\delta_k}, A_{h_k}) \right) + \delta_k \leq \\
&\leq h_k \left\| f^\dagger \right\|_X + \left(\delta_k + h_k \left\| f^\dagger \right\|_X + \hat{\mu}^{\kappa_1}_{\eta_k}(g_{\delta_k}, A_{h_k}) \right) + \delta_k = \\
&= 2 \left(\delta_k + h_k \left\| f^\dagger \right\|_X \right) + \hat{\mu}^{\kappa_1}_{\eta_k}(g_{\delta_k}, A_{h_k}),
\end{aligned} \tag{1.301}$$

whence, remembering relation (1.258), we immediately get

$$\limsup_{k\to\infty} \left\| Af^{\eta_k}_{\alpha^*_{1,k}} - g \right\|_Y \leq \mu. \tag{1.302}$$

Substituting this result into (1.300), we find that

$$\left\| Af^* - g \right\|_Y \leq \mu; \tag{1.303}$$

if we now compare inequality (1.303) with relation (1.254), we straightforwardly deduce that equality has to hold, i.e.:

$$\|Af^* - g\|_Y = \mu. \tag{1.304}$$

Moreover, we remember that the generalized solution is the unique minimum norm solution and, on the other hand, we have found (see (1.299)) $\|f^*\|_X \leq \|f^\dagger\|_X$: hence, we conclude that $f^* = f^\dagger$ and so, again from (1.299), we immediately get

$$\lim_{k\to\infty} \left\| f^{\eta_k}_{\alpha^*_{1,k}} \right\|_X = \left\| f^\dagger \right\|_X . \tag{1.305}$$

Summing up, by virtue of the H-property (see footnote 22), the two relations (1.298) (with f^* replaced by $f^\dagger$) and (1.305) imply that

$$\lim_{k\to\infty} \left\| f^{\eta_k}_{\alpha^*_{1,k}} - f^\dagger \right\|_X = 0. \tag{1.306}$$

Now, it is not difficult to see that equation (1.306) implies that

$$\lim_{\eta\to 0^+} \sup_{g_\delta, A_h} \left\{ \left\| f^{\eta}_{\alpha^*_1(\eta, g_\delta, A_h)} - f^\dagger \right\|_X \mid \|g_\delta - g\|_Y \leq \delta;\ \|A_h - A\| \leq h;\ \hat{\rho}^{\kappa_1}_\eta(\alpha^*_1(\eta, g_\delta, A_h)) = 0 \right\} = 0, \tag{1.307}$$

which is just our first thesis, i.e. relation (1.280). In fact, let us suppose, by absurd, that (1.307) is not true; then, there exist an $\varepsilon > 0$ and a sequence[32] $\{\eta_n(\varepsilon)\}_{n=0}^\infty \equiv \{\eta_n\}_{n=0}^\infty$, with $\eta_n > 0\ \forall n \in \mathbb{N}$, such that $\lim_{n\to\infty} \eta_n = 0$ and $\forall n \in \mathbb{N}$ it holds

$$\sup_{g_{\delta_n}, A_{h_n}} \left\{ \left\| f^{\eta_n}_{\alpha^*_{1,n}} - f^\dagger \right\|_X \mid \|g_{\delta_n} - g\|_Y \leq \delta_n;\ \|A_{h_n} - A\| \leq h_n;\ \hat{\rho}^{\kappa_1}_{\eta_n}(\alpha^*_1(\eta_n, g_{\delta_n}, A_{h_n})) = 0 \right\} > \varepsilon. \tag{1.308}$$

Inequality (1.308) clearly implies that $\forall n \in \mathbb{N}$ there exists a 4-uple $(\delta_n, h_n, \tilde{g}_{\delta_n}, \tilde{A}_{h_n}) \in \mathcal{F}$ and, consequently, a (unique) $\tilde{\alpha}^*_{1,n} := \alpha^*_1(\eta_n, \tilde{g}_{\delta_n}, \tilde{A}_{h_n})$ satisfying $\hat{\rho}^{\kappa_1}_{\eta_n}(\tilde{\alpha}^*_{1,n}) = 0$, such that the corresponding regularized solution

$$\tilde{f}^{\eta_n}_{\tilde{\alpha}^*_{1,n}} := \left(\tilde{A}^*_{h_n} \tilde{A}_{h_n} + \tilde{\alpha}^*_{1,n} I \right)^{-1} \tilde{A}^*_{h_n} \tilde{g}_{\delta_n} \tag{1.309}$$

verifies the inequality $\left\| \tilde{f}^{\eta_n}_{\tilde{\alpha}^*_{1,n}} - f^\dagger \right\|_X \geq \frac{\varepsilon}{2}$; in other terms, there exists a sequence $\left\{ \tilde{f}^{\eta_n}_{\tilde{\alpha}^*_{1,n}} \right\}_{n=0}^\infty$ such that

$$\left\| \tilde{f}^{\eta_n}_{\tilde{\alpha}^*_{1,n}} - f^\dagger \right\|_X \geq \frac{\varepsilon}{2} \qquad \forall n \in \mathbb{N}. \tag{1.310}$$

On the other hand, as we have seen above, from the sequence $\left\{ \tilde{f}^{\eta_n}_{\tilde{\alpha}^*_{1,n}} \right\}_{n=0}^\infty$ itself we can extract a subsequence, say $\left\{ \tilde{f}^{\eta_k}_{\tilde{\alpha}^*_{1,k}} \right\}_{k=0}^\infty$, that verifies relation (1.306), i.e. such that

$$\lim_{k\to\infty} \left\| \tilde{f}^{\eta_k}_{\tilde{\alpha}^*_{1,k}} - f^\dagger \right\|_X = 0. \tag{1.311}$$

Relations (1.310) and (1.311) are obviously contradictory, so equation (1.307) must hold. ■

[32] See also the last part of proof No 2 of theorem 1.7.4

Remark 1.8.4. Strictly speaking, we should admit that the generalized discrepancy principle introduced above is not a regularization algorithm, in the sense that it does not completely fulfil the requirements of definition 1.6.1: indeed, as we have just seen also in the proof of the previous theorem 1.8.5, such requirements are fully satisfied as long as $f^\dagger \neq 0$ and the noise level η is small enough; otherwise, we simply choose $f^\eta = 0$ as our approximation of $f^\dagger$ and, in such a case, it is even meaningless to speak of a value for the regularization parameter; at most, remembering definition (1.246) and limit No 4 in lemma 1.8.1, one might imagine that when condition (1.278) is not satisfied, the "regularized" solution $f^\eta = 0$ is obtained by taking the limit as $\alpha \to +\infty$ of the really Tikhonov regularized solution with a generic α (cf. definition (1.173)), i.e.

$$f_\alpha^\eta := \operatorname{argmin} \Phi_\alpha^\eta[f; g_\delta], \tag{1.312}$$

but the requirements of definition 1.6.1 would be not fulfilled anyway. However, these are minor details: as a matter of fact, the generalized discrepancy principle is very useful in the applications and its properties are, in practice, more than sufficient to regularize an inverse problem: hence we shall always commit a slight abuse of language and keep on calling it a *regularization algorithm* in the sense of definition 1.6.1. □

Remark 1.8.5. Of course, any numerical implementation of the generalized discrepancy principle requires the computation of an estimate $\hat{\mu}_\eta^{\kappa_1}(g_\delta, A_h)$ of the modified incompatibility measure $\hat{\mu}_\eta(g_\delta, A_h)$ (cf. relations (1.250) and (1.257)). For sake of brevity, here we shall not face such a problem: we simply refer to [67] (pp. 58-63), in which some theorems and consequent numerical recipes are proposed in order to satisfactorily solve it. □

1.8.3. The compatible case

In the previous subsection we have never assumed that the exact datum $g \in Y$ should belong to the range $\mathcal{R}(A)$ of the exact operator A. However, in the case in which it actually holds $g \in \mathcal{R}(A)$ (i.e. in the *compatible*[33] case), one can formulate an alternative generalized discrepancy principle, which is a little simpler than the previous one and will be sketched out in the current subsection. Owing to the substantial analogy with the incompatible case and for sake of brevity, we shall omit all the proofs (except the one, particularly simple, of the following lemma 1.8.6), referring to [67] for them (and, more generally, for a deeper treatment).

Lemma 1.8.6. *If $g \in \mathcal{R}(A)$, the following relation holds:*

$$\lim_{\eta \to 0^+} \sup_{g_\delta, A_h} \{\mu_\eta(g_\delta, A_h) \mid \|g_\delta - g\|_Y \leq \delta;\ \|A_h - A\| \leq h\} = 0. \tag{1.313}$$

[33] We point out that the compatibility is necessary for the validity of the arguments of the current subsection.

Proof. Recalling definition (1.249), denoting, as usual, with $f^\dagger$ the generalized solution of the exact problem (so that, in the current compatible case, it holds $Af^\dagger = g$, i.e. $\mu = 0$, where μ is given by (1.254)), using the triangle inequality and remembering the usual conditions $\|g_\delta - g\|_Y \leq \delta$, $\|A_h - A\| \leq h$, we easily get:

$$\mu_\eta(g_\delta, A_h) = \inf_{f \in X} \|A_h f - g_\delta\|_Y \leq \left\|A_h f^\dagger - g_\delta\right\|_Y \leq \left\|A_h f^\dagger - A f^\dagger\right\|_Y + \|g - g_\delta\|_Y \leq h \left\|f^\dagger\right\|_X + \delta, \tag{1.314}$$

whence limit (1.313) immediately follows. ■

However, we point out that, in general, $\mu_\eta(g_\delta, A_h)$ may not be computed exactly, but rather with error $\kappa_2 \geq 0$, which is supposed to match with the noise η, in the sense that $\kappa_2 = \kappa_2(\eta) \to 0$ as $\eta \to 0^+$ (for example, $\kappa_2(\eta) \equiv \kappa_2(\delta, h) := \delta + h$). We shall denote with $\mu_\eta^{\kappa_2}(g_\delta, A_h)$ the approximate estimate of $\mu_\eta(g_\delta, A_h)$ and assume that

$$\mu_\eta(g_\delta, A_h) \leq \mu_\eta^{\kappa_2}(g_\delta, A_h) \leq \mu_\eta(g_\delta, A_h) + \kappa_2. \tag{1.315}$$

We easily observe that, if $\kappa_2(\eta) \to 0$ as $\eta \to 0^+$, from relations (1.313) and (1.315) it immediately follows:

$$\lim_{\eta \to 0^+} \sup_{g_\delta, A_h} \left\{\mu_\eta^{\kappa_2}(g_\delta, A_h) \mid \|g_\delta - g\|_Y \leq \delta;\ \|A_h - A\| \leq h\right\} = 0. \tag{1.316}$$

Definition 1.8.3. *The function of α defined as*

$$\rho_\eta^{\kappa_2}(\alpha) := \|A_h f_\alpha^\eta - g_\delta\|_Y^2 - (\delta + h\|f_\alpha^\eta\|_X)^2 - \left(\mu_\eta^{\kappa_2}(g_\delta, A_h)\right)^2, \tag{1.317}$$

i.e., recalling definitions (1.246) and (1.247),

$$\rho_\eta^{\kappa_2}(\alpha) := \beta_\eta(\alpha) - \left(\delta + h\sqrt{\gamma_\eta(\alpha)}\right)^2 - \left(\mu_\eta^{\kappa_2}(g_\delta, A_h)\right)^2, \tag{1.318}$$

is called generalized discrepancy (for the compatible case).

Then we can state the *generalized discrepancy principle* (for the compatible case) as follows. Given the noisy version

$$A_h f = g_\delta \tag{1.319}$$

of the exact problem $Af = g$ (with $g \in \mathcal{R}(A)$ and, as usual, $\|g_\delta - g\|_Y \leq \delta$, $\|A_h - A\| \leq h$),

1. if it holds

$$\|g_\delta\|_Y^2 \leq \delta^2 + \left(\mu_\eta^{\kappa_2}(g_\delta, A_h)\right)^2, \tag{1.320}$$

let $f^\eta = 0$ be the selected approximation of the generalized solution $f^\dagger$ of the exact problem;

2. if it holds

$$\|g_\delta\|_Y^2 > \delta^2 + \left(\mu_\eta^{\kappa_2}(g_\delta, A_h)\right)^2, \tag{1.321}$$

there exists a unique $\alpha_2^*(\eta, g_\delta, A_h) > 0$ such that $\rho_\eta^{\kappa_2}(\alpha_2^*(\eta, g_\delta, A_h)) = 0$ and then we take $f^\eta_{\alpha_2^*(\eta,g_\delta,A_h)}$ as approximation of $f^\dagger$.

Now, let us put:

$$f^\eta_{[\alpha_2^*(\eta,g_\delta,A_h)]} := \begin{cases} f^\eta = 0 & \text{if } \|g_\delta\|_Y^2 \le \delta^2 + \left(\mu_\eta^{\kappa_2}(g_\delta, A_h)\right)^2, \\ f^\eta_{\alpha_2^*(\eta,g_\delta,A_h)} & \text{if } \|g_\delta\|_Y^2 > \delta^2 + \left(\mu_\eta^{\kappa_2}(g_\delta, A_h)\right)^2. \end{cases} \tag{1.322}$$

Theorem 1.8.7. *The generalized discrepancy principle (for the compatible case) is a regularizing algorithm for solving $Af = g$, that is, remembering definition 1.6.1, the following limits hold:*

$$\lim_{\eta\to 0^+} \sup_{g_\delta, A_h} \left\{ \left\| f^\eta_{[\alpha_2^*(\eta,g_\delta,A_h)]} - f^\dagger \right\|_X \mid \|g_\delta - g\|_Y \le \delta;\ \|A_h - A\| \le h;\ \rho_\eta^{\kappa_2}(\alpha_2^*(\eta, g_\delta, A_h)) = 0 \right\} = 0, \tag{1.323}$$

$$\lim_{\eta\to 0^+} \sup_{g_\delta, A_h} \left\{ \alpha_2^*(\eta, g_\delta, A_h) \mid \|g_\delta - g\|_Y \le \delta;\ \|A_h - A\| \le h;\ \rho_\eta^{\kappa_2}(\alpha_2^*(\eta, g_\delta, A_h)) = 0 \right\} = 0. \tag{1.324}$$

Proof. As already hinted at the very beginning of the current subsection, we omit this proof (owing to the substantial analogy with the one of theorem 1.8.5), and directly refer to [67] for it. However, for future purpose, we point out that, instead of inequality (1.297), it now holds:

$$\left\| f^{\eta_n}_{\alpha_{2,n}^*} \right\|_X \le \left\| f^\dagger \right\|_X \qquad \forall n \in \mathbb{N}, \tag{1.325}$$

while equality (1.292) should be replaced by the following one:

$$\left\| A_{h_n} f^{\eta_n}_{\alpha_{2,n}^*} - g_{\delta_n} \right\|_Y^2 = \left(\delta_n + h_n \left\| f^{\eta_n}_{\alpha_{2,n}^*} \right\|_X \right)^2 + \left(\mu_{\eta_n}^{\kappa_2}(g_{\delta_n}, A_{h_n}) \right)^2. \tag{1.326}$$

■

Remark 1.8.6. A comment analogous to the one in remark 1.8.4 would obviously hold also for the generalized discrepancy principle in the current compatible case. □

Remark 1.8.7. Of course, analogously to what observed in remark 1.8.5, any numerical implementation of the generalized discrepancy principle (for the compatible case) requires the computation of an estimate $\mu_\eta^{\kappa_2}(g_\delta, A_h)$ of the (simple) incompatibility measure $\mu_\eta(g_\delta, A_h)$ (cf. relations (1.249) and (1.315)). However, unlike the incompatible case, an estimate of

$\mu_\eta(g_\delta, A_h)$ itself is, in principle, quite easy: indeed, by virtue of the direct sum decomposition $Y = \overline{\mathcal{R}(A_h)} \oplus \mathcal{R}(A_h)^\perp$, we can uniquely write $g_\delta = g_{1,\delta} + g_{2,\delta}$, with $g_{1,\delta} \in \overline{\mathcal{R}(A_h)}$ and $g_{2,\delta} \in \mathcal{R}(A_h)^\perp$, so that it holds:

$$\begin{aligned}\mu_\eta(g_\delta, A_h) &= \inf_{f\in X} \|A_h f - g_\delta\|_Y = \inf_{f\in X} \left(\|A_h f - g_{1,\delta}\|_Y + \|g_{2,\delta}\|_Y\right) = \\ &= \inf_{f\in X} \|A_h f - g_{1,\delta}\|_Y + \|g_{2,\delta}\|_Y = 0 + \|g_{2,\delta}\|_Y = \|g_{2,\delta}\|_Y. \end{aligned} \tag{1.327}$$

□

1.8.4. A mixed approach (in the compatible case)

From a merely theoretical viewpoint, at least in the compatible case (i.e. when $g \in \mathcal{R}(A)$) the two generalized discrepancy principles introduced above seem to be somehow equivalent, in the sense that both of them, as stated by theorems 1.8.5 and 1.8.7, satisfy the two general conditions (1.68) and (1.69), so that they actually give rise to a regularization method. However, it should be pointed out that such conditions concern the behaviour of the regularization algorithm itself for *vanishing* noise levels δ and h, while in practice δ and h never vanish, but rather they are somehow estimated and then *fixed*. Hence, the above hinted equivalence between the two generalized discrepancy principles may not hold in effective numerical implementations with fixed noise levels, since in general they provide two different values of the regularization parameter which are not equally suitable to the approximate reconstruction of $f^\dagger$ we are looking for.

To be more precise, we firstly recall that, in the (not necessarily) incompatible case, the value $\alpha_1^*(\eta, g_\delta, A_h)$ of the regularization parameter α is chosen as the unique solution to the equation $\hat\rho_\eta^{\kappa_1}(\alpha) = 0$; in other terms, recalling definition (1.260) and adopting the shorthand notation

$$\alpha_1^* := \alpha_1^*(\eta, g_\delta, A_h), \tag{1.328}$$

we have that α_1^* itself is uniquely defined by the condition

$$\beta_\eta(\alpha_1^*) = \left(\delta + h\sqrt{\gamma_\eta(\alpha_1^*)} + \hat\mu_\eta^{\kappa_1}(g_\delta, A_h)\right)^2, \tag{1.329}$$

or, equivalently,

$$\beta_\eta(\alpha_1^*) - \left(\delta + h\sqrt{\gamma_\eta(\alpha_1^*)}\right)^2 = \left(\hat\mu_\eta^{\kappa_1}(g_\delta, A_h)\right)^2 + 2\hat\mu_\eta^{\kappa_1}(g_\delta, A_h)\left(\delta + h\sqrt{\gamma_\eta(\alpha_1^*)}\right). \tag{1.330}$$

Analogously, we recall that, in the compatible case, the value $\alpha_2^*(\eta, g_\delta, A_h)$ of the regularization parameter α is chosen as the unique solution to the equation $\rho_\eta^{\kappa_2}(\alpha) = 0$; in other terms, recalling definition (1.318) and adopting the shorthand notation

$$\alpha_2^* := \alpha_2^*(\eta, g_\delta, A_h), \tag{1.331}$$

we have that α_2^* itself is uniquely defined by the condition

$$\beta_\eta(\alpha_2^*) = \left(\delta + h\sqrt{\gamma_\eta(\alpha_2^*)}\right)^2 + \left(\mu_\eta^{\kappa_2}(g_\delta, A_h)\right)^2, \tag{1.332}$$

or, equivalently,

$$\beta_\eta(\alpha_2^*) - \left(\delta + h\sqrt{\gamma_\eta(\alpha_2^*)}\right)^2 = \left(\mu_\eta^{\kappa_2}(g_\delta, A_h)\right)^2. \tag{1.333}$$

We can now observe that, by virtue of inequality (1.253), in the particular case $\kappa_1 = \kappa_2 = 0$ the right-hand side of equation (1.330) would be certainly greater than or equal to the right-hand side of equation (1.333); then, the same inequality would hold also for the respective left-hand sides of the same equations, i.e.

$$\beta_\eta(\alpha_1^*) - \left(\delta + h\sqrt{\gamma_\eta(\alpha_1^*)}\right)^2 \geq \beta_\eta(\alpha_2^*) - \left(\delta + h\sqrt{\gamma_\eta(\alpha_2^*)}\right)^2, \tag{1.334}$$

and, consequently, remembering the (strictly) increasing monotonicity of the function $\beta_\eta(\alpha) - \left(\delta + h\sqrt{\gamma_\eta(\alpha)}\right)^2$, we would get

$$\alpha_1^* \geq \alpha_2^*. \tag{1.335}$$

Coming back to the most general case, insofar as we assume that there is no relation at all between κ_1 and κ_2, inequality (1.335) cannot be proved, nor shall we use it in the following. However, it is worthwhile noticing that in practice, as intuition itself suggests, the right-hand side of equation (1.330) is strictly greater than the right-hand side of equation (1.333), so that, as a matter of fact, it always holds:

$$\alpha_1^* > \alpha_2^*. \tag{1.336}$$

Hence, a problem naturally arises: how can we choose, between α_1^* and α_2^*, the best value, i.e. the value providing the most satisfactory regularized solution? The answer to this question is not at all immediate and it mostly depends on the noise level affecting the datum and/or the operator.

In order to suggest and justify a possible approach to this problem, we are now going to summarize our experience in numerical simulation by proposing the following discussion, which, although absolutely heuristic, could be made more precise by using tools of numerical analysis and more convincing by means of tables collecting some numerical results. However, for sake of brevity, we shall limit ourselves to the following remarks.

To fix our ideas, let us consider the typical situation (occurring, e.g., in implementing the linear sampling method, as we shall see in sections 2.5 and 3.1) in which an inverse problem set in a finite-dimensional context (maybe after an appropriate discretization of a continuous problem) is studied and the noise affects only the operator (i.e. $\delta = 0$): in such a case, the exact problem can be written as

$$\mathbf{A}\mathbf{f} = \mathbf{g}, \tag{1.337}$$

where $\mathbf{A}$ is a square $N \times N$ matrix (having complex entries and regarded as a linear continuous operator from $X = \mathbb{C}^N$ to $Y = \mathbb{C}^N$), $\mathbf{g} \in \mathbb{C}^N$ is the datum in the form of a column vector with N complex-valued components and $\mathbf{f} \in \mathbb{C}^N$ is the unknown column vector (with N complex-valued components too), while the corresponding noisy version of the exact problem (1.337) is:

$$\mathbf{A}_h \mathbf{f} = \mathbf{g}, \tag{1.338}$$

where, as usual, we assume to know that $\|\mathbf{A}_h - \mathbf{A}\| \leq h$.

As a compact operator, $\mathbf{A}_h$ obviously admits its own singular representation (cf. definition 1.5.6), i.e.:

$$\mathbf{A}_h \mathbf{w} = \sum_{p=0}^{r_h - 1} \sigma_p^h (\mathbf{w}, \mathbf{u}_p^h)_{\mathbb{C}^N} \mathbf{v}_p^h \qquad \forall \mathbf{w} \in \mathbb{C}^N, \tag{1.339}$$

where r_h is the rank of $\mathbf{A}_h$, $\{\sigma_p^h, \mathbf{u}_p^h, \mathbf{v}_p^h\}_{p=0}^{r_h - 1}$ is the singular system of $\mathbf{A}_h$ and $(\cdot, \cdot)_{\mathbb{C}^N}$ is the canonical scalar product in $\mathbb{C}^N$. However, since the compact operator $\mathbf{A}_h$ is also a *matrix*, a representation like (1.339) turns out to be one the possible ways of writing its SVD (Singular Value Decomposition): for the reader's convenience, here we briefly recall this concept, referring, e.g., to [9] for all the proofs and, more generally, for a much deeper treatment.

1. In its standard formulation, the general concept of SVD of a matrix[34] is the following: *let* $\mathbf{T}$ *be a rectangular matrix* $M \times N$*, with rank* r*; then there exist a* $r \times r$ *diagonal matrix* $\mathbf{\Sigma}$*, with positive diagonal elements, and two isometric matrices* $\mathbf{U}$ *and* $\mathbf{V}$*, respectively* $M \times r$ *and* $N \times r$*, such that the following representation for* $\mathbf{T}$*, called* singular value decomposition of the matrix $\mathbf{T}$, *holds:*

$$\mathbf{T} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^*, \tag{1.340}$$

where $\mathbf{V}^*$ *denotes the adjoint, i.e. the conjugate transposed of* $\mathbf{A}$. The diagonal elements of $\mathbf{\Sigma}$ are called *singular values* of the matrix $\mathbf{T}$ and denoted (in decreasing order) with $\sigma_0 \geq \sigma_1 \geq \ldots \geq \sigma_{r-1}$; moreover[35] it holds $\|\mathbf{T}\| = \sigma_0$. We also recall that a rectangular matrix $\mathbf{V}$ is called *isometric* if it satisfies the condition $\mathbf{V}^* \mathbf{V} = \mathbf{I}_r$, where $\mathbf{I}_r$ is the $r \times r$ identity matrix.

2. If the SVD of the matrix $\mathbf{T}$ is given by (1.340), then the SVD of its generalized inverse matrix $\mathbf{T}^\dagger$ is given by

$$\mathbf{T}^\dagger = \mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{U}^* \tag{1.341}$$

and, consequently, $\|\mathbf{T}^\dagger\| = \dfrac{1}{\sigma_{r-1}}$.

[34] If not specified otherwise, all the matrices we are speaking of have *complex* entries.

[35] Cf. also remark 1.5.3.

3. A particular, but important, case is the one in which $\mathbf{T}$ is a square $N \times N$ matrix with rank $r = N$ or, equivalently, with non-zero determinant: then representations (1.340) and (1.341) keep on holding, but now $\mathbf{T}$, $\mathbf{T}^\dagger$, $\mathbf{U}$, $\mathbf{U}^*$, $\mathbf{\Sigma}$, $\mathbf{\Sigma}^{-1}$, $\mathbf{V}$ and $\mathbf{V}^*$ are all square $N \times N$ matrices and it is easy to realize that $\mathbf{T}^\dagger = \mathbf{T}^{-1}$. Moreover, a square isometric matrix is unitary: then, $\mathbf{U}$ and $\mathbf{V}$ are unitary and, consequently, their determinant is a complex number of modulus 1. Hence, recalling Binet's theorem, we immediately have:

$$|\det\mathbf{T}| = \prod_{p=0}^{N-1} \sigma_p \qquad \text{and} \qquad \left|\det\mathbf{T}^{-1}\right| = \prod_{p=0}^{N-1} \frac{1}{\sigma_p}. \tag{1.342}$$

Coming back to our exact inverse problem (1.337) and its noisy version (1.338), let us firstly assume that the determinant of the exact matrix $\mathbf{A}$ is different from zero (and, consequently, $\mathcal{R}(\mathbf{A}) = \mathbb{C}^N$): then, we are in the compatible case and, consequently, we can apply both the generalized discrepancy principles introduced above. However, it may happen that the last[36] singular values $\sigma_{N-1} \leq \sigma_{N-2} \leq \sigma_{N-3} \leq \ldots$ of the matrix $\mathbf{A}$ are extremely small, as well as much smaller than the first[37] ones $\sigma_0 \geq \sigma_1 \geq \sigma_2 \geq \ldots$, so that:

1. the modulus itself of the determinant of the matrix $\mathbf{A}$ is extremely small, i.e. "nearly" zero: this means, roughly speaking, that the rows or the columns of the matrix $\mathbf{A}$ are "nearly" linearly dependent or, equivalently, that we are in a "nearly" incompatible case;

2. the exact problem (1.337), although well-posed, is strongly ill-conditioned: indeed the condition number of the matrix $\mathbf{A}$, which is[38]

$$C(\mathbf{A}) = \|\mathbf{A}\| \, \|\mathbf{A}^{-1}\| = \frac{\sigma_0}{\sigma_{N-1}}, \tag{1.343}$$

turns out to be dramatically large, owing to the prevailing smallness of σ_{N-1} in comparison with the value of σ_0.

Now, if we begin adding noise to $\mathbf{A}$ with different and increasing levels $h_1 < h_2 < h_3 < \ldots$, we can typically observe the following effects:

1. the modulus of the determinant of the matrix increases, i.e.

$$|\det\mathbf{A}| < |\det\mathbf{A}_{h_1}| < |\det\mathbf{A}_{h_2}| < |\det\mathbf{A}_{h_3}| < \ldots, \tag{1.344}$$

and such an increase is very strong; in other terms, roughly speaking, adding noise significantly decreases the probability that the rows or the columns of the matrix are linearly dependent and, consequently, increases a lot the compatibility of the problem;

[36] "Last" is clearly to be intended with respect to the decreasing order, as hinted at the end of the previous point No 1.

[37] Also "first" is clearly to be intended with respect to the decreasing order, as hinted at the end of the previous point No 1.

[38] Cf. definition (1.10) and the previous points No 1-3.

2. while the largest singular value σ_0 remains substantially unaltered, the smallest one undergoes a (notable) increase, so that the condition number decreases, i.e.

$$C(\mathbf{A}) > C(\mathbf{A}_{h_1}) > C(\mathbf{A}_{h_2}) > C(\mathbf{A}_{h_3}) > \dots, \tag{1.345}$$

and such a decrease is very strong; in other terms, adding noise significantly reduces the ill-posedness of the problem, so that, incredibly enough, it turns out to be a rough (although completely inadequate) form of regularization.

In such a situation, it is not surprising to observe that for sufficiently small noise levels the value α_1^* provides regularized solutions which are more satisfactory than the ones provided by α_2^*: the latter value is, indeed, too small and, consequently, undersmoothing; on the contrary, for sufficiently large noise levels, it is α_2^* that provides more satisfactory regularized solutions than the ones provided by α_1^*: the latter value is, indeed, too large and, consequently, oversmoothing.

Hence the idea may arise of blending somehow the two regularization parameters[39] or the two regularized solutions, in such a way that when the noise is small or large enough, one takes into account, as a matter of fact, only the regularized solution coming from the incompatible or the compatible approach respectively, while for intermediate noise levels one considers an appropriate combination of the two. We are now going to show that this task can be accomplished in such a way that the resulting algorithm is still a regularization method.

Blending the two values of the regularization parameter

We consider the following convex combination of $\alpha_1^*(\eta, g_\delta, A_h)$ and $\alpha_2^*(\eta, g_\delta, A_h)$:

$$\alpha_b^*(\eta, g_\delta, A_h) := c(\eta)\, \alpha_1^*(\eta, g_\delta, A_h) + [1 - c(\eta)]\, \alpha_2^*(\eta, g_\delta, A_h), \tag{1.346}$$

where the function

$$\begin{array}{rccc} c: & (\mathbb{R}^+ \cup \{0\}) \times (\mathbb{R}^+ \cup \{0\}) & \longrightarrow & [0,1] \\ & \eta = (\delta, h) & \longmapsto & c(\eta) \end{array} \tag{1.347}$$

plays the role of blending-tuner; hence, if we put

$$\alpha_m^*(\eta, g_\delta, A_h) := \min\{\alpha_1^*(\eta, g_\delta, A_h), \alpha_2^*(\eta, g_\delta, A_h)\} \qquad \forall n \in \mathbb{N}, \tag{1.348}$$

$$\alpha_M^*(\eta, g_\delta, A_h) := \max\{\alpha_1^*(\eta, g_\delta, A_h), \alpha_2^*(\eta, g_\delta, A_h)\} \qquad \forall n \in \mathbb{N}, \tag{1.349}$$

it clearly holds:

$$\alpha_m^*(\eta, g_\delta, A_h) \le \alpha_b^*(\eta, g_\delta, A_h) \le \alpha_M^*(\eta, g_\delta, A_h). \tag{1.350}$$

[39] An example of such a procedure will be briefly presented at the end of section 3.1 (cf. relation (3.73)) and illustrated by figures B.11, B.12.

Moreover, we also put:

$$\nu_1 = \nu_1(\eta, g_\delta, A_h) := \delta + \hat{\mu}_\eta^{\kappa_1}(g_\delta, A_h), \tag{1.351}$$

$$\nu_2 = \nu_2(\eta, g_\delta, A_h) := \sqrt{\delta^2 + \left(\mu_\eta^{\kappa_2}(g_\delta, A_h)\right)^2}\,. \tag{1.352}$$

Then we can state the *mixed discrepancy principle* (for the compatible case) as follows. Given the noisy version

$$A_h f = g_\delta \tag{1.353}$$

of the exact problem $Af = g$ (with $g \in \mathcal{R}(A)$ and, as usual, $\|g_\delta - g\|_Y \leq \delta$, $\|A_h - A\| \leq h$),

1. if it holds
$$\|g_\delta\|_Y \leq \min\{\nu_1, \nu_2\}, \tag{1.354}$$
let $f^\eta = 0$ be the selected approximation of the generalized solution $f^\dagger$ of the exact problem;

2. if it holds
$$\nu_1 < \|g_\delta\|_Y \leq \nu_2, \tag{1.355}$$
let $f^\eta_{\alpha_1^*(\eta, g_\delta, A_h)}$ be the selected approximation of the generalized solution $f^\dagger$ of the exact problem;

3. if it holds
$$\nu_2 < \|g_\delta\|_Y \leq \nu_1, \tag{1.356}$$
let $f^\eta_{\alpha_2^*(\eta, g_\delta, A_h)}$ be the selected approximation of the generalized solution $f^\dagger$ of the exact problem;

4. if it holds
$$\|g_\delta\|_Y^2 > \max\{\nu_1, \nu_2\}, \tag{1.357}$$
let $f^\eta_{\alpha_b^*(\eta, g_\delta, A_h)}$ be the selected approximation of the generalized solution $f^\dagger$ of the exact problem.

Now, let us put:

$$f^\eta_{[\alpha_b^*(\eta, g_\delta, A_h)]} := \begin{cases} f^\eta = 0 & \text{if } \|g_\delta\|_Y \leq \min\{\nu_1, \nu_2\}, \\ f^\eta_{\alpha_1^*(\eta, g_\delta, A_h)} & \text{if } \nu_1 < \|g_\delta\|_Y \leq \nu_2, \\ f^\eta_{\alpha_2^*(\eta, g_\delta, A_h)} & \text{if } \nu_2 < \|g_\delta\|_Y \leq \nu_1, \\ f^\eta_{\alpha_b^*(\eta, g_\delta, A_h)} & \text{if } \|g_\delta\|_Y > \max\{\nu_1, \nu_2\}. \end{cases} \tag{1.358}$$

Theorem 1.8.8. *The mixed generalized discrepancy principle (for the compatible case) is a regularizing algorithm for solving $Af = g$, that is, remembering definition 1.6.1, the following limits hold:*

$$\lim_{\eta\to 0^+} \sup_{g_\delta, A_h} \left\{ \left\| f^{\eta}_{[\alpha^*_b(\eta, g_\delta, A_h)]} - f^\dagger \right\|_X \mid \|g_\delta - g\|_Y \le \delta;\ \|A_h - A\| \le h \right\} = 0, \tag{1.359}$$

$$\lim_{\eta\to 0^+} \sup_{g_\delta, A_h} \{\alpha^*_b(\eta, g_\delta, A_h) \mid \|g_\delta - g\|_Y \le \delta;\ \|A_h - A\| \le h\} = 0. \tag{1.360}$$

Proof. We have already seen, at the very beginning of the proof of theorem 1.8.5, that if $f^\dagger = 0$, then it holds (with the current notation) $\|g_\delta\|_Y \le \nu_1$. On the other hand, if we had proved also theorem 1.8.7, we would have shown that if $f^\dagger = 0$, then it also holds (with the current notation) $\|g_\delta\|_Y \le \nu_2$: indeed $f^\dagger = 0$ implies, in the compatible case, that $g = 0$ and, consequently, the usual error bound $\|g_\delta - g\|_Y \le \delta$ simply becomes $\|g_\delta\|_Y \le \delta$, whence one immediately gets $\|g_\delta\|_Y \le \nu_2$. Summing up, if $f^\dagger = 0$, we then take $f^\eta = 0$ as our approximation of $f^\dagger$, which is zero too, and nothing else needs to be proved.

We have also seen, in the proof of the same theorem 1.8.5, that if $f^\dagger \neq 0$, then, at least for η small enough, it holds (with the current notation) $\|g_\delta\|_Y > \nu_1$. On the other hand, if we had proved also theorem 1.8.7, we would have shown that if $f^\dagger \neq 0$, then, at least for η small enough, it holds (with the current notation) $\|g_\delta\|_Y > \nu_2$: indeed $f^\dagger \neq 0$ implies that $g \neq 0$, then, by observing that

$$|\, \|g_\delta\|_Y - \|g\|_Y| \le \|g_\delta - g\| \le \delta, \tag{1.361}$$

we immediately find

$$\lim_{\delta\to 0} \inf_{g_\delta} \{\|g_\delta\|_Y \mid \|g_\delta - g\|_Y \le \delta\} = \|g\|_Y > 0, \tag{1.362}$$

while, on the other hand, remembering limit (1.316), we easily get (with the current notation)

$$\lim_{\eta\to 0^+} \sup_{g_\delta, A_h} \{\nu_2(\eta, g_\delta, A_h)|\ \|g_\delta - g\|_Y \le \delta;\ \|A_h - A\| \le h\} = 0. \tag{1.363}$$

Hence, taking into account relations (1.362) and (1.363), we have that condition $\|g_\delta\|_Y > \nu_2$ holds, at least for η small enough. Summing up, if $f^\dagger \neq 0$, inequality $\|g_\delta\|_Y > \max\{\nu_1, \nu_2\}$ is true (at least for η small enough); then, for vanishing η, we can actually take $f^\eta_{\alpha^*_b(\eta, g_\delta, A_h)}$ as approximation of $f^\dagger$ and prove limit (1.359) in the case $f^\eta_{[\alpha^*_b(\eta, g_\delta, A_h)]} = f^\eta_{\alpha^*_b(\eta, g_\delta, A_h)}$. Moreover, by virtue of theorem 1.7.7, the proof of (1.359) in such a case implies that limit (1.360) holds too.

Now, by virtue of the (strict) monotonicity of the functions $\beta_\eta(\alpha)$ and $\gamma_\eta(\alpha)$ in $(0, +\infty)$, it is not difficult to realize that the remainder of the current proof can be obtained by simply "recycling" the correspondent part of the proof of theorem 1.8.5. More precisely, let $\mathcal{F}$ be as

in definition (1.287) and let us put:

$$\alpha^*_{1,n} := \alpha^*_1(\eta_n, g_{\delta_n}, A_{h_n}) \qquad \forall n \in \mathbb{N}, \tag{1.364}$$

$$\alpha^*_{2,n} := \alpha^*_2(\eta_n, g_{\delta_n}, A_{h_n}) \qquad \forall n \in \mathbb{N}, \tag{1.365}$$

$$\alpha^*_{b,n} := c(\eta_n)\,\alpha^*_{1,n} + [1 - c(\eta_n)]\,\alpha^*_{2,n} \qquad \forall n \in \mathbb{N}, \tag{1.366}$$

$$\alpha^*_{m,n} := \min\{\alpha^*_{1,n}, \alpha^*_{2,n}\} \qquad \forall n \in \mathbb{N}, \tag{1.367}$$

$$\alpha^*_{M,n} := \max\{\alpha^*_{1,n}, \alpha^*_{2,n}\} \qquad \forall n \in \mathbb{N}. \tag{1.368}$$

Now, since it holds (cf. inequalities (1.297) and (1.325)):

$$\left\| f^{\eta_n}_{\alpha^*_{1,n}} \right\|_X \leq \left\| f^\dagger \right\|_X \qquad \text{and} \qquad \left\| f^{\eta_n}_{\alpha^*_{2,n}} \right\|_X \leq \left\| f^\dagger \right\|_X \qquad \forall n \in \mathbb{N}, \tag{1.369}$$

by virtue of inequalities (1.350) and the (strictly) decreasing monotonicity of $\gamma_\eta(\alpha)$ in $(0, +\infty)$ just recalled, we have

$$\left\| f^{\eta_n}_{\alpha^*_{M,n}} \right\|_X \leq \left\| f^{\eta_n}_{\alpha^*_{b,n}} \right\|_X \leq \left\| f^{\eta_n}_{\alpha^*_{m,n}} \right\|_X \leq \left\| f^\dagger \right\|_X \qquad \forall n \in \mathbb{N}, \tag{1.370}$$

i.e. the sequence $\left\{ f^{\eta_n}_{\alpha^*_{b,n}} \right\}_{n=0}^{\infty}$ is bounded in the Hilbert space X, and therefore [13] it has a subsequence $\left\{ f^{\eta_{n(k)}}_{\alpha^*_{b,n(k)}} \right\}_{k=0}^{\infty}$ (which we shall denote, for notational convenience, with $\left\{ f^{\eta_k}_{\alpha^*_{b,k}} \right\}_{k=0}^{\infty}$) that is weakly convergent to an element $f^* \in X$, i.e.

$$f^{\eta_k}_{\alpha^*_{b,k}} \rightharpoonup f^*. \tag{1.371}$$

Now, with exactly the same arguments used for passing from relation (1.298) to inequalities (1.300), we have:

$$\|f^*\|_X \leq \liminf_{k\to\infty} \left\| f^{\eta_k}_{\alpha^*_{b,k}} \right\|_X \leq \limsup_{k\to\infty} \left\| f^{\eta_k}_{\alpha^*_{b,k}} \right\|_X \leq \left\| f^\dagger \right\|_X \tag{1.372}$$

and

$$\|Af^* - g\|_Y \leq \liminf_{k\to\infty} \left\| Af^{\eta_k}_{\alpha^*_{b,k}} - g \right\|_Y \leq \limsup_{k\to\infty} \left\| Af^{\eta_k}_{\alpha^*_{b,k}} - g \right\|_Y. \tag{1.373}$$

Then, using, in particular, the triangle inequality and relations (1.288), (1.289), as well as the (strictly) increasing monotonicity of $\beta_\eta(\alpha)$ in $(0, +\infty)$, we have:

$$\begin{aligned} \left\| Af^{\eta_k}_{\alpha^*_{b,k}} - g \right\|_Y &\leq \left\| Af^{\eta_k}_{\alpha^*_{b,k}} - A_{h_k} f^{\eta_k}_{\alpha^*_{b,k}} \right\|_Y + \left\| A_{h_k} f^{\eta_k}_{\alpha^*_{b,k}} - g_{\delta_k} \right\|_Y + \|g_{\delta_k} - g\|_Y \leq \\ &\leq h_k \left\| f^{\eta_k}_{\alpha^*_{b,k}} \right\|_X + \left\| A_{h_k} f^{\eta_k}_{\alpha^*_{M,k}} - g_{\delta_k} \right\|_Y + \delta_k. \end{aligned} \tag{1.374}$$

Now, if $\alpha^*_{M,k} = \alpha^*_{1,k}$, we have (cf. relation (1.292)):

$$\left\| A_{h_k} f^{\eta_k}_{\alpha^*_{M,k}} - g_{\delta_k} \right\|_Y \equiv \left\| A_{h_k} f^{\eta_k}_{\alpha^*_{1,k}} - g_{\delta_k} \right\|_Y = \delta_k + h_k \left\| f^{\eta_k}_{\alpha^*_{1,k}} \right\|_X + \hat{\mu}^{\kappa_1}_{\eta_k}(g_{\delta_k}, A_{h_k}); \tag{1.375}$$

then, remembering the first of inequalities (1.369), we can complete the chain of inequalities (1.374) as follows:

$$\left\|Af_{\alpha^*_{b,k}}^{\eta_k} - g\right\|_Y \leq h_k \left\|f^\dagger\right\|_X + \left(\delta_k + h_k \left\|f^\dagger\right\|_X + \hat{\mu}_{\eta_k}^{\kappa_1}(g_{\delta_k}, A_{h_k})\right) + \delta_k = \\ = 2\left(\delta_k + h_k \left\|f^\dagger\right\|_X\right) + \hat{\mu}_{\eta_k}^{\kappa_1}(g_{\delta_k}, A_{h_k}). \tag{1.376}$$

On the other hand, if $\alpha^*_{M,k} = \alpha^*_{2,k}$, we have (cf. relation (1.326)):

$$\left\|A_{h_k} f_{\alpha^*_{M,k}}^{\eta_k} - g_{\delta_k}\right\|_Y \equiv \left\|A_{h_k} f_{\alpha^*_{2,k}}^{\eta_k} - g_{\delta_k}\right\|_Y = \sqrt{\left(\delta_k + h_k \left\|f_{\alpha^*_{2,k}}^{\eta_k}\right\|_X\right)^2 + \left(\mu_{\eta_k}^{\kappa_2}(g_{\delta_k}, A_{h_k})\right)^2}; \tag{1.377}$$

then, remembering the second of inequalities (1.369), we can complete the chain of inequalities (1.374) as follows:

$$\left\|Af_{\alpha^*_{b,k}}^{\eta_k} - g\right\|_Y \leq h_k \left\|f^\dagger\right\|_X + \sqrt{\left(\delta_k + h_k \left\|f^\dagger\right\|_X\right)^2 + \left(\mu_{\eta_k}^{\kappa_2}(g_{\delta_k}, A_{h_k})\right)^2} + \delta_k. \tag{1.378}$$

By virtue of relations (1.258), (1.316) and recalling that in the compatible case it holds $\mu = 0$, it follows from either (1.376) or (1.378) that it holds:

$$\limsup_{k\to\infty} \left\|Af_{\alpha^*_{b,k}}^{\eta_k} - g\right\|_Y \leq 0. \tag{1.379}$$

Substituting this result into (1.373), we find that

$$\left\|Af^* - g\right\|_Y \leq 0, \tag{1.380}$$

i.e., as a consequence of the definition itself of norm,

$$\left\|Af^* - g\right\|_Y = 0. \tag{1.381}$$

At this point, in order to prove thesis (1.359) we can follow, with minor adjustments, the same arguments explained in the proof of theorem 1.8.5 from relation (1.304) (with $\mu = 0$) to the end. ■

Remark 1.8.8. Of course, one has in practice to appropriately determine the explicit form of the blending function $c(\eta) \equiv c(\delta, h)$: however, except the obvious requirement of non-decreasing monotonicity of $c(\delta_0, \cdot)$ for each fixed $\delta_0 > 0$ and, in the case of strongly ill-posed exact problems, the trivial condition

$$\lim_{h\to 0^+} c(\delta_0, h) = 0 \qquad \forall \delta_0 > 0, \tag{1.382}$$

we have here no other suggestion to offer. In all the numerical experiments performed by us, it was $\delta = 0$ and the explicit form of $c(\eta) = c(h)$ has been always chosen in an absolutely heuristic way, even taking into account[40] the value h_s of the norm of the specific noise matrix $\mathbf{H}_s$ which, in the numerical experiment under consideration, we add to the exact operator $\mathbf{A}$ in order to simulate measurement errors. □

[40] Cf. relation (3.73).

Remark 1.8.9. Hence, a natural criticism to all this blending approach could be that we have simply transformed the problem of finding, for any fixed linear inverse problem, an optimal value of the regularization parameter α into the somehow equivalent problem of finding an optimal value of the blending parameter c: but, in any case, the real optimality problem has not been solved (and, in general, cannot actually be solved, as pointed out in remark 1.6.7). However, we can obviously object that the two parameters α and c are not at all on the same level: indeed, first of all we should remember that, in many cases, at least one of the two values $\alpha_1^*(\eta, g_\delta, A_h)$ and $\alpha_2^*(\eta, g_\delta, A_h)$ of α provided by the generalized discrepancy principle in its two possible forms are satisfactory enough or, at least, give a first approximation of the optimal value of α, while determining the best, or, at least, a suitable value of c represents a subsequent procedure of fine tuning. Moreover, a purely heuristic (i.e. "by hand") choice of the value of α does not give rise, in general, to a regularization algorithm (in the sense of definition 1.6.1), while an analogous choice of the value of c does not prevent the algorithm itself from being a regularization method. In other terms, if, on the one hand, we cannot give a general recipe for determining, for any linear inverse problem, *a suitable and specific value* of c (i.e. we can give neither an explicit expression of the blending function $c = c(\eta)$ introduced in (1.347), nor a general computational procedure to numerically determine it), we have at least proved that *the blending procedure in itself* constitutes a regularization algorithm. □

Blending the two regularized solutions

We consider the following convex combination of the two regularized solutions:

$$f_b^\eta := c(\eta)\, f^\eta_{\alpha_1^*(\eta,g_\delta,A_h)} + [1 - c(\eta)]\, f^\eta_{\alpha_2^*(\eta,g_\delta,A_h)} \tag{1.383}$$

(where $c = c(\eta)$ is as in (1.347)) and define (cf. relation (1.358)):

$$f^\eta_{[b]} := \begin{cases} f^\eta = 0 & \text{if } \|g_\delta\|_Y \le \min\{\nu_1, \nu_2\}, \\ f^\eta_{\alpha_1^*(\eta,g_\delta,A_h)} & \text{if } \nu_1 < \|g_\delta\|_Y \le \nu_2, \\ f^\eta_{\alpha_2^*(\eta,g_\delta,A_h)} & \text{if } \nu_2 < \|g_\delta\|_Y \le \nu_1, \\ f_b^\eta & \text{if } \|g_\delta\|_Y > \max\{\nu_1, \nu_2\}. \end{cases} \tag{1.384}$$

Of course, the really non-trivial case is the fourth one, i.e.:

$$f^\eta_{[b]} = f_b^\eta. \tag{1.385}$$

It is not difficult to prove that, in general, f_b^η is not a Tikhonov regularized solution of the inverse noisy problem $A_h f = g_\delta$, in the sense that there does not exist any $\hat\alpha > 0$ such that (cf. definitions (1.172) and (1.173)) it holds:

$$f_b^\eta = \operatorname{argmin} \Phi^\eta_{\hat\alpha}[f; g_\delta], \tag{1.386}$$

or, equivalently, such that f_b^η is the unique solution to the Euler equation (cf. (1.174)):

$$(A_h^* A_h + \hat{\alpha} I)\, f_b^\eta = A_h^*\, g_\delta. \tag{1.387}$$

To this end, recalling the shorthand notations (1.328) and (1.331), we firstly remember that $f_{\alpha_1^*}^\eta := f_{\alpha_1^*(\eta, g_\delta, A_h)}^\eta$ is the unique solution to the Euler equation

$$(A_h^* A_h + \alpha_1^* I)\, f_{\alpha_1^*}^\eta = A_h^*\, g_\delta, \tag{1.388}$$

while, analogously, $f_{\alpha_2^*}^\eta := f_{\alpha_2^*(\eta, g_\delta, A_h)}^\eta$ is the unique solution to the Euler equation

$$(A_h^* A_h + \alpha_2^* I)\, f_{\alpha_2^*}^\eta = A_h^*\, g_\delta. \tag{1.389}$$

Hence, by multiplying for $c(\eta)$ both sides of equation (1.388) and for $[1 - c(\eta)]$ both sides of (1.389), and then summing member by member the two equations so obtained, one easily gets:

$$A_h^* A_h \left\{ c(\eta)\, f_{\alpha_1^*}^\eta + [1 - c(\eta)]\, f_{\alpha_2^*}^\eta \right\} + c(\eta)\, \alpha_1^*\, f_{\alpha_1^*}^\eta + [1 - c(\eta)]\, \alpha_2^*\, f_{\alpha_2^*}^\eta = A_h^* \left\{ c(\eta)\, g_\delta + [1 - c(\eta)]\, g_\delta \right\}, \tag{1.390}$$

i.e., recalling definition (1.383),

$$A_h^* A_h f_b^\eta + c(\eta)\, \alpha_1^*\, f_{\alpha_1^*}^\eta + [1 - c(\eta)]\, \alpha_2^*\, f_{\alpha_2^*}^\eta = A_h^* g_\delta. \tag{1.391}$$

Hence, let us now suppose, by absurd, that equation (1.387) holds for a certain $\hat{\alpha} > 0$; since, on the other hand, also equation (1.391) has to hold, we can subtract member by member the two equations (1.387) and (1.391) themselves, so that we find:

$$\hat{\alpha} f_b^\eta - \left\{ c(\eta)\, \alpha_1^*\, f_{\alpha_1^*}^\eta + [1 - c(\eta)]\, \alpha_2^*\, f_{\alpha_2^*}^\eta \right\} = 0, \tag{1.392}$$

i.e., recalling definition (1.383),

$$\hat{\alpha} \left\{ c(\eta)\, f_{\alpha_1^*}^\eta + [1 - c(\eta)]\, f_{\alpha_2^*}^\eta \right\} - \left\{ c(\eta)\, \alpha_1^*\, f_{\alpha_1^*}^\eta + [1 - c(\eta)]\, \alpha_2^*\, f_{\alpha_2^*}^\eta \right\} = 0, \tag{1.393}$$

or, equivalently,

$$(\hat{\alpha} - \alpha_1^*)\, c(\eta)\, f_{\alpha_1^*}^\eta + (\hat{\alpha} - \alpha_2^*)[1 - c(\eta)]\, f_{\alpha_2^*}^\eta = 0. \tag{1.394}$$

We now observe that if $\alpha_1^* \neq \alpha_2^*$ (as we are actually assuming, otherwise definition (1.383) would be trivial), then, in general, $f_{\alpha_1^*}^\eta$ and $f_{\alpha_2^*}^\eta$ are linearly independent[41]: in such a case, equality (1.394) holds if and only if

$$(\hat{\alpha} - \alpha_1^*)\, c(\eta) = 0 \qquad \wedge \qquad (\hat{\alpha} - \alpha_2^*)[1 - c(\eta)] = 0; \tag{1.395}$$

[41] To this end, it suffices to remember representation (1.177), holding in the case in which A_h is compact: if we replace the generic α in (1.177) with two different values α_1^* and α_2^*, we easily realize that the coefficients multiplying the orthonormal elements u_k^h when $\alpha = \alpha_1^*$ are not proportional with the same proportionality constant to the corresponding ones, obtained for $\alpha = \alpha_2^*$: this clearly implies the linear independence of $f_{\alpha_1^*}^\eta$ and $f_{\alpha_2^*}^\eta$.

the previous relations (1.395), in turn, hold if and only if

$$\hat{\alpha} = \alpha_1^* \ \wedge \ c(\eta) = 1 \qquad \vee \qquad \hat{\alpha} = \alpha_2^* \ \wedge \ c(\eta) = 0. \tag{1.396}$$

But now, from definition (1.383) we immediately realize that if $c(\eta) = 1$, then $f_b^\eta = f_{\alpha_1^*}^\eta$, while, if $c(\eta) = 0$, then $f_b^\eta = f_{\alpha_2^*}^\eta$: in both cases, the convex linear combination (1.383) is degenerate.

The previous arguments show that blending the two regularized solutions cannot be considered as a generalization or a particular case of blending the two regularization parameters: the two procedures are intrinsically different. In other terms, blending the two regularization parameters gives rise to a Tikhonov regularized solution $f_{\alpha_b^*(\eta,g_\delta,A_h)}^\eta$, with its proper value $\alpha_b^*(\eta, g_\delta, A_h)$ for the regularization parameter, while blending the two regularized solutions gives rise to a new regularized solution f_b^η which is not of Tikhonov type and for which it is even meaningless to speak of a correspondent value for the regularization parameter. However, we can easily state the following theorem.

Theorem 1.8.9. *Let $f^\dagger$ denote, as usual, the generalized solution of the exact inverse problem $Af = g$, with $g \in \mathcal{R}(A)$; then it holds:*

$$\lim_{\eta \to 0^+} \sup_{g_\delta, A_h} \left\{ \left\| f_{[b]}^\eta - f^\dagger \right\|_X \ | \ \|g_\delta - g\|_Y \leq \delta; \ \|A_h - A\| \leq h \right\} = 0. \tag{1.397}$$

Proof. By virtue of the same argument used at the very beginning of the proof of theorem 1.8.8, we can show that if $f^\dagger = 0$, then $f_{[b]}^\eta = f^\eta = 0$ (and nothing else needs to be proved), while if $f^\dagger \neq 0$, we can always assume (at least for η small enough) that $f_{[b]}^\eta = f_b^\eta$. Then, recalling definition (1.383) and using the triangle inequality, we can write:

$$\begin{aligned} \left\| f_b^\eta - f^\dagger \right\|_X &= \left\| c(\eta) \, f_{\alpha_1^*(\eta,g_\delta,A_h)}^\eta + [1 - c(\eta)] \, f_{\alpha_2^*(\eta,g_\delta,A_h)}^\eta - \left\{ c(\eta) \, f^\dagger + [1 - c(\eta)] \, f^\dagger \right\} \right\|_X \leq \\ &\leq c(\eta) \, \left\| f_{\alpha_1^*(\eta,g_\delta,A_h)}^\eta - f^\dagger \right\|_X + [1 - c(\eta)] \, \left\| f_{\alpha_2^*(\eta,g_\delta,A_h)}^\eta - f^\dagger \right\|_X . \end{aligned} \tag{1.398}$$

Limits (1.280) and (1.323), together with the fact that $0 \leq c(\eta) \leq 1 \ \forall \eta \geq 0$, now suffice to prove our thesis (1.397). ■

We can conclude this chapter recalling, also in this case, remark 1.8.8.

CHAPTER 2

The direct and the inverse scattering problem

In this chapter we introduce the direct and the inverse electromagnetic scattering problem we want to deal with: we focus on the case of a penetrable medium, endowed with suitable symmetry properties, but an analogous approach, with the proper changes, can be followed also for many other kinds of scattering (impenetrable objects with various boundary conditions, acoustic waves, etc.): in particular, what can be considered, from our point of view, the main theoretical results of this chapter, i.e. the theses (2.215)-(2.220) of the *general theorem* 2.4.10 in section 2.4, hold substantially unchanged in all situations, provided that the hypotheses are suitably adapted to each case.

The aim of sections 2.1, 2.2 and 2.3 is to introduce the mathematical framework in which our direct and inverse scattering problem is posed and studied; appendix A has been written with the principal purpose to make them more easily readable, by collecting in few pages the most important definitions, notations, theorems and properties which form the theoretical basis of such a framework. In general, as far as the proofs of the theorems stated in these three sections are concerned, we simply refer to the existing literature (and mainly to [15]). However, we occasionally give an explicit proof of some results: it happens when we have not been able to refer to any book of our common use or when the proof itself is unusually simple or can be helpful in the following.

In section 2.4 all the theorems are explicitly proved: this is true, in particular, for the *general theorem* 2.4.10, whose results form the mathematical basis inspiring the linear sampling method. The latter is fully explained, from a technical viewpoint, in section 2.5, while the reasons for the importance of such a visualization method, as well as its main features, are illustrated in the preface of this PhD thesis.

Finally, since in [15], which is our main reference book for sections 2.1, 2.2, 2.3 and 2.4, the support of the scatterer is always supposed to take up a spatial region which is the closure

of a C^2 domain D, we maintain this smoothness hypothesis; however, the only reason for which the authors of [15] make such an assumption, which is actually a bit restrictive, seems to be the concern of simplifying, as far as possible, the introduction of the Sobolev spaces appearing in the mathematical framework they need to explain. Hence, all the theorems and results stated in this chapter should keep on holding unaltered also when D is assumed to be a Lipschitz domain[1]; anyway, since for the proofs of several theorems we simply refer to [15], without checking in detail if all the arguments are still true for less regular domains, we need to maintain the smoothness hypothesis made in that book.

2.1. Formulation of the direct scattering problem

The aim of this section is to mathematically formulate the direct problem concerning the scattering of a time-harmonic electromagnetic wave by a penetrable, orthotropic[2], inhomogeneous and cylinder-shaped medium. Since we have in mind a weak formulation of the problem in suitable Sobolev spaces, at first we shall not be interested in specifying all the regularity properties of the functions we are going to consider and we shall simply assume that they are smooth enough for all the passages to make sense. Of course, we shall finally state a theorem of well-posedness, in which all the hypotheses will be explicitly declared.

Moreover, we shall set up the problem in such a way that it will be endowed with cylindrical symmetry: this will enable us to pass from a 3D to a 2D formulation by considering a plane section orthogonal to the axis of the cylinder; the latter will be assumed to be parallel to the x_3-axis of a cartesian orthogonal system of coordinates (x_1, x_2, x_3) in $\mathbb{R}^3$. Such reduction of the original 3D problem to a 2D one implies that the variable $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ will be replaced by its orthogonal projection $x' = (x_1, x_2) \in \mathbb{R}^2$. As usual, we shall denote with $\nabla\times$, ∇_2, $\nabla_2\cdot$, Δ_2 respectively the curl operator in $\mathbb{R}^3$, the gradient, divergence and laplacian operators in $\mathbb{R}^2$, all expressed in spatial cartesian orthogonal coordinates. Finally, if $\xi = (\xi_1, \xi_2)$ and $\zeta = (\zeta_1, \zeta_2)$ are two elements of $\mathbb{C}^2$, we define the operation

$$\xi \cdot \zeta := \xi_1\zeta_1 + \xi_2\zeta_2, \tag{2.1}$$

which is clearly not a scalar product on $\mathbb{C}^2$, but it is actually the canonical scalar product on $\mathbb{R}^2$ if restricted to $\mathbb{R}^2$ itself. Of course, it holds:

$$\|\xi\|_{\mathbb{C}^2} = \sqrt{\xi \cdot \bar{\xi}} \qquad \forall \xi \in \mathbb{C}^2, \tag{2.2}$$

[1] We point out that the case of Lipschitz domains is explicitly taken into account, e.g., in the following papers: [18], [19], [20].

[2] The adjective "orthotropic" stands for "orthogonally anisotropic"; hence, in general, a material is orthotropic if its physical properties are different along mutually orthogonal directions. Of course, this is only a quite generic definition by words; for a precise characterization in mathematical terms of what we mean by "orthotropic", see relations from (2.7) to (2.10) in the following.

having denoted with $\|\xi\|_{\mathbb{C}^2}$ the usual norm of any element $\xi \in \mathbb{C}^2$. More generally, if $n \in \mathbb{N} \setminus \{0,1\}$, we denote with $\|\xi\|_{\mathbb{C}^n}$ and $\|x\|_{\mathbb{R}^n}$ the norm of any element $\xi \in \mathbb{C}^n$ and $x \in \mathbb{R}^n$ respectively, while the more familiar notation $|\xi|$ or $|x|$ will be used to indicate the modulus of any element $\xi \in \mathbb{C}$ or $x \in \mathbb{R}$ respectively.

Let us now consider electromagnetic wave propagation in an inhomogeneous anisotropic medium which is geometrically described in coordinates as a subset of $\mathbb{R}^3$ and physically characterized by the three following tensors: electric permittivity $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(x)$, magnetic permeability $\boldsymbol{\mu} = \boldsymbol{\mu}(x)$ and electric conductivity $\boldsymbol{\sigma} = \boldsymbol{\sigma}(x)$. For the moment, we simply assume that each of these tensors is represented in our cartesian orthogonal coordinate system by a $\mathbb{C}^{3\times 3}$ matrix depending (continuously) on x. Some conditions involving the three matrices will be imposed later. The electromagnetic wave is described by the $\mathbb{C}^3$-valued functions $\mathbf{E} = \mathbf{E}(x,t)$ and $\mathbf{H} = \mathbf{H}(x,t)$ (called respectively *electric field* and *magnetic field*) satisfying the system of Maxwell equations:

$$\begin{cases} \nabla \times \mathbf{E} + \boldsymbol{\mu} \dfrac{\partial \mathbf{H}}{\partial t} = 0 \\ \nabla \times \mathbf{H} - \boldsymbol{\varepsilon} \dfrac{\partial \mathbf{E}}{\partial t} = \boldsymbol{\sigma} \mathbf{E}, \end{cases} \tag{2.3}$$

where we have used the constitutive relation (Ohm's law) $\mathbf{J} = \boldsymbol{\sigma}\mathbf{E}$ for the density current $\mathbf{J} = \mathbf{J}(x,t)$. If we consider electromagnetic waves which are time harmonic, i.e. of the form

$$\mathbf{E}(x,t) = \tilde{\mathbf{E}}(x) e^{-\mathrm{i}\omega t}, \tag{2.4}$$

$$\mathbf{H}(x,t) = \tilde{\mathbf{H}}(x) e^{-\mathrm{i}\omega t}, \tag{2.5}$$

with frequency $\omega > 0$, and we substitute expressions (2.4) and (2.5) into the equations of system (2.3), we find that the complex vector-valued and space dependent functions $\tilde{\mathbf{E}} = \tilde{\mathbf{E}}(x)$ and $\tilde{\mathbf{H}} = \tilde{\mathbf{H}}(x)$ (called *spatial electric field* and *spatial magnetic field* respectively) satisfy the following system:

$$\begin{cases} \nabla \times \tilde{\mathbf{E}} - \mathrm{i}\omega \boldsymbol{\mu} \tilde{\mathbf{H}} = 0, \\ \nabla \times \tilde{\mathbf{H}} + (\mathrm{i}\omega \boldsymbol{\varepsilon} - \boldsymbol{\sigma})\, \tilde{\mathbf{E}} = 0. \end{cases} \tag{2.6}$$

We now suppose that the inhomogeneity consists of an infinitely long cylinder of penetrable, i.e. imperfectly conductor, material. Let $D \subset \mathbb{R}^2$ be the (open) cross section of the cylinder, having a C^2 boundary ∂D, and let ν be the unit outward normal to ∂D. Moreover, we assume that the cylinder has its axis coinciding with the x_3-axis (i.e. it is the set $\bar{D} \times \mathbb{R} \subset \mathbb{R}^3$) and that it is embedded in a non-conducting homogeneous background, i.e. the electric permittivity and the magnetic permeability of the background medium are positive constants $\varepsilon_0 > 0$ and $\mu_0 > 0$ respectively, while its conductivity is $\sigma_0 = 0$.

Then we define the matrices $\mathbf{A} = \mathbf{A}(x)$ and $\mathbf{N} = \mathbf{N}(x)$ as

$$\mathbf{A}(x) := \frac{1}{\varepsilon_0}\left(\boldsymbol{\varepsilon}(x) + \mathrm{i}\frac{\boldsymbol{\sigma}(x)}{\omega}\right), \tag{2.7}$$

$$\mathbf{N}(x) := \frac{1}{\mu_0}\boldsymbol{\mu}(x), \tag{2.8}$$

and we further assume that the interior cylinder-shaped medium is orthotropic: this means, in our case, that the matrices $\mathbf{A}(x)$ and $\mathbf{N}(x)$ are independent of the x_3-coordinate and are of the form:

$$\mathbf{A}(x) = \begin{pmatrix} a_{11}(x) & a_{12}(x) & 0 \\ a_{21}(x) & a_{22}(x) & 0 \\ 0 & 0 & a(x) \end{pmatrix}, \tag{2.9}$$

$$\mathbf{N}(x) = \begin{pmatrix} n_{11}(x) & n_{12}(x) & 0 \\ n_{21}(x) & n_{22}(x) & 0 \\ 0 & 0 & n(x) \end{pmatrix}. \tag{2.10}$$

If we denote with $\tilde{\mathbf{E}}^{int} = \tilde{\mathbf{E}}^{int}(x)$, $\tilde{\mathbf{H}}^{int} = \tilde{\mathbf{H}}^{int}(x)$ and $\tilde{\mathbf{E}}^{ext} = \tilde{\mathbf{E}}^{ext}(x)$, $\tilde{\mathbf{H}}^{ext} = \tilde{\mathbf{H}}^{ext}(x)$ the spatial electric and magnetic fields inside the orthotropic medium and outside it respectively, and define the *relative* spatial electric and magnetic fields $\hat{\mathbf{E}}^{int} = \hat{\mathbf{E}}^{int}(x)$, $\hat{\mathbf{E}}^{ext} = \hat{\mathbf{E}}^{ext}(x)$, $\hat{\mathbf{H}}^{int} = \hat{\mathbf{H}}^{int}(x)$, $\hat{\mathbf{H}}^{ext} = \hat{\mathbf{H}}^{ext}(x)$ as

$$\hat{\mathbf{E}}^{int} := \sqrt{\varepsilon_0}\,\tilde{\mathbf{E}}^{int}, \qquad \hat{\mathbf{E}}^{ext} := \sqrt{\varepsilon_0}\,\tilde{\mathbf{E}}^{ext}, \tag{2.11}$$

$$\hat{\mathbf{H}}^{int} := \sqrt{\mu_0}\,\tilde{\mathbf{H}}^{int}, \qquad \hat{\mathbf{H}}^{ext} := \sqrt{\mu_0}\,\tilde{\mathbf{H}}^{ext}, \tag{2.12}$$

it follows from system (2.6) and definitions (2.7), (2.8) that the fields $\hat{\mathbf{E}}^{int}$, $\hat{\mathbf{H}}^{int}$ inside the cylinder satisfy the system:

$$\begin{cases} \nabla \times \hat{\mathbf{E}}^{int} - \mathrm{i}k\mathbf{N}\hat{\mathbf{H}}^{int} = 0 \\ \nabla \times \hat{\mathbf{H}}^{int} + \mathrm{i}k\mathbf{A}\hat{\mathbf{E}}^{int} = 0, \end{cases} \tag{2.13}$$

while the fields $\hat{\mathbf{E}}^{ext}$, $\hat{\mathbf{H}}^{ext}$ outside the cylinder satisfy the system:

$$\begin{cases} \nabla \times \hat{\mathbf{E}}^{ext} - \mathrm{i}k\hat{\mathbf{H}}^{ext} = 0 \\ \nabla \times \hat{\mathbf{H}}^{ext} + \mathrm{i}k\hat{\mathbf{E}}^{ext} = 0, \end{cases} \tag{2.14}$$

where we have denoted with

$$k := \omega\sqrt{\varepsilon_0\mu_0} \tag{2.15}$$

the wavenumber in the background medium. Of course, across the boundary of the cylinder one has to impose the continuity of the tangential components of both the (relative spatial) electric and magnetic fields.

If we now assume that the matrix $\mathbf{A}(x)$ is invertible $\forall x \in \bar{D} \times \mathbb{R}$, we can get an expression for $\hat{\mathbf{E}}^{int}(x)$ from the second equation of system (2.13) and substitute it into the first one, so finding the equation

$$\nabla \times \left[\mathbf{A}^{-1}\left(\nabla \times \hat{\mathbf{H}}^{int}\right)\right] - k^2\,\mathbf{N}\,\hat{\mathbf{H}}^{int} = 0 \tag{2.16}$$

for the relative spatial magnetic field inside the cylinder. Analogously, we can get an expression for $\hat{\mathbf{E}}^{ext}$ from the second equation of system (2.14) and substitute it into the first one, so finding the equation

$$\nabla \times \left(\nabla \times \hat{\mathbf{H}}^{ext}\right) - k^2\,\hat{\mathbf{H}}^{ext} = 0 \tag{2.17}$$

for the relative spatial magnetic field outside the cylinder. Of course, the same equation as (2.17) holds for $\hat{\mathbf{E}}^{ext}$ too.

Till now we have simply treated the propagation of time harmonic electromagnetic waves in two different media; however, our aim is more specific, since it requires one to find and study the equations describing a scattering experiment. More precisely, we now assume that a time harmonic electromagnetic incident wave, characterized by the relative spatial electric and magnetic fields $\hat{\mathbf{E}}^i = \hat{\mathbf{E}}^i(x)$, $\hat{\mathbf{H}}^i = \hat{\mathbf{H}}^i(x)$, is propagated in the background medium and then scattered by the cylinder. This implies, in particular, that the following expressions hold:

$$\hat{\mathbf{E}}^{ext} = \hat{\mathbf{E}}^s + \hat{\mathbf{E}}^i, \tag{2.18}$$

$$\hat{\mathbf{H}}^{ext} = \hat{\mathbf{H}}^s + \hat{\mathbf{H}}^i, \tag{2.19}$$

where we have respectively denoted with $\hat{\mathbf{E}}^s = \hat{\mathbf{E}}^s(x)$ and $\hat{\mathbf{H}}^s = \hat{\mathbf{H}}^s(x)$ the relative spatial electric and magnetic fields scattered by the cylinder. In general, the fields $\hat{\mathbf{E}}^i$, $\hat{\mathbf{H}}^i$ characterizing the incident wave are entire (i.e. defined onto all $\mathbb{R}^3$) solutions of system (2.14) (and hence of equation (2.17) too), while the scattered fields $\hat{\mathbf{E}}^s$, $\hat{\mathbf{H}}^s$ satisfy the Silver-Müller radiation condition (see [54], [59]), which can be written in the following form:

$$\lim_{r\to\infty} \sup_{\substack{\theta\,\in\,[0,\,\pi] \\ \varphi\,\in\,[0,\,2\pi]}} \left[\hat{\mathbf{H}}^s(r,\theta,\varphi) \times \mathbf{x}(r,\theta,\varphi) - r\hat{\mathbf{E}}^s(r,\theta,\varphi)\right] = 0, \tag{2.20}$$

where we have denoted with (r,θ,φ) the spherical coordinates in $\mathbb{R}^3$, with $\mathbf{x}$ the position vector of cartesian components $(r\sin\theta\cos\varphi, r\sin\theta\sin\varphi, r\cos\theta)$, and with $\hat{\mathbf{E}}^s(r,\theta,\varphi)$, $\hat{\mathbf{H}}^s(r,\theta,\varphi)$ the expression in spherical coordinates of the fields $\hat{\mathbf{E}}^s(x)$, $\hat{\mathbf{H}}^s(x)$ respectively[3].

Let us now assume that the incident wave propagates along a direction perpendicular to the axis of the cylinder and is TE polarized[4] in such a way that

$$\hat{\mathbf{H}}^i(x) = (0, 0, u^i(x)), \tag{2.21}$$

[3] Of course, we are making a notational abuse: for example, writing $\hat{\mathbf{E}}^s(r,\theta,\varphi)$ is simply a shorthand for $\hat{\mathbf{E}}^s(r\sin\theta\cos\varphi, r\sin\theta\sin\varphi, r\cos\theta)$.

[4] "TE" is a shorthand for "Transverse Electric": then an electromagnetic wave is TE polarized if its electric field is orthogonal to the axis of the cylinder-shaped medium scattering the wave itself.

being the function $u^i(x)$ independent of the third coordinate x_3; from the cylindrical symmetry of the problem, it immediately follows that there exist two x_3-independent functions $u^s(x)$ and $v(x)$ such that

$$\hat{\mathbf{H}}^s(x) = (0, 0, u^s(x)), \tag{2.22}$$

$$\hat{\mathbf{H}}^{int}(x) = (0, 0, v(x)). \tag{2.23}$$

If we now substitute relation (2.23) into (2.16) and remember that all the involved functions do not depend on x_3, we find that the equation (2.16) itself is equivalent to the following one:

$$\nabla_2 \cdot \mathbf{A}' \nabla_2 v + k^2 n v = 0 \qquad \text{in } D, \tag{2.24}$$

where $v = v(x')$, $n = n(x')$ with $x' \in D$ and

$$\mathbf{A}' = \mathbf{A}'(x') := \frac{1}{a_{11}(x')a_{22}(x') - a_{12}(x')a_{21}(x')} \begin{pmatrix} a_{11}(x') & a_{12}(x') \\ a_{21}(x') & a_{22}(x') \end{pmatrix}. \tag{2.25}$$

Analogously, substituting relations (2.22), (2.21), (2.19) into equation (2.17) and remembering that $\hat{\mathbf{H}}^i$ is an entire solution of (2.17) itself, we find that the latter is equivalent to the Helmholtz equation:

$$\Delta_2 u^s + k^2 u^s = 0 \qquad \text{in } \mathbb{R}^2 \setminus \bar{D}, \tag{2.26}$$

where $u^s = u^s(x')$ with $x' \in \mathbb{R}^2 \setminus \bar{D}$. Moreover, the transmission conditions requiring the continuity of the tangential component of both the (relative spatial) electric and magnetic fields can be written as:

$$\begin{cases} \nu \times \hat{\mathbf{H}}^{ext} = \nu \times \hat{\mathbf{H}}^{int} \\ \nu \times \hat{\mathbf{E}}^{ext} = \nu \times \hat{\mathbf{E}}^{int} \end{cases} \qquad \text{on } \partial D \times \mathbb{R}, \tag{2.27}$$

i.e., recalling relations (2.18), (2.19) and the second equations of systems (2.13), (2.14),

$$\begin{cases} \nu \times \left(\hat{\mathbf{H}}^s + \hat{\mathbf{H}}^i \right) = \nu \times \hat{\mathbf{H}}^{int} \\ \nu \times \left[\nabla \times \left(\hat{\mathbf{H}}^s + \hat{\mathbf{H}}^i \right) \right] = \nu \times \left[\mathbf{A}^{-1} \left(\nabla \times \hat{\mathbf{H}}^{int} \right) \right] \end{cases} \qquad \text{on } \partial D \times \mathbb{R}. \tag{2.28}$$

If we now substitute representations (2.21), (2.22), (2.23) into system (2.28), the transmission conditions turn out to be equivalent to the following ones:

$$\begin{cases} v - u^s = u^i \\ \nu \cdot \mathbf{A}' \nabla_2 v - \nu \cdot \nabla_2 u^s = \nu \cdot \nabla_2 u^i \end{cases} \qquad \text{on } \partial D, \tag{2.29}$$

where clearly $v = v(x')$, $u^s = u^s(x')$, $u = u^i(x')$, $\nu = \nu(x')$ with $x' \in \partial D$.

Furthermore, if we define the normal and conormal derivative respectively for functions $u \in C^1(\mathbb{R}^2 \setminus D)$ and $v \in C^1(\bar{D})$ as:

$$\frac{\partial u}{\partial \nu}(x') := \lim_{h \to 0^+} [\nu(x') \cdot \nabla_2 u(x' + h\nu(x'))] \qquad \forall x' \in \partial D, \tag{2.30}$$

$$\frac{\partial v}{\partial \nu_{\mathbf{A}'}}(x') := \lim_{h \to 0^+} [\nu(x') \cdot \mathbf{A}'(x') \nabla_2 v(x' - h\nu(x'))] \qquad \forall x' \in \partial D, \tag{2.31}$$

we can rewrite the second condition of system (2.29) as:

$$\frac{\partial v}{\partial \nu_{\mathbf{A}'}} - \frac{\partial u^s}{\partial \nu} = \frac{\partial u^i}{\partial \nu} \qquad \text{on } \partial D. \tag{2.32}$$

Finally, the $\mathbb{R}^2$ analogue of the Silver-Müller radiation condition (2.20) is the Sommerfeld radiation condition [63], which in our case, denoting with $u^s(\rho', \theta')$ the expression in polar coordinates of the function $u^s(x')$ (and then making a little notational abuse, cf. footnote 3), reads

$$\lim_{\rho' \to \infty} \sup_{\theta' \in [0, 2\pi]} \left[\sqrt{\rho'} \left(\frac{\partial u^s}{\partial \rho'}(\rho', \theta') - \mathrm{i}k u^s(\rho', \theta') \right) \right] = 0. \tag{2.33}$$

In the following, we shall rewrite the previous limit in the more familiar form (which is actually a shorthand):

$$\lim_{r \to \infty} \left[\sqrt{r} \left(\frac{\partial u^s}{\partial r} - \mathrm{i}k u^s \right) \right] = 0. \tag{2.34}$$

Summing up and recalling relations (2.24), (2.26), (2.29), (2.32), (2.34), it turns out that the electromagnetic scattering of a known time harmonic incident field by an orthotropic inhomogeneity, as described above, can be mathematically formulated, roughly speaking, as the following problem in $\mathbb{R}^2$: given the function $u^i = u^i(x')$, find two functions $v = v(x')$ and $u^s = u^s(x')$ such that

$$\begin{cases} \nabla_2 \cdot \mathbf{A}' \nabla_2 v + k^2 n v = 0 & \text{in } D & \text{(a)} \\ \Delta_2 u^s + k^2 u^s = 0 & \text{in } \mathbb{R}^2 \setminus \bar{D} & \text{(b)} \\ v - u^s = u^i & \text{on } \partial D & \text{(c)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} - \dfrac{\partial u^s}{\partial \nu} = \dfrac{\partial u^i}{\partial \nu} & \text{on } \partial D & \text{(d)} \\ \lim\limits_{r \to \infty} \left[\sqrt{r} \left(\dfrac{\partial u^s}{\partial r} - \mathrm{i}k u^s \right) \right] = 0. & & \text{(e)} \end{cases} \tag{2.35}$$

The next step is now to refine this first formulation of the problem, by making precise assumptions on the domain D, on the coefficients $\mathbf{A}'$, n in equation (2.35)(a) and on the functional (Sobolev) spaces to which the functions v, u^s, u^i have to belong.

Referring to Appendix A (in particular, to sections A.1, A.8, A.9, A.10, A.13, A.14, A.15, A.17) for a definition of all the Sobolev spaces involved, it turns out (see [15], section 5.2) that a good mathematical (and more general) formulation of problem (2.35) is the following one.

Problem 2.1.1. *Let $D \subset \mathbb{R}^2$ be a nonempty, open and bounded set such that its boundary ∂D is of class C^2 and the exterior domain $\mathbb{R}^2 \setminus \bar{D}$ is connected. Let ν be the unit normal vector to ∂D, directed into the exterior of D. Let the matrix-valued function $\mathbf{A}' : \bar{D} \to \mathbb{C}^{2\times 2}$, with $\mathbf{A}' = \left(a'_{jk}\right)_{j,k=1,2}$, satisfy the following properties:*

1. *the functions a'_{jk} are continuously differentiable, i.e. $a'_{jk} \in C^1(\bar{D})$ $\forall j,k = 1,2$;*
2. *the matrix-valued function $\mathrm{Re}(\mathbf{A}') : \bar{D} \to \mathbb{R}^{2\times 2}$, defined by $(\mathrm{Re}(\mathbf{A}')(x'))_{i,j} := \mathrm{Re}(a'_{jk}(x'))$ $\forall j,k = 1,2$ and $\forall x' \in \bar{D}$, is symmetric and verifies the condition:*

$$\exists\, \gamma > 0 \quad | \quad \bar{\xi} \cdot \mathrm{Re}(\mathbf{A}'(x'))\, \xi \geq \gamma \, \|\xi\|^2_{\mathbb{C}^2} \qquad \forall \xi \in \mathbb{C}^2,\ \forall x' \in \bar{D}; \tag{2.36}$$

3. *the matrix-valued function $\mathrm{Im}(\mathbf{A}') : \bar{D} \to \mathbb{R}^{2\times 2}$, defined by $(\mathrm{Im}(\mathbf{A}')(x'))_{i,j} := \mathrm{Im}(a'_{jk}(x'))$ $\forall j,k = 1,2$ and $\forall x' \in \bar{D}$, is symmetric and verifies the condition:*

$$\bar{\xi} \cdot \mathrm{Im}(\mathbf{A}'(x'))\, \xi \leq 0 \qquad \forall \xi \in \mathbb{C}^2,\ \forall x' \in \bar{D}. \tag{2.37}$$

Finally, let $n \in C(\bar{D})$ be such that $\mathrm{Im}(n(x)) \geq 0$ for all $x' \in \bar{D}$. Then, the problem can be formulated in the following way:

- *given $(f,h) \in H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$, find $(v,u^s) \in H^1(D) \oplus H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ such that*

$$\begin{cases} \nabla_2 \cdot \mathbf{A}' \nabla_2 v + k^2 n v = 0 & \text{in } D & \text{(a)} \\ \Delta_2 u^s + k^2 u^s = 0 & \text{in } \mathbb{R}^2 \setminus \bar{D} & \text{(b)} \\ v - u^s = f & \text{on } \partial D & \text{(c)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} - \dfrac{\partial u^s}{\partial \nu} = h & \text{on } \partial D & \text{(d)} \\ \lim\limits_{r\to\infty} \left[\sqrt{r}\left(\dfrac{\partial u^s}{\partial r} - \mathrm{i}k u^s\right)\right] = 0, & & \text{(e)} \end{cases} \tag{2.38}$$

where the equations (2.38)(a)-(b) are to be intended in the weak sense, the boundary conditions (2.38)(c)-(d) are written in the sense of the trace operators[5] and the radial derivative in (2.38)(e) can be intended in the classical sense (cf. remark 2.2.1 in the following).

Remark 2.1.1. The scattering problem (2.35) is a particular case of problem 2.1.1. More precisely, the interior field v and the scattered field u^s satisfy system (2.38) by making in the latter the following identifications:

$$f = u^i|_{\partial D}, \qquad h = \left.\frac{\partial u^i}{\partial \nu}\right|_{\partial D}, \tag{2.39}$$

[5] See theorems A.15.2, A.17.1 and remarks A.17.1, A.17.2.

where the incident field u^i is such that

$$\Delta_2 u^i + k^2 u^i = 0 \qquad \text{in } \mathbb{R}^2, \tag{2.40}$$

i.e. u^i is an entire solution of the Helmholtz equation, and the right-hand sides of both relations (2.39) are to be intended in the sense of the trace operators. □

Remark 2.1.2. Owing to the symmetry of $\mathrm{Re}(\mathbf{A}'(x'))$ $\forall x' \in \bar{D}$, it holds:

$$\bar{\xi} \cdot \mathrm{Re}(\mathbf{A}'(x'))\, \xi = \mathrm{Re}\left(\bar{\xi} \cdot \mathbf{A}'(x')\, \xi\right) \qquad \forall x' \in \bar{D}, \quad \forall \xi \in \mathbb{C}^2; \tag{2.41}$$

analogously, owing to the symmetry of $\mathrm{Im}(\mathbf{A}'(x'))$ $\forall x' \in \bar{D}$, it holds:

$$\bar{\xi} \cdot \mathrm{Im}(\mathbf{A}'(x'))\, \xi = \mathrm{Im}\left(\bar{\xi} \cdot \mathbf{A}'(x')\, \xi\right) \qquad \forall x' \in \bar{D}, \quad \forall \xi \in \mathbb{C}^2. \tag{2.42}$$

□

Remark 2.1.3. In formulating problem 2.1.1 we have made some technical hypotheses on $\mathbf{A}'$ and n; then, a problem naturally arises: how much do these mathematical hypotheses match real scattering phenomena? In order to answer this question, we should investigate the physical meaning of our assumptions on $\mathbf{A}'$ and n; however, this kind of analysis, although extremely interesting, would lead us too far from our current purposes. Hence, we limit ourselves to pointing out that such hypotheses are general enough to be verified by a large gamut of real materials and we directly refer to [48] (in particular, to chapter XI) for a detailed treatment of electromagnetic wave propagation in anisotropic media. □

We now observe that, in general, any boundary value problem arising in scattering theory (like problem 2.1.1) is formulated in an unbounded domain. However, in order to solve such a problem by means of variational methods providing weak solutions, we need to reformulate it as an *equivalent*[6] problem in a bounded domain. Typically, to this end an open disc Ω_R, centred at the origin and large enough to contain $\bar{D}$, is introduced and at first the problem is solved in $\Omega_R \setminus \bar{D}$. Then, the next step is to extend the solution outside Ω_R in such a way as to get a solution of the original problem. In order to enable such an extension, the crucial point is to choose an appropriate condition to be satisfied on the artificial boundary $\partial\Omega_R$ by the initial solution defined in $\Omega_R \setminus \bar{D}$. It turns out (see [15], section 5.3) that such a condition involves the so-called *Dirichlet to Neumann map*, which we now introduce and describe by means of the two following definitions and the subsequent theorem.

[6]Some different concepts of equivalence between two problems are formulated in the following definition 2.1.3.

Definition 2.1.1. *Let* $\Omega_R := \{x' \in \mathbb{R}^2 \mid \|x'\|_{\mathbb{R}^2} < R\}$*, with* $R > 0$*, and let* $w \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{\Omega}_R)$ *be a weak solution to the Helmholtz equation* $\Delta_2 w + k^2 w = 0$*; then* w *is called* radiating *if it satisfies the* Sommerfeld radiation condition *(2.34), i.e.*

$$\lim_{r\to\infty} \left[\sqrt{r}\left(\frac{\partial w}{\partial r} - \mathrm{i}kw\right)\right] = 0, \tag{2.43}$$

where, as we shall see in remark 2.2.1, the radial derivative can be always intended in the classical sense.

Definition 2.1.2. *Let* w *be a weak radiating solution to the Helmholtz equation,* $\partial\Omega_R$ *the boundary of an open and origin-centred disk* Ω_R *of radius* R*, and* ν_R *the outward unit normal to* $\partial\Omega_R$*. The map* T *such that*

$$T : w|_{\partial\Omega_R} \mapsto T(w|_{\partial\Omega_R}) := \left.\frac{\partial w}{\partial r}\right|_{\partial\Omega_R}, \tag{2.44}$$

where the radial derivative is intended in the classical sense, is called Dirichlet to Neumann map. *In the following, we shall simply write* $Tw = \dfrac{\partial w}{\partial \nu_R}$ *instead of* $T(w|_{\partial\Omega_R}) = \left.\dfrac{\partial w}{\partial r}\right|_{\partial\Omega_R}$.

It is worthwhile observing that in the previous definition 2.1.2 the restrictions $w|_{\partial\Omega_R}$ and $\left.\dfrac{\partial w}{\partial r}\right|_{\partial\Omega_R}$ are not at all defined in terms of the trace operators introduced in theorems A.15.2 and A.17.1; on the contrary, by expressing the radiating solution w as a suitable series expansion involving the Hankel functions of the first kind, it is possible to determine $\left.\dfrac{\partial w}{\partial r}\right|_{\partial\Omega_R}$ (intended in the classical sense) starting from the knowledge of $w|_{\partial\Omega_R}$. See theorem 5.20 in [15] for details.

Theorem 2.1.1. *The Dirichlet to Neumann map* T *is a bounded linear operator from* $H^{\frac{1}{2}}(\partial\Omega_R)$ *to* $H^{-\frac{1}{2}}(\partial\Omega_R)$*.*

Proof. See theorem 5.20 in [15]. ■

Before giving an equivalent formulation of problem 2.1.1, we first fix the concept of equivalence between two problems by means of the following definition.

Definition 2.1.3. *Given two different problems, formulated in the same hypotheses and with the same data, they are said*

1. equivalent with regard to existence *(or, briefly,* existence-equivalent*) if from a solution (assumed existing) of one of the problems, one can get a solution of the other one, and conversely;*

2. equivalent with regard to uniqueness *(or, briefly,* uniqueness-equivalent*) if assuming the uniqueness of the solution (if it exists) of one of the problem, one can prove the uniqueness of the solution (if it exists) of the other one, and conversely;*

3. equivalent *if they are both existence-equivalent and uniqueness-equivalent.*

Now we can consider the following problem, set on a bounded domain, and state its equivalence with the previous problem 2.1.1 by means of the subsequent theorems.

Problem 2.1.2. *Let D, ν, $\mathbf{A}'$, n verify the same assumptions as in problem 2.1.1; let $\Omega_R \supset \bar{D}$ and $T : H^{\frac{1}{2}}(\partial\Omega_R) \to H^{-\frac{1}{2}}(\partial\Omega_R)$ be as in definition 2.1.2. Then,*

- *given $(f, h) \in H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$, find $(v, u^s) \in H^1(D) \oplus H^1(\Omega_R \setminus \bar{D})$ such that*

$$\begin{cases} \nabla_2 \cdot \mathbf{A}'\nabla_2 v + k^2 n v = 0 & \text{in } D & \text{(a)} \\ \Delta_2 u^s + k^2 u^s = 0 & \text{in } \Omega_R \setminus \bar{D} & \text{(b)} \\ v - u^s = f & \text{on } \partial D & \text{(c)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} - \dfrac{\partial u^s}{\partial \nu} = h & \text{on } \partial D & \text{(d)} \\ \dfrac{\partial u^s}{\partial r} = T u^s & \text{on } \partial\Omega_R & \text{(e)} \end{cases} \tag{2.45}$$

where, as before, the equations (2.45)(a)-(b) are to be intended in the weak sense, the boundary conditions (2.45)(c)-(d) are written in the sense of the trace operators and the radial derivative in (2.45)(e) can be intended in the classical sense.

Theorem 2.1.2. *Problems 2.1.1 and 2.1.2 are existence-equivalent; more precisely:*

1. *if $(v, u^s) \in H^1(D) \oplus H^1(\Omega_R \setminus \bar{D})$ is a solution to system (2.45), then u^s can be extended to a function $u^s_{ext} \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ such that (v, u^s_{ext}) is a solution to system (2.38);*
2. *conversely, if $(v, u^s) \in H^1(D) \oplus H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ is a solution to system (2.38), then u^s can be restricted to the function $u^s|_{\Omega_R \setminus \bar{D}} \in H^1(\Omega_R \setminus \bar{D})$ and $(v, u^s|_{\Omega_R \setminus \bar{D}})$ solves system (2.45).*

Proof. See section 5.4 in [15]. ■

Theorem 2.1.3. *Both problems 2.1.1 and 2.1.2 have at most one solution.*

Proof. See section 5.4 in [15]. ■

The previous theorem 2.1.3 trivially implies that problems 2.1.1 and 2.1.2 are uniqueness equivalent; then, remembering theorem 2.1.2, we immediately get the following result.

Theorem 2.1.4. *Problems 2.1.1 and 2.1.2 are equivalent.*

In order to establish the existence of a (unique) solution to problem 2.1.2 and, even more, its well-posedness, we shall resort to an equivalent variational formulation of it. To this end, we still need the concept of *Dirichlet eigenvalue*, which is introduced in the following definition, and an auxiliary lemma, enunciated soon after.

Definition 2.1.4. *Let D be as in problem 2.1.1. The values of k^2 for which there exists a nonzero function $u \in H_0^1(D)$ satisfying*[7]

$$\Delta_2 u + k^2 u = 0 \qquad \text{in} \quad D \tag{2.46}$$

(in the weak sense) are called the Dirichlet eigenvalues *of $-\Delta_2$ and the corresponding nonzero solutions are called the* eigensolutions *for $-\Delta_2$.*

Lemma 2.1.5. *Let D be as in problem 2.1.1; let Ω_R be as in definition 2.1.2 and such that k^2 is not a Dirichlet eigenvalue*[8] *for $-\Delta_2$ in $\Omega_R \setminus \bar{D}$; finally, let $f \in H^{\frac{1}{2}}(\partial D)$. Then there exists a unique solution $u_f \in H^1(\Omega_R \setminus \bar{D})$ to the following weak Dirichlet boundary value problem:*

$$\begin{cases} \Delta_2 u_f + k^2 u_f = 0 & \text{in} \quad \Omega_R \setminus \bar{D} \qquad \text{(a)} \\ u_f = f & \text{on} \quad \partial D \qquad \text{(b)} \\ u_f = 0 & \text{on} \quad \partial\Omega_R. \qquad \text{(c)} \end{cases} \tag{2.47}$$

Proof. See sections 5.3 and 5.4 in [15]. ■

We can now give the equivalent variational formulation of problem 2.1.2; the equivalence in question is stated by the subsequent theorem.

Problem 2.1.3. *Let D, $\mathbf{A}'$, n, Ω_R, T be as in problem 2.1.2. Additionally, let Ω_R be such that k^2 is not a Dirichlet eigenvalue for $-\Delta_2$ in $\Omega_R \setminus \bar{D}$. Then,*

- *given $(f, h) \in H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$, find $w \in H^1(\Omega_R)$ such that*

$$\int_D \left(\nabla_2 \bar{\varphi} \cdot \mathbf{A}' \nabla_2 w - k^2 n\, \bar{\varphi}\, w\right) dx' + \int_{\Omega_R \setminus \bar{D}} \left(\nabla_2 \bar{\varphi} \cdot \nabla_2 w - k^2 \bar{\varphi}\, w\right) dx' - \int_{\partial\Omega_R} \bar{\varphi}\, T w \, d\sigma(x') = \tag{2.48}$$

$$= \int_{\partial D} \bar{\varphi}\, h \, d\sigma(x') - \int_{\partial\Omega_R} \bar{\varphi}\, T u_f \, d\sigma(x') + \int_{\Omega_R \setminus \bar{D}} \left(\nabla_2 \bar{\varphi} \cdot \nabla_2 u_f - k^2 \bar{\varphi}\, u_f\right) dx' \qquad \forall \varphi \in H^1(\Omega_R),$$

where the boundary integrals are to be intended in the pairing sense (as explained about definition (A.122)), u_f is the same function as in lemma 2.1.5 and dx', $d\sigma(x')$ denote the Lebesgue measure respectively on $\mathbb{R}^2$ and on the proper contour.

[7] Note that the zero boundary condition is incorporated in the space $H_0^1(D)$; cf. theorem A.15.3.

[8] It can be shown that this is always possible.

Theorem 2.1.6. *Problems 2.1.2 and 2.1.3 are equivalent*[9]*; more precisely:*

- *if $w \in H^1(\Omega_R)$ is the unique solution to equation (2.48), then the functions v, u^s defined as*

$$v := w|_D, \tag{2.49}$$

$$u^s := w|_{\Omega_R \setminus \bar{D}} - u_f \tag{2.50}$$

  are such that $(v, u^s) \in H^1(D) \oplus H^1(\Omega_R \setminus \bar{D})$ is the unique solution to system (2.45);

- *conversely, if $(v, u^s) \in H^1(D) \oplus H^1(\Omega_R \setminus \bar{D})$ is the unique solution to system (2.45), then the function w defined as*

$$w := \begin{cases} v & \text{in } D \\ u^s + u_f & \text{in } \Omega_R \setminus \bar{D} \end{cases} \tag{2.51}$$

  is such that $w \in H^1(\Omega_R)$ is the unique solution to equation (2.48).

Proof. See section 5.4 in [15]. ■

Finally, by means of the previous variational formulation, stated in problem 2.1.3, it is possible to prove the following theorems, which state the *well-posedness* of the 2D direct scattering problem for an orthotropic medium.

Theorem 2.1.7. *Problem 2.1.3 is well-posed.*

Proof. See section 5.4 in [15]. ■

Theorem 2.1.8. *For any data $(f, h) \in H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$, problem 2.1.2 has a unique solution $(v, u^s) \in H^1(D) \oplus H^1(\Omega_R \setminus \bar{D})$ which satisfies:*

$$\|v\|_{H^1(D)} + \|u^s\|_{H^1(\Omega_R \setminus \bar{D})} \leq C \left(\|f\|_{H^{\frac{1}{2}}(\partial D)} + \|h\|_{H^{-\frac{1}{2}}(\partial D)} \right), \tag{2.52}$$

being $C > 0$ a positive constant independent of f and h.

Proof. See section 5.4 in [15]. ■

[9] Obviously, we are assuming that also in problem 2.1.2 the disk Ω_R is chosen, as it is always possible to do, in such a way that k^2 is not a Dirichlet eigenvalue for $-\Delta_2$ in $\Omega_R \setminus \bar{D}$.

Remark 2.1.4. Let us put, according to the notations of chapter 1:

$$X := H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D), \qquad Y := H^1(D) \oplus H^1(\Omega_R \setminus \bar{D}). \tag{2.53}$$

If not otherwise specified, the direct sum of Hilbert spaces (like X and Y), which is in turn a Hilbert space (cf. section A.1), is regarded as equipped with the norm (A.7) induced by the scalar product (A.6) in it. However, such a norm is equivalent to the infinitely many others defined in (A.4), (A.5); in particular, it is equivalent to the norm obtained by family (A.4) for $p = 1$. Hence, since it is easy to realize that the equations and the boundary conditions forming system (2.45) imply the linear dependence of its solution (v, u^s) on the data (f, h), we can paraphrase the statement of theorem 2.1.8 by saying that solving problem 2.1.2 defines a linear continuous operator from the Hilbert space X to the Hilbert space Y:

$$\begin{array}{rccc} A: & X & \longrightarrow & Y \\ & (f,h) & \longmapsto & (v,u^s). \end{array} \tag{2.54}$$

□

2.2. The far-field pattern of the scattered field

Our aim is now to state some mathematical properties of the scattered field $u^s \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ (cf. system (2.38) and identifications (2.39)), in particular at very large distances from the boundary ∂D.

From now on, we shall denote a generic point in $\mathbb{R}^2$ simply with x instead of x': indeed we shall never come back to a genuine 3D framework, then there is no risk of notational ambiguity.

For all the statements and properties we are going to recall here below (up to theorem 2.2.1), we refer to sections 3.2, 3.3 in [15] and to sections 2.2, 2.3, 2.4, 3.4 in [27].

Let $H_0^{(1)}(t)$ (with $t > 0$) denote the *Hankel function* of the first kind and of order 0; it is not difficult to show that, for a fixed $y \in \mathbb{R}^2$, the function of x defined as

$$\Phi(x, y) := \frac{\mathrm{i}}{4} H_0^{(1)}(k\|x - y\|_{\mathbb{R}^2}), \qquad x \neq y, \tag{2.55}$$

satisfies the Helmholtz equation

$$\Delta_2 u + k^2 u = 0 \tag{2.56}$$

in $\mathbb{R}^2 \setminus \{y\}$; more precisely, $\Phi(x, y)$ is a *fundamental solution* of the Helmholtz equation, i.e. it satisfies the distributional equation:

$$\Delta_2 \Phi(x, y) + k^2 \Phi(x, y) = \delta(x - y), \tag{2.57}$$

where, as usual, we have indicated with $\delta(x - y)$ the Dirac delta set in y.

It is worthwhile pointing out that $\Phi(x,y)$ satisfies the Sommerfeld radiation condition (2.34) and that it is a real-analytic function[10] of the cartesian coordinates x_1, x_2 of the point $x=(x_1,x_2)$, provided that $x \neq y$. Moreover, unlike the 3D case, the analytical expression of $\Phi(x,y)$ in our 2D case is quite involved, so we only recall its asymptotic behaviour for $\|x-y\|_{\mathbb{R}^2} \to 0$ and for $\|x-y\|_{\mathbb{R}^2} \to \infty$:

$$\Phi(x,y) = \frac{1}{2\pi}\log\frac{1}{\|x-y\|_{\mathbb{R}^2}} + \frac{\mathrm{i}}{4} - \frac{1}{2\pi}\log\frac{k}{2} - \frac{C}{2\pi} + O\left(\|x-y\|^2_{\mathbb{R}^2}\log\frac{1}{\|x-y\|_{\mathbb{R}^2}}\right) \quad \text{as } \|x-y\|_{\mathbb{R}^2} \to 0; \tag{2.58}$$

$$\Phi(x,y) = \frac{\mathrm{i}}{4}\sqrt{\frac{2}{\pi k\|x-y\|_{\mathbb{R}^2}}}\exp\left[\mathrm{i}\left(k\|x-y\|_{\mathbb{R}^2} - \frac{\pi}{4}\right)\right] + O\left(\|x-y\|_{\mathbb{R}^2}^{-\frac{3}{2}}\right) \quad \text{as } \|x-y\|_{\mathbb{R}^2} \to \infty, \tag{2.59}$$

where, in relation (2.58), $C>0$ denotes the Euler's constant.

We can now state the following theorem, which provides two useful representation formulas for solutions of the Helmholtz equation considered respectively in a bounded domain or in its complementary.

Theorem 2.2.1. *Let $D \subset \mathbb{R}^2$ be a nonempty, open and bounded set having C^2 boundary ∂D such that the exterior domain $\mathbb{R}^2 \setminus \bar{D}$ is connected. Then,*

1. *if $u^s \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ is a weak solution of the Helmholtz equation in the exterior of D and satisfies the Sommerfeld radiation condition, the following representation for u^s holds:*

$$u^s(x) = \int_{\partial D}\left(u^s(y)\frac{\partial\Phi(x,y)}{\partial\nu(y)} - \frac{\partial u^s}{\partial\nu}(y)\Phi(x,y)\right)d\sigma(y) \qquad \text{for } \ x \in \mathbb{R}^2\setminus\bar{D}; \tag{2.60}$$

2. *if $u \in H^1(D)$ is a weak solution of the Helmholtz equation in D, the following representation for u holds:*

$$u(x) = \int_{\partial D}\left(\frac{\partial u}{\partial\nu}(y)\Phi(x,y) - u(y)\frac{\partial\Phi(x,y)}{\partial\nu(y)}\right)d\sigma(y) \qquad \text{for } \ x \in D. \tag{2.61}$$

Proof. See section 3.3 and remark 5.8 in [15]. ■

Remark 2.2.1. Representations (2.60), (2.61) need some comments. First of all, we observe that by virtue of trace theorems A.15.2 (for $k=2$, $s=1$), A.17.1 and remarks A.17.1, A.17.2,

[10] In section A.4 the reader can find the definition of *real-analytic function*, as well as some theorems about such kind of functions.

it turns out that[11] $u^s|_{\partial D}$, $u|_{\partial D}$, $\Phi(x,\cdot)|_{\partial D} \in H^{\frac{1}{2}}(\partial D)$ and $\frac{\partial u^s}{\partial \nu}$, $\frac{\partial u}{\partial \nu}$, $\frac{\partial \Phi(x,\cdot)}{\partial \nu(\cdot)} \in H^{-\frac{1}{2}}(\partial D)$: this implies that, in general, the boundary integrals in (2.60) and (2.61) are to be intended in the pairing sense, as explained about definition (A.122). In order to clarify this point and to sketch the proof of the fact that representations (2.60) and (2.61) allow one to conclude the analyticity of u^s and u, let us focus on the boundary integral in (2.61) (an analogous discussion will hold also for that in (2.60)).

As a preliminary remark, we begin by remembering that, as observed above, for a fixed y the fundamental solution $\Phi(x,y)$ is a real-analytic function of x_1, x_2 (with $x = (x_1, x_2) \neq y$): this implies, in particular (see section A.4), that for any fixed $y \in \mathbb{R}^2$ and for each $x_0 \neq y$ there exists a neighbourhood $U_{x_0,y}$ of x_0 such that the following Taylor series representation holds:

$$\Phi(x,y) = \sum_{\alpha \in \mathbb{N}^2} \frac{1}{\alpha!} \left(\partial_x^\alpha \Phi(x_0,y)\right) (x - x_0)^\alpha \qquad \forall x \in U_{x_0,y}, \tag{2.62}$$

where the subscript x in ∂_x^α reminds one that the partial derivatives are calculated with respect to $x = (x_1, x_2)$. Of course, all the partial derivatives of $\Phi(x,y)$ enjoy an analogous representation. Moreover, since we want to focus on the boundary integral in (2.61), we choose in representation (2.62) $y \in \partial D$, $x_0 \in D$ and we can always assume that $U_{x_0,y} \cap \partial D = \emptyset$.

In whatever sense the boundary integral in (2.61) is to be intended, by linearity we have that:

$$u(x) = \int\limits_{\partial D} \frac{\partial u}{\partial \nu}(y) \Phi(x,y) d\sigma(y) - \int\limits_{\partial D} u(y) \frac{\partial \Phi(x,y)}{\partial \nu(y)} d\sigma(y) \qquad \forall x \in D. \tag{2.63}$$

Now, let us put:

$$F_1(x) := \int\limits_{\partial D} \frac{\partial u}{\partial \nu}(y) \Phi(x,y) d\sigma(y); \qquad F_2(x) := \int\limits_{\partial D} u(y) \frac{\partial \Phi(x,y)}{\partial \nu(y)} d\sigma(y). \tag{2.64}$$

Recalling definition (A.122), we know that, in general, the previous boundary integrals are to be intended only as a different notation for the pairing between $H^{-\frac{1}{2}}(\partial D)$ and $H^{\frac{1}{2}}(\partial D)$, i.e.:

$$F_1(x) = \left\langle \frac{\partial u}{\partial \nu}, \Phi(x,\cdot)|_{\partial D} \right\rangle_{\partial D}; \qquad F_2(x) = \left\langle \frac{\partial \Phi(x,\cdot)}{\partial \nu(\cdot)}, u|_{\partial D} \right\rangle_{\partial D}. \tag{2.65}$$

We now remember that $H^{\frac{1}{2}}(\partial D)$ is actually (see definition (A.117)) a subset of $L^2(\partial D)$; hence $u|_{\partial D} \in L^2(\partial D)$; on the other hand, by virtue of the analyticity of $\Phi(x,\cdot)$ regarded now as a function of y, one easily realizes that also $\frac{\partial \Phi(x,\cdot)}{\partial \nu(\cdot)} \in L^2(\partial D)$. Hence the integral defining $F_2(x)$ in the second of relations (2.64) does make sense on its own as a Lebesgue integral and, by virtue of what observed just below definition (A.122), is equal to the pairing in the second of relations (2.65). Now, if we arbitrarily choose $x_0 \in D$, write a Taylor series representation

[11] Of course, the restrictions to ∂D are to be intended in the sense of the trace operator.

for $\dfrac{\partial\Phi(x,y)}{\partial\nu(y)}$ (regarded as a function of x) in a neighbourhood of x_0 and substitute such a representation into the (Lebesgue) integral defining $F_2(x)$, it is possible to show, by means of an argument based on the Lebesgue's dominated convergence theorem[12], that the signs $\int_{\partial D}$ and $\sum_{\alpha\in\mathbb{N}^2}$ commute: this yields a power series representation for $F_2(x)$ in a neighbourhood of x_0, i.e. $F_2(\cdot)$ is real-analytic in x_0; since this holds for all $x_0 \in D$, $F_2(\cdot)$ is real-analytic in D.

The proof of the analyticity of $F_1(\cdot)$ is more involved: indeed, the integral defining $F_1(x)$ in the first of relations (2.64) can only been intended in the pairing sense (according to the first of relations (2.65)), since now $\dfrac{\partial u}{\partial\nu} \in H^{-\frac{1}{2}}(\partial D)$ and, as observed just below definition (A.122), in general it holds $H^{-\frac{1}{2}}(\partial D) \supset L^2(\partial D)$. However, we can proceed as follows. As an element of $H^{-\frac{1}{2}}(\partial D)$, $\dfrac{\partial u}{\partial\nu}$ can be identified with a $\mathbb{C}$-valued linear continuous functional on $H^{\frac{1}{2}}(\partial D)$ (see section A.14); but the latter is an Hilbert space, then, by virtue of the Riesz representation theorem, there exists a unique element, say $f_{\partial_\nu u}$, of $H^{\frac{1}{2}}(\partial D)$ such that:

$$\left\langle \frac{\partial u}{\partial\nu}, v|_{\partial D} \right\rangle_{\partial D} = \left(f_{\partial_\nu u}, v\right)_{H^{\frac{1}{2}}(\partial D)} \qquad \forall v \in H^{\frac{1}{2}}(\partial D), \tag{2.66}$$

where the right-hand side of relation (2.66) denotes the scalar product in $H^{\frac{1}{2}}(\partial D)$ between $f_{\partial_\nu u}$ and v. Hence, in particular, it holds:

$$\left\langle \frac{\partial u}{\partial\nu}, \Phi(x,\cdot)|_{\partial D} \right\rangle_{\partial D} = \left(f_{\partial_\nu u}, \Phi(x,\cdot)|_{\partial D}\right)_{H^{\frac{1}{2}}(\partial D)}. \tag{2.67}$$

Moreover, recalling the notations and topics of section A.14 and, in particular, definition (A.118), we have that[13]

$$\left(f_{\partial_\nu u}, \Phi(x,\cdot)|_{\partial D}\right)_{H^{\frac{1}{2}}(\partial D)} = \left([f_{\partial_\nu u}]_\zeta , [\Phi(x,\cdot)|_{\partial D}]_\zeta \right)_{H^{\frac{1}{2}}(\mathbb{R})}. \tag{2.68}$$

Since $H^{\frac{1}{2}}(\mathbb{R})$ is a Hilbert space (see section A.9), by virtue of the Riesz representation theorem we can regard the scalar product at the right-hand side of relation (2.68) as the action of a linear continuous functional $T_{\partial_\nu u} : H^{\frac{1}{2}}(\mathbb{R}) \to \mathbb{C}$ (uniquely determined by $f_{\partial_\nu u}$) on $[\Phi(x,\cdot)|_{\partial D}]_\zeta$, i.e.:

$$\left([f_{\partial_\nu u}]_\zeta , [\Phi(x,\cdot)|_{\partial D}]_\zeta \right)_{H^{\frac{1}{2}}(\mathbb{R})} = T_{\partial_\nu u}\, [\Phi(x,\cdot)|_{\partial D}]_\zeta . \tag{2.69}$$

On the other hand, from theorem A.10.1 (statement No 1) we know that $H^{\frac{1}{2}}(\mathbb{R}) = W^{\frac{1}{2},2}(\mathbb{R})$ with equivalent norms; hence, a linear continuous functional on $H^{\frac{1}{2}}(\mathbb{R})$ is also a linear continuous functional on $W^{\frac{1}{2},2}(\mathbb{R})$. This allows us to regard the right-hand side of relation (2.69) as the action of a functional in $\left[W^{\frac{1}{2},2}(\mathbb{R})\right]^*$ on $[\Phi(x,\cdot)|_{\partial D}]_\zeta \in W^{\frac{1}{2},2}(\mathbb{R})$; but the latter is an

[12] See, for example, [35], p. 43.

[13] For sake of simplicity, we can assume that D is a C^2 hypograph; if it is not, an analogous argument holds by introducing a suitable partition of unity for ∂D.

Hilbert space, then, as before, such a functional is representable in the form of a scalar product in $W^{\frac{1}{2},2}(\mathbb{R})$ between an element $Q_{\partial_\nu u}$ (uniquely determined by $T_{\partial_\nu u}$) with $[\Phi(x,\cdot)|_{\partial D}]_\zeta$ itself, i.e.:

$$T_{\partial_\nu u}\,[\Phi(x,\cdot)|_{\partial D}]_\zeta = \left(Q_{\partial_\nu u}, [\Phi(x,\cdot)|_{\partial D}]_\zeta\right)_{W^{\frac{1}{2},2}(\mathbb{R})}. \tag{2.70}$$

Finally, by virtue of definition (A.72) (with $r = 0$, $s = 1/2$, $n = 1$, $\Omega = \mathbb{R}$), the scalar product at the right-hand side of equality (2.70) can be expressed in terms of the Lebesgue integral of a function constructed in terms of $[\Phi(x,\cdot)|_{\partial D}]_\zeta$; analogously to the case of $F_2(x)$, this allows one to use an argument based on the analyticity (in x) of $[\Phi(x,\cdot)|_{\partial D}]_\zeta$ and on the Lebesgue's dominated convergence theorem in order to prove, in turn, the analyticity of $F_1(\cdot)$ in D.

Summing up, it is possible to deduce from representations (2.60), (2.61) that if the hypotheses of theorem 2.2.1 are satisfied, then weak solutions $u^s \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ or $u \in H^1(D)$ of the Helmholtz equation are real-analytic functions of their independent variables x_1, x_2. Hence these solutions, although defined, a priori, up to a set of zero measure in their domains, are such that also their punctual values are meaningful. Moreover, the analyticity of u^s clearly implies that the radial derivative operator in the Sommerfeld radiation condition (2.34) can be always intended in the classical sense. □

An important consequence of representation formula (2.60) is stated in the following theorem.

Theorem 2.2.2. *If $u \in H^1_{\partial D,loc}(\mathbb{R}^2)$ is an entire weak solution of the Helmholtz equation and satisfies the Sommerfeld radiation condition, it is identically zero onto all $\mathbb{R}^2$.*

Proof. Let x be an arbitrary, but fixed, point in $\mathbb{R}^2$ and let $\Omega_{x,r} := \{y \in \mathbb{R}^2 \mid \|y - x\|_{\mathbb{R}^2} < r\}$, with $r > 0$: then we have that $u \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{\Omega}_{x,r})$; on the other hand, also $\Phi(x,\cdot) \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{\Omega}_{x,r})$. Moreover, both u and $\Phi(x,\cdot)$ satisfy the Sommerfeld radiation condition. Hence, we can apply theorem A.18.2: its thesis (A.160) in our case reads:

$$\int_{\partial D} \left(u(y)\frac{\partial \Phi(x,y)}{\partial \nu(y)} - \Phi(x,y)\frac{\partial u}{\partial \nu}(y)\right) d\sigma(y) = 0. \tag{2.71}$$

If we now compare relations (2.60) and (2.71), we immediately find that $u(x) = 0$. Since such an argument holds for any x in $\mathbb{R}^2$, the thesis follows. ■

Another important result following from the representation formula (2.60) and the asymptotic behaviour (2.59) is stated in the following theorem, which gives a mathematical characterization of a radiating solution of the Helmholtz equation at large distances from the scatterer.

Theorem 2.2.3. *Let D, u^s be as in theorem 2.2.1; let (r,φ) be the polar coordinates of $x = (x_1,x_2) \in \mathbb{R}^2$ and $\hat{x} := (\cos\varphi, \sin\varphi)$ the observation direction; then, the following asymptotic representation holds uniformly in all directions $\hat{x}$:*

$$u^s(x) = \frac{e^{\mathrm{i}kr}}{\sqrt{r}} u_\infty(\varphi) + O(r^{-3/2}) \qquad \text{as} \quad r \to \infty, \tag{2.72}$$

where the $\mathbb{C}$-valued function $u_\infty \in L^2[0,2\pi]$, called the far-field pattern *of u^s, is given by:*

$$u_\infty(\varphi) = \frac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}} \int_{\partial D} \left(u^s(y) \frac{\partial e^{-\mathrm{i}k\hat{x}\cdot y}}{\partial \nu(y)} - e^{-\mathrm{i}k\hat{x}\cdot y} \frac{\partial u^s(y)}{\partial \nu} \right) d\sigma(y). \tag{2.73}$$

Proof. See sections 4.1, 6.1 in [15] and sections 2.2, 3.4 in [27]. ■

In particular, for each $y \in \mathbb{R}^2$, the far-field pattern of the fundamental solution $\Phi(x,y)$ is given by (see section 4.3 in [15]):

$$\Phi_\infty(\varphi, y) = \frac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}}\, e^{-\mathrm{i}k(\cos\varphi, \sin\varphi)\cdot y}, \tag{2.74}$$

which we shall often write, with a slight notational abuse, as:

$$\Phi_\infty(\hat{x}, y) = \gamma e^{-\mathrm{i}k\hat{x}\cdot y}, \tag{2.75}$$

where, obviously, $\gamma := \dfrac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}}$.

Remark 2.2.2. Since, for a fixed y, $e^{-\mathrm{i}k\hat{x}\cdot y}$ is a real-analytic function of φ (being $\hat{x} = (\cos\varphi, \sin\varphi)$), by means of an argument similar to that used in remark 2.2.1 it is possible to deduce from representation (2.73) (where the boundary integral is clearly to be intended in the pairing sense) that if D is as in theorem 2.2.1, then the far-field pattern u_∞ of a radiating solution $u^s \in H^1_{\partial D, loc}(\mathbb{R}^2 \setminus \bar{D})$ of the Helmholtz equation is a real-analytic function of its variable φ.

Moreover, the same representation (2.73) allows one to demonstrate the continuous dependence of the far-field pattern u_∞ on the scattered field u^s; our aim is now to sketch a proof of this property. Following the same approach and using notations analogous to those of remark 2.2.1, we can write:

$$u_\infty(\varphi) = C\,(G_1(\varphi) - G_2(\varphi)) \qquad \forall \varphi \in [0, 2\pi], \tag{2.76}$$

where, having now denoted, for notational convenience, the scalar product in $\mathbb{R}^2$ as $(\cdot,\cdot)_{\mathbb{R}^2}$ instead of "$\cdot$", we have put:

$$C := \frac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}}, \quad G_1(\varphi) := \left\langle \frac{\partial e^{-\mathrm{i}k(\hat{x},\cdot)_{\mathbb{R}^2}}}{\partial \nu(\cdot)}, u^s|_{\partial D} \right\rangle_{\partial D}, \quad G_2(\varphi) := \left\langle \frac{\partial u^s}{\partial \nu}, e^{-\mathrm{i}k(\hat{x},\cdot)_{\mathbb{R}^2}}|_{\partial D} \right\rangle_{\partial D}; \tag{2.77}$$

the pairings in (2.77) are obviously between an element of $H^{-\frac{1}{2}}(\partial D)$ and an element of $H^{\frac{1}{2}}(\partial D)$. Relation (2.76) clearly implies that

$$|u_\infty(\varphi)| \leq |C| \left(|G_1(\varphi)| + |G_2(\varphi)|\right) \qquad \forall \varphi \in [0, 2\pi]. \tag{2.78}$$

Moreover, remembering relation (A.169), we have:

$$|G_1(\varphi)| \leq \left\| \frac{\partial e^{-\mathrm{i}k(\hat{x},\cdot)_{\mathbb{R}^2}}}{\partial \nu(\cdot)} \right\|_{H^{-\frac{1}{2}}(\partial D)} \|u^s|_{\partial D}\|_{H^{\frac{1}{2}}(\partial D)}, \tag{2.79}$$

$$|G_2(\varphi)| \leq \left\| \frac{\partial u^s}{\partial \nu} \right\|_{H^{-\frac{1}{2}}(\partial D)} \left\| e^{-\mathrm{i}k(\hat{x},\cdot)_{\mathbb{R}^2}}|_{\partial D} \right\|_{H^{\frac{1}{2}}(\partial D)}. \tag{2.80}$$

If we put:

$$f_1(\varphi) := \left\| \frac{\partial e^{-\mathrm{i}k(\hat{x},\cdot)_{\mathbb{R}^2}}}{\partial \nu(\cdot)} \right\|_{H^{-\frac{1}{2}}(\partial D)}, \qquad f_2(\varphi) := \left\| e^{-\mathrm{i}k(\hat{x},\cdot)_{\mathbb{R}^2}}|_{\partial D} \right\|_{H^{\frac{1}{2}}(\partial D)} \qquad \forall \varphi \in [0, 2\pi], \tag{2.81}$$

the smoothness (i.e. analyticity) of $e^{-\mathrm{i}k(\hat{x},\cdot)_{\mathbb{R}^2}}$ easily implies that $f_1, f_2 \in L^2[0, 2\pi]$.

Now, let us take $\Omega_R := \{x \in \mathbb{R}^2 \mid \|x\|_{\mathbb{R}^2} < R\}$ large enough, so that $\Omega_R \supset \bar{D}$, and consider the bounded and C^2 domain $\Omega := \Omega_R \setminus \bar{D}$. Since $u^s \in H^1_{\partial D, loc}(\mathbb{R}^2 \setminus \bar{D})$, from the definition itself (A.144) of $H^1_{\partial D, loc}(\mathbb{R}^2 \setminus \bar{D})$ it is evident that $u^s \in H^1(\Omega)$; then we can define, in the sense of the trace operators (see theorems A.15.2 and A.17.1), $u^s|_{\partial\Omega} \in H^{\frac{1}{2}}(\partial\Omega)$ and $\dfrac{\partial u^s}{\partial \nu} \in H^{-\frac{1}{2}}(\partial\Omega)$. Since $\partial\Omega = \partial\Omega_r \cup \partial D$, it is not difficult to realize that

$$\|u^s|_{\partial D}\|_{H^{\frac{1}{2}}(\partial D)} \leq \|u^s|_{\partial D}\|_{H^{\frac{1}{2}}(\partial\Omega)}, \qquad \left\| \frac{\partial u^s}{\partial \nu} \right\|_{H^{-\frac{1}{2}}(\partial D)} \leq \left\| \frac{\partial u^s}{\partial \nu} \right\|_{H^{-\frac{1}{2}}(\partial\Omega)}. \tag{2.82}$$

Substituting relations (2.81), (2.82) into (2.79),(2.80), we immediately get:

$$|G_1(\varphi)| \leq f_1(\varphi) \|u^s|_{\partial D}\|_{H^{\frac{1}{2}}(\partial\Omega)}, \qquad |G_2(\varphi)| \leq f_2(\varphi) \left\| \frac{\partial u^s}{\partial \nu} \right\|_{H^{-\frac{1}{2}}(\partial\Omega)}. \tag{2.83}$$

Moreover, by virtue of relations (A.131) and (A.150), it respectively holds:

$$\|u^s|_{\partial D}\|_{H^{\frac{1}{2}}(\partial\Omega)} \leq C_1 \|u^s\|_{H^1(\Omega)}, \qquad \left\| \frac{\partial u^s}{\partial \nu} \right\|_{H^{-\frac{1}{2}}(\partial\Omega)} \leq C_3 \|u^s\|_{H^1(\Omega)}, \tag{2.84}$$

being C_1 and C_3 two real positive constants (depending, in general, on Ω). By comparing inequalities (2.83) and (2.84), we have:

$$|G_1(\varphi)| \leq C_1 f_1(\varphi) \|u^s\|_{H^1(\Omega)}, \qquad |G_2(\varphi)| \leq C_3 f_2(\varphi) \|u^s\|_{H^1(\Omega)}. \tag{2.85}$$

Hence, if we put $A_1 := |C| C_1$ and $A_3 := |C| C_3$, by virtue of inequalities (2.78) and (2.85) we find:

$$|u_\infty(\varphi)|^2 \leq \left(A_1 f_1(\varphi) + A_2 f_2(\varphi)\right)^2 \|u^s\|^2_{H^1(\Omega)} \qquad \forall \varphi \in [0, 2\pi]. \tag{2.86}$$

Finally, by integrating on $[0, 2\pi]$ both members of the previous inequality (2.86) and then taking their square roots, we find:

$$\|u_\infty\|_{L^2[0,2\pi]} \leq L \, \|u^s\|_{H^1(\Omega_R \setminus \bar{D})} \,, \tag{2.87}$$

having remembered that $\Omega = \Omega_R \setminus \bar{D}$ and having put:

$$L := \left[\int_0^{2\pi} \left(A_1 \, f_1(\varphi) + A_2 \, f_2(\varphi) \right)^2 \, d\varphi \right]^{\frac{1}{2}} . \tag{2.88}$$

Relation (2.87) expresses the continuous dependence of the far-field pattern $u_\infty \in L^2[0, 2\pi]$ on the scattered field $u^s \in H^1_{\partial D, loc}(\mathbb{R}^2 \setminus \bar{D})$. □

The next theorem is a classic result in scattering theory and it was first proved by Rellich [58] and Vekua [66] in 1943. Owing perhaps to wartime conditions, Vekua's paper remained unknown to western scientific community and the theorem is commonly attributed only to Rellich. In its proof, the analyticity of u^s plays an important role.

Theorem 2.2.4. (Rellich's Lemma) *Let D be as in theorem 2.2.1 and let $u^s \in H^1_{\partial D, loc}(\mathbb{R}^2 \setminus \bar{D})$ a solution of the Helmholtz equation such that*

$$\lim_{R \to \infty} \int_{\|y\|_{\mathbb{R}^2} = R} |u^s(y)|^2 \, d\sigma(y) = 0. \tag{2.89}$$

Then $u^s = 0$ in $\mathbb{R}^2 \setminus \bar{D}$.

Proof. See section 3.3 in [15]. ■

A straightforward consequence of representation (2.72) and theorem 2.2.4 is given by the following proposition.

Theorem 2.2.5. *Let u_∞ be the far-field pattern of a radiating solution $u^s \in H^1_{\partial D, loc}(\mathbb{R}^2 \setminus \bar{D})$ of the Helmholtz equation; if $u_\infty(\varphi) = 0 \ \forall \varphi \in [0, 2\pi]$, then $u^s(x) = 0 \ \forall x \in \mathbb{R}^2 \setminus \bar{D}$.*

Proof. By virtue of (2.72), we have:

$$\int_{\|y\|_{\mathbb{R}^2} = R} |u^s(y)|^2 \, d\sigma(y) = \int_0^{2\pi} |u_\infty(\varphi)|^2 d\varphi + O\left(\frac{1}{R}\right) \qquad \text{as} \quad R \to \infty. \tag{2.90}$$

Since $u_\infty(\varphi) = 0 \ \forall \varphi \in [0, 2\pi]$, by Rellich's lemma 2.2.4 we immediately get that $u^s = 0$ in $\mathbb{R}^2 \setminus \bar{D}$. ■

Of course, by virtue of the analyticity of u_∞ stated above, we can replace the hypothesis $u_\infty(\varphi) = 0 \ \forall \varphi \in [0, 2\pi]$ in theorem 2.2.5 with the weaker one: $u_\infty = 0$ on a subset of $[0, 2\pi]$ containing at least an accumulation point.

Remark 2.2.3. We observe that if two radiating fields are equal, they trivially have the same far-field pattern; conversely, by virtue of the previous theorem 2.2.5, we easily realize that if two radiating fields have the same far-field pattern, they are equal. Indeed, it suffices to consider the difference of the two far-field patterns: it is zero and, on the other hand, by superposition, it represents the far-field pattern of the radiating field obtained as difference of the two given radiating fields; but the latter difference is zero by virtue of theorem 2.2.5, i.e. the two radiating fields are equal. Summing up, we can say that theorem 2.2.5 establishes a one-to-one correspondence between radiating fields and their far-field patterns.

For future reference, we state also the following theorem.

Theorem 2.2.6. *Let $u^s \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ be a radiating solution of the Helmholtz equation, $u_\infty \in L^2[0,2\pi]$ its far-field pattern and Ω_R an open disc centred at the origin and containing $\bar{D}$. Then the linear operator K, defined as*

$$\begin{array}{rccc} K: & H^{\frac{1}{2}}(\partial\Omega_R) \oplus H^{-\frac{1}{2}}(\partial\Omega_R) & \longrightarrow & L^2[0,2\pi] \\ & \left(u^s|_{\partial\Omega_R}, \left.\dfrac{\partial u^s}{\partial \nu}\right|_{\partial\Omega_R} \right) & \longmapsto & u_\infty, \end{array} \tag{2.91}$$

is compact.

Proof. See theorems 4.8 and 6.22 in [15]. ■

Having introduced the concept of *far-field pattern*, we can now consider a modified and simplified version of the direct scattering problem 2.1.1, which reads as follows.

Problem 2.2.1. *Let the same hypotheses of problem 2.1.1 hold; then,*

- *given $(f,h) \in H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$, find the far-field pattern $u_\infty \in L^2[0,2\pi]$ of the scattered field u^s uniquely determined by system (2.38).*

Such a problem obviously allows one to define a linear operator B mapping the data (f,h) into the far-field pattern u_∞ of u^s; by virtue of the well-posedness of problem 2.1.1 (as stated by theorem 2.1.8) and of the continuous dependence of u_∞ on u^s (as stated by relation (2.87)), also problem 2.2.1 is well-posed and, in particular, the operator B is continuous. We fix this idea in the following definition.

Definition 2.2.1. *We denote with B the bounded linear operator which maps the data of problem 2.2.1 into its solution:*

$$\begin{array}{rccc} B: & H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D) & \longrightarrow & L^2[0,2\pi] \\ & (f,h) & \longmapsto & u_\infty. \end{array} \tag{2.92}$$

As we shall see in section 2.4, the operator B will play an important role in introducing the linear sampling method. However, for the moment we limit ourselves to pointing out that we are mostly interested in a particular case of the previous problem 2.2.1, in which the incident field is chosen as a *plane wave*. More precisely, we should consider system (2.38) in the particular case in which

$$f = e^{\mathrm{i}kx\cdot d}, \qquad h = \frac{\partial e^{\mathrm{i}kx\cdot d}}{\partial \nu}, \tag{2.93}$$

where $d := (\cos\theta, \sin\theta)$ is the incidence direction. Summing up, we separately formulate the following problem.

Problem 2.2.2. *Let the same hypotheses of problem 2.1.1 hold; then,*

- *find the far-field pattern $u_\infty \in L^2[0, 2\pi]$ of the scattered field $u^s \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ uniquely determined by the following system:*

$$\begin{cases} \nabla_2 \cdot \mathbf{A}'\nabla_2 v + k^2 n v = 0 & \text{in } D & \text{(a)} \\ \Delta_2 u^s + k^2 u^s = 0 & \text{in } \mathbb{R}^2 \setminus \bar{D} & \text{(b)} \\ v - u^s = e^{\mathrm{i}kx\cdot d} & \text{on } \partial D & \text{(c)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} - \dfrac{\partial u^s}{\partial \nu} = \dfrac{\partial e^{\mathrm{i}kx\cdot d}}{\partial \nu} & \text{on } \partial D & \text{(d)} \\ \lim\limits_{r\to\infty} \sqrt{r}\left(\dfrac{\partial u^s}{\partial r} - \mathrm{i}k u^s\right) = 0, & & \text{(e)} \end{cases} \tag{2.94}$$

where, as usual, the equations (2.94)(a)-(b) are to be intended in the weak sense, the boundary conditions (2.94)(c)-(d) are written in the sense of the trace operators[14] *and the radial derivative in (2.94)(e) can be intended in the classical sense.*

From system (2.94) it clearly turns out that the interior field $v(\cdot) := v(\cdot, \theta)$, the scattered field $u^s(\cdot) := u^s(\cdot, \theta)$ and consequently its far-field pattern $u_\infty(\cdot) := u_\infty(\cdot, \theta)$ depend on the incidence angle θ; for example, we could rewrite relation (2.72) as:

$$u^s(x, \theta) = \frac{e^{\mathrm{i}kr}}{\sqrt{r}} u_\infty(\varphi, \theta) + O(r^{-3/2}) \qquad \text{as} \quad r \to \infty. \tag{2.95}$$

Moreover, using Green's second identity and representation (2.73), one can obtain (see [15],

[14] Since the incident field is now a plane wave, i.e. a real-analytic function, one would expect (and it actually can be shown) that under suitable smoothness assumptions on the boundary of the scatterer and on the coefficients of the equations, the fields v and u^s are smooth too, e.g. C^2 inside the domain and C^1 up to the boundary, so that all the derivation operators can be intended in the classical sense. This general statement falls in the class of the so-called regularity results for the solutions of boundary value problems for elliptic partial differential equations. Precise formulations of such results can be found, e.g., in [36] and [50].

section 4.1) the so-called *reciprocity relation*, i.e.[15]

$$u_\infty(\varphi,\theta) = u_\infty(\theta+\pi,\varphi+\pi) \qquad \forall \varphi,\theta \in [0,2\pi], \tag{2.96}$$

which states the (physically obvious) fact that there is no difference if we send an incident field in the direction $d = (\cos\theta, \sin\theta)$ and observe the scattered field in the direction $\hat{x} = (\cos\varphi, \sin\varphi)$ or, conversely, we send the same incident field in the direction $-\hat{x}$ and observe the scattered field in the direction $-d$.

2.3. Formulation of the inverse scattering problem

The inverse scattering problem we have in mind is not one of the possible inverse ones obtained from problem 2.1.1 (nor, strictly speaking, from problem 2.2.2) interchanging the roles of the unknowns and of the data; anyway, it is actually a simplified version of one of the possible inverse problems arising from the direct problem 2.2.2. It reads as follows.

Problem 2.3.1. *Let the same hypotheses of problem 2.1.1 hold and let $u_\infty(\cdot,\theta)$ denote the far-field pattern of the scattered field $u^s(\cdot,\theta)$ determined by system (2.94); then,*

- *determine the support D of the scatterer from a knowledge of the incident plane wave and of the far-field pattern $u_\infty(\varphi,\theta)$ for all the incidence angles $\theta \in [0,2\pi]$ and all the observation angles $\varphi \in [0,2\pi]$.*

We remark that for an orthotropic medium standard examples (see [38], [56]) show that $\mathbf{A}'$ and n are not in fact uniquely determined by knowing the incident plane wave and the far-field pattern $u_\infty(\varphi,\theta)$ for all $\theta \in [0,2\pi]$ and $\varphi \in [0,2\pi]$, but rather, as we shall see in this section (cf. theorem 2.3.14), what is possible to determine is the support D of the inhomogeneity.

We now introduce the following operator, which will play a central role in solving the inverse problem 2.3.1.

Definition 2.3.1. *The linear operator $F : L^2[0,2\pi] \to L^2[0,2\pi]$ defined as*

$$(Fg)(\varphi) := \int_0^{2\pi} u_\infty(\varphi,\theta) g(\theta) d\theta, \tag{2.97}$$

where $u_\infty(\varphi,\theta)$ is the far-field pattern of the radiating field u^s determined by system (2.94), is called the far-field operator (corresponding to (2.94)).

[15] In writing equality (2.96), we are clearly making a little notational abuse: indeed, if, for each $\theta \in [0,2\pi]$, $u_\infty(\cdot,\theta)$ is defined in $[0,2\pi]$, the right-hand side is, in general, meaningful if and only if $\varphi,\theta \in [0,\pi]$. However, such domains of definition are merely conventional and one can easily regard u_∞ as a 2π-periodic function, defined on $\mathbb{R}^2$, of both the variables φ, θ.

Remark 2.3.1. The far-field operator F is not only continuous, but also compact and we shall indirectly prove this property of F in the following: indeed, we shall show (see relation (2.193)) that the factorization $F = B_0 \circ \mathcal{H}$ holds, where B_0 is a compact restriction of B (see theorem 2.4.5) and $\mathcal{H}$ is a continuous operator (see definition 2.4.1 and theorem 2.4.6). □

The next step is to study the injectivity and the denseness of the range of the far-field operator. To this end, we introduce the definition of *Herglotz wave function.*

Definition 2.3.2. *Given $g \in L^2[0, 2\pi]$ and $d = (\cos\theta, \sin\theta)$, the function defined as*

$$v_g(x) := \int_0^{2\pi} e^{\mathrm{i}kx\cdot d} g(\theta) d\theta \tag{2.98}$$

is called a Herglotz wave function (with kernel g).

Taking into account expression (2.98), an easy argument, based on the analyticity (in x) of $e^{\mathrm{i}kx\cdot d}$ and on the Lebesgue's dominated convergence theorem, shows that v_g is a real-analytic function that solves the Helmholtz equation in all $\mathbb{R}^2$.

Moreover, on the one hand, if two Herglotz wave functions have the same kernel g, they are obviously equal; on the other, the following theorem implies that if two Herglotz wave functions are equal, then they have the same kernel g: in other terms, there is a one-to-one correspondence between Herglotz wave functions and their kernels.

Theorem 2.3.1. *If the Herglotz wave function v_g is such that $v_g(x) = 0\ \forall x \in \mathbb{R}^2$, then its kernel g is the zero element in $L^2[0, 2\pi]$.*

Proof. See sections 3.3 and 3.4 in [27]. ■

For future reference, we introduce the following linear operator and state its main properties in the subsequent theorem.

Definition 2.3.3. *Let v_g be a Herglotz wave function with kernel $g \in L^2[0, 2\pi]$; then we define the linear operator $\mathcal{H}_1$ as:*

$$\begin{array}{rccc} \mathcal{H}_1: & L^2[0,2\pi] & \longrightarrow & H^{-\frac{1}{2}}(\partial D) \\ & g & \longmapsto & \left(\dfrac{\partial v_g}{\partial \nu} + \mathrm{i} v_g\right)\Big|_{\partial D}. \end{array} \tag{2.99}$$

Theorem 2.3.2. *The operator $\mathcal{H}_1$ is bounded, injective and has dense range in $H^{-\frac{1}{2}}(\partial D)$.*

Proof. See section 4.3 in [15]. ■

The following theorem is a particular formulation of the superposition principle.

Theorem 2.3.3. *Let $u^s(\cdot,\theta)$ and $u_\infty(\cdot,\theta)$ be the scattered field determined by system (2.94) and its far-field pattern respectively; let $g \in L^2[0,2\pi]$ be given. If we replace the incident plane wave $e^{\mathrm{i}kx\cdot d}$ in system (2.94) with the incident field given by the Herglotz wave function (2.98), then the corresponding scattered field is given by*

$$v_g^s(x) := \int_0^{2\pi} u^s(x,\theta) g(\theta) d\theta \tag{2.100}$$

and its far-field pattern is

$$v_\infty(\varphi) = \int_0^{2\pi} u_\infty(\varphi,\theta) g(\theta) d\theta. \tag{2.101}$$

Proof. Cf. sections 3.3 and 10.4 in [27]. ■

Comparing relations (2.97) and (2.101), we immediately find

$$(Fg)(\varphi) = v_\infty(\varphi) \qquad \forall \varphi \in [0,2\pi]. \tag{2.102}$$

Moreover, for future reference, we observe that the function defined as

$$\tilde{v}_g(x) := \int_0^{2\pi} e^{-\mathrm{i}kx\cdot d} g(\theta) d\theta \tag{2.103}$$

is also a Herglotz wave function with kernel $g(\theta - \pi)$.

Now we have all the ingredients needed to state the following theorem, which gives a necessary and sufficient condition for the far-field operator F to be injective with dense range. For future reference, we give the proof.

Theorem 2.3.4. *The far-field operator F corresponding to (2.94) is injective with dense range if and only if there does not exist a Herglotz wave function v_g such that the pair $(v, v_g) \in H^1(D) \oplus H^1(D)$ is a solution to the following system*[16]*:*

$$\begin{cases} \nabla_2 \cdot \mathbf{A}' \nabla_2 v + k^2 n v = 0 & \text{in } D \quad \text{(a)} \\ \Delta_2 v_g + k^2 v_g = 0 & \text{in } D \quad \text{(b)} \\ v = v_g & \text{on } \partial D \quad \text{(c)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} = \dfrac{\partial v_g}{\partial \nu} & \text{on } \partial D. \quad \text{(d)} \end{cases} \tag{2.104}$$

[16] From now on, when we shall consider a boundary value problem, we shall not repeat that the partial differential equations are to be intended in the weak sense, the boundary conditions are written in the sense of the trace operators and, if any, the radial derivatives can be intended in the classical sense.

Proof. First of all, we prove that the far-field operator F is injective if and only if its adjoint operator F^* is injective. By definition of adjoint, we have that

$$(Fg, h)_{L^2[0,2\pi]} = (g, F^*h)_{L^2[0,2\pi]} \qquad \forall g, h \in L^2[0, 2\pi], \tag{2.105}$$

i.e.

$$\int_0^{2\pi} \left[\int_0^{2\pi} u_\infty(\varphi, \theta) g(\theta) d\theta \right] \overline{h(\varphi)}\, d\varphi = \int_0^{2\pi} g(\varphi) \overline{(F^*h)(\varphi)}\, d\varphi \qquad \forall g, h \in L^2[0, 2\pi]. \tag{2.106}$$

If we now treat the left-hand side of the previous relation (2.106) by applying Tonelli's and Fubini's theorems[17], regarding u_∞ and h as periodic functions of period 2π, using the reciprocity relation (2.96), interchanging the name of the integration variables and remembering the definition (2.97) of the operator F, we get:

$$\begin{aligned}
(Fg, h)_{L^2[0,2\pi]} &= \int_0^{2\pi}\int_0^{2\pi} u_\infty(\varphi, \theta) g(\theta) \overline{h(\varphi)} d\theta d\varphi = \int_0^{2\pi}\int_0^{2\pi} g(\theta) u_\infty(\varphi, \theta) \overline{h(\varphi)} d\theta d\varphi = \\
&= \int_0^{2\pi}\int_0^{2\pi} g(\theta) u_\infty(\theta + \pi, \varphi + \pi) \overline{h(\varphi)} d\theta d\varphi = \int_0^{2\pi}\int_0^{2\pi} g(\varphi) u_\infty(\varphi + \pi, \theta + \pi) \overline{h(\theta)} d\varphi d\theta = \\
&= \int_0^{2\pi} g(\varphi) \left[\int_0^{2\pi} u_\infty(\varphi + \pi, \theta) \overline{h(\theta - \pi)} d\theta \right] d\varphi = \int_0^{2\pi} g(\varphi) (Fp)(\varphi + \pi) d\varphi,
\end{aligned} \tag{2.107}$$

where $p(\theta) := \overline{h(\theta - \pi)}$. Hence, by comparison between the right-hand side of definition (2.106) and the last member of equalities (2.107), we obtain:

$$(F^*h)(\varphi) = \overline{(Fp)(\varphi + \pi)}. \tag{2.108}$$

The previous relation (2.108) proves that F is injective if and only if F^* is injective. Moreover, since $\mathcal{N}(F^*)^\perp = \overline{\mathcal{R}(F)}$, to prove the theorem it suffices to show that F is injective if and only if there does not exist a Herglotz wave function v_g such that the pair $(v, v_g) \in H^1(D) \oplus H^1(D)$ is a solution to system (2.104), or, equivalently, that F is not injective if and only if there exists a Herglotz wave function v_g such that the pair $(v, v_g) \in H^1(D) \oplus H^1(D)$ is a solution to system (2.104). We shall follow the latter way.

Then, let us assume that F is not injective; now, $Fg = 0$ with $g \neq 0$ is equivalent, by virtue of theorems 2.3.1, 2.3.3 and relation (2.102), to the existence of a nonzero Herglotz wave function v_g (with kernel g) such that the scattered field v_g^s determined by system (2.94) written replacing $e^{ikx\cdot d}$ with v_g has zero far-field pattern v_∞. By theorem 2.2.5, we have that $v_g^s = 0$ in $\mathbb{R}^2 \setminus \bar{D}$; hence the transmission conditions (2.94)(c)-(d) become in our case

$$v = v_g \qquad \text{and} \qquad \frac{\partial v}{\partial \nu_{\mathbf{A}'}} = \frac{\partial v_g}{\partial \nu} \quad \text{on} \quad \partial D, \tag{2.109}$$

[17] See, for example, [35], p. 45.

i.e. exactly conditions (2.104)(c)-(d). Finally, since the differential equations (2.94)(a) and (2.104)(a) for v are identical and, on the other hand, any Herglotz wave function v_g satisfies the Helmholtz equation, i.e. (2.104)(b), we have that the pair $(v, v_g) \in H^1(D) \oplus H^1(D)$ satisfies equations (2.104)(a)-(b) as well. This concludes the proof. ■

Motivated by the previous theorem 2.3.4, we now state, in a general form, the *interior transmission problem* associated with problem 2.1.1.

Problem 2.3.2. (Interior transmission problem) *Let the same hypotheses of problem 2.1.1 hold; then,*

- *given* $(f, h) \in H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$*, find two functions* $(v, w) \in H^1(D) \oplus H^1(D)$ *solving the system:*

$$\begin{cases} \nabla_2 \cdot \mathbf{A}' \nabla_2 v + k^2 n v = 0 & \text{in } D \quad \text{(a)} \\ \Delta_2 w + k^2 w = 0 & \text{in } D \quad \text{(b)} \\ v - w = f & \text{on } \partial D \quad \text{(c)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} - \dfrac{\partial w}{\partial \nu} = h & \text{on } \partial D. \quad \text{(d)} \end{cases} \tag{2.110}$$

Definition 2.3.4. *The boundary value problem (2.110) with* $f = 0$ *and* $h = 0$ *is named the* homogeneous interior transmission problem*. Values of* k^2 *for which the homogeneous interior transmission problem has a nontrivial solution are called* transmission eigenvalues*.*

Hence, theorem 2.3.4 can be partially paraphrased saying that if k^2 is not a transmission eigenvalue for the homogeneous interior transmission problem (2.104), then the far-field operator corresponding to (2.94) is injective with dense range.

Since, as we shall see in section 2.4, the linear sampling method is based on the assumption that k^2 is not a transmission eigenvalue, it will be of particular interest to establish if and when transmission eigenvalues exist. In the final part of this section we shall give some partial answers to this problem. Anyway, we now begin by establishing (under suitable hypotheses) the uniqueness of the solution to problem 2.3.2.

Theorem 2.3.5. *Let the same hypotheses of problem 2.1.1 hold; moreover, let us assume that there exists a point* $x_0 \in D$ *such that either*

$$\mathrm{Im}(n(x_0)) > 0 \tag{2.111}$$

or

$$\mathrm{Im}(\bar{\xi} \cdot \mathbf{A}'(x_0)\, \xi) < 0 \qquad \forall \xi \in \mathbb{C}^2 \setminus (0,0). \tag{2.112}$$

Then the interior transmission problem 2.3.2 has at most one solution.

Proof. See section 6.2 in [15]. ■

In order to establish also the existence of a solution to problem 2.3.2, we firstly need to study the following intermediate problem.

Problem 2.3.3. (Modified interior transmission problem) *Let D, $\mathbf{A}'$ be as in problem 2.3.2; let the real-valued function $m \in C(\bar{D})$ and the (complex-valued) functions ρ_1, $\rho_2 \in L^2[0, 2\pi]$ be assigned. Then,*

- *given $(f, g) \in H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$, find $(v, w) \in H^1(D) \oplus H^1(D)$ solving the system:*

$$\begin{cases} \nabla_2 \cdot \mathbf{A}' \nabla_2 v - mv = \rho_1 & \text{in } D \quad \text{(a)} \\ \Delta_2 w - w = \rho_2 & \text{in } D \quad \text{(b)} \\ v - w = f & \text{on } \partial D \quad \text{(c)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} - \dfrac{\partial w}{\partial \nu} = h & \text{on } \partial D. \quad \text{(d)} \end{cases} \tag{2.113}$$

Our aim is now to reformulate system (2.113) as an equivalent variational problem and state the well-posedness of the latter. To this end, we define the Hilbert space[18]

$$W(D) := \left\{ \mathbf{w} \in L^2(D) \oplus L^2(D) \mid \nabla_2 \cdot \mathbf{w} \in L^2(D) \text{ and } \nabla_2 \times \mathbf{w} = 0 \right\}, \tag{2.115}$$

equipped with the inner product

$$(\mathbf{w}_1, \mathbf{w}_2)_W := (\mathbf{w}_1, \mathbf{w}_2)_{L^2(D)\oplus L^2(D)} + (\nabla_2 \cdot \mathbf{w}_1, \nabla_2 \cdot \mathbf{w}_2)_{L^2(D)} \tag{2.116}$$

and the norm

$$\|\mathbf{w}\|_W^2 := \|\mathbf{w}\|_{L^2(D)\oplus L^2(D)}^2 + \|\nabla_2 \cdot \mathbf{w}\|_{L^2(D)}^2. \tag{2.117}$$

Now, let $(\varphi, \boldsymbol{\psi}) \in H^1(D) \oplus W(D)$ and, as usual, let ν be the unit outward normal to ∂D: then $\varphi|_{\partial D} \in H^{\frac{1}{2}}(\partial D)$ by virtue of theorem A.15.2; on the other hand, it is possible to prove[19] that $(\boldsymbol{\psi} \cdot \nu)|_{\partial D} \in H^{-\frac{1}{2}}(\partial D)$ and to apply a generalized version of the divergence theorem to the field $\varphi\, \boldsymbol{\psi}$. Since $\nabla_2 \cdot (\varphi\, \boldsymbol{\psi}) = \nabla_2 \varphi \cdot \boldsymbol{\psi} + \varphi\, \nabla_2 \cdot \boldsymbol{\psi}$, this allows one to write:

$$\int_D \varphi\, \nabla_2 \cdot \boldsymbol{\psi}\, dx + \int_D \nabla_2 \varphi \cdot \boldsymbol{\psi}\, dx = \int_{\partial D} \varphi\, \boldsymbol{\psi} \cdot \nu\, d\sigma. \tag{2.118}$$

[18] In definition (2.115), we use two shorthands: if $\mathbf{w} = (w_1, w_2) \in L^2(D) \oplus L^2(D)$, we have (obviously, in the weak sense):

$$\nabla_2 \cdot \mathbf{w} := \frac{\partial w_1}{\partial x_1} + \frac{\partial w_2}{\partial x_2}, \qquad \nabla_2 \times \mathbf{w} := \frac{\partial w_1}{\partial x_2} - \frac{\partial w_2}{\partial x_1}. \tag{2.114}$$

[19] We refer to chapter 5 in [15] for details.

If we observe that, by virtue of definition (A.122), the right-hand side of identity (2.118) is just the duality pairing between $\varphi|_{\partial D} \in H^{\frac{1}{2}}(\partial D)$ and $(\boldsymbol{\psi} \cdot \nu)|_{\partial D} \in H^{-\frac{1}{2}}(\partial D)$, we immediately find:

$$\langle \varphi, \boldsymbol{\psi} \cdot \nu \rangle_{\partial D} = \int_D \varphi \, \nabla_2 \cdot \boldsymbol{\psi} \, dx + \int_D \nabla_2 \varphi \cdot \boldsymbol{\psi} \, dx, \tag{2.119}$$

where, in the left-hand side of equality (2.119), we have omitted the sign of restriction to ∂D as subscript of the elements involved, keeping it only as subscript of the pairing sign.

We can now give the equivalent variational formulation of problem 2.3.3; the equivalence in question is stated by the subsequent theorem.

Problem 2.3.4. *Let D, $\mathbf{A}'$, m, ρ_1 and ρ_2 be as in problem 2.3.3. Then,*

- *given $(f, h) \in H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$, find $(v, \mathbf{w}) \in H^1(D) \oplus W(D)$ such that, for all $(\varphi, \boldsymbol{\psi}) \in H^1(D) \oplus W(D)$, it holds:*

$$\int_D \mathbf{A}' \nabla_2 v \cdot \nabla_2 \bar{\varphi} \, dx + \int_D m v \bar{\varphi} \, dx + \int_D \nabla_2 \cdot \mathbf{w} \, \nabla_2 \bar{\boldsymbol{\psi}} \, dx + \int_D \mathbf{w} \cdot \bar{\boldsymbol{\psi}} \, dx - \langle v, \bar{\boldsymbol{\psi}} \cdot \nu \rangle_{\partial D} + \\ - \langle \bar{\varphi}, \mathbf{w} \cdot \nu \rangle_{\partial D} = \int_D \left(\rho_1 \bar{\varphi} + \rho_2 \nabla_2 \cdot \bar{\boldsymbol{\psi}} \right) dx + \langle \bar{\varphi}, h \rangle_{\partial D} - \langle f, \bar{\boldsymbol{\psi}} \cdot \nu \rangle_{\partial D}. \tag{2.120}$$

Theorem 2.3.6. *Problems 2.3.3 and 2.3.4 are equivalent; more precisely:*

- *if $(v, w) \in H^1(D) \oplus H^1(D)$ is the unique solution to system (2.113), then $(v, \nabla_2 w) \in H^1(D) \oplus W(D)$ is the unique solution to equation (2.120);*
- *conversely, if $(v, \mathbf{w}) \in H^1(D) \oplus W(D)$ is the unique solution to equation (2.120), then the unique solution $(v, w) \in H^1(D) \oplus H^1(D)$ to system (2.113) is such that $\mathbf{w} = \nabla_2 w$.*

Proof. See section 6.2 in [15]. ■

The following theorem states, under suitable hypotheses, the well-posedness[20] of problem 2.3.4.

Theorem 2.3.7. *Let D, $\mathbf{A}'$, m, ρ_1 and ρ_2 be as in problem 2.3.4; besides, let us assume that there exists a constant $\gamma > 1$ such that, $\forall x \in D$, it holds:*

$$\mathrm{Re}\left(\bar{\xi} \cdot \mathbf{A}'(x)\, \xi\right) \geq \gamma \|\xi\|^2_{\mathbb{C}^2} \qquad \forall \xi \in \mathbb{C}^2; \qquad m(x) \geq \gamma. \tag{2.121}$$

Then problem 2.3.4 has a unique solution $(v, \mathbf{w}) \in H^1(D) \oplus W(D)$, which satisfies the a priori estimate

$$\|v\|_{H^1(D)} + \|\mathbf{w}\|_W \leq 2C \frac{\gamma + 1}{\gamma - 1} \left(\|\rho_1\|_{L^2(D)} + \|\rho_2\|_{L^2(D)} + \|f\|_{H^{\frac{1}{2}}(\partial D)} + \|h\|_{H^{-\frac{1}{2}}(\partial D)} \right), \tag{2.122}$$

where the constant $C > 0$ is independent of ρ_1, ρ_2, f, h and γ.

[20] As regards the well-posedness, from theorem 2.3.7 to theorem 2.3.11 one should remember the question of the equivalence of the norms, as already discussed in remark 2.1.4.

Proof. See section 6.2 in [15]. ■

As a consequence of theorems 2.3.6 and 2.3.7, we can now state the well-posedness of problem 2.3.3 as follows.

Theorem 2.3.8. *Let the same hypotheses of theorem 2.3.7 hold; then problem 2.3.3 has a unique solution* $(v, w) \in H^1(D) \oplus H^1(D)$*, which satisfies the a priori estimate:*

$$\|v\|_{H^1(D)} + \|w\|_{H^1(D)} \leq C\frac{\gamma+1}{\gamma-1}\left(\|\rho_1\|_{L^2(D)} + \|\rho_2\|_{L^2(D)} + \|f\|_{H^{\frac{1}{2}}(\partial D)} + \|h\|_{H^{-\frac{1}{2}}(\partial D)}\right), \quad (2.123)$$

where the constant $C > 0$ *is independent of* ρ_1*,* ρ_2*,* f*,* h *and* γ*.*

Proof. See section 6.2 in [15]. ■

Theorems 2.3.5 and 2.3.8 now allow one to state the well-posedness of the interior transmission problem 2.3.2.

Theorem 2.3.9. *Let the same hypotheses of theorem 2.3.5 hold; moreover, let us assume that there exists a constant* $\gamma > 1$ *such that,* $\forall x \in D$*, it holds:*

$$\mathrm{Re}\left(\bar{\xi} \cdot \mathbf{A}'(x)\,\xi\right) \geq \gamma\|\xi\|^2_{\mathbb{C}^2} \qquad \forall \xi \in \mathbb{C}^2. \quad (2.124)$$

Then problem 2.3.2 has a unique solution $(v, w) \in H^1(D) \oplus H^1(D)$*, which satisfies the a priori estimate:*

$$\|v\|_{H^1(D)} + \|w\|_{H^1(D)} \leq C\left(\|f\|_{H^{\frac{1}{2}}(\partial D)} + \|h\|_{H^{-\frac{1}{2}}(\partial D)}\right), \quad (2.125)$$

where the constant $C > 0$ *is independent of* f *and* h*.*

Proof. See section 6.2 in [15]. ■

We point out that hypothesis (2.124) implies, in particular, that $\|\mathrm{Re}(\mathbf{A}')(x)\| > 1$ $\forall x \in D$: hence, the case $\mathbf{A}'(x) = \mathbf{I}$ $\forall x \in D$ (where $\mathbf{I}$ is the identity matrix) is not contemplated by the previous theorem. Anyway, the case of $\mathrm{Re}(\mathbf{A}')(x)$ positive definite such that $\|\mathrm{Re}(\mathbf{A}')(x)\| < 1$ $\forall x \in D$ (which excludes again the possibility $\mathbf{A}'(x) = \mathbf{I}$ $\forall x \in D$) is considered in [21]: in this paper it is shown that, by modifying the variational approach of theorems 2.3.6 and 2.3.7, one can prove the two following results.

Theorem 2.3.10. *Let* D*,* $\mathbf{A}'$*,* m*,* ρ_1 *and* ρ_2 *be as in problem 2.3.3; besides, let us assume that there exists a constant* $\gamma > 1$ *such that,* $\forall x \in D$*, it holds:*

$$\mathrm{Re}\left(\bar{\xi} \cdot (\mathbf{A}'(x))^{-1}\,\xi\right) \geq \gamma\|\xi\|^2_{\mathbb{C}^2} \qquad \forall \xi \in \mathbb{C}^2; \qquad \gamma^{-1} \leq m(x) < 1. \quad (2.126)$$

Then problem 2.3.3 has a unique solution $(v, w) \in H^1(D) \oplus H^1(D)$*, which satisfies the a priori estimate:*

$$\|v\|_{H^1(D)} + \|w\|_{H^1(D)} \leq C \left(\|\rho_1\|_{L^2(D)} + \|\rho_2\|_{L^2(D)} + \|f\|_{H^{\frac{1}{2}}(\partial D)} + \|h\|_{H^{-\frac{1}{2}}(\partial D)} \right), \tag{2.127}$$

where the constant $C > 0$ *is independent of* ρ_1*,* ρ_2*,* f *and* h*.*

Theorem 2.3.11. *Let the same hypotheses of theorem 2.3.5 hold; moreover, let us assume that there exists a constant* $\gamma > 1$ *such that,* $\forall x \in D$*, it holds:*

$$\mathrm{Re}\left(\bar{\xi} \cdot (\mathbf{A}'(x))^{-1} \xi \right) \geq \gamma \|\xi\|^2_{\mathbb{C}^2} \qquad \forall \xi \in \mathbb{C}^2. \tag{2.128}$$

Then problem 2.3.2 has a unique solution $(v, w) \in H^1(D) \oplus H^1(D)$*, which satisfies the a priori estimate:*

$$\|v\|_{H^1(D)} + \|w\|_{H^1(D)} \leq C \left(\|f\|_{H^{\frac{1}{2}}(\partial D)} + \|h\|_{H^{-\frac{1}{2}}(\partial D)} \right), \tag{2.129}$$

where the constant $C > 0$ *is independent of* f *and* h*.*

Of course, the previous theorem 2.3.11 states the well-posedness of the interior transmission problem 2.3.2 under the specified hypotheses.

In general, if $\mathbf{A}'$ and n do not satisfy the assumptions of either theorem 2.3.9 or theorem 2.3.11, one cannot state the well-posedness of problem 2.3.2. In particular, if the hypotheses of theorem 2.3.5 (which are included in those of theorems 2.3.9 and 2.3.11) are not verified, it might happen that k^2 is a transmission eigenvalue: this would clearly imply, by linearity, the non-uniqueness of the solution of problem 2.3.2 itself. Anyway, do transmission eigenvalues exist and, if so, do they form a discrete set? In general, it is not known if transmission eigenvalues exist. The only known result concerning the existence of transmission eigenvalues is for the case in which the matrix-valued function $\mathbf{A}'(x)$ is the identity matrix $\mathbf{I}$ for all $x \in D$ and $n(x)$ has the particular radial form $n(r)$ (see theorem 8.13 in [27]).

Here below we give two theorems providing partial answers to the questions just asked.

Theorem 2.3.12. *Let the same hypotheses of problem 2.3.2 hold; moreover, let us assume that*

$$\mathrm{Im}\,(n(x)) = 0, \qquad \mathrm{Im}\,(\mathbf{A}'(x)) = 0 \qquad \forall x \in D, \tag{2.130}$$

and that there exists a constant $\gamma > 1$ *such that,* $\forall x \in D$*, it holds:*

$$\bar{\xi} \cdot \mathbf{A}'(x)\, \xi \geq \gamma \|\xi\|_{\mathbb{C}^2} \qquad \forall \xi \in \mathbb{C}^2; \qquad\qquad n(x) \geq \gamma. \tag{2.131}$$

Then the set of transmission eigenvalues is either empty or discrete.

Proof. See section 6.2 in [15]. ■

Theorem 2.3.13. *Let the same hypotheses of problem 2.3.2 hold; moreover, let us assume that*

$$\mathrm{Im}\,(n(x)) = 0, \qquad \mathrm{Im}\,(\mathbf{A}'(x)) = 0 \qquad \forall x \in D, \tag{2.132}$$

and that there exists a constant $\gamma > 1$ such that, $\forall x \in D$, it holds:

$$\bar{\xi} \cdot (\mathbf{A}'(x))^{-1}\, \xi \geq \gamma \|\xi\|_{\mathbb{C}^2} \qquad \forall \xi \in \mathbb{C}^2; \qquad \gamma^{-1} \leq n(x) < 1. \tag{2.133}$$

Then the set of transmission eigenvalues is either empty or discrete.

Proof. See section 6.2 in [15]. ■

For the reasons explained above, theorems 2.3.12 and 2.3.13, as well as theorems 2.3.9 and 2.3.11, exclude the case $\mathbf{A}'(x) = \mathbf{I}\ \forall x \in D$; such a case can be treated either by rewriting the system (2.110) of the interior transmission problem 2.3.2 as a boundary value problem for a fourth order partial differential equation for the difference $v - w \in H^2(D)$ (see [60]) or by using analytic projection operators (see section 8.6 in [27]). The case in which $\|\mathrm{Re}\,(\mathbf{A}'(x))\| > 1$ for $x \in D_0 \subset D$ and $\|\mathrm{Re}\,(\mathbf{A}'(x))\| < 1$ for $x \in D \setminus \bar{D}_0$ is still an open problem.

We conclude this section by stating the uniqueness for the inverse medium scattering problem 2.3.1.

Theorem 2.3.14. *Let the domains D_1 and D_2, the matrix-valued functions $\mathbf{A}'_\mathbf{1}$ and $\mathbf{A}'_\mathbf{2}$, the functions n_1 and n_2 satisfy the hypotheses of problem 2.1.1 (and, consequently, of problem 2.2.2). Moreover, let us assume that there exists a constant $\gamma > 1$ such that, $\forall x \in D$, the two following conditions hold:*

1. *either* $\bar{\xi} \cdot \mathrm{Re}(\mathbf{A}'_\mathbf{1}(x))\, \xi \geq \gamma \|\xi\|^2_{\mathbb{C}^2}$ *or* $\bar{\xi} \cdot \mathrm{Re}(\mathbf{A}'_\mathbf{1}(x))^{-1}\, \xi \geq \gamma \|\xi\|^2_{\mathbb{C}^2} \qquad \forall \xi \in \mathbb{C}^2$;
2. *either* $\bar{\xi} \cdot \mathrm{Re}(\mathbf{A}'_\mathbf{2}(x))\, \xi \geq \gamma \|\xi\|^2_{\mathbb{C}^2}$ *or* $\bar{\xi} \cdot \mathrm{Re}(\mathbf{A}'_\mathbf{2}(x))^{-1}\, \xi \geq \gamma \|\xi\|^2_{\mathbb{C}^2} \qquad \forall \xi \in \mathbb{C}^2$.

Finally, denoting with $u^1_\infty(\cdot,\theta)$ and $u^2_\infty(\cdot,\theta)$ the far-field patterns of the radiating fields $u^s_1(\cdot,\theta)$ and $u^s_2(\cdot,\theta)$ respectively determined by system (2.94) written for either D_1, $\mathbf{A}'_\mathbf{1}$, n_1 or D_2, $\mathbf{A}'_\mathbf{2}$, n_2, let us assume that

$$u^1_\infty(\varphi,\theta) = u^2_\infty(\varphi,\theta) \qquad \forall \varphi,\theta \in [0,2\pi]. \tag{2.134}$$

Then it holds $D_1 = D_2$.

Proof. See section 6.3 in [15]. ■

This theorem clearly implies that the far-field pattern contains information enough to allow one to determine, at least in principle, the support D of the inhomogeneity: in other terms, under the hypotheses of theorem 2.3.14, the solution to problem 2.3.1, if it exists, is unique. However, such a problem is an inverse one, then it involves the pathologies and the consequent regularization methods explained in chapter 1.

2.4. The general theorem

The next step is now to determine the range of the operator B introduced by definition 2.2.1. To this end, it is more convenient to consider its transpose[21] rather than its adjoint, since operating with the duality relation between $H^{\frac{1}{2}}(\partial D)$ and $H^{-\frac{1}{2}}(\partial D)$ is much simpler then using the corresponding inner products. By means of theorem A.19.6, we can prove the following result for the operator B.

Theorem 2.4.1. *The bounded linear operator $B : H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D) \to L^2[0, 2\pi]$ has dense range in $L^2[0, 2\pi]$.*

Proof. The transpose operator of B is

$$\begin{array}{rccc} B^T : & L^2[0,2\pi] & \longrightarrow & H^{-\frac{1}{2}}(\partial D) \oplus H^{\frac{1}{2}}(\partial D) \\ & g & \longmapsto & (\tilde{f}, \tilde{h}), \end{array} \tag{2.135}$$

where the functions $(\tilde{f}, \tilde{h})$ are such that the general definition (A.174) of transpose operator is satisfied. The next step is to give a specific form for such a definition in our case and consequently to obtain an explicit expression of $\tilde{f}, \tilde{h}$ in terms of g. For notational convenience, from now on we put

$$X := H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D), \tag{2.136}$$

$$Y := L^2[0, 2\pi]; \tag{2.137}$$

this implies (see theorem A.1.4 and relation (A.51)):

$$X^* = H^{-\frac{1}{2}}(\partial D) \oplus H^{\frac{1}{2}}(\partial D), \tag{2.138}$$

$$Y^* = L^2[0, 2\pi]. \tag{2.139}$$

Then definition (A.174) in our case reads

$$\left\langle B^T g, (f, h) \right\rangle_{X^*, X} = \langle g, B(f, h) \rangle_{Y^*, Y} \qquad \forall (f, h) \in X, \quad \forall g \in Y^*, \tag{2.140}$$

and it can also be written, recalling (2.135), as

$$\left\langle (\tilde{f}, \tilde{h}), (f, h) \right\rangle_{X^*, X} = \langle g, B(f, h) \rangle_{Y^*, Y} \qquad \forall (f, h) \in X, \quad \forall g \in Y^*, \tag{2.141}$$

i.e.

$$\left\langle \tilde{f}, f \right\rangle_{H^{-\frac{1}{2}}(\partial D), H^{\frac{1}{2}}(\partial D)} + \left\langle \tilde{h}, h \right\rangle_{H^{\frac{1}{2}}(\partial D), H^{-\frac{1}{2}}(\partial D)} = \langle g, B(f, h) \rangle_{Y^*, Y} \qquad \forall (f, h) \in X, \quad \forall g \in Y^*. \tag{2.142}$$

[21] See definition A.174 and, more generally, section A.19 for some elements about transpose operators.

Now, let us consider the Herglotz wave function $\tilde{v}_g(x)$ defined in (2.103); if the boundary data functions (f,h) in problem 2.1.1 are given in terms of $\tilde{v}_g(x)$ as

$$(f,h) := \left(\tilde{v}_g|_{\partial D}, \frac{\partial \tilde{v}_g}{\partial \nu}\Big|_{\partial D}\right) \in H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D), \tag{2.143}$$

we denote with $(\tilde{v}, \tilde{u}^s) \in H^1(D) \oplus H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ the unique solution to system (2.94) with data (2.143).

On the other hand, let, as usual, $(v, u^s) \in H^1(D) \oplus H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ be the unique solution to system (2.94) itself for generic data $(f,h) \in H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$ and let u_∞ be the far-field pattern of the scattered field u^s.

Substituting representation (2.73) for the far-field pattern of a radiating field u^s into the explicit expression of the duality pairing $\langle g, B(f,h)\rangle_{Y^*,Y}$ and applying Tonelli's and Fubini's theorems, we have:

$$\begin{aligned}
\langle g, B(f,h)\rangle_{Y^*,Y} &= \int_0^{2\pi} u_\infty(\varphi) g(\varphi) d\varphi = \\
&= \int_0^{2\pi} \frac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}} \left[\int_{\partial D} \left(u^s(y) \frac{\partial e^{-\mathrm{i}k\hat{x}\cdot y}}{\partial \nu(y)} - e^{-\mathrm{i}k\hat{x}\cdot y} \frac{\partial u^s(y)}{\partial \nu} \right) d\sigma(y) \right] g(\varphi) d\varphi = \\
&= \frac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}} \int_{\partial D} \left\{ u^s(y) \int_0^{2\pi} \frac{\partial e^{-\mathrm{i}k\hat{x}\cdot y}}{\partial \nu(y)} g(\varphi) d\varphi - \left[\int_0^{2\pi} e^{-\mathrm{i}k\hat{x}\cdot y} g(\varphi) d\varphi \right] \frac{\partial u^s(y)}{\partial \nu} \right\} d\sigma(y) = \\
&= \frac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}} \int_{\partial D} \left\{ u^s(y) \frac{\partial \tilde{v}_g(y)}{\partial \nu} - \tilde{v}_g(y) \frac{\partial u^s(y)}{\partial \nu} \right\} d\sigma(y),
\end{aligned} \tag{2.144}$$

where $\hat{x} = (\cos\varphi, \sin\varphi)$. We now observe that $u^s, \tilde{u}^s \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ are weak solutions of the Helmholtz equation in $\mathbb{R}^2 \setminus \bar{D}$ and satisfy the Sommerfeld radiation condition; then we can apply theorem A.18.2 and, by virtue of its thesis (A.160), we have:

$$\int_{\partial D} \left[u^s(y) \frac{\partial \tilde{u}^s(y)}{\partial \nu} - \tilde{u}^s(y) \frac{\partial u^s(y)}{\partial \nu} \right] d\sigma(y) = 0. \tag{2.145}$$

Hence, multiplying this zero term by $\frac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}}$ and adding it to (2.144), as well as using the transmission conditions (2.38)(c)-(d) first for $(\tilde{v}, \tilde{u}^s)$ and then for (v, u^s), we have:

$$\begin{aligned}
\langle g, B(f,h)\rangle_{Y^*,Y} &= \frac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}} \int_{\partial D} \left\{ u^s(y) \left[\frac{\partial \tilde{v}_g(y)}{\partial \nu} + \frac{\partial \tilde{u}^s(y)}{\partial \nu} \right] - \left[\tilde{v}_g(y) + \tilde{u}^s(y) \right] \frac{\partial u^s(y)}{\partial \nu} \right\} d\sigma(y) = \\
&= \frac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}} \int_{\partial D} \left\{ u^s(y) \frac{\partial \tilde{v}(y)}{\partial \nu_{\mathbf{A}'}} - \tilde{v}(y) \frac{\partial u^s(y)}{\partial \nu} \right\} d\sigma(y) =
\end{aligned}$$

$$= \frac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}} \int\limits_{\partial D} \left\{ [v(y) - f(y)] \frac{\partial \tilde{v}(y)}{\partial \nu_{\mathbf{A}'}} - \tilde{v}(y) \left[\frac{\partial v(y)}{\partial \nu_{\mathbf{A}'}} - h(y) \right] \right\} d\sigma(y). \tag{2.146}$$

We now observe that $v, \tilde{v} \in H^1(D)$ are weak solutions of equation (2.38)(a) in D; then we can apply theorem A.18.1 (whose hypotheses are clearly verified, since the mere definition 2.2.1 of B presupposes the hypotheses of problem 2.2.1, i.e. of problem 2.1.1) and, by virtue of its thesis (A.155), we have:

$$\int\limits_{\partial D} \left[v(y) \frac{\partial \tilde{v}(y)}{\partial \nu_{\mathbf{A}'}} - \tilde{v}(y) \frac{\partial v(y)}{\partial \nu_{\mathbf{A}'}} \right] d\sigma(y) = 0. \tag{2.147}$$

Substituting this result into the last member of (2.146), we immediately get:

$$\begin{aligned} \langle g, B(f,h) \rangle_{Y^*,Y} &= \frac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}} \int\limits_{\partial D} \left\{ -\frac{\partial \tilde{v}(y)}{\partial \nu_{\mathbf{A}'}} f(y) + \tilde{v}(y) h(y) \right\} d\sigma(y) = \\ &= \frac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}} \left\langle \left(-\frac{\partial \tilde{v}}{\partial \nu_{\mathbf{A}'}}\Big|_{\partial D}, \tilde{v}|_{\partial D} \right), (f,h) \right\rangle_{X^*,X} \qquad \forall (f,h) \in X,\ \forall g \in Y^*. \end{aligned} \tag{2.148}$$

This means that the dual operator B^T can be characterized as

$$B^T g = \frac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}} \left(-\frac{\partial \tilde{v}}{\partial \nu_{\mathbf{A}'}}\Big|_{\partial D}, \tilde{v}|_{\partial D} \right). \tag{2.149}$$

The next step is now to show that the operator B^T is injective. To this end, let us suppose that $B^T g = 0$, with $g \in L^2[0, 2\pi]$. This implies that

$$\tilde{v} = 0 \quad \text{and} \quad \frac{\partial \tilde{v}}{\partial \nu_{\mathbf{A}'}} = 0 \quad \text{on } \partial D. \tag{2.150}$$

Therefore $\tilde{u}^s$ satisfies the Helmholtz equation in $\mathbb{R}^2 \setminus \bar{D}$, the Sommerfeld radiation condition and the transmission conditions (2.38)(c)(d), which in our case, taking also into account relations (2.150), read:

$$\tilde{u}^s = -\tilde{v}_g \quad \text{and} \quad \frac{\partial \tilde{u}^s}{\partial \nu} = -\frac{\partial \tilde{v}_g}{\partial \nu} \quad \text{on } \partial D. \tag{2.151}$$

As we have already observed just below definition (2.98), a Herglotz wave function (as $-\tilde{v}_g$ is) solves the Helmholtz equation in all $\mathbb{R}^2$ and then, in particular, in a domain D; hence, from transmission conditions (2.151), it follows that $\tilde{u}^s$ can be extended to an entire solution $\tilde{u}^s_{ext}$ of the Helmholtz equation by means of the definition:

$$\tilde{u}^s_{ext} := \begin{cases} -\tilde{v}_g & \text{in } D \\ \tilde{u}^s & \text{in } \mathbb{R}^2 \setminus \bar{D} \end{cases} \tag{2.152}$$

Of course, $\tilde{u}^s_{ext} \in H^1_{\partial D,loc}(\mathbb{R}^2)$ is an entire solution that satisfies the Sommerfeld radiation condition: remembering theorem 2.2.2, this can only happen if $\tilde{u}^s_{ext}$ is identically zero on $\mathbb{R}^2$,

which implies that $\tilde{v}_g$ is zero too on D, i.e., by analyticity, on $\mathbb{R}^2$; by virtue of theorem 2.3.1, it follows that $g = 0$, whence the injectivity of B^T. Now theorem A.19.6 suffices to conclude that the range of B is dense in $L^2[0, 2\pi]$. ■

Using the notations of the previous theorem 2.4.1 and remembering equality (2.149), from the second of relations (A.187) we also have:

$$\mathcal{N}(B) = {}^a[\mathcal{R}(B^T)] = \left\{(f_0, h_0) \in X \mid \Big\langle (\tilde{f}, \tilde{h}), (f_0, h_0) \Big\rangle_{X^*, X} = 0 \quad \forall\, (\tilde{f}, \tilde{h}) \in \mathcal{R}(B^T) \right\} =$$

$$= \left\{ (f_0, h_0) \in X \mid \frac{e^{\mathrm{i}\pi/4}}{\sqrt{8\pi k}} \int\limits_{\partial D} \left[-\frac{\partial \tilde{v}(y)}{\partial \nu_{\mathbf{A}'}} f_0(y) + \tilde{v}(y) h_0(y) \right] d\sigma(y) = 0 \;\; \forall g \in L^2[0, 2\pi] \right\}. \tag{2.153}$$

For the same reasons that led us to relation (2.147), we see that the pairs $\left(v|_{\partial D}, \left.\frac{\partial v}{\partial \nu_{\mathbf{A}'}}\right|_{\partial D} \right)$, where $v \in H^1(D)$ is a solution of equation (2.38)(a) in D, are in $\mathcal{N}(B)$. Hence, B is not injective; however, we shall restrict the operator B in such a way that the restriction is injective and still has dense range.

To this end, let us firstly introduce the two following vector spaces:

$$H := \left\{ v_g \in C^\infty(\bar{D}) \mid \exists g \in L^2[0, 2\pi] \text{ such that } v_g(x) = \int\limits_0^{2\pi} e^{\mathrm{i}kx\cdot d} g(\theta) d\theta \right\}, \tag{2.154}$$

$$S(D) := \left\{ u \in C^2(D) \cap C^1(\bar{D}) \mid \Delta_2 u + k^2 u = 0 \text{ in } D \right\}, \tag{2.155}$$

and let us denote with $\overline{H}$ and $\overline{S(D)}$ their closure in $H^1(D)$. Of course, H is the space of the Herglotz wave functions restricted to D, while $S(D)$ and $\overline{S(D)}$ are respectively the space of the classical solutions and the space of the H^1 weak solutions to the Helmholtz equation in D. As already observed soon below definition (2.98), any Herglotz wave function is real-analytic and solves the Helmholtz equation in all $\mathbb{R}^2$; hence $H \subset S(D)$ and, consequently, $\overline{H} \subset \overline{S(D)}$.

Theorem 2.4.2. *It holds $\overline{H} = \overline{S(D)}$. In other terms, any weak solution to the Helmholtz equation in a bounded domain $D \subset \mathbb{R}^2$ with C^2 boundary ∂D can be approximated in the $H^1(D)$-norm by a Herglotz wave function.*

Proof. Since $\overline{H} \subset \overline{S(D)}$, it suffices to prove that $\overline{S(D)} \subset \overline{H}$.

Then, let $u \in \overline{S(D)}$; it turns out that it is possible to regard any $u \in \overline{S(D)}$ as a weak solution of the so-called *interior mixed boundary value problem* in one of its particular cases

(see system (8.10)-(8.12) in [15] and put in it $\lambda = 1$, $\Gamma_D = \emptyset$); hence, by virtue of theorem 8.4 in [15], there exists a positive constant $C > 0$ such that

$$\|u\|_{H^1(D)} \leq C \left\| \frac{\partial u}{\partial \nu} + \mathrm{i}u \right\|_{H^{-\frac{1}{2}}(\partial D)}. \tag{2.156}$$

We now remember that the operator $\mathcal{H}_1 : L^2[0, 2\pi] \to H^{-\frac{1}{2}}(\partial D)$ defined by relation (2.99) has dense range, as stated by theorem 2.3.2. Then, for any $\varepsilon > 0$, there exists a Herglotz wave function v_g with kernel $g \in L^2[0, 2\pi]$ such that

$$\left\| \left(\frac{\partial u}{\partial \nu} + \mathrm{i}u \right) - \left(\frac{\partial v_g}{\partial \nu} + \mathrm{i}v_g \right) \right\|_{H^{-\frac{1}{2}}(\partial D)} < \varepsilon, \tag{2.157}$$

i.e.

$$\left\| \frac{\partial (u - v_g)}{\partial \nu} + \mathrm{i}(u - v_g) \right\|_{H^{-\frac{1}{2}}(\partial D)} < \varepsilon. \tag{2.158}$$

By virtue of relation (2.156) written for $(u - v_g) \in \overline{S(D)}$, the previous inequality (2.158) obviously implies

$$\|u - v_g\|_{H^1(D)} < C\varepsilon. \tag{2.159}$$

This means that $u \in \overline{H}$. ■

We still need to define another space:

$$H(\partial D) := \left\{ \left(u|_{\partial D}, \left. \frac{\partial u}{\partial \nu} \right|_{\partial D} \right) \mid u \in \overline{H} \right\}. \tag{2.160}$$

Lemma 2.4.3. *$H(\partial D)$ is a closed subspace of $H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$.*

Proof. By virtue of trace theorems A.15.2 and A.17.1, it is obvious that $H(\partial D)$ is a subspace of $H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$, so we have to prove only its closedness.

To this end, let $(f, h) \in \overline{H(\partial D)}$; then, there exists a sequence $\left\{ \left(u_n|_{\partial D}, \left. \frac{\partial u_n}{\partial \nu} \right|_{\partial D} \right) \right\}_{n=0}^{\infty} \subset H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$ such that $u_n \in \overline{H}$ $\forall n \in \mathbb{N}$ and

$$\lim_{n \to \infty} \left\| \left(u_n|_{\partial D}, \left. \frac{\partial u_n}{\partial \nu} \right|_{\partial D} \right) - (f, h) \right\|_{H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)} = 0. \tag{2.161}$$

Since the sequence $\left\{ \left(u_n|_{\partial D}, \left. \frac{\partial u_n}{\partial \nu} \right|_{\partial D} \right) \right\}_{n=0}^{\infty}$ converges, it is bounded in $H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$; hence, analogously to the proof of theorem 2.4.2, by regarding $u_n \in \overline{H} = \overline{S(D)}$ $\forall n \in \mathbb{N}$ as a weak solution of the so-called *interior mixed boundary value problem* in one of its particular cases (see system (8.10)-(8.12) in [15] and put in it $\lambda = 1$, $\Gamma_D = \emptyset$), it turns out that u_n satisfies inequality (2.156) and then the sequence $\{u_n\}_{n=0}^{\infty}$ is bounded in $H^1(D)$. Hence, there exists a

subsequence $\left\{u_{n(k)}\right\}_{k=0}^{\infty} \subset \overline{H}$ that converges weakly in $H^1(D)$ to a function[22] $u \in \overline{H}$. From the (weak)[23] continuity of the trace operators (see theorems A.15.2 and A.17.1), we deduce that $\left\{\left(u_{n(k)}|_{\partial D}, \left.\frac{\partial u_{n(k)}}{\partial \nu}\right|_{\partial D}\right)\right\}_{k=0}^{\infty}$ converges weakly in $H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$ to $\left(u|_{\partial D}, \left.\frac{\partial u}{\partial \nu}\right|_{\partial D}\right) \in H(\partial D)$. By the uniqueness of the limit, we have that

$$(f,h) = \left(u|_{\partial D}, \left.\frac{\partial u}{\partial \nu}\right|_{\partial D}\right) \in H(\partial D). \tag{2.162}$$

Summing up, we have found that $(f,h) \in \overline{H(\partial D)} \Rightarrow (f,h) \in H(\partial D)$, which concludes the proof. ■

By virtue of the previous lemma 2.4.3, $H(\partial D)$ equipped with the scalar product induced by $H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$ is a Hilbert space.

Now, let us denote with B_0 the restriction of the operator B to $H(\partial D)$.

Theorem 2.4.4. *Let us assume that k^2 is not a transmission eigenvalue; then the bounded linear operator $B_0 : H(\partial D) \to L^2[0, 2\pi]$ is injective. Furthermore, if the hypotheses of either theorem 2.3.9 or theorem 2.3.11 are also satisfied, B_0 has dense range.*

Proof. Let $(f,h) \in H(\partial D)$ be such that $B_0(f,h) = 0$ and let $(v, u^s) \in H^1(D) \oplus H^1_{\partial D, loc}(\mathbb{R}^2 \setminus \bar{D})$ the solution to system (2.38) corresponding to these boundary data. Then u^s has zero far-field pattern, whence $u^s = 0$ in $\mathbb{R}^2 \setminus \bar{D}$ by virtue of theorem 2.2.5. Hence, we see from system (2.38) that v satisfies the relations:

$$\begin{cases} \nabla_2 \cdot \mathbf{A}' \nabla_2 v + k^2 n v = 0 & \text{in } D & \text{(a)} \\ v = f & \text{on } \partial D & \text{(b)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} = h & \text{on } \partial D. & \text{(c)} \end{cases} \tag{2.163}$$

By definition of $H(\partial D)$, the functions f and h are the traces on ∂D of a $H^1(D)$ solution w to the Helmholtz equation and of its normal derivative respectively, i.e. there exists $w \in H^1(D)$ such that

$$\begin{cases} \Delta_2 w + k^2 w = 0 & \text{in } D & \text{(a)} \\ f = w|_{\partial D} & \text{on } \partial D & \text{(b)} \\ h = \left.\dfrac{\partial w}{\partial \nu}\right|_{\partial D} & \text{on } \partial D. & \text{(c)} \end{cases} \tag{2.164}$$

[22] In general, while any set which is closed in the strong topology is closed also in the weak topology, the converse is false. However, for convex sets the strong closedness coincides with the weak one (see [13], p. 57); hence, by virtue of the convexity of $\overline{H}$ (due to the fact that it is even a vector space), we have $u \in \overline{H}$, and not merely $u \in H^1(D)$.

[23] See, for example, [13], p.58.

From systems (2.163) and (2.164), it follows that $(v, w) \in H^1(D) \oplus H^1(D)$ solves the homogeneous interior transmission problem, i.e.

$$\begin{cases} \nabla_2 \cdot \mathbf{A}' \nabla_2 v + k^2 n v = 0 & \text{in } D \quad \text{(a)} \\ \Delta_2 w + k^2 w = 0 & \text{in } D \quad \text{(b)} \\ v - w = 0 & \text{on } \partial D \quad \text{(c)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} - \dfrac{\partial w}{\partial \nu} = 0 & \text{on } \partial D. \quad \text{(d)} \end{cases} \tag{2.165}$$

Since, by hypothesis, k^2 is not a transmission eigenvalue, system (2.165) has only the trivial solution $(v, w) \equiv (0, 0)$ in D and then, by virtue of relations (2.164)(b)-(c), we get $f = h = 0$, i.e. B_0 is injective.

We still have to prove that the set $\mathcal{R}(B_0)$ is dense in $L^2[0, 2\pi]$. To this end, it clearly suffices to demonstrate that $\mathcal{R}(B) \subset \mathcal{R}(B_0)$, since we have already proved in theorem 2.4.1 that $\mathcal{R}(B)$ is dense in $L^2[0, 2\pi]$. We can observe that the other inclusion $\mathcal{R}(B_0) \subset \mathcal{R}(B)$ trivially holds: hence we shall actually prove that $\mathcal{R}(B_0) = \mathcal{R}(B)$.

Then, let $u_\infty \in \mathcal{R}(B)$: this means that u_∞ is the far-field pattern of the radiating field u^s which, together with the interior field v, form the unique solution $(v, u^s) \in H^1(D) \oplus H^1_{\partial D, loc}(\mathbb{R}^2 \setminus \bar{D})$ of system (2.38) with certain boundary data (f, h). On the other hand, by virtue of either theorem 2.3.9 or theorem 2.3.11, the same interior field v, together with a suitable function $w \in H^1(D)$, can be regarded as forming the unique solution $(v, w) \in H^1(D) \oplus H^1(D)$ of the interior transmission problem 2.3.2 with boundary data $\left(u^s|_{\partial D}, \left. \dfrac{\partial u^s}{\partial \nu} \right|_{\partial D} \right) \in H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$, i.e.:

$$\begin{cases} \nabla_2 \cdot \mathbf{A}' \nabla_2 v + k^2 n v = 0 & \text{in } D \quad \text{(a)} \\ \Delta_2 w + k^2 w = 0 & \text{in } D \quad \text{(b)} \\ v - w = u^s & \text{on } \partial D \quad \text{(c)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} - \dfrac{\partial w}{\partial \nu} = \dfrac{\partial u^s}{\partial \nu} & \text{on } \partial D. \quad \text{(d)} \end{cases} \tag{2.166}$$

Hence, by means of a comparison between systems (2.38) and (2.166), we deduce that $(v, u^s) \in H^1(D) \oplus H^1_{\partial D, loc}(\mathbb{R}^2 \setminus \bar{D})$ is the unique solution of system (2.38) itself with boundary data

$\left(w|_{\partial D}, \left.\frac{\partial w}{\partial \nu}\right|_{\partial D}\right) \in H(\partial D)$, i.e.:

$$\begin{cases} \nabla_2 \cdot \mathbf{A}' \nabla_2 v + k^2 n v = 0 & \text{in } D & \text{(a)} \\ \Delta_2 u^s + k^2 u^s = 0 & \text{in } \mathbb{R}^2 \setminus \bar{D} & \text{(b)} \\ v - u^s = w & \text{on } \partial D & \text{(c)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} - \dfrac{\partial u^s}{\partial \nu} = \dfrac{\partial w}{\partial \nu} & \text{on } \partial D & \text{(d)} \\ \lim\limits_{r\to\infty} \sqrt{r}\left(\dfrac{\partial u^s}{\partial r} - \mathrm{i}k u^s\right) = 0. & & \text{(e)} \end{cases} \tag{2.167}$$

This clearly means that

$$B_0\left(w|_{\partial D}, \left.\frac{\partial w}{\partial \nu}\right|_{\partial D}\right) = u_\infty, \tag{2.168}$$

i.e. $u_\infty \in \mathcal{R}(B_0)$. Hence it holds $\mathcal{R}(B) \subset \mathcal{R}(B_0)$, and this concludes the proof. ■

Theorem 2.4.5. *The operator $B_0 : H(\partial D) \to L^2[0, 2\pi]$ is compact.*

Proof. Given $w \in \overline{H}$, let $(v, u^s) \in H^1(D) \oplus H^1_{\partial D, loc}(\mathbb{R}^2 \setminus \bar{D})$ be the unique solution of system (2.38) with boundary data $f := w|_{\partial D}$ and $h := \left.\frac{\partial w}{\partial \nu}\right|_{\partial D}$. Let Ω_R be an open disk of radius R, centred at the origin and containing $\bar{D}$, and let $\partial\Omega_R$ be its boundary. Then, we can define the linear operator

$$\begin{array}{rccc} E: & H^1(D) \oplus H^1(\Omega_R \setminus \bar{D}) & \longrightarrow & H^{\frac{1}{2}}(\partial\Omega_R) \oplus H^{-\frac{1}{2}}(\partial\Omega_R) \\ & (v, u^s) & \longmapsto & \left(u^s|_{\partial\Omega_R}, \left.\dfrac{\partial u^s}{\partial \nu}\right|_{\partial\Omega_R}\right). \end{array} \tag{2.169}$$

Since $\Omega_R \supset \bar{D}$, by putting $\Omega := \Omega_R \setminus \bar{D}$ we have $\partial\Omega = \partial\Omega_R \cup \partial D \supset \partial D$; this implies that

$$\|u^s\|_{H^{\frac{1}{2}}(\partial\Omega_R)} \leq \|u^s\|_{H^{\frac{1}{2}}(\partial\Omega)}, \qquad \left\|\frac{\partial u^s}{\partial \nu}\right\|_{H^{-\frac{1}{2}}(\partial\Omega_R)} \leq \left\|\frac{\partial u^s}{\partial \nu}\right\|_{H^{-\frac{1}{2}}(\partial\Omega)}. \tag{2.170}$$

By virtue of the trace theorems A.15.2, A.17.1 (i.e., more precisely, by virtue of inequalities (A.131) and (A.150) applied in the bounded and C^2 domain Ω) and of inequalities (2.170), we have that

$$\|u^s\|_{H^{\frac{1}{2}}(\partial\Omega_R)} + \left\|\frac{\partial u^s}{\partial \nu}\right\|_{H^{-\frac{1}{2}}(\partial\Omega_R)} \leq C_1 \|u^s\|_{H^1(\Omega_R \setminus \bar{D})} + C_3 \|u^s\|_{H^1(\Omega_R \setminus \bar{D})}, \tag{2.171}$$

which obviously implies

$$\|u^s\|_{H^{\frac{1}{2}}(\partial\Omega_R)} + \left\|\frac{\partial u^s}{\partial \nu}\right\|_{H^{-\frac{1}{2}}(\partial\Omega_R)} \leq \|v\|_{H^1(D)} + (C_1 + C_3)\|u^s\|_{H^1(\Omega_R \setminus \bar{D})}. \tag{2.172}$$

The last inequality clearly expresses the continuity[24] of the operator E. On the other hand, also the operator A defined by (2.54) in remark 2.1.4 is continuous and its continuity is expressed by relation (2.52); the restriction $A|_{H(\partial D)} : H(\partial D) \to H^1(D) \oplus H^1(\Omega_R \setminus \bar{D})$ is obviously continuous too and its continuity is expressed by relation (2.52) itself written for our case, i.e.:

$$\|v\|_{H^1(D)} + \|u^s\|_{H^1(\Omega_R \setminus \bar{D})} \leq C \left(\|w\|_{H^{\frac{1}{2}}(\partial D)} + \left\| \frac{\partial w}{\partial \nu} \right\|_{H^{-\frac{1}{2}}(\partial D)} \right). \tag{2.173}$$

Hence, the operator G, defined as

$$\begin{aligned} G := E \circ A|_{H(\partial D)} : \qquad H(\partial D) \qquad &\longrightarrow \quad H^{\frac{1}{2}}(\partial \Omega_R) \oplus H^{-\frac{1}{2}}(\partial \Omega_R) \\ \left(w|_{\partial D}, \left. \frac{\partial w}{\partial \nu} \right|_{\partial D} \right) &\longmapsto \quad \left(u^s|_{\partial \Omega_R}, \left. \frac{\partial u^s}{\partial \nu} \right|_{\partial \Omega_R} \right), \end{aligned} \tag{2.174}$$

is continuous, since it is the composition of two continuous operators; the relation expressing its continuity follows from (2.172) and (2.173), and it can be written as

$$\|u^s\|_{H^{\frac{1}{2}}(\partial \Omega_R)} + \left\| \frac{\partial u^s}{\partial \nu} \right\|_{H^{-\frac{1}{2}}(\partial \Omega_R)} \leq C_4 \left(\|w\|_{H^{\frac{1}{2}}(\partial D)} + \left\| \frac{\partial w}{\partial \nu} \right\|_{H^{-\frac{1}{2}}(\partial D)} \right). \tag{2.175}$$

If we now remember the compact operator K defined by (2.91) in theorem 2.2.6, we immediately realize that $B_0 = K \circ G$; hence B_0 is compact, since it consists of the composition of a continuous operator with a compact one. ■

Definition 2.4.1. *Let v_g be a Herglotz wave function with kernel $g \in L^2[0, 2\pi]$; then we define the linear operator $\mathcal{H}$ as:*

$$\begin{aligned} \mathcal{H} : \quad L^2[0, 2\pi] \quad &\longrightarrow \quad H(\partial D) \\ g \qquad &\longmapsto \quad \left(v_g|_{\partial D}, \left. \frac{\partial v_g}{\partial \nu} \right|_{\partial D} \right). \end{aligned} \tag{2.176}$$

Theorem 2.4.6. *The linear operator $\mathcal{H} : L^2[0, 2\pi] \to H(\partial D)$ is injective and continuous.*

Proof. The injectivity of $\mathcal{H}$ can be easily proved observing that if $v_g|_{\partial D} = 0$, then $v_g = 0$ in all $\mathbb{R}^2$ by analyticity: by virtue of theorem 2.3.1, this implies $g = 0$.

As regards the continuity of $\mathcal{H}$, we firstly observe that, by definition of norm in $H(\partial D)$ and by virtue of relations (A.131), (A.150), we have:

$$\begin{aligned} \left\| \left(v_g|_{\partial D}, \left. \frac{\partial v_g}{\partial \nu} \right|_{\partial D} \right) \right\|^2_{H(\partial D)} &= \|v_g|_{\partial D}\|^2_{H^{\frac{1}{2}}(\partial D)} + \left\| \left. \frac{\partial v_g}{\partial \nu} \right|_{\partial D} \right\|^2_{H^{-\frac{1}{2}}(\partial D)} \leq \\ &\leq C_1^2 \, \|v_g\|^2_{H^1(D)} + C_3^2 \, \|v_g\|^2_{H^1(D)} = C_4^2 \, \|v_g\|^2_{H^1(D)}, \end{aligned} \tag{2.177}$$

[24] Here and in the remainder of the current proof one should remember the question of the equivalence of the norms, as already discussed in remark 2.1.4.

having obviously put $C_4 := \sqrt{C_1^2 + C_3^2}$. Moreover, remembering the definitions (2.98) of Herglotz wave function and (A.50) of scalar product in $L^2[0, 2\pi]$ and applying the Cauchy-Schwarz inequality, we have:

$$|v_g(x)| = \left| \int_0^{2\pi} e^{\mathrm{i}kx\cdot d}\, g(\theta) d\theta \right| = \left| \left(g, e^{-\mathrm{i}kx\cdot d}\right)_{L^2[0,2\pi]} \right| \leq \|g\|_{L^2[0,2\pi]}\ \|e^{-\mathrm{i}kx\cdot d}\|_{L^2[0,2\pi]} = \sqrt{2\pi}\, \|g\|_{L^2[0,2\pi]}. \tag{2.178}$$

We now remember (see theorem A.13.2) that $H^1(D) = W^{1,2}(D)$ with equivalent norms; hence, recalling the definition of norm in $W^{1,2}(D)$ (see (A.66)) and using the previous inequality (2.178), we can write[25]:

$$\begin{aligned}
&\|v_g\|_{H^1(D)} \leq M_2 \left\{ \|v_g\|^2_{L^2(D)} + \|\partial_1 v_g\|^2_{L^2(D)} + \|\partial_2 v_g\|^2_{L^2(D)} \right\}^{\frac{1}{2}} = \\
&= M_2 \left\{ \left[1 + k^2(|d_1|^2 + |d_2|^2)\right]\ \|v_g\|^2_{L^2(D)} \right\}^{\frac{1}{2}} = M_2 \left\{ \left[1 + k^2(|d_1|^2 + |d_2|^2)\right] \int_D |v_g(x)|^2 dx \right\}^{\frac{1}{2}} \leq \\
&\leq M_2 \left[1 + k^2(|d_1|^2 + |d_2|^2)\right]^{\frac{1}{2}} \sqrt{2\pi}\, \|g\|_{L^2[0,2\pi]} \sqrt{\mathrm{mis}\, D} = A\, \|g\|_{L^2[0,2\pi]},
\end{aligned} \tag{2.179}$$

where M_2 is a real positive constant (cf. (A.3)), $(d_1, d_2) \in \mathbb{R}^2$ are the components of the unit vector d and $A := M_2 \left[1 + k^2(|d_1|^2 + |d_2|^2)\right]^{\frac{1}{2}} \sqrt{2\pi}\, \sqrt{\mathrm{mis}\, D}$.

Now, by comparing relations (2.177) and (2.179), we easily get:

$$\left\| \left(v_g|_{\partial D},\ \left.\frac{\partial v_g}{\partial \nu}\right|_{\partial D} \right) \right\|_{H(\partial D)} \leq C_4\, A\, \|g\|_{L^2[0,2\pi]}; \tag{2.180}$$

the previous inequality (2.180) expresses the continuity of $\mathcal{H}$. ■

Theorem 2.4.7. *If the far-field pattern $u_\infty \in L^2[0, 2\pi]$ is in the range of B_0, then, for every $\varepsilon > 0$, there exists a $g^\varepsilon \in L^2[0, 2\pi]$ such that $\mathcal{H}g^\varepsilon \in H(\partial D)$ satisfies the inequality*

$$\|B_0(\mathcal{H}g^\varepsilon) - u_\infty\|_{L^2[0,2\pi]} \leq \varepsilon. \tag{2.181}$$

Proof. By hypothesis, $u_\infty \in \mathcal{R}(B_0)$: this means that there exists a function $w \in \overline{H}$ such that

$$B_0 \left(w|_{\partial D},\ \left.\frac{\partial w}{\partial \nu}\right|_{\partial D} \right) = u_\infty. \tag{2.182}$$

On the other hand, by definition of $\mathcal{H}$, for any $g \in L^2[0, 2\pi]$ it holds:

$$B_0\left(\mathcal{H}g\right) = B_0 \left(v_g|_{\partial D},\ \left.\frac{\partial v_g}{\partial \nu}\right|_{\partial D} \right). \tag{2.183}$$

[25] According to the notations of section A.2, we denote with ∂_1 [resp. ∂_2] the partial derivative operator $\frac{\partial}{\partial x_1}$ [resp. $\frac{\partial}{\partial x_2}$].

Then, by virtue of the linearity of B_0 and of the trace operators γ_1, γ_2 defined in theorems A.15.2, A.17.1 respectively, we have

$$B_0\left(\mathcal{H}g\right) - u_\infty = B_0\left(\left(v_g - w\right)|_{\partial D},\, \left(\frac{\partial v_g}{\partial \nu} - \frac{\partial w}{\partial \nu}\right)\Big|_{\partial D}\right); \tag{2.184}$$

the continuity of B_0 and relation (2.184) now imply:

$$\|B_0(\mathcal{H}g) - u_\infty\|_{L^2[0,2\pi]} \leq C\left(\|v_g - w\|_{H^{\frac{1}{2}}(\partial D)} + \left\|\frac{\partial v_g}{\partial \nu} - \frac{\partial w}{\partial \nu}\right\|_{H^{-\frac{1}{2}}(\partial D)}\right). \tag{2.185}$$

Moreover, by definition of $\overline{H}$, for every $\varepsilon' > 0$ there exists a Herglotz wave function $v_{g^{\varepsilon'}}$ such that

$$\|v_{g^{\varepsilon'}} - w\|_{H^1(D)} \leq \varepsilon'. \tag{2.186}$$

By virtue of trace theorems A.15.2 and A.17.1 (i.e., more precisely, by virtue of inequalities (A.131) and (A.150)), from relation (2.186) it immediately follows

$$\|v_{g^{\varepsilon'}} - w\|_{H^{\frac{1}{2}}(\partial D)} \leq C_1\,\varepsilon', \tag{2.187}$$

$$\left\|\frac{\partial v_{g^{\varepsilon'}}}{\partial \nu} - \frac{\partial w}{\partial \nu}\right\|_{H^{-\frac{1}{2}}(\partial D)} \leq C_3\,\varepsilon'. \tag{2.188}$$

Substituting inequalities (2.187) and (2.188) into (2.185), we find:

$$\left\|B_0(\mathcal{H}g^{\varepsilon'}) - u_\infty\right\|_{L^2[0,2\pi]} \leq C\left(C_1\,\varepsilon' + C_3\,\varepsilon'\right). \tag{2.189}$$

If we put $C' := \max\{C\,C_1, C\,C_3\}$, from inequality (2.189) we get:

$$\left\|B_0(\mathcal{H}g^{\varepsilon'}) - u_\infty\right\|_{L^2[0,2\pi]} \leq 2C'\varepsilon'. \tag{2.190}$$

The last relation (2.190) becomes exactly the thesis (2.181) by choosing $\varepsilon'(\varepsilon) := \dfrac{\varepsilon}{2C'}$ and simply writing g^ε instead of $g^{\varepsilon'(\varepsilon)}$. ■

Now we can turn our attention to the main goal of this section, i.e. that of finding an approximation to the scattering inhomogeneity D; to this end, for each $z \in \mathbb{R}^2$ we consider the so-called *far-field equation* in the unknown $g \in L^2[0, 2\pi]$:

$$(Fg)(\varphi) = \Phi_\infty(\varphi, z), \qquad z \in \mathbb{R}^2, \tag{2.191}$$

where F is the far-field operator corresponding to system (2.94) (cf. definition (2.97)), while $\Phi_\infty(\varphi, z)$ is the far-field pattern (given by relation (2.74)) of the fundamental solution $\Phi(x, z)$

to the Helmholtz equation; substituting the expressions (2.97) and (2.75) into the left-hand and the right-hand side of equation (2.191) respectively, we obtain its explicit form:

$$\int_0^{2\pi} u_\infty(\varphi,\theta)g(\theta)d\theta = \gamma e^{-\mathrm{i}k\hat{x}\cdot z}, \qquad z \in \mathbb{R}^2, \tag{2.192}$$

where $\hat{x} = (\cos\varphi, \sin\varphi)$. The explicit form (2.192) of the far-field equation shows that, for each $z \in \mathbb{R}^2$, it is a Fredholm integral equation of the first kind, in which the data function is a known real-analytic function and the integral kernel is the far-field pattern of the radiating field u^s determined by system (2.94). If we now remember theorem 2.3.3, relation (2.102), definitions 2.2.1, 2.4.1 and that $B_0 := B|_{H(\partial D)}$, we deduce that the far-field operator F can be factored as

$$F = B_0 \circ \mathcal{H}, \tag{2.193}$$

so that the far-field equation (2.191) can be written in the form:

$$(B_0(\mathcal{H}g))(\varphi) = \Phi_\infty(\varphi, z), \qquad z \in \mathbb{R}^2. \tag{2.194}$$

In other terms, both $(Fg)(\varphi)$ and $(B_0(\mathcal{H}g))(\varphi)$ are the far-field pattern of the scattered field $u^s = v_g^s$ (cf. (2.100), (2.101) and (2.102)) determined by system (2.38) written for boundary data $(f,h) := \mathcal{H}g$. Hence, by virtue of the one-to-one correspondence between radiating fields and their far-field patterns (see theorem 2.2.5 and the comment soon below), the far-field equation implies, for $z \in D$, that this v_g^s coincides with $\Phi(\cdot,z)$ in $\mathbb{R}^2 \setminus \bar{D}$. It follows that, for $z \in D$, $g \in L^2[0,2\pi]$ solves the far-field equation (2.194) if and only if the pair $(v, \Phi(\cdot,z)) \in H^1(D) \oplus H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ solves system (2.38), which, for the current case, reads:

$$\begin{cases} \nabla_2 \cdot \mathbf{A}'\nabla_2 v + k^2 n v = 0 & \text{in } D & \text{(a)} \\ \Delta_2 \Phi(\cdot,z) + k^2 \Phi(\cdot,z) = 0 & \text{in } \mathbb{R}^2 \setminus \bar{D} & \text{(b)} \\ v - \Phi(\cdot,z) = v_g & \text{on } \partial D & \text{(c)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} - \dfrac{\partial \Phi(\cdot,z)}{\partial \nu} = \dfrac{\partial v_g}{\partial \nu} & \text{on } \partial D & \text{(d)} \\ \lim\limits_{r\to\infty} \sqrt{r}\left(\dfrac{\partial \Phi(\cdot,z)}{\partial r} - \mathrm{i}k\Phi(\cdot,z)\right) = 0. & & \text{(e)} \end{cases} \tag{2.195}$$

Actually, since $\Phi(\cdot,z)$ is a fundamental solution to the Helmholtz equation and satisfies the Sommerfeld radiation condition, conditions (2.195)(b) and (2.195)(e) are identically satisfied. Moreover, according to the same approach followed in the last part of the proof of theorem 2.3.4, we can remember that any Herglotz wave function v_g solves the Helmholtz equation in

all $\mathbb{R}^2$. Hence, system (2.195) is equivalent to the following interior transmission problem:

$$\begin{cases} \nabla_2 \cdot \mathbf{A}'\nabla_2 v + k^2 n v = 0 & \text{in } D \quad \text{(a)} \\ \Delta_2 v_g + k^2 v_g = 0 & \text{in } D \quad \text{(b)} \\ v - v_g = \Phi(\cdot, z) & \text{on } \partial D \quad \text{(c)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} - \dfrac{\partial v_g}{\partial \nu} = \dfrac{\partial \Phi(\cdot, z)}{\partial \nu} & \text{on } \partial D. \quad \text{(d)} \end{cases} \tag{2.196}$$

Summing up, if $z \in D$, $g \in L^2[0, 2\pi]$ solves the far-field equation (2.194) if and only if the pair $(v, v_g) \in H^1(D) \oplus H^1(D)$ solves system (2.196); if $z \in \mathbb{R}^2 \setminus \bar{D}$, this is no more true.

Now we are nearly ready to state and prove the *general theorem*, from which the linear sampling method gets its inspiration. However, we still need two preliminary results, which we give in the two following lemmas.

Lemma 2.4.8. *Let z be an arbitrary point in $\mathbb{R}^2$ and U_z an open neighbourhood of z; then $\Phi(\cdot, z) \notin H^1(U_z)$.*

Proof. First of all, it is not restrictive to assume that $z = (0, 0)$ and $U_z = \Omega_\varepsilon := \{x \in \mathbb{R}^2 \mid \|x\|_{\mathbb{R}^2} < \varepsilon\}$ for some $\varepsilon > 0$ small enough. Secondly, if we remember definition (2.55) and use the relation[26]

$$\frac{d}{dr} H_0^{(1)}(r) = H_1^{(1)}(r) \qquad \forall r > 0, \tag{2.197}$$

taking into account the asymptotic behaviour[27] of $H_1^{(1)}(r)$ for $r \to 0^+$ we easily get:

$$\frac{\partial \Phi(\cdot, (0,0))}{\partial r}(x) = \frac{1}{2\pi} \frac{1}{\|x\|_{\mathbb{R}^2}} + O(\|x\|_{\mathbb{R}^2} \log \|x\|_{\mathbb{R}^2}) \qquad \text{as } \|x\|_{\mathbb{R}^2} \to 0, \tag{2.198}$$

where we have denoted with $\dfrac{\partial \Phi(\cdot, (0,0))}{\partial r}(x)$ the value in $x \neq (0, 0)$ of the classical radial derivative of $\Phi(\cdot, (0, 0))$.

Now, let us suppose, by absurd, that $\Phi(\cdot, (0, 0)) \in H^1(\Omega_\varepsilon)$: this implies, in particular, that the weak derivatives $\partial_1 \Phi(\cdot, (0, 0))$, $\partial_2 \Phi(\cdot, (0, 0))$ exist and belong to $L^2(\Omega_\varepsilon)$; consequently, for each $\theta \in [0, 2\pi]$ also their linear combination

$$f_\theta(\cdot) := \cos\theta \, \partial_1 \Phi(\cdot, (0,0)) + \sin\theta \, \partial_2 \Phi(\cdot, (0,0)) \tag{2.199}$$

belongs to $L^2(\Omega_\varepsilon)$. On the other hand, if we take $\varepsilon' < \varepsilon$ and consider the open circular corona $\Omega(\varepsilon', \varepsilon) := \Omega_\varepsilon \setminus \Omega_{\varepsilon'}$, we obviously have that $\Phi(\cdot, (0, 0)) \in C^1(\Omega(\varepsilon', \varepsilon)) \cap L^2(\Omega(\varepsilon', \varepsilon))$ with classical partial derivatives $\dfrac{\partial \Phi(\cdot, (0,0))}{\partial x_1}$, $\dfrac{\partial \Phi(\cdot, (0,0))}{\partial x_2} \in L^2(\Omega(\varepsilon', \varepsilon))$: then, by virtue of remark A.8.1, the weak partial derivatives coincide with the classical ones in $\Omega(\varepsilon', \varepsilon)$, i.e.

$$\partial_1 \Phi(\cdot, (0,0)) = \frac{\partial \Phi(\cdot, (0,0))}{\partial x_1}, \qquad \partial_2 \Phi(\cdot, (0,0)) = \frac{\partial \Phi(\cdot, (0,0))}{\partial x_2} \qquad \text{in } \Omega(\varepsilon', \varepsilon). \tag{2.200}$$

[26] See sections 3.2 and 3.3 in [15].

[27] See sections 3.2 and 3.3 in [15].

Hence, by comparing definition (2.199) with equalities (2.200), we also have

$$f_\theta(\cdot) = \cos\theta\,\frac{\partial\Phi(\cdot,(0,0))}{\partial x_1} + \sin\theta\,\frac{\partial\Phi(\cdot,(0,0))}{\partial x_2} = \frac{\partial\Phi(\cdot,(0,0))}{\partial r} \qquad \text{in } \Omega(\varepsilon',\varepsilon). \tag{2.201}$$

Since the previous equality holds for all $\varepsilon' \in (0,\varepsilon)$, it actually holds in $\Omega_\varepsilon \setminus \{(0,0)\}$ (i.e. almost everywhere in Ω_ε). Now, the behaviour of $\dfrac{\partial\Phi(\cdot,(0,0))}{\partial r}$ in $\Omega_\varepsilon \setminus \{(0,0)\}$ is given by relation (2.198), whose right-hand side involves the term $\dfrac{1}{\|x\|_{\mathbb{R}^2}}$, i.e. a singularity in $(0,0)$. On the other hand, a well-known general result in classical analysis is that

$$\frac{1}{\|\cdot\|_{\mathbb{R}^n}} \in L^p(\Omega_r) \iff 1 \le p < n, \tag{2.202}$$

being $\Omega_r := \{x \in \mathbb{R}^n \mid \|x\|_{\mathbb{R}^n} < r\}$ for some $r > 0$. Since in our case $p = n = 2$, we have that $\dfrac{\partial\Phi(\cdot,(0,0))}{\partial r} \notin L^2(\Omega_\varepsilon)$, i.e., by virtue of equality (2.201), $f_\theta(\cdot) \notin L^2(\Omega_\varepsilon)$. This contradicts what previously stated about f_θ and then the proof is complete. ■

Lemma 2.4.9. *Let, as usual, D be a nonempty, bounded and open subset of $\mathbb{R}^2$ with C^2 boundary; let z^* be an arbitrary point of ∂D and $\{z_j\}_{j=0}^\infty \subset D$ a sequence of points of D such that $\lim_{j\to\infty} \|z_j - z^*\|_{\mathbb{R}^2} = 0$. Moreover, let $\Omega_R := \{x \in \mathbb{R}^2 \mid \|x\|_{\mathbb{R}^2} < R\}$ be large enough to contain $\bar{D}$, i.e. $\Omega_R \supset \bar{D}$. Then it holds:*

$$\lim_{j\to\infty} \|\Phi(\cdot, z_j)\|_{H^1(\Omega_R\setminus\bar{D})} = \infty. \tag{2.203}$$

Proof. We firstly observe that if U_{z^*} is an open neighbourhood of z^*, then, by virtue of the previous lemma 2.4.8, $\Phi(\cdot, z^*) \notin H^1(U_{z^*})$: this clearly implies that $\Phi(\cdot, z^*) \notin H^1(\Omega_R \setminus \bar{D})$, i.e.

$$\|\Phi(\cdot, z^*)\|_{H^1(\Omega_R\setminus\bar{D})} = \infty. \tag{2.204}$$

Then, let us suppose, by absurd, that limit (2.203) does not hold: this means that there exist a constant $M > 0$ and a subsequence $\{z_{j_k}\}_{k=0}^\infty \subset D$ such that

$$\|\Phi(\cdot, z_{j_k})\|_{H^1(\Omega_R\setminus\bar{D})} \le M \qquad \forall k \in \mathbb{N}. \tag{2.205}$$

Remembering that $H^1(\Omega_R \setminus \bar{D}) = W^{1,2}(\Omega_R \setminus \bar{D})$ with equivalent norms (see theorem A.13.2, statement No 2) and recalling definition (A.66) (with $n = 2$, $r = 1$, $p = 2$), relation (2.205) implies that there exists a constant $M' > 0$ such that

$$\int_{\Omega_R\setminus\bar{D}} \sum_{|\alpha|_{\mathbb{N}^2}\le 1} |\partial^\alpha\Phi(x, z_{j_k})|^2\,dx \le (M')^2 \qquad \forall k \in \mathbb{N}. \tag{2.206}$$

On the other hand, it clearly holds:

$$\lim_{k\to\infty} \partial^\alpha \Phi(x, z_{j_k}) = \partial^\alpha \Phi(x, z^*) \qquad \forall x \in \mathbb{R}^2 \setminus \{z^*\}, \quad \forall \alpha \in \mathbb{N}^2, \tag{2.207}$$

and then, in particular:

$$\lim_{k\to\infty} \sum_{|\alpha|_{\mathbb{N}^2}\leq 1} |\partial^\alpha \Phi(x, z_{j_k})|^2 = \sum_{|\alpha|_{\mathbb{N}^2}\leq 1} |\partial^\alpha \Phi(x, z^*)|^2 \qquad \forall x \in \mathbb{R}^2 \setminus \{z^*\}. \tag{2.208}$$

Now, relations (2.206) and (2.208) together imply, by Fatou's lemma[28], that

$$\exists \int_{\Omega_R \setminus \bar{D}} \sum_{|\alpha|_{\mathbb{N}^2}\leq 1} |\partial^\alpha \Phi(x, z^*)|^2 \, dx \leq (M')^2. \tag{2.209}$$

Remembering again the equivalence holding between the norm in $W^{1,2}(\Omega_R \setminus \bar{D})$ and the one in $H^1(\Omega_R \setminus \bar{D})$, relation (2.209) means that $\|\Phi(\cdot, z^*)\|^2_{H^1(\Omega_R\setminus\bar{D})}$ exists finite, against (2.204). This concludes the proof. ■

Theorem 2.4.10. [general theorem] *Let $D \subset \mathbb{R}^2$ be a nonempty, open and bounded set such that its boundary ∂D is of class C^2 and the exterior domain $\mathbb{R}^2 \setminus \bar{D}$ is connected. Let the matrix-valued function $\mathbf{A}' : \bar{D} \to \mathbb{C}^{2\times 2}$, with $\mathbf{A}' = \left(a'_{jk}\right)_{j,k=1,2}$, satisfy the following properties:*

1. *the functions a'_{jk} are continuously differentiable, i.e. $a'_{jk} \in C^1(\bar{D})$ $\forall j, k = 1, 2$;*
2. *the matrix-valued function $\mathrm{Re}(\mathbf{A}') : \bar{D} \to \mathbb{R}^{2\times 2}$, defined by $(\mathrm{Re}(\mathbf{A}')(x))_{i,j} := \mathrm{Re}(a'_{jk}(x))$ $\forall j, k = 1, 2$ and $\forall x \in \bar{D}$, is symmetric and verifies the condition:*

   either

   $$\exists\, \gamma > 1 \quad | \quad \bar{\xi} \cdot \mathrm{Re}(\mathbf{A}'(x))\, \xi \geq \gamma \, \|\xi\|^2_{\mathbb{C}^2} \qquad \forall \xi \in \mathbb{C}^2, \ \forall x \in \bar{D}, \tag{2.210}$$

   or

   $$\exists\, \gamma > 1 \quad | \quad \bar{\xi} \cdot \mathrm{Re}(\mathbf{A}'(x))^{-1}\, \xi \geq \gamma \, \|\xi\|^2_{\mathbb{C}^2} \qquad \forall \xi \in \mathbb{C}^2, \ \forall x \in \bar{D}; \tag{2.211}$$

3. *the matrix-valued function $\mathrm{Im}(\mathbf{A}') : \bar{D} \to \mathbb{R}^{2\times 2}$, defined by $(\mathrm{Im}(\mathbf{A}')(x))_{i,j} := \mathrm{Im}(a'_{jk}(x))$ $\forall j, k = 1, 2$ and $\forall x \in \bar{D}$, is symmetric and verifies the condition:*

   $$\bar{\xi} \cdot \mathrm{Im}(\mathbf{A}'(x))\, \xi \leq 0 \qquad \forall \xi \in \mathbb{C}^2, \ \forall x \in \bar{D}. \tag{2.212}$$

Moreover, let $n \in C(\bar{D})$ be such that $\mathrm{Im}(n(x)) \geq 0$ for all $x \in \bar{D}$. Finally, let us assume that there exists a point $x_0 \in D$ such that either

$$\mathrm{Im}(n(x_0)) > 0 \tag{2.213}$$

[28] See, for example, [45], p. 300.

or

$$\bar{\xi} \cdot \mathrm{Im}(\mathbf{A}'(x_0))\,\xi < 0 \qquad \forall \xi \in \mathbb{C}^2 \setminus (0,0). \tag{2.214}$$

Then, if F is the far-field operator (2.97) corresponding to system (2.94), the following facts hold.

1. *If $z \in D$, then for every $\varepsilon > 0$ there exists a solution $g_z^\varepsilon \in L^2[0,2\pi]$ of the inequality*

$$\|Fg_z^\varepsilon - \Phi_\infty(\cdot,z)\|_{L^2[0,2\pi]} \leq \varepsilon; \tag{2.215}$$

moreover this solution is such that

$$\lim_{z\to\partial D} \|g_z^\varepsilon\|_{L^2[0,2\pi]} = \infty \tag{2.216}$$

and

$$\lim_{z\to\partial D} \|v_{g_z^\varepsilon}\|_{H^1(D)} = \infty, \tag{2.217}$$

where $v_{g_z^\varepsilon}$ is the Herglotz wave function with kernel g_z^ε.

2. *If $z \in \mathbb{R}^2 \setminus \bar{D}$, then for every $\varepsilon > 0$ and $\delta > 0$ there exists a solution $g_z^{\varepsilon,\delta} \in L^2[0,2\pi]$ of the inequality*

$$\left\|Fg_z^{\varepsilon,\delta} - \Phi_\infty(\cdot,z)\right\|_{L^2[0,2\pi]} \leq \varepsilon + \delta; \tag{2.218}$$

moreover this solution is such that

$$\lim_{\delta\to 0^+} \|g_z^{\varepsilon,\delta}\|_{L^2[0,2\pi]} = \infty \tag{2.219}$$

and

$$\lim_{\delta\to 0^+} \|v_{g_z^{\varepsilon,\delta}}\|_{H^1(D)} = \infty, \tag{2.220}$$

where $v_{g_z^{\varepsilon,\delta}}$ is the Herglotz wave function with kernel $g_z^{\varepsilon,\delta}$.

Proof. We obviously distinguish two cases: $z \in D$ and $z \in \mathbb{R}^2 \setminus \bar{D}$.

1) Let us assume that $z \in D$. We formulate the following interior transmission problem: find $(v,w) \in H^1(D) \oplus H^1(D)$ such that

$$\begin{cases} \nabla_2 \cdot \mathbf{A}' \nabla_2 v + k^2 n v = 0 & \text{in } D \quad \text{(a)} \\ \Delta_2 w + k^2 w = 0 & \text{in } D \quad \text{(b)} \\ v - w = \Phi(\cdot,z) & \text{on } \partial D \quad \text{(c)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} - \dfrac{\partial w}{\partial \nu} = \dfrac{\partial \Phi(\cdot,z)}{\partial \nu} & \text{on } \partial D. \quad \text{(d)} \end{cases} \tag{2.221}$$

Since the hypotheses of either theorem 2.3.9 or theorem 2.3.11 are clearly satisfied, system (2.221) has a unique solution $(v,w) \in H^1(D) \oplus H^1(D)$; since, as already observed, equations (2.195)(b) and (2.195)(e) are identically satisfied, from system (2.221) it easily follows that

$(v, \Phi(\cdot, z))$ solves system (2.38) with boundary data $(f, h) \in H(\partial D)$ defined as $f := w|_{\partial D}$, $h := \left.\dfrac{\partial w}{\partial \nu}\right|_{\partial D}$, i.e.

$$\begin{cases} \nabla_2 \cdot \mathbf{A}' \nabla_2 v + k^2 n v = 0 & \text{in } D \quad \text{(a)} \\ \Delta_2 \Phi(\cdot, z) + k^2 \Phi(\cdot, z) = 0 & \text{in } \mathbb{R}^2 \setminus \bar{D} \quad \text{(b)} \\ v - \Phi(\cdot, z) = w & \text{on } \partial D \quad \text{(c)} \\ \dfrac{\partial v}{\partial \nu_{\mathbf{A}'}} - \dfrac{\partial \Phi(\cdot, z)}{\partial \nu} = \dfrac{\partial w}{\partial \nu} & \text{on } \partial D \quad \text{(d)} \\ \lim\limits_{r \to \infty} \sqrt{r} \left(\dfrac{\partial \Phi(\cdot, z)}{\partial r} - \mathrm{i}k \Phi(\cdot, z) \right) = 0. & \text{(e)} \end{cases} \tag{2.222}$$

Since the scattered field is, in this case, $\Phi(\cdot, z)$, which has far-field pattern $\Phi_\infty(\cdot, z)$, we can conclude that $\Phi_\infty(\cdot, z)$ is in the range of B_0, that is, recalling (2.182),

$$B_0 \left(w|_{\partial D}, \left.\frac{\partial w}{\partial \nu}\right|_{\partial D} \right) = \Phi_\infty(\cdot, z). \tag{2.223}$$

Then, by virtue of theorem 2.4.7, for any $\varepsilon > 0$, we can find a g_z^ε such that

$$\| B_0(\mathcal{H} g_z^\varepsilon) - \Phi_\infty(\cdot, z) \|_{L^2[0, 2\pi]} \leq \varepsilon, \tag{2.224}$$

which, remembering factorization (2.193), is just thesis (2.215); moreover, recalling the proof of theorem 2.4.7 itself (in particular, relation (2.186)), we have that the corresponding Herglotz wave function $v_{g_z^\varepsilon}$ approximates w in the $H^1(D)$-norm.

Our aim is now to show that if z approaches the boundary ∂D from the interior of D, then both g_z^ε and $v_{g_z^\varepsilon}$ blow up in their respective norms. To this end, let z^* be any point of the boundary of ∂D and let $\{z_j\}_{j=0}^\infty \subset D$ be any sequence of points in D such that $\lim\limits_{j \to \infty} \| z_j - z^* \|_{\mathbb{R}^2} = 0$; for example, for $L > 0$ small enough, we can define such a sequence in the following way:

$$z_j := z^* - \frac{L}{j} \nu(z^*), \qquad j \in \mathbb{N}, \tag{2.225}$$

where $\nu(z^*)$ is the unit outward normal at z^*. We denote with (v_j, w_j) the unique solution to system (2.221) corresponding to $z = z_j$. As $j \to \infty$, the points z_j approach the boundary point z^*; hence, by virtue of lemma 2.4.9, we have

$$\lim_{j \to \infty} \| \Phi(\cdot, z_j) \|_{H^1(\Omega_R \setminus \bar{D})} = \infty, \tag{2.226}$$

where Ω_R is, as in lemma 2.4.9 itself, an open and origin-centred disk containing $\bar{D}$.

Now, by means of relation (2.226) it is possible to prove that

$$\lim_{j \to \infty} \| w_j \|_{H^1(D)} = \infty. \tag{2.227}$$

Indeed, let us assume, by absurd, that limit (2.227) does not hold: hence, there exist a subsequence $\{z_{j_k}\}_{k=0}^{\infty}$ of $\{z_j\}_{j=0}^{\infty}$ and a positive constant C_0 such that

$$\|w_{j_k}\|_{H^1(D)} \leq C_0 \qquad \forall k \in \mathbb{N}. \tag{2.228}$$

From the trace theorems A.15.2 and A.17.1 (i.e., more precisely, from inequalities (A.131) and (A.150)), we immediately get:

$$\|w_{j_k}\|_{H^{\frac{1}{2}}(\partial D)} \leq C_1 C_0 \qquad \text{and} \qquad \left\|\frac{\partial w_{j_k}}{\partial \nu}\right\|_{H^{-\frac{1}{2}}(\partial D)} \leq C_3 C_0 \qquad \forall k \in \mathbb{N}. \tag{2.229}$$

We now recall that, for any $k \in \mathbb{N}$, the pair $(v_{j_k}, \Phi(\cdot, z_{j_k}))$ is the unique solution to system (2.38) with boundary data $(f,h) := \left(w_{j_k}|_{\partial D}, \left.\frac{\partial w_{j_k}}{\partial \nu}\right|_{\partial D}\right)$; hence we can use the a priori estimate (2.52), which, together with relations (2.229), gives:

$$\|v_{j_k}\|_{H^1(D)} + \|\Phi(\cdot, z_{j_k})\|_{H^1(\Omega_R \setminus \bar{D})} \leq C \left(\|w_{j_k}\|_{H^{\frac{1}{2}}(\partial D)} + \left\|\frac{\partial w_{j_k}}{\partial \nu}\right\|_{H^{-\frac{1}{2}}(\partial D)} \right) \leq C C_0 (C_1 + C_3), \tag{2.230}$$

which clearly contradicts relation (2.226). Hence limit (2.227) has to hold.

On the other hand, by virtue of what already observed about inequality (2.224), for any $j \in \mathbb{N}$ there exists a Herglotz wave function $v_{g^{\varepsilon}_{z_j}}$ (with kernel $g^{\varepsilon}_{z_j}$ satisfying inequality (2.224) written for z_j instead of z) that approximates w_j in the $H^1(D)$-norm; more precisely, for each $\varepsilon > 0$ chosen for inequality (2.224), there exists an $\varepsilon'(\varepsilon)$, with $\lim_{\varepsilon \to 0^+} \varepsilon'(\varepsilon) = 0$, such that

$$\|v_{g^{\varepsilon}_{z_j}} - w_j\|_{H^1(D)} \leq \varepsilon'(\varepsilon). \tag{2.231}$$

Since, by the continuity of the norm, it holds:

$$\left| \|v_{g^{\varepsilon}_{z_j}}\|_{H^1(D)} - \|w_j\|_{H^1(D)} \right| \leq \|v_{g^{\varepsilon}_{z_j}} - w_j\|_{H^1(D)}, \tag{2.232}$$

relations (2.227) and (2.231) clearly imply that

$$\lim_{j \to \infty} \|v_{g^{\varepsilon}_{z_j}}\|_{H^1(D)} = \infty. \tag{2.233}$$

Moreover, remembering the definitions (2.98) of Herglotz wave function and (A.50) of scalar product in $L^2[0, 2\pi]$ and applying the Cauchy-Schwarz inequality, we have:

$$\begin{aligned} |v_{g^{\varepsilon}_{z_j}}(x)| &= \left| \int_0^{2\pi} e^{\mathrm{i}kx\cdot d}\, g^{\varepsilon}_{z_j}(\theta) d\theta \right| = \left| \left(g^{\varepsilon}_{z_j}, e^{-\mathrm{i}kx\cdot d} \right)_{L^2[0,2\pi]} \right| \leq \\ &\leq \|g^{\varepsilon}_{z_j}\|_{L^2[0,2\pi]}\ \|e^{-\mathrm{i}kx\cdot d}\|_{L^2[0,2\pi]} = \sqrt{2\pi}\, \|g^{\varepsilon}_{z_j}\|_{L^2[0,2\pi]}. \end{aligned} \tag{2.234}$$

We now remember (see theorem A.13.2) that $H^1(D) = W^{1,2}(D)$ with equivalent norms; hence, recalling the definition of norm in $W^{1,2}(D)$ (see (A.66)) and using the previous inequality (2.234), we can write[29]:

$$\|v_{g^\varepsilon_{z_j}}\|_{H^1(D)} \leq M_2 \left\{ \|v_{g^\varepsilon_{z_j}}\|^2_{L^2(D)} + \|\partial_1 v_{g^\varepsilon_{z_j}}\|^2_{L^2(D)} + \|\partial_2 v_{g^\varepsilon_{z_j}}\|^2_{L^2(D)} \right\}^{\frac{1}{2}} =$$

$$= M_2 \left\{ \left[1 + k^2(|d_1|^2 + |d_2|^2)\right] \|v_{g^\varepsilon_{z_j}}\|^2_{L^2(D)} \right\}^{\frac{1}{2}} = \left\{ \left[1 + k^2(|d_1|^2 + |d_2|^2)\right] \int_D |v_{g^\varepsilon_{z_j}}(x)|^2 dx \right\}^{\frac{1}{2}} \leq$$

$$\leq M_2 \left[1 + k^2(|d_1|^2 + |d_2|^2)\right]^{\frac{1}{2}} \sqrt{2\pi}\, \|g^\varepsilon_{z_j}\|_{L^2[0,2\pi]} \sqrt{\operatorname{mis} D} = A\, \|g\|_{L^2[0,2\pi]}, \tag{2.235}$$

where M_2 is a real positive constant (cf. (A.3)), $(d_1, d_2) \in \mathbb{R}^2$ are the components of the unit vector d and $A := M_2 \left[1 + k^2(|d_1|^2 + |d_2|^2)\right]^{\frac{1}{2}} \sqrt{2\pi} \sqrt{\operatorname{mis} D}$.

Finally, from limit (2.233) and relation (2.235), we immediately get:

$$\lim_{j\to\infty} \|g^\varepsilon_{z_j}\|_{L^2[0,2\pi]} = \infty. \tag{2.236}$$

Since limits (2.236) and (2.233) hold for any sequence $\{z_j\}_{j=0}^\infty \subset D$ such that $\lim_{j\to\infty} \|z_j - z^*\|_{\mathbb{R}^2} = 0$, we conclude that the thesis limits (2.216) and (2.217) hold too.

2) Let us now assume that $z \in \mathbb{R}^2 \setminus \bar{D}$. Then, by virtue of lemma 2.4.8, we have that $\Phi(\cdot, z) \notin H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ and then $\Phi(\cdot, z)$ cannot be a weak (i.e. in $H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$) solution to the Helmholtz equation in the exterior of D (moreover, from remark 2.2.1, we know that if it were such a solution, $\Phi(\cdot, z)$ would be analytic in $\mathbb{R}^2 \setminus \bar{D}$, which is clearly not the case, owing to its singularity in z).

This implies, in particular, that $\Phi_\infty(\cdot, z)$ cannot belong to the (dense) range of the operator B_0, but only to its closure. Indeed, let us suppose, by absurd, that $\Phi_\infty(\cdot, z) \in \mathcal{R}(B_0)$: this means that there exists a weak solution $(v, u^s) \in H^1(D) \oplus H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ of system (2.38) (for some boundary data $(f, h) \in H(\partial D)$) such that the far-field pattern u_∞ of the radiating scattered field u^s coincides with $\Phi_\infty(\cdot, z)$. Now, let us consider a disc $\Omega_R := \{(x, y) \in \mathbb{R}^2 \mid \|(x, y)\|_{\mathbb{R}^2} < R\}$ large enough to contain $\bar{D} \cup \{z\}$: of course, u_∞ and $\Phi_\infty(\cdot, z)$ are respectively the far-field patterns of the radiating solutions $u^s, \Phi(\cdot, z) \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{\Omega}_R)$ of the Helmholtz equation. Hence, remembering remark 2.2.3, it holds $u^s = \Phi(\cdot, z)$, i.e. $u^s - \Phi(\cdot, z) = 0$, in $\mathbb{R}^2 \setminus \bar{\Omega}_R$ and consequently, by virtue of the unique continuation property enjoyed by real-analytic functions (see theorem A.4.3), $u^s - \Phi(\cdot, z) = 0$, i.e. $u^s = \Phi(\cdot, z)$, in $\mathbb{R}^2 \setminus \left\{\bar{D} \cup U_z\right\}$, where U_z is an arbitrarily small neighbourhood of z. This implies that u^s cannot be a real-analytic function in $\mathbb{R}^2 \setminus \bar{D}$, in contradiction (remembering remark 2.2.1) with the previous assumption that $u^s \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ is a weak and radiating solution of the Helmholtz equation.

[29] According to the notations of section A.2, we denote with ∂_1 [resp. ∂_2] the partial derivative operator $\frac{\partial}{\partial x_1}$ [resp. $\frac{\partial}{\partial x_2}$].

Summing up, the equation

$$B_0(f,h) = \Phi_\infty(\cdot,z) \tag{2.237}$$

for the unknown $(f,h) \in H(\partial D)$ is impossible for any $z \in \mathbb{R}^2 \setminus \bar{D}$. Such an impossibility does not derive from the presence of noise on the "datum" $\Phi_\infty(\cdot,z)$ (which is properly not a datum, since it is rather a known analytic function) or on the operator B_0 (which is now supposed to be known exactly): in fact, equation (2.237), just as the far-field equation (2.191), has been introduced a priori and it does not describe any physical phenomenon, so it does not need to be solvable in absence of noise (cf. remark 1.6.2). However, there is nothing preventing us from computing the Tikhonov regularized solution $(f_{\alpha(z)}, h_{\alpha(z)})$ of equation (2.237) for any $z \in \mathbb{R}^2 \setminus \bar{D}$; but then the question arises: how to choose the value $\alpha^*(z)$ of the regularization parameter $\alpha(z)$? In the current situation, it would be meaningless to apply the generalized discrepancy principle or any other criterion taking into account the noise levels δ and h on the datum and on the operator respectively: indeed, this would imply no regularization at all, since here $\delta = h = 0$, and, consequently, no kind of solution at all to equation (2.237), which does not admit even the generalized solution, owing to the fact that $\Phi_\infty(\cdot,z) \in \overline{\mathcal{R}(B_0)} \setminus \mathcal{R}(B_0)$. Then, the next step is to give a class of parameter choice rules which enable us to go on with the proof of the theorem.

To this end, we recall property No 5 of lemma 1.8.1, which we rewrite here below for the reader's convenience:

$$\lim_{\alpha\to 0^+} \|A_h f_\alpha^\eta - g_\delta\|_Y = \inf_{f\in X} [A_h f - g_\delta]_Y \,, \tag{2.238}$$

and which in our case reads

$$\lim_{\alpha(z)\to 0^+} \left\|B_0(f_{\alpha(z)}, h_{\alpha(z)}) - \Phi_\infty(\cdot,z)\right\|_{L^2[0,2\pi]} = 0 \qquad \forall z \in \mathbb{R}^2 \setminus \bar{D}, \tag{2.239}$$

having remembered that now $\delta = h = 0$ (i.e. neither the "datum" $\Phi_\infty(\cdot,z)$ nor the operator B_0 respectively are affected by noise) and $\overline{\mathcal{R}(B_0)} = L^2[0,2\pi]$. The previous limit (2.239) implies that for any $z \in \mathbb{R}^2 \setminus \bar{D}$ and for any $\delta > 0$ (which has clearly *nothing* to do with the noise level on the "datum" $\Phi_\infty(\cdot,z)$) there exists an $\alpha_0(\delta,z) > 0$ such that

$$\left\|B_0(f_{\alpha(z)}, h_{\alpha(z)}) - \Phi_\infty(\cdot,z)\right\|_{L^2[0,2\pi]} \le \delta \qquad \forall \alpha(z) \in (0, \alpha_0(\delta,z)). \tag{2.240}$$

It follows that for any $z \in \mathbb{R}^2 \setminus \bar{D}$ it is always possible to choose, as parameter choice rule, a function $\alpha^*(\cdot,z) : \mathbb{R}^+ \to \mathbb{R}^+$ mapping δ into $\alpha^*(\delta,z)$ and verifying the two following properties:

$$\left\|B_0(f_{\alpha^*(\delta,z)}, h_{\alpha^*(\delta,z)}) - \Phi_\infty(\cdot,z)\right\|_{L^2[0,2\pi]} \le \delta \qquad \forall \delta > 0,\ \forall z \in \mathbb{R}^2 \setminus \bar{D}; \tag{2.241}$$

$$\lim_{\delta\to 0^+} \alpha^*(\delta,z) = 0 \qquad \forall z \in \mathbb{R}^2 \setminus \bar{D}. \tag{2.242}$$

All the functions $\alpha^*(\cdot,z)$ satisfying conditions (2.241) and (2.242) form the family of the parameter choice rules we are interested in: from now on, we arbitrarily choose one of them and denote it simply with $\alpha^*(\cdot,z)$.

Furthermore, by virtue of theorem 2.4.7, for any $z \in \mathbb{R}^2 \setminus \bar{D}$ and for any $\varepsilon > 0$ there exists a function $g^{\varepsilon}_{\alpha^*(\delta,z)} \in L^2[0, 2\pi]$, denoted from now on as $g^{\varepsilon,\delta}_z$, such that:

$$\left\| B_0(\mathcal{H} g^{\varepsilon,\delta}_z) - B_0(f_{\alpha^*(\delta,z)}, h_{\alpha^*(\delta,z)}) \right\|_{L^2[0,2\pi]} \leq \varepsilon. \tag{2.243}$$

Hence, by means of the triangle inequality and of relations (2.241), (2.243), for any $z \in \mathbb{R}^2 \setminus \bar{D}$ it holds:

$$\begin{aligned} &\left\| B_0(\mathcal{H} g^{\varepsilon,\delta}_z) - \Phi_\infty(\cdot, z) \right\|_{L^2[0,2\pi]} \leq \\ &\quad \leq \left\| B_0(\mathcal{H} g^{\varepsilon,\delta}_z) - B_0(f_{\alpha^*(\delta,z)}, h_{\alpha^*(\delta,z)}) \right\|_{L^2[0,2\pi]} + \left\| B_0(f_{\alpha^*(\delta,z)}, h_{\alpha^*(\delta,z)}) - \Phi_\infty(\cdot, z) \right\|_{L^2[0,2\pi]} \leq \varepsilon + \delta, \end{aligned} \tag{2.244}$$

which, remembering factorization (2.193), is just thesis (2.218).

We now observe that since $(f_{\alpha^*(\delta,z)}, h_{\alpha^*(\delta,z)}) \in D(B_0) = H(\partial D)$ for any $z \in \mathbb{R}^2 \setminus \bar{D}$ and for any $\delta > 0$, by definition of $H(\partial D)$ there exists a function $w_{\alpha^*(\delta,z)} \in \overline{H}$ such that

$$(f_{\alpha^*(\delta,z)}, h_{\alpha^*(\delta,z)}) = \left(w_{\alpha^*(\delta,z)}|_{\partial D}, \left. \frac{\partial w_{\alpha^*(\delta,z)}}{\partial \nu} \right|_{\partial D} \right). \tag{2.245}$$

Recalling the proof of theorem 2.4.7 (in particular, relation (2.186)), we have that such a function $w_{\alpha^*(\delta,z)}$ is approximated in the $H^1(D)$-norm by the Herglotz wave function $v_{g^{\varepsilon,\delta}_z}$ with kernel $g^{\varepsilon,\delta}_z$ introduced to write relation (2.243); more precisely, for any $\varepsilon > 0$ chosen for inequality (2.243), there exists an $\varepsilon'(\varepsilon)$, with $\lim_{\varepsilon \to 0^+} \varepsilon'(\varepsilon) = 0$, such that

$$\left\| v_{g^{\varepsilon,\delta}_z} - w_{\alpha^*(\delta,z)} \right\|_{H^1(D)} \leq \varepsilon'(\varepsilon). \tag{2.246}$$

By virtue of trace theorems A.15.2, A.17.1 (i.e., more precisely, by virtue of inequalities (A.131) and (A.150)), from relations (2.245) and (2.246) it follows:

$$\left\| v_{g^{\varepsilon,\delta}_z} - f_{\alpha^*(\delta,z)} \right\|_{H^{\frac{1}{2}}(\partial D)} = \left\| v_{g^{\varepsilon,\delta}_z} - w_{\alpha^*(\delta,z)} \right\|_{H^{\frac{1}{2}}(\partial D)} \leq C_1\, \varepsilon'(\varepsilon), \tag{2.247}$$

$$\left\| \frac{\partial v_{g^{\varepsilon,\delta}_z}}{\partial \nu} - h_{\alpha^*(\delta,z)} \right\|_{H^{-\frac{1}{2}}(\partial D)} = \left\| \frac{\partial v_{g^{\varepsilon,\delta}_z}}{\partial \nu} - \frac{\partial w_{\alpha^*(\delta,z)}}{\partial \nu} \right\|_{H^{-\frac{1}{2}}(\partial D)} \leq C_3\, \varepsilon'(\varepsilon). \tag{2.248}$$

Hence, taking the square of relations (2.247), (2.248) and summing them, we have:

$$\begin{aligned} &\left\| \left(v_{g^{\varepsilon,\delta}_z}, \frac{\partial v_{g^{\varepsilon,\delta}_z}}{\partial \nu} \right) - \left(f_{\alpha^*(\delta,z)}, h_{\alpha^*(\delta,z)} \right) \right\|^2_{H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)} = \\ &\quad = \left\| v_{g^{\varepsilon,\delta}_z} - f_{\alpha^*(\delta,z)} \right\|^2_{H^{\frac{1}{2}}(\partial D)} + \left\| \frac{\partial v_{g^{\varepsilon,\delta}_z}}{\partial \nu} - h_{\alpha^*(\delta,z)} \right\|^2_{H^{-\frac{1}{2}}(\partial D)} \leq \left(C_1^2 + C_3^2 \right) [\varepsilon'(\varepsilon)]^2. \end{aligned} \tag{2.249}$$

We now remember that the norm on $H(\partial D)$ is the restriction to $H(\partial D)$ itself of the norm on $H^{\frac{1}{2}}(\partial D) \oplus H^{-\frac{1}{2}}(\partial D)$ (cf. the brief comment soon below lemma 2.4.3) and observe that both the

pairs $\left(v_{g_z^{\varepsilon,\delta}}, \dfrac{\partial v_{g_z^{\varepsilon,\delta}}}{\partial \nu}\right)\Big|_{\partial D}$ and $\left(f_{\alpha^*(\delta,z)}, h_{\alpha^*(\delta,z)}\right)|_{\partial D}$ belong to $H(\partial D)$. Hence, as a consequence of the continuity of such a norm, we can write:

$$\left| \left\| \left(v_{g_z^{\varepsilon,\delta}}, \frac{\partial v_{g_z^{\varepsilon,\delta}}}{\partial \nu} \right) \right\|_{H(\partial D)} - \left\| \left(f_{\alpha^*(\delta,z)}, h_{\alpha^*(\delta,z)} \right) \right\|_{H(\partial D)} \right| \leq$$

$$\leq \left\| \left(v_{g_z^{\varepsilon,\delta}}, \frac{\partial v_{g_z^{\varepsilon,\delta}}}{\partial \nu} \right) - \left(f_{\alpha^*(\delta,z)}, h_{\alpha^*(\delta,z)} \right) \right\|_{H(\partial D)}. \tag{2.250}$$

By means of a comparison between relations (2.249) and (2.250), we immediately get

$$\left| \left\| \left(v_{g_z^{\varepsilon,\delta}}, \frac{\partial v_{g_z^{\varepsilon,\delta}}}{\partial \nu} \right) \right\|_{H(\partial D)} - \left\| \left(f_{\alpha^*(\delta,z)}, h_{\alpha^*(\delta,z)} \right) \right\|_{H(\partial D)} \right| \leq \sqrt{C_1^2 + C_3^2}\, \varepsilon'(\varepsilon). \tag{2.251}$$

On the other hand, we have that

$$\left(f_{\alpha^*(\delta,z)}, h_{\alpha^*(\delta,z)} \right) = R_{\alpha^*(\delta,z)} \Phi_\infty(\cdot, z), \tag{2.252}$$

where, for each $\delta > 0$ and for each $z \in \mathbb{R}^2 \setminus \bar{D}$, $R_{\alpha^*(\delta,z)}$ is an operator belonging to the family of (linear) Tikhonov's regularization operators $\{R_\alpha\}_{\alpha>0}$, defined, in our case, as (cf. equality (1.128)):

$$R_\alpha = B_0^*(B_0 B_0^* + \alpha I)^{-1} \qquad \forall \alpha > 0. \tag{2.253}$$

Moreover, rewriting relation (1.158) for our case gives:

$$\left\| B_0 B_0^* (B_0 B_0^* + \alpha I)^{-1} \right\| \leq 1 \qquad \forall \alpha > 0, \tag{2.254}$$

then we have:

$$\sup_{\alpha>0} \|B_0 R_\alpha\| < \infty, \tag{2.255}$$

i.e. hypothesis (1.97) of theorem 1.6.7 is verified; hence also its thesis (1.98) holds. In particular, since, as already observed, $\Phi_\infty(\cdot, z) \notin D(B_0^\dagger)$, we have that

$$\lim_{\sigma \to 0^+} \|R_\sigma \Phi_\infty(\cdot, z)\|_{H(\partial D)} = \infty \qquad \forall z \in \mathbb{R}^2 \setminus \bar{D}, \tag{2.256}$$

where limit (2.256) is to be intended over the set of σ-values which are in the range of $\alpha^*(\cdot, z)$. By virtue of relation (2.242), we can rewrite limit (2.256) as

$$\lim_{\delta \to 0^+} \left\| R_{\alpha^*(\delta,z)} \Phi_\infty(\cdot, z) \right\|_{H(\partial D)} = \infty \qquad \forall z \in \mathbb{R}^2 \setminus \bar{D}. \tag{2.257}$$

Remembering now relations (2.251), (2.252) and (2.257), we easily get:

$$\lim_{\delta \to 0^+} \left\| \left(v_{g_z^{\varepsilon,\delta}}, \frac{\partial v_{g_z^{\varepsilon,\delta}}}{\partial \nu} \right) \right\|_{H(\partial D)} = \infty, \tag{2.258}$$

i.e.

$$\lim_{\delta\to 0^+}\left\{\left\|v_{g_z^{\varepsilon,\delta}}\right\|_{H^{\frac{1}{2}}(\partial D)}+\left\|\frac{\partial v_{g_z^{\varepsilon,\delta}}}{\partial\nu}\right\|_{H^{\frac{1}{2}}(\partial D)}\right\}=\infty. \tag{2.259}$$

Furthermore, by applying trace theorems A.15.2 and A.17.1 again (i.e., more precisely, by virtue of inequalities (A.131) and (A.150)), we have:

$$\left\|v_{g_z^{\varepsilon,\delta}}\right\|_{H^{\frac{1}{2}}(\partial D)}\leq C_1\left\|v_{g_z^{\varepsilon,\delta}}\right\|_{H^1(D)}, \tag{2.260}$$

$$\left\|\frac{\partial v_{g_z^{\varepsilon,\delta}}}{\partial\nu}\right\|_{H^{-\frac{1}{2}}(\partial D)}\leq C_3\left\|v_{g_z^{\varepsilon,\delta}}\right\|_{H^1(D)}. \tag{2.261}$$

Summing inequalities (2.260) and (2.261), we get:

$$\left\|v_{g_z^{\varepsilon,\delta}}\right\|_{H^{\frac{1}{2}}(\partial D)}+\left\|\frac{\partial v_{g_z^{\varepsilon,\delta}}}{\partial\nu}\right\|_{H^{\frac{1}{2}}(\partial D)}\leq (C_1+C_3)\|v_{g_z^{\varepsilon,\delta}}\|_{H^1(D)}. \tag{2.262}$$

Taking now into account relations (2.259) and (2.262), we easily find:

$$\lim_{\delta\to 0^+}\left\|v_{g_z^{\varepsilon,\delta}}\right\|_{H^1(D)}=\infty, \tag{2.263}$$

which is exactly thesis (2.220).

Finally, thesis (2.219) follows from (2.220) using an argument analogous to the one employed to prove that limit (2.236) follows from (2.233). ■

We conclude this section by pointing out that although we have stated and proved the general theorem for the case of a penetrable, orthotropic, inhomogeneous and cylinder-shaped medium scattering a TE polarized electromagnetic wave, analogous theorems hold not only for TM polarization, but also for the obstacle (i.e. impenetrable) case (with various boundary conditions and with both TE and TM polarization), as well as for genuine 3D scatterers. We refer to [15] (and its bibliography) for all these (and other, more general) cases.

2.5. The linear sampling method

The general theorem 2.4.10 states, in particular, that, under suitable hypotheses, an approximate solution for the far-field equation exists whose $L^2[0,2\pi]$-norm blows up to infinity for all points approaching the boundary of the scatterer from inside and stays arbitrarily large outside. The existence of such an approximate solution is the mathematical key idea inspiring the linear sampling method [26]: the latter, indeed, is a qualitative method for the visualization of scatterers (under fixed-frequency scattering conditions) based on the plot of a regularized

solution of the far-field equation, written in an angle-discretized version obtained from a finite number of (noisy) measurements of the far-field pattern of the scattered field.

More precisely, the linear sampling method applies the regularization theory for linear inverse problems in order to provide a regularized solution of an angle-discretized version of the far-field equation according to the following algorithm [29]: given $P \times Q$ (noisy) measurements of the far-field pattern at P observation angles and for Q incident fixed-frequency fields:

- take a grid of points in $\mathbb{R}^2$ (or $\mathbb{R}^3$) covering a region in which the scatterer is known to be located;
- for each grid point, determine the Tikhonov regularized solution of the linear system obtained as a discretization of the far-field equation over the incidence and observation angles;
- for each grid point, choose the optimal[30] value of the regularization parameter by means of the generalized discrepancy principle;
- for each grid point, map the Euclidean norm of the optimal regularized solution.

Hence, the linear sampling method turns out to be a qualitative method for the solution of inverse scattering problems based on the observation that the scatterer profile can be detected by all grid points where the Euclidean norm of the optimal regularized solution is mostly large. Of course a visualization of the scatterer profile can be obtained by mapping the values of an appropriate monotonically increasing or decreasing function $I : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ of the norm itself. The composition of I with the Euclidean norm of the optimal regularized solution represents the so-called *indicator function*, defined over the grid. The scatterer profile is given by the grid points where the indicator function is respectively small or large, depending on the decreasing or increasing monotonicity of I.

We can now give a detailed description of the effective implementation of the linear sampling method. To this end, we firstly observe that, in real experiments, the far-field pattern is measured for P observation angles $\{\varphi_i\}_{i=0}^{P-1}$ and Q incidence angles $\{\theta_j\}_{j=0}^{Q-1}$, i.e. for observation directions $\{\hat{x}_i = (\cos\varphi_i, \sin\varphi_i)\}_{i=0}^{P-1}$ and incidence directions $\{d_j = (\cos\theta_j, \sin\theta_j)\}_{j=0}^{Q-1}$. In the following we shall assume $P = Q = N$, the generalization to rectangular problems being

[30] From now on, for sake of brevity we decide to use the adjective *optimal* (put before expressions like "regularization parameter", "regularized solution" or similar ones) only to intend that the value of the regularization parameter has been chosen by means of the generalized discrepancy principle in one of its possible formulations, as explained in section 1.8. Hence, in our context, the meaning of this adjective is different from the one of remark 1.6.7 and has clearly nothing to do with the most common one (in regularization theory), which concerns the rate of convergence of the regularized solution to the generalized one (see, for example, chapter 3 in [33]).

straightforward. Furthermore, for sake of simplicity, we shall take

$$\varphi_i = \frac{2\pi i}{N}, \qquad \theta_j = \frac{2\pi j}{N}, \qquad i,j = 0,\dots,N-1. \tag{2.264}$$

These values are placed into the *far-field matrix* $\mathbf{F}$, whose elements are defined as

$$\mathbf{F}_{ij} := u_\infty(\hat{x}_i, d_j). \tag{2.265}$$

In practical applications the far-field matrix is affected by the measurement noise, and therefore only a noisy version $\mathbf{F}_h$ of the far-field matrix is at disposal, such that

$$\mathbf{F}_h = \mathbf{F} + \mathbf{H}, \tag{2.266}$$

where $\mathbf{H}$ is the noise matrix. If, as usual, we denote with $\|\cdot\|$ the operatorial norm of a linear continuous operator and regard the matrix $\mathbf{H}$ as a linear continuous operator in $\mathbb{C}^N$, we assume to know that $\|\mathbf{H}\| \leq h$ (cf. also relation (1.64)).

Remark 2.5.1. In our numerical simulations, we firstly compute the exact $N \times N$ far-field matrix $\mathbf{F}$ by using the Nyström method [27] and define the two matrices $\mathbf{F}_{\mathrm{Re}}$ and $\mathbf{F}_{\mathrm{Im}}$ whose elements are given by:

$$(\mathbf{F}_{\mathrm{Re}})_{ij} := \mathrm{Re}(\mathbf{F}_{ij}), \qquad (\mathbf{F}_{\mathrm{Im}})_{ij} := \mathrm{Im}(\mathbf{F}_{ij}) \qquad \forall i,j = 0,\dots,N-1. \tag{2.267}$$

Then we construct the $N \times N$ noise matrix $\mathbf{H}$ and add it to $\mathbf{F}$, following the procedure we are going to explain. We form two distinct $N \times N$ random and real-valued matrices $\mathbf{G}_{\mathrm{Re}}$ and $\mathbf{G}_{\mathrm{Im}}$ in such a way that each entry is randomly chosen according to a Gaussian (or normal) distribution

$$f_{X,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-X)^2}{2\sigma^2}\right] \tag{2.268}$$

having mean value $X = 0$ and standard deviation $\sigma = 1$: such a random process requires an initializing numerical value, which is arbitrarily chosen from time to time by the code user. Then we define the two matrices $\mathbf{H}_{\mathrm{Re}}$ and $\mathbf{H}_{\mathrm{Im}}$ with elements given by:

$$(\mathbf{H}_{\mathrm{Re}})_{ij} := n\,(\mathbf{F}_{\mathrm{Re}})_{ij} \cdot (\mathbf{G}_{\mathrm{Re}})_{ij}, \qquad (\mathbf{H}_{\mathrm{Im}})_{ij} := n\,(\mathbf{F}_{\mathrm{Im}})_{ij} \cdot (\mathbf{G}_{\mathrm{Im}})_{ij} \qquad \forall i,j = 0,\dots,N-1, \tag{2.269}$$

where $n \in \mathbb{R}^+ \cup \{0\}$ represents the noise level (e.g. $n = 1\%$, $n = 5\%$ and so on) and the dot products between the matrix elements are to be intended just entry by entry, having nothing to do with the usual rows $\times$ columns matrix product. Then we form our noise matrix $\mathbf{H}$ as

$$\mathbf{H} := \mathbf{H}_{\mathrm{Re}} + \mathrm{i}\mathbf{H}_{\mathrm{Im}}, \tag{2.270}$$

where obviously $\mathrm{i} = \sqrt{-1}$. Finally we can construct the noisy far-field matrix $\mathbf{F}_h$ just as in equality (2.266).

When, in the following, we shall have to explain our numerical experiments, in order to refer to such a procedure as briefly as possible we shall simply say that we add $n\%$ of Gaussian noise to the (exact) far-field matrix and $n\%$ will be called the *noise level.*

Obviously, for any fixed noise level, there are infinite possible realizations of the noise matrix $\mathbf{H}$. Let $\mathbf{H}_s$ be the specific and known noise matrix used in a certain numerical experiment and let $h_s := \|\mathbf{H}_s\|$: then, in the general relation $\|\mathbf{F}_h - \mathbf{F}\| \leq h$, we shall assume that the estimated bound h to the noise affecting the operator has just the value h_s. □

Remark 2.5.2. We point out that the ones just described are neither the physical way in which noise affects far-field measurements, nor the realistic way in which the bound to the noise affecting the operator is estimated. Indeed, in real experiments, one only measures modulus and phase (each of them with its own error), rather than real and imaginary part of a *noisy* acoustic or electromagnetic field *at finite (however long) distances* from the scatterer. In fact, one of our next tasks is to face the problem of simulating as realistically as possible the overall measurement process: in particular, we need to determine nature and typical percentage values of noise, as well as to find a way of estimating the latter by means of arguments involving only the noisy far-field matrix $\mathbf{F}_h$ and based on an assessment of how far $\mathbf{F}_h$ is from satisfying the reciprocity relation (2.96), which, in absence of noise, should be exactly verified. □

Getting on in our description of the implementation of the linear sampling method, we then create in $\mathbb{R}^2$ a grid $\mathcal{Z} := \{z_l\}_{l=0}^{L-1}$ of L sampling points, covering the region in which the scatterer is located. For each $z_l = r_l(\cos\psi_l, \sin\psi_l) \in \mathcal{Z}$, we perform a discretization of $\Phi_\infty(\hat{x}, z_l)$ (cf. relation (2.75)) by defining the column vector

$$\mathbf{\Phi}_\infty(z_l) := \frac{e^{\mathrm{i}\frac{\pi}{4}}}{\sqrt{8\pi k}} \left[e^{-\mathrm{i}kr_l\cos(\varphi_0-\psi_l)}, \ldots, e^{-\mathrm{i}kr_l\cos(\varphi_{N-1}-\psi_l)}\right]^{\mathrm{T}}. \tag{2.271}$$

Analogously, for each $z_l \in \mathcal{Z}$, the unknown vector $\mathbf{g}(z_l)$ is an element of $\mathbb{C}^N$ with the i-th component given by $g_i(z_l) = g_{z_l}(d_i)$. Therefore the discretized noisy version of the far-field equation (1.13) is given, for each $z_l \in \mathcal{Z}$, by the square linear system

$$\mathbf{F}_h\mathbf{g}(z_l) = \frac{N}{2\pi}\mathbf{\Phi}_\infty(z_l). \tag{2.272}$$

This linear system is ill-conditioned and the numerical instabilities due to the presence of noise can be reduced by applying Tikhonov regularization method, i.e. by determining (cf. definitions (1.172) and (1.173))

$$\mathbf{g}_{\alpha(z_l)}(z_l) = \operatorname{argmin}\left\{\left\|\mathbf{F}_h\mathbf{g}(z_l) - \frac{N}{2\pi}\mathbf{\Phi}_\infty(z_l)\right\|^2_{\mathbb{C}^N} + \alpha(z_l)\,\|\mathbf{g}(z_l)\|^2_{\mathbb{C}^N}\right\}, \tag{2.273}$$

where we have denoted with $\|\cdot\|_{\mathbb{C}^N}$ the Euclidean norm on $\mathbb{C}^N$. In order to determine an explicit form for $\mathbf{g}_{\alpha(z_l)}(z_l)$, we introduce the Singular Value Decomposition (SVD) of the far-

field matrix $\mathbf{F}_h$, i.e.:

$$\mathbf{F}_h\mathbf{w} = \sum_{p=0}^{r_h-1} \sigma_p^h (\mathbf{w}, \mathbf{u}_p^h)_{\mathbb{C}^N} \mathbf{v}_p^h \qquad \forall \mathbf{w} \in \mathbb{C}^N, \tag{2.274}$$

where r_h is the rank of $\mathbf{F}_h$, $\{\sigma_p^h, \mathbf{u}_p^h, \mathbf{v}_p^h\}_{p=0}^{r_h-1}$ is the singular system[31] of $\mathbf{F}_h$ (regarded as a compact linear operator from $\mathbb{C}^N$ in itself) and $(\cdot,\cdot)_{\mathbb{C}^N}$ is the scalar product in $\mathbb{C}^N$.

Remark 2.5.3. We point out that the SVD of the matrix $\mathbf{F}_h$ is nothing else than its singular representation (cf. relation (1.48)) when regarded as a compact linear operator $\mathbf{F}_h : \mathbb{C}^N \to \mathbb{C}^N$. For future purpose, we also recall[32] that representation (2.274) implies

$$\|\mathbf{F}_h\| = \sigma_0^h \tag{2.275}$$

and that the generalized solution of system (2.272) is given by[33]

$$\mathbf{g}_h^\dagger(z_l) = \frac{N}{2\pi} \sum_{p=0}^{r_h-1} \frac{\left(\mathbf{\Phi}_\infty(z_l), \mathbf{v}_p^h\right)_{\mathbb{C}^N}}{\sigma_p^h} \mathbf{u}_p^h. \tag{2.276}$$

□

Now, the Tikhonov regularized solution of system (2.272) can be written in terms of the singular system of the matrix $\mathbf{F}_h$ as (cf. relation (1.177)):

$$\mathbf{g}_{\alpha(z_l)}(z_l) = \frac{N}{2\pi} \sum_{p=0}^{r_h-1} \frac{\sigma_p^h}{(\sigma_p^h)^2 + \alpha(z_l)} (\mathbf{\Phi}_\infty(z_l), \mathbf{v}_p^h)_{\mathbb{C}^N} \mathbf{u}_p^h, \tag{2.277}$$

and consequently its Euclidean norm is

$$\left\|\mathbf{g}_{\alpha(z_l)}(z_l)\right\|_{\mathbb{C}^N} = \frac{N}{2\pi} \sqrt{\sum_{p=0}^{r_h-1} \frac{(\sigma_p^h)^2}{[(\sigma_p^h)^2 + \alpha(z_l)]^2} \left|\left(\mathbf{\Phi}_\infty(z_l), \mathbf{v}_p^h\right)_{\mathbb{C}^N}\right|^2}. \tag{2.278}$$

Then, the optimal regularized solution (and, consequently, its Euclidean norm) is obtained by fixing, for each grid point z_l, the value $\alpha^*(z_l)$ of the regularization parameter $\alpha(z_l)$ by means of one of the different versions of the generalized discrepancy principle, as explained in section 1.8.

Remark 2.5.4. For example, let us focus on the generalized discrepancy function (1.317), holding in the compatible case (analogous remarks can obviously be repeated for the incompatible case or for the mixed approach): in the current context, being $\delta = 0$, such a function can be written in the form

$$\rho_h^{\kappa_2}(\alpha; z_l) = \left\| \mathbf{F}_h \mathbf{g}_\alpha(z_l) - \frac{N}{2\pi} \mathbf{\Phi}_\infty(z_l) \right\|_{\mathbb{C}^N}^2 - h^2 \|\mathbf{g}_\alpha(z_l)\|_{\mathbb{C}^N}^2 - \left[\mu_h^{\kappa_2} \left(\frac{N}{2\pi} \mathbf{\Phi}_\infty(z_l), \mathbf{F}_h \right) \right]^2, \tag{2.279}$$

[31] Cf. definition 1.5.6.

[32] See remark 1.5.3.

[33] See representation (1.56).

where $\mu_h^{\kappa_2}\left(\frac{N}{2\pi}\mathbf{\Phi}_\infty(z_l), \mathbf{F}_h\right)$ is an approximate estimate of the *(simple) incompatibility measure* (see definition 1.8.1):

$$\mu_h\left(\frac{N}{2\pi}\mathbf{\Phi}_\infty(z_l), \mathbf{F}_h\right) = \inf_{\mathbf{g}(z_l)\in\mathbb{C}^N} \left\|\mathbf{F}_h\mathbf{g}(z_l) - \frac{N}{2\pi}\mathbf{\Phi}_\infty(z_l)\right\|_{\mathbb{C}^N}, \tag{2.280}$$

in such a way that (cf. inequalities (1.315))

$$\mu_h\left(\frac{N}{2\pi}\mathbf{\Phi}_\infty(z_l), \mathbf{F}_h\right) \leq \mu_h^{\kappa_2}\left(\frac{N}{2\pi}\mathbf{\Phi}_\infty(z_l), \mathbf{F}_h\right) \leq \mu_h\left(\frac{N}{2\pi}\mathbf{\Phi}_\infty(z_l), \mathbf{F}_h\right) + \kappa_2(h), \tag{2.281}$$

and $\lim_{h\to 0^+}\kappa_2(h) = 0$. Then, the optimal regularization parameter $\alpha^*(z_l) \equiv \alpha_2^*(z_l)$ must be fixed, in general, L times (one for each grid point z_l) by imposing that $\rho_h^{\kappa_2}(\alpha_2^*(z_l); z_l) = 0$.

Of course, according to the generalized discrepancy principle (for the compatible case), as formulated in subsection 1.8.3 from equation (1.319) to inequality (1.321), the problem of finding (the unique) $\alpha_2^*(z_l)$ such that $\rho_h^{\kappa_2}(\alpha_2^*(z_l); z_l) = 0$ is to be faced if and only if condition (1.321) is satisfied, i.e., in our case (being $\delta = 0$), if and only if it holds:

$$\left\|\frac{N}{2\pi}\mathbf{\Phi}_\infty(z_l)\right\|_{\mathbb{C}^N} > \mu_h^{\kappa_2}\left(\frac{N}{2\pi}\mathbf{\Phi}_\infty(z_l), \mathbf{F}_h\right); \tag{2.282}$$

otherwise, the selected approximation of the generalized solution of equation (2.272) is simply zero. □

Coming back to the implementation of the linear sampling method, we point out that the latter, in general, visualizes the scatterer profile by plotting the value of $I\left(\|\mathbf{g}_{\alpha^*(z_l)}(z_l)\|_{\mathbb{C}^N}\right)$ for each $z_l \in \mathcal{Z}$, where $I : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ is a suitable monotonic continuous function. In other terms, the indicator function is defined as

$$\begin{aligned} &\Psi_I : \mathcal{Z} \longrightarrow \mathbb{R} \\ &\qquad z_l \longmapsto \Psi_I(z_l) := I\left(\|\mathbf{g}_{\alpha^*(z_l)}(z_l)\|_{\mathbb{C}^N}\right). \end{aligned} \tag{2.283}$$

Of course, a three-dimensional plot of $\Psi_I(z_l)$ is only possible for a two-dimensional scatterer; if the latter is three-dimensional, one can only represent a three-dimensional section of the four-dimensional plot of $\Psi_I(z_l)$. However, both for two-dimensional and three-dimensional scatterers, the following problem arises: once the indicator function is available, which general criterion can suggest the thresholding level for its values? Or, in other terms, what can be considered "large" or "small" for the indicator function, respectively depending on the increasing or decreasing monotonicity of I? This is actually one of the open issues in the implementation of the linear sampling method[34]. A heuristic answer to such a question is suggested in [28] together with some numerical validations: we can summarize this approach as follows.

[34]We shall briefly present some of the open problems concerning the linear sampling method at the end of the current section.

Let us consider a certain scattering experiment to be performed: the scatterer, the wavelength, the number of incidence/observation angles, the noise level and the boundary conditions are then fixed. Now, let us substitute the original scatterer with a known disk (or, in the three-dimensional case, a known sphere), keeping unaltered all the other physical parameters, and compute the corresponding indicator function $\Psi_I^{(d)}(z_l)$ (where the superscript (d) reminds one of the disk case), having chosen a suitable I. The next step is to determine the cut-off section for the plot of the indicator function $\Psi_I^{(d)}(z_l)$ in such a way that the profile so obtained is as similar as possible to the true circular one: this procedure determines a certain height a at which the plot of $\Psi_I^{(d)}(z_l)$ is sectioned. Finally, when the original (non-circular) scatterer is considered and the corresponding indicator function $\Psi_I^{(s)}(z_l)$ (where the superscript (s) now reminds one of the original scatterer) is computed (for the same I as before), we decide to choose the same height a as thresholding level for the plot of $\Psi_I^{(s)}(z_l)$.

However, in the following we shall address the problem of the choice of the cut-off level for the indicator function by explicitly taking into account the knowledge of the scatterers to be reconstructed: in other terms, in all the numerical experiments we shall present (except the ones of section 3.5), the selected visualization profile will be given by the level curve of the indicator function containing an area equal to the one contained by the theoretical profile; on the other hand, in section 3.5 itself, a promising approach to the cut-off problem will be outlined: as we shall see, it is based on the numerical implementation of deformable contour models.

Some figures (with their respective captions) illustrating the implementation of the linear sampling method, as described just above, have been collected in section B.1. In the latter we present the reconstruction of three different impenetrable scatterers in the case of Dirichlet boundary conditions, with a wavenumber $k = 1$ and for two values of the noise level n, i.e. 1% and 10%, by using the generalized discrepancy principle in the compatible case (cf. subsection 1.8.3 and, in particular, definition (1.317) as well as its specific form (2.279) for the current context) and by choosing

$$\begin{aligned} \Psi_{-2\ln} : \mathcal{Z} &\longrightarrow \mathbb{R} \\ z_l &\longmapsto \Psi_{-2\ln}(z_l) := -\ln \left\| \mathbf{g}_{\alpha_2^*(z_l)}(z_l) \right\|_{\mathbb{C}^N}^2 \end{aligned} \tag{2.284}$$

as indicator function (cf. definition (2.283) with $I(t) := -\ln t^2 \equiv -2\ln t$).

Figure B.1 is concerned with an ellipse having its centre in $(0, 0)$ and semiaxes equal to 1 and 2 respectively, figure B.2 considers the case of a kite described by the parametric equation

$$x_1(t) = 1.5 \cdot \sin t, \qquad x_2(t) = \cos t + 0.65 \cdot \cos(2t) - 0.65, \qquad t \in [0, 2\pi], \tag{2.285}$$

while figure B.3 involves a double scatterer, i.e. the same kite described by equation (2.285) but centred in $(-4, -4)$ and rotated of 45° clockwise, together with an ellipse again with semiaxes

1 and 2 but centred in $(4, 4)$ and rotated of 45° counterclockwise. More details can be found in the respective captions.

We conclude the current section by observing that some crucial points concerning both the mathematical foundation and the numerical implementation of the linear sampling method still need to be completely understood. More precisely, as concerns a mainly theoretical viewpoint, the state-of-the-art for the general theorem presents some open problems [16] of notable mathematical interest: a typical example is the determination of the formal connections between the Tikhonov regularized solution of the (angle-discretized) far-field equation and the approximate solution whose properties are described by the general theorem (this connection is discussed in [4] for certain scalar cases); on the other hand, there are some general issues concerning the implementation and performances of the linear sampling method which, if solved, could greatly improve the effectiveness of the method and widen its applicability in the field of inverse scattering problems of interest in applied sciences. Summing up, four of these issues (some of which already introduced above) are touched by the following questions:

(i) is there a criterion suggesting how to choose the parameters of an "optimal" grid containing the scatterer (i.e. number of points and sampling distance)?

(ii) is it possible to give a characterization of the indicator function in terms of its physical meaning or analytical properties?

(iii) which is the spatial resolution power achievable by means of the linear sampling method?

(iiii) once the visualization map, i.e. the indicator function, is available, which general criterion can suggest the thresholding level for its values? Or, in other terms, what can be considered "large" or "small" for the indicator function, respectively depending on the increasing or decreasing monotonicity of the function I?

As far as we know, item (i) has never been considered by anyone. Interesting results concerning item (ii) are in [22], where a physical interpretation of the Euclidean norm of the optimal regularized solution is given. As far as item (iii) is concerned, we can mention [62], in which an interesting discussion of super-resolution related to the factorization method [44] (considered as a modified version of the linear sampling method) is proposed. Finally, we have already mentioned [28] for its contribution in treating item (iiii).

In the next chapter we shall deal with all four questions (i), (ii), (iii) and (iiii): more precisely, we are going to present a new implementation of the linear sampling method in which the set of the angle-discretized far-field equations for all sampling points is replaced by a single functional equation formulated in a Hilbert space defined as a direct sum of L^2 spaces, so that the problem of choicing a grid of sampling points is removed; then the squared norm of the regularized solution of such equation is used as indicator function and is analytically determined together with its Fourier transform: this provides some theoretical hints about the

spatial resolution achievable by the method. Finally, as just announced above, we shall try to face the cut-off problem by using deformable contour models.

CHAPTER 3

The linear sampling method without sampling

Following [3], in this chapter we present a new (no-sampling) implementation of the linear sampling method whereby the regularization parameter does not depend any longer on the sampling point and an analytical representation for any indicator function is therefore possible without any sampling in the space. Then, for sake of simplicity, we choose a particular indicator function whose analytical expression allows one to show that it is band-limited and, consequently, to obtain some theoretical information about the spatial resolution achievable by the method. Finally, we propose two further applications of our no-sampling implementation: we discuss the possibility of using a different family of indicator functions (with no apparent improvement in reconstruction accuracy) and we outline the technique of deformable contour models in order to face the problem of finding a suitable cut-off value for the visualization maps provided by the linear sampling method.

3.1. A new implementation of the linear sampling method

As we have seen in section 2.5, the indicator function (2.283) is known only on the grid $\mathcal{Z}$. On the other hand, the knowledge of its analytic form on $\mathbb{R}^2$ or, better, over a rectangle $T_A^B := (-A, A) \times (-B, B) \subset \mathbb{R}^2$ containing the scatterer (i.e. such that $T_A^B \supset \bar{D}$), would open new perspectives on both the computational effectiveness of the method and the quantitative assessment of its performances in terms of spatial resolution, as we shall see in the following. Then we are interested in regarding expression (2.283) as a sampled version of a function $\Psi_I(z)$ defined over T_A^B; nevertheless, this is not at all a straightforward task, since, once the monotonic function I is chosen, the dependence of $\Psi_I(z_l)$ (or, equivalently, of $\|\mathbf{g}_{\alpha^*(z_l)}(z_l)\|_{\mathbb{C}^N}$) on z_l, and therefore on any z, is explicit for $\mathbf{\Phi}_\infty(z_l)$ (cf. (2.278)), but only implicit, and in general not

known explicitly, for $\alpha^*(z_l)$.

In order to overcome this drawback, in the present section we derive a new implementation of the algorithm, again based on the general theorem 2.4.10, whereby the optimal value of the regularization parameter does not depend on $z \in T_A^B$. The starting point is to replace the finite set of equations (2.272) by an infinite set of equations

$$\mathbf{F}_h \mathbf{g}(z) = \frac{N}{2\pi} \mathbf{\Phi}_\infty(z) \qquad \forall z \in T_A^B. \tag{3.1}$$

In this framework T_A^B can be regarded as a continuous grid whereby the generic sampling point z_l has become a continuous variable $z \in T_A^B$. Then we want to modify the approach to the method from the pointwise algebraic setting represented in equation (3.1) to a unifying functional context, whereby regularization consists of a single procedure, which gives rise to a single value of the regularization parameter.

This result can be accomplished within the following mathematical framework. Let us consider the direct sum of Hilbert spaces:

$$\left[L^2(T_A^B)\right]^N := \underbrace{L^2(T_A^B) \oplus \ldots \oplus L^2(T_A^B)}_{N \text{ times}}, \tag{3.2}$$

where $L^2(T_A^B)$ denotes the usual set of Lebesgue square-integrable functions defined for almost all (f.a.a.) $z \in T_A^B$ and with values in $\mathbb{C}$. It is convenient to adopt here a notation which is slightly different from that of section A.6: more precisely, we shall denote with $f(\cdot)$ or $g(\cdot)$ a generic element of $L^2(T_A^B)$, with $(f(\cdot), g(\cdot))_2$ the scalar product of two functions $f(\cdot)$, $g(\cdot)$in $L^2(T_A^B)$ and with $\|f(\cdot)\|_2$ the norm of a function $f(\cdot)$ in $L^2(T_A^B)$.

Moreover, if we denote with $\mathbf{f}(\cdot) = \{f_i(\cdot)\}_{i=0}^{N-1}$ or $\mathbf{g}(\cdot) = \{g_i(\cdot)\}_{i=0}^{N-1}$ two generic elements of $\left[L^2(T_A^B)\right]^N$, then[1] the latter is a Hilbert space equipped with the scalar product

$$(\mathbf{f}(\cdot), \mathbf{g}(\cdot))_{2,N} := \sum_{i=0}^{N-1} (f_i(\cdot), g_i(\cdot))_2 \qquad \forall \mathbf{f}(\cdot), \mathbf{g}(\cdot) \in \left[L^2(T_A^B)\right]^N, \tag{3.3}$$

and the induced norm $\|\mathbf{f}(\cdot)\|_{2,N} := \sqrt{(\mathbf{f}(\cdot), \mathbf{f}(\cdot))_{2,N}}$, which is easily seen to be given by:

$$\|\mathbf{f}(\cdot)\|_{2,N} = \sqrt{\int_{T_A^B} \|\mathbf{f}(z)\|_{\mathbb{C}^N}^2 dz}. \tag{3.4}$$

Remark 3.1.1. For future purpose, we can immediately prove that:

$$\|\mathbf{g}(\cdot)\|_{2,N} \leq 1 \;\Rightarrow\; \|g_i(\cdot)\|_2 \leq 1 \qquad \forall i = 0, \ldots, N-1. \tag{3.5}$$

[1] See section A.1

Indeed, it holds:

$$\|\mathbf{g}(\cdot)\|_{2,N} = \sqrt{\sum_{i=0}^{N-1} (g_i(\cdot), g_i(\cdot))_2} = \sqrt{\sum_{i=0}^{N-1} \|g_i(\cdot)\|_2^2}, \tag{3.6}$$

whence implication (3.5) follows. □

We now need to define an operator acting on N-tuples of functions in order to replace the far-field matrix acting on N-tuples of numbers.

Definition 3.1.1. *The linear operator* $\mathsf{F}_h : \left[L^2(T_A^B)\right]^N \to \left[L^2(T_A^B)\right]^N$ *is given by*

$$\left[\mathsf{F}_h \mathbf{g}(\cdot)\right](\cdot) := \left\{ \sum_{j=0}^{N-1} (\mathbf{F}_h)_{ij}\, g_j(\cdot) \right\}_{i=0}^{N-1} \qquad \forall \mathbf{g}(\cdot) \in \left[L^2(T_A^B)\right]^N, \tag{3.7}$$

where the $(\mathbf{F}_h)_{ij}$ *are the elements of the noisy far-field matrix.*

The next step is to study some properties of F_h, in order to be able to work with it. This task is accomplished by the following results, stated from theorem 3.1.1 to theorem 3.1.5.

Theorem 3.1.1. *The linear operator* F_h *is bounded.*

Proof. Remembering that one of the possible definitions of the norm of a linear continuous operator $\mathsf{T} : X \to Y$ from a normed space X to a normed space Y is (cf. remark 1.3.1)

$$\|\mathsf{T}\| := \sup_{\|x\|_X = 1} \|\mathsf{T}(x)\|_Y, \tag{3.8}$$

let $\mathbf{g}(\cdot)$ be any element of $\left[L^2(T_A^B)\right]^N$ such that $\|\mathbf{g}(\cdot)\|_{2,N} = 1$. Then, for such a $\mathbf{g}(\cdot) \in \left[L^2(T_A^B)\right]^N$, we consider $\|[\mathsf{F}_h \mathbf{g}(\cdot)](\cdot)\|_{2,N}$ and the consequent chain of equalities or inequalities (obtained using most of the previous definitions, the fact that $\|\mathbf{g}(\cdot)\|_{2,N} = 1$, property (3.5) and the Cauchy-Schwarz inequality):

$$\begin{aligned}
&\|[\mathsf{F}_h \mathbf{g}(\cdot)](\cdot)\|_{2,N} = \\
&= \sqrt{([\mathsf{F}_h \mathbf{g}(\cdot)](\cdot), [\mathsf{F}_h \mathbf{g}(\cdot)](\cdot))_{2,N}} = \sqrt{\left(\left\{ \sum_{j=0}^{N-1} (\mathbf{F}_h)_{ij}\, g_j(\cdot) \right\}_{i=0}^{N-1}, \left\{ \sum_{p=0}^{N-1} (\mathbf{F}_h)_{ip}\, g_p(\cdot) \right\}_{i=0}^{N-1} \right)_{2,N}} = \\
&= \sqrt{\sum_{i=0}^{N-1} \left(\sum_{j=0}^{N-1} (\mathbf{F}_h)_{ij}\, g_j(\cdot), \sum_{p=0}^{N-1} (\mathbf{F}_h)_{ip}\, g_p(\cdot) \right)_2} = \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sum_{p=0}^{N-1} (\mathbf{F}_h)_{ij} (\bar{\mathbf{F}}_h)_{ip} \left(g_j(\cdot), g_p(\cdot)\right)_2} = \\
&= \sqrt{\sum_{i=0}^{N-1} \sum_{p=j=0}^{N-1} (\mathbf{F}_h)_{ij} (\bar{\mathbf{F}}_h)_{ij} \left(g_j(\cdot), g_j(\cdot)\right)_2 + \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sum_{j \neq p=0}^{N-1} (\mathbf{F}_h)_{ij} (\bar{\mathbf{F}}_h)_{ip} \left(g_j(\cdot), g_p(\cdot)\right)_2} \leq \\
&\leq \sqrt{\sum_{i=0}^{N-1} \max_{i,j} |(\mathbf{F}_h)_{ij}|^2 \sum_{j=0}^{N-1} \left(g_j(\cdot), g_j(\cdot)\right)_2 + \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sum_{j \neq p=0}^{N-1} |(\mathbf{F}_h)_{ij} (\bar{\mathbf{F}}_h)_{ip}| \, |\left(g_j(\cdot), g_p(\cdot)\right)_2|} \leq
\end{aligned}$$

$$\leq \sqrt{N \max_{i,j} |(\mathbf{F}_h)_{ij}|^2 \cdot 1 + N \max_{i,j} |(\mathbf{F}_h)_{ij}|^2 \sum_{j=0}^{N-1} \sum_{j \neq p=0}^{N-1} \|g_j(\cdot)\|_2 \, \|g_p(\cdot)\|_2} \leq$$
$$\leq \sqrt{N \max_{i,j} |(\mathbf{F}_h)_{ij}|^2 + N^2(N-1) \max_{i,j} |(\mathbf{F}_h)_{ij}|^2} =$$
$$\leq \max_{i,j} |(\mathbf{F}_h)_{ij}| \sqrt{(N^3 - N^2 + N)} =: \mathcal{M}. \tag{3.9}$$

Hence the linear operator $\mathsf{F}_h : \left[L^2(T_A^B)\right]^N \to \left[L^2(T_A^B)\right]^N$ is bounded and

$$\|\mathsf{F}_h\| = \sup_{\|\mathbf{g}(\cdot)\|_{2,N}=1} \|[\mathsf{F}_h \mathbf{g}(\cdot)](\cdot)\|_{2,N} \leq \mathcal{M}. \tag{3.10}$$

This concludes the proof. ■

Theorem 3.1.2. *Denoting, as usual, with $\mathcal{N}(\mathsf{F}_h)$ and $\mathcal{N}(\mathbf{F}_h)$ the kernels of the linear operators F_h and $\mathbf{F}_h$ respectively, it holds:*

$$\mathcal{N}(\mathsf{F}_h) = \left\{ \mathbf{f}(\cdot) \in \left[L^2(T_A^B)\right]^N \mid \mathbf{f}(z) \in \mathcal{N}(\mathbf{F}_h) \text{ f.a.a. } z \in T_A^B \right\}. \tag{3.11}$$

Furthermore, if $\mathbf{g}(\cdot) \in \left[L^2(T_A^B)\right]^N$ is such that $\mathbf{g}(z) \in \mathcal{N}(\mathbf{F}_h)^\perp$ f.a.a. $z \in T_A^B$, then $\mathbf{g}(\cdot) \in \mathcal{N}(\mathsf{F}_h)^\perp$, where the orthogonality must be intended with respect to the corresponding scalar product[2]*.*

Proof. The characterization (3.11) is obvious from definition (3.7) itself.

Now, let us take any $\mathbf{f}(\cdot) \in \mathcal{N}(\mathsf{F}_h)$: by virtue of relation (3.11), we have that $\mathbf{f}(z) \in \mathcal{N}(\mathbf{F}_h)$ f.a.a. $z \in T_A^B$; on the other hand, if $\mathbf{g}(\cdot) \in \left[L^2(T_A^B)\right]^N$ is such that $\mathbf{g}(z) \in \mathcal{N}(\mathbf{F}_h)^\perp$ f.a.a. $z \in T_A^B$, then it holds (remembering definition (3.3))

$$(\mathbf{f}(\cdot), \mathbf{g}(\cdot))_{2,N} = \int_{T_A^B} \sum_{i=0}^{N-1} f_i(z) \bar{g}_i(z) dz = \int_{T_A^B} (\mathbf{f}(z), \mathbf{g}(z))_{\mathbb{C}^N} dz = 0, \tag{3.12}$$

since $(\mathbf{f}(z), \mathbf{g}(z))_{\mathbb{C}^N} = 0$ f.a.a. $z \in T_A^B$. This concludes the proof. ■

Theorem 3.1.3. *The linear operator F_h is not compact.*

Proof. Let us consider a countable infinite orthonormal set $\{e_i(\cdot)\}_{i=0}^{\infty}$ of $L^2(T_A^B)$ and then, for each $n \in \mathbb{N}$, define the following element of $\left[L^2(T_A^B)\right]^N$:

$$\mathbf{b}_n(\cdot) := \frac{1}{\sqrt{N}} \{e_{i+nN}(\cdot)\}_{i=0}^{N-1}. \tag{3.13}$$

[2] That is $(\cdot,\cdot)_{\mathbb{C}^N}$ for $\mathcal{N}(\mathbf{F}_h)^\perp$ and $(\cdot,\cdot)_{2,N}$ for $\mathcal{N}(\mathsf{F}_h)^\perp$.

Clearly, the set $B := \{\mathbf{b}_n(\cdot)\}_{n=0}^{\infty} \subset \left[L^2(T_A^B)\right]^N$ is bounded and, more precisely, orthonormal: indeed, remembering definitions (3.3) and (3.13), we can write:

$$(\mathbf{b}_n(\cdot), \mathbf{b}_m(\cdot))_{2,N} = \frac{1}{N}\sum_{i=0}^{N-1}(e_{i+nN}(\cdot), e_{i+mN}(\cdot))_2 = \frac{1}{N}\sum_{i=0}^{N-1}\delta_{nm} = \frac{1}{N}N\delta_{nm} = \delta_{nm}. \tag{3.14}$$

Now we show that the set $\mathsf{F}_h(B)$, which is just the sequence $\{[\mathsf{F}_h\mathbf{b}_n(\cdot)](\cdot)\}_{n=0}^{\infty}$, is not relatively compact in $\left[L^2(T_A^B)\right]^N$, since it is impossible to extract by it a subsequence verifying the Cauchy criterion. Indeed, let us consider the following chain of equalities:

$$\begin{aligned}
&\|[\mathsf{F}_h\mathbf{b}_n(\cdot)](\cdot) - [\mathsf{F}_h\mathbf{b}_m(\cdot)](\cdot)\|_{2,N}^2 = \\
&= \left\| \left\{ \sum_{j=0}^{N-1} (\mathbf{F}_h)_{ij} \frac{e_{j+nN}(\cdot)}{\sqrt{N}} \right\}_{i=0}^{N-1} - \left\{ \sum_{j=0}^{N-1} (\mathbf{F}_h)_{ij} \frac{e_{j+mN}(\cdot)}{\sqrt{N}} \right\}_{i=0}^{N-1} \right\|_{2,N}^2 = \\
&= \left\| \frac{1}{\sqrt{N}} \left\{ \sum_{j=0}^{N-1} (\mathbf{F}_h)_{ij} \left[e_{j+nN}(\cdot) - e_{j+mN}(\cdot)\right] \right\}_{i=0}^{N-1} \right\|_{2,N}^2 = \\
&= \frac{1}{N} \left(\left\{ \sum_{j=0}^{N-1} (\mathbf{F}_h)_{ij} \left[e_{j+nN}(\cdot) - e_{j+mN}(\cdot)\right] \right\}_{i=0}^{N-1}, \left\{ \sum_{j=0}^{N-1} (\mathbf{F}_h)_{ij} \left[e_{j+nN}(\cdot) - e_{j+mN}(\cdot)\right] \right\}_{i=0}^{N-1} \right)_{2,N} = \\
&= \frac{1}{N} \sum_{i=0}^{N-1} \left(\sum_{j=0}^{N-1} (\mathbf{F}_h)_{ij} \left[e_{j+nN}(\cdot) - e_{j+mN}(\cdot)\right], \sum_{p=0}^{N-1} (\mathbf{F}_h)_{ip} \left[e_{p+nN}(\cdot) - e_{p+mN}(\cdot)\right] \right)_2 = \\
&= \frac{1}{N} \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sum_{p=0}^{N-1} (\mathbf{F}_h)_{ij}(\bar{\mathbf{F}}_h)_{ip} \left(e_{j+nN}(\cdot) - e_{j+mN}(\cdot),\, e_{p+nN}(\cdot) - e_{p+mN}(\cdot)\right)_2 = \\
&= \frac{1}{N} \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sum_{p=0}^{N-1} (\mathbf{F}_h)_{ij}(\bar{\mathbf{F}}_h)_{ip} \left[(e_{j+nN}(\cdot), e_{p+nN}(\cdot))_2 - (e_{j+nN}(\cdot), e_{p+mN}(\cdot))_2 + \right. \\
&\qquad\qquad \left. - (e_{j+mN}(\cdot), e_{p+nN}(\cdot))_2 + (e_{j+mN}(\cdot), e_{p+mN}(\cdot))_2\right] = \\
&= \frac{1}{N} \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sum_{p=0}^{N-1} (\mathbf{F}_h)_{ij}(\bar{\mathbf{F}}_h)_{ip} \left(\delta_{jp} - 2\delta_{nm}\delta_{jp} + \delta_{jp}\right) = \\
&= \frac{1}{N} \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sum_{p=0}^{N-1} (\mathbf{F}_h)_{ij}(\bar{\mathbf{F}}_h)_{ip} \left(1 - 2\delta_{nm} + 1\right)\delta_{jp} = \frac{2}{N} \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sum_{p=0}^{N-1} (\mathbf{F}_h)_{ij}(\bar{\mathbf{F}}_h)_{ip}\,\delta_{jp}(1 - \delta_{nm}).
\end{aligned} \tag{3.15}$$

Hence, if $n = m$, we obviously have that $\|[\mathsf{F}_h\mathbf{b}_n(\cdot)](\cdot) - [\mathsf{F}_h\mathbf{b}_m(\cdot)](\cdot)\|_{2,N}^2 = 0$; otherwise, we

get:

$$\|[\mathsf{F}_h\mathbf{b}_n(\cdot)](\cdot) - [\mathsf{F}_h\mathbf{b}_m(\cdot)](\cdot)\|^2_{2,N} = \frac{2}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sum_{p=0}^{N-1}(\mathbf{F}_h)_{ij}(\bar{\mathbf{F}}_h)_{ip}\,\delta_{jp} = \frac{2}{N}\sum_{i=0}^{N-1}\sum_{p=0}^{N-1}(\mathbf{F}_h)_{ip}(\bar{\mathbf{F}}_h)_{ip} =$$
$$= \frac{2}{N}\sum_{i=0}^{N-1}\sum_{p=0}^{N-1}|(\mathbf{F}_h)_{ip}|^2. \tag{3.16}$$

Of course, the last quantity is strictly positive and independent of n, m (with $n \neq m$); hence, the sequence $\{[\mathsf{F}_h\mathbf{b}_n(\cdot)](\cdot)\}_{n=0}^{\infty}$ cannot verify the Cauchy criterion. ■

Remark 3.1.2. The non-compactness of the operator F_h is not in contradiction with the well-known compactness of the usual far-field operator F defined in (2.97) (see remark 2.3.1): indeed F_h acts upon N-tuples of functions of the spatial variable z, while F acts upon functions of the angular variable θ. □

Definition 3.1.2. *For any* $\mathbf{g}(\cdot) \in \left[L^2(T_A^B)\right]^N$ *and any* $\mathbf{w} \in \mathbb{C}^N$ *(with components* w_i*, for* $i = 0, \ldots, N-1$*), we define (with a little notational misuse) the following element of* $L^2(T_A^B)$*:*

$$\begin{aligned}(\mathbf{g}(\cdot), \mathbf{w})_{\mathbb{C}^N} : T_A^B &\longrightarrow \mathbb{C} \\ z &\longmapsto (\mathbf{g}(z), \mathbf{w})_{\mathbb{C}^N} \qquad \text{f.a.a. } z \in T_A^B,\end{aligned} \tag{3.17}$$

or equivalently, in components:

$$(\mathbf{g}(\cdot), \mathbf{w})_{\mathbb{C}^N} := \sum_{i=0}^{N-1} \bar{w}_i\, g_i(\cdot). \tag{3.18}$$

Remark 3.1.3. Remembering the singular representation (2.274) of the far-field matrix $\mathbf{F}_h$, the definition (3.7) of F_h and using the notation (3.18), we easily realize that the following representation for F_h holds:

$$[\mathsf{F}_h\mathbf{g}(\cdot)](\cdot) = \left\{\sum_{p=0}^{r-1}\sigma_p^h\, v_{p,i}^h\, (\mathbf{g}(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N}\right\}_{i=0}^{N-1} \qquad \forall \mathbf{g}(\cdot) \in \left[L^2(T_A^B)\right]^N, \tag{3.19}$$

where $v_{p,i}^h$ is the i-th component of the vector $\mathbf{v}_p^h \in \mathbb{C}^N$. □

By virtue of the previous relation (3.19), we can prove the following result.

Theorem 3.1.4. *It holds:*

$$\|\mathsf{F}_h\| = \sigma_0^h. \tag{3.20}$$

Proof. As in the proof of theorem 3.1.1, let $\mathbf{g}(\cdot)$ be any element of $\left[L^2(T_A^B)\right]^N$ such that $\|\mathbf{g}(\cdot)\|_{2,N} = 1$. Then, for such a $\mathbf{g}(\cdot) \in \left[L^2(T_A^B)\right]^N$, we consider $\|[\mathsf{F}_h\mathbf{g}(\cdot)](\cdot)\|_{2,N}$: by using most of the previous definitions, the orthonormality of vectors $\mathbf{v}_p^h$, the Cauchy-Schwarz inequality, relation (3.4) and the fact that $\|\mathbf{g}(\cdot)\|_{2,N} = 1$, we get:

$$
\begin{aligned}
\|[\mathsf{F}_h\mathbf{g}(\cdot)](\cdot)\|_{2,N} &= \sqrt{(\mathsf{F}_h\mathbf{g}(\cdot), \mathsf{F}_h\mathbf{g}(\cdot))_{2,N}} = \\
&= \sqrt{\left(\left\{ \sum_{p=0}^{r-1} \sigma_p^h \, v_{p,i}^h \, (\mathbf{g}(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} \right\}_{i=0}^{N-1}, \ \left\{ \sum_{q=0}^{r-1} \sigma_q^h \, v_{q,i}^h \, (\mathbf{g}(\cdot), \mathbf{u}_q^h)_{\mathbb{C}^N} \right\}_{i=0}^{N-1} \right)_{2,N}} = \\
&= \sqrt{\sum_{i=0}^{N-1} \left(\sum_{p=0}^{r-1} \sigma_p^h \, v_{p,i}^h \, (\mathbf{g}(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N}, \ \sum_{q=0}^{r-1} \sigma_q^h \, v_{q,i}^h \, (\mathbf{g}(\cdot), \mathbf{u}_q^h)_{\mathbb{C}^N} \right)_2} = \\
&= \sqrt{\sum_{i=0}^{N-1} \sum_{p=0}^{r-1} \sum_{q=0}^{r-1} \sigma_p^h \, v_{p,i}^h \, \sigma_q^h \, \bar{v}_{q,i}^h \, \left((\mathbf{g}(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N}, \ (\mathbf{g}(\cdot), \mathbf{u}_q^h)_{\mathbb{C}^N} \right)_2} = \\
&= \sqrt{\sum_{p=0}^{r-1} \sum_{q=0}^{r-1} \sum_{i=0}^{N-1} (\mathbf{v}_p^h, \mathbf{v}_q^h)_{\mathbb{C}^N} \, \sigma_p^h \, \sigma_q^h \, \left((\mathbf{g}(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N}, \ (\mathbf{g}(\cdot), \mathbf{u}_q^h)_{\mathbb{C}^N} \right)_2} = \\
&= \sqrt{\sum_{p=0}^{r-1} \sum_{q=0}^{r-1} \delta_{pq} \, \sigma_p^h \, \sigma_q^h \, \left((\mathbf{g}(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N}, \ (\mathbf{g}(\cdot), \mathbf{u}_q^h)_{\mathbb{C}^N} \right)_2} = \\
&= \sqrt{\sum_{p=0}^{r-1} (\sigma_p^h)^2 \, \left((\mathbf{g}(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N}, \ (\mathbf{g}(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} \right)_2} \leq \qquad (3.21) \\
&\leq \sigma_0^h \sqrt{\sum_{p=0}^{r-1} \int_{T_A^B} \left| (\mathbf{g}(z), \mathbf{u}_p^h)_{\mathbb{C}^N} \right|^2 dz} = \sigma_0^h \sqrt{\int_{T_A^B} \sum_{p=0}^{r-1} \left| (\mathbf{g}(z), \mathbf{u}_p^h)_{\mathbb{C}^N} \right|^2 dz} \leq \\
&\leq \sigma_0^h \sqrt{\int_{T_A^B} \|\mathbf{g}(z)\|_{\mathbb{C}^N}^2 dz} = \sigma_0^h \, \|\mathbf{g}(\cdot)\|_{2,N} = \sigma_0^h, \qquad (3.22)
\end{aligned}
$$

where equality always holds in all the previous passages if, e.g., we choose $\mathbf{g}(\cdot) \in \left[L^2(T_A^B)\right]^N$ such that it is a constant $\mathbb{C}^N$-valued function equal to $\frac{\mathbf{u}_0^h}{\sqrt{2A \times 2B}}$. ■

Remark 3.1.4. Since also the norm of the matrix $\mathbf{F}_h$, regarded as an operator from $\mathbb{C}^N$ to $\mathbb{C}^N$, is equal to σ_0^h (see relation (2.275) in remark 2.5.3), the previous result (3.20) states that $\|\mathbf{F}_h\| = \|\mathsf{F}_h\|$. Now, if we remember relation (2.266) and define, in terms of the matrices $\mathbf{F}$, $\mathbf{H}$ respectively, the linear continuous operators F, H in the same way followed to define the linear continuous operator F_h in terms of the matrix $\mathbf{F}_h$, we easily realize that

$$
\mathbf{F}_h = \mathbf{F} + \mathbf{H} \implies \mathsf{F}_h = \mathsf{F} + \mathsf{H}. \qquad (3.23)
$$

By means of an argument fully analogous to that used to prove relation (3.20), from the previous implication (3.23) we easily obtain that

$$\|\mathbf{F}_h - \mathbf{F}\| = \|\mathbf{H}\| = \|\mathsf{H}\| = \|\mathsf{F}_h - \mathsf{F}\| \tag{3.24}$$

and hence

$$\|\mathbf{F}_h - \mathbf{F}\| \leq h \implies \|\mathsf{F}_h - \mathsf{F}\| \leq h. \tag{3.25}$$

This means that the bound h on the noise affecting the matrix $\mathbf{F}_h$ is the same as the one affecting the operator F_h. □

Theorem 3.1.5. *The range $\mathcal{R}(\mathsf{F}_h)$ of the operator F_h is closed.*

Proof. Let us consider any $\mathbf{h}(\cdot) \in \overline{\mathcal{R}(\mathsf{F}_h)} \setminus \mathcal{R}(\mathsf{F}_h)$; thus, there exists a sequence $\{\mathbf{g}_n(\cdot)\}_{n=0}^{\infty} \subset \left[L^2(T_A^B)\right]^N$ such that:

$$\lim_{n\to\infty} \|[\mathsf{F}_h \mathbf{g}_n(\cdot)](\cdot) - \mathbf{h}(\cdot)\|_{2,N} = 0; \tag{3.26}$$

we want to show that:

$$\exists \mathbf{t}(\cdot) \in \left[L^2(T_A^B)\right]^N \quad \text{such that} \quad [\mathsf{F}_h \mathbf{t}(\cdot)](\cdot) = \mathbf{h}(\cdot). \tag{3.27}$$

Firstly, we observe that, for a generic $\mathbf{g}(\cdot) \in \left[L^2(T_A^B)\right]^N$, it holds, by virtue of (3.21):

$$\begin{aligned} \|[\mathsf{F}_h \mathbf{g}(\cdot)](\cdot)\|_{2,N} &= \sqrt{\sum_{p=0}^{r-1} (\sigma_p^h)^2 \left((\mathbf{g}(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N},\ (\mathbf{g}(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} \right)_2} \geq \\ &\geq \sigma_{r-1}^h \sqrt{\sum_{p=0}^{r-1} \left((\mathbf{g}(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N},\ (\mathbf{g}(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} \right)_2} = \\ &= \sigma_{r-1}^h \sqrt{\sum_{p=0}^{r-1} \left\| (\mathbf{g}(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} \right\|_2^2}. \end{aligned} \tag{3.28}$$

Now we point out that $\{[\mathsf{F}_h \mathbf{g}_n(\cdot)](\cdot)\}_{n=0}^{\infty} \subset \left[L^2(T_A^B)\right]^N$ is a Cauchy sequence, since it converges by virtue of (3.26); then, $\forall \varepsilon > 0$, there exists $N > 0$ such that, for $n, m \geq N$, it holds:

$$\|[\mathsf{F}_h \mathbf{g}_n(\cdot)](\cdot) - \mathsf{F}_h \mathbf{g}_m(\cdot)](\cdot)\|_{2,N} < \varepsilon, \tag{3.29}$$

i.e.

$$\|[\mathsf{F}_h (\mathbf{g}_n(\cdot) - \mathbf{g}_m(\cdot))](\cdot)\|_{2,N} < \varepsilon. \tag{3.30}$$

Recalling relation (3.28), inequality (3.30) implies:

$$\sigma_{r-1}^h \sqrt{\sum_{p=0}^{r-1} \left\| (\mathbf{g}_n(\cdot) - \mathbf{g}_m(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} \right\|_2^2} < \varepsilon, \tag{3.31}$$

whence we immediately get:

$$\sum_{p=0}^{r-1} \left\| (\mathbf{g}_n(\cdot) - \mathbf{g}_m(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} \right\|_2^2 < \left(\frac{\varepsilon}{\sigma_{r-1}^h} \right)^2, \tag{3.32}$$

and then, putting $\varepsilon' := \dfrac{\varepsilon}{\sigma_{r-1}^h}$,

$$\left\| (\mathbf{g}_n(\cdot) - \mathbf{g}_m(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} \right\|_2 < \varepsilon' \qquad \forall p = 0, \ldots, r-1, \tag{3.33}$$

i.e.

$$\left\| (\mathbf{g}_n(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} - (\mathbf{g}_m(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} \right\|_2 < \varepsilon' \qquad \forall p = 0, \ldots, r-1. \tag{3.34}$$

This means that $\{(\mathbf{g}_n(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N}\}_{n=0}^{\infty} \subset L^2(T_A^B)$ is a Cauchy sequence $\forall p = 0, \ldots, r-1$ and then, by virtue of the completeness of $L^2(T_A^B)$, it converges to an element $f_p(\cdot) \in L^2(T_A^B)$. In other terms, $\forall p = 0, \ldots, r-1$, there exists $f_p(\cdot) \in L^2(T_A^B)$ such that

$$\lim_{n \to \infty} \left\| (\mathbf{g}_n(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} - f_p(\cdot) \right\|_2 = 0. \tag{3.35}$$

Now we define the following elements of $\left[L^2(T_A^B)\right]^N$:

$$\mathbf{t}_n(\cdot) := \sum_{p=0}^{r-1} (\mathbf{g}_n(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} \mathbf{u}_p^h \qquad \forall n \in \mathbb{N}; \tag{3.36}$$

$$\mathbf{t}(\cdot) := \sum_{p=0}^{r-1} f_p(\cdot) \mathbf{u}_p^h. \tag{3.37}$$

The next step is to show that:

$$\lim_{n \to \infty} \|\mathbf{t}_n(\cdot) - \mathbf{t}(\cdot)\|_{2,N}^2 = 0. \tag{3.38}$$

To this purpose, remembering relation (3.6), let us consider the following chain of equalities or inequalities:

$$\begin{aligned}
\|\mathbf{t}_n(\cdot) - \mathbf{t}(\cdot)\|_{2,N}^2 &= \sum_{i=0}^{N-1} \|t_{n,i}(\cdot) - t_i(\cdot)\|_2^2 = \sum_{i=0}^{N-1} \left\| \sum_{p=0}^{r-1} (\mathbf{g}_n(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} u_{p,i}^h - \sum_{p=0}^{r-1} f_p(\cdot) u_{p,i}^h \right\|_2^2 = \\
&= \sum_{i=0}^{N-1} \left\| \sum_{p=0}^{r-1} \left[(\mathbf{g}_n(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} - f_p(\cdot) \right] u_{p,i}^h \right\|_2^2 \leq \\
&\leq \sum_{i=0}^{N-1} \left\{ \sum_{p=0}^{r-1} \left\| \left[(\mathbf{g}_n(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} - f_p(\cdot) \right] u_{p,i}^h \right\|_2 \right\}^2 \leq \\
&\leq \sum_{i=0}^{N-1} \left\{ \sum_{p=0}^{r-1} |u_{p,i}^h| \left\| (\mathbf{g}_n(\cdot), \mathbf{u}_p^h)_{\mathbb{C}^N} - f_p(\cdot) \right\|_2 \right\}^2.
\end{aligned} \tag{3.39}$$

Recalling relation (3.35), inequality (3.39) implies limit (3.38), whence, by virtue of the continuity of F_h, we immediately get:

$$\lim_{n\to\infty} \|[\mathsf{F}_h \mathbf{t}_n(\cdot)](\cdot) - [\mathsf{F}_h \mathbf{t}(\cdot)](\cdot)\|_{2,N} = 0. \tag{3.40}$$

On the other hand, recalling representation (3.19) and the orthonormality of vectors $\mathbf{u}_p^h$, it is easy to realize that

$$[\mathsf{F}_h \mathbf{t}_n(\cdot)](\cdot) = [\mathsf{F}_h \mathbf{g}_n(\cdot)](\cdot), \tag{3.41}$$

so that relation (3.40) can also be written as

$$\lim_{n\to\infty} \|[\mathsf{F}_h \mathbf{g}_n(\cdot)](\cdot) - [\mathsf{F}_h \mathbf{t}(\cdot)](\cdot)\|_{2,N} = 0. \tag{3.42}$$

By comparison of the two limits (3.42) and (3.26), we finally get:

$$[\mathsf{F}_h \mathbf{t}(\cdot)](\cdot) = \mathbf{h}(\cdot), \tag{3.43}$$

which is just our thesis (3.27). ■

From a practical viewpoint, the introduction of the operator F_h allows one to express the infinitely many algebraic systems (3.1) as the single functional equation in $[L^2(T_A^B)]^N$

$$[\mathsf{F}_h \mathbf{g}(\cdot)](\cdot) = \frac{N}{2\pi} \mathbf{\Phi}_\infty(\cdot), \tag{3.44}$$

where $\mathbf{\Phi}_\infty(\cdot)$ is the element in $[L^2(T_A^B)]^N$ trivially obtained from $\mathbf{\Phi}_\infty(z)$ simply regarding z as a variable on T_A^B instead of a fixed point in $\mathbb{R}^2$.

Since, according to theorem 3.1.5, $\mathcal{R}(\mathsf{F}_h)$ is closed, then, by virtue of theorem 1.5.6, the generalized inverse operator $\mathsf{F}_h^\dagger$ is continuous, i.e. the problem of determining the generalized solution of the functional equation (3.44) is well-posed. Nevertheless, we know (see the last part of section 1.3) that such a problem, as an inverse one, is, in general, ill-conditioned and therefore a regularization method is anyway necessary. Hence, the novelty in comparison with the traditional implementation of the linear sampling method is rather the fact that now the regularization of equation (3.44) occurs in a way which is independent from z and therefore provides a single value of the regularization parameter[3].

At this stage there is a final computational open issue to address, which is concerned with how to determine the regularized solution of equation (3.44) in practice. But this problem is solved by the following theorem: indeed, it shows that, starting from the generalized and Tikhonov regularized[4] solutions of system (3.1), which for each $z \in T_A^B$ admit the representations[5]

$$\mathbf{g}_h^\dagger(z) = \frac{N}{2\pi} \sum_{p=0}^{r_h-1} \frac{(\mathbf{\Phi}_\infty(z), \mathbf{v}_p^h)_{\mathbb{C}^N}}{\sigma_p^h} \mathbf{u}_p^h \tag{3.45}$$

[3] We point out, however, that such a value will depend on the choice of the rectangle T_A^B.

[4] Computed for a generic (and, in particular, z-independent) value $\alpha \in \mathbb{R}^+$ of the regularization parameter.

[5] See relations (2.276) and (2.277).

and

$$\mathbf{g}_\alpha(z) = \frac{N}{2\pi}\sum_{p=0}^{r_h-1}\frac{\sigma_p^h}{(\sigma_p^h)^2+\alpha}(\mathbf{\Phi}_\infty(z),\mathbf{v}_p^h)_{\mathbb{C}^N}\mathbf{u}_p^h, \tag{3.46}$$

it is possible to determine, in the new functional context, the generalized and Tikhonov regularized[6] solutions of problem (3.44) simply regarding z as an independent variable.

Theorem 3.1.6. *The generalized and Tikhonov regularized solutions of problem (3.44) are respectively given by*

$$\mathbf{g}_h^\dagger(\cdot) = \frac{N}{2\pi}\sum_{p=0}^{r_h-1}\frac{(\mathbf{\Phi}_\infty(\cdot),\mathbf{v}_p^h)_{\mathbb{C}^N}}{\sigma_p^h}\mathbf{u}_p^h \tag{3.47}$$

and

$$\mathbf{g}_\alpha(\cdot) = \frac{N}{2\pi}\sum_{p=0}^{r_h-1}\frac{\sigma_p^h}{(\sigma_p^h)^2+\alpha}(\mathbf{\Phi}_\infty(\cdot),\mathbf{v}_p^h)_{\mathbb{C}^N}\mathbf{u}_p^h, \tag{3.48}$$

where α is a generic real positive number[7].

Proof. The generalized solution $\mathbf{g}_h^\dagger(z)$ of equation (3.1) is the unique least-squares solution in $\mathcal{N}(\mathbf{F}_h)^\perp$. Therefore, for any $\mathbf{g}(\cdot)\in\left[L^2(T_A^B)\right]^N$ and f.a.a. $z\in T_A^B$, it holds:

$$\left\|\mathbf{F}_h\mathbf{g}_h^\dagger(z)-\frac{N}{2\pi}\mathbf{\Phi}_\infty(z)\right\|_{\mathbb{C}^N}^2 \le \left\|\mathbf{F}_h\mathbf{g}(z)-\frac{N}{2\pi}\mathbf{\Phi}_\infty(z)\right\|_{\mathbb{C}^N}^2. \tag{3.49}$$

From (3.49), we immediately get:

$$\int_{T_A^B}\left\|\mathbf{F}_h\mathbf{g}_h^\dagger(z)-\frac{N}{2\pi}\mathbf{\Phi}_\infty(z)\right\|_{\mathbb{C}^N}^2 dz \le \int_{T_A^B}\left\|\mathbf{F}_h\mathbf{g}(z)-\frac{N}{2\pi}\mathbf{\Phi}_\infty(z)\right\|_{\mathbb{C}^N}^2 dz. \tag{3.50}$$

We now observe that, for any $\mathbf{g}(\cdot)\in[L^2(T_A^B)]^N$, definition (3.7) implies that

$$\mathbf{F}_h\mathbf{g}(z) = [\mathsf{F}_h\mathbf{g}(\cdot)](z); \tag{3.51}$$

therefore relation (3.50) can be written as

$$\int_{T_A^B}\left\|[\mathsf{F}_h\mathbf{g}_h^\dagger(\cdot)](z)-\frac{N}{2\pi}\mathbf{\Phi}_\infty(z)\right\|_{\mathbb{C}^N}^2 dz \le \int_{T_A^B}\left\|[\mathsf{F}_h\mathbf{g}(\cdot)](z)-\frac{N}{2\pi}\mathbf{\Phi}_\infty(z)\right\|_{\mathbb{C}^N}^2 dz, \tag{3.52}$$

where $\mathbf{g}_h^\dagger(\cdot)$ is the element in $[L^2(T_A^B)]^N$ obtained from $\mathbf{g}_h^\dagger(z)$, as given by (3.45), when z varies in T_A^B, i.e. $\mathbf{g}_h^\dagger(\cdot)$ has just the form (3.47). Then relation (3.4) leads to

$$\left\|[\mathsf{F}_h\mathbf{g}_h^\dagger(\cdot)](\cdot)-\frac{N}{2\pi}\mathbf{\Phi}_\infty(\cdot)\right\|_{2,N}^2 \le \left\|[\mathsf{F}_h\mathbf{g}(\cdot)](\cdot)-\frac{N}{2\pi}\mathbf{\Phi}_\infty(\cdot)\right\|_{2,N}^2. \tag{3.53}$$

[6] Again, computed for a generic and *obviously* z-independent value $\alpha\in\mathbb{R}^+$ of the regularization parameter.

[7] The problem of choosing a suitable value α^* for the regularization parameter α will be addressed soon below the proof of the current theorem 3.1.6: see, e.g., relation (3.69).

Since this inequality holds $\forall \mathbf{g}(\cdot) \in \left[L^2(T_A^B)\right]^N$, it immediately follows that

$$\mathbf{g}_h^\dagger(\cdot) = \operatorname{argmin}\left\{\left\|[\mathsf{F}_h\mathbf{g}(\cdot)](\cdot) - \frac{N}{2\pi}\mathbf{\Phi}_\infty(\cdot)\right\|_{2,N}^2\right\}. \tag{3.54}$$

Moreover, since $\mathbf{g}_h^\dagger(z) \in \mathcal{N}(\mathbf{F}_h)^\perp \ \forall z \in T_A^B$, from theorem 3.1.2 we have that

$$\mathbf{g}_h^\dagger(\cdot) \in \mathcal{N}(\mathsf{F}_h)^\perp. \tag{3.55}$$

Relations (3.54) and (3.55) together imply that $\mathbf{g}_h^\dagger(\cdot)$ is the generalized solution of problem (3.44), posed in the Hilbert space $\left[L^2(T_A^B)\right]^N$.

As regards the Tikhonov regularized solution of the same problem (3.44), we firstly consider the Tikhonov regularized solution of problem (3.1) for a generic $\alpha \in \mathbb{R}^+$: we know that it is defined as the unique element in $\mathbb{C}^N$ such that

$$\mathbf{g}_\alpha(z) := \operatorname{argmin}\left\{\left\|\mathbf{F}_h\mathbf{g}(z) - \frac{N}{2\pi}\mathbf{\Phi}_\infty(z)\right\|_{\mathbb{C}^N}^2 + \alpha\|\mathbf{g}(z)\|_{\mathbb{C}^N}^2\right\} \qquad \forall z \in T_A^B \tag{3.56}$$

and that it admits the representation (3.46).

Now, let us consider any $\mathbf{g}(\cdot) \in \left[L^2(T_A^B)\right]^N$: by virtue of relation (3.56), we can write the following inequality, holding f.a.a. $z \in T_A^B$:

$$\left\|\mathbf{F}_h\mathbf{g}_\alpha(z) - \frac{N}{2\pi}\mathbf{\Phi}_\infty(z)\right\|_{\mathbb{C}^N}^2 + \alpha\|\mathbf{g}_\alpha(z)\|_{\mathbb{C}^N}^2 \leq \left\|\mathbf{F}_h\mathbf{g}(z) - \frac{N}{2\pi}\mathbf{\Phi}_\infty(z)\right\|_{\mathbb{C}^N}^2 + \alpha\|\mathbf{g}(z)\|_{\mathbb{C}^N}^2. \tag{3.57}$$

From the previous inequality (3.57), we immediately get:

$$\begin{aligned}\int_{T_A^B}\left[\left\|\mathbf{F}_h\mathbf{g}_\alpha(z) - \frac{N}{2\pi}\mathbf{\Phi}_\infty(z)\right\|_{\mathbb{C}^N}^2 + \alpha\|\mathbf{g}_\alpha(z)\|_{\mathbb{C}^N}^2\right]dz \leq \\ \leq \int_{T_A^B}\left[\left\|\mathbf{F}_h\mathbf{g}(z) - \frac{N}{2\pi}\mathbf{\Phi}_\infty(z)\right\|_{\mathbb{C}^N}^2 + \alpha\|\mathbf{g}(z)\|_{\mathbb{C}^N}^2\right]dz.\end{aligned} \tag{3.58}$$

Using the linearity of the integral and remembering relation (3.51), we can rewrite inequality (3.58) as:

$$\begin{aligned}\int_{T_A^B}\left\|[\mathsf{F}_h\mathbf{g}_\alpha(\cdot)](z) - \frac{N}{2\pi}\mathbf{\Phi}_\infty(z)\right\|_{\mathbb{C}^N}^2 dz + \alpha\int_{T_A^B}\|\mathbf{g}_\alpha(z)\|_{\mathbb{C}^N}^2 dz \leq \\ \leq \int_{T_A^B}\left\|[\mathsf{F}_h\mathbf{g}(\cdot)](z) - \frac{N}{2\pi}\mathbf{\Phi}_\infty(z)\right\|_{\mathbb{C}^N}^2 dz + \alpha\int_{T_A^B}\|\mathbf{g}(z)\|_{\mathbb{C}^N}^2 dz,\end{aligned} \tag{3.59}$$

where we have denoted with $\mathbf{g}_\alpha(\cdot)$ the element of $\left[L^2(T_A^B)\right]^N$ trivially obtained from $\mathbf{g}_\alpha(z)$, as given by (3.46), simply regarding z as a variable on T_A^B instead of a fixed point in $\mathbb{R}^2$: then, $\mathbf{g}_\alpha(\cdot)$ has just the form (3.48).

Recalling relation (3.4), the previous inequality can be rewritten as:

$$\left\|[\mathsf{F}_h\mathbf{g}_\alpha(\cdot)](\cdot)-\frac{N}{2\pi}\mathbf{\Phi}_\infty(\cdot)\right\|_{2,N}^2+\alpha\|\mathbf{g}_\alpha(\cdot)\|_{2,N}^2\leq\left\|[\mathsf{F}_h\mathbf{g}(\cdot)](\cdot)-\frac{N}{2\pi}\mathbf{\Phi}_\infty(\cdot)\right\|_{2,N}^2+\alpha\|\mathbf{g}(\cdot)\|_{2,N}^2. \quad (3.60)$$

Since the last inequality holds $\forall\mathbf{g}(\cdot)\in\left[L^2(T_A^B)\right]^N$, it immediately follows that

$$\mathbf{g}_\alpha(\cdot)=\operatorname{argmin}\left\{\left\|[\mathsf{F}_h\mathbf{g}(\cdot)](\cdot)-\frac{N}{2\pi}\mathbf{\Phi}_\infty(\cdot)\right\|_{2,N}^2+\alpha\|\mathbf{g}(\cdot)\|_{2,N}^2\right\}. \quad (3.61)$$

This clearly means that $\mathbf{g}_\alpha(\cdot)$, as given by (3.48), is the Tikhonov regularized solution of problem (3.44), posed in the Hilbert space $\left[L^2(T_A^B)\right]^N$. This concludes the proof. ■

In expression (3.48), α is a generic real positive number whose optimal[8] value α^* can be fixed by applying one of the different versions of the generalized discrepancy principle, as explained in section 1.8.

Remark 3.1.5. For example, let us focus on the generalized discrepancy function (1.317), holding in the compatible case (analogous remarks can obviously be repeated for the incompatible case or for the mixed approach): in the current functional context, being $\delta=0$, such a function can be written in the form[9]

$$\rho_h^{\kappa_2}(\alpha)=\left\|[\mathsf{F}_h\mathbf{g}_\alpha(\cdot)](\cdot)-\frac{N}{2\pi}\mathbf{\Phi}_\infty(\cdot)\right\|_{2,N}^2-h^2\|\mathbf{g}_\alpha(\cdot)\|_{2,N}^2-\left[\mu_h^{\kappa_2}\left(\frac{N}{2\pi}\mathbf{\Phi}_\infty(\cdot),\mathsf{F}_h\right)\right]^2, \quad (3.62)$$

where $\mu_h^{\kappa_2}\left(\frac{N}{2\pi}\mathbf{\Phi}_\infty(\cdot),\mathsf{F}_h\right)$ is an approximate estimate of the *(simple) incompatibility measure* (see definition 1.8.1):

$$\mu_h\left(\frac{N}{2\pi}\mathbf{\Phi}_\infty(\cdot),\mathsf{F}_h\right)=\inf_{\mathbf{g}(\cdot)\in\left[L^2(T_A^B)\right]^N}\left\|\mathsf{F}_h\mathbf{g}(\cdot)-\frac{N}{2\pi}\mathbf{\Phi}_\infty(\cdot)\right\|_{2,N}, \quad (3.63)$$

in such a way that (cf. inequalities (1.315))

$$\mu_h\left(\frac{N}{2\pi}\mathbf{\Phi}_\infty(\cdot),\mathsf{F}_h\right)\leq\mu_h^{\kappa_2}\left(\frac{N}{2\pi}\mathbf{\Phi}_\infty(\cdot),\mathsf{F}_h\right)\leq\mu_h\left(\frac{N}{2\pi}\mathbf{\Phi}_\infty(\cdot),\mathsf{F}_h\right)+\kappa_2(h), \quad (3.64)$$

and $\lim_{h\to0^+}\kappa_2(h)=0$. Then, the optimal regularization parameter $\alpha^*\equiv\alpha_2^*$ must be fixed, in general, only once by imposing that $\rho_h^{\kappa_2}(\alpha_2^*)=0$.

[8] Cf. footnote 30 in section 2.5.

[9] Cf. expression (2.279).

Of course, according to the generalized discrepancy principle (for the compatible case), as formulated in subsection 1.8.3 from equation (1.319) to inequality (1.321), the problem of finding (the unique) α_2^* such that $\rho_h^{\kappa_2}(\alpha_2^*) = 0$ is to be faced if and only if condition (1.321) is satisfied, i.e., in our case (being $\delta = 0$), if and only if it holds:

$$\left\| \frac{N}{2\pi} \mathbf{\Phi}_\infty(\cdot) \right\|_{2,N} > \mu_h^{\kappa_2} \left(\frac{N}{2\pi} \mathbf{\Phi}_\infty(\cdot), \mathsf{F}_h \right); \tag{3.65}$$

otherwise, the selected approximation of the generalized solution of equation (3.44) is simply zero.

For future purpose, we write here below also the explicit form assumed in the current functional context by the generalized discrepancy function (1.259), holding for the incompatible case: since $\delta = 0$, such a function can be written as

$$\hat{\rho}_h^{\kappa_1}(\alpha) = \left\| [\mathsf{F}_h \mathbf{g}_\alpha(\cdot)](\cdot) - \frac{N}{2\pi} \mathbf{\Phi}_\infty(\cdot) \right\|_{2,N}^2 - \left[h \left\| \mathbf{g}_\alpha(\cdot) \right\|_{2,N} + \hat{\mu}_h^{\kappa_1} \left(\frac{N}{2\pi} \mathbf{\Phi}_\infty(\cdot), \mathsf{F}_h \right) \right]^2, \tag{3.66}$$

where $\hat{\mu}_h^{\kappa_1} \left(\frac{N}{2\pi} \mathbf{\Phi}_\infty(\cdot), \mathsf{F}_h \right)$ is an approximate estimate of the *modified incompatibility measure* (see definition 1.8.1):

$$\hat{\mu}_h \left(\frac{N}{2\pi} \mathbf{\Phi}_\infty(\cdot), \mathsf{F}_h \right) = \inf_{\mathbf{g}(\cdot) \in \left[L^2(T_A^B) \right]^N} \left(h \left\| \mathbf{g}(\cdot) \right\|_{2,N} + \left\| \mathsf{F}_h \mathbf{g}(\cdot) - \frac{N}{2\pi} \mathbf{\Phi}_\infty(\cdot) \right\|_{2,N} \right), \tag{3.67}$$

in such a way that (cf. inequalities (1.257))

$$\hat{\mu}_h \left(\frac{N}{2\pi} \mathbf{\Phi}_\infty(\cdot), \mathsf{F}_h \right) \leq \hat{\mu}_h^{\kappa_1} \left(\frac{N}{2\pi} \mathbf{\Phi}_\infty(\cdot), \mathsf{F}_h \right) \leq \hat{\mu}_h \left(\frac{N}{2\pi} \mathbf{\Phi}_\infty(\cdot), \mathsf{F}_h \right) + \kappa_1(h), \tag{3.68}$$

and $\lim_{h \to 0^+} \kappa_1(h) = 0$. □

Now, if α^* is the optimal value provided by the generalized discrepancy principle in one of its possible versions, then

$$\mathbf{g}_{\alpha^*}(\cdot) = \frac{N}{2\pi} \sum_{p=0}^{r_h - 1} \frac{\sigma_p^h}{(\sigma_p^h)^2 + \alpha^*} (\mathbf{\Phi}_\infty(\cdot), \mathbf{v}_p^h)_{\mathbb{C}^N} \mathbf{u}_p^h \tag{3.69}$$

is the optimal regularized solution of the functional problem (3.44); in particular, it clearly holds:

$$\left\| \mathbf{g}_{\alpha^*}(z) \right\|_{\mathbb{C}^N} = \frac{N}{2\pi} \sqrt{ \sum_{p=0}^{r_h - 1} \frac{(\sigma_p^h)^2}{[(\sigma_p^h)^2 + \alpha^*]^2} \left| \left(\mathbf{\Phi}_\infty(z), \mathbf{v}_p^h \right)_{\mathbb{C}^N} \right|^2 } \qquad \forall z \in T_A^B. \tag{3.70}$$

Moreover, once a suitable monotonic continuous function $I : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ is chosen, we can now define our new indicator function on all T_A^B as (cf. definition (2.283)):

$$\begin{aligned} \Psi_I : T_A^B &\longrightarrow \mathbb{R} \\ z &\longmapsto \Psi_I(z) := I \left(\left\| \mathbf{g}_{\alpha^*}(z) \right\|_{\mathbb{C}^N} \right). \end{aligned} \tag{3.71}$$

Remark 3.1.6. We point out that, independently of the choice of I, in the analytic expression of $\Psi_I(z)$, as given by (3.71), no sampling is performed, since α^* does not depend on z and all the other terms in expression (3.70) are known once the measured far-field matrix is at disposal. From a computational viewpoint, this removes the problem of deciding the parameters of the optimal grid $\mathcal{Z}$ containing the scatterer (number of points, sampling distance). □

Remark 3.1.7. The mathematical framework for this new implementation of the linear sampling method naturally implies that the optimal regularization parameter is unique, i.e. independent of z. On the other hand, it is worthwhile noticing that, although in a completely different context, a sampling-point-independent choice of the regularization parameter is also suggested in [22], by giving a heuristic recipe based on physical considerations. □

Theorem 3.1.6 provides a new implementation of the linear sampling method, whereby the contour of the scatterer is detected by all the points where $\Psi_I(z)$ becomes mostly large or small, depending on the increasing or decreasing monotonicity of I. Hence the most natural question is now: does this new implementation yield the same results as the traditional one based on a sampling in the space? We can answer this question only from an empirical point of view, i.e. by making several numerical experiments changing, from time to time, the scatterer, the noise level, the boundary conditions, the wavenumber, the number of incidence/observation angles, the indicator function, the criterion for its cut-off and so on. Of course, it is not possible here to present all the numerical experiments we have made to this purpose: anyway, as far as our experience is concerned, no significant difference between the two implementations has ever been observed. Here we can only support and clarify such a general observation by means of few specific experiments, which we are going to illustrate. As a starting point, we consider exactly the same inverse scattering problems we chose in section 2.5 (with figures in section B.1) to illustrate the traditional implementation of the linear sampling method, changing only the mathematical framework and, consequently, the regularization procedure, as explained just above: this means that we use the generalized discrepancy principle in the compatible case (cf. subsection 1.8.3 and, in particular, definition (1.317), as well as its specific form (3.62) for the current context) and we choose

$$\begin{aligned} \Psi_{-2\ln} : T_A^B &\longrightarrow \mathbb{R} \\ z &\longmapsto \Psi_{-2\ln}(z) := -\ln \|\mathbf{g}_{\alpha_2^*}(z)\|_{\mathbb{C}^N}^2 \end{aligned} \tag{3.72}$$

as indicator function (cf. definition (3.71) with $I(t) := -\ln t^2 \equiv -2\ln t$); the cut-off value of the latter, as in the traditional implementation, is fixed by making the area defined by its level curves equal to the one of the true scatterer.

The results are shown in figures B.4, B.5 and B.6: it can be clearly seen, mainly by comparing panels (c) and (c') of these figures, on the one hand, with the homologous panels of figures B.1, B.2 and B.3 respectively, on the other one, that the two implementations give nearly

indistinguishable results. Their substantial equivalence is even more evident in figure B.7, in which, for the same scattering experiments and the same choice of $I(t) := -\ln t^2 \equiv -2\ln t$ as before, we have essentially[10] overlapped the corresponding panels (c) [(c')] of all the previous figures B.1-B.6: for each of the six panels in figure B.7, we provide (solid line) the profile of the true object and, superimposed, the reconstructed contours provided by the two implementations: the traditional one (dashed line) and the no-sampling one (dotted line). From these plots the differences between the two implementations turn out to be completely negligible. Analogous results occur when the noise level affecting the far-field pattern, the number of incidence/observation angles and the wavenumber are changed. Hence, from now on we shall deal only with the no-sampling implementation.

Let us now see what happens, in the new functional context, if we consider the same scattering experiments and the same choice of $I(t) := -\ln t^2 \equiv -2\ln t$ as before, but using this time the generalized discrepancy principle in the incompatible case (cf. subsection 1.8.2 and, in particular, definition (1.259), as well as its specific form (3.66) for the current context): the results are presented in figures B.8, B.9 and B.10, which, by virtue of a comparison with the corresponding figures B.4, B.5 and B.6 (and with the respective values of the regularization parameters), clearly show that the generalized discrepancy principle in the incompatible case tends to be strongly oversmoothing, mainly for sufficiently high noise levels: indeed the only satisfactory reconstructions are the ones of panels (c) in figures B.8 and B.9, holding for a noise level $n = 1\%$; on the other hand, figure B.9 itself provides, for the same noise level $n = 1\%$, a reconstruction of the kite which seems to be somehow better than the one provided by figure B.5, as suggested, in particular, by comparing the darkest internal regions in panels (b) of figures B.5 and B.9. On the contrary, for a noise level $n = 10\%$, a comparison between panels (a'), (b') and (c') of figure B.5 (as well as of figures B.4 and B.6) on the one hand, and the homologous panels of figure B.9 (as well as of figures B.8 and B.10), on the other one, clearly shows that the generalized discrepancy principle works much worse in the incompatible case than in the compatible one.

If we now remember the discussion of subsection 1.8.4, the idea may arise, roughly speaking,

[10] To tell the truth, the panels in figure B.7 have been realized some days later than figures B.1-B.6: this implies that, although all the physical parameters (noise levels included) have been obviously set to the same values as in figures B.1-B.6, it has not been possible to blur the exact far-field matrix $\mathbf{F}$ with just the same noise matrix $\mathbf{H}$ (cf. equalities (2.265) and (2.266)), since the latter, as explained in remark 2.5.1, is a random matrix depending on an initializing "seed" which, in general, is chosen differently and randomly every time, so that we could not remember its value in past numerical experiments. This explains the very slight differences occurring in the reconstructed profiles and the not complete concordance of the values of the regularization parameters when one compares figures B.1-B.6, on the one hand, and figure B.7 (with its associated table B.1), on the other one. Of course, we could have remade all the figures using, at least for the same scatterer, the same seed, but we have considered more interesting and more convincing this kind of "double" confirm of the substantial equivalence between the two implementations of the linear sampling method.

of blending the reconstructions obtained by using the two generalized discrepancy principles, at least in situations in which the noise level has an intermediate value, for example $n = 5\%$. Hence, we have repeated, for the ellipse and the kite, the previous scattering experiments: all the physical and geometrical parameters are exactly the same as before (as well as the choice of $I(t) := -\ln t^2 \equiv -2\ln t$), except that the noise level is now $n = 5\%$. The results are shown in figures B.11 and B.12, in which we compare, for each scatterer, the three reconstructions obtained by using three different values of the regularization parameter: α_2^* (i.e. the zero of the generalized discrepancy function (3.62), holding for the compatible case), α_1^* (i.e. the zero of the generalized discrepancy function (3.66), holding for the incompatible case) and $\alpha_b^* := c\,\alpha_1^* + (1-c)\,\alpha_2^*$ (the latter is a shorthand for definition (1.346)), having heuristically[11] chosen, for both the scatterers, the value of $c = c(h_s)$ as

$$c(h_s) := \frac{2}{\pi}\arctan(40\,h_s), \tag{3.73}$$

where h_s denotes the norm of the specific noise matrix $\mathbf{H}_s$ added to the exact far-field matrix $\mathbf{F}$ in that particular numerical experiment. From panels (c), (c'), (c") of figures B.11 and B.12, we can observe that the reconstructions of both the ellipse and, even more, the kite are improved by the blending regularization.

3.2. Band-limitedness of the indicator function

From now on, we shall choose $I(t) := t^2$ in definition (3.71): then the selected indicator function will be

$$\Psi_2(z) := \|\mathbf{g}_{\alpha^*}(z)\|^2_{\mathbb{C}^N} = \frac{N^2}{4\pi^2}\sum_{p=0}^{r_h-1}\frac{(\sigma_p^h)^2}{[(\sigma_p^h)^2+\alpha^*]^2}\left|\left(\mathbf{\Phi}_\infty(z),\mathbf{v}_p^h\right)_{\mathbb{C}^N}\right|^2. \tag{3.74}$$

The reason for the choice of (3.74) as indicator function is that it leads to feasible analytical results, as we are going to see. However, in section 3.3 we shall motivate this choice in a more general framework.

Our aim is now to compute the Fourier transform of the indicator function (3.74) analytically continued onto all $\mathbb{R}^2$ (and still denoted with $\Psi_2(z)$).

We begin by remembering that (cf. relations (2.75) and (2.271)):

$$\begin{aligned}\mathbf{\Phi}_\infty(z) &= \frac{e^{\mathrm{i}\frac{\pi}{4}}}{\sqrt{8\pi k}}\left[e^{-\mathrm{i}kz\cdot\hat{x}_0},\ldots,e^{-\mathrm{i}kz\cdot\hat{x}_{N-1}}\right]^{\mathrm{T}} = \\ &= \frac{e^{\mathrm{i}\frac{\pi}{4}}}{\sqrt{8\pi k}}\left[e^{-\mathrm{i}k(z_1\cos\varphi_0+z_2\sin\varphi_0)},\ldots,e^{-\mathrm{i}k(z_1\cos\varphi_{N-1}+z_2\sin\varphi_{N-1})}\right]^{\mathrm{T}},\end{aligned} \tag{3.75}$$

[11] Cf. remark 1.8.8.

where $z = (z_1, z_2) \in T_A^B$. Hence, by inserting the previous expression (3.75) of $\mathbf{\Phi}_\infty(z)$ into (3.74), we obtain

$$\Psi_2(z) = \frac{N^2}{32\pi^3 k} \sum_{p=0}^{r_h-1} \frac{(\sigma_p^h)^2}{[(\sigma_p^h)^2 + \alpha^*]^2} \cdot \tag{3.76}$$
$$\cdot \left| \bar{v}_{p,0}^h e^{-\mathrm{i}k(z_1 \cos\varphi_0 + z_2 \sin\varphi_0)} + \cdots + \bar{v}_{p,N-1}^h e^{-\mathrm{i}k(z_1 \cos\varphi_{N-1} + z_2 \sin\varphi_{N-1})} \right|^2,$$

where $\bar{v}_{p,0}^h, \ldots, \bar{v}_{p,N-1}^h$ denote the complex conjugate of the N components of the vector $\mathbf{v}_p^h$ for each $p = 0, \ldots, r_h - 1$. If we now consider, in particular, the term in absolute value at the right-hand side of equality (3.76) and use the polar coordinates of the N components of each vector $\mathbf{v}_p^h$, i.e.:

$$\begin{aligned} v_{p,0}^h &= (\rho_{p,0}^h, \varepsilon_{p,0}^h) = \rho_{p,0}^h e^{\mathrm{i}\varepsilon_{p,0}^h}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ v_{p,N-1}^h &= (\rho_{p,N-1}^h, \varepsilon_{p,N-1}^h) = \rho_{p,N-1}^h e^{\mathrm{i}\varepsilon_{p,N-1}^h}, \end{aligned} \tag{3.77}$$

we get

$$\left| \bar{v}_{p,0}^h e^{-\mathrm{i}k(z_1 \cos\varphi_0 + z_2 \sin\varphi_0)} + \cdots + \bar{v}_{p,N-1}^h e^{-\mathrm{i}k(z_1 \cos\varphi_{N-1} + z_2 \sin\varphi_{N-1})} \right|^2 = \tag{3.78}$$

$$= \left| \rho_{p,0}^h e^{-\mathrm{i}(kz_1 \cos\varphi_0 + kz_2 \sin\varphi_0 + \varepsilon_{p,0}^h)} + \cdots + \rho_{p,N-1}^h e^{-\mathrm{i}(kz_1 \cos\varphi_{N-1} + kz_2 \sin\varphi_{N-1} + \varepsilon_{p,N-1}^h)} \right|^2 =$$

$$= \left| \rho_{p,0}^h \cos\left[k\left(z_1 \cos\varphi_0 + z_2 \sin\varphi_0\right) + \varepsilon_{p,0}^h\right] + \ldots \right.$$
$$\left. \cdots + \rho_{p,N-1}^h \cos\left[k\left(z_1 \cos\varphi_{N-1} + z_2 \sin\varphi_{N-1}\right) + \varepsilon_{p,N-1}^h\right] + \right.$$
$$- \mathrm{i}\left\{ \rho_{p,0}^h \sin\left[k\left(z_1 \cos\varphi_0 + z_2 \sin\varphi_0\right) + \varepsilon_{p,0}^h\right] + \ldots \right.$$
$$\left. \cdots + \rho_{p,N-1}^h \sin\left[k\left(z_1 \cos\varphi_{N-1} + z_2 \sin\varphi_{N-1}\right) + \varepsilon_{p,N-1}^h\right] \right\} \Big|^2 =$$

$$= \left| \sum_{i=0}^{N-1} \rho_{p,i}^h \cos\left[k\left(z_1 \cos\varphi_i + z_2 \sin\varphi_i\right) + \varepsilon_{p,i}^h\right] - \mathrm{i} \sum_{i=0}^{N-1} \rho_{p,i}^h \sin\left[k\left(z_1 \cos\varphi_i + z_2 \sin\varphi_i\right) + \varepsilon_{p,i}^h\right] \right|^2 =$$

$$= \left\{ \sum_{i=0}^{N-1} \rho_{p,i}^h \cos\left[k\left(z_1 \cos\varphi_i + z_2 \sin\varphi_i\right) + \varepsilon_{p,i}^h\right] \right\}^2 +$$
$$+ \left\{ \sum_{i=0}^{N-1} \rho_{p,i}^h \sin\left[k\left(z_1 \cos\varphi_i + z_2 \sin\varphi_i\right) + \varepsilon_{p,i}^h\right] \right\}^2 =$$

$$= \sum_{i,j=0}^{N-1} \rho_{p,i}^h \rho_{p,j}^h \cos\left[k\left(z_1 \cos\varphi_i + z_2 \sin\varphi_i\right) + \varepsilon_{p,i}^h\right] \cos\left[k\left(z_1 \cos\varphi_j + z_2 \sin\varphi_j\right) + \varepsilon_{p,j}^h\right] +$$
$$+ \sum_{i,j=0}^{N-1} \rho_{p,i}^h \rho_{p,j}^h \sin\left[k\left(z_1 \cos\varphi_i + z_2 \sin\varphi_i\right) + \varepsilon_{p,i}^h\right] \sin\left[k\left(z_1 \cos\varphi_j + z_2 \sin\varphi_j\right) + \varepsilon_{p,j}^h\right] =$$

$$= \sum_{i,j=0}^{N-1} \rho_{p,i}^h \rho_{p,j}^h \left\{ \cos\left[k\left(z_1\cos\varphi_i + z_2\sin\varphi_i\right) + \varepsilon_{p,i}^h\right] \cos\left[k\left(z_1\cos\varphi_j + z_2\sin\varphi_j\right) + \varepsilon_{p,j}^h\right] + \right.$$

$$\left. + \sin\left[k\left(z_1\cos\varphi_i + z_2\sin\varphi_i\right) + \varepsilon_{p,i}^h\right] \sin\left[k\left(z_1\cos\varphi_j + z_2\sin\varphi_j\right) + \varepsilon_{p,j}^h\right] \right\} =$$

$$= \sum_{i,j=0}^{N-1} \rho_{p,i}^h \rho_{p,j}^h \cos\left\{k\left[z_1\left(\cos\varphi_i - \cos\varphi_j\right) + z_2\left(\sin\varphi_i - \sin\varphi_j\right)\right] + \varepsilon_{p,i}^h - \varepsilon_{p,j}^h\right\}, \tag{3.79}$$

where, in the last passage, we have used the well-known identity:

$$\cos\alpha\,\cos\beta + \sin\alpha\,\sin\beta = \cos(\alpha-\beta) \qquad \forall\alpha,\beta\in\mathbb{R}. \tag{3.80}$$

Hence, if we put

$$\hat{\omega}_{1,ij} := k(\cos\varphi_i - \cos\varphi_j), \tag{3.81}$$

$$\hat{\omega}_{2,ij} := k(\sin\varphi_i - \sin\varphi_j), \tag{3.82}$$

$$\Delta\varepsilon_{p,ij}^h := \varepsilon_{p,i}^h - \varepsilon_{p,j}^h \tag{3.83}$$

and substitute (3.81), (3.82), (3.83) into (3.79), from relation (3.76) we can obtain the following expression for the indicator function:

$$\Psi_2(z) = \frac{N^2}{32\pi^3 k} \sum_{p=0}^{r_h-1} \frac{(\sigma_p^h)^2}{[(\sigma_p^h)^2+\alpha^*]^2} \sum_{i,j=0}^{N-1} \rho_{p,i}^h \rho_{p,j}^h \cos\left(\hat{\omega}_{1,ij}\,z_1 + \hat{\omega}_{2,ij}\,z_2 + \Delta\varepsilon_{p,ij}^h\right). \tag{3.84}$$

Hence, $\Psi_2(z)$ consists of a finite linear combination of cosine-like functions in the variables z_1 and z_2. We now remember that, starting from the very beginning of the general framework we have conceived for our no-sampling implementation of the linear sampling method, $\Psi_2(z)$ is defined, a priori, only in T_A^B; however, the right-hand side of expression (3.84) is mathematically well-defined for all $z\in\mathbb{R}^2$ and, as such, it is easily seen to be real-analytic in $\mathbb{R}^2$ by applying theorem A.4.2. Then, by virtue of the unique continuation property enjoyed by real-analytic functions (see theorem A.4.3), we can analytically extend $\Psi_2(z)$ to all $\mathbb{R}^2$ simply regarding expression (3.84) as holding not only in T_A^B, but also, by definition, in all $\mathbb{R}^2$. We shall keep on denoting with $\Psi_2(z)$ the analytical extension in $\mathbb{R}^2$ of the original $\Psi_2(z)$ defined in T_A^B.

The next step is to compute the Fourier transform $[\mathcal{F}(\Psi_2(z))](\omega) \equiv [\mathcal{F}(\Psi_2)](\omega)$ of $\Psi_2(z)$ (with $\omega = (\omega_1,\omega_2)\in\mathbb{R}^2$); of course, $\Psi_2(z)$ belongs neither to $L^1(\mathbb{R}^2)$ nor to $L^2(\mathbb{R}^2)$, but it should be rather regarded as an element of $\mathcal{S}^*(\mathbb{R}^2)$ (see the brief comment after relation (A.53)) and, consequently, the Fourier transform we need for our purposes is the operator $\mathcal{F}:\mathcal{S}^*(\mathbb{R}^2)\to\mathcal{S}^*(\mathbb{R}^2)$ defined by condition (A.62).

However, in order to compute $[\mathcal{F}(\Psi_2)](\omega)$ we can still use several tools typical of the classical Lebesgue integration of functions, although with the proper care. Our following calculations can be rigorously justified by the theory of distributions: for sake of brevity, here we simply refer, e.g., to [35] (chapters III and V).

We firstly observe that, by virtue of expression (3.84), $[\mathcal{F}(\Psi_2)](\omega)$ is determined, by linearity, once we have computed the Fourier transform of the functions $\cos\left(\hat{\omega}_{1,ij}\, z_1 + \hat{\omega}_{2,ij}\, z_2 + \Delta\varepsilon^h_{p,ij}\right)$ for all $p = 0, \ldots, r_h - 1$ and for all $i, j = 0, \ldots, N-1$. To this end, we use the well-known identity

$$\cos(\alpha + \beta) = \cos\alpha\, \cos\beta - \sin\alpha\, \sin\beta \qquad \forall \alpha, \beta \in \mathbb{R} \tag{3.85}$$

to write:

$$\begin{aligned}\cos\left(\hat{\omega}_{1,ij}\, z_1 + \hat{\omega}_{2,ij}\, z_2 + \Delta\varepsilon^h_{p,ij}\right) = &\cos\left(\hat{\omega}_{1,ij}\, z_1 + \hat{\omega}_{2,ij}\, z_2\right)\, \cos\left(\Delta\varepsilon^h_{p,ij}\right) + \\ &- \sin\left(\hat{\omega}_{1,ij}\, z_1 + \hat{\omega}_{2,ij}\, z_2\right)\, \sin\left(\Delta\varepsilon^h_{p,ij}\right).\end{aligned} \tag{3.86}$$

moreover, by applying again identity (3.85) and the following one:

$$\sin(\alpha + \beta) = \sin\alpha\, \cos\beta + \cos\alpha\, \sin\beta \qquad \forall \alpha, \beta \in \mathbb{R}, \tag{3.87}$$

we easily get:

$$\cos\left(\hat{\omega}_{1,ij}\, z_1 + \hat{\omega}_{2,ij}\, z_2\right) = \cos\left(\hat{\omega}_{1,ij}\, z_1\right)\, \cos\left(\hat{\omega}_{2,ij}\, z_2\right) - \sin\left(\hat{\omega}_{1,ij}\, z_1\right)\, \sin\left(\hat{\omega}_{2,ij}\, z_2\right), \tag{3.88}$$

$$\sin\left(\hat{\omega}_{1,ij}\, z_1 + \hat{\omega}_{2,ij}\, z_2\right) = \sin\left(\hat{\omega}_{1,ij}\, z_1\right)\, \cos\left(\hat{\omega}_{2,ij}\, z_2\right) + \cos\left(\hat{\omega}_{1,ij}\, z_1\right)\, \sin\left(\hat{\omega}_{2,ij}\, z_2\right). \tag{3.89}$$

Substituting (3.88) and (3.89) into (3.86), we have:

$$\begin{aligned}\cos\left(\hat{\omega}_{1,ij}\, z_1 + \hat{\omega}_{2,ij}\, z_2 + \Delta\varepsilon^h_{p,ij}\right) = &\\ = \cos\left(\Delta\varepsilon^h_{p,ij}\right)\, &\left[\cos\left(\hat{\omega}_{1,ij}\, z_1\right)\, \cos\left(\hat{\omega}_{2,ij}\, z_2\right) - \sin\left(\hat{\omega}_{1,ij}\, z_1\right)\, \sin\left(\hat{\omega}_{2,ij}\, z_2\right)\right] + \\ - \sin\left(\Delta\varepsilon^h_{p,ij}\right)\, &\left[\sin\left(\hat{\omega}_{1,ij}\, z_1\right)\, \cos\left(\hat{\omega}_{2,ij}\, z_2\right) + \cos\left(\hat{\omega}_{1,ij}\, z_1\right)\, \sin\left(\hat{\omega}_{2,ij}\, z_2\right)\right];\end{aligned} \tag{3.90}$$

then, by linearity again, it suffices to compute, for all $i, j = 0, \ldots, N-1$, the Fourier transform of the following functions (regarded as elements of $\mathcal{S}^*(\mathbb{R}^2)$):

$$f_1(z_1, z_2) := \cos\left(\hat{\omega}_{1,ij}\, z_1\right)\, \cos\left(\hat{\omega}_{2,ij}\, z_2\right), \tag{3.91}$$

$$f_2(z_1, z_2) := \sin\left(\hat{\omega}_{1,ij}\, z_1\right)\, \sin\left(\hat{\omega}_{2,ij}\, z_2\right), \tag{3.92}$$

$$f_3(z_1, z_2) := \sin\left(\hat{\omega}_{1,ij}\, z_1\right)\, \cos\left(\hat{\omega}_{2,ij}\, z_2\right), \tag{3.93}$$

$$f_4(z_1, z_2) := \cos\left(\hat{\omega}_{1,ij}\, z_1\right)\, \sin\left(\hat{\omega}_{2,ij}\, z_2\right), \tag{3.94}$$

so that we can write:

$$\begin{aligned}\left[\mathcal{F}\left(\cos\left(\hat{\omega}_{1,ij}\, z_1 + \hat{\omega}_{2,ij}\, z_2 + \Delta\varepsilon^h_{p,ij}\right)\right)\right](\omega_1, \omega_2) = &\\ = \cos\left(\Delta\varepsilon^h_{p,ij}\right)\, &\left\{[\mathcal{F}(f_1)](\omega_1, \omega_2) - [\mathcal{F}(f_2)](\omega_1, \omega_2)\right\} + \\ - \sin\left(\Delta\varepsilon^h_{p,ij}\right)\, &\left\{[\mathcal{F}(f_3)](\omega_1, \omega_2) + [\mathcal{F}(f_4)](\omega_1, \omega_2)\right\}.\end{aligned} \tag{3.95}$$

We can now remember a general result in Fourier transform theory[12]: if $g(z_1, z_2)$ admits the factorization $g(z_1, z_2) = g_1(z_1)\, g_2(z_2)$ and there exist $[\mathcal{F}(g_1)](\omega_1)$, $[\mathcal{F}(g_2)](\omega_2)$, then it holds:

$$[\mathcal{F}(g)](\omega_1, \omega_2) = [\mathcal{F}(g_1)](\omega_1)\, [\mathcal{F}(g_2)](\omega_2), \tag{3.96}$$

[12] See, e.g., [35], p. 233 and p. 284, or [55], p. 91.

where obviously $\mathcal{F}$ is to be intended as acting on $\mathcal{S}^*(\mathbb{R}^2)$ at the left-hand side and on $\mathcal{S}^*(\mathbb{R})$ at the right-hand side of identity (3.96).

If we use, in general, the specific definition (A.54) of the Fourier transform, two elementary results in one-dimensional distribution theory are[13]:

$$[\mathcal{F}(\cos(\omega_0 t))](\omega) = \pi\left[\delta(\omega-\omega_0)+\delta(\omega+\omega_0)\right] \qquad \forall\omega_0\in\mathbb{R}, \tag{3.97}$$

$$[\mathcal{F}(\sin(\omega_0 t))](\omega) = \mathrm{i}\pi\left[\delta(\omega+\omega_0)-\delta(\omega-\omega_0)\right] \qquad \forall\omega_0\in\mathbb{R}, \tag{3.98}$$

where we have obviously denoted with $\delta(\omega-\omega_0)$ the Dirac delta set in ω_0. Hence, by virtue of relations (3.96), (3.97) and (3.98), we can easily compute the Fourier transforms of the functions given by (3.91)-(3.94) as follows:

$$[\mathcal{F}(f_1)](\omega_1,\omega_2) = \pi\left[\delta(\omega_1-\hat{\omega}_{1,ij})+\delta(\omega_1+\hat{\omega}_{1,ij})\right]\cdot\pi\left[\delta(\omega_2-\hat{\omega}_{2,ij})+\delta(\omega_2+\hat{\omega}_{2,ij})\right], \tag{3.99}$$

$$[\mathcal{F}(f_2)](\omega_1,\omega_2) = \mathrm{i}\pi\left[\delta(\omega_1+\hat{\omega}_{1,ij})-\delta(\omega_1-\hat{\omega}_{1,ij})\right]\cdot\mathrm{i}\pi\left[\delta(\omega_2+\hat{\omega}_{2,ij})-\delta(\omega_2-\hat{\omega}_{2,ij})\right], \tag{3.100}$$

$$[\mathcal{F}(f_3)](\omega_1,\omega_2) = \mathrm{i}\pi\left[\delta(\omega_1+\hat{\omega}_{1,ij})-\delta(\omega_1-\hat{\omega}_{1,ij})\right]\cdot\pi\left[\delta(\omega_2-\hat{\omega}_{2,ij})+\delta(\omega_2+\hat{\omega}_{2,ij})\right], \tag{3.101}$$

$$[\mathcal{F}(f_4)](\omega_1,\omega_2) = \pi\left[\delta(\omega_1-\hat{\omega}_{1,ij})+\delta(\omega_1+\hat{\omega}_{1,ij})\right]\cdot\mathrm{i}\pi\left[\delta(\omega_2+\hat{\omega}_{2,ij})-\delta(\omega_2-\hat{\omega}_{2,ij})\right]. \tag{3.102}$$

From relations (3.99) and (3.100) we easily find:

$$[\mathcal{F}(f_1)](\omega_1,\omega_2)-[\mathcal{F}(f_2)](\omega_1,\omega_2) = 2\pi^2[\delta(\omega_1-\hat{\omega}_{1,ij})\delta(\omega_2-\hat{\omega}_{2,ij})+\delta(\omega_1+\hat{\omega}_{1,ij})\delta(\omega_2+\hat{\omega}_{2,ij})], \tag{3.103}$$

while from relations (3.101) and (3.102) we get:

$$[\mathcal{F}(f_3)](\omega_1,\omega_2)+[\mathcal{F}(f_4)](\omega_1,\omega_2) = -2\mathrm{i}\pi^2[\delta(\omega_1-\hat{\omega}_{1,ij})\delta(\omega_2-\hat{\omega}_{2,ij})-\delta(\omega_1+\hat{\omega}_{1,ij})\delta(\omega_2+\hat{\omega}_{2,ij})]. \tag{3.104}$$

Finally, substituting relations (3.103) and (3.104) into equality (3.95) and remembering expression (3.84), we find that the Fourier transform of $\Psi_2(z)$ is:

$$\begin{aligned}[\mathcal{F}(\Psi_2)](\omega_1,\omega_2) =& \frac{N^2}{32\pi^3 k}\sum_{p=0}^{r_h-1}\frac{(\sigma_p^h)^2}{[(\sigma_p^h)^2+\alpha^*]^2}\sum_{i,j=0}^{N-1}\rho_{p,i}^h\,\rho_{p,j}^h\cdot \\ &\cdot\left\{2\pi^2\cos(\Delta\varepsilon_{p,ij}^h)[\delta(\omega_1-\hat{\omega}_{1,ij})\delta(\omega_2-\hat{\omega}_{2,ij})+\delta(\omega_1+\hat{\omega}_{1,ij})\delta(\omega_2+\hat{\omega}_{2,ij})]+\right. \\ &\left.+2\mathrm{i}\pi^2\sin(\Delta\varepsilon_{p,ij}^h)[\delta(\omega_1-\hat{\omega}_{1,ij})\delta(\omega_2-\hat{\omega}_{2,ij})-\delta(\omega_1+\hat{\omega}_{1,ij})\delta(\omega_2+\hat{\omega}_{2,ij})]\right\}.\end{aligned} \tag{3.105}$$

This expression shows that the Fourier transform of the indicator function $\Psi_2(z)$ analytically extended to all $\mathbb{R}^2$ is a distribution, which can be named "Dirac brush", whose support[14] (which is, by definition, the band of $\Psi_2(z)$) is the compact set

$$\mathsf{S} = \{(\hat{\omega}_{1,ij},\hat{\omega}_{2,ij})\}_{i,j=0}^{N-1}. \tag{3.106}$$

[13] See, e.g., [35], p. 256 and p. 609, or [55], p. 100.

[14] The definition of support of a distribution is given in section A.5; for some useful properties concerning the support of a sum or of a product of distributions, see, e.g., [35], p. 143-150.

From relations (3.81), (3.82) we easily have that

$$\sqrt{(\hat{\omega}_{1,ij})^2 + (\hat{\omega}_{2,ij})^2} \leq 2k \qquad \forall i,j = 0,\ldots,N-1, \tag{3.107}$$

i.e. the support of the Dirac brush is contained in a circle of radius $2k$ in the frequency space. Moreover, if we define

$$\Omega_1 := \max_{i,j}[\hat{\omega}_{1,ij}], \qquad \Omega_2 := \max_{i,j}[\hat{\omega}_{2,ij}], \tag{3.108}$$

we can say that the indicator function (3.74) is (Ω_1, Ω_2)-bandlimited. Hence, if we put

$$\mathrm{sinc}(t) := \begin{cases} \dfrac{\sin(\pi t)}{\pi t} & \text{if } t \neq 0 \\ 1 & \text{if } t = 0, \end{cases} \tag{3.109}$$

the following Shannon-Nyquist representation[15] of $\Psi_2(z) \equiv \Psi_2(z_1, z_2)$

$$\Psi_2(z) = \sum_{n_1,n_2=-\infty}^{+\infty} \Psi_2(n_1 d_1, n_2 d_2)\,\mathrm{sinc}\left[\frac{z_1 - n_1 d_1}{d_1}\right]\mathrm{sinc}\left[\frac{z_2 - n_2 d_2}{d_2}\right] \tag{3.110}$$

holds, provided that the sampling distances d_1 and d_2 along the z_1-axis and z_2-axis respectively satisfy the following conditions[16]:

$$d_1 < \frac{\pi}{\Omega_1}, \qquad d_2 < \frac{\pi}{\Omega_2}, \tag{3.111}$$

where $\dfrac{\pi}{\Omega_1}$ and $\dfrac{\pi}{\Omega_2}$ are called *Nyquist distances*. For example, if N is a multiple of 4, then it holds $\Omega_1 = \Omega_2 = 2k$ and consequently the Nyquist distances are:

$$\frac{\pi}{\Omega_1} = \frac{\pi}{\Omega_2} = \frac{\lambda}{4}, \tag{3.112}$$

where $\lambda = \dfrac{2\pi}{k}$ is the wavelength. If N is not a multiple of 4, then $\dfrac{\lambda}{4}$ is a strict lower bound for the Nyquist distance.

We remark that the support (3.106) of distribution (3.105) is independent of the scatterer and only depends on the wavenumber and on the number N of the observation/incidence angles (of course, the Fourier transform of the indicator function does depend on the scatterer characteristics, in particular through the singular system of the far-field matrix). As examples, figure B.13 shows this support in the case of $k = 5$ and $N = 8, 16, 32, 64$. In order to validate these results, we considered the scattering of $N = 8$ plane waves for $k = 5$ with the conducting kite (2.285) in the case of Dirichlet boundary conditions, for $N = 8$ observation angles. In figure B.14 we computed the numerical Fourier transform of the corresponding indicator function (panel (a)) and compared it with the theoretical support for the same N (panel (b)): the position of the peaks of the numerical Fourier transform coincides with the peaks of the Dirac brush.

[15] See, e.g., [9], p. 23-25.

[16] See, e.g., [9], p. 23, or also [12], p. 83.

3.3. Spatial resolution

Spatial resolution is the main concept in image formation theory and practice. In this regard, the results of the previous two sections provide some hints about how to heuristically estimate the ability of the method to recover close objects from the superposition of their noisy discretized far-field patterns. Indeed, the Shannon-Nyquist representation (3.110) implies that $\Psi_2(z)$ cannot vary significantly on length scales smaller than the Nyquist distance $\lambda/4$, since such representation consists of a superposition of sinc-like functions that are peaked at a distance smaller then $\lambda/4$ from one another and are very smooth between adjacent sampling points: hence, when $\Psi_2(z)$ is chosen as indicator function, one is tempted to regard $\lambda/4$ at least as a rough estimate of the distance under which two distinct objects begin to be seen as a single one. However, a satisfactory solution to the problem of determining a reliable assessment of the spatial resolution achievable by the method is still missing, since one should also take into account the following crucial points:

1. in whatever implementation of the linear sampling method, the visualization of the scatterer profile is obtained by choosing a cut-off section for the 3D plot of $\Psi_2(z)$ and the relation between the Nyquist distance for $\Psi_2(z)$ and the spatial resolution achievable for the scatterer profile on this section still needs to be clarified; in particular, when the cut-off criterion consists in selecting the level curve of the indicator function containing an area equal to the one contained by the theoretical profile (i.e. the cut-off criterion adopted by us), numerical experience suggests that the Nyquist distance often represents a pessimistic estimate of the spatial resolution in two dimensions, i.e. situations frequently occur where the cut-off section produces quite featured profiles although $\Psi_2(z)$ varies very smoothly;

2. the way in which the cut-off section itself is performed has, in principle, a strong influence in determining whether two close objects are distinguishable or not[17]: if we remember that, till now, no general and satisfactory cut-off criterion has been formulated, we can better realize the importance of this issue;

3. noise only affects the singular system of the far-field matrix and its presence does not change the band of the indicator function, i.e. does not change the support of the distribution $[\mathcal{F}(\Psi_2)](\omega)$. Therefore this is a typical application where the presence of noise does not affect, at least in principle, the theoretical spatial resolution of the method. On

[17] Let us think, e.g., to a three-dimensional plot formed, roughly speaking, by two similar and positive bells blending at half-height, say $a/2 > 0$: if the cut-off value is greater than $a/2$ (and less than a), two distinct objects will be detected, while for a cut-off value less than $a/2$ (and greater than 0), the corresponding section will provide a single object.

the other hand, the presence of noise affects the overall shape of the indicator function and thus contributes to a general deterioration of the visualization accuracy. In other terms, when two scatterers, as it happens in figure B.15 (which we are going to introduce soon below), are made closer and closer (while all the other physical and geometrical parameters are maintained unaltered), their progressive deformation is due to two different contributions: a growing one, deriving from the increasing proximity of the two scatterers, and a roughly constant one, deriving from the presence of noise: then, the latter can actually worsen the theoretically estimated resolution owing to an additional deformation in the reconstruction of the two objects, which consequently may begin to touch each other also when their spatial separation is greater than the Nyquist distance.

Let us now consider the following numerical experiment. We choose two conducting objects with Dirichlet boundary conditions, an ellipse and a peanut, and we bring them nearer and nearer: the ellipse has semiaxes equal to 1 and 2 respectively, while the peanut is obtained from a rotation of 45° counterclockwise and a suitable translation[18] of the prototype and zero-centred peanut, described by the following equations:

$$x_1(t) = f_0(t)\cos t, \qquad x_2(t) = f_0(t)\sin t, \qquad t \in [0, 2\pi], \tag{3.113}$$

where

$$f_0(t) = \sqrt{\cos^2 t + 4\sin^2 t}. \tag{3.114}$$

The exact far-field matrix $\mathbf{F}$ is computed, from time to time, by means of the Nyström method [27] in the case of $N = 32$ incidence and observation angles, and 0.5% of Gaussian noise is added by means of a suitable 32×32 noise matrix $\mathbf{H}$ summed to $\mathbf{F}$ (see relation (2.266) and remark 2.5.1). The wavenumber is $k = 1$, i.e. $\lambda/4 \simeq 1.57$. We define the distance d between the two objects as the distance between the closest points of the two boundaries. Then we consider four cases: $d = 5.3, 2.3, 1.8, 1.3$, i.e. in two cases the distance between the objects is bigger than the Nyquist distance, in one case it is very close to (but still bigger than) the Nyquist distance and in the last case it is slightly smaller. We compute the indicator function $\Psi_2(z)$ (3.74) with $\alpha^* = \alpha_1^*$ provided by the generalized discrepancy principle in the incompatible case[19] (cf. (3.66)) and determine the optimal level curves in such a way that the sum of the areas of the true scatterers and the sum of the areas described by the level curves are equal. The results of this experiment are plotted in figure B.15, where our theoretical assessment of the spatial resolution is confirmed by the computational outcome.

Remark 3.3.1. All the previous results have been obtained by using a specific form for the indicator function, i.e. for the monotonic function I. In principle, the resolution power of the

[18] Specified in the caption of figure B.15.

[19] The choice of the incompatible case in implementing the generalized discrepancy principle is explained by the low noise level: cf. discussion in subsection 1.8.4.

method depends on the form of I. By instance, an easy way to indefinitely widen the band is to choose $I(t) = t^{2n}$ instead of $I(t) = t^2$, with an arbitrarily large $n \in \mathbb{N}\setminus\{0\}$, so that our indicator function $\Psi_2(z)$, given by (3.74), is to be replaced by $\Psi_{2n}(z) \equiv \Psi_2^n(z)$. However this implies that, in general, also the range of $\Psi_2^n(z)$ becomes arbitrarily large, so that an appropriate visualization of a 3D plot of $\Psi_2^n(z)$ would require a (not necessarily linear) rescaling of the Cartesian axis perpendicular to the z_1Oz_2-plane (and this procedure would be essentially equivalent to utilize a different indicator function with a smaller n). From a 2D perspective, after a cut-off criterion for the indicator function is applied, we observe that, despite regularization, random oscillations due to the presence of noise still affect the regularized solution $\mathbf{g}_{\alpha^*}(z)$ and, consequently, its Euclidean norm $\|\mathbf{g}_{\alpha^*}(z)\|_{\mathbb{C}^N}$. Hence, for increasing values of n the correspondently decreasing Nyquist distances tend to become comparable or smaller than the length scale on which such oscillations deteriorate the detected scatterer profile. In other terms, if n is too large, the theoretically estimated resolution power (heuristically identified with the Nyquist distance itself) may become unrealistic, since it concerns length scales on which the smoothing effect of the regularization procedure is not completely satisfactory.

Anyway, in general, the implementation of the linear sampling method proposed in section 3.1 allows one to compute the analytical form for all possible indicator functions. From this, it is possible to infer information on the Fourier transform and therefore on the achievable resolution. By instance, for the "traditional" choice $\sqrt{\Psi_2(z)}$ for the indicator function, it is easy to show that the Nyquist distance is $\lambda/2$.

Summing up, in sections 3.1 and 3.2 we have chosen $\Psi_2(z)$, given by (3.74), as indicator function, since it involves feasible analytic computations and consequently provides a detailed and fully worked out example of the potential applications of our "no-sampling" implementation. However, $\Psi_2(z)$ is not necessarily the best indicator function, as well as our "equal-areas" cut-off criterion is not necessarily the best way to detect the scatterer profile. □

3.4. Using a new family of indicator functions

Both the traditional and the no-sampling implementations of the linear sampling method get their inspiration from the general theorem 2.4.10, which states, in particular, that, under suitable hypotheses, an approximate solution for the far-field equation exists whose $L^2[0, 2\pi]$-norm blows up to infinity for all points approaching the boundary of the scatterer from inside and stays arbitrarily large outside. However, one should say, more precisely, that both the implementations are based on two particular statements of the general theorem 2.4.10, i.e. the pairs of relations (2.215)-(2.216) and (2.218)-(2.219); in such a way, we have completely disregarded limits (2.217) and (2.220), which are part too of the statement of theorem 2.4.10 itself.

The reason for this exclusion is evident: indeed, relations (2.217) and (2.220) require the computation of the $H^1(D)$-norm of a suitable Herglotz wave function: but the domain D, where the scatterer is placed, a priori (and, above all, in real experiments) is obviously unknown, then such a computation would be actually impossible.

In the following, we propose a way to overcome such a drawback and, consequently, to use also Herglotz wave functions in implementing the linear sampling method.

To this end, we firstly recall[20] that, for any $g \in L^2[0, 2\pi]$, the Herglotz wave function of kernel g

$$v_g(x) := \int_0^{2\pi} e^{\mathrm{i}kx\cdot d} g(\theta) d\theta \qquad \forall x \in \mathbb{R}^2 \tag{3.115}$$

(with $d = (\cos\theta, \sin\theta)$) is real-analytic on $\mathbb{R}^2$ and, consequently, it is also an element of $H^1(E)$ for any non-empty, bounded, open and Lebesgue measurable subset $E \subset \mathbb{R}^2$. In particular, we can choose $E = T_A^B$, where, as in section 3.1, T_A^B is a known rectangle containing the unknown domain D; then it clearly holds:

$$\|v_g\|_{H^1(D)} \leq \|v_g\|_{H^1(T_A^B)}. \tag{3.116}$$

Hence:

1. if $z \in D$, we can consider the function $g_z^\varepsilon \in L^2[0, 2\pi]$ of statement No 1 in theorem 2.4.10: then we have that $v_{g_z^\varepsilon} \in H^1(T_A^B)$ and, remembering relations (2.217) and (3.116), it holds:

$$\lim_{z \to \partial D} \left\| v_{g_z^\varepsilon} \right\|_{H^1(T_A^B)} = \infty; \tag{3.117}$$

2. if $z \in \mathbb{R}^2 \setminus \bar{D}$, we can consider the function $g_z^{\varepsilon,\delta} \in L^2[0, 2\pi]$ of statement No 2 in theorem 2.4.10: then we have that $v_{g_z^{\varepsilon,\delta}} \in H^1(T_A^B)$ and, remembering relations (2.220) and (3.116), it holds:

$$\lim_{\delta \to 0^+} \left\| v_{g_z^{\varepsilon,\delta}} \right\|_{H^1(T_A^B)} = \infty. \tag{3.118}$$

Now, relations (3.117) and (3.118) suggest two new possible implementations of the linear sampling method:

(i) the first one consists in reproducing, as far as possible, the traditional implementation: in such a case, one considers the Tikhonov pointwise-regularized solution $\mathbf{g}_{\alpha^*(z_l)}(z_l)$, as given by expression[21] (2.277), of the discretized far-field equation (2.272). Then, putting, as usual, $\theta_j = \dfrac{2\pi j}{N}$, $d_j = (\cos\theta_j, \sin\theta_j)$ and denoting with $g^j_{\alpha^*(z_l)}(z_l)$ the j-th component of

[20] See definition (2.98) and the brief comment soon below.

[21] Obviously, with $\alpha(z_l)$ replaced by $\alpha^*(z_l)$.

the vector $\mathbf{g}_{\alpha^*(z_l)}(z_l) \in \mathbb{C}^N$ with respect to the canonical basis $\{\mathbf{e}_j\}_{j=0}^{N-1}$ of $\mathbb{R}^N$, one defines the following function of $z_l \in \mathcal{Z}$:

$$\left\| v_{[\mathbf{g}_{\alpha^*(z_l)}(z_l)]}(\cdot) \right\|_{H^1(T_A^B)} := \left\| \frac{2\pi}{N} \sum_{j=0}^{N-1} e^{\mathrm{i}k(\cdot,d_j)_{\mathbb{R}^2}} g^j_{\alpha^*(z_l)}(z_l) \right\|_{H^1(T_A^B)} \qquad \forall z_l \in \mathcal{Z}, \tag{3.119}$$

where we have denoted with $(\cdot, d_j)_{\mathbb{R}^2}$ the scalar product in $\mathbb{R}^2$ between the independent variable[22] (i.e. $x \in \mathbb{R}^2$) and the vector $d_j \in \mathbb{R}^2$. Then, the final step is to consider, as indicator function, the following one:

$$\begin{aligned} &\Xi_I : \mathcal{Z} \longrightarrow \mathbb{R} \\ &\qquad z_l \longmapsto \Xi_I(z_l) := I\left(\left\| v_{[\mathbf{g}_{\alpha^*(z_l)}(z_l)]}(\cdot) \right\|_{H^1(T_A^B)} \right), \end{aligned} \tag{3.120}$$

where, as usual, $I : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ is a suitable monotonic continuous function.

(ii) the second one consists in following the no-sampling approach: in such a case, one considers the Tikhonov no-sampling-regularized solution $\mathbf{g}_{\alpha^*}(\cdot)$, as given by (3.69), of the functional equation (3.44). Then, putting again $\theta_j = \dfrac{2\pi j}{N}$, $d_j = (\cos\theta_j, \sin\theta_j)$ and denoting with $g^j_{\alpha^*}(z)$ the value in z of the j-th component of the N-tuple $\mathbf{g}_{\alpha^*}(\cdot) \in \left[L^2(T_A^B)\right]^N$, one defines the following function of $z \in T_A^B$:

$$\left\| v_{[\mathbf{g}_{\alpha^*}(z)]}(\cdot) \right\|_{H^1(T_A^B)} := \left\| \frac{2\pi}{N} \sum_{j=0}^{N-1} e^{\mathrm{i}k(\cdot,d_j)_{\mathbb{R}^2}} g^j_{\alpha^*}(z) \right\|_{H^1(T_A^B)} \qquad \forall z \in T_A^B, \tag{3.121}$$

where, as before, we have denoted with $(\cdot, d_j)_{\mathbb{R}^2}$ the scalar product in $\mathbb{R}^2$ between the independent variable (i.e. $x \in \mathbb{R}^2$) of the function whose $H^1(T_A^B)$-norm is to be computed and the vector $d_j \in \mathbb{R}^2$. Then, the final step is to consider, as indicator function, the following one:

$$\begin{aligned} &\Xi_I : T_A^B \longrightarrow \mathbb{R} \\ &\qquad z \longmapsto \Xi_I(z) := I\left(\left\| v_{[\mathbf{g}_{\alpha^*}(z)]}(\cdot) \right\|_{H^1(T_A^B)} \right), \end{aligned} \tag{3.122}$$

where, as usual, $I : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ is a suitable monotonic continuous function.

A comparison between the two new implementations (i) and (ii) would clearly require that one made several numerical experiments changing, from time to time, the scatterer, the noise level, the boundary conditions, the wavenumber, the number of incidence/observation angles, the indicator function, the criterion for its cut-off and so on. At present, we cannot show and

[22] The dot "$\cdot$" instead of x points out which the independent variable is for the function whose $H^1(T_A^B)$-norm is to be computed.

discuss any numerical experiment concerning such a comparison: however, by virtue of the substantial equivalence between the traditional and the previous no-sampling implementation already observed at the end of section 3.1, it is reasonable to assume that such an equivalence also holds for the implementations (i) and (ii) we are now considering, all the more that the two different no-sampling implementations we are now going to compare, one using $\Psi_2(z)$ and the other one using $\Xi_2(z)$ as indicator functions respectively, will prove, in turn, to be substantially equivalent[23].

To this end, we are now going to compute the Fourier transform of the indicator function (3.122) for $I(t) = t^2$, but, for sake of simplicity, we shall replace the $H^1(T_A^B)$-norm with the $W^{1,2}(T_A^B)$-norm: this is justified by the equivalence of the two norms (see theorem A.13.2), which implies that the two limits (3.117), (3.118) inspiring the current alternative no-sampling implementation also hold when, instead of the $H^1(T_A^B)$-norm, one considers the $W^{1,2}(T_A^B)$-norm.

To make such a computation, we firstly need to obtain a handy analytic expression of the right-hand side of definition (3.121) (rewritten using the $W^{1,2}(T_A^B)$-norm) as a function of z. We begin by putting:

$$f(x;z) := \frac{2\pi}{N} \sum_{j=0}^{N-1} e^{\mathrm{i}k(x,d_j)_{\mathbb{R}^2}} g_{\alpha^*}^j(z) \qquad \forall x, z \in T_A^B; \tag{3.123}$$

then, recalling relation (3.69) and remembering that $u_{p,j}^h = (\mathbf{u}_p^h, \mathbf{e}_j)_{\mathbb{C}^N}$, we have:

$$\begin{aligned} f(x;z) &= \frac{2\pi}{N} \sum_{j=0}^{N-1} e^{\mathrm{i}k(x,d_j)_{\mathbb{R}^2}} (\mathbf{g}_{\alpha^*}(z), \mathbf{e}_j)_{\mathbb{C}^N} = \\ &= \sum_{j=0}^{N-1} e^{\mathrm{i}k(x,d_j)_{\mathbb{R}^2}} \sum_{p=0}^{r_h-1} \frac{\sigma_p^h}{(\sigma_p^h)^2 + \alpha^*} (\mathbf{\Phi}_\infty(z), \mathbf{v}_p^h)_{\mathbb{C}^N} (\mathbf{u}_p^h, \mathbf{e}_j)_{\mathbb{C}^N} = \\ &= \sum_{j=0}^{N-1} e^{\mathrm{i}k(x,d_j)_{\mathbb{R}^2}} \sum_{p=0}^{r_h-1} \frac{u_{p,j}^h \, \sigma_p^h}{(\sigma_p^h)^2 + \alpha^*} (\mathbf{\Phi}_\infty(z), \mathbf{v}_p^h)_{\mathbb{C}^N} = \\ &= \sum_{j=0}^{N-1} e^{\mathrm{i}k(x_1 \cos\theta_j + x_2 \sin\theta_j)} \, a_j^h(z), \end{aligned} \tag{3.124}$$

where, in the last passage, we have written in explicit form the scalar product in the exponential and we have put:

$$a_j^h(z) := \sum_{p=0}^{r_h-1} \frac{u_{p,j}^h \, \sigma_p^h}{(\sigma_p^h)^2 + \alpha^*} (\mathbf{\Phi}_\infty(z), \mathbf{v}_p^h)_{\mathbb{C}^N} \qquad \forall j = 0, \ldots, N-1, \quad \forall z \in T_A^B. \tag{3.125}$$

[23] Compare also figures (B.15) and (B.16): the latter will be introduced at the end of the current section.

From equality (3.124) we easily get:

$$\frac{\partial f}{\partial x_1}(x;z) = \sum_{j=0}^{N-1} \mathrm{i}k\cos\theta_j\, e^{\mathrm{i}k(x_1\cos\theta_j + x_2\sin\theta_j)}\, a_j^h(z) = \mathrm{i}k \sum_{j=0}^{N-1} e^{\mathrm{i}k(x_1\cos\theta_j + x_2\sin\theta_j)}\, b_j^h(z), \tag{3.126}$$

$$\frac{\partial f}{\partial x_2}(x;z) = \sum_{j=0}^{N-1} \mathrm{i}k\sin\theta_j\, e^{\mathrm{i}k(x_1\cos\theta_j + x_2\sin\theta_j)}\, a_j^h(z) = \mathrm{i}k \sum_{j=0}^{N-1} e^{\mathrm{i}k(x_1\cos\theta_j + x_2\sin\theta_j)}\, c_j^h(z), \tag{3.127}$$

having obviously put:

$$b_j^h(z) := \cos\theta_j\, a_j^h(z), \qquad c_j^h(z) := \sin\theta_j\, a_j^h(z) \qquad \forall j = 0,\dots,N-1, \quad \forall z \in T_A^B. \tag{3.128}$$

From equalities (3.124), (3.126) and (3.127) it easily follows:

$$|f(x;z)|^2 = \left| \sum_{j=0}^{N-1} e^{\mathrm{i}k(x_1\cos\theta_j + x_2\sin\theta_j)}\, a_j^h(z) \right|^2, \tag{3.129}$$

$$\left| \frac{\partial f}{\partial x_1}(x;z) \right|^2 = k^2 \left| \sum_{j=0}^{N-1} e^{\mathrm{i}k(x_1\cos\theta_j + x_2\sin\theta_j)}\, b_j^h(z) \right|^2, \tag{3.130}$$

$$\left| \frac{\partial f}{\partial x_2}(x;z) \right|^2 = k^2 \left| \sum_{j=0}^{N-1} e^{\mathrm{i}k(x_1\cos\theta_j + x_2\sin\theta_j)}\, c_j^h(z) \right|^2. \tag{3.131}$$

Now, by virtue of relations (3.121), (3.123) and remembering the definition (A.66) of $W^{1,2}(T_A^B)$-norm, it holds:

$$\|v_{[\mathbf{g}_{\alpha^*}(z)]}(\cdot)\|^2_{W^{1,2}(T_A^B)} = \int_{T_A^B} |f(x;z)|^2\, dx + \int_{T_A^B} \left| \frac{\partial f}{\partial x_1}(x;z) \right|^2 dx + \int_{T_A^B} \left| \frac{\partial f}{\partial x_2}(x;z) \right|^2 dx, \tag{3.132}$$

i.e., remembering equalities (3.129), (3.130) and (3.131),

$$\|v_{[\mathbf{g}_{\alpha^*}(z)]}(\cdot)\|^2_{W^{1,2}(T_A^B)} = \int_{T_A^B} \left| \sum_{j=0}^{N-1} e^{\mathrm{i}k(x_1\cos\theta_j + x_2\sin\theta_j)}\, a_j^h(z) \right|^2 dx + \tag{3.133}$$

$$+ k^2 \int_{T_A^B} \left| \sum_{j=0}^{N-1} e^{\mathrm{i}k(x_1\cos\theta_j + x_2\sin\theta_j)}\, b_j^h(z) \right|^2 dx + k^2 \int_{T_A^B} \left| \sum_{j=0}^{N-1} e^{\mathrm{i}k(x_1\cos\theta_j + x_2\sin\theta_j)}\, c_j^h(z) \right|^2 dx.$$

Let us now focus on the first integral at the right-hand side of relation (3.133): we can easily realize that, for each $z \in T_A^B$, the algebraic structure of the function to be integrated is essentially the same of the square modulus in (3.78); hence, if we use polar coordinates (r, β) for all the coefficients $a_j^h(z)$, i.e. if we put:

$$a_j^h(z) = (r_j^h(z), \beta_j^h(z)) = r_j^h(z)\, e^{\mathrm{i}\beta_j^h(z)} \qquad \forall j = 0,\dots,N-1, \quad \forall z \in T_A^B, \tag{3.134}$$

we can make computations analogous to the ones that allowed us to pass from (3.78) to the internal sum inside representation (3.84): this means that, by putting

$$\tilde{\omega}_{1,ij} := k(\cos\theta_i - \cos\theta_j), \tag{3.135}$$

$$\tilde{\omega}_{2,ij} := k(\sin\theta_i - \sin\theta_j), \tag{3.136}$$

$$\Delta\beta^h_{ij}(z) := \beta^h_i(z) - \beta^h_j(z), \tag{3.137}$$

it holds:

$$\left|\sum_{j=0}^{N-1} e^{\mathrm{i}k(x_1\cos\theta_j + x_2\sin\theta_j)}\, a^h_j(z)\right|^2 = \sum_{i,j=0}^{N-1} r^h_i(z) r^h_j(z) \cos\left(\tilde{\omega}_{1,ij}\, x_1 + \tilde{\omega}_{2,ij}\, x_2 + \Delta\beta^h_{ij}(z)\right), \tag{3.138}$$

i.e., recalling equalities (3.129) and (3.90),

$$\begin{aligned} |f(x;z)|^2 = \sum_{i,j=0}^{N-1} & r^h_i(z) r^h_j(z) \cdot \\ & \cdot \left\{\cos\left(\Delta\beta^h_{ij}(z)\right)\left[\cos\left(\tilde{\omega}_{1,ij}\, x_1\right)\cos\left(\tilde{\omega}_{2,ij}\, x_2\right) - \sin\left(\tilde{\omega}_{1,ij}\, x_1\right)\sin\left(\tilde{\omega}_{2,ij}\, x_2\right)\right] + \right. \\ & \left. - \sin\left(\Delta\beta^h_{ij}(z)\right)\left[\sin\left(\tilde{\omega}_{1,ij}\, x_1\right)\cos\left(\tilde{\omega}_{2,ij}\, x_2\right) + \cos\left(\tilde{\omega}_{1,ij}\, x_1\right)\sin\left(\tilde{\omega}_{2,ij}\, x_2\right)\right]\right\}. \end{aligned} \tag{3.139}$$

Hence, by linearity and by Fubini's theorem (recalling that $T^B_A = (-A, A) \times (-B, B)$), in order to compute $\int\limits_{T^B_A} |f(x;z)|^2\, dx$ it suffices to determine the values of the following integrals:

$$\tilde{A}_{ij} := \int_{-A}^{A} \cos\left(\tilde{\omega}_{1,ij}\, x_1\right) dx_1 \int_{-B}^{B} \cos\left(\tilde{\omega}_{2,ij}\, x_2\right) dx_2, \tag{3.140}$$

$$\tilde{B}_{ij} := \int_{-A}^{A} \sin\left(\tilde{\omega}_{1,ij}\, x_1\right) dx_1 \int_{-B}^{B} \sin\left(\tilde{\omega}_{2,ij}\, x_2\right) dx_2, \tag{3.141}$$

$$\tilde{C}_{ij} := \int_{-A}^{A} \sin\left(\tilde{\omega}_{1,ij}\, x_1\right) dx_1 \int_{-B}^{B} \cos\left(\tilde{\omega}_{2,ij}\, x_2\right) dx_2, \tag{3.142}$$

$$\tilde{D}_{ij} := \int_{-A}^{A} \cos\left(\tilde{\omega}_{1,ij}\, x_1\right) dx_1 \int_{-B}^{B} \sin\left(\tilde{\omega}_{2,ij}\, x_2\right) dx_2. \tag{3.143}$$

Since the sine function is odd, while the integration domains $(-A, A)$ and $(-B, B)$ are even, we immediately conclude that

$$\tilde{B}_{ij} = \tilde{C}_{ij} = \tilde{D}_{ij} = 0 \qquad \forall i, j = 0, \ldots, N-1; \tag{3.144}$$

as regards the quantities $\tilde{A}_{ij}$, they are, in general, different from zero (e.g. when $i = j$) and their values are easily computable. Then, by virtue of relations (3.139), (3.140)-(3.143) and (3.144), we get:

$$\int\limits_{T_A^B} |f(x;z)|^2 \, dx = \sum_{i,j=0}^{N-1} r_i^h(z) r_j^h(z) \tilde{A}_{ij} \cos\left(\Delta\beta_{ij}^h(z)\right), \tag{3.145}$$

i.e., remembering definition (3.137) and identity (3.80),

$$\int\limits_{T_A^B} |f(x;z)|^2 \, dx = \sum_{i,j=0}^{N-1} \tilde{A}_{ij} \left[r_i^h(z) \cos\beta_i^h(z) \, r_j^h(z) \cos\beta_j^h(z) + r_i^h(z) \sin\beta_i^h(z) \, r_j^h(z) \sin\beta_j^h(z) \right], \tag{3.146}$$

or, equivalently, recalling the polar coordinates representation (3.134),

$$\int\limits_{T_A^B} |f(x;z)|^2 \, dx = \sum_{i,j=0}^{N-1} \tilde{A}_{ij} \left[\mathrm{Re}\{a_i^h(z)\} \, \mathrm{Re}\{a_j^h(z)\} + \mathrm{Im}\{a_i^h(z)\} \, \mathrm{Im}\{a_j^h(z)\} \right]. \tag{3.147}$$

In order to compute the real and imaginary parts of the complex numbers $a_i^h(z)$, let us firstly introduce, analogously to (3.77), polar coordinates for the N components of each vector $\mathbf{u}_p^h$, i.e.

$$\begin{aligned} u_{p,0}^h &= (\lambda_{p,0}^h, \gamma_{p,0}^h) = \lambda_{p,0}^h e^{\mathrm{i}\gamma_{p,0}^h}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots, \\ u_{p,N-1}^h &= (\lambda_{p,N-1}^h, \gamma_{p,N-1}^h) = \lambda_{p,N-1}^h e^{\mathrm{i}\gamma_{p,N-1}^h}; \end{aligned} \tag{3.148}$$

then, remembering definition (3.125) and using relations (3.75), (3.77), (3.148), we can write:

$$\begin{aligned} a_i^h(z) &= \sum_{p=0}^{r_h-1} \frac{\lambda_{p,i}^h \, \sigma_p^h}{(\sigma_p^h)^2 + \alpha^*} \, e^{\mathrm{i}\gamma_{p,i}^h} (\mathbf{\Phi}_\infty(z), \mathbf{v}_p^h)_{\mathbb{C}^N} = \\ &= \sum_{p=0}^{r_h-1} \frac{\lambda_{p,i}^h \, \sigma_p^h}{(\sigma_p^h)^2 + \alpha^*} \, e^{\mathrm{i}\gamma_{p,i}^h} \left(\bar{v}_{p,0}^h e^{-\mathrm{i}k(z_1\cos\varphi_0 + z_2\sin\varphi_0)} + \cdots + \bar{v}_{p,N-1}^h e^{-\mathrm{i}k(z_1\cos\varphi_{N-1} + z_2\sin\varphi_{N-1})} \right) = \\ &= \sum_{p=0}^{r_h-1} \frac{\lambda_{p,i}^h \, \sigma_p^h}{(\sigma_p^h)^2 + \alpha^*} \, e^{\mathrm{i}\gamma_{p,i}^h} \sum_{s=0}^{N-1} \rho_{p,s}^h \, e^{-\mathrm{i}\varepsilon_{p,s}^h} e^{-\mathrm{i}k(z_1\cos\varphi_s + z_2\sin\varphi_s)} = \\ &= \sum_{p=0}^{r_h-1} \frac{\lambda_{p,i}^h \, \sigma_p^h}{(\sigma_p^h)^2 + \alpha^*} \sum_{s=0}^{N-1} \rho_{p,s}^h \, e^{-\mathrm{i}(\varepsilon_{p,s}^h - \gamma_{p,i}^h)} e^{-\mathrm{i}k(z_1\cos\varphi_s + z_2\sin\varphi_s)} = \\ &= \sum_{p=0}^{r_h-1} \frac{\lambda_{p,i}^h \, \sigma_p^h}{(\sigma_p^h)^2 + \alpha^*} \sum_{s=0}^{N-1} \rho_{p,s}^h \, e^{-\mathrm{i}\left[k(z_1\cos\varphi_s + z_2\sin\varphi_s) + \varepsilon_{p,s}^h - \gamma_{p,i}^h\right]}. \end{aligned} \tag{3.149}$$

Equality (3.149) clearly implies that

$$\mathrm{Re}\left\{a_i^h(z)\right\} = \sum_{p=0}^{r_h-1} \frac{\lambda_{p,i}^h \, \sigma_p^h}{(\sigma_p^h)^2 + \alpha^*} \sum_{s=0}^{N-1} \rho_{p,s}^h \, \cos\left[k(z_1\cos\varphi_s + z_2\sin\varphi_s) + \varepsilon_{p,s}^h - \gamma_{p,i}^h\right], \tag{3.150}$$

$$\mathrm{Im}\left\{a_i^h(z)\right\} = -\sum_{p=0}^{r_h-1} \frac{\lambda_{p,i}^h\,\sigma_p^h}{(\sigma_p^h)^2+\alpha^*} \sum_{s=0}^{N-1} \rho_{p,s}^h\,\sin\left[k(z_1\cos\varphi_s + z_2\sin\varphi_s) + \varepsilon_{p,s}^h - \gamma_{p,i}^h\right]. \tag{3.151}$$

By virtue of the previous relations (3.150), (3.151) and remembering identity (3.80), as well as definitions (3.81), (3.82), we get:

$$\begin{aligned}
&\mathrm{Re}\{a_i^h(z)\}\,\mathrm{Re}\{a_j^h(z)\} + \mathrm{Im}\{a_i^h(z)\}\,\mathrm{Im}\{a_j^h(z)\} = \\
&= \sum_{p,q=0}^{r_h-1} \frac{\lambda_{p,i}^h\,\sigma_p^h}{(\sigma_p^h)^2+\alpha^*} \frac{\lambda_{q,j}^h\,\sigma_q^h}{(\sigma_q^h)^2+\alpha^*} \sum_{s,t=0}^{N-1} \rho_{p,s}^h\,\rho_{q,t}^h \cdot \\
&\qquad \left\{\cos\left[k(z_1\cos\varphi_s + z_2\sin\varphi_s) + \varepsilon_{p,s}^h - \gamma_{p,i}^h\right]\cos\left[k(z_1\cos\varphi_t + z_2\sin\varphi_t) + \varepsilon_{q,t}^h - \gamma_{q,j}^h\right] + \right. \\
&\qquad \left. + \sin\left[k(z_1\cos\varphi_s + z_2\sin\varphi_s) + \varepsilon_{p,s}^h - \gamma_{p,i}^h\right]\sin\left[k(z_1\cos\varphi_t + z_2\sin\varphi_t) + \varepsilon_{q,t}^h - \gamma_{q,j}^h\right]\right\} = \\
&= \sum_{p,q=0}^{r_h-1} \frac{\lambda_{p,i}^h\,\sigma_p^h}{(\sigma_p^h)^2+\alpha^*} \frac{\lambda_{q,j}^h\,\sigma_q^h}{(\sigma_q^h)^2+\alpha^*} \sum_{s,t=0}^{N-1} \rho_{p,s}^h\,\rho_{q,t}^h \cdot \\
&\qquad \cos\left\{k\left[z_1(\cos\varphi_s - \cos\varphi_t) + z_2(\sin\varphi_s - \sin\varphi_t)\right] + \eta_{p,si}^h - \eta_{q,tj}^h\right\} = \\
&= \sum_{p,q=0}^{r_h-1} \frac{\lambda_{p,i}^h\,\sigma_p^h}{(\sigma_p^h)^2+\alpha^*} \frac{\lambda_{q,j}^h\,\sigma_q^h}{(\sigma_q^h)^2+\alpha^*} \sum_{s,t=0}^{N-1} \rho_{p,s}^h\,\rho_{q,t}^h \cos\left(\hat\omega_{1,st}\,z_1 + \hat\omega_{2,st}\,z_2 + \Delta\eta_{pq,sitj}^h\right),
\end{aligned} \tag{3.152}$$

where, in the last two passages, we have obviously put:

$$\eta_{p,si}^h := \varepsilon_{p,s}^h - \gamma_{p,i}^h, \qquad \eta_{q,tj}^h := \varepsilon_{q,t}^h - \gamma_{q,j}^h, \qquad \Delta\eta_{pq,sitj}^h := \eta_{p,si}^h - \eta_{q,tj}^h. \tag{3.153}$$

Hence, substituting relation (3.152) into (3.147), we find:

$$\begin{aligned}
&\int_{T_A^B} |f(x;z)|^2\,dx = \\
&\qquad = \sum_{i,j=0}^{N-1} \tilde A_{ij} \sum_{p,q=0}^{r_h-1} \frac{\lambda_{p,i}^h\,\sigma_p^h}{(\sigma_p^h)^2+\alpha^*} \frac{\lambda_{q,j}^h\,\sigma_q^h}{(\sigma_q^h)^2+\alpha^*} \sum_{s,t=0}^{N-1} \rho_{p,s}^h\,\rho_{q,t}^h \cos\left(\hat\omega_{1,st}\,z_1 + \hat\omega_{2,st}\,z_2 + \Delta\eta_{pq,sitj}^h\right),
\end{aligned} \tag{3.154}$$

i.e., recalling equality (3.90),

$$\begin{aligned}
\int_{T_A^B} |f(x;z)|^2\,dx = \sum_{i,j=0}^{N-1} \tilde A_{ij} \sum_{p,q=0}^{r_h-1} \frac{\lambda_{p,i}^h\,\sigma_p^h}{(\sigma_p^h)^2+\alpha^*} \frac{\lambda_{q,j}^h\,\sigma_q^h}{(\sigma_q^h)^2+\alpha^*} \sum_{s,t=0}^{N-1} \rho_{p,s}^h\,\rho_{q,t}^h \cdot \\
\cdot \left\{\cos\left(\Delta\eta_{pq,sitj}^h\right)\left[\cos\left(\hat\omega_{1,st}\,z_1\right)\,\cos\left(\hat\omega_{2,st}\,z_2\right) - \sin\left(\hat\omega_{1,st}\,z_1\right)\,\sin\left(\hat\omega_{2,st}\,z_2\right)\right] + \right. \\
\left. - \sin\left(\Delta\eta_{pq,sitj}^h\right)\left[\sin\left(\hat\omega_{1,st}\,z_1\right)\,\cos\left(\hat\omega_{2,st}\,z_2\right) + \cos\left(\hat\omega_{1,st}\,z_1\right)\,\sin\left(\hat\omega_{2,st}\,z_2\right)\right]\right\}.
\end{aligned} \tag{3.155}$$

The second and the third integral at the right-hand side of relation (3.133) are treated in a way which is fully analogous to that followed for the first one; in particular, remembering

definitions (3.128) and relations (3.130), (3.131), one easily realizes that the two following representations hold:

$$\int\limits_{T_A^B} \left| \frac{\partial f}{\partial x_1}(x;z) \right|^2 dx = k^2 \sum_{i,j=0}^{N-1} \tilde{A}_{ij} \cos\theta_i \, \cos\theta_j \sum_{p,q=0}^{r_h-1} \frac{\lambda_{p,i}^h \, \sigma_p^h}{(\sigma_p^h)^2 + \alpha^*} \, \frac{\lambda_{q,j}^h \, \sigma_q^h}{(\sigma_q^h)^2 + \alpha^*} \sum_{s,t=0}^{N-1} \rho_{p,s}^h \, \rho_{q,t}^h \cdot \tag{3.156}$$

$$\cdot \left\{ \cos\left(\Delta\eta_{pq,sitj}^h\right) \left[\cos\left(\hat{\omega}_{1,st} \, z_1\right) \, \cos\left(\hat{\omega}_{2,st} \, z_2\right) - \sin\left(\hat{\omega}_{1,st} \, z_1\right) \, \sin\left(\hat{\omega}_{2,st} \, z_2\right) \right] + \right.$$
$$\left. - \sin\left(\Delta\eta_{pq,sitj}^h\right) \left[\sin\left(\hat{\omega}_{1,st} \, z_1\right) \, \cos\left(\hat{\omega}_{2,st} \, z_2\right) + \cos\left(\hat{\omega}_{1,st} \, z_1\right) \, \sin\left(\hat{\omega}_{2,st} \, z_2\right) \right] \right\} .$$

$$\int\limits_{T_A^B} \left| \frac{\partial f}{\partial x_2}(x;z) \right|^2 dx = k^2 \sum_{i,j=0}^{N-1} \tilde{A}_{ij} \sin\theta_i \, \sin\theta_j \sum_{p,q=0}^{r_h-1} \frac{\lambda_{p,i}^h \, \sigma_p^h}{(\sigma_p^h)^2 + \alpha^*} \, \frac{\lambda_{q,j}^h \, \sigma_q^h}{(\sigma_q^h)^2 + \alpha^*} \sum_{s,t=0}^{N-1} \rho_{p,s}^h \, \rho_{q,t}^h \cdot \tag{3.157}$$

$$\cdot \left\{ \cos\left(\Delta\eta_{pq,sitj}^h\right) \left[\cos\left(\hat{\omega}_{1,st} \, z_1\right) \, \cos\left(\hat{\omega}_{2,st} \, z_2\right) - \sin\left(\hat{\omega}_{1,st} \, z_1\right) \, \sin\left(\hat{\omega}_{2,st} \, z_2\right) \right] + \right.$$
$$\left. - \sin\left(\Delta\eta_{pq,sitj}^h\right) \left[\sin\left(\hat{\omega}_{1,st} \, z_1\right) \, \cos\left(\hat{\omega}_{2,st} \, z_2\right) + \cos\left(\hat{\omega}_{1,st} \, z_1\right) \, \sin\left(\hat{\omega}_{2,st} \, z_2\right) \right] \right\} .$$

Finally, if we substitute relations (3.155), (3.156) and (3.157) into (3.132), we find that

$$\Xi_2(z) := \left\| v_{[\mathbf{g}_{\alpha^*}(z)]}(\cdot) \right\|_{W^{1,2}(T_A^B)}^2 \tag{3.158}$$

consists in a finite linear combination of the four functions (3.91)-(3.94), just as it happened for $\Psi_2(z) = \|\mathbf{g}_{\alpha^*}(z)\|_{\mathbb{C}^N}^2$; although the two linear combinations are very different in the two cases (the one for $\Xi_2(z)$ is obviously much more involved), this analogy easily implies that the supports of the Fourier transforms of $\Xi_2(z)$ and $\Psi_2(z)$ are the same[24]: in other terms, $[\mathcal{F}(\Xi_2)](\omega_1, \omega_2)$ and $[\mathcal{F}(\Psi_2)](\omega_1, \omega_2)$ are two different Dirac brushes having the same support, given by (3.106). This implies that, from the viewpoint of the resolution achievable, using $\Xi_2(z)$ or $\Psi_2(z)$ as indicator function should be equivalent. This is confirmed by a comparison between figures (B.15) and (B.16): the latter shows the results obtained by repeating exactly the same numerical experiment we made to create figure (B.15), except that the indicator function is now $\Xi_2(z)$ instead of $\Psi_2(z)$.

3.5. Facing the cut-off problem: deformable models

In this section we want to test the application of deformable contour models to the visualization maps provided by the linear sampling method in order to extract the scatterer profile: indeed, we think that this is a very promising approach to the problem of finding a cut-off criterion for the plots of the indicator function provided by the method itself.

In the last twenty years there has been an increasing research activity in order to be able to reconstruct the boundary of an object starting from a given image reproducing somehow

[24] Obviously, like $\Psi_2(z)$, also $\Xi_2(z)$ can be analytically continued onto all $\mathbb{R}^2$.

the object itself; although there are now several works on edge detection techniques or active contours, we shall just cite the following three: [23], [24], [42], to which we refer for a much deeper and more complete treatment than the very brief one we are going to outline in the following.

A deformable contour is a curve[25] $\gamma_0 : [0,1] \to \mathbb{R}^2$; a deformable model is a couple formed by a space A_d of admissible deformations of γ_0 and a functional $\mathcal{E} : A_d \to \mathbb{R}$ to be minimized. This functional represents the energy of the model and has the following form:

$$\mathcal{E}(\gamma) := \int_0^1 \left[\frac{1}{2} \left(w_1(s) \left\| \gamma'(s) \right\|^2_{\mathbb{R}^2} + w_2(s) \left\| \gamma''(s) \right\|^2_{\mathbb{R}^2} \right) + E_{ext}(\gamma(s)) \right] ds, \tag{3.159}$$

where the prime sign "$'$" denotes the ordinary derivative with respect to s and

1. the maps $w_1 : [0,1] \to \mathbb{R}$ and $w_2 : [0,1] \to \mathbb{R}$ are weight functions that respectively control the importance of the first-order and second-order terms imposing the regularity of the curve; their choice determines the mechanical properties or, more precisely, the internal forces, i.e. elasticity and rigidity respectively, of the model[26]

2. E_{ext} denotes the potential energy associated to the external forces deriving from the image map and pushing the curve to the significant lines which correspond to the desired attributes (i.e., in our case, edges); in our LSM-oriented application, we choose

$$E_{ext} := - \left\| \nabla_2 \Psi_I \right\|^2_{\mathbb{R}^2}, \tag{3.160}$$

where Ψ_I is clearly the generic indicator function (3.71) and ∇_2 denotes the gradient operator: indeed, if we want the deformable contour to be attracted by edge points, the potential should depend on the gradient of the image.

The functions w_1, w_2, γ and E_{ext} are assumed to be smooth enough, in such a way that all the previous and following computations involving them make sense from a classical viewpoint. Moreover, we shall restrict the space A_d of admissible deformations by assigning the boundary

[25] Although in the applications we have in mind the curve is *closed*, the general approach we are sketching does not require this hypothesis.

[26] The first-order term makes the deformable contour behave like an elastic string of zero rest-length, since $\frac{1}{2} \int_0^1 w_1(s) \left\| \gamma'(s) \right\|^2_{\mathbb{R}^2}$ is just the elastic energy of such a string, characterized by an elastic constant $w_1(s)$ depending on the point.

The second-order term imposes some rigidity to the string itself, which can be understood if we imagine that for a certain $\hat{s} \in [0,1]$ the string has a corner: this implies that $\gamma'(s)$ is discontinuous in $\hat{s}$ and $\gamma''(s)$ is a Dirac delta in $\hat{s}$, then such a corner is pointwise penalized by means of a term proportional to $w_2(\hat{s})$ itself (and of course no penalization is given if $w_2(\hat{s}) = 0$). In any case, we can say that the second-order term tends to make the deformable contour behave like a stick.

Summing up, tuning the weight functions w_1 and w_2 controls the relative importance of the elastic string and stick features respectively; the resulting overall behaviour is the one of a spline.

conditions $\gamma(0)$, $\gamma'(0)$, $\gamma(1)$ and $\gamma'(1)$; we can also use periodic curves or, in general, other kinds of boundary conditions.

Minimizing $\mathcal{E}(\gamma)$ is a variational problem: a necessary condition for γ to be a local minimum for the functional $\mathcal{E}(\gamma)$ is that [31] γ itself satisfies the Euler (vectorial) equation:

$$(w_1(s)\,\gamma'(s))' - (w_2(s)\,\gamma''(s))'' - \nabla_2 E_{ext}(\gamma(s)) = 0 \tag{3.161}$$

with given boundary conditions $\gamma(0)$, $\gamma'(0)$, $\gamma(1)$ and $\gamma'(1)$. Since, in general, the energy $\mathcal{E}(\gamma)$ is not convex, it may have several local minima. However, we point out that finding the global minimum of the energy does not necessarily have a meaning: indeed, e.g., if $P \in \mathbb{R}^2$ is a point of the plane where E_{ext} has a global minimum, then the degenerate constant curve $\gamma(s) = P$ is a global minimum for the energy with periodic boundary conditions. On the other hand, we are rather interested in finding a good contour in a given area: in fact, we suppose to have at disposal a rough estimate, i.e. γ_0, of the contour we want to reconstruct. Then, in order to determine the solution to (3.161) we are interested in, we assume that the unknown curve γ becomes dynamic, by artificially regarding it as a function of time t as well as of the original variable s, i.e. $\gamma = \gamma(s,t)$, so that the latter is a solution of the associated evolution problem:

$$\begin{cases} \dfrac{\partial}{\partial t}\gamma(s,t) = (w_1(s)\,\gamma'(s,t))' - (w_2(s)\,\gamma''(s,t))'' - \nabla_2 E_{ext}(\gamma(s,t)), & \text{(a)} \\ \gamma(s,0) = \gamma_0(s), & \text{(b)} \\ \gamma(0,t) = \gamma_0(0), \quad \gamma(1,t) = \gamma_0(1), & \text{(c)} \\ \gamma'(0,t) = \gamma_0'(0), \quad \gamma'(1,t) = \gamma_0'(0), & \text{(d)} \end{cases} \tag{3.162}$$

where the prime sign "$'$" is now to be read, in general, as $\frac{\partial}{\partial s}$, and the boundary or initial conditions (3.162)(b)-(d) impose that γ is "close" enough to the initial guess γ_0. When the solution γ of (3.162) stabilizes, the term $\frac{\partial \gamma}{\partial t}$ goes to zero and we get a solution of the static problem (3.161). Therefore a numerical solution to (3.161) can be found by discretizing, in both s and t, problem (3.162) and solving iteratively (in t) the discretized system so obtained, till a stable solution is found. An accurate implementation of this iterative procedure requires some clever devices and technicalities, which will be omitted here. We only point out that discretization may introduce numerical instabilities, which, however, can be reduced by replacing equation (3.162)(a) with the following one,

$$\frac{\partial}{\partial t}\gamma(s,t) = (w_1(s)\,\gamma'(s,t))' - (w_2(s)\,\gamma''(s,t))'' + F(\gamma(s,t)), \tag{3.163}$$

where

$$F := -\kappa \frac{\nabla_2 E_{ext}}{\|\nabla_2 E_{ext}\|_{\mathbb{R}^2}}, \tag{3.164}$$

for a suitable $\kappa \in \mathbb{R}^+$, which can be chosen in a standard way.

It is worthwhile noticing that, unlike the general case (in which one has at disposal only an image as a set of pixels), in the framework of the no-sampling implementation of the linear sampling method the external energy term $\nabla_2 E_{ext}$ does not need to be numerically computed by means of finite-increments methods, but can be analytically determined, thus increasing the accuracy of the procedure.

As numerical applications, we consider the following scattering experiments in the case of Dirichlet boundary conditions, with wavenumber $k = 1$, indicator function $\Psi_{-2\ln}(z) := -\ln \|\mathbf{g}_{\alpha^*}(z)\|^2_{\mathbb{C}^N}$ (cf. definition (3.72)) and $N = 32$ incidence/observation angles:

1. an ellipse having its centre in $(0, 0)$ and semiaxes equal to 1 and 2 respectively, with $n = 5\%$ of Gaussian noise added to the far-field matrix and with blended regularization[27] providing a value $\alpha_b^* = 1.2 \cdot 10^{-1}$ for the regularization parameter α (figure B.17, panel (a));

2. a kite described by the parametric equation (2.285), with $n = 5\%$ of Gaussian noise added to the far-field matrix and with blended regularization providing a value $\alpha_b^* = 1.1 \cdot 10^{-1}$ for the regularization parameter α (figure B.17, panel (b));

3. a double scatterer, i.e. the same kite described by equation (2.285) but centred in $(-4, -4)$ and rotated of 45° clockwise, together with an ellipse again with semiaxes 1 and 2 but centred in $(4, 4)$ and rotated of 45° counterclockwise; the noise level is $n = 10\%$ and the generalized discrepancy principle for the compatible case[28] is used, this providing a value $\alpha_2^* = 4.9 \cdot 10^{-2}$ for the regularization parameter α (figure B.18).

In each figure we show the true profile of the scatterer (white line), the initial guess γ_0 (red line) and reconstructed profile (blue line) obtained by applying the deformable model described just above in this section, i.e., more precisely, by iteratively solving problem (3.162) (with the corrections pointed out in (3.163)-(3.164)), where we have chosen (in a standard way[29]) constant weight functions $w_1(s) = 0.02$, $w_2(s) = 0$ and put $\kappa = 11.8$; the number of iterations has been stopped to 100, since a greater number would have provided identical visualizations. We point out that for the double scatterer of figure B.18 all the procedure has been implemented twice: the first time by choosing as initial guess γ_0 the single circle around

[27] See subsection 1.8.4.

[28] See subsection 1.8.3: the choice of the compatible case in implementing the generalized discrepancy principle is explained by the high noise level: cf. discussion in subsection 1.8.4.

[29] In any case, the algorithm is quite robust with respect to the choice of the values to be assigned to the various parameters involved. We also point out that the standardization of the procedure according to which such values can be fixed can represent a strong objection to the possible criticism that might be moved against the overall approach, i.e. that the problem of choosing the cut-off threshold for the visualization maps provided by the linear sampling method has been simply converted into the (possibly more difficult) one of determining suitable values for the parameters appearing in the implementation of a deformable model.

the ellipse and, after 100 iterations, the second one by choosing as initial guess γ_0 the single circle around the kite and then stopping the procedure itself after 100 iterations again. For more details, read the captions of the figures themselves.

We think that the results shown in figures B.17 and B.18 are quite promising, all the more that they are probably improvable with a little more careful choice of the parameters $w_1(s)$ and $w_2(s)$; in any case, by means of the deformable contour technique just described above, we have reconstructed, with no effort, boundaries which are closer to the true profiles than the ones obtained in [14] by using a traditional Canny edge-detection algorithm.

3.6. Conclusions

This chapter presents a new viewpoint for the implementation of the linear sampling method when the far-field matrix is discretized according to the same number of equidistant incidence and observation angles. In this new framework the sampling procedure of the previous implementation is replaced by a single functional equation which is regularized by means of a single optimization process. The advantages of this approach are two-fold. First, at a computational level, pointwise regularization is avoided together with a notable number of time consuming zero-finding processes for the generalized discrepancy function (as pointed out in [29], in the traditional implementation the optimal values of the regularization parameter chosen by means of the generalized discrepancy principle significantly depend on the sampling points and in some cases can be even used as further indicator function for the visualization of the scatterer profile). This fact may have important implications in the case of three-dimensional anisotropic scattering, when the inversion requires a notable computational effort. Second, from the point of view of applications, the knowledge of the analytic expression of the indicator function allows one to deduce an approximate estimate of the resolution power of the linear sampling method from the Shannon-Nyquist representation theorem and to improve the computational effectiveness when deformable contour techniques are implemented in order to extract the scatterer profile from the visualization maps provided by the method itself.

We observe that the no-sampling approach to linear sampling can be extended to other visualization methods requiring the solution of many ill-conditioned linear systems parametrized over a sampling grid containing the scatterer, like for example the factorization method [44] or methods which use more strongly singular sources than $\Phi(\cdot, z)$. In particular, the same no-sampling approach introduced in this paper can be applied to theoretically assess whether the use, for example, of a derivative of the fundamental solution at the source point z provides a better spatial resolution. On the other hand, establishing the relation between the regularized solution of equation (3.44) and the approximate solution introduced by the general theorem 2.4.10 is still an open problem.

APPENDIX A

Mathematical miscellany

The aim of this appendix A is to collect in few pages a good number of definitions, notations, theorems and properties which are often used in this PhD thesis, mainly in chapter 2. The various arguments are subdivided into (usually brief) sections; each section is not, in general, self-contained, but it rather presupposes the previous ones. With few exceptions, of no theorem we give the proof; however, just below the title of each section we indicate the bibliographical source(s) from which the material has been drawn. This one may not be such a precise way of giving references, but the treated topics are nearly always of common use in functional analysis and can be found in several books: then the cited references are generally to be intended as mere possibilities or suggestions.

A.1. Direct sum of vector spaces

References: [1], [7], [35], [68].

It is often useful to consider vector spaces constructed by means of other vector spaces. The simplest case is represented by the direct sum of a finite number of vector spaces.

Definition A.1.1. *For all $j = 1, 2, \dots, n$, let X_j be a vector space. We define the* direct sum $X := \bigoplus_{j=1}^{n} X_j$ *of these vector spaces as their Cartesian product $\prod_{j=1}^{n} X_j$ endowed, in turn, with the structure of vector space by means of the following internal operations:*

1. *if $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ are two elements of $\prod_{j=1}^{n} X_j$, we define their sum as*
$$x + y := (x_1 + y_1, \dots, x_n + y_n); \tag{A.1}$$
2. *if $x = (x_1, \dots, x_n) \in \prod_{j=1}^{n} X_j$, we define its product by a complex number c as*
$$c(x_1, \dots, x_n) := (cx_1, \dots, cx_n). \tag{A.2}$$

Of particular interest in functional analysis is the case in which X_j is a Banach space for each $j = 1, \dots, n$. To this purpose, we firstly recall the following definition.

Definition A.1.2. *Two norms $\|\cdot\|_{(1)}$ and $\|\cdot\|_{(2)}$ on a vector space V are said* equivalent *if they induce the same topology on V or, in other terms, if there exist two constants $M_1, M_2 > 0$ such that*

$$M_1 \|v\|_{(1)} \leq \|v\|_{(2)} \leq M_2 \|v\|_{(1)} \qquad \forall v \in V. \tag{A.3}$$

Then we can state the following theorem.

Theorem A.1.1. *For all $j = 1, 2, \dots, n$, let X_j be a Banach space with norm $\|\cdot\|_{X_j}$; then the direct sum $X := \bigoplus_{j=1}^{n} X_j$ is a Banach space too with respect to any one of the infinitely many equivalent norms defined as*

$$\|x\|_{X;(p)} := \left(\sum_{j=1}^{n} \|x_j\|_{X_j}^p \right)^{\frac{1}{p}} \qquad \forall p \in [1, \infty), \quad \forall x = (x_1, \dots, x_n) \in X, \tag{A.4}$$

$$\|x\|_{X;(\infty)} := \max_{1 \leq j \leq n} \|x_j\|_{X_j} \qquad \forall x = (x_1, \dots, x_n) \in X. \tag{A.5}$$

Finally, if X_j is separable [reflexive] for all $j = 1, \dots, n$, then X is separable [reflexive] too.

The particular case in which X_j is a Hilbert space for each $j = 1, \dots, n$ deserves a theorem apart.

Theorem A.1.2. *For all $j = 1, 2, \dots, n$, let X_j be a Hilbert space with scalar product $(\cdot, \cdot)_{X_j}$ and corresponding induced norm $\|\cdot\|_{X_j}$; then the direct sum $X := \bigoplus_{j=1}^{n} X_j$ is a Hilbert space too if equipped with the scalar product defined as*

$$(x, y)_X := \sum_{j=1}^{n} (x_j, y_j)_{X_j} \qquad \forall\ x = (x_1, \dots, x_n),\ y = (y_1, \dots, y_n) \in X. \tag{A.6}$$

Of course, the scalar product (A.6) induces a corresponding norm:

$$\|x\|_X := \sqrt{(x, x)_X} = \sqrt{\sum_{j=1}^{n} (x_j, x_j)_{X_j}} = \left(\sum_{j=1}^{n} \|x_j\|_{X_j}^2 \right)^{\frac{1}{2}}; \tag{A.7}$$

this norm is clearly obtained by the family of norms (A.4) for $p = 2$ and is equivalent to all the other ones belonging to the same family and defined for different values of p, as well as to the norm (A.5).

Coming back to the case of Banach spaces, we recall the concept of dual space.

Definition A.1.3. *Let X be a normed space; then we define its* dual space *as $X^* := \mathcal{B}(X, \mathbb{C})$, i.e.*

$$X^* := \{f : X \to \mathbb{C} \mid \ f \text{ is linear and bounded}\}. \tag{A.8}$$

It can be proved that X^*, when equipped with the operatorial norm

$$\|f\|_{X^*} := \sup_{\|x\|_X \neq 0} \frac{|f(x)|}{\|x\|_X} \qquad \forall f \in X^*, \tag{A.9}$$

is complete, i.e. it is a Banach space, whether or not X is. Moreover, if X is a Hilbert space, X^* is a Hilbert space too and they can be identified by virtue of the *Riesz representation theorem*, which we recall here below.

Theorem A.1.3. [Riesz] *Let X be a Hilbert space and f a linear functional on X. Then $f \in X^*$ if and only if there exists $y_f \in X$ such that it holds*

$$f(x) = (x, y_f)_X \qquad \forall x \in X. \tag{A.10}$$

Moreover, y_f is uniquely determined by $f \in X^*$ and $\|y_f\|_X = \|f\|_{X^*}$.

As regards the direct sum of Banach spaces, we can now state the following theorem.

Theorem A.1.4. *For all $j = 1, 2, \ldots, n$, let X_j be a Banach space with norm $\|\cdot\|_{X_j}$ and X_j^* its dual with norm $\|\cdot\|_{X_j^*}$; let us consider the direct sums*

$$X := \bigoplus_{j=1}^{n} X_j \quad \text{with norm} \quad \|x\|_{X;(1)} := \sum_{j=1}^{n} \|x_j\|_{X_j} \quad \forall x = (x_1, \ldots, x_n) \in X, \tag{A.11}$$

$$Y := \bigoplus_{j=1}^{n} X_j^* \quad \text{with norm} \quad \|f\|_{Y;(\infty)} := \max_{1 \leq j \leq n} \|f_j\|_{X_j^*} \quad \forall f = (f_1, \ldots, f_n) \in Y; \tag{A.12}$$

then Y and X^* are normisomorphic, i.e. there exists an isomorphism $\mathcal{I} : Y \to X^*$, defined as

$$[\mathcal{I}(f)](x) := \sum_{j=1}^{n} f_j(x_j) \qquad \forall x = (x_1, \ldots, x_n) \in X,\ \forall f = (f_1, \ldots, f_n) \in Y, \tag{A.13}$$

which preserves the norms, i.e. such that

$$\|f\|_{Y;(\infty)} = \|\mathcal{I}(f)\|_{X^*} \qquad \forall f = (f_1, \ldots, f_n) \in Y. \tag{A.14}$$

The previous theorem clearly allows one to identify Y with X^*.

A.2. Multi-index notation

References: [41].

The so-called *multi-index* notation is very useful for its conciseness. With the term "multi-index" we mean an n-tuple of non-negative integers, i.e. an element $\alpha = (\alpha_1, \ldots, \alpha_n)$ of $\mathbb{N}^n$. For each $\alpha \in \mathbb{N}^n$, we define the scalars

$$|\alpha|_{\mathbb{N}^n} := \alpha_1 + \ldots + \alpha_n \tag{A.15}$$

and

$$\alpha! := \alpha_1! \cdot \ldots \cdot \alpha_n!, \tag{A.16}$$

while we shall denote with c_α a (complex) numerical coefficient depending on n non-negative integers $\alpha_1, \ldots, \alpha_n$:

$$c_\alpha := c_{\alpha_1,\ldots,\alpha_n}. \tag{A.17}$$

Moreover, for any $x = (x^1, \ldots, x^n) \in \mathbb{R}^n$ and any $\alpha \in \mathbb{N}^n$, we define the monomial

$$x^\alpha := x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n}. \tag{A.18}$$

Finally, if, for each $k = 1, \ldots, n$ and for each $m \in \mathbb{N}$, we put $\partial_k^m := \dfrac{\partial^m}{\partial x_k^m}$ (where for $m = 0$ the derivation operator is clearly the identity), we can define the general partial differentiation operator of order $m \in \mathbb{N}$ for functions defined on open subsets of $\mathbb{R}^n$ as

$$\partial^\alpha := \partial_1^{\alpha_1} \ldots \partial_n^{\alpha_n} = \frac{\partial^m}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}, \tag{A.19}$$

where α is such that $|\alpha|_{\mathbb{N}^n} = m$.

A.3. Spaces of continuous functions

References: [15], [35].

Now, let Ω be a non-empty open subset of $\mathbb{R}^n$; then, for any $r \in \mathbb{N}$, we define:

$$C^r(\Omega) := \left\{u : \Omega \to \mathbb{C} \mid \partial^\alpha u \text{ exists and is continuous on } \Omega \ \ \forall \alpha \ : \ |\alpha|_{\mathbb{N}^n} \leq r\right\}, \tag{A.20}$$

$$C^r(\bar{\Omega}) := \left\{u \in C^k(\Omega) \mid \partial^\alpha u \text{ can be continuously extended onto } \bar{\Omega} \ \ \forall \alpha \ : \ |\alpha|_{\mathbb{N}^n} \leq r\right\} \tag{A.21}$$

and

$$C^\infty(\Omega) := \bigcap_{r\geq 0} C^r(\Omega), \tag{A.22}$$

$$C^\infty(\bar{\Omega}) := \bigcap_{r\geq 0} C^r(\bar{\Omega}). \tag{A.23}$$

Of course, if $r \leq s$, then it holds $C^s(\Omega) \subset C^r(\Omega)$. For $u \in C^0(\Omega)$, we define the support of u as

$$\operatorname{supp} u := \overline{\{x \in \Omega \mid u(x) \neq 0\}}. \tag{A.24}$$

From now on, let K denote a compact subset of the open set Ω (this condition will be written as $K \Subset \Omega$); then, for a fixed $K \Subset \Omega$, we define

$$C^r_K(\Omega) := \{u \in C^r(\Omega) \mid \operatorname{supp} u \subset K\} \tag{A.25}$$

and

$$C^\infty_K(\Omega) := \bigcap_{r\geq 0} C^r_K(\Omega). \tag{A.26}$$

Moreover, we can also define[1]:

$$C^r_{\text{comp}}(\Omega) := \{u : \Omega \to \mathbb{C} \mid \exists\, K \Subset \Omega \text{ such that } u \in C^r_K(\Omega)\} \tag{A.27}$$

and

$$C^\infty_{\text{comp}}(\Omega) := \{u : \Omega \to \mathbb{C} \mid \exists\, K \Subset \Omega \text{ such that } u \in C^\infty_K(\Omega)\}\,. \tag{A.28}$$

A.4. Real-analytic functions

References: [41].

Definition A.4.1. *Let Ω be a non-empty open subset of $\mathbb{R}^n$ and let $f : \Omega \to \mathbb{R}$ be a real-valued function defined in Ω. For any $x_0 \in \Omega$, we say that f is real-analytic at x_0 if there exist coefficients $(c_{x_0})_\alpha \in \mathbb{R}$ and a neighbourhood U_{x_0} of x_0 (all depending on x_0) such that it holds:*

$$f(x) = \sum_{\alpha\in\mathbb{N}^n} (c_{x_0})_\alpha\, (x - x_0)^\alpha \qquad \forall x \in U_{x_0}. \tag{A.29}$$

Moreover, we say that f is real-analytic in Ω if it is real-analytic at each $x_0 \in \Omega$ and we put

$$C^\omega(\Omega) := \{f : \Omega \to \mathbb{R} \mid f \text{ is real-analytic in } \Omega\}. \tag{A.30}$$

[1]Of course, in definitions (A.27), (A.28) the compact subset K of Ω may be different for each function u.

Remark A.4.1. It is not restrictive to assume[2] that the multiple power series (in several real variables) at the right-hand side of relation (A.29) converges absolutely in U_{x_0}: of course, this allows us not to worry about the order in which the infinitely many addenda in the series itself are arranged. □

As stated by the following theorems A.4.1 and A.4.2, real-analytic functions form a subset of $C^\infty(\Omega)$ and they can be characterized either by the property of local representability by Taylor power series, or by the way their partial derivatives increase with increasing order.

Theorem A.4.1. *If $f \in C^\omega(\Omega)$, then $f \in C^\infty(\Omega)$; moreover, for any $x_0 \in \Omega$ there exist a neighbourhood U_{x_0} of x_0 and positive real numbers M_{x_0}, r_{x_0} such that it holds*

$$f(x) = \sum_{\alpha \in \mathbb{N}^n} \frac{1}{\alpha!} (\partial^\alpha f(x_0))(x - x_0)^\alpha \qquad \forall x \in U_{x_0} \tag{A.31}$$

and

$$|\partial^\alpha f(x)| \leq M_{x_0} \, |\alpha|_{\mathbb{N}^n}! \, r_{x_0}^{-|\alpha|_{\mathbb{N}^n}} \qquad \forall x \in U_{x_0}, \quad \forall \alpha \in \mathbb{N}^n. \tag{A.32}$$

Theorem A.4.2. *Let f be a real-valued function defined in the open subset $\Omega \subset \mathbb{R}^n$. Then necessary and sufficient condition for $f \in C^\omega(\Omega)$ is that $f \in C^\infty(\Omega)$ and that for every compact $K \Subset \Omega$ there exist positive real numbers M_K, r_K such that it holds*

$$|\partial^\alpha f(x_0)| \leq M_K \, |\alpha|_{\mathbb{N}^n}! \, r_K^{-|\alpha|_{\mathbb{N}^n}} \qquad \forall x_0 \in K, \quad \forall \alpha \in \mathbb{N}^n. \tag{A.33}$$

Real analytic functions enjoy the property of *unique continuation* expressed by the following theorem.

Theorem A.4.3. *Let Ω be a connected open subset of $\mathbb{R}^n$; let $f \in C^\omega(\Omega)$ and $x_0 \in \Omega$. Then f is determined uniquely in Ω if we know $\partial^\alpha f(x_0)$ for all $\alpha \in \mathbb{N}^n$. In particular, f is determined uniquely in Ω by its restriction to any non-empty open subset of Ω.*

A.5. Distributions

References: [50].

Following the notation introduced by Schwartz [61], we denote with

[2] See p. 100-101 in [46] for details.

1. $\mathcal{E}(\Omega)$ the space $C^\infty(\Omega)$ equipped with the following definition of convergence of sequences: if $\{\varphi_j\}_{j=0}^\infty$ is a sequence in $C^\infty(\Omega)$, we say that

$$\varphi_j \to 0 \ \text{ in } \ \mathcal{E}(\Omega) \ \text{ as } j \to \infty \tag{A.34}$$

   if for each compact $K \Subset \Omega$ and for each multi-index α it holds:

$$\partial^\alpha \varphi_j \to 0 \quad \text{uniformly on } K \ \text{ as } j \to \infty; \tag{A.35}$$

2. $\mathcal{D}_K(\Omega)$ the space $C_K^\infty(\Omega)$ (for a fixed $K \Subset \Omega$) equipped with the following definition of convergence of sequences: if $\{\varphi_j\}_{j=0}^\infty$ is a sequence in $C_K^\infty(\Omega)$, we say that

$$\varphi_j \to 0 \ \text{ in } \ \mathcal{D}_K(\Omega) \ \text{ as } j \to \infty \tag{A.36}$$

   if for each multi-index α it holds:

$$\partial^\alpha \varphi_j \to 0 \quad \text{uniformly on } K \ \text{ as } j \to \infty; \tag{A.37}$$

3. $\mathcal{D}(\Omega)$ the space $C_{\text{comp}}^\infty(\Omega)$ equipped with the following definition of convergence of sequences: if $\{\varphi_j\}_{j=0}^\infty$ is a sequence in $C_{\text{comp}}^\infty(\Omega)$, we say that

$$\varphi_j \to 0 \ \text{ in } \ \mathcal{D}(\Omega) \ \text{ as } j \to \infty \tag{A.38}$$

   if there exists a compact $K \Subset \Omega$ such that for each multi-index α it holds:

$$\partial^\alpha \varphi_j \to 0 \quad \text{uniformly on } K \ \text{ as } j \to \infty. \tag{A.39}$$

Moreover, we define the space of rapidly decreasing, C^∞ functions:

$$\mathcal{S}(\mathbb{R}^n) := \left\{ \varphi \in C^\infty(\mathbb{R}^n) \mid \sup_{x \in \mathbb{R}^n} \left| x^\alpha \partial^\beta \varphi(x) \right| < \infty \ \ \forall \alpha, \beta \in \mathbb{N}^n \right\}, \tag{A.40}$$

equipped with the following definition of convergence of sequences: if $\{\varphi_j\}_{j=0}^\infty$ is a sequence in $C^\infty(\mathbb{R}^n)$, we say that

$$\varphi_j \to 0 \ \text{ in } \ \mathcal{S}(\mathbb{R}^n) \ \text{ as } j \to \infty \tag{A.41}$$

if for all multi-indices $\alpha, \beta \in \mathbb{N}^n$ it holds:

$$x^\alpha \partial^\beta \varphi_j \to 0 \quad \text{uniformly on } \mathbb{R}^n \ \text{ as } j \to \infty. \tag{A.42}$$

Now, let us consider a linear functional $F : \mathcal{D}(\Omega) \to \mathbb{C}$ and assume that it is sequentially continuous, i.e. for every sequence $\{\varphi_j\}_{j=0}^\infty$ in $\mathcal{D}(\Omega)$ it holds:

$$\varphi_j \to 0 \ \text{ in } \mathcal{D}(\Omega) \ \text{ as } j \to \infty \implies F(\varphi_j) \to 0 \ \text{ in } \mathbb{C} \ \text{ as } j \to \infty; \tag{A.43}$$

then F is called a (Schwartz) *distribution* on Ω. In this context, the elements of $\mathcal{D}(\Omega)$ are referred to as *test functions* on Ω. The set of all distributions on Ω is denoted by $\mathcal{D}^*(\Omega)$. The value $F(\varphi)$ of F at $\varphi \in \mathcal{D}(\Omega)$ will be often denoted by means of the *pairing* $\langle F, \varphi \rangle_\Omega$.

Let Ω_1 be an open subset of Ω and, for any $\varphi \in \mathcal{D}(\Omega_1)$, let $\tilde{\varphi} \in \mathcal{D}(\Omega)$ denote the extension of φ by zero. For any distribution $u \in \mathcal{D}^*(\Omega)$, we define the *restriction* $u|_{\Omega_1} \in \mathcal{D}^*(\Omega_1)$ by means of the condition

$$\langle u|_{\Omega_1}, \varphi \rangle_{\Omega_1} = \langle u, \tilde{\varphi} \rangle_\Omega \qquad \forall \varphi \in \mathcal{D}(\Omega_1). \tag{A.44}$$

Moreover, we say that $u = 0$ on Ω_1 if $u|_{\Omega_1} = 0$, and define the *support* $\operatorname{supp} u$ of u to be the largest closed subset of Ω such that $u = 0$ on $\Omega \setminus \operatorname{supp} u$.

Analogously to $\mathcal{D}^*(\Omega)$, one can define also the spaces $\mathcal{E}^*(\Omega)$ and $\mathcal{S}^*(\mathbb{R}^n)$. It can be proved that $\mathcal{E}^*(\Omega)$ coincides with the space of distributions having compact support, i.e.

$$\mathcal{E}^*(\Omega) = \{u \in \mathcal{D}^*(\Omega) \mid \operatorname{supp} u \Subset \Omega\}. \tag{A.45}$$

Furthermore, it can be shown that the inclusions $\mathcal{D}(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n) \subset \mathcal{E}(\mathbb{R}^n)$ are continuous with dense image, so we have

$$\mathcal{E}^*(\mathbb{R}^n) \subset \mathcal{S}^*(\mathbb{R}^n) \subset \mathcal{D}^*(\mathbb{R}^n). \tag{A.46}$$

The elements of $\mathcal{S}^*(\mathbb{R}^n)$, i.e. the sequentially continuous linear functionals on $\mathcal{S}(\mathbb{R}^n)$, are called *temperate distributions*.

One can define partial differentiation for distributions getting inspiration from the procedure of integration by parts, i.e.

$$\langle \partial^\alpha u, \varphi \rangle_\Omega := (-1)^{|\alpha|_{\mathbb{N}^n}} \langle u, \partial^\alpha \varphi \rangle_\Omega \qquad \forall u \in \mathcal{D}^*(\Omega),\ \forall \varphi \in \mathcal{D}(\Omega); \tag{A.47}$$

here, the sequential continuity of $\partial^\alpha u$, i.e. the fact that $\partial^\alpha u$ itself is a distribution, immediately follows from the obvious fact that if $\varphi_j \to 0$ in $\mathcal{D}(\Omega)$ as $j \to \infty$, then $\partial^\alpha \varphi_j \to 0$ in $\mathcal{D}(\Omega)$ as $j \to \infty$ (cf. (A.38), (A.39)).

We can also define the complex conjugate $\bar{u} \in \mathcal{D}^*(\Omega)$ of a distribution $u \in \mathcal{D}^*(\Omega)$ by putting

$$\langle \bar{u}, \varphi \rangle_\Omega := \overline{\langle u, \bar{\varphi} \rangle_\Omega} \qquad \forall \varphi \in \mathcal{D}(\Omega). \tag{A.48}$$

A.6. Spaces of Lebesgue integrable functions

References: [1], [50], [51].

From now on, let Ω denote any Lebesgue measurable (and not necessarily open) subset of $\mathbb{R}^n$ with strictly positive measure and let $p \in [1, \infty)$; then we define $L^p(\Omega)$ as the set of the

measurable functions[3] $u : \Omega \to \mathbb{C}$ such that the function $|u|^p$ is Lebesgue integrable in Ω. In $L^p(\Omega)$ we introduce the norm

$$\|u\|_{L^p(\Omega)} := \left(\int_\Omega |u(x)|^p dx \right)^{\frac{1}{p}}, \qquad \forall u \in L^p(\Omega). \tag{A.49}$$

It turns out that for each $p \in [1, \infty)$, the set $L^p(\Omega)$ is a Banach space; in particular, for $p = 2$ it is a Hilbert space, with scalar product defined as

$$(u, v)_{L^2(\Omega)} := \int_\Omega u(x)\, \overline{v(x)}\, dx \qquad \forall u, v \in L^2(\Omega); \tag{A.50}$$

another remarkable property of $L^2(\Omega)$ is that, like any Hilbert space, it can be identified with its dual by virtue of the Riesz representation theorem, i.e.

$$L^2(\Omega) = [L^2(\Omega)]^*. \tag{A.51}$$

More precisely, any $v \in L^2(\Omega)$ can be regarded as an element $f_v \in [L^2(\Omega)]^*$ by putting:

$$f_v(u) := (u, \bar{v})_{L^2(\Omega)} = \int_\Omega u(x)\, v(x)\, dx \qquad \forall u \in L^2(\Omega); \tag{A.52}$$

moreover, it holds $\|v\|_{L^2(\Omega)} = \|f_v\|_{[L^2(\Omega)]^*}$.

In the same hypotheses (on Ω and p) assumed to define $L^p(\Omega)$, we denote with $L^p_{loc}(\Omega)$ the set of the measurable functions $u : \Omega \to \mathbb{C}$ such that $u \in L^p(K)$ for every compact $K \subset \Omega$.

Each function $u \in L^1_{loc}(\Omega)$ can be regarded as a distribution, i.e. an element of $\mathcal{D}^*(\Omega)$, by means of the imbedding $\iota : L^1_{loc}(\Omega) \to \mathcal{D}^*(\Omega)$ defined by the following condition:

$$\langle \iota u, \varphi \rangle_\Omega := \int_\Omega u(x)\varphi(x) dx \qquad \forall \varphi \in \mathcal{D}(\Omega); \tag{A.53}$$

such an imbedding allows one to identify $L^1_{loc}(\Omega)$ with a subspace of $\mathcal{D}^*(\Omega)$.

Moreover, a sufficient condition for a function $u \in L^1_{loc}(\mathbb{R}^n)$ to be identifiable with a temperate distribution $\iota u \in \mathcal{S}^*(\mathbb{R}^n)$ is that it is *slowly growing*, i.e. that there exists $r \in \mathbb{R}$ such that $u(x) = O(\|x\|^r_{\mathbb{R}^n})$ as $\|x\|_{\mathbb{R}^n} \to \infty$: if such a condition is satisfied, definition (A.53) (with $\Omega = \mathbb{R}^n$) can be extended to the case in which $\varphi \in \mathcal{S}(\mathbb{R}^n)$.

Finally, since it is possible to show that $L^p(\Omega) \subset L^1_{loc}(\Omega)$ $\forall p \in [1, \infty)$, we have that any function $u \in L^p(\Omega)$ can be regarded as a distribution $\iota u \in \mathcal{D}^*(\Omega)$.

[3] More properly, it is the set formed by the equivalence classes of these functions, the equivalence relation being such that two functions defined on Ω are equivalent if and only if they are equal almost everywhere in Ω, i.e. they (possibly) differ only on a set of zero measure. However, with a slight and useful abuse of terminology, the elements of $L^p(\Omega)$ are commonly called *functions*.

A.7. Fourier transform

References: [35], [50], [51].

If $u \in L^1(\mathbb{R}^n)$, we define its Fourier transform as:

$$[\mathcal{F}(u)](\omega) := \int_{\mathbb{R}^n} u(x)\, e^{-\mathrm{i}\omega\cdot x} dx \qquad \forall \omega \in \mathbb{R}^n; \tag{A.54}$$

if $v \in L^1(\mathbb{R}^n)$, we define its Fourier anti-transform as:

$$[\mathcal{F}_{(-1)}(v)](x) := \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} v(\omega)\, e^{\mathrm{i}\omega\cdot x} d\omega \qquad \forall x \in \mathbb{R}^n. \tag{A.55}$$

In general, if $u \in L^1(\mathbb{R}^n)$ and no additional hypotheses are assumed, the natural inversion formula one has in mind, i.e.

$$u(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} [\mathcal{F}(u)](\omega)\, e^{\mathrm{i}\omega\cdot x} d\omega, \tag{A.56}$$

it is meaningless, since the function to be integrated does not necessarily belong to $L^1(\mathbb{R}^n)$: in this sense, we cannot say that the Fourier anti-transform is the inverse of the Fourier transform. However, it can be shown that if both u and $\mathcal{F}(u)$ belong to $L^1(\mathbb{R}^n)$, then the inversion formula (A.56) is valid at every point x in which u is continuous; even more, if both u and $\mathcal{F}(u)$ are continuous on all $\mathbb{R}^n$ and belong to $L^1(\mathbb{R}^n)$, then

$$\{\mathcal{F}_{(-1)}[\mathcal{F}(u)]\}(x) = u(x) = \{\mathcal{F}[\mathcal{F}_{(-1)}(u)]\}(x) \qquad \forall x \in \mathbb{R}^n. \tag{A.57}$$

Since it can be proved that $\mathcal{F}(\mathcal{S}(\mathbb{R}^n)) \subset \mathcal{S}(\mathbb{R}^n)$, $\mathcal{F}_{(-1)}(\mathcal{S}(\mathbb{R}^n)) \subset \mathcal{S}(\mathbb{R}^n)$ and, on the other hand, it trivially holds $\mathcal{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n) \cap C^\infty(\mathbb{R}^n)$, it follows that, in particular, the inversion formula (A.56) is valid for any $u = \varphi \in \mathcal{S}(\mathbb{R}^n)$.

Put $M^\alpha(x) := x^\alpha$ for all $x \in \mathbb{R}^n$, easy calculations show that if $\varphi \in \mathcal{S}(\mathbb{R}^n)$, then

$$[\mathcal{F}(\partial^\alpha \varphi)](\omega) = \mathrm{i}^{|\alpha|_{\mathbb{N}^n}} M^\alpha(\omega)[\mathcal{F}(\varphi)](\omega) \qquad \forall \alpha \in \mathbb{N}^n \tag{A.58}$$

and

$$[\mathcal{F}(M^\alpha \varphi)](\omega) = \mathrm{i}^{|\alpha|_{\mathbb{N}^n}} \partial^\alpha [\mathcal{F}(\varphi)](\omega) \qquad \forall \alpha \in \mathbb{N}^n. \tag{A.59}$$

Relations (A.58) and (A.59) imply that the Fourier transform defines a (sequentially) continuous linear operator

$$\mathcal{F} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n). \tag{A.60}$$

Relations analogous to (A.58) and (A.59) also hold for $\mathcal{F}_{(-1)} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$, which is then (sequentially) continuous too and, by virtue of relations (A.57), is the inverse of $\mathcal{F}$ on $\mathcal{S}(\mathbb{R}^n)$.

Of course, any element $\varphi \in \mathcal{S}(\mathbb{R}^n)$ can be regarded as a slowly growing function belonging to $L^1_{loc}(\mathbb{R}^n)$, so that $\iota\varphi \in \mathcal{S}^*(\mathbb{R}^n)$. Since it clearly holds

$$\langle \iota[\mathcal{F}(\varphi)], \psi \rangle_{\mathbb{R}^n} = \langle \varphi, \iota[\mathcal{F}(\psi)] \rangle_{\mathbb{R}^n} \qquad \forall \varphi, \psi \in \mathcal{S}(\mathbb{R}^n), \tag{A.61}$$

we can easily define an extension of $\mathcal{F}$ to an operator, still denoted with $\mathcal{F}$, from $\mathcal{S}^*(\mathbb{R}^n)$ in itself, by means of the condition[4]

$$\langle \mathcal{F}(u), \varphi \rangle_{\mathbb{R}^n} := \langle u, \mathcal{F}(\varphi) \rangle_{\mathbb{R}^n} \qquad \forall u \in \mathcal{S}^*(\mathbb{R}^n), \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^n). \tag{A.62}$$

An analogous procedure allows one to define an extension $\mathcal{F}_{(-1)} : \mathcal{S}^*(\mathbb{R}^n) \to \mathcal{S}^*(\mathbb{R}^n)$.

Furthermore, with arguments based on the denseness of $\mathcal{S}(\mathbb{R}^n)$ in $L^2(\mathbb{R}^n)$, Plancherel's theorem shows that there exists a unique operator (still denoted with $\mathcal{F}$) from $L^2(\mathbb{R}^n)$ in itself that extends by linearity and continuity the Fourier transform restricted to $\mathcal{S}(\mathbb{R}^n)$; an analogous result (and notation) clearly holds for $\mathcal{F}_{(-1)}$. Moreover, the operator $\mathcal{F}_{(-1)}$ itself is the inverse of $\mathcal{F}$ on $L^2(\mathbb{R}^n)$, i.e. $\mathcal{F}_{(-1)}\mathcal{F} = \mathcal{F}\mathcal{F}_{(-1)} = I_{L^2(\mathbb{R}^n)}$, and the generalized Parseval equality holds

$$(2\pi)^n (\mathcal{F}^{-1}(u), \mathcal{F}^{-1}(v))_{L^2(\mathbb{R}^n)} = (u, v)_{L^2(\mathbb{R}^n)} = \frac{1}{(2\pi)^n} (\mathcal{F}(u), \mathcal{F}(v))_{L^2(\mathbb{R}^n)} \qquad \forall u, v \in L^2(\mathbb{R}^n), \tag{A.63}$$

which clearly becomes, in the particular case $u = v$,

$$(2\pi)^{n/2} \|\mathcal{F}^{-1}(u)\|_{L^2(\mathbb{R}^n)} = \|u\|_{L^2(\mathbb{R}^n)} = \frac{1}{(2\pi)^{n/2}} \|\mathcal{F}(u)\|_{L^2(\mathbb{R}^n)} \qquad \forall u \in L^2(\mathbb{R}^n). \tag{A.64}$$

A.8. Sobolev spaces (first family)

References: [1], [13], [50].

Let Ω be any Lebesgue measurable and open subset of $\mathbb{R}^n$ with strictly positive measure and let $p \in [1, \infty)$; since $L^p(\Omega) \subset L^1_{loc}(\Omega)$, any element $u \in L^p(\Omega)$ is such that, for each $\alpha \in \mathbb{N}^n$, the distribution ιu (cf. definition (A.53)) has its partial derivative $\partial^\alpha \iota u$ defined according to condition (A.47). It may happen that the distribution $\partial^\alpha \iota u$ is, in turn, identifiable with a function $g_\alpha \in L^p(\Omega)$, i.e. that there exists a function $g_\alpha \in L^p(\Omega)$ such that $\partial^\alpha \iota u = \iota g_\alpha$. In such a case, we say that g_α is a *weak partial derivative* of u and we simply write, with a slight notational abuse, that $\partial^\alpha u = g_\alpha$.

[4] Of course, at the left-hand side of definition (A.62) the operator $\mathcal{F}$ is to be intended as mapping $\mathcal{S}^*(\mathbb{R}^n)$ in itself, while at the right-hand side has to be regarded as mapping $\mathcal{S}(\mathbb{R}^n)$ in itself.

We can now define the *Sobolev space* $W^{r,p}(\Omega)$ of order $r \in \mathbb{N}$ based on $L^p(\Omega)$ as

$$W^{r,p}(\Omega) := \{u \in L^p(\Omega) \mid \partial^\alpha u \in L^p(\Omega) \ \ \forall \alpha \ : \ |\alpha|_{\mathbb{N}^n} \leq r\}. \tag{A.65}$$

The completeness of $L^p(\Omega)$ implies that $W^{r,p}(\Omega)$ becomes a Banach space with the norm defined as:

$$\|u\|_{W^{r,p}(\Omega)} := \left(\sum_{|\alpha|_{\mathbb{N}^n} \leq r} \int_\Omega |\partial^\alpha u(x)|^p dx \right)^{\frac{1}{p}}. \tag{A.66}$$

It is possible to prove that the previous norm is equivalent to the following one:

$$\|u\|'_{W^{r,p}(\Omega)} := \sum_{|\alpha|_{\mathbb{N}^n} \leq r} \|\partial^\alpha u(x)\|_{L^p(\Omega)}. \tag{A.67}$$

Remark A.8.1. If $u \in C^1(\Omega) \cap L^p(\Omega)$ and if $\dfrac{\partial u}{\partial x_i} \in L^p(\Omega)$ for all $i = 1, \ldots, n$ (where $\dfrac{\partial u}{\partial x_i}$ denotes the partial derivative of u in the classical sense), then $u \in W^{1,p}(\Omega)$ and the partial derivatives of u in the classical sense coincide with the weak partial derivatives of u. □

We also need to define Sobolev spaces of non-integer order $s \in [0, +\infty)$; to this end, we introduce the Slobodeckiĭ seminorm

$$|u|_{\mu,p,\Omega} := \left(\int_\Omega \int_\Omega \frac{|u(x) - u(y)|^p}{\|x - y\|^{n+p\mu}_{\mathbb{R}^n}} dx\, dy \right)^{\frac{1}{p}} \qquad \forall \mu \in (0,1); \tag{A.68}$$

then, for $s = r + \mu$, we define the Sobolev space

$$W^{s,p}(\Omega) := \{u \in W^{r,p}(\Omega) \ \mid \ |\partial^\alpha u|_{\mu,p,\Omega} < \infty \ \ \forall \alpha \ : \ |\alpha|_{\mathbb{N}^n} = r\} \tag{A.69}$$

equipped with the norm

$$\|u\|_{W^{s,p}(\Omega)} := \left(\|u\|^p_{W^{r,p}(\Omega)} + \sum_{|\alpha|_{\mathbb{N}^n} = r} |\partial^\alpha u|^p_{\mu,p,\Omega} \right)^{\frac{1}{p}}. \tag{A.70}$$

The case $p = 2$ is of particular interest: indeed, $W^{s,2}(\Omega)$ turns out to be a Hilbert space for all $s \in [0, +\infty)$. More precisely, for any integer $r \geq 0$ the norm in $W^{r,2}(\Omega)$ is induced by the scalar product

$$(u, v)_{W^{r,2}(\Omega)} := \sum_{|\alpha|_{\mathbb{N}^n} \leq r} \int_\Omega \partial^\alpha u(x)\, \overline{\partial^\alpha v(x)}\, dx, \tag{A.71}$$

while for $s = r + \mu$ the norm in $W^{s,2}(\Omega)$ is induced by the scalar product

$$(u, v)_{W^{s,2}(\Omega)} := (u, v)_{W^{r,2}(\Omega)} + \sum_{|\alpha|_{\mathbb{N}^n} = r} \int_\Omega \int_\Omega \frac{[\partial^\alpha u(x) - \partial^\alpha u(y)]\, \overline{[\partial^\alpha v(x) - \partial^\alpha v(y)]}}{\|x - y\|^{n+2\mu}_{\mathbb{R}^n}} dx\, dy. \tag{A.72}$$

For any integer $r \geq 1$, we also define the negative-order Sobolev space $W^{-r,p}(\Omega)$ as the set of distributions $u \in \mathcal{D}^*(\Omega)$ that admit a representation of the form

$$u = \sum_{|\alpha|_{\mathbb{N}^n} \leq r} \partial^\alpha f_\alpha \qquad \text{with} \quad f_\alpha \in L^p(\Omega) \quad \forall \alpha \ : \ |\alpha|_{\mathbb{N}^n} \leq r, \tag{A.73}$$

equipped with the norm

$$\|u\|_{W^{-r,p}(\Omega)} := \inf \left(\sum_{|\alpha|_{\mathbb{N}^n} \leq r} \|f_\alpha\|^p_{L^p(\Omega)} \right)^{\frac{1}{p}}, \tag{A.74}$$

where the infimum is taken over all the representations of the form (A.73).

A.9. Sobolev spaces (second family)

References: [1], [50].

For any $s \in \mathbb{R}$, we define a continuous linear operator $\mathcal{J}^s : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$, called the *Bessel potential* of order s, by

$$[\mathcal{J}^s(\varphi)](x) := \mathcal{F}^{-1} \left\{ \left(1 + \|\omega\|^2_{\mathbb{R}^n}\right)^{\frac{s}{2}} [\mathcal{F}(\varphi)](\omega) \right\} (x) \qquad \forall x \in \mathbb{R}^n, \tag{A.75}$$

i.e., more explicitly,

$$[\mathcal{J}^s(\varphi)](x) := \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \left(1 + \|\omega\|^2_{\mathbb{R}^n}\right)^{\frac{s}{2}} [\mathcal{F}(\varphi)](\omega)\, e^{\mathrm{i}\omega \cdot x} d\omega \qquad \forall x \in \mathbb{R}^n. \tag{A.76}$$

Definitions (A.54) and (A.75) easily imply that

$$\{\mathcal{F}[\mathcal{J}^s(\varphi)]\}(\omega) = \left(1 + \|\omega\|^2_{\mathbb{R}^n}\right)^{\frac{s}{2}} [\mathcal{F}(\varphi)](\omega), \tag{A.77}$$

then, under Fourier transformation, the action of $\mathcal{J}^s$ is to multiply $[\mathcal{F}(\varphi)](\omega)$ by a function that is $O(\|\omega\|^s_{\mathbb{R}^n})$ as $\|\omega\|^s_{\mathbb{R}^n} \to \infty$. Therefore, remembering relation (A.58), we can regard $\mathcal{J}^s$ as a sort of differential operator of order s. We also point out that, for all $s, t \in \mathbb{R}$, it holds:

$$\mathcal{J}^{s+t} = \mathcal{J}^s \mathcal{J}^t, \qquad (\mathcal{J}^s)^{-1} = \mathcal{J}^{-s}, \qquad \mathcal{J}^0 = I_{\mathcal{S}(\mathbb{R}^n)}. \tag{A.78}$$

Moreover, it follows from (A.54), (A.76) and (A.77) a relation analogous to (A.61), i.e.

$$\langle \iota[\mathcal{J}^s(\varphi)], \psi \rangle_{\mathbb{R}^n} = \langle \varphi, \iota[\mathcal{J}^s(\psi)] \rangle_{\mathbb{R}^n} \qquad \forall \varphi, \psi \in \mathcal{S}(\mathbb{R}^n). \tag{A.79}$$

Relation (A.79) suggests and actually allows a natural extension of the Bessel potential to a linear operator $\mathcal{J}^s : \mathcal{S}^*(\mathbb{R}^n) \to \mathcal{S}^*(\mathbb{R}^n)$ on the space of temperate distributions, defined, analogously to relation (A.62), by means of the following condition[5]:

$$\langle \mathcal{J}^s u, \psi \rangle := \langle u, \mathcal{J}^s \psi \rangle \qquad \forall u \in \mathcal{S}^*(\mathbb{R}^n), \quad \forall \psi \in \mathcal{S}(\mathbb{R}^n). \tag{A.80}$$

For any $s \in \mathbb{R}$, we now define the Sobolev space[6]

$$H^s(\mathbb{R}^n) := \{u \in \mathcal{S}^*(\mathbb{R}^n) \mid \mathcal{J}^s u \in L^2(\mathbb{R}^n)\} \tag{A.81}$$

equipped with the scalar product

$$(u, v)_{H^s(\mathbb{R}^n)} := (\mathcal{J}^s u, \mathcal{J}^s v)_{L^2(\mathbb{R}^n)} \qquad \forall u, v \in \mathcal{S}^*(\mathbb{R}^n) \tag{A.82}$$

and the induced norm

$$\|u\|_{H^s(\mathbb{R}^n)} := \sqrt{(u, u)_{H^s(\mathbb{R}^n)}} = \|\mathcal{J}^s u\|_{L^2(\mathbb{R}^n)} \qquad \forall u \in \mathcal{S}^*(\mathbb{R}^n). \tag{A.83}$$

Then the Bessel potential

$$\mathcal{J}^s : H^s(\mathbb{R}^n) \to L^2(\mathbb{R}^n) \tag{A.84}$$

turns out to be a unitary isomorphism and, in particular, since $\mathcal{J}^0 u = u$, it holds

$$H^0(\mathbb{R}^n) = L^2(\mathbb{R}^n). \tag{A.85}$$

By virtue of relations (A.75), (A.83) and (A.64), we easily find that

$$\|u\|^2_{H^s(\mathbb{R}^n)} = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \left(1 + \|\omega\|^2_{\mathbb{R}^n}\right)^s \, |\, [\mathcal{F}(u)](\omega) \,|^2 \, d\omega; \tag{A.86}$$

therefore if $s \leq t$ and $\|u\|_{H^t(\mathbb{R}^n)} < \infty$, then $\|u\|_{H^s(\mathbb{R}^n)} \leq \|u\|_{H^t(\mathbb{R}^n)}$ and, consequently, $H^t(\mathbb{R}^n) \subset H^s(\mathbb{R}^n)$; it can be proved that this inclusion is continuous with dense image.

Several properties of $H^s(\mathbb{R}^n)$ straightforwardly follow from standard properties of $L^2(\mathbb{R}^n)$. For example, $H^s(\mathbb{R}^n)$ is a (separable) Hilbert space, and $\mathcal{D}(\mathbb{R}^n)$ is dense in $H^s(\mathbb{R}^n)$ (i.e. $H^s(\mathbb{R}^n)$ is the closure of $\mathcal{D}(\mathbb{R}^n)$ in the norm given by (A.86)), since $\mathcal{J}^s[\mathcal{S}(\mathbb{R}^n)] = \mathcal{S}(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n)$ and the inclusion $\mathcal{D}(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n)$ is continuous with dense image; moreover, by virtue of the identification $L^2(\Omega) = [L^2(\Omega)]^*$, it is possible to prove that $H^{-s}(\mathbb{R}^n)$ is an isometric realization of the dual space of $H^s(\mathbb{R}^n)$, i.e.

$$H^{-s}(\mathbb{R}^n) = [H^s(\mathbb{R}^n)]^* \qquad \forall s \in \mathbb{R} \tag{A.87}$$

[5]Of course, at the left-hand side of definition (A.80) the operator $\mathcal{J}^s$ is to be intended as mapping $\mathcal{S}^*(\mathbb{R}^n)$ in itself, while at the right-hand side has to be regarded as mapping $\mathcal{S}(\mathbb{R}^n)$ in itself.

[6]Of course, the condition $\mathcal{J}^s u \in L^2(\mathbb{R}^n)$ in definition (A.81) is to be intended as a shorthand for the following one: there exists a function $v \in L^2(\mathbb{R}^n) \subset L^1_{loc}(\mathbb{R}^n)$ such that $\iota v = \mathcal{J}^s(u) \in \mathcal{S}^*(\mathbb{R}^n)$.

and

$$\|u\|_{H^{-s}(\mathbb{R}^n)} = \sup_{0\neq v\in H^s(\mathbb{R}^n)} \frac{|\langle u, v\rangle_{\mathbb{R}^n}|}{\|v\|_{H^s(\mathbb{R}^n)}} \qquad \forall u \in H^{-s}(\mathbb{R}^n). \tag{A.88}$$

For any closed set $F \subset \mathbb{R}^n$, we also define the associated Sobolev space of order $s \in \mathbb{R}$ as

$$H^s_F := \{u \in H^s(\mathbb{R}^n) \mid \operatorname{supp} u \subset F\}. \tag{A.89}$$

The space H^s_F easily turns out to be a closed subspace of $H^s(\mathbb{R}^n)$ and therefore it is a Hilbert space when equipped with the restriction of the scalar product in $H^s(\mathbb{R}^n)$.

Moreover, for any non-empty open set $\Omega \subset \mathbb{R}^n$, we define the associated Sobolev space of order $s \in \mathbb{R}$ as

$$H^s(\Omega) = \{u \in \mathcal{D}^*(\Omega) \mid \exists\, U \in H^s(\mathbb{R}^n) \text{ such that } u = U|_\Omega\}. \tag{A.90}$$

Also $H^s(\Omega)$ can be endowed with a Hilbert space structure, although not straightforwardly: in particular, it is possible to prove that the norm induced by the scalar product on $H^s(\Omega)$ is such that

$$\|u\|_{H^s(\Omega)} = \min_{U\in A_u} \|U\|_{H^s(\mathbb{R}^n)}, \tag{A.91}$$

where $A_u := \{U \in H^s(\mathbb{R}^n) \mid U|_\Omega = u\}$; for our purposes, equality (A.91) can be regarded as a definition, although improperly.

Relation (A.91) itself shows that the restriction operator $R_\Omega : H^s(\mathbb{R}^n) \to H^s(\Omega)$ mapping U into $U|_\Omega$ is continuous and therefore, by virtue of the denseness of $\mathcal{D}(\mathbb{R}^n)$ in $H^s(\mathbb{R}^n)$, the space defined as

$$\mathcal{D}(\bar\Omega) := \{\varphi \in C^\infty(\Omega) \mid \exists\, \Phi \in \mathcal{D}(\mathbb{R}^n) \text{ such that } \varphi = \Phi|_U\} \tag{A.92}$$

is dense in $H^s(\Omega)$.

We also define two other Sobolev spaces on Ω, i.e.

$$\widetilde{H}^s(\Omega) := \text{closure of } \mathcal{D}(\Omega) \text{ in } H^s(\mathbb{R}^n), \tag{A.93}$$

$$H^s_0(\Omega) := \text{closure of } \mathcal{D}(\Omega) \text{ in } H^s(\Omega); \tag{A.94}$$

they are easily endowed with a Hilbert space structure by respectively restricting the scalar products in $H^s(\mathbb{R}^n)$ and in $H^s(\Omega)$. It is not difficult to realize that

$$\widetilde{H}^s(\Omega) \subset H^s_{\bar\Omega} \qquad \text{and} \qquad \widetilde{H}^s(\Omega) \subset H^s_0(\Omega), \tag{A.95}$$

while the reverse inclusions hold only under suitable conditions on Ω and s (see theorems A.13.1 and A.13.3 in the following).

An element of $H^s_{\bar\Omega}$ is a distribution on $\mathbb{R}^n$, but, provided the n-dimensional Lebesgue measure of the boundary of Ω is zero, the restriction operator $u \mapsto u|_\Omega$ defines an imbedding

$$H^s_{\bar\Omega} \subset L^2(\Omega) \qquad \forall s \geq 0. \tag{A.96}$$

A.10. Links between the two families of Sobolev spaces (1)

References: [1], [50].

Before any further assumptions are made on Ω, it is already possible to show the following relations between the two families of Sobolev spaces introduced above.

Theorem A.10.1. *The following statements hold:*

1. *if $s \geq 0$, then $W^{s,2}(\mathbb{R}^n) = H^s(\mathbb{R}^n)$ with equivalent norms;*
2. *for any non-empty open subset $\Omega \subset \mathbb{R}^n$, there is a continuous inclusion*
$$H^s(\Omega) \subset W^{s,2}(\Omega) \qquad \forall s \geq 0; \tag{A.97}$$
3. *for any non-empty open subset $\Omega \subset \mathbb{R}^n$ and for any integer $r \geq 0$, it holds*
$$H^{-r}(\Omega) = W^{-r,2}(\Omega) \tag{A.98}$$
with equivalent norms.

A.11. Partition of unity

References: [50].

In the following, we shall need the notion of *partition of unity*.

Definition A.11.1. *A* partition of unity *for an open subset $E \subset \mathbb{R}^n$ is a (finite or infinite) set of functions $\{\varphi_j\}_{j\in\mathcal{J}}$, with $\varphi_j \in C^\infty(\mathbb{R}^n)$ $\forall j \in \mathcal{J}$, such that*

1. *$\varphi_j(x) \geq 0$ $\forall x \in \mathbb{R}^n$ and $\forall j \in \mathcal{J}$;*
2. *each point of E has a neighbourhood that intersects $\operatorname{supp}\varphi_j$ for only finitely many $j \in \mathcal{J}$;*
3. *$\sum_{j\in\mathcal{J}} \varphi_j(x) = 1$ $\forall x \in E$.*

We point out that condition No 2 implies that the sum in condition No 3 is finite for each $x \in E$. If E is not open, then we say that the set of functions $\{\varphi_j\}_{j\in\mathcal{J}}$ forms a partition of unity for E if it forms a partition of unity for some open neighbourhood of E.

Moreover, if $\{W_p\}_{p\in\mathcal{P}}$ is an open cover for E, i.e. a family of open sets such that $E \subset \bigcup_{p\in\mathcal{P}} W_p$, we say that a partition of unity $\{\varphi_j\}_{j\in\mathcal{J}}$ is *subordinate to* $\{W_p\}_{p\in\mathcal{P}}$ if for each $j \in \mathcal{J}$ there exists $p_j \in \mathcal{P}$ such that $\operatorname{supp}\varphi_j \subset W_{p_j}$.

Theorem A.11.1. *Given any open cover $\{W_p\}_{p\in\mathcal{P}}$ of a subset $E \subset \mathbb{R}^n$, there exists a partition of unity $\{\varphi_j\}_{j\in\mathcal{J}}$ for E subordinate to $\{W_p\}_{p\in\mathcal{P}}$; moreover, $\{\varphi_j\}_{j\in\mathcal{J}}$ can be chosen in such a way that* $\operatorname{supp}\varphi_j$ *is compact for each $j \in \mathcal{J}$.*

Corollary A.11.2. *Given any countable open cover $\{W_p\}_{p\in\mathbb{N}}$ of a subset $E \subset \mathbb{R}^n$, there exists a countable partition of unity $\{\varphi_j\}_{j\in\mathbb{N}}$ for E such that* $\operatorname{supp}\varphi_j \subset W_j$ *for each $j \in \mathbb{N}$.*

A.12. Lipschitz domains and C^k domains

References: [50].

We denote the boundary of the open subset $\Omega \subset \mathbb{R}^n$ by

$$\partial\Omega = \bar{\Omega} \cap (\mathbb{R}^n \setminus \Omega). \tag{A.99}$$

Till now, we have made no regularity assumptions on $\partial\Omega$, but from now on we shall require that, roughly speaking, $\partial\Omega$ can be locally represented as the graph of a Lipschitz function (using different systems of Cartesian coordinates for different parts of $\partial\Omega$, if necessary). The simplest case occurs when there exists a function $\zeta : \mathbb{R}^{n-1} \to \mathbb{R}$ such that

$$\Omega = \left\{x = (x', x_n) \in \mathbb{R}^n \mid x_n < \zeta(x') \ \ \forall x' := (x_1, \dots, x_{n-1}) \in \mathbb{R}^{n-1}\right\}; \tag{A.100}$$

if ζ is a Lipschitz function, i.e. if there exists a constant L such that

$$|\zeta(x') - \zeta(y')| \le L \, \|x' - y'\|_{\mathbb{R}^{n-1}} \qquad \forall x', y' \in \mathbb{R}^{n-1}, \tag{A.101}$$

then we say that Ω is a *Lipschitz hypograph*. Clearly, if Ω is given by (A.100), then its boundary is

$$\partial\Omega := \{x = (x', x_n) \in \mathbb{R}^n \mid x_n = \zeta(x') \ \ \forall x' := (x_1, \dots, x_{n-1}) \in \mathbb{R}^{n-1}\}. \tag{A.102}$$

Definition A.12.1. *The open subset $\Omega \subset \mathbb{R}^n$ is called a* Lipschitz domain *if its boundary $\partial\Omega$ is compact and if there exist finite families $\{W_j\}_{j=0}^J$ and $\{\Omega_j\}_{j=0}^J$ of subsets of $\mathbb{R}^n$ having the following properties:*

1. *the family $\{W_j\}_{j=0}^J$ is a finite open cover of $\partial\Omega$, i.e. each W_j is an open subset of $\mathbb{R}^n$ and $\partial\Omega \subset \bigcup_j W_j$;*
2. *each Ω_j can be transformed into a Lipschitz hypograph by a rigid motion, i.e. by a rototranslation;*
3. *the condition $W_j \cap \Omega = W_j \cap \Omega_j$ holds for each $j = 0, \dots J$.*

It is worthwhile observing that although, according to definition A.12.1, the boundary of a Lipschitz domain must be compact, the domain itself may be unbounded. In particular, if Ω is a bounded Lipschitz domain, then $\mathbb{R}^n \setminus \bar{\Omega}$ is an unbounded Lipschitz domain. In any case, it is clear that a Lipschitz hypograph Ω, given by (A.100), cannot be a Lipschitz domain, since its boundary $\partial\Omega$, as shown by relation (A.102), is not compact, even if ζ were compactly supported.

Sometimes, a different smoothness condition will be needed, so we extend the above terminology in the following way. For any $k \in \mathbb{N}$, we firstly say that the set (A.100) is a C^k *hypograph* if the function $\zeta : \mathbb{R}^{n-1} \to \mathbb{R}$ is an element of $C^k(\mathbb{R}^n)$ and $\partial^\alpha \zeta$ is bounded for all α such that $|\alpha|_{\mathbb{N}^n} \leq k$. Then, we obviously define a C^k domain by replacing "Lipschitz" with "C^k" throughout definition A.12.1.

Moreover, for any $\mu \in (0, 1]$, we can analogously define a $C^{k,\mu}$ domain as a C^k domain endowed with the additional property that the kth-order partial derivatives of each function ζ_j relative to the C^k hypograph Ω_j are Hölder-continuous with exponent μ, i.e.

$$|\partial^\alpha \zeta_j(x') - \partial^\alpha \zeta_j(y')| \leq L \, \|x' - y'\|^\mu_{\mathbb{R}^{n-1}} \qquad \forall x', y' \in \mathbb{R}^{n-1}, \ \forall \alpha \ \mid \ |\alpha|_{\mathbb{N}^n} = k. \tag{A.103}$$

Hence, a Lipschitz domain is the same thing as a $C^{0,1}$ domain; we also point out that if $k \geq 1$, then a C^k domain is also a Lipschitz domain. Moreover, if we want, we can always regard each ζ_j as compactly supported, since $\partial\Omega$ is always assumed to be compact.

Finally, we point out that in chapter 2 we sometimes use the expression "domain with $C^{k,\mu}$ boundary" (or similar ones) as synonymous of "$C^{k,\mu}$ domain".

A.13. Links between the two families of Sobolev spaces (2)

References: [50].

Making suitable assumptions on the smoothness of the boundary $\partial\Omega$, it is possible to prove some further relations, besides the ones stated in section A.10, between the two families of Sobolev spaces above introduced.

Theorem A.13.1. *If Ω is a C^0 domain, then*

1. *$\mathcal{D}(\bar{\Omega})$ is dense in $W^s(\Omega)$ $\forall s \geq 0$;*
2. *$\mathcal{D}(\Omega)$ is dense in $H^s_{\bar{\Omega}}$ or, in other words, $\widetilde{H}^s(\Omega) = H^s_{\bar{\Omega}}$ $\forall s \in \mathbb{R}$.*

Theorem A.13.2. *If Ω is a Lipschitz domain, then*

1. *$[H^s(\Omega)]^* = \widetilde{H}^{-s}(\Omega)$ and $[\widetilde{H}^s(\Omega)]^* = H^{-s}(\Omega)$ $\forall s \in \mathbb{R}$;*

2. $W^{s,2}(\Omega) = H^s(\Omega)\ \forall s \geq 0$ *with equivalent norms.*

Theorem A.13.3. *Let* $s \geq 0$*. If* Ω *is a Lipschitz domain, then it holds*

$$\widetilde{H}^s(\Omega) = \{u \in L^2(\Omega) \mid \tilde{u} \in H^s(\mathbb{R}^n)\} \subset H_0^s(\Omega), \tag{A.104}$$

where $\tilde{u}$ *denotes the extension of* u *by zero, i.e.*

$$\tilde{u}(x) := \begin{cases} u(x) & \text{if } x \in \Omega \\ 0 & \text{if } x \in \mathbb{R}^n \setminus \Omega. \end{cases} \tag{A.105}$$

In fact, it holds

$$\widetilde{H}^s(\Omega) = H_0^s(\Omega) \qquad \forall s \in [0, +\infty) \setminus \left\{ q \in \mathbb{Q} \mid \exists n \in \mathbb{N} \text{ such that } q = \frac{2n+1}{2} \right\}. \tag{A.106}$$

A.14. Sobolev spaces on the boundary

References: [15], [47], [50].

Any Lipschitz domain Ω has a surface measure σ and an outward unit normal ν that exists σ-almost everywhere on $\partial\Omega$. In fact, by Rademacher's theorem, if $\zeta : \mathbb{R}^{n-1} \to \mathbb{R}$ is a Lipschitz function, then ζ is Fréchet-differentiable almost everywhere and

$$\text{ess sup} \, \|\nabla_{n-1}\zeta(x')\|_{\mathbb{R}^{n-1}} \leq \sqrt{n-1}\ L, \tag{A.107}$$

where we have denoted with "ess sup" the essential supremum, with ∇_{n-1} the classical gradient operator in $\mathbb{R}^{n-1}$ and with L any Lipschitz constant for ζ, according to (A.101).

If Ω is the Lipschitz hypograph (A.100), then, for any $x = (x', \zeta(x')) \in \partial\Omega$, it holds:

$$d\sigma(x) = \sqrt{1 + \|\nabla_{n-1}\zeta(x')\|^2_{\mathbb{R}^{n-1}}}\, dx' \tag{A.108}$$

and

$$\nu(x) = \frac{(-\nabla_{n-1}\zeta(x'), 1)}{\sqrt{1 + \|\nabla_{n-1}\zeta(x')\|^2_{\mathbb{R}^{n-1}}}}; \tag{A.109}$$

in particular, relation (A.108) allows one to define the boundary integral of a function $u : \partial\Omega \to \mathbb{C}$ as

$$\int_{\partial\Omega} u(x) d\sigma(x) := \int_{\mathbb{R}^{n-1}} u(x', \zeta(x'))\, \sqrt{1 + \|\nabla_{n-1}\zeta(x')\|^2_{\mathbb{R}^{n-1}}}\, dx' \tag{A.110}$$

whenever the integral at the right-hand side exists.

If Ω is a Lipschitz domain, let $\{W_j\}_{j=0}^J$ be a finite open cover of $\partial\Omega$ as in definition A.12.1. By virtue of corollary A.11.2, we can choose a partition of unity $\{\varphi_j\}_{j=0}^J$ for $\partial\Omega$ such that

$\varphi_j|_{W_j} \in \mathcal{D}(W_j)$ for each $j = 0, \dots J$. Since it holds, by definition of partition of unity, $\sum_{j=0}^{J} \varphi_j(x) = 0$ for all $x \in \partial\Omega$, the natural extension of definition (A.110) reads as:

$$\int_{\partial\Omega} u(x) d\sigma(x) := \sum_{j=0}^{J} \int_{\mathbb{R}^{n-1}} \varphi_j(x', \zeta_j(x'))\, u(x', \zeta_j(x')) \sqrt{1 + \|\nabla_{n-1}\zeta_j(x')\|^2_{\mathbb{R}^{n-1}}}\, dx'. \tag{A.111}$$

By the way, we observe that relations (A.108), (A.109), the concept of partition of unity and the definition of boundary integral allow one to prove the divergence theorem in more general hypotheses than the ones traditionally assumed. Also for future purpose, we state this result here below, denoting with $\nabla_n \cdot$ the classical divergence operator in $\mathbb{R}^n$.

Theorem A.14.1. *If $\Omega \subset \mathbb{R}^n$ is a Lipschitz domain, ν is the outward unit normal to $\partial\Omega$ existing σ-almost everywhere on $\partial\Omega$ and $\mathbf{V} : \mathbb{R}^n \to \mathbb{R}^n$ is a C^1 vector field with compact support, then*

$$\int_{\Omega} \nabla_n \cdot \mathbf{V}\, dx = \int_{\partial\Omega} \mathbf{V} \cdot \nu\, d\sigma(x). \tag{A.112}$$

Now, as regards the Sobolev spaces defined on the boundary of a domain Ω, we firstly deal with the simplest case, in which Ω is a Lipschitz hypograph: then we can construct Sobolev spaces on its boundary $\partial\Omega$ in terms of Sobolev spaces on $\mathbb{R}^{n-1}$, in the following way. Let us consider the space $L^2(\partial\Omega) \equiv L^2(\partial\Omega, \sigma)$, i.e. the set of all the functions $u : \partial\Omega \to \mathbb{C}$ satisfying

$$\int_{\partial\Omega} |u(x)|^2\, d\sigma(x) \equiv \int_{\mathbb{R}^{n-1}} |u(x', \zeta(x'))|^2 \sqrt{1 + \|\nabla_{n-1}\zeta(x')\|^2_{\mathbb{R}^{n-1}}}\, dx' < \infty, \tag{A.113}$$

which is a Hilbert space when equipped with the scalar product

$$\begin{aligned}(u, v)_{L^2(\partial\Omega)} &:= \int_{\partial\Omega} u(x)\bar{v}(x) d\sigma(x) \equiv \\ &\equiv \int_{\mathbb{R}^{n-1}} u(x', \zeta(x'))\, \overline{v(x', \zeta(x'))} \sqrt{1 + \|\nabla_{n-1}\zeta(x')\|^2_{\mathbb{R}^{n-1}}}\, dx' \qquad \forall u, v \in L^2(\partial\Omega)\end{aligned} \tag{A.114}$$

and the induced norm

$$\|u\|_{L^2(\partial\Omega)} := \sqrt{(u, u)_{L^2(\partial\Omega)}}\,. \tag{A.115}$$

For any $u \in L^2(\partial\Omega)$, we put

$$u_\zeta(x') := u(x', \zeta(x')) \qquad \text{for } x' \in \mathbb{R}^{n-1}; \tag{A.116}$$

then we define the space

$$H^s(\partial\Omega) := \{u \in L^2(\partial\Omega) \mid u_\zeta \in H^s(\mathbb{R}^{n-1})\} \qquad \forall s \in [0, 1], \tag{A.117}$$

which is a Hilbert space when equipped with the scalar product

$$(u, v)_{H^s(\partial\Omega)} := (u_\zeta, v_\zeta)_{H^s(\mathbb{R}^{n-1})} \tag{A.118}$$

and the induced norm

$$\|u\|_{H^s(\partial\Omega)} := \sqrt{(u, u)_{H^s(\partial\Omega)}} \ . \tag{A.119}$$

If $u \in L^2(\partial\Omega)$, it follows that $u_\zeta \sqrt{1 + \|\nabla_{n-1}\zeta\|^2_{\mathbb{R}^{n-1}}} \in H^{-s}(\mathbb{R}^{n-1})$ for all $s \in (0, 1]$, so one can put

$$\|u\|_{H^{-s}(\partial\Omega)} := \left\| u_\zeta \sqrt{1 + \|\nabla_{n-1}\zeta\|^2_{\mathbb{R}^{n-1}}} \right\|_{H^{-s}(\mathbb{R}^{n-1})} \qquad \forall s \in (0, 1] \tag{A.120}$$

and define $H^{-s}(\partial\Omega)$ as the completion of $L^2(\partial\Omega)$ in this norm. Then, it can be shown that $H^{-s}(\partial\Omega)$ is actually a realization of the dual space of $H^s(\partial\Omega)$ and that the norm defined in (A.120) is equivalent to that of $[H^s(\partial\Omega)]^*$, clearly defined as

$$\|u\|_{[H^s(\partial\Omega)]^*} := \sup_{0 \neq v \in H^s(\partial\Omega)} \frac{|\langle u, v\rangle_{\partial\Omega}|}{\|v\|_{H^s(\partial\Omega)}} \qquad \forall u \in [H^s(\partial\Omega)]^*, \tag{A.121}$$

where we have obviously denoted with $\langle u, v\rangle_{\partial\Omega}$ the pairing between an element $u \in [H^s(\partial\Omega)]^* = H^{-s}(\partial\Omega)$ and an element $v \in H^s(\partial\Omega)$. We can now introduce the following useful notation:

$$\int_{\partial\Omega} u(x)\, v(x)\, d\sigma(x) := \langle u, v\rangle_{\partial\Omega} \qquad \forall u \in H^{-s}(\partial\Omega),\ \forall v \in H^s(\partial\Omega). \tag{A.122}$$

To this purpose, we observe that while v is always an element of $L^2(\partial\Omega)$ (since, by virtue of definition (A.117), it holds $H^s(\partial\Omega) \subset L^2(\partial\Omega)$), in general u is not, since it is a distribution belonging to $H^{-s}(\partial\Omega) \supset L^2(\partial\Omega)$: hence the left-hand side of (A.122) is, in general, only a different notation for the pairing at the right-hand side. However, when $u \in L^2(\partial\Omega) \cap H^{-s}(\partial\Omega)$, the integral in (A.122) is well-defined on its own and it is possible to prove that it is equal to the pairing at the right-hand side.

All the previous results can be immediately extended to the case in which not Ω itself, but rather $\kappa(\Omega)$ is a Lipschitz hypograph for some rototranslation $\kappa : \mathbb{R}^n \to \mathbb{R}^n$: to this end, it suffices to put $u_\zeta(x') := u[\kappa^{-1}(x', \zeta(x'))]$ and then to define $H^s(\partial\Omega)$ in the same way as above.

Finally, if Ω is a Lipschitz domain, we can proceed as follows. Using also the notation of definition A.12.1, let $\kappa_j : \mathbb{R}^n \to \mathbb{R}^n$ be a rototranslation that transforms each Ω_j into a Lipschitz hypograph and let $(x', \zeta_j(x'))$, with $\zeta_j : \mathbb{R}^{n-1} \to \mathbb{R}$, be a parametric representation of $\partial\Omega_j$; finally, let $u_{\zeta_j}(x') := u[\kappa_j^{-1}(x', \zeta_j(x'))]$. Then we define

$$H^s(\partial\Omega) := \{u \in L^2(\partial\Omega) \mid u_{\zeta_j} \in H^s(\mathbb{R}^{n-1}) \ \ \forall j\}. \tag{A.123}$$

In order to define a scalar product in $H^s(\partial\Omega)$, let $\{W_j\}_{j=0}^J$ be a finite open cover of $\partial\Omega$ as in definition A.12.1. By virtue of corollary A.11.2, we can choose a partition of unity $\{\varphi_j\}_{j=0}^J$ for

$\partial\Omega$ such that $\varphi_j|_{W_j} \in \mathcal{D}(W_j)$ for each $j = 0, \dots J$; then we put

$$(u, v)_{H^s(\partial\Omega)} := \sum_{j=0}^{J} (\varphi_j u, \varphi_j v)_{H^s(\partial\Omega_j)}. \tag{A.124}$$

It is possible to prove that definitions (A.123) and (A.124) are well-posed, in the sense that a different choice of $\{\Omega_j\}_{j=0}^J$, $\{W_j\}_{j=0}^J$ and $\{\varphi_j\}_{j=0}^J$ would yield the same space $H^s(\partial\Omega)$ with an equivalent norm, for all s such that $|s| \leq 1$.

Finally, if Ω is a $C^{k-1,1}$ domain for $k \in \mathbb{N} \setminus \{0\}$, then it is possible to analogously define $H^s(\partial\Omega)$ for each s such that $|s| \leq k$.

A.15. Trace operators (1)

References: [15], [50].

In studying boundary value problems, one often needs to give a meaning to the restriction $u|_{\partial\Omega}$ as an element of a Sobolev space on $\partial\Omega$ when u belongs to a Sobolev space on Ω. To this purpose, the main idea is expressed by the following theorem.

Theorem A.15.1. *Let γ be the* trace operator *defined as*

$$\begin{aligned} \gamma : \mathcal{D}(\mathbb{R}^n) &\longrightarrow \mathcal{D}(\mathbb{R}^{n-1}) \\ u \quad &\longmapsto [\gamma u](x') := u(x', 0) \qquad \forall x' \in \mathbb{R}^{n-1}; \end{aligned} \tag{A.125}$$

if $s > \frac{1}{2}$, then γ has a unique extension (still denoted with γ) to a bounded linear operator

$$\gamma : H^s(\mathbb{R}^n) \longrightarrow H^{s-\frac{1}{2}}(\mathbb{R}^{n-1}), \tag{A.126}$$

and this extension has a continuous right inverse. In other terms, if $s > \frac{1}{2}$:

1. *there exists a constant $C > 0$ such that*

$$\|\gamma u\|_{H^{s-\frac{1}{2}}(\mathbb{R}^{n-1})} \leq C \|u\|_{H^s(\mathbb{R}^n)} \qquad \forall u \in H^s(\mathbb{R}^n); \tag{A.127}$$

2. *there exist a linear operator $\eta : H^{s-\frac{1}{2}}(\mathbb{R}^{n-1}) \to H^s(\mathbb{R}^n)$ and a constant $C' > 0$ such that for any $f \in H^{s-\frac{1}{2}}(\mathbb{R}^{n-1})$ it holds $\gamma(\eta f) = f$ and*

$$\|\eta f\|_{H^s(\mathbb{R}^n)} \leq C' \|f\|_{H^{s-\frac{1}{2}}(\mathbb{R}^{n-1})}. \tag{A.128}$$

For Sobolev spaces on domains $\Omega \subset \mathbb{R}^n$, the following result holds.

Theorem A.15.2. *Let Ω be an open subset of $\mathbb{R}^n$ and let γ_1 be the* trace operator *defined as*

$$\begin{aligned} \gamma_1 : \mathcal{D}(\bar{\Omega}) &\longrightarrow \mathcal{D}(\partial\Omega) \\ u \quad &\longmapsto \gamma_1 u := u|_{\partial\Omega}. \end{aligned} \tag{A.129}$$

If Ω is a $C^{k-1,1}$ domain for a certain $k \in \mathbb{N} \setminus \{0\}$ and if $\frac{1}{2} < s \leq k$, then γ_1 has a unique extension (still denoted with γ_1) to a bounded linear operator

$$\gamma_1 : H^s(\Omega) \longrightarrow H^{s-\frac{1}{2}}(\partial\Omega), \tag{A.130}$$

and this extension has a continuous right inverse. In other terms, under the specified hypotheses:

1. *there exists a constant $C_1 > 0$ such that*

$$\|\gamma_1 u\|_{H^{s-\frac{1}{2}}(\partial\Omega)} \leq C_1 \|u\|_{H^s(\Omega)} \qquad \forall u \in H^s(\Omega); \tag{A.131}$$

2. *there exist a linear operator $\eta_1 : H^{s-\frac{1}{2}}(\partial\Omega) \to H^s(\Omega)$ and a constant $C_1' > 0$ such that for any $f \in H^{s-\frac{1}{2}}(\partial\Omega)$ it holds $\gamma_1(\eta_1 f) = f$ and*

$$\|\eta_1 f\|_{H^s(\Omega)} \leq C_1' \|f\|_{H^{s-\frac{1}{2}}(\partial\Omega)}. \tag{A.132}$$

The previous theorem holds, in particular, for $\frac{1}{2} < s \leq 1$ if Ω is a $C^{0,1}$ domain, i.e. a Lipschitz domain: in such a case, however, it possible to prove a stronger result, due to Costabel [30], establishing the boundedness of the trace operator (A.130) for $\frac{1}{2} < s < \frac{3}{2}$.

Finally, we give the following result, which states when and in which sense the elements of $H_0^s(\Omega)$ are functions vanishing at the boundary of Ω.

Theorem A.15.3. *Let $\Omega \subset \mathbb{R}^n$ be a $C^{k-1,1}$ domain, with $k \in \mathbb{N} \setminus \{0\}$; thus*

1. *if $0 \leq s \leq \frac{1}{2}$, then $H_0^s(\Omega) = H^s(\Omega)$;*
2. *if $\frac{1}{2} < s \leq k$, then $H_0^s(\Omega) = \{u \in H^s(\Omega) \;\mid\; \gamma_1(\partial^\alpha u) = 0 \;\; \forall \alpha \in \mathbb{N}^n : |\alpha|_{\mathbb{N}^n} < s - \frac{1}{2}\}$.*

A.16. Green identities

References: [27], [47], [50].

We denote with ∇_n, $\nabla_n \cdot$ and Δ_n the classical gradient, divergence and laplacian operators in $\mathbb{R}^n$. Moreover, if $\xi = (\xi_1, \ldots, \xi_n)$ and $\tau = (\tau_1, \ldots, \tau_n)$ are two elements of $\mathbb{C}^n$, we define the operation

$$\xi \cdot \tau := \sum_{i=1}^{n} \xi_i \tau_i, \tag{A.133}$$

which is clearly not a scalar product on $\mathbb{C}^n$, but it is actually the canonical scalar product on $\mathbb{R}^n$ if restricted to $\mathbb{R}^n$ itself[7].

Theorem A.16.1. *Let $\Omega \subset \mathbb{R}^n$ be a Lipschitz domain and ν the outward unit normal to $\partial\Omega$ existing σ-almost everywhere on $\partial\Omega$. Moreover, let $u, v : \mathbb{R}^n \to \mathbb{C}$ be two functions such that at least one of them is compactly supported. Then:*

1. *if $u \in C^1(\mathbb{R}^n)$ and $v \in C^2(\mathbb{R}^n)$, the* first Green identity *holds:*

$$\int_\Omega (\nabla_n u \cdot \nabla_n v + u\Delta_n v)\, dx = \int_{\partial\Omega} u \frac{\partial v}{\partial \nu} d\sigma; \tag{A.134}$$

2. *if $u, v \in C^2(\mathbb{R}^n)$, the* second Green identity *holds:*

$$\int_\Omega (u\Delta_n v - v\Delta_n u)\, dx = \int_{\partial\Omega} \left(u\frac{\partial v}{\partial \nu} - v\frac{\partial u}{\partial \nu} \right) d\sigma. \tag{A.135}$$

Proof. To establish (A.134), it suffices to substitute in relation (A.112), expressing the divergence theorem, the compactly supported, C^1 vector field $\mathbf{V} := u\nabla_n v$ and use the well-known identity

$$\nabla_m \cdot (u\nabla_n v) = \nabla_n u \cdot \nabla_n v + u\nabla_n \cdot \nabla_n v. \tag{A.136}$$

To establish (A.135), it suffices to rewrite the first Green identity interchanging u and v and subtract it to the original one, written as in (A.134). ■

Remark A.16.1. A more traditional formulation of the previous theorem A.16.1 states the same identities (A.134) and (A.135) under the hypothesis that Ω is a bounded, C^1 domain and that $u \in C^1(\bar{\Omega})$, $v \in C^2(\bar{\Omega})$ or, respectively, that $u, v \in C^2(\bar{\Omega})$. □

A.17. Trace operators (2)

References: [15].

Besides theorem A.15.2, we need another, more technical, trace theorem, involving the Sobolev space $H^1(D, \Delta_{\mathbf{A}'})$, which we are going to define. Throughout this section, we deal with bounded, C^2 domains[8] in $\mathbb{R}^2$.

[7] For example, in the left-hand side integral of identity (A.134), the expression $\nabla_n u \cdot \nabla_n v$ is to be intended as $\sum_{i=1}^n \partial_i u\ \partial_i v$.

[8] We recall that a C^k domain, by definition, is open (cf. definition A.12.1 and the subsequent comment).

Then, let $D \subset \mathbb{R}^2$ be a bounded, C^2 domain and let $\mathbf{A}' : \bar{D} \to \mathbb{C}^{2\times 2}$, with $\mathbf{A}' = (a'_{jk})_{j,k=1,2}$, be a matrix-valued function such that $a'_{jk} \in C^1(\bar{D})\ \forall j,k = 1,2$. Moreover, let $\nabla_2 \cdot$ and ∇_2 the *weak* divergence and gradient operators in $\mathbb{R}^2$ respectively. Then we define the Sobolev space

$$H^1(D, \Delta_{\mathbf{A}'}) := \left\{ u \in H^1(D) \mid \nabla_2 \cdot \mathbf{A}' \nabla_2 u \in L^2(D) \right\}, \tag{A.137}$$

equipped with the norm

$$\|u\|^2_{H^1(D,\Delta_{\mathbf{A}'})} := \|u\|^2_{H^1(D)} + \|\nabla_2 \cdot \mathbf{A}' \nabla_2 u\|^2_{L^2(D)}. \tag{A.138}$$

Moreover, if ν denotes the unit normal vector to ∂D, directed into the exterior of D, we can define the *conormal derivative* of a function $u \in \mathcal{D}(\bar{D})$ as

$$\frac{\partial u}{\partial \nu_{\mathbf{A}'}} := (\nu \cdot \mathbf{A}' \nabla_2 u)\,|_{\partial D}, \tag{A.139}$$

where the gradient ∇_2 at the right-hand side of relation (A.139) is to be intended in the classical sense.

Theorem A.17.1. *Let $D \subset \mathbb{R}^2$ be a bounded and C^2 domain, let $\mathbf{A}' : \bar{D} \to \mathbb{C}^{2\times 2}$, with $\mathbf{A}' = (a'_{jk})_{j,k=1,2}$, be a matrix-valued function such that $a'_{jk} \in C^1(\bar{D})\ \forall j,k = 1,2$ and let $\gamma_2 : \mathcal{D}(\bar{D}) \to C^1(\partial D)$ be the trace operator defined as*

$$\gamma_2 u := \frac{\partial u}{\partial \nu_{\mathbf{A}'}} \qquad \forall u \in \mathcal{D}(\bar{D}). \tag{A.140}$$

Then γ_2 has a unique extension to a bounded linear operator, still denoted with γ_2,

$$\gamma_2 : H^1(D, \Delta_{\mathbf{A}'}) \to H^{-\frac{1}{2}}(\partial D). \tag{A.141}$$

In other terms, there exists a constant $C_2 > 0$ such that

$$\left\| \frac{\partial u}{\partial \nu_{\mathbf{A}'}} \right\|_{H^{-\frac{1}{2}}(\partial D)} \leq C_2 \|u\|_{H^1(D,\Delta_{\mathbf{A}'})} \qquad \forall u \in H^1(D, \Delta_{\mathbf{A}'}). \tag{A.142}$$

Proof. See theorem 5.5 in [15]. ■

Remark A.17.1. By means of arguments similar to the ones used in the proof of theorem A.17.1 (see remark 5.7 in [15]) and with the help of a cutoff function for a neighbourhood of ∂D (see lemma 5.4 in [15]), it is possible to define $\dfrac{\partial u}{\partial \nu_{\mathbf{A}'}} \in H^{-\frac{1}{2}}(\partial D)$ for functions u belonging to the Sobolev space

$$H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D}, \Delta_{\mathbf{A}'}) := \left\{ u \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D}) \mid \nabla_2 \cdot \mathbf{A}' \nabla_2 u \in L^2_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D}) \right\}, \tag{A.143}$$

where, in turn, being $\Omega_R := \{x \in \mathbb{R}^2 \mid \|x\|_{\mathbb{R}^2} < R\}$:

$$H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D}) := \{u : \mathbb{R}^2 \setminus \bar{D} \to \mathbb{C} \mid u \in H^1\left((\mathbb{R}^2 \setminus \bar{D}) \cap \Omega_R\right) \ \ \forall R > 0 : (\mathbb{R}^2 \setminus \bar{D}) \cap \Omega_R \neq \emptyset\}, \tag{A.144}$$

$$L^2_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D}) := \{u : \mathbb{R}^2 \setminus \bar{D} \to \mathbb{C} \mid u \in L^2\left((\mathbb{R}^2 \setminus \bar{D}) \cap \Omega_R\right) \ \ \forall R > 0 : (\mathbb{R}^2 \setminus \bar{D}) \cap \Omega_R \neq \emptyset\}. \tag{A.145}$$

□

Remark A.17.2. By putting $\mathbf{A}' = \mathbf{I}$ in theorem A.17.1 (where $\mathbf{I}$ is obviously the identity matrix), we have that $\dfrac{\partial u}{\partial \nu}$ is well defined as an element of $H^{-\frac{1}{2}}(\partial D)$ for functions u belonging to the Sobolev space

$$H^1(D, \Delta_2) := \left\{u \in H^1(D) \mid \Delta_2 u \in L^2(D)\right\}, \tag{A.146}$$

equipped with the norm

$$\|u\|^2_{H^1(D,\Delta_2)} := \|u\|^2_{H^1(D)} + \|\Delta_2 u\|^2_{L^2(D)}. \tag{A.147}$$

In such a case, relation (A.142) obviously becomes:

$$\left\|\frac{\partial u}{\partial \nu}\right\|_{H^{-\frac{1}{2}}(\partial D)} \leq C_2 \|u\|_{H^1(D,\Delta_2)} \qquad \forall u \in H^1(D, \Delta_2). \tag{A.148}$$

Of notable interest for various applications is the case in which $u \in H^1(D)$ is a weak solution of the Helmholtz equation $\Delta_2 u + k^2 u = 0$ in D (with $k^2 > 0$): since it clearly holds $\Delta_2 u = -k^2 u$ almost everywhere in D, we have that $\Delta_2 u \in L^2(D)$ and then $u \in H^1(D, \Delta_2)$. Moreover, from definition (A.147) we easily get

$$\|u\|^2_{H^1(D,\Delta_2)} = \|u\|^2_{H^1(D)} + \| - k^2 u\|^2_{L^2(D)} \leq (1 + k^4)\|u\|^2_{H^1(D)}. \tag{A.149}$$

Put $C_3 := C_2\sqrt{1 + k^4}$, from relations (A.148) and (A.149) it immediately follows

$$\left\|\frac{\partial u}{\partial \nu}\right\|_{H^{-\frac{1}{2}}(\partial D)} \leq C_3 \|u\|_{H^1(D)} \tag{A.150}$$

for any $u \in H^1(D, \Delta_2)$ satisfying the Helmholtz equation in D; relation (A.150) is often useful in chapter 2.

By virtue of the previous remark A.17.1, analogous results clearly hold if, instead of the bounded domain D, the unbounded domain $\mathbb{R}^2 \setminus \bar{D}$ is considered. □

A.18. Generalized Green identities and their consequences

References: [15].

As a consequence of theorem A.17.1, it is possible to extend Gauss divergence theorem to a wider space of functions and consequently to obtain the following generalization of the first and second Green identities: if D and $\mathbf{A}'$ are as in theorem A.17.1, then it holds (see theorem 5.5 and corollary 5.6 in [15]):

$$\int_D (\nabla_2 u \cdot \mathbf{A}' \nabla_2 v + u \nabla_2 \cdot \mathbf{A}' \nabla_2 v)\, dx = \int_{\partial D} u \frac{\partial v}{\partial \nu_{\mathbf{A}'}}\, d\sigma \quad \forall u \in H^1(D),\ \forall v \in H^1(D, \Delta_{\mathbf{A}'}); \tag{A.151}$$

$$\int_D (\nabla_2 u \cdot \mathbf{A}' \nabla_2 v - \nabla_2 v \cdot \mathbf{A}' \nabla_2 u + u \nabla_2 \cdot \mathbf{A}' \nabla_2 v - v \nabla_2 \cdot \mathbf{A}' \nabla_2 u)\, dx =$$

$$= \int_{\partial D} \left(u \frac{\partial v}{\partial \nu_{\mathbf{A}'}} - v \frac{\partial u}{\partial \nu_{\mathbf{A}'}} \right) d\sigma \qquad \forall u, v \in H^1(D, \Delta_{\mathbf{A}'}). \tag{A.152}$$

Of course, if we assume that the matrix $\mathbf{A}'(x)$ is symmetric for all $x \in \bar{D}$, the second Green identity (A.152) can be written in the simpler form:

$$\int_D (u \nabla_2 \cdot \mathbf{A}' \nabla_2 v - v \nabla_2 \cdot \mathbf{A}' \nabla_2 u)\, dx = \int_{\partial D} \left(u \frac{\partial v}{\partial \nu_{\mathbf{A}'}} - v \frac{\partial u}{\partial \nu_{\mathbf{A}'}} \right) d\sigma \qquad \forall u, v \in H^1(D, \Delta_{\mathbf{A}'}). \tag{A.153}$$

We remark that, in general, the boundary integrals at the right-hand side of identities (A.151), (A.152) and (A.153) are to be intended in the pairing sense, as explained about definition (A.122).

We can now state the following theorem.

Theorem A.18.1. *Let D, $\mathbf{A}'$ be as in theorem A.17.1, let $\mathbf{A}'(x)$ be symmetric for all $x \in \bar{D}$ and let $n \in C^0(\bar{D})$; moreover, let us assume that $u, v \in H^1(D)$ are weak solutions of the equation:*

$$\nabla_2 \cdot \mathbf{A}' \nabla_2 w + k^2 n w = 0 \quad \text{in } D; \tag{A.154}$$

then, u and v verify the identity

$$\int_{\partial D} \left(u \frac{\partial v}{\partial \nu_{\mathbf{A}'}} - v \frac{\partial u}{\partial \nu_{\mathbf{A}'}} \right) d\sigma = 0. \tag{A.155}$$

Proof. Since, by hypothesis, $u, v \in H^1(D)$, $n \in C^0(\bar{D})$ and

$$\nabla_2 \cdot \mathbf{A}' \nabla_2 u = -k^2 n u, \quad \nabla_2 \cdot \mathbf{A}' \nabla_2 v = -k^2 n v \quad \text{a.e.}^9 \text{ in } D, \tag{A.156}$$

[9] To be read: almost everywhere.

it easily follows that $\nabla_2 \cdot \mathbf{A}'\nabla_2 u$, $\nabla_2 \cdot \mathbf{A}'\nabla_2 v \in L^2(D)$ and, consequently, $u, v \in H^1(D, \Delta_{\mathbf{A}'})$. Hence, by virtue of the assumed symmetry of the matrix $\mathbf{A}'(x)$ for all $x \in \bar{D}$, we can use identity (A.153) for the functions u, v and the domain D, and the thesis (A.155) easily follows from observing that, by virtue of relations (A.156), the function to be integrated at the left-hand side of (A.153) is zero almost everywhere in D. ■

Remark A.18.1. By putting $\mathbf{A}' = \mathbf{I}$ in identities (A.151), (A.152), we obtain a simpler and more familiar generalization of the first and second Green identities (A.134), (A.135), respectively in the form:

$$\int_D (\nabla_2 u \cdot \nabla_2 v + u\Delta_2 v)\, dx = \int_{\partial D} u \frac{\partial v}{\partial \nu} d\sigma \qquad \forall u \in H^1(D), \quad \forall v \in H^1(D, \Delta_2); \tag{A.157}$$

$$\int_D (u\Delta_2 v - v\Delta_2 u)\, dx = \int_{\partial D} \left(u\frac{\partial v}{\partial \nu} - v\frac{\partial u}{\partial \nu} \right) d\sigma \qquad \forall u, v \in H^1(D, \Delta_2). \tag{A.158}$$

□

From the generalized second Green identity (A.158) one can deduce some technical results, which we are going to illustrate in the next theorem.

Theorem A.18.2. *Let $D \subset \mathbb{R}^2$ be a bounded and C^2 domain; moreover, let us assume that*

1. *either $u, v \in H^1(D)$ are weak solutions of the Helmholtz equation $\Delta_2 w + k^2 w = 0$ in D,*
2. *or $u, v \in H^1_{\partial D, loc}(\mathbb{R}^2 \setminus \bar{D})$ are weak solutions of the Helmholtz equation $\Delta_2 w + k^2 w = 0$ in $\mathbb{R}^2 \setminus \bar{D}$ and satisfy the so-called* Sommerfeld radiation condition[10]

$$\lim_{r \to \infty} \sup_{\partial \Omega_r} \left[\sqrt{r} \left(\frac{\partial w}{\partial r} - \mathrm{i}kw \right) \right] = 0, \tag{A.159}$$

where r denotes the first of the polar coordinates in $\mathbb{R}^2$ and $\Omega_r := \{x \in \mathbb{R}^2 \mid \|x\|_{\mathbb{R}^2} < r\}$;

then, in both cases, u and v verify the identity:

$$\int_{\partial D} \left(u\frac{\partial v}{\partial \nu} - v\frac{\partial u}{\partial \nu} \right) d\sigma = 0. \tag{A.160}$$

[10] For some hints about it, we refer to relations (2.20), (2.33), (2.34), definition 2.1.1 and remark 2.2.1: the latter, in particular, states that the radial derivative in limit (A.159) can be always intended in the classical sense.

Proof. *Case No 1.* This is simply a particular case, with $\mathbf{A}' = \mathbf{I}$ and $n = 1$, of theorem A.18.1. *Case No 2.* Since, by hypothesis, $u, v \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ and

$$\Delta_2 u = -k^2 u, \quad \Delta_2 v = -k^2 v \quad \text{a.e. in } \mathbb{R}^2 \setminus \bar{D}, \tag{A.161}$$

it easily follows that $\Delta_2 u, \Delta_2 v \in L^2(D)$ and, consequently, $u, v \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D}, \Delta_2)$, where the latter space is defined in remark A.17.1, relation (A.143), putting $\mathbf{A}' = \mathbf{I}$.

If we now take $\Omega_r := \{x \in \mathbb{R}^2 \mid \|x\|_{\mathbb{R}^2} < r\}$ large enough, so that $\Omega_r \supset \bar{D}$, and we put $\Omega := \Omega_r \setminus \bar{D}$, from the definition itself of the spaces $H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D}, \Delta_2)$, $H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ and $L^2_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ it follows that $u|_\Omega, v|_\Omega \in H^1(\Omega, \Delta_2)$. Hence we can use identity (A.158) for the functions u, v and the domain Ω; observing that $\partial\Omega = \partial\Omega_r \cup \partial D$, we get:

$$\int_\Omega (u\Delta_2 v - v\Delta_2 u)\, dx = \int_{\partial\Omega_r} \left(u\frac{\partial v}{\partial \nu_r} - v\frac{\partial u}{\partial \nu_r}\right) d\sigma - \int_{\partial D} \left(u\frac{\partial v}{\partial \nu} - v\frac{\partial u}{\partial \nu}\right) d\sigma, \tag{A.162}$$

where we have denoted with ν_r the unit outward normal to $\partial\Omega_r$, and with ν the unit *outward*[11] normal to ∂D. Since, by virtue of relations (A.161), the function to be integrated at the left-hand side of (A.162) is zero almost everywhere in Ω, we can rewrite relation (A.162) itself as

$$\int_{\partial\Omega_r} \left(u\frac{\partial v}{\partial \nu_r} - v\frac{\partial u}{\partial \nu_r}\right) d\sigma = \int_{\partial D} \left(u\frac{\partial v}{\partial \nu} - v\frac{\partial u}{\partial \nu}\right) d\sigma. \tag{A.163}$$

The next step is now to prove that, by virtue of the Sommerfeld radiation condition (A.159), it holds:

$$\lim_{r\to\infty} \left| \int_{\partial\Omega_r} \left(u\frac{\partial v}{\partial \nu_r} - v\frac{\partial u}{\partial \nu_r}\right) d\sigma \right| = 0. \tag{A.164}$$

To this end, we firstly observe that it trivially holds $\dfrac{\partial v}{\partial \nu_r} = \dfrac{\partial v}{\partial r}\Big|_{\partial\Omega_R}$, $\dfrac{\partial u}{\partial \nu_r} = \dfrac{\partial u}{\partial r}\Big|_{\partial\Omega_R}$; then, we can rewrite the left-hand member of equality (A.163) as:

$$\begin{aligned}\int_{\partial\Omega_r} \left(u\frac{\partial v}{\partial r} - v\frac{\partial u}{\partial r}\right) d\sigma &= \int_{\partial\Omega_r} \left(u\frac{\partial v}{\partial r} - \mathrm{i}kuv + \mathrm{i}kuv - v\frac{\partial u}{\partial r}\right) d\sigma = \\ &= \int_{\partial\Omega_r} u\left(\frac{\partial v}{\partial r} - \mathrm{i}kv\right) d\sigma - \int_{\partial\Omega_r} v\left(\frac{\partial u}{\partial r} - \mathrm{i}ku\right) d\sigma.\end{aligned} \tag{A.165}$$

Now, it is possible to show, in general, that if $w \in H^1_{\partial D,loc}(\mathbb{R}^2 \setminus \bar{D})$ is a weak solution of the Helmholtz equation and satisfies the Sommerfeld radiation condition, then it is analytic in $\mathbb{R}^2 \setminus \bar{D}$ (see remark 2.2.1); since $\Omega_r \supset \bar{D}$, this implies, in particular, that[12] $u|_{\partial\Omega_r}, v|_{\partial\Omega_r} \in L^2(\partial\Omega_r)$ and

[11] This explains the minus sign before the second integral at right-hand side of relation (A.162).

[12] In the following calculations, it is convenient to regard $L^2(\partial\Omega_r)$ as identifiable with $L^2[0, 2\pi]$, so that $\int_{\partial\Omega_r} u(x)d\sigma(x) = \int_0^{2\pi} u(r\cos\theta, r\sin\theta)rd\theta$.

$\left.\frac{\partial v}{\partial r}\right|_{\partial\Omega_r}, \left.\frac{\partial u}{\partial r}\right|_{\partial\Omega_r} \in L^2(\partial\Omega_r)$. Hence, it is possible to regard each of the two integrals appearing in (A.165) as a scalar product in $L^2(\partial\Omega_r)$ and to apply the Cauchy-Schwarz inequality; developing our calculations for the first integral (the second one is treated analogously), we get:

$$\left| \int_{\partial\Omega_r} u \left(\frac{\partial v}{\partial r} - \mathrm{i}kv \right) d\sigma \right| = \left| \left(\frac{\partial v}{\partial r} - \mathrm{i}kv, \bar{u} \right)_{L^2(\partial\Omega_r)} \right| \leq \|u\|_{L^2(\partial\Omega_r)} \left\| \frac{\partial v}{\partial r} - \mathrm{i}kv \right\|_{L^2(\partial\Omega_r)}. \quad \text{(A.166)}$$

Now, on the one hand it is possible to show[13] (see theorem 2.4 in [27]) that

$$\|u\|_{L^2(\partial\Omega_r)} = O(1) \qquad \text{as } r \to \infty; \quad \text{(A.167)}$$

on the other hand, one can easily write:

$$\begin{aligned} \left\| \frac{\partial v}{\partial r} - \mathrm{i}kv \right\|^2_{L^2(\partial\Omega_r)} &= \int_{\partial\Omega_r} \left| \frac{\partial v}{\partial r} - \mathrm{i}kv \right|^2 d\sigma \leq \mathrm{mis}(\partial\Omega_r) \sup_{\partial\Omega_r} \left| \frac{\partial v}{\partial r} - \mathrm{i}kv \right|^2 = \\ &= 2\pi r \sup_{\partial\Omega_r} \left| \frac{\partial v}{\partial r} - \mathrm{i}kv \right|^2 = 2\pi \sup_{\partial\Omega_r} \left[r \left| \frac{\partial v}{\partial r} - \mathrm{i}kv \right|^2 \right] \to 0 \quad \text{as } r \to \infty, \end{aligned} \quad \text{(A.168)}$$

where the last passage is due to condition (A.159). Hence, recalling relations (A.165), (A.166), (A.167) and (A.168), we just get limit (A.164).

Finally, we can observe that in identity (A.163) only the left-hand side depends, a priori, on r, while the right-hand side does not; since identity (A.163) has to hold for any r such that $\Omega_r \supset \bar{D}$, this means that the integral at left-hand side is actually constant with respect to r. By virtue of relation (A.164), we can conclude that the value of this integral is zero for any r such that $\Omega_r \supset \bar{D}$; hence we can rewrite identity (A.163) just as thesis (A.160). ■

A.19. Transpose operators

References: [15], [50].

Remembering definition A.1.3, we shall write $\langle f, u \rangle$ to the denote the value $f(u)$ of the functional $f \in X^*$ at the vector $u \in X$. By virtue of the definition (A.9) of the norm in $\mathcal{B}(X, \mathbb{C})$, it holds:

$$|\langle f, u \rangle| \leq \|f\|_{X^*} \|u\|_X. \quad \text{(A.169)}$$

[13] The proof of this fact is quite involved, so, for sake of brevity, we do not give it here.

By definition of X^*, if $f \in X^*$, then the linear functional $\langle f, \cdot \rangle : X \to \mathbb{C}$ is continuous, i.e.

$$\lim_{n\to\infty} \|u_n - u\|_X = 0 \quad \Rightarrow \quad \lim_{n\to\infty} |\langle f, u_n\rangle - \langle f, u\rangle| = 0; \tag{A.170}$$

on the other hand, it is easy to realize that, for each $u \in X$, also the linear functional $\langle \cdot, u\rangle : X^* \to \mathbb{C}$ is continuous, i.e.

$$\lim_{n\to\infty} \|f_n - f\|_{X^*} = 0 \quad \Rightarrow \quad \lim_{n\to\infty} |\langle f_n, u\rangle - \langle f, u\rangle| = 0. \tag{A.171}$$

Indeed, by virtue of the linearity of the functionals and remembering inequality (A.169), we get:

$$|\langle f_n, u\rangle - \langle f, u\rangle| = |\langle f_n - f, u\rangle| \leq \|f_n - f\|_{X^*}\|u\|_X \to 0 \quad \text{as } n \to \infty. \tag{A.172}$$

Lemma A.19.1. *If $0 \neq u \in X$, then there exists a functional $g \in X^*$ such that*

$$\langle g, u\rangle = \|u\|_X \quad \text{and} \quad \|g\|_{X^*} = 1. \tag{A.173}$$

Proof. See corollary 2.6 in [50]. ■

Definition A.19.1. *Let X, Y be two normed spaces and let X^*, Y^* be their dual spaces. For any linear operator $A : X \to Y$, the* transpose *$A^T : Y^* \to X^*$ is the linear operator defined by*

$$\langle A^T v, u\rangle_{X^*,X} = \langle v, Au\rangle_{Y^*,Y} \qquad \forall u \in X, \quad \forall v \in Y^*, \tag{A.174}$$

where $\langle \cdot, \cdot \rangle$ denotes the duality pairing between the spaces denoted at subscript[14].

Theorem A.19.2. *The transpose operator A^T is bounded if and only if the operator A is bounded; moreover, it holds:*

$$\|A^T\|_{\mathcal{B}(Y^*,X^*)} = \|A\|_{\mathcal{B}(X,Y)}. \tag{A.175}$$

Proof. "⇐": if A is bounded, then the definition (A.174) of A^T and inequality (A.169) imply:

$$\left|\left\langle A^T v, u\right\rangle_{X^*,X}\right| = \left|\langle v, Au\rangle_{Y^*,Y}\right| \leq \|v\|_{Y^*}\|Au\|_Y \leq \|v\|_{Y^*}\|A\|_{\mathcal{B}(X,Y)}\|u\|_X \quad \forall u \in X, \quad \forall v \in Y^*. \tag{A.176}$$

From inequalities (A.176) we immediately get:

$$\frac{\left|\left\langle A^T v, u\right\rangle_{X^*,X}\right|}{\|u\|_X} \leq \|v\|_{Y^*}\|A\|_{\mathcal{B}(X,Y)} \quad \forall\, 0 \neq u \in X, \quad \forall v \in Y^*, \tag{A.177}$$

and then, recalling (A.9),

$$\|A^T v\|_{X^*} = \sup_{\|u\|_X \neq 0} \frac{\left|\left\langle A^T v, u\right\rangle_{X^*,X}\right|}{\|u\|_X} \leq \|A\|_{\mathcal{B}(X,Y)}\|v\|_{Y^*} \quad \forall v \in Y^*. \tag{A.178}$$

[14] Subscripts will often be omitted, with no risk of ambiguity.

Relation (A.178) clearly implies the boundedness of A^T and

$$\|A^T\|_{\mathcal{B}(Y^*,X^*)} \leq \|A\|_{\mathcal{B}(X,Y)}. \tag{A.179}$$

"⇒": conversely, let us suppose that A^T is bounded, and let $u \in X$. If $Au \neq 0$, then, by virtue of lemma A.19.1, there is a $v \in Y^*$ such that

$$\langle v, Au\rangle_{Y^*,Y} = \|Au\|_X \quad \text{and} \quad \|v\|_{Y^*} = 1. \tag{A.180}$$

Hence, by means of relations (A.169), (A.174), (A.180) and remembering that A^T is bounded by hypothesis, we can easily write the following chain of equalities or inequalities:

$$\|Au\|_X = \left|\langle A^T v, u\rangle_{X^*,X}\right| \leq \|A^T v\|_{X^*}\|u\|_X \leq \|A^T\|_{\mathcal{B}(Y^*,X^*)}\|v\|_{Y^*}\|u\|_X = \|A^T\|_{\mathcal{B}(Y^*,X^*)}\|u\|_X. \tag{A.181}$$

Summing up, for all $u \in X$ such that $Au \neq 0$, we have found the inequality

$$\|Au\|_X \leq \|A^T\|_{\mathcal{B}(Y^*,X^*)}\|u\|_X, \tag{A.182}$$

which is clearly trivial for $Au = 0$: hence relation (A.182) holds for all $u \in X$, then it implies the boundedness of A and:

$$\|A\|_{\mathcal{B}(X,Y)} \leq \|A^T\|_{\mathcal{B}(Y^*,X^*)}. \tag{A.183}$$

Finally, relations (A.179) and (A.183) together imply equality (A.175). ■

To describe the relation between the range and the kernel of A and A^T we use the following terminology.

Definition A.19.2. *For any subset $W \subset X$, we define its* annihilator W^a *as the subset of X^* given by*

$$W^a := \{g \in X^* \mid \langle g, u\rangle = 0 \ \ \forall u \in W\}; \tag{A.184}$$

similarly, for any subset $V \subset X^*$, we define its annihilator ${}^aV$ *as the subset of X given by*

$${}^aV := \{u \in X \mid \langle g, u\rangle = 0 \ \ \forall g \in V\}. \tag{A.185}$$

Theorem A.19.3. *The subsets W^a and ${}^aV$ previously defined are closed subspaces of X^* and X respectively.*

Proof. The linearity of the functionals $\langle \cdot, u\rangle : X^* \to \mathbb{C}$ and $\langle g, \cdot\rangle : X \to \mathbb{C}$ trivially implies that W^a and ${}^aV$ are subspaces of X^* and X respectively.

As regards their closedness, we prove it for W^a (the proof for ${}^aV$ is completely analogous). Let $g \in \overline{W^a} \subset X^*$: we are going to show that $g \in W^a$. To this end, let us consider a sequence

$\{g_n\}_{n=0}^{\infty} \subset W^a$ such that $\lim_{n\to\infty} \|g_n - g\|_{X^*} = 0$. Then, by using relation (A.171) and observing that $\langle g_n, u\rangle = 0\ \forall n \in \mathbb{N}$ and $\forall u \in W$, we easily get:

$$\langle g, u\rangle = \langle \lim_{n\to\infty} g_n, u\rangle = \lim_{n\to\infty} \langle g_n, u\rangle = 0 \qquad \forall u \in W, \tag{A.186}$$

i.e. $g \in W^a$. ■

Theorem A.19.4. *The kernels and the ranges of A and A^T satisfy*

$$\mathcal{N}(A^T) = [\mathcal{R}(A)]^a \quad \textit{and} \quad \mathcal{N}(A) = {}^a[\mathcal{R}(A^T)]. \tag{A.187}$$

Proof. Applying definitions (A.184) and (A.174), we obtain:

$$\begin{aligned}
[\mathcal{R}(A)]^a &= \{g \in Y^* \mid \langle g, v\rangle_{Y^*,Y} = 0 \ \ \forall v \in \mathcal{R}(A)\} = \\
&= \{g \in Y^* \mid \langle g, Au\rangle_{Y^*,Y} = 0 \ \ \forall u \in X\} = \\
&= \left\{g \in Y^* \mid \langle A^T g, u\rangle_{X^*,X} = 0 \ \ \forall u \in X\right\} = \\
&= \left\{g \in Y^* \mid A^T g = 0\right\} = \mathcal{N}(A^T),
\end{aligned} \tag{A.188}$$

that is the first of equalities (A.187); the second one is proved by means of an analogous argument.

■

Theorem A.19.5. *Let X be a Hilbert space; then a subspace $W \subset X$ is dense in X if and only if $W^a = \{0\}$.*

Proof. "⇒": if $\overline{W} = X$, for any $v \in X$ we can find a sequence $\{v_n\}_{n=0}^{\infty} \subset W$ such that $\lim_{n\to\infty} \|v_n - v\|_X = 0$. Then, if $g \in W^a$, by using relation (A.170) and observing that $\langle g, v_n\rangle = 0\ \forall n \in \mathbb{N}$, we easily get:

$$\langle g, v\rangle = \langle g, \lim_{n\to\infty} v_n\rangle = \lim_{n\to\infty} \langle g, v_n\rangle = 0. \tag{A.189}$$

Since the previous argument can be used for all $v \in X$, we conclude:

$$\langle g, v\rangle = 0 \qquad \forall v \in X, \tag{A.190}$$

i.e. $g = 0$ and hence $W^a = \{0\}$. We note that in this part of the proof neither the structure of Hilbert space of X nor the fact that W is a subspace are involved.
"⇐": let us suppose that $W^a = \{0\}$. This means that, if $g \in X^*$,

$$\langle g, u\rangle = 0 \quad \forall u \in W \qquad \Rightarrow \qquad g = 0. \tag{A.191}$$

Using Riesz representation theorem for linear bounded functionals acting on a Hilbert space X, we can reformulate the previous implication (A.191) as

$$(f_g, u)_X = 0 \quad \forall u \in W \qquad \Rightarrow \qquad f_g = 0, \tag{A.192}$$

where we have denoted with $(\cdot,\cdot)_X$ the scalar product in X. This means that $W^\perp = \{0\}$; on the other hand, since W is a subspace, it holds $X = \overline{W} \oplus W^\perp$: but $W^\perp = \{0\}$, hence $X = \overline{W}$. ■

Since $\mathcal{R}(A)$ is a subspace, the first of relations (A.187) and theorem A.19.5 together imply the following proposition.

Theorem A.19.6. *Let X, Y two Hilbert spaces; the linear bounded operator $A : X \to Y$ has dense range if and only if the transpose $A^T : Y^* \to X^*$ is injective.*

APPENDIX B

Figures

As already pointed out at the end of the preface, in order to make the written text more easily readable we have collected all the figures of this PhD thesis in a special appendix, i.e. the current one. They have been subdivided into sections having the same title (and, between square brackets, the same number) of the correspondent sections in chapters 2 and 3 (chapter 1 has no figures).

B.1. [2.5] The linear sampling method

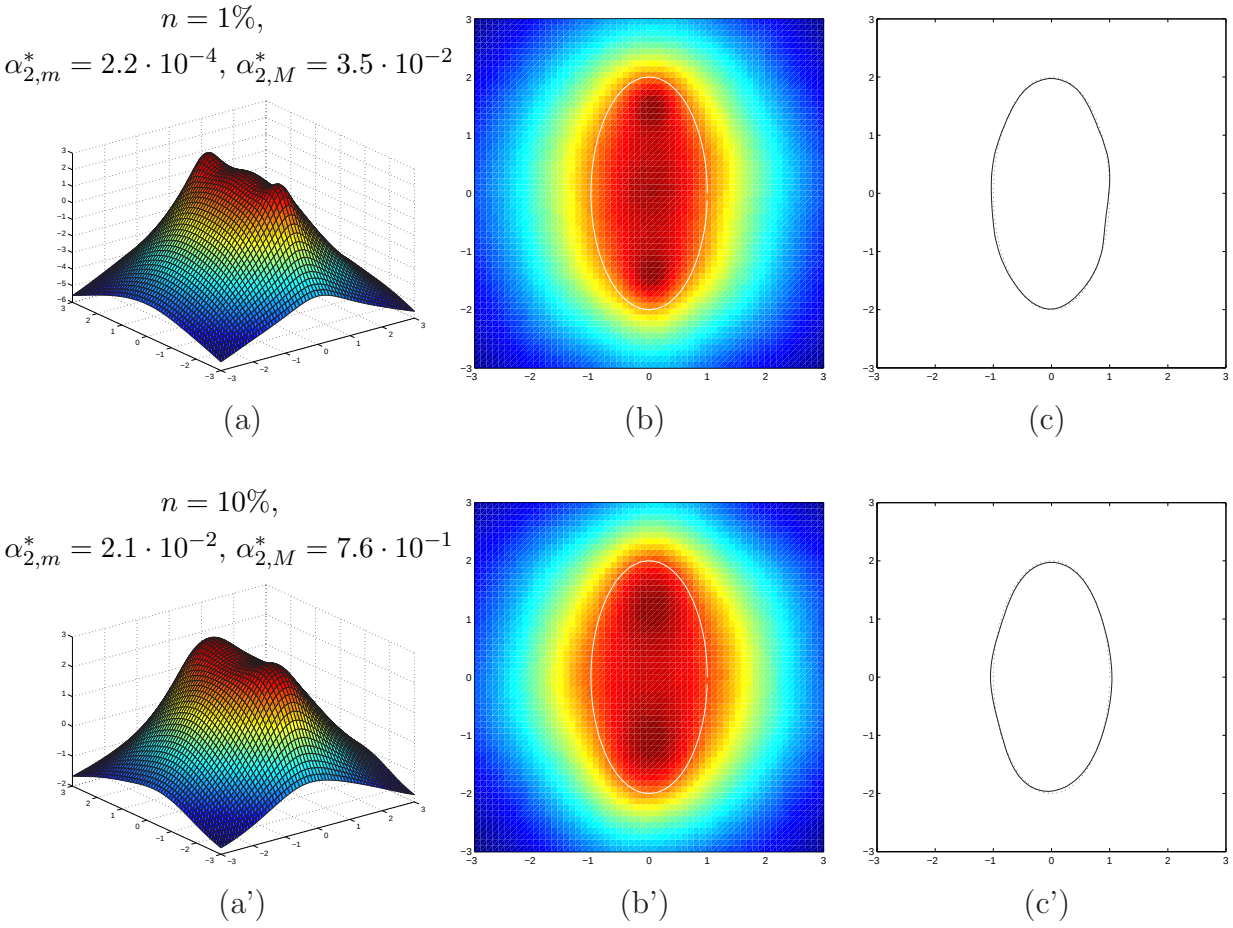

Figure B.1: Reconstruction of the profile of a conducting ellipse-shaped scatterer (in the case of Dirichlet boundary conditions) by means of the traditional implementation of the linear sampling method. The wavenumber is $k = 1$, the number of incidence/observation angles is $N = 32$, the sampling grid $\mathcal{Z}$ is 60×60 points $z_l \in \mathcal{Z}$ uniformly chosen in the square $[-3, 3] \times [-3, 3]$ and a level $n = 1\%$ (for panels (a), (b), (c)) or $n = 10\%$ (for panels (a'), (b'), (c')) of Gaussian noise is respectively added to the exact far-field matrix, which is computed by using the Nyström method (cf. remark 2.5.1). Panel (a) [(a')] shows the three-dimensional plot of the indicator function $\Psi_{-2\ln}(z_l) := -\ln \left\| \mathbf{g}_{\alpha^*_2(z_l)}(z_l) \right\|^2_{\mathbb{C}^N}$ (cf. definition (2.284)), where, for each z_l, the optimal value $\alpha^*_2(z_l)$ of the regularization parameter is fixed by using the generalized discrepancy principle in the compatible case (cf. subsection 1.8.3 and, in particular, definition (1.317), as well as its specific form (2.279) for the current context); over the three-dimensional plot itself, we have written the corresponding noise level n and the minimum and maximum values of the regularization parameter, according to the definitions $\alpha^*_{2,m} := \min_{z_l \in \mathcal{Z}} \alpha^*_2(z_l)$ and $\alpha^*_{2,M} := \max_{z_l \in \mathcal{Z}} \alpha^*_2(z_l)$. Panel (b) [(b')] shows a two-dimensional projection (i.e., roughly speaking, a view from above) of the plot in panel (a) [(a')] together with the profile (in white colour) of the true scatterer. Panel (c) [(c')] shows the profile of the true scatterer (dotted line) and, superimposed, the reconstructed profile (solid line) obtained as the level curve of the plot in panel (a) [(a')] containing an area equal to the one contained by the true profile.

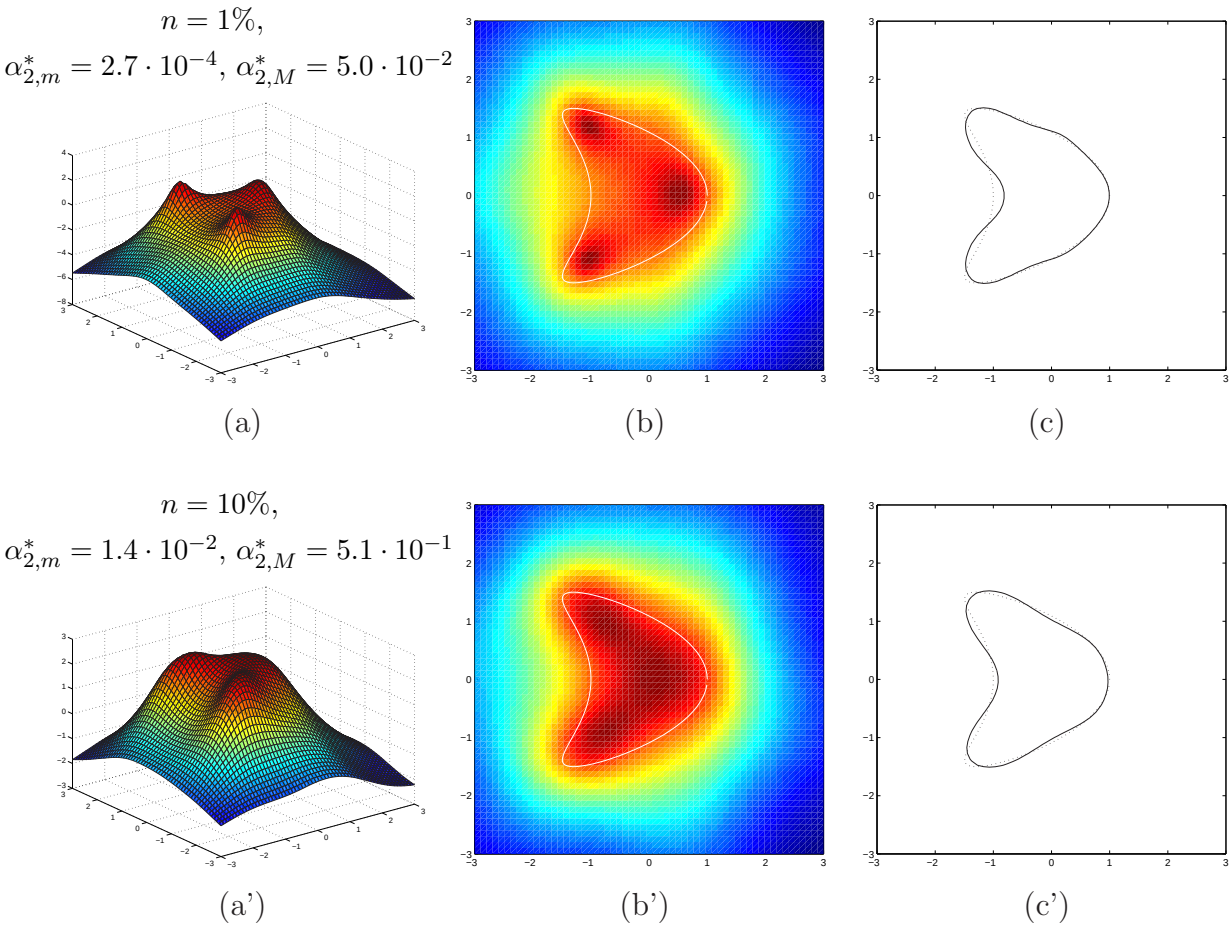

Figure B.2: Reconstruction of the profile of a conducting kite-shaped scatterer (in the case of Dirichlet boundary conditions) by means of the traditional implementation of the linear sampling method. The wavenumber is $k = 1$, the number of incidence/observation angles is $N = 32$, the sampling grid $\mathcal{Z}$ is 60×60 points $z_l \in \mathcal{Z}$ uniformly chosen in the square $[-3, 3] \times [-3, 3]$ and 1% (for panels (a), (b), (c)) or 10% (for panels (a'), (b'), (c')) of Gaussian noise is respectively added to the exact far-field matrix, which is computed by using the Nyström method (cf. remark 2.5.1). Panel (a) [(a')] shows the three-dimensional plot of the indicator function $\Psi_{-2\ln}(z_l) := -\ln \left\| \mathbf{g}_{\alpha^*_2(z_l)}(z_l) \right\|^2_{\mathbb{C}^N}$ (cf. definition (2.284)), where, for each z_l, the optimal value $\alpha^*_2(z_l)$ of the regularization parameter is fixed by using the generalized discrepancy principle in the compatible case (cf. subsection 1.8.3 and, in particular, definition (1.317), as well as its specific form (2.279) for the current context); over the three-dimensional plot itself, we have written the corresponding noise level n and the minimum and maximum values of the regularization parameter, according to the definitions $\alpha^*_{2,m} := \min_{z_l \in \mathcal{Z}} \alpha^*_2(z_l)$ and $\alpha^*_{2,M} := \max_{z_l \in \mathcal{Z}} \alpha^*_2(z_l)$. Panel (b) [(b')] shows a two-dimensional projection (i.e., roughly speaking, a view from above) of the plot in panel (a) [(a')] together with the profile (in white colour) of the true scatterer. Panel (c) [(c')] shows the profile of the true scatterer (dotted line) and, superimposed, the reconstructed profile (solid line) obtained as the level curve of the plot in panel (a) [(a')] containing an area equal to the one contained by the true profile.

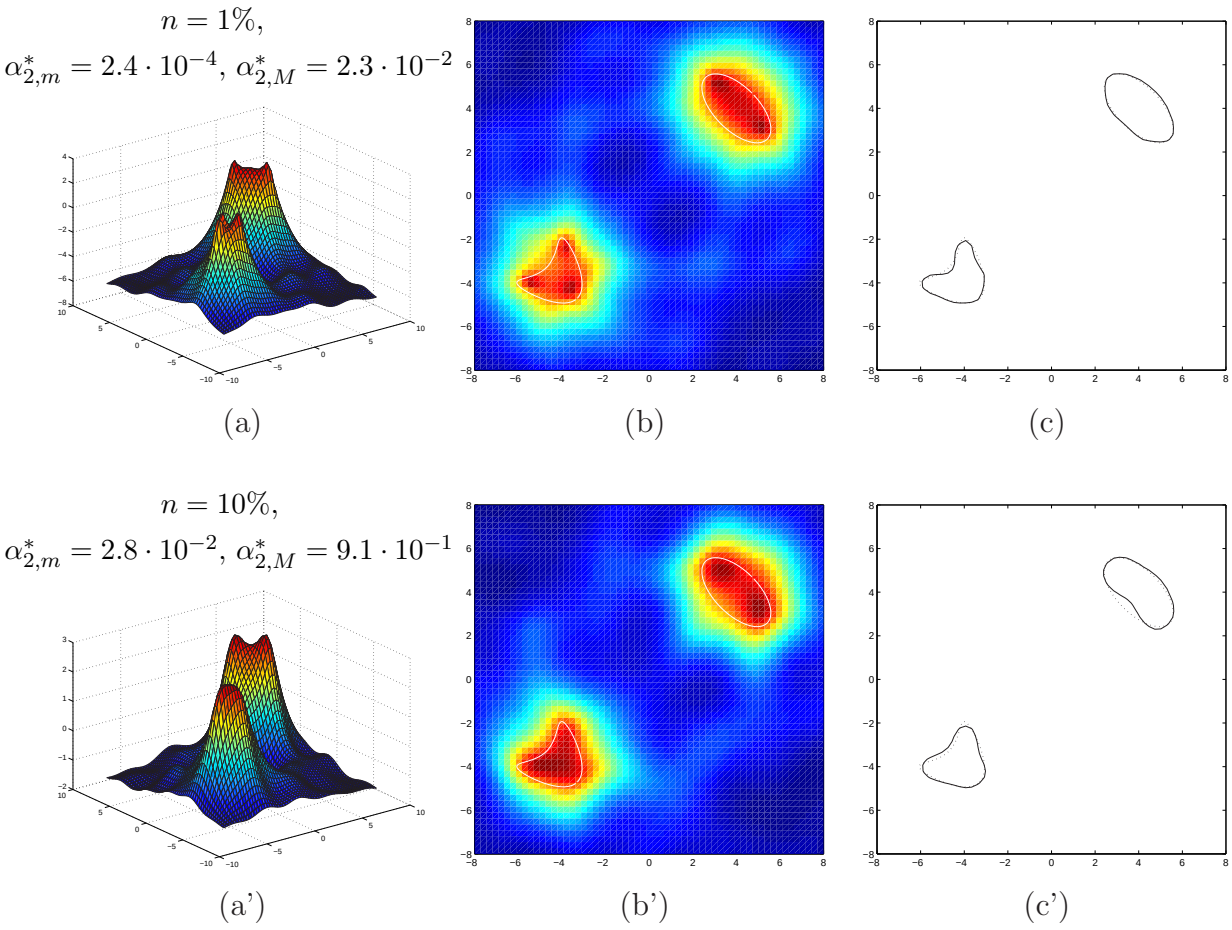

Figure B.3: Reconstruction of the profile of a conducting double scatterer (in the case of Dirichlet boundary conditions) by means of the traditional implementation of the linear sampling method: the scatterer consists of a kite and an ellipse. The wavenumber is $k = 1$, the number of incidence/observation angles is $N = 32$, the sampling grid $\mathcal{Z}$ is 60×60 points $z_l \in \mathcal{Z}$ uniformly chosen in the square $[-8, 8] \times [-8, 8]$ and 1% (for panels (a), (b), (c)) or 10% (for panels (a'), (b'), (c')) of Gaussian noise is respectively added to the exact far-field matrix, which is computed by using the Nyström method (cf. remark 2.5.1). Panel (a) [(a')] shows the three-dimensional plot of the indicator function $\Psi_{-2\ln}(z_l) := -\ln \left\| \mathbf{g}_{\alpha^*_2(z_l)}(z_l) \right\|^2_{\mathbb{C}^N}$ (cf. definition (2.284)), where, for each z_l, the optimal value $\alpha^*_2(z_l)$ of the regularization parameter is fixed by using the generalized discrepancy principle in the compatible case (cf. subsection 1.8.3 and, in particular, definition (1.317), as well as its specific form (2.279) for the current context); over the three-dimensional plot itself, we have written the corresponding noise level n and the minimum and maximum values of the regularization parameter, according to the definitions $\alpha^*_{2,m} := \min_{z_l \in \mathcal{Z}} \alpha^*_2(z_l)$ and $\alpha^*_{2,M} := \max_{z_l \in \mathcal{Z}} \alpha^*_2(z_l)$. Panel (b) [(b')] shows a two-dimensional projection (i.e., roughly speaking, a view from above) of the plot in panel (a) [(a')] together with the profile (in white colour) of the true scatterer. Panel (c) [(c')] shows the profile of the true scatterer (dotted line) and, superimposed, the reconstructed profile (solid line) obtained as the level curve of the plot in panel (a) [(a')] containing an area equal to the one contained by the true profile.

B.2. [3.1] A new implementation of the linear sampling method

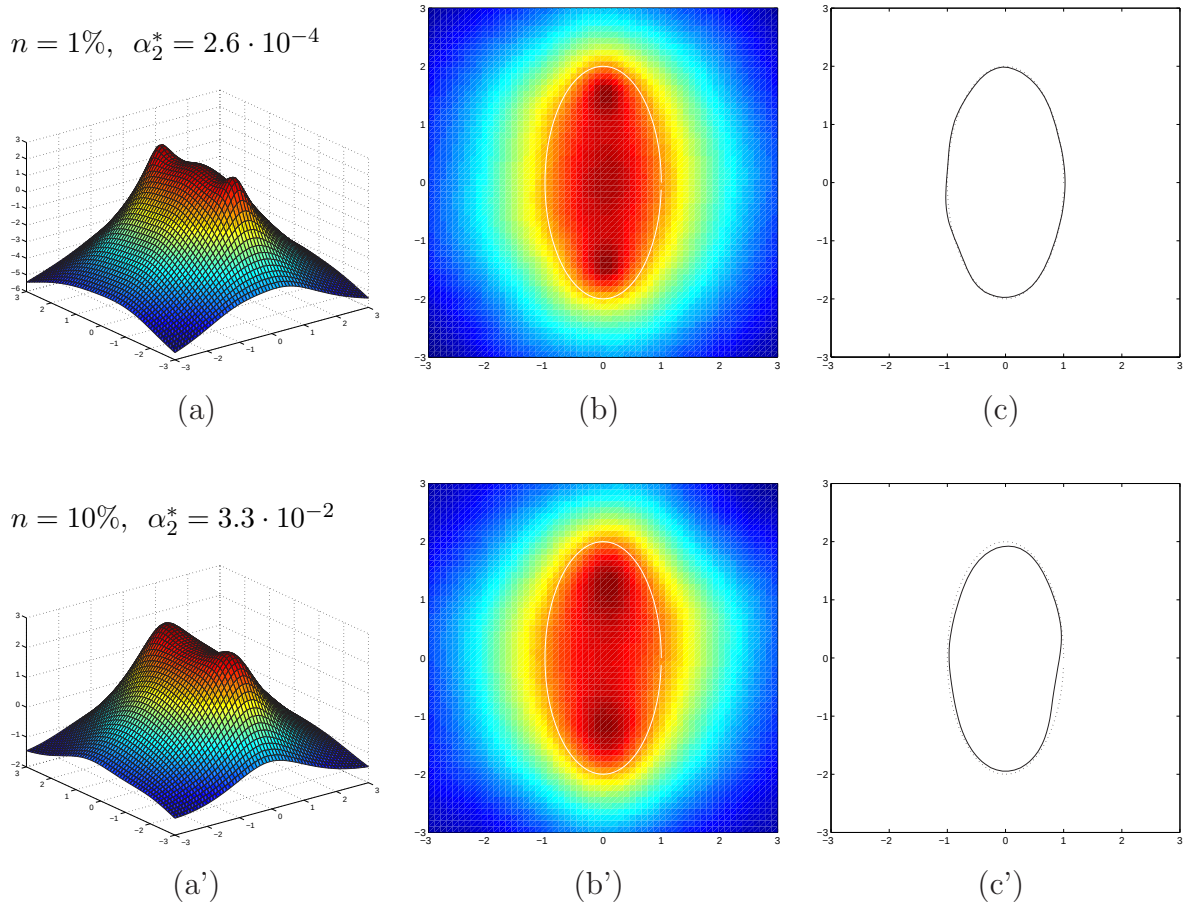

Figure B.4: Reconstruction of the profile of a conducting ellipse-shaped scatterer (in the case of Dirichlet boundary conditions) by means of the new (i.e. no-sampling) implementation of the linear sampling method. The wavenumber is $k = 1$, the number of incidence/observation angles is $N = 32$ and 1% (for panels (a), (b), (c)) or 10% (for panels (a'), (b'), (c')) of Gaussian noise is respectively added to the exact far-field matrix, which is computed by using the Nyström method (cf. remark 2.5.1). Panel (a) [(a')] shows the three-dimensional plot of the indicator function $\Psi_{-2\ln}(z) := -\ln \left\| \mathbf{g}_{\alpha_2^*}(z) \right\|^2_{\mathbb{C}^N}$ considered in the open square $T_A^B := (-3,3) \times (-3,3)$ (cf. definition (3.72)), where the (unique) optimal value α_2^* of the regularization parameter is fixed by using the generalized discrepancy principle in the compatible case (cf. subsection 1.8.3 and, in particular, definition (1.317), as well as its specific form (3.62) for the current context); over the three-dimensional plot itself, we have written the corresponding noise level n and the (unique) value α_2^* of the regularization parameter. Panel (b) [(b')] shows a two-dimensional projection (i.e., roughly speaking, a view from above) of the plot in panel (a) [(a')] together with the profile (in white colour) of the true scatterer. Panel (c) [(c')] shows the profile of the true scatterer (dotted line) and, superimposed, the reconstructed profile (solid line) obtained as the level curve of the plot in panel (a) [(a')] containing an area equal to the one contained by the true profile.

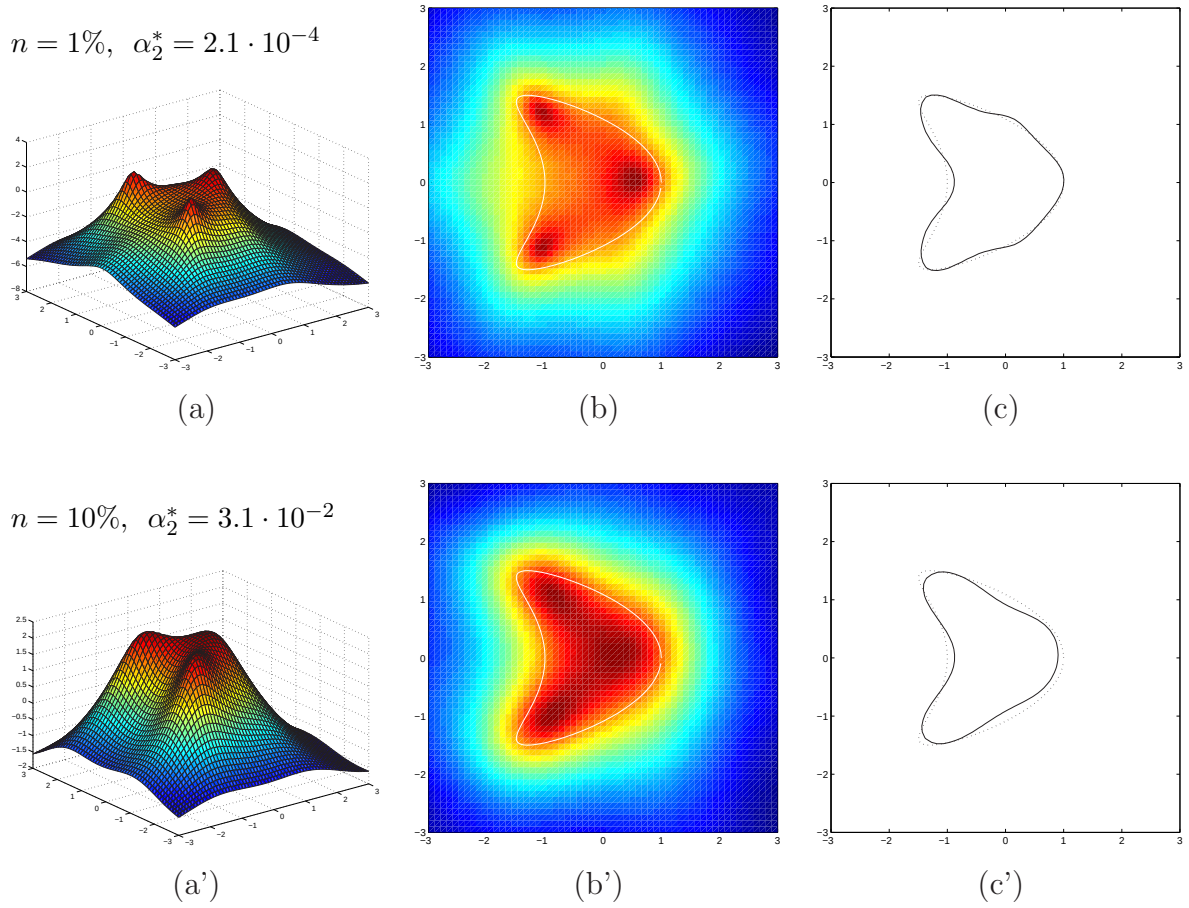

Figure B.5: Reconstruction of the profile of a conducting kite-shaped scatterer (in the case of Dirichlet boundary conditions) by means of the new (i.e. no-sampling) implementation of the linear sampling method. The wavenumber is $k = 1$, the number of incidence/observation angles is $N = 32$ and 1% (for panels (a), (b), (c)) or 10% (for panels (a'), (b'), (c')) of Gaussian noise is respectively added to the exact far-field matrix, which is computed by using the Nyström method (cf. remark 2.5.1). Panel (a) [(a')] shows the three-dimensional plot of the indicator function $\Psi_{-2\ln}(z) := -\ln \left\| \mathbf{g}_{\alpha_2^*}(z) \right\|^2_{\mathbb{C}^N}$ considered in the open square $T_A^B := (-3,3) \times (-3,3)$ (cf. definition (3.72)), where the (unique) optimal value α_2^* of the regularization parameter is fixed by using the generalized discrepancy principle in the compatible case (cf. subsection 1.8.3 and, in particular, definition (1.317), as well as its specific form (3.62) for the current context); over the three-dimensional plot itself, we have written the corresponding noise level n and the (unique) value α_2^* of the regularization parameter. Panel (b) [(b')] shows a two-dimensional projection (i.e., roughly speaking, a view from above) of the plot in panel (a) [(a')] together with the profile (in white colour) of the true scatterer. Panel (c) [(c')] shows the profile of the true scatterer (dotted line) and, superimposed, the reconstructed profile (solid line) obtained as the level curve of the plot in panel (a) [(a')] containing an area equal to the one contained by the true profile.

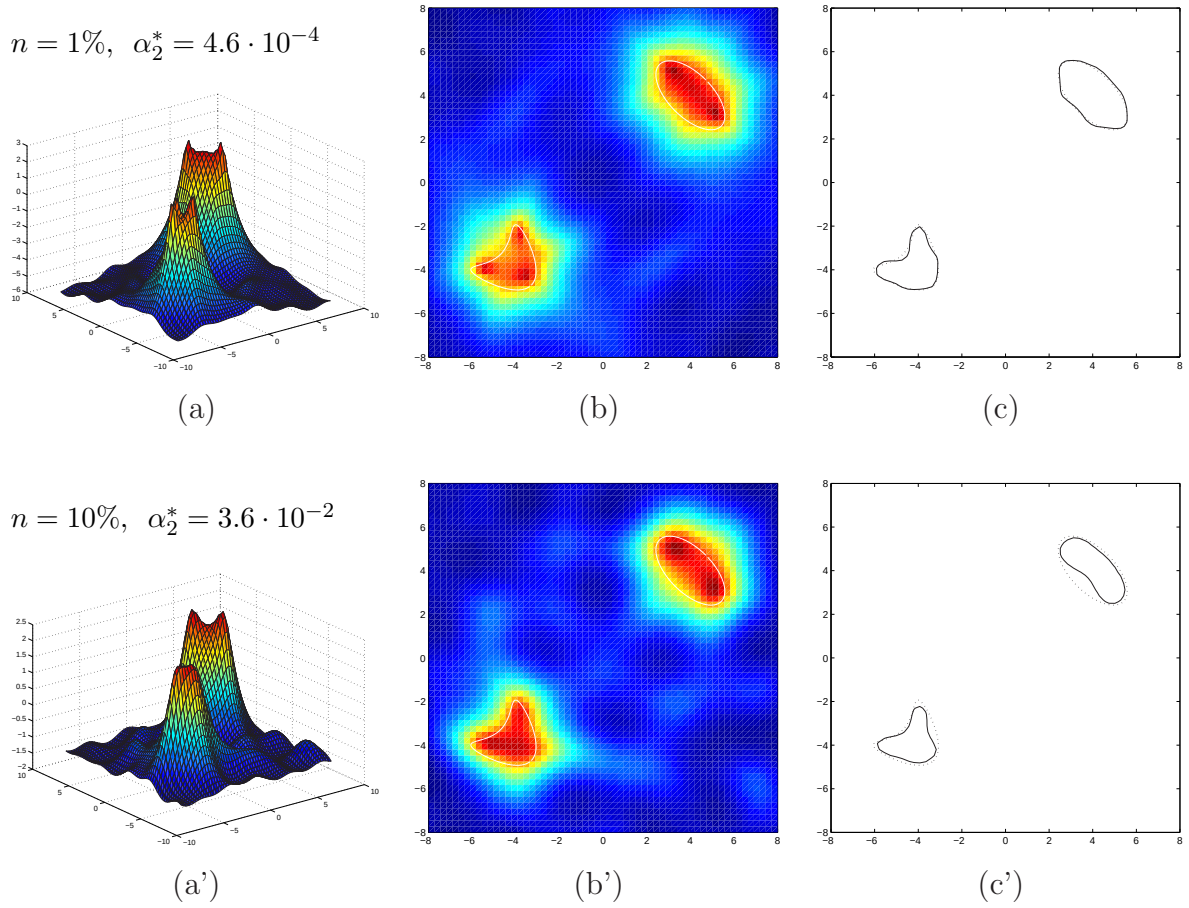

Figure B.6: Reconstruction of the profile of a conducting double scatterer (in the case of Dirichlet boundary conditions) by means of the new (i.e. no-sampling) implementation of the linear sampling method: the scatterer consists of a kite and an ellipse. The wavenumber is $k = 1$, the number of incidence/observation angles is $N = 32$ and 1% (for panels (a), (b), (c)) or 10% (for panels (a'), (b'), (c')) of Gaussian noise is respectively added to the exact far-field matrix, which is computed by using the Nyström method (cf. remark 2.5.1). Panel (a) [(a')] shows the three-dimensional plot of the indicator function $\Psi_{-2\ln}(z) := -\ln \left\| \mathbf{g}_{\alpha_2^*}(z) \right\|^2_{\mathbb{C}^N}$ considered in the open square $T_A^B := (-8, 8) \times (-8, 8)$ (cf. definition (3.72)), where the (unique) optimal value α_2^* of the regularization parameter is fixed by using the generalized discrepancy principle in the compatible case (cf. subsection 1.8.3 and, in particular, definition (1.317), as well as its specific form (3.62) for the current context); over the three-dimensional plot itself, we have written the corresponding noise level n and the (unique) value α_2^* of the regularization parameter. Panel (b) [(b')] shows a two-dimensional projection (i.e., roughly speaking, a view from above) of the plot in panel (a) [(a')] together with the profile (in white colour) of the true scatterer. Panel (c) [(c')] shows the profile of the true scatterer (dotted line) and, superimposed, the reconstructed profile (solid line) obtained as the level curve of the plot in panel (a) [(a')] containing an area equal to the one contained by the true profile.

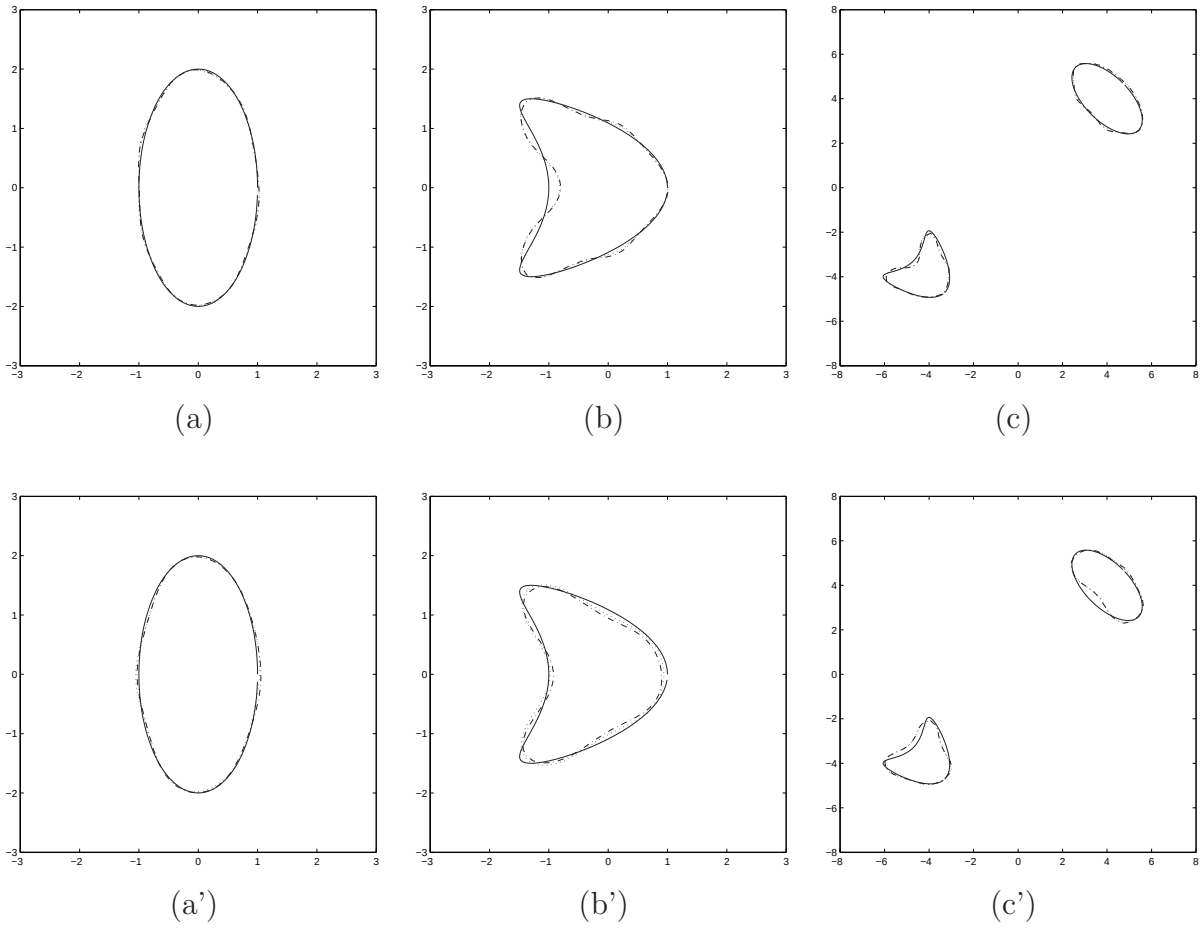

Figure B.7: A comparison between the traditional and the no-sampling implementation of the linear sampling method. This figure is essentially a collection of the panels (c) and (c') of all the previous figures B.1-B.6. For each of the six panels, we provide (solid line) the profile of the true object and, superimposed, the reconstructed contours provided by the two implementations: the traditional one (dashed line) and the no-sampling one (dotted line). This caption is completed by the following table B.1 and by footnote 10 in section 3.1.

scattering object	panel	n	$\alpha^*_{2,m}$	$\alpha^*_{2,M}$	α^*_2
ellipse	(a)	1%	$2.3 \cdot 10^{-4}$	$3.4 \cdot 10^{-2}$	$2.9 \cdot 10^{-4}$
kite	(b)	1%	$1.7 \cdot 10^{-4}$	$1.8 \cdot 10^{-2}$	$2.6 \cdot 10^{-4}$
kite+ellipse	(c)	1%	$2.4 \cdot 10^{-4}$	$2.3 \cdot 10^{-2}$	$3.9 \cdot 10^{-4}$
ellipse	(a')	10%	$2.4 \cdot 10^{-2}$	$7.1 \cdot 10^{-1}$	$3.5 \cdot 10^{-2}$
kite	(b')	10%	$1.9 \cdot 10^{-2}$	$6.5 \cdot 10^{-1}$	$3.2 \cdot 10^{-2}$
kite+ellipse	(c')	10%	$2.5 \cdot 10^{-2}$	$9.5 \cdot 10^{-1}$	$4.1 \cdot 10^{-2}$

Table B.1: This numerical table completes the caption of the previous figure B.7. Obviously, the meaning of the parameters n, $\alpha^*_{2,m}$, $\alpha^*_{2,M}$, α^*_2 is the same as the one in figures B.1-B.6. Cf. also footnote 10 in section 3.1.

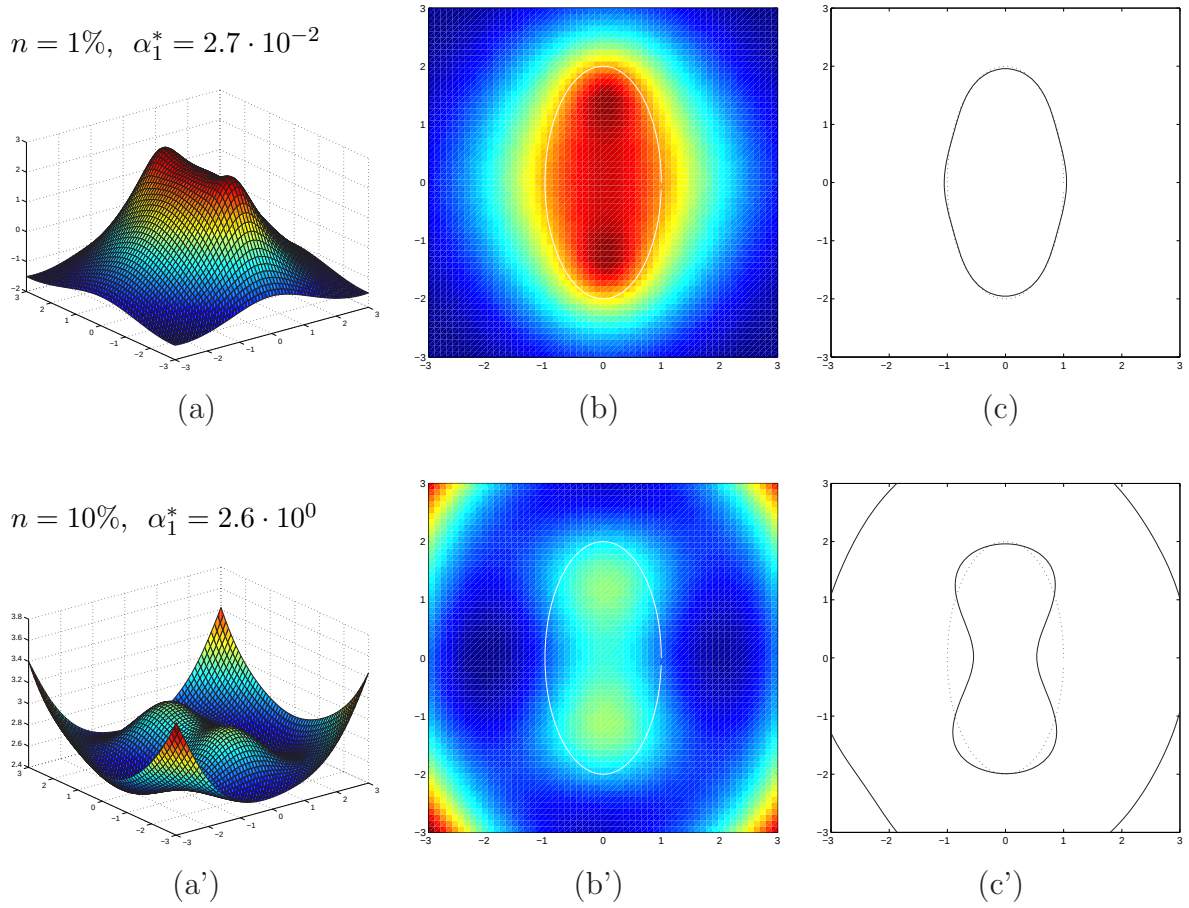

Figure B.8: Reconstruction of the profile of a conducting ellipse-shaped scatterer (in the case of Dirichlet boundary conditions) by means of the new (i.e. no-sampling) implementation of the linear sampling method. The wavenumber is $k = 1$, the number of incidence/observation angles is $N = 32$ and 1% (for panels (a), (b), (c)) or 10% (for panels (a'), (b'), (c')) of Gaussian noise is respectively added to the exact far-field matrix, which is computed by using the Nyström method (cf. remark 2.5.1). Panel (a) [(a')] shows the three-dimensional plot of the indicator function $\Psi_{-2\ln}(z) := -\ln \left\| \mathbf{g}_{\alpha_1^*}(z) \right\|^2_{\mathbb{C}^N}$ considered in the open square $T_A^B := (-3,3) \times (-3,3)$ (cf. definition (3.72)), where the (unique) optimal value α_1^* of the regularization parameter is fixed by using the generalized discrepancy principle in the incompatible case (cf. subsection 1.8.2 and, in particular, definition (1.259), as well as its specific form (3.66) for the current context). Panel (b) [(b')] shows a two-dimensional projection (i.e., roughly speaking, a view from above) of the plot in panel (a) [(a')] together with the profile (in white colour) of the true scatterer. Panel (c) [(c')] shows the profile of the true scatterer (dotted line) and, superimposed, the reconstructed profile (solid line) obtained as the level curve of the plot in panel (a) [(a')] containing an area equal to the one contained by the true profile. Without going into details, we only point out that the computational procedure implemented by our code in order to produce the third kind of panels (i.e. (c) and (c')) by means of the equal-areas criterion is such that when the three-dimensional plot (like the one in panel (a')) is too far from being the expected "hill-shaped" diagram, the corresponding profile (like the one in panel (c')), reconstructed by using such a criterion, is somehow meaningless or, anyway, unacceptable for our purposes.

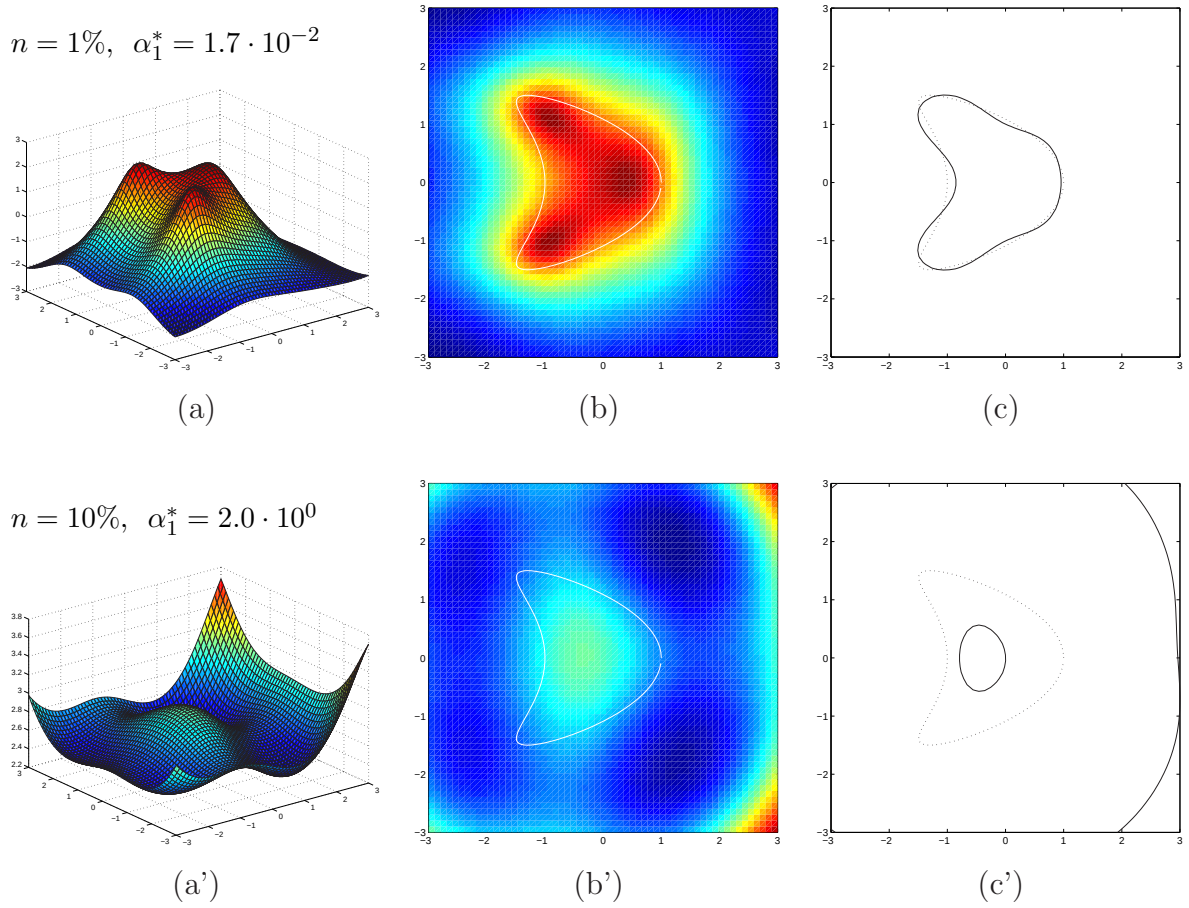

Figure B.9: Reconstruction of the profile of a conducting kite-shaped scatterer (in the case of Dirichlet boundary conditions) by means of the new (i.e. no-sampling) implementation of the linear sampling method. The wavenumber is $k = 1$, the number of incidence/observation angles is $N = 32$ and 1% (for panels (a), (b), (c)) or 10% (for panels (a'), (b'), (c')) of Gaussian noise is respectively added to the exact far-field matrix, which is computed by using the Nyström method (cf. remark 2.5.1). Panel (a) [(a')] shows the three-dimensional plot of the indicator function $\Psi_{-2\ln}(z) := -\ln \left\| \mathbf{g}_{\alpha_1^*}(z) \right\|^2_{\mathbb{C}^N}$ considered in the open square $T_A^B := (-3,3) \times (-3,3)$ (cf. definition (3.72)), where the (unique) optimal value α_1^* of the regularization parameter is fixed by using the generalized discrepancy principle in the incompatible case (cf. subsection 1.8.2 and, in particular, definition (1.259), as well as its specific form (3.66) for the current context). Panel (b) [(b')] shows a two-dimensional projection (i.e., roughly speaking, a view from above) of the plot in panel (a) [(a')] together with the profile (in white colour) of the true scatterer. Panel (c) [(c')] shows the profile of the true scatterer (dotted line) and, superimposed, the reconstructed profile (solid line) obtained as the level curve of the plot in panel (a) [(a')] containing an area equal to the one contained by the true profile. Without going into details, we only point out that the computational procedure implemented by our code in order to produce the third kind of panels (i.e. (c) and (c')) by means of the equal-areas criterion is such that when the three-dimensional plot (like the one in panel (a')) is too far from being the expected "hill-shaped" diagram, the corresponding profile (like the one in panel (c')), reconstructed by using such a criterion, is somehow meaningless or, anyway, unacceptable for our purposes.

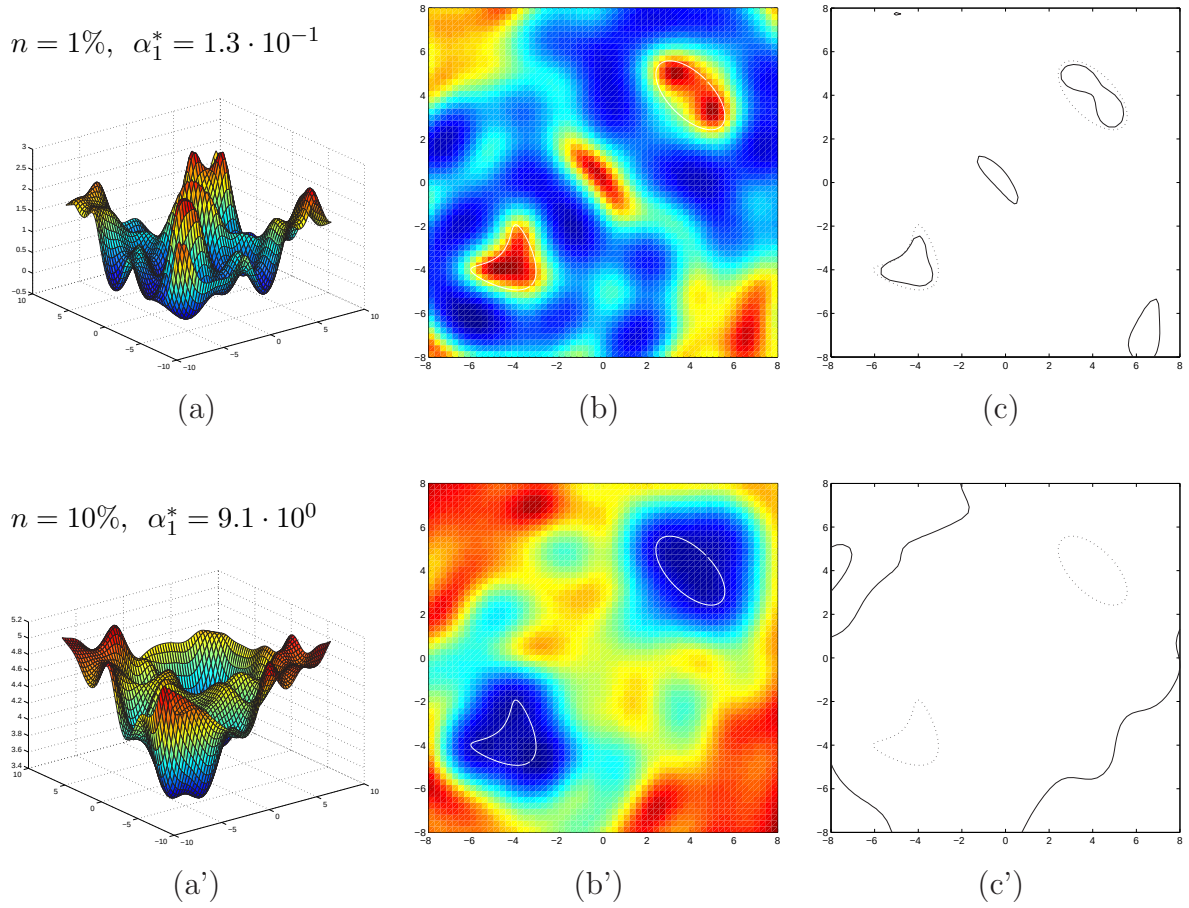

Figure B.10: Reconstruction of the profile of a conducting double scatterer (in the case of Dirichlet boundary conditions) by means of the new (i.e. no-sampling) implementation of the linear sampling method: the scatterer consists of a kite and an ellipse. The wavenumber is $k = 1$, the number of incidence/observation angles is $N = 32$ and 1% (for panels (a), (b), (c)) or 10% (for panels (a'), (b'), (c')) of Gaussian noise is respectively added to the exact far-field matrix, which is computed by using the Nyström method (cf. remark 2.5.1). Panel (a) [(a')] shows the three-dimensional plot of the indicator function $\Psi_{-2\ln}(z) := -\ln \left\| \mathbf{g}_{\alpha_1^*}(z) \right\|^2_{\mathbb{C}^N}$ considered in the open square $T_A^B := (-8, 8) \times (-8, 8)$ (cf. definition (3.72)), where the (unique) optimal value α_1^* of the regularization parameter is fixed by using the generalized discrepancy principle in the incompatible case (cf. subsection 1.8.2 and, in particular, definition (1.259), as well as its specific form (3.66) for the current context). Panel (b) [(b')] shows a two-dimensional projection (i.e., roughly speaking, a view from above) of the plot in panel (a) [(a')] together with the profile (in white colour) of the true scatterer. Panel (c) [(c')] shows the profile of the true scatterer (dotted line) and, superimposed, the reconstructed profile (solid line) obtained as the level curve of the plot in panel (a) [(a')] containing an area equal to the one contained by the true profile. Without going into details, we only point out that the computational procedure implemented by our code in order to produce the third kind of panels (i.e. (c) and (c')) by means of the equal-areas criterion is such that when the three-dimensional plot (like the one in panel (a')) is too far from being the expected "hill-shaped" diagram, the corresponding profile (like the one in panel (c')), reconstructed by using such a criterion, is somehow meaningless or, anyway, unacceptable for our purposes.

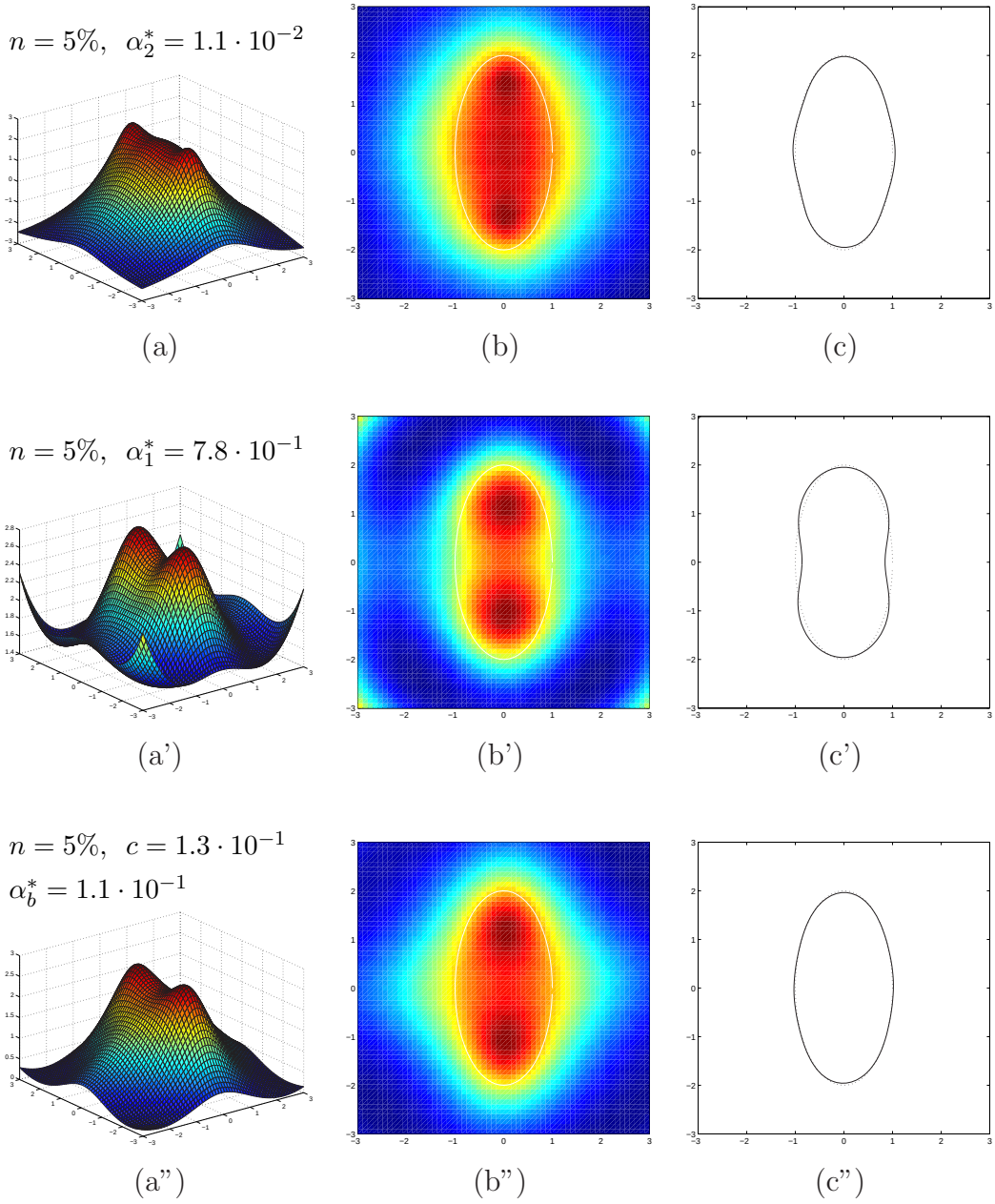

Figure B.11: The captions of panels (a), (b), (c) and (a'), (b'), (c') are respectively the same as the ones of figures B.4 and B.8, except that the noise level is now 5%. Panels (a"), (b"), (c") show the analogous results obtained by means of the blended regularization, consisting in choosing $\alpha_b^* := c\,\alpha_1^* + (1-c)\,\alpha_2^*$ (the latter is a shorthand for definition (1.346)) as optimal value of the regularization parameter. The value of $c = c(h_s)$ is heuristically chosen as $c(h_s) := [2\arctan(40\,h_s)]/\pi$, where h_s denotes the norm of the specific noise matrix $\mathbf{H}_s$ added to the exact far-field matrix $\mathbf{F}$ in this particular numerical experiment. A comparison between panels (c), (c') and (c") seems to indicate the latter as the one providing the best reconstruction of the ellipse.

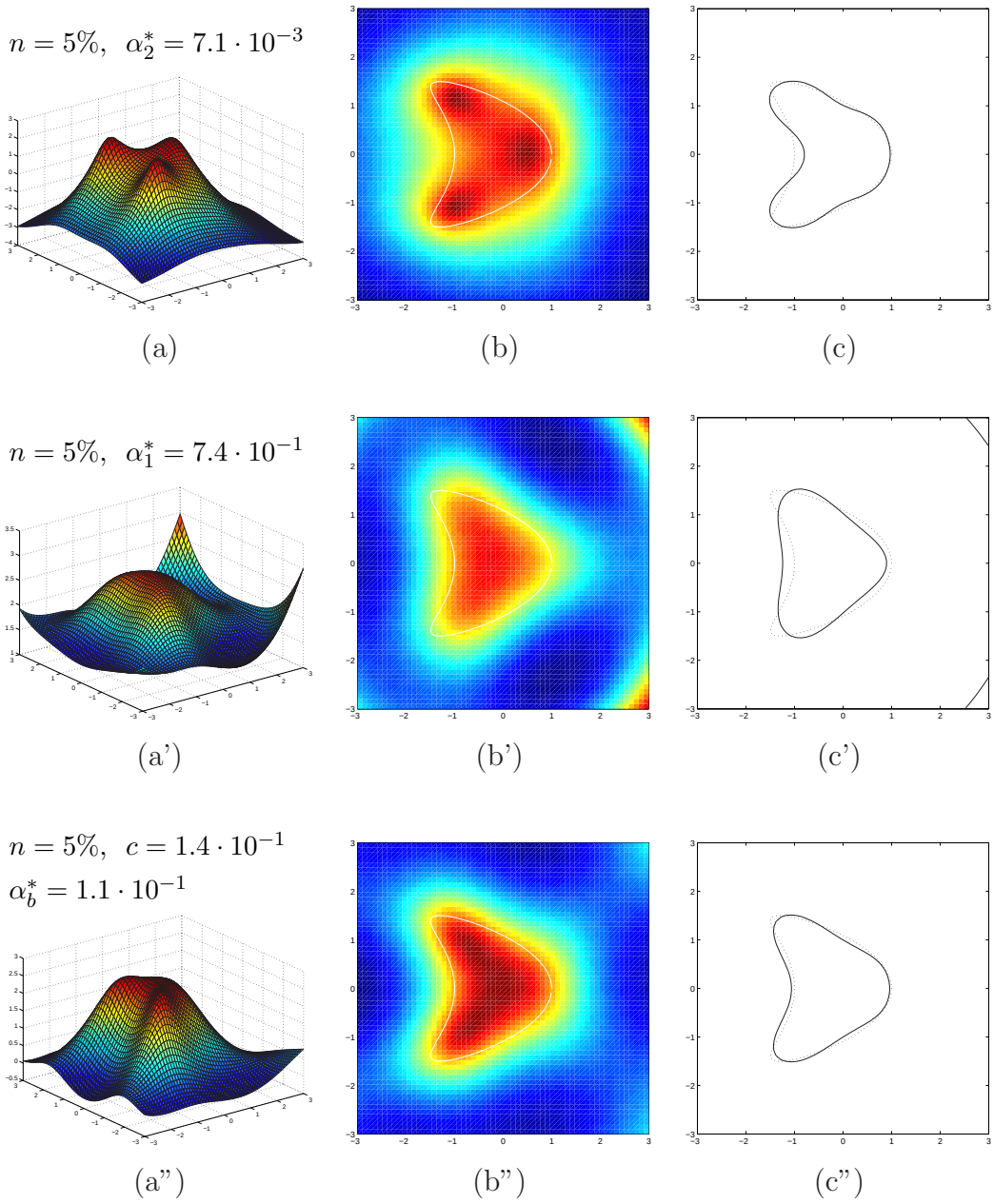

Figure B.12: The captions of panels (a), (b), (c) and (a'), (b'), (c') are respectively the same as the ones of figures B.5 and B.9, except that the noise level is now 5%. Panels (a"), (b"), (c") show the analogous results obtained by means of the blended regularization, consisting in choosing $\alpha_b^* := c\,\alpha_1^* + (1-c)\,\alpha_2^*$ (the latter is a shorthand for definition (1.346)) as optimal value of the regularization parameter. The value of $c = c(h_s)$ is heuristically chosen as $c(h_s) := [2\arctan(40\,h_s)]/\pi$, where h_s denotes the norm of the specific noise matrix $\mathbf{H}_s$ added to the exact far-field matrix $\mathbf{F}$ in this particular numerical experiment. A comparison between panels (c), (c') and (c") clearly indicates the latter as the one providing the best reconstruction of the kite.

B.3. [3.2] Band-limitedness of the indicator function

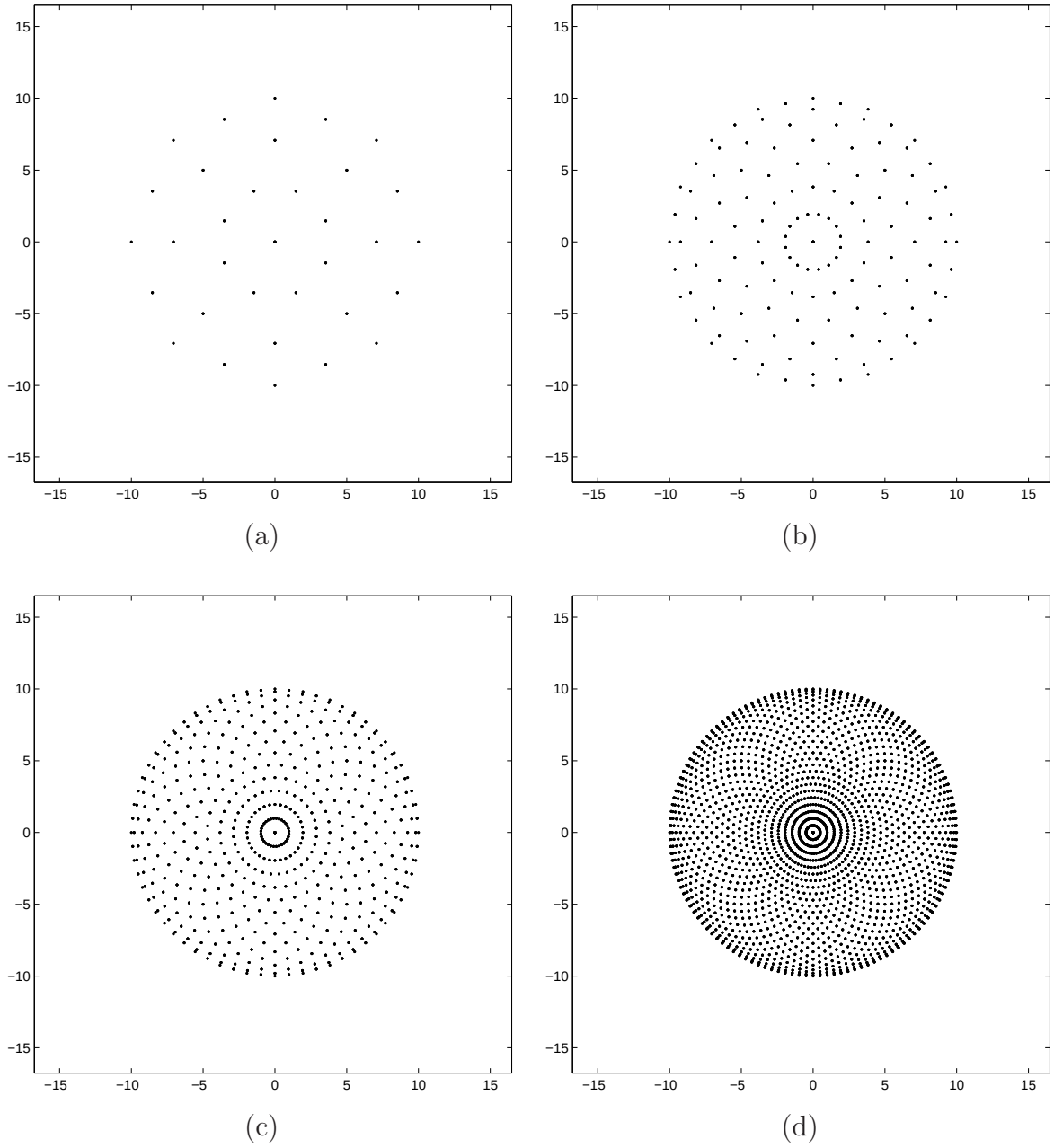

Figure B.13: Support (3.106) of the Fourier transform (3.105) of the indicator function (3.74) in the case in which the wavenumber is $k = 5$ and the number of incidence/observation angles is $N = 8$ (panel (a)), $N = 16$ (panel (b)), $N = 32$ (panel (c)), $N = 64$ (panel (d)).

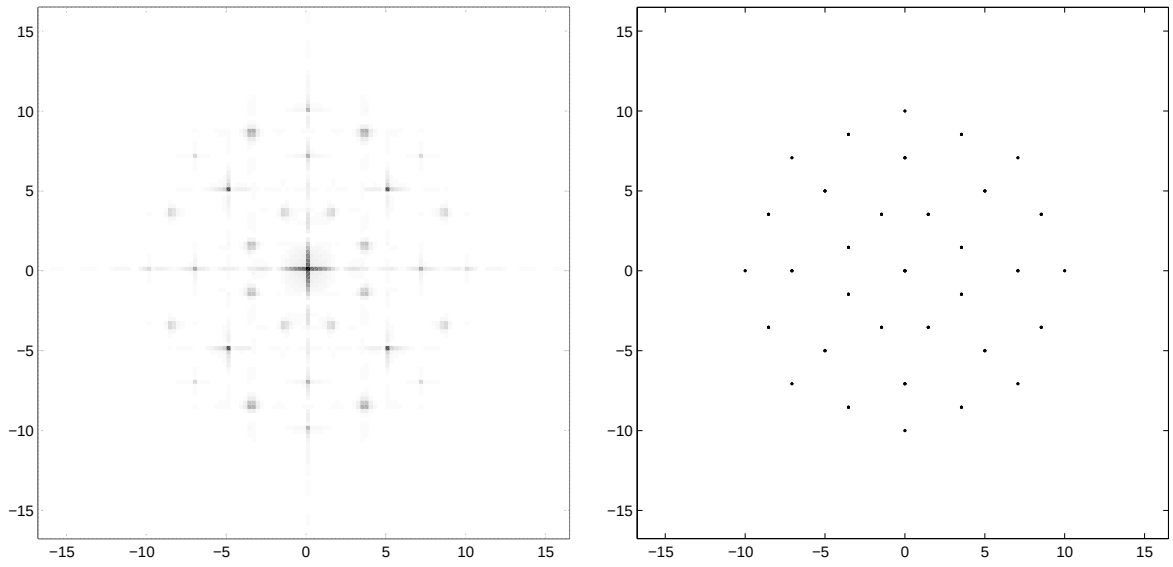

Figure B.14: Scattering of $N = 8$ plane waves with a conducting kite in the case of Dirichlet boundary conditions, for a wavenumber $k = 5$ and $N = 8$ observation angles. The numerical Fourier transform of the corresponding indicator function (panel (a)) is computed and is compared with the Dirac brush in the same scattering situation (panel (b)).

B.4. [3.3] Spatial resolution

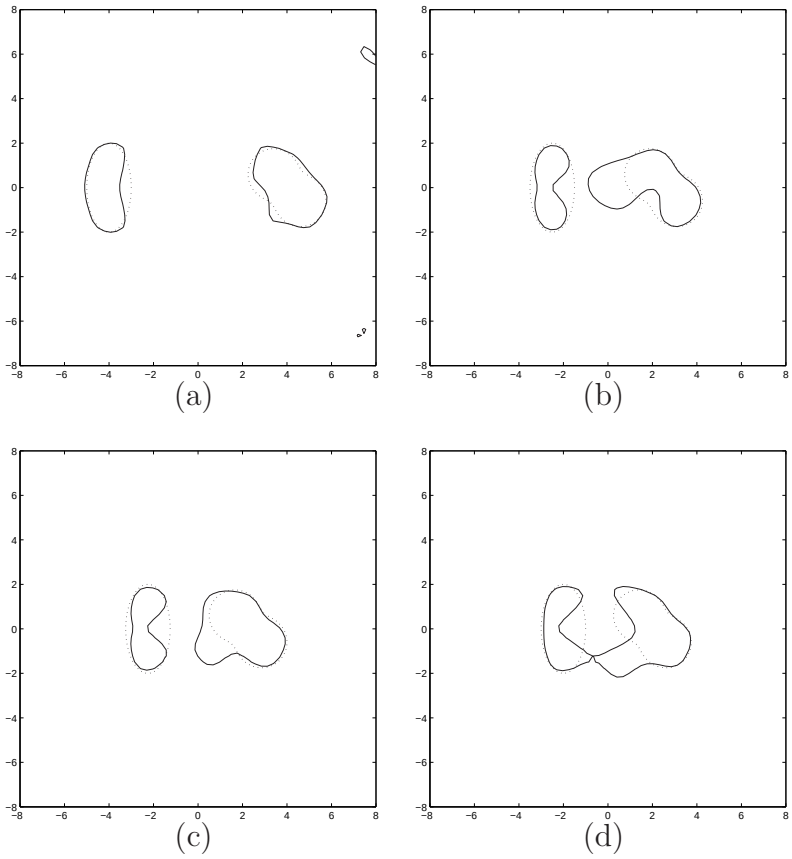

Figure B.15: Validation of the theoretical assessment of the spatial resolution in the case of two conducting scatterers: an ellipse and a peanut (dashed lines). The far-field pattern is computed in the case of $N = 32$ incidence and observation angles, and 0.5% Gaussian noise is added (cf. remark 2.5.1). The wavenumber is $k = 1$, i.e. $\lambda/4 \simeq 1.57$. The indicator function is $\Psi_2(z)$ (given by (3.74)) considered in the open square $T_A^B := (-8, 8) \times (-8, 8)$, with $\alpha^* = \alpha_1^*$ provided by the generalized discrepancy principle in the incompatible case (cf. subsection 1.8.2 and, in particular, definition (1.259), as well as its specific form (3.66) for the current context). The reconstructed profiles of the two scatterers (solid lines) are obtained by sectioning the plot of $\Psi_2(z)$ in such a way that the sum of the areas of the true scatterers and the sum of the areas described by the level curves are equal. This caption is completed by the following table B.2.

panel	C_e	C_p	d	α_1^*
(a)	$(-4; 0)$	$(4; 0)$	5.3	$3.3 \cdot 10^{-2}$
(b)	$(-2.5; 0)$	$(2.5; 0)$	2.3	$3.9 \cdot 10^{-2}$
(c)	$(-2.25; 0)$	$(2.25; 0)$	1.8	$3.2 \cdot 10^{-2}$
(d)	$(-2; 0)$	$(2; 0)$	1.3	$2.8 \cdot 10^{-2}$

Table B.2: This numerical table completes the caption of the previous figure B.15. For each panel, we give the centres of the ellipse (C_e) and of the peanut (C_p), the distance d between the two scatterers and the value α_1^* of the regularization parameter.

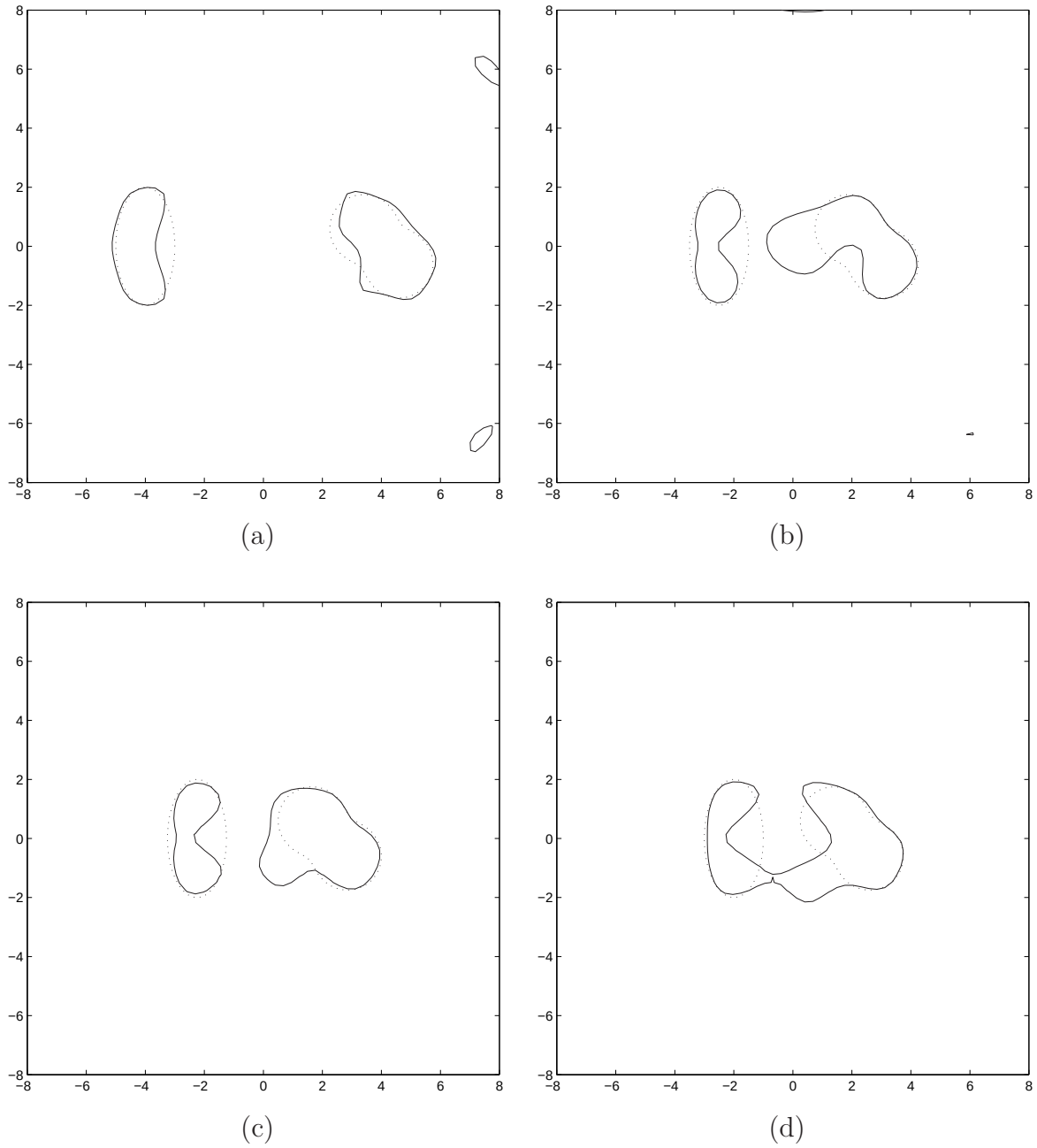

Figure B.16: The caption of this figure is the same as the one of the previous figure B.15 (table B.2 included), except that the indicator function is now $\Xi_2(z)$ (given by (3.158)) instead of $\Psi_2(z)$ (given by (3.74)).

B.5. [3.5] Deformable models

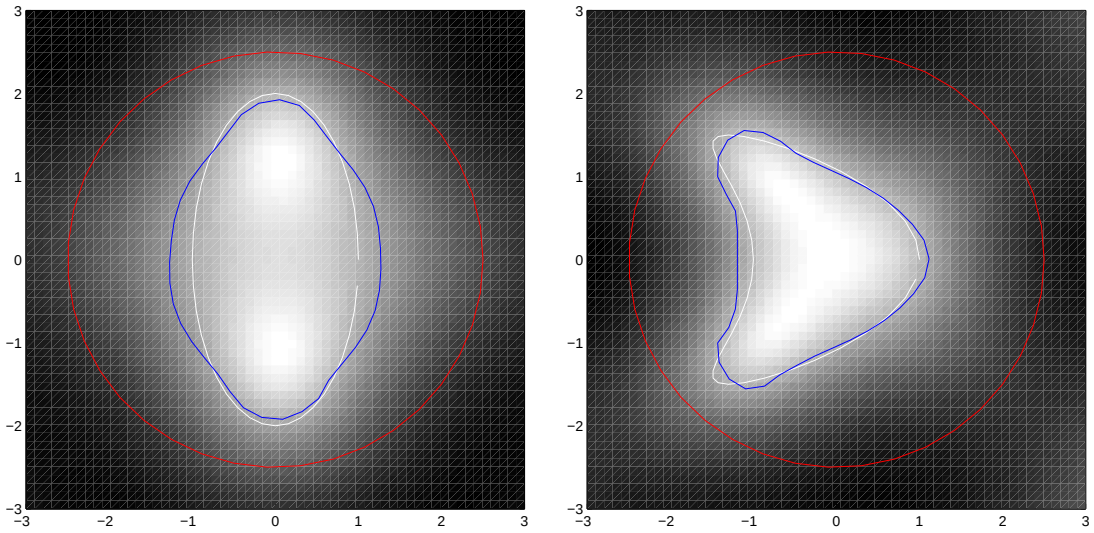

Figure B.17: Implementation of the deformable contour technique to extract the profile of a conducting ellipse-shaped (panel (a)) or kite-shaped (panel (b)) scatterer (in the case of Dirichlet boundary conditions). The wavenumber is $k = 1$, the number of incidence/observation angles is $N = 32$ and a level $n = 5\%$ of Gaussian noise is added to the respective exact far-field matrices, which are computed by using the Nyström method (cf. remark 2.5.1). Each panel consists, first of all, of a two-dimensional projection "from above" of the indicator function $\Psi_{-2\ln}(z) := -\ln \|\mathbf{g}_{\alpha_b^*}(z)\|^2_{\mathbb{C}^N}$ considered in the open square $T_A^B := (-3,3) \times (-3,3)$ (cf. definition (3.72)), where the (unique) optimal value α_b^* of the regularization parameter is fixed by means of the blended regularization (see subsection 1.8.4), i.e. as $\alpha_b^* := c\,\alpha_1^* + (1-c)\,\alpha_2^*$ (the latter is a shorthand for definition (1.346)). The value of $c = c(h_s)$ is heuristically chosen as $c(h_s) := [2\arctan(40\,h_s)]/\pi$, where h_s denotes the norm of the specific noise matrix $\mathbf{H}_s$ added to exact far-field matrix $\mathbf{F}$ in the particular numerical experiment which is being performed. Some numerical values of interest are written in the following table B.3, which completes the caption of this figure. The physical and geometrical parameters of the two experiments made to realize panels (a) and (b) are the same as the ones made for figures B.11 and B.12 respectively; however, for the same reasons explained by footnote 10 in section 3.1, the values of the regularization parameters α_1^*, α_2^* and α_b^* contained in table B.3 are slightly different from the corresponding ones in figures B.11 and B.12. Finally, in each panel of this figure B.17 we also show the true profile of the scatterer (white line), the initial guess γ_0 (red line: it is a circle of radius 2.5) and the reconstructed profile (blue line) obtained by applying the deformable model described in section 3.5 and stopping the procedure after 100 iterations (since a greater number would have provided identical visualizations).

scattering object	panel	n	α_1^*	α_2^*	c	α_b^*
ellipse	(a)	5%	$9.0 \cdot 10^{-1}$	$1.3 \cdot 10^{-2}$	$1.2 \cdot 10^{-1}$	$1.2 \cdot 10^{-1}$
kite	(b)	5%	$7.9 \cdot 10^{-1}$	$9.0 \cdot 10^{-3}$	$8.7 \cdot 10^{-1}$	$1.1 \cdot 10^{-1}$

Table B.3: This numerical table completes the caption of the previous figure B.17: as explained here, the values of the parameters α_1^*, α_2^*, α_b^* are very similar to the corresponding ones in figures B.11 and B.12.

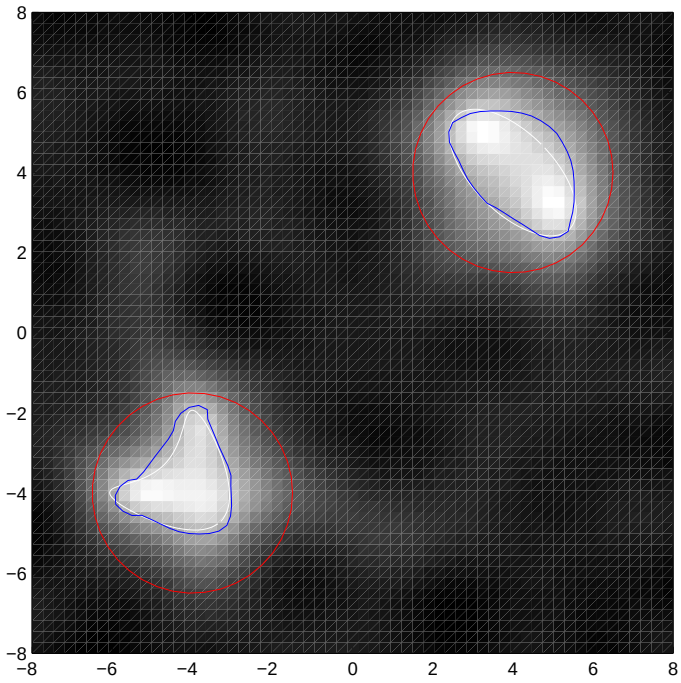

Figure B.18: Implementation of the deformable contour technique to extract the profile of a conducting double scatterer (in the case of Dirichlet boundary conditions). The wavenumber is $k = 1$, the number of incidence/observation angles is $N = 32$ and a level $n = 10\%$ of Gaussian noise is added to the exact far-field matrix, which is computed by using the Nyström method (cf. remark 2.5.1). This figure consists, first of all, of a two-dimensional projection "from above" of the indicator function $\Psi_{-2\ln}(z) := -\ln \|\mathbf{g}_{\alpha_2^*}(z)\|^2_{\mathbb{C}^N}$ considered in the open square $T_A^B := (-8, 8) \times (-8, 8)$ (cf. definition (3.72)), where the (unique) optimal value $\alpha_2^* = 4.9 \cdot 10^{-2}$ of the regularization parameter is fixed by using the generalized discrepancy principle in the compatible case (cf. subsection 1.8.3 and, in particular, definition (1.317), as well as its specific form (3.62) for the current context). The physical and geometrical parameters of the experiment made to realize this figure are the same as the ones made for panels (a'), (b') and (c') in figure B.6; however, for the same reasons explained by footnote 10 in section 3.1, the value α_2^* of the regularization parameter is different from the corresponding one in panel (a') of figure B.6. Finally, in this figure B.18 we also show the true profile of the scatterer (white line), the initial guess γ_0 (red line: two circles of radius 2.5 and centres in $(-4, -4)$ and $(4, 4)$) and the reconstructed profile (blue line) obtained by applying the deformable model described in section 3.5 and adapted to the case of two objects; more precisely, all the procedure has been implemented twice: the first time by choosing as initial guess γ_0 the single circle around the ellipse and, after 100 iterations, the second one by choosing as initial guess γ_0 the single circle around the kite and then stopping the procedure itself after 100 iterations again (since a greater number would have provided identical visualizations).

Bibliography

[1] Adams R. A. 1975 *Sobolev Spaces* (Academic Press, New York).

[2] Albanese R. A., Medina R. L. and Penn J. W. 1994 Mathematics, medicine and microwaves *Inverse Problems* **10** 995-1007.

[3] Aramini R., Brignone M. and Piana M. 2006 The linear sampling method without sampling *Inverse Problems* **22** 2237-2254.

[4] Arens T. 2004 Why linear sampling works? *Inverse Problems* **20** 163-173.

[5] Bakushinsky A. B. 1984 Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion, *USSR Compt. Math. Math. Phys. 24,4; 181-182.*

[6] Bakushinsky A. and Goncharsky A. 1994 *Ill-Posed Problems: Theory and Applications* (Kluwer Academic Publishers).

[7] Balakrishnan A. V. 1976 *Applied Functional Analysis* (Springer, New York).

[8] Bertero M. 1986 Regularization methods for linear inverse problems in *Inverse Problems* (ed. G. Talenti) (Lect. Notes in Math. vol.1225) (Springer, Berlin).

[9] Bertero M. and Boccacci P. 1998 *Introduction to inverse problems in imaging* (Institute of Physics Publishing).

[10] Bertero M., Poggio T. A. and Torre V. 1988 Ill-posed problems in early vision *Proc. IEEE* **76** 8.

[11] Birman M. S. and Solomjak M. Z. 1987 *Spectral theory of self-adjoint operators in Hilbert space* (D. Reidel Publishing Company).

[12] Brémaud P 2002 *Mathematical Principles of Signal Processing* (Springer, Berlin).

[13] Brezis H. 1986 *Analisi funzionale* (Liguori, Napoli).

[14] Brignone M. and Piana M. 2005 The use of constraints for solving inverse scattering problems: physical optics and the linear sampling method *Inverse Problems* **21** 207-222.

[15] Cakoni F. and Colton D. 2005 *Qualitative Methods in Inverse Scattering Theory* (Springer, Berlin).

[16] Cakoni F. and Colton D. 2005 Open problems in the qualitative approach to inverse electromagnetic scattering theory *Europ. J. Appl. Math.* **16** 411-425.

[17] Cakoni F., Colton D. and Haddar H. 2002 The linear sampling method for anisotropic media *J. Comput. Appl. Math.* **146** 285-9.

[18] Cakoni F., Colton D. and Monk P. 2001 The direct and inverse scattering problems for partially coated obstacles *Inverse Problems* **17** 1997-2015.

[19] Cakoni F., Colton D. and Monk P. 2004 Electromagnetic inverse scattering problem for partially coated Lipschitz Domains *Proc. Royal Society of Edinburgh* **134A** 661-682.

[20] Cakoni F., Colton D. and Monk P. 2006 The inverse electromagnetic scattering problem for a partially coated dielectric *J. Comput. Appl. Math.* (to appear).

[21] Cakoni F. and Haddar H. 2003 Interior transmission problem for anisotropic media in *Mathematical and Numerical Aspects of Wave Propagation* (ed. Cohen et al.) (Springer, Berlin) 613-618.

[22] Catapano I., Crocco L. and Isernia T. 2006 On *simple methods* for shape reconstruction of unknown scatterers IEEE *Trans. Ant. Prop.* (in press).

[23] Cohen I., Cohen L. D. and Ayache N. 1992 Using deformable surfaces to segment 3D images and infer differential structures, *CVGIP: Image Understanding* **56** 242-263.

[24] Cohen L. D. 1991 On active contour models and balloons *CVGP: Image Understanding* **52** 242-263.

[25] Colli Franzone P., Taccardi B. and Viganotti C. 1977 *Adv. Cardiol.* **21** 167.

[26] Colton D. and Kirsch A. 1996 A simple method for solving inverse scattering problems in the resonance region *Inverse Problems* **12** 383-393.

[27] Colton D. and Kress R. 1998 *Inverse Acoustic and Electromagnetic Scattering Theory* (Springer, Berlin; 2nd edition).

[28] Colton D. and Monk P. 2006 Target identification of coated objects *IEEE Trans. Ant. Prop.* **54 (4)** 4: 1232-42.

[29] Colton D., Piana M. and Potthast R. 1997 A simple method using Morozov's discrepancy principle for solving inverse scattering problems *Inverse Problems* **12** 1477-1493.

[30] Costabel M. 1988 Boundary integral operators on Lipschitz domains: elementary results *SIAM J. Math. Anal.* **19** 613-626.

[31] Courant R. and Hilbert D. 1953 *Method of Mathematical Physics* (Interscience Publishers, Inc., New York).

[32] Dunford N. and Schwartz J. T. 1957 *Linear Operators* (Interscience Publishers, Inc., New York).

[33] Engl H. W., Hanke M. and Neubauer A. 1996 *Regularization of Inverse Problems* (Kluwer Academic Publishers).

[34] Evans L. C. 1998 *Partial Differential Equations* (AMS, Rhode Island).

[35] Gilardi G. 1994 *Analisi tre* (McGraw-Hill, Milano).

[36] Gilbarg D. and Trudinger N. S. 1983 *Elliptic Partial Differential Equations of Second Order* (Springer, Berlin; 2nd edition).

[37] Groetsch C. W. 1993 *Inverse Problems in the Mathematical Sciences* (Vieweg, Wiesbaden).

[38] Gylys-Colwell F. 2000 An inverse problem for the Helmholtz equation *Inverse Problems* **16** 139-156.

[39] Hadamard J. 1902 *Bull. Univ. Princeton* **13** 49.

[40] Hadamard J. 1923 *Lectures on Cauchy's Problem in Linear Partial Differential Equations* (New Haven, CT: Yale University Press).

[41] John F. 1982 *Partial Differential Equations* (Springer, Berlin).

[42] Kass M., Witkin A. and Terzopoulos A. R. 1987 Snakes: Active contour models, *Int. J. Computer Vision* Vol. 1 **4** 321-331.

[43] Keller J. B. 1976 *Am. Math. Monthly* **83** 107.

[44] Kirsch A. 1998 Characterization of the shape of a scattering obstacle using the spectral data of the far-field operator *Inverse Problems* **14** 1489-1512.

[45] Kolmogorov A. N. and Fomin S. V. 1980 *Elementi di teoria delle funzioni e di analisi funzionale* (Edizioni Mir, Mosca).

[46] Krantz S. G. 2001 *Function Theory of Several Complex Variables* (AMS Chelsea Publishing, Rhode Island; 2nd edition).

[47] Kress R. 1999 *Linear Integral Equations* (Springer, Berlin; 2nd edition).

[48] Landau L. D., Lifšits E. M. and Pitaevskij L. P. 1986 *Elettrodinamica dei mezzi continui* (Editori Riuniti, Roma).

[49] Lang S. 1993 *Real and functional analysis* (Springer, Berlin).

[50] McLean W. 2000 *Strongly Elliptic Systems and Boundary Integral Equations* (Cambridge University Press).

[51] Moretti V. 2003 *Struttura Matematica delle Teorie Quantistiche e Teoria Spettrale in Spazi di Hilbert* at the e-address: http://www.science.unitn.it/˜moretti/dispense.html

[52] Morozov V. A. 1984 *Methods for Solving Incorrectly Posed Problems* (Springer, New York).

[53] Morozov V. A. 1966 On the solution of functional equations by the method of regularization, *Soviet Math. Dokl.* **7** 414-417.

[54] Müller C. 1945/46 Zur mathematischen Theorie elektromagnetischer Schwingungen, *Abh. deutsch. Akad. Wiss. Berlin* **3** 5-56.

[55] Papoulis A. 1968 *Systems and Transforms with Applications in Optics* (McGraw-Hill Book Company, New York).

[56] Piana M. 1998 On uniqueness for anisotropic inhomogeneous inverse scattering problems *Inverse Problems* **14** 1565-1579.

[57] Reed M. and Simon B. 1972 *Methods of modern mathematical physics I* (Academic, New York).

[58] Rellich F. 1943 Über das asymptotische Verhalten der Lösungen von $\Delta u + \lambda u = 0$ im unendlichen Gebieten, *Jber. Deutsch. Math. Verein.* **53** 57-65.

[59] Silver S. 1949 *Microwave Antenna Theory and Design.* (M.I.T. Radiation Laboratory Series vol. 12, McGraw-Hill, New York).

[60] Rynne B. P. and Sleeman B. D. 1991 The interior transmission problem and inverse scattering from inhomogeneous media, *SIAM J. Math. Anal.* **22** 1755-1762.

[61] Schwartz L. 1966 *Theorie des Distributions* (Hermann, Paris).

[62] Simonetti F. 2006 Multiple scattering: the key to unravel the subwavelength world from the far-field pattern of a scattered wave *Phys. Rev.* E **73** 1-13.

[63] Sommerfeld A. 1912 Die Greensche Funktion der Schwingungsgleichung, *Jber. Deutsch. Math. Verein.* **21** 309-353.

[64] Tikhonov A. N. 1963 Regularization of incorrectly posed problems *Soviet Math. Dokl.* **4** 1624.

[65] Tikhonov A. N. 1964 Solution of linear integral equations of the first kind *Soviet Math. Dokl.* **5** 835.

[66] Vekua I. N. 1943 Metaharmonic functions, *Trudy Tbilisskogo Matematichesgo Instituta* **12** 105-174.

[67] Tikhonov A. N., Goncharsky A. V., Stepanov V. V. and Yagola A. G. 1995 *Numerical Methods for the Solution of Ill-Posed Problems* (Kluwer Academic Publishers).

[68] Zeidler E. 1995 *Applied Functional Analysis* (Springer, Berlin).