Reprint from

Topics in Current Physics
Volume 20: **Inverse Scattering Problems** in Optics
Editor: H. P. Baltes

© by Springer-Verlag Berlin Heidelberg 1980
Printed in Germany. Not for Sale.

Springer-Verlag
Berlin Heidelberg New York

Inverse Scattering Problems in Optics

Editor: H. P. Baltes

5. The Stability of Inverse Problems

M. Bertero, C. De Mol, and G. A. Viano

With 7 Figures

Many inverse problems arising in optics and other fields like geophysics, medical diagnostics and remote sensing, present numerical instability: the noise affecting the data may produce arbitrarily large errors in the solutions. In other words, these problems are *ill-posed* in the sense of Hadamard.

The basic point, in the study of ill-posed problems, is that the development of adequate computational methods, leading to stable results, requires *prior knowledge* of properties of the admissible solutions: global bounds, smoothness conditions, positivity constraints, statistical properties, etc. The problem is first to incorporate the supplementary constraints in the computational algorithm, and secondly to estimate the accuracy of the solutions for a given prior knowledge and data accuracy. General methods are available only for linear inverse problems.

This chapter begins with an outline of the main features of ill-posed problems, of their connection with inverse problems and of the basic ideas enabling one to solve them. Next we discuss *regularization theory* where the supplementary constraints are prescribed bounds on the class of admissible solutions. Then we analyze the application to ill-posed problems of the method of linear mean square estimation (*optimum filtering*), when prior knowledge of statistical properties of the solutions is available. Finally, we review the applications of the previous methods to some linear inverse problems in optics and scattering theory.

5.1 Ill-Posedness in Inverse Problems

The concept of *ill-posedness* was introduced by HADAMARD [5.1] in the field of partial differential equations. For years, ill-posed problems have been considered as mere mathematical anomalies. Indeed, it was believed that physical situations were leading only to well-posed problems like, for instance, the Dirichlet problem for elliptic equations of potential theory, or the Cauchy problem for hyperbolic equations describing wave motion. However, it appeared later that this attitude was erroneous and that many ill-posed problems, generally inverse problems, were arising from practical situations. Nowadays there is no doubt that a systematic study of these problems is of great relevance in many fields of applied physics.

5.1.1 Well-Posed and Ill-Posed Problems

It is rather difficult to give a precise and exhaustive definition of an ill-posed problem. Indeed this term covers a lot of various problems presenting many common features but also differences so important that a global and unified theory is not yet available. The best characterization is perhaps a negative one: ill-posed problems do not fulfill all the required conditions for well-posedness [5.1], i.e., *existence*, *uniqueness* and *continuity* of the solution on the data (requirement of stability). As clearly stated by COURANT and HILBERT [Ref.5.2, p.227], *"the third requirement, particularly incisive, is necessary if the mathematical formulation is to describe observable natural phenomena. Data in nature cannot possibly be conceived as rigidly fixed; the mere process of measuring them involves small errors. Therefore a mathematical problem cannot be considered as realistically corresponding to physical phenomena unless a variation of the given data in a sufficiently small range leads to an arbitrary small change in the solution. This requirement of "stability" is not only essential for meaningful problems in mathematical physics, but also for approximation methods"*.

An example of a well-posed problem is to find a solution u of the Laplace equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \qquad (5.1)$$

in some domain D of the plane, with the condition u = g on the boundary of D (Dirichlet problem). It is well known that there exists a unique solution which depends continuously on the data. Indeed, the maximum principle [Ref.5.2, p.255] guarantees that when g is slightly perturbed into g', the corresponding solution u' is in a neighborhood of u. More precisely, $|g - g'| \leq \epsilon$ implies $|u - u'| \leq \epsilon$.

Any problem failing to satisfy one or more of the three requirements quoted above might be called an ill-posed (or improperly posed) problem. Nevertheless, this term is usually reserved to those problems for which the second requirement (uniqueness) is fulfilled, but not the first and the third ones. Indeed, as we shall see below, existence and continuity are in general closely related.

The first who pointed out the concepts of well- and ill-posedness was J. Hadamard. Let us recall his famous example showing the lack of continuity on the data in the Cauchy problem for elliptic partial differential equations. Consider (5.1) with the boundary conditions

$$u(x, 0) = 0 \quad , \quad \frac{\partial u}{\partial y}(x, 0) = \frac{1}{n} \sin(nx) \quad . \qquad (5.2)$$

It is straightforward to verify that this problem has the following solution:

$$u(x, y) = \frac{1}{n^2} \sin(nx) \sinh(ny) \quad . \qquad (5.3)$$

The term $n^{-1} \sin(nx)$ departs from zero on the x axis in an imperceptible way for n sufficiently large. However, because of the hyperbolic sine, the solution (5.3) becomes enormous at any given distance from the x axis, provided that n is sufficiently large.

Related to the Cauchy problem for the Laplace equation is the analytic continuation of functions of a complex variable. In fact, let the values of the harmonic function u, i.e., the solution of (5.1), and its normal derivative $\partial u/\partial n$ be known on some curve Γ. We denote by f(z), z = x + iy, the analytic function f = u + iv, where v is the function conjugate to u. Then, on the curve Γ, v is related to u as follows

$$v(z) = \int_{z_0}^{z} \frac{\partial u}{\partial n}(z')ds + \text{constant} \ , \qquad (5.4)$$

where z_0 is one of the endpoints of Γ. Hence, if u and $\partial u/\partial n$ are known on Γ, one may consider that the values of the analytic function f(z) on Γ are known. This shows that the solution of the Cauchy problem for the Laplace equation gives the analytic continuation of f outside Γ, which is therefore also an ill-posed problem.

Moreover, it is worth noting that the determination of an analytic function from its values on a curve Γ, inside the domain of regularity, is a problem which can be reduced to the solution of a Fredholm integral equation of the first kind, by means of the well-known Cauchy formula. Therefore, it is quite natural to guess that also integral equations of the first kind give rise to ill-posed problems. This is indeed true, as we shall show in Sect.5.1.2.

To be convinced of the practical relevance of ill-posed problems, it is sufficient to have a glance at the enormous amount of literature devoted to this field. Many references may be found for instance in the books by LAVRENTIEV [5.3], TIKHONOV and ARSENINE [5.4] and PAYNE [5.5].

5.1.2 Ill-Posedness and Numerical Instability

Let us consider the following Fredholm integral equation of the first kind:

$$\int_a^b K(x, y)\bar{f}(y)dy = \bar{g}(x) \ , \quad c \leq x \leq d \quad , \qquad (5.5)$$

where the kernel K(x, y) is supposed to be continuous. Assuming that there exists a unique solution $\bar{f}$ corresponding to $\bar{g}$, we might add to that solution a function $f^{(n)}(x) = C \sin(nx)$ where C is an arbitrary constant. From the Riemann-Lebesgue theorem we know that

$$\lim_{n \to +\infty} \int_a^b K(x, y) \sin(ny)dy = 0 \quad . \qquad (5.6)$$

Hence, taking the constant C and the integer n sufficiently large, we see that widely different functions $\bar{f}$ give approximately the same $\bar{g}$. As in the case of the Cauchy problem for the Laplace equation, small modifications of $\bar{g}$ can alter radically the solution of (5.5).

Without being conscious of the ill-posedness of this problem, one could try to solve numerically (5.5) by discretizing it. By means of some N-point quadrature formula, the integral in (5.5) may be approximated by a finite sum. Then, supposing that $\bar{g}$ is given in M points, the integral equation becomes a linear algebraic system

$$[K]\underline{f} = \underline{g} \quad , \tag{5.7}$$

where [K] is a M×N matrix of components $K_{mn} = K(x_m, y_n)w_n$ (the w_n are the weight factors depending upon the quadrature formula used) while $\underline{f} = (\bar{f}(y_n))$ and $\underline{g} = (\bar{g}(x_m))$ are vectors in euclidean spaces of dimension N and M, respectively. At this point, let us introduce the usual euclidean scalar product between two M-dimensional vectors

$$(\underline{g}, \underline{h})_M = \sum_{m=1}^{M} g_m h_m^* \tag{5.8}$$

and the corresponding euclidean norm $\| \underline{g} \|_M^2 = (\underline{g}, \underline{g})_M$. Now, when $\underline{g}$ is affected by errors, one could always add to a given solution $\underline{f}$ a spurious vector $\underline{u}$ such that

$$\| [K]\underline{u} \|_M^2 = ([K]\underline{u}, [K]\underline{u})_M \leq \varepsilon^2 \quad , \tag{5.9}$$

where ε is an estimate of data accuracy. Let us now investigate the shape of the set of those $\underline{u}$ satisfying (5.9). To this purpose let us put the quadratic form (5.9) in a somewhat different form

$$([K]^*[K]\underline{u}, \underline{u})_N \leq \varepsilon^2 \quad , \tag{5.10}$$

where [K]* is a N×M matrix denoting the adjoint (or hermitian conjugate) matrix of [K]. Even if [K] is not a square matrix, [K]* [K] is a N×N symmetric, nonnegative matrix, so that it can be diagonalized. Let us denote by λ_n^2 the eigenvalues of [K]* [K] (λ_n is also called a singular value of [K]) and assume that they are all strictly positive. Of course this can happen only if N≤M. Then inequality (5.10) defines the interior of a N-dimensional nondegenerate ellipsoid with center at the origin and axes directed along the eigenvectors of [K]* [K]. The length of each axis is given by $a_n = \varepsilon/\lambda_n$, n = 1, ..., N, and when the eigenvalues λ_n^2 are ordered in decreasing magnitude, the length of the greatest axis is ε/λ_N, while the length of the shortest one is ε/λ_1. The ratio between the two lengths, $\alpha = \lambda_1/\lambda_N$ is the so-called *condition number* of the matrix [K]. When α is much greater than one, the ellipsoid (5.10) contains, along certain principal directions, vectors whose euclidean norm is very large. A small change in the data vector $\underline{g}$ may produce a large error in the solution (or pseudo-solution) of (5.7). The algebraic system (5.7) is then said to be *ill-conditioned*.

In general this actually arises when discretizing Fredholm equations of the first kind. Indeed, let us consider for simplicity an integral operator whose kernel K(x, y) is symmetric, and let us assume that it does not have the eigenvalue zero [of course, we also assume a = c and b = d in (5.5)]. Then, as it is well known, such an operator admits an infinite sequence of real eigenvalues (with finite multiplicity) accumulating to zero [Ref.5.6, Chap.2]. Hence it is easy to understand that the finer the discretization of (5.5) is (i.e., the larger N and M), the worse conditioned the resulting system (5.7) is.

5.1.3 General Formulation of Linear Inverse Problems

In order to make precise the concepts illustrated in the previous sections concerning instability, we must specify the sets to which the data and the solutions belong. Moreover, we must define what is meant by "closeness" in each set. This can be done by introducing a norm and defining a distance between two functions of the set as the norm of their difference. Particularly important in many applications is a norm induced by a scalar product (or inner product) like the norm of a vector in Euclidean space. In that way one may speak about angles and perpendiculars and perform the familiar geometrical constructions even for infinite dimensional spaces. A typical and very important example is the space of square integrable functions on some interval (a, b). This space, called $L^2(a, b)$, is equipped with the following scalar product

$$(f, g) = \int_a^b f(x)g^*(x)dx \tag{5.11}$$

and the induced norm is

$$\| f \| = (f, f)^{\frac{1}{2}} = \left(\int_a^b |f(x)|^2 dx \right)^{\frac{1}{2}} \quad . \tag{5.12}$$

The space $L^2(a, b)$ is not only a normed space, but also a *Hilbert space* [Ref.5.7, Chap.1]. This means that it is *complete* with respect to the norm, i.e., that every Cauchy sequence converges to an element of the space. Moreover, it is a separable space: there exists a countably infinite orthonormal sequence (u_n) such that every element of the space can be indefinitely approximated in norm by linear combinations of the vectors u_n. Such a sequence is called a basis and every function f can thus be written as

$$f = \sum_{n=0}^{+\infty} f_n u_n \quad , \tag{5.13}$$

where $f_n = (f, u_n)$ are the Fourier components of f with respect to the basis (u_n). In the following we shall often use the so-called *Parseval equality* which expresses the scalar product of two functions in terms of their Fourier components

$$(f, g) = \sum_{n=0}^{+\infty} f_n g_n^* \quad . \tag{5.14}$$

Another norm, which is often used in the case of continuous functions on the closed interval [a, b], is the so-called *uniform norm* defined as follows:

$$\| f \| = \max_{a \leq x \leq b} |f(x)| \quad , \tag{5.15}$$

i.e., the maximal value of the modulus of f on the interval [a, b]. Convergence with respect to the norm (5.15) is uniform convergence and the space of continuous functions is complete with respect to this norm.

After these few preliminaries, we can give a more precise meaning to the concept of *ill-posed linear inverse problems*.

First let us define the *direct problem*: it is a mapping of a space F of functions, called by CHADAN and SABATIER [5.8] "parameters" and by BALTES [Ref.5.9, p.1] "source functions", into a space $\bar{G}$ of functions, called "results" or "data". In the analysis of imaging systems a function of F is called an "object" and a function of $\bar{G}$ a "noiseless image". We assume that F, $\bar{G}$ are normed spaces and that the mapping is given by a linear operator A. We write $A: F \to \bar{G}$ and, in mathematical language, the space $\bar{G}$ is called the range of the operator A.

Usually the operator A is continuous. This means that to any sequence of elements of F, say $\{f^{(n)}\}$, converging to the null element, there corresponds a sequence $\{Af^{(n)}\}$ which converges to the null element of $\bar{G}$. This property ensures the stability of the direct problem: any perturbation of $\bar{g}$ vanishes when the inducing perturbation of f tends to zero. Besides it is always possible to introduce a norm in $\bar{G}$ such that $\bar{G}$ becomes a complete normed space. Let us assume now that the inverse mapping A^{-1} exists, which is equivalent to require that the equation Af = 0 has only the trivial solution f = 0. Then a theorem of Banach [Ref.5.10, p.83] implies that A^{-1} is also continuous. At this point one could try to define the inverse problem as the problem of solving the functional equation

$$A\bar{f} = \bar{g} \quad , \tag{5.16}$$

where $\bar{g}$ is a given function of $\bar{G}$. The continuity of A^{-1} would ensure the stability of the solution.

However, this approach is inadequate for the following reason. The operator A has usually a smoothing effect. Consider, for instance, the integral operator of (5.5): if the kernel K(x, y) has continuous derivatives with respect to x up to a certain order, then the same property holds for $\bar{g}(x)$. In any case the operator attenuates the higher frequencies - see (5.6). Now, in general, measurement errors or noise destroy the smoothness properties of $\bar{g}$: the "measured result" g is no longer a function of $\bar{G}$ (in the case of imaging systems g is the "noisy image"). In other words,

as remarked by SABATIER [Ref.5.11, p.5], one has to extend the space $\bar{G}$ into a larger space G containing all possible results of measurements. The space G must be equipped with a norm suitable for describing experimental errors: a L^2- space with the L^2-norm for instance, when one considers mean-squared errors, or a space of continuous functions with the uniform norm (5.15), when one considers maximal absolute errors. It happens that $\bar{G}$ is no longer a complete space with respect to the norm of G and the operator A^{-1} is no longer continuous. *The inverse problem turns out to be an ill-posed problem.* Besides the equation Af = g might have no solution, because g does not necessarily belong to $\bar{G}$. We see that the questions of existence and continuity are closely connected.

We shall call F the *solution space* and G the *data space* . If we assume a simple additive model for noise and measurements errors, then we have

$$Af + h = g \quad . \tag{5.17}$$

Since both f (the solution) and h (the noise) are unknown and since the equation Af = g might have no solution, it follows that:

I) the best we can do is to search for some f reproducing the given g within a tolerable uncertainty. The problem is then reformulated as follows: find f such that

$$\| Af - g \|_G \leq \varepsilon \quad , \tag{5.18}$$

where ε is the "size" of the noise, measured with the norm of G.
II) the previous formulation is adequate if the set H of all the functions f satisfying (5.18) is bounded and sufficiently "small" so that any element of H might be taken as an approximation of the "true" solution. However, when A^{-1} is not continuous, H is not bounded. In other words, given an arbitrary number Δ, one can find two functions $f^{(1)}$, $f^{(2)}$ satisfying (5.18) and such that $\| f^{(1)} - f^{(2)} \|_F > \Delta$. This is precisely the meaning of Hadamard's example discussed in Sect. 5.1.1. In such a case, as we shall see below, some supplementary constraints on the solution are necessary.

The situation illustrated above is quite similar to that of ill-conditioned systems as described in Sect.5.1.2. Evidently, in the finite dimensional case the set H is always bounded, but it is very large along some directions.

Finally we want to remark that, when the inverse operator does not exist, the previous analysis can be repeated considering for (5.16) only solutions of minimal norm. These solutions can be expressed in terms of the *generalized inverse* (or pseudo-inverse) of the operator A [5.12]. The generalized inverse is an extension, for operators in functional spaces, of the Moore-Penrose inverse for matrices. When the operator A has a smoothing effect, it happens that its generalized inverse is not continuous with respect to the norm of the data space G and therefore we get again an ill-posed problem.

5.1.4 Prior Knowledge as a Remedy to Ill-Posedness

In Sect.5.1.3 we saw that, for an ill-posed problem, the set H is unbounded so that (5.18) is not sufficient for determining meaningful approximate solutions. The main idea, common to most available methods for curing ill-posedness, is to restrict the class of admissible solutions by means of suitable *prior knowledge*. In the following we shall always assume that the inverse operator A^{-1} exists.

In *regularization methods*, a subset M of the solution space F is defined in such a way that the intersection of M with the set H, should be a set K of reasonably small size (see Fig.5.1).

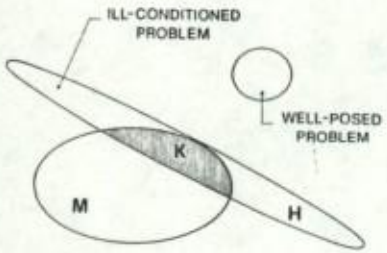

Fig. 5.1. Illustrating the difference between ill-posed and well-posed problems, and the basic idea of regularization

The problem is then said to be regularized if K collapses around the element $A^{-1}\bar{g}$, when g tends to an element $\bar{g}$ of G, i.e., when the noise tends to zero. In such a case one also says that *continuous dependence of the solution on the data has been restored*.

The set M can be defined by imposing global constraints on the class of admissible solutions. For example, one might ask for nonnegative solutions, or for solutions satisfying prescribed bounds, lying in compact sets, etc. The relevance of compactness as a way for restoring continuity was emphasized by TIKHONOV [5.13] who also introduced the concept of regularization. Besides he showed, in the case of Fredholm integral equations of the first kind, how to incorporate the constraints into the computational algorithm.

The role played by prescribed bounds in ill-posed problems for partial differential equations, has been particularly emphasized by JOHN [5.14] and PUCCI [5.15] (in Sect.5.1.5 we shall give a simple example of this approach). Later, along these lines, MILLER [5.16] formulated a regularization algorithm, having in mind the problem of analytic continuation (for an application of this method to the analytic continuation of scattering amplitudes, see also MILLER and VIANO [5.17]). Since it is formulated in the framework of Hilbert spaces, the method of MILLER [5.16] is easily adapted to the general formulation of inverse problems given in Sect.5.1.3. The prescribed bound on the solution is then expressed by means of a linear operator B as

follows

$$\| Bf \|_F \leq E \ , \tag{5.19}$$

where E is a given positive constant. The operator B is called by Miller the *constraint operator*. The method of TIKHONOV [5.13] and the method of MILLER [5.16] are not exactly equivalent even if, in many cases, they lead practically to the same results. We shall discuss these points in Sect.5.2.3.

The global bound (5.19) should express some expected properties of the solution and has to be prescribed according to the physical character of the problem one considers. If the solution represents for instance a signal, then one may know some realistic upper bound for its energy. Very often also, some smoothness condition on the solution can be prescribed by bounding its derivatives. The role of this prior knowledge is to discriminate between interesting solutions and spurious solutions generated by uncontrolled propagation of data errors. The principle of regularization methods is to include the additional conditions explicitly, at the start, instead of resorting consciously or not, during the computations, to some tricks eliminating the instability. The essential drawbacks of such apposite tricks is indeed that their implications, on the class of admissible solutions, often remain in the dark.

Another route to regularization is provided by the theory of stochastic processes. The idea is to associate random processes both to the class of admissible solutions and to the data set. Again we must have some prior knowledge about the solutions. For linear mean-square estimation (which is in general the best one can do) it is enough to know expectation values (mean values), autocorrelation and cross-correlation functions of data and solutions.

Along these lines most work has been done on ill-conditioned algebraic systems, arising from the discretization of ill-posed problems. A good review on this subject is the paper of TURCHIN et al. [5.18]. Stochastic regularization for ill-posed problems has been formulated by LAVRENTIEV [5.3] , MOROZOV [5.19] and FRANKLIN [5.20].

When data and solutions belong to Hilbert spaces, the fundamental mathematical tool is given by the theory of *weak random variables* [Ref.5.7, Chap.6]. Then linear mean-square estimation (optimum filtering) is performed, allowing a comparison with regularization theory based on the constraint (5.19), as presented in [5.21,22]. In the stochastic approach, one says that a continuous dependence of the solution on the data (i.e., stability) has been restored if the mean-square error on the solution tends to zero when the noise tends to zero. This requirement imposes conditions on the autocorrelation functions (covariance operators) of the solutions and of the noise. Stochastic regularization methods will be analyzed in Sect.5.3.

5.1.5 Hölder and Logarithmic Continuity

In Sects.5.2.4 and 5.3.3, we will review some convergence results ensuring that continuity on the data has truly been restored, by means of prior knowledge. This means that the error in the solution converges to zero when the data error ε vanishes. However, convergence theorems are not necessarily enough for practical purposes: one has to know whether the convergence is fast enough or not in order to allow efficient numerical computations. In this connection, we have to distinguish between two different types of continuity. For some problems, the error on the solution is proportional to ε^{α}, $0<\alpha<1$. Such a continuity, called *Hölder continuity*, may in general be considered as fairly satisfactory [5.14]. Indeed, the number of significant digits in the solution is then a fixed percentage of the number of significant digits in the data. However, there are some problems where the optimal bounds for the solution error are proportional only to $|\ln \varepsilon|^{-\beta}$, $\beta>0$. Then a lowering of the data noise by many orders of magnitude, does not improve significantly the solution accuracy. This is *logarithmic continuity* which appears very poor for numerical computations. In other words, a real improvement of the solution is only possible if further constraints may be introduced.

As an illustration of these different types of continuity, let us consider the problem of analytic continuation of functions holomorphic in the unit disk. Suppose that the data are given on a curve Γ contained in the interior of the disk. The domain whose boundary is the unit circle and the curve Γ, may be conformally mapped in an annulus with inner radius a and outer radius R. The data are then given on the inner circle. Let f(z) be a function holomorphic in the annulus; if we denote by M(ρ) the maximum of the modulus of f[ρ exp(iφ)] on the circle of radius ρ, $a \le \rho \le R$, then by Hadamard's three circle theorem [Ref.5.23, Chap.5], we have

$$M(\rho) \le [M(a)]^{\alpha}[M(R)]^{1-\alpha} \quad , \quad \alpha = \frac{\ln(\rho/R)}{\ln(a/R)} \quad . \tag{5.20}$$

From this inequality it follows that analytic continuation to points within the annulus is stable if prior knowledge assures that the admissible functions are bounded by E on the outer circle. Indeed, let us take two such functions $f^{(1)}$, $f^{(2)}$ and consider their difference $f = f^{(1)} - f^{(2)}$. The modulus of f is bounded by 2ε on the inner circle and by 2E on the outer circle, so that, by (5.20), it is bounded by $2E(\varepsilon/E)^{\alpha}$ on any circle of radius ρ. This result implies Hölder continuity for analytic continuation to points within the domain of analyticity. However, α is equal to zero for ρ = R, and therefore (5.20) does not ensure the stability up to the outer circle. If one pretends to continue a function precisely up to the boundary of its analyticity domain, then a more restrictive bound is necessary. For instance, one requires that also the first derivative is bounded. In such a case it is possible to show that, at the boundary, one gets logarithmic continuity [5.14]. This fact is not surprising since analytic functions are smooth and well behaved deep inside their holomorphy domain, but may grow rough and oscillatory when approaching the boundary.

For general inverse problems, one expects that the type of restored continuity will depend upon the smoothing or filtering effect of the operator A. Consider for instance a Fredholm integral operator. Then the regularity properties of its kernel are related to the decreasing rate of its eigenvalues. In particular, for analytic kernels, the eigenvalues tend exponentially to zero [5.24] and therefore, if the constraint is not too restrictive, we get only logarithmic continuity. In other words, some information contained in the "true" solution is lost in the data. Accordingly one expects that there are severe limitations on the restoration of fine details in the solution. To check this point it is convenient to consider the reconstruction of a "blurred solution", i.e., the restoration of local weighted averages (Sect.5.2.4). Then it is possible to define a "resolution limit", practically noise independent, giving a measure of the size of the finest details which can be restored. This will be illustrated by many examples in Sect.5.4.

5.2 Regularization Theory

The concept of *regularization* was introduced by TIKHONOV [5.13] in the study of Fredholm integral equations of the first kind. The basic ideas have already been discussed in Sect.5.1.4. A similar method was developed by MILLER [5.16] for improperly posed problems in a Hilbert space setting. We chose the latter method for the following reasons. Firstly because, using the geometrical properties of Hilbert spaces, it is possible to justify, by means of elementary arguments, the main points of the theory. Secondly because Miller's theory allows precise estimations of the solution accuracy (Sect.5.2.4).

5.2.1 An Outline of Miller's Theory

As seen in Sects.5.1.3,4, the regularization of a linear inverse problem can be formulated as follows: to search for functions f satisfying both constraints

$$\| Af - g \|_G \le \varepsilon \tag{5.21}$$

and

$$\| Bf \|_F \le E \quad . \tag{5.22}$$

The spaces F and G are Hilbert spaces, A:F→G is a known continuous operator, ε is an estimate of the data accuracy, E is a prescribed constant and, finally, B:F→F is the constraint operator.

Many different choices are possible for B, according to the available prior knowledge. The simplest one is B = I (the identity operator in F); then (5.22) is a constraint on the norm of f. Another usual choice is to let B be a differential operator (see Sect.5.2.3) and then the bound (5.22) is a smoothness requirement on the solution. However, for the general formulation of the theory, it is not necessary to specify B. It is only required that B is densely defined in F and that it has a *continuous inverse* B^{-1}. In other words, there must exist a constant s such

that $\|Bf\|_F \geq \beta \|f\|_F$. Therefore, the set M of the functions satisfying (5.22) is bounded, and hence also the set K of the functions satisfying (5.21,22). Of course the set K is not allowed to be empty, and this property depends on the values of the numbers ε and E. A couple (ε, E) is said to be *permissible* [5.16] if there exists at least one function f which satisfies (5.21,22).

Let us denote by Π the set of permissible couples. It can be proved [5.16] that this set is convex, i.e., it contains the segment joining any two of its points. Its boundary can be drawn as follows. Let us introduce the functional

$$\Phi(\alpha; f) = \|Af - g\|_G^2 + \alpha \|Bf\|_F^2 , \tag{5.23}$$

where α is a positive parameter. Then, since B^{-1} is bounded, there exists a unique function $\bar{f}_\alpha$ minimizing the functional $\Phi(\alpha; f)$ [see also the subsequent discussion from (5.25) to (5.27)]. If we write

$$\varepsilon_\alpha = \|A\bar{f}_\alpha - g\|_G , \quad E_\alpha = \|B\bar{f}_\alpha\|_F \tag{5.24}$$

it can be proved that ε_α is a continuously increasing and E_α a continuously decreasing function of α, when α runs from 0 to $+\infty$. Moreover, since $\bar{f}_\alpha$ minimizes $\|Af - g\|_G$ under to the constraint $\|Bf\|_F = E_\alpha$ and likewise minimizes $\|Bf\|_F$ under the constraint $\|Af - g\|_G = \varepsilon_\alpha$, then Π is exactly the set of points which are above and to the right of the curve $(\varepsilon_\alpha, E_\alpha)$, $0 < \alpha < +\infty$. Since Π is a convex set, the computation of only a few points on its boundary curve, coupled with linear interpolation in between, would give a good idea of its shape.

Let us assume now that the couple (ε, E) is permissible; then if K is not too "large", any function of K may be taken as a satisfactory estimate of the unknown solution. We are faced with the following two problems:

a) how to exhibit at least one function of K;

b) how to estimate the accuracy of the solution.

For solving these problems it is convenient to find out a simpler and more symmetric geometry than that of the set K. To this purpose, one can introduce two sets K_0 and K_1 sandwiching K [5.25,26]. Indeed, if we consider the functional defined in (5.23) with $\alpha = (\varepsilon/E)^2$

$$\Phi(f) = \|Af - g\|_G^2 + \left(\frac{\varepsilon}{E}\right)^2 \|Bf\|_F^2 , \tag{5.25}$$

then the set K_0 of the functions f satisfying the condition $\Phi(f) \leq \varepsilon^2$ is contained in K, while the set K_1 of the functions f such that $\Phi(f) \leq 2\varepsilon^2$ contains K.

In order to show that the sets K_0 and K_1 have a simpler geometrical structure than the set K, we must consider the following operator:

$$C = A^*A + (\varepsilon/E)^2 B^*B , \tag{5.26}$$

where A^* and B^* are the adjoints (hermitian conjugates) of A and B, respectively. Observe that $A^* : G \to F$, so that $A^*A : F \to F$.

The operator $C : F \to F$ is defined on the domain of B^*B and has the following properties: it is a positive definite operator, i.e., for any f in its domain $(Cf, f)_F > 0$; it is self-adjoint, i.e., $C^* = C$; it has a continuous inverse since, from the analogous property assumed for B, it follows that there exists a positive constant γ^2 such that $\|Cf\|_F \geq \gamma^2 \|f\|_F$. The last property implies that, for any given g, we can introduce the function

$$\bar{f} = C^{-1}A^*g . \tag{5.27}$$

Then the functional (5.25) may be rewritten as follows:

$$\Phi(f) = (C[f - \bar{f}], [f - \bar{f}])_F + \|g\|_G^2 - (g, A\bar{f})_G . \tag{5.28}$$

It is clear that the function $\bar{f}$ minimizes the functional and that

$$\Phi(\bar{f}) = \|g\|_G^2 - (g, A\bar{f})_G \geq 0 . \tag{5.29}$$

Now, in order to investigate the geometrical structure of the sets K_0, K_1, assume for simplicity that the operator C has a complete orthonormal set of eigenfunctions (u_n); then the condition $\Phi(f) \leq \varepsilon^2$ can be written in the following form:

$$\sum_{n=0}^{+\infty} \gamma_n^2 |f_n - \bar{f}_n|^2 \leq \varepsilon^2 - [\|g\|_G^2 - (g, A\bar{f})_G] , \tag{5.30}$$

where (γ_n^2) is the set of the eigenvalues of C and f_n, $\bar{f}_n$ are the Fourier components of f, $\bar{f}$ in the basis (u_n), i.e., $f_n = (f, u_n)_F$, $\bar{f}_n = (\bar{f}, u_n)_F$. At this point it is clear that the sets K_0, K_1 are infinite-dimensional "ellipsoids" having the same center $\bar{f}$ and the same principal axes, the latter being given by the eigenvectors of C.

Now, does $\bar{f}$ belong to the set K? A sufficient condition for this is $\Phi(\bar{f}) \leq \varepsilon^2$, which can be easily verified by numerical computation.

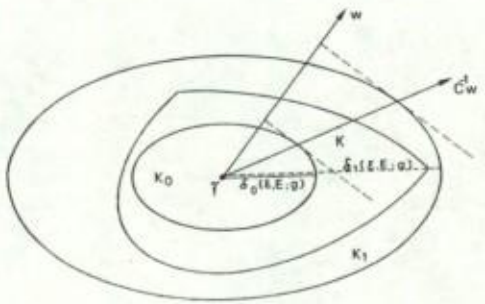

Fig. 5.2. Schematic representation of the relation between the sets K_0, K, K_1. The sets K_0, K_1 are represented as homothetic ellipses with center $\bar{f}$

Besides the set K_0 is nonvoid if and only if this condition is satisfied. In such a case, the situation is schematically represented in Fig.5.2. It is then clear that we may take $\bar{f}$ as an estimate of the unknown solution.

5.2.2 Eigenfunction Expansions and Numerical Filtering

Let us suppose now that A is a compact operator. A typical example is an integral operator over a finite interval and with continuous kernel. The corresponding inverse problem is the solution of a Fredholm equation like (5.5). Existence theory for these equations was developed by PICARD [5.27], using expansions in terms of the eigenfunctions of the operators A*A and AA*. Picard's theory generalizes immediately to the case of compact operators in Hilbert space. The operator A*A is a compact, self-adjoint, nonnegative operator and its inverse exists since A^{-1} exists. From the spectral theory for compact operators [Ref.5.7, Chap.3], it follows that A*A admits a countably infinite set of positive eigenvalues (α_n^2), and that the set (u_n) of the corresponding eigenfunctions is a basis in F. Each eigenvalue has finite multiplicity and (α_n^2) can be ordered as follows: $\alpha_0^2 \geq \alpha_1^2 \geq \alpha_2^2 \geq \dots$. Moreover $\alpha_n^2 \to 0$ when $n \to +\infty$. The $\alpha_n (\alpha_n > 0)$ and the u_n are called, respectively, singular values and singular functions of A. If we introduce the vectors

$$v_n = \alpha_n^{-1} A u_n \tag{5.31}$$

it is easy to check that

$$A u_n = \alpha_n u_n \ , \quad A^* v_n = \alpha_n u_n \tag{5.32}$$

and that

$$A^* A u_n = \alpha_n^2 u_n \ , \quad A A^* v_n = \alpha_n^2 v_n \ . \tag{5.33}$$

The set (v_n) is a complete orthonormal set in the closure of the range of the operator A [Ref.5.28, Chap.5.2], i.e., in the closure of $\bar{G}$. Therefore (v_n) is a basis for representing noise-free data.

Solution (5.27) takes on a simple form when the u_n diagonalize B*B. In such a case, we have

$$B^* B f = \sum_{n=0}^{+\infty} \beta_n^2 f_n u_n \tag{5.34}$$

and B has a continuous inverse if and only if $\beta_n \geq \beta > 0$ for any n. The u_n diagonalize also the operator C, and the corresponding eigenvalues are given by $\gamma_n^2 = \alpha_n^2 + (\varepsilon/E)^2 \beta_n^2$. Now, from (5.32) it follows that

$$A^* g = \sum_{n=0}^{+\infty} \alpha_n g_n u_n \ , \tag{5.35}$$

where $g_n = (g, v_n)_G$, and from (5.27) we get

$$\bar{f} = \sum_{n=0}^{+\infty} \frac{\alpha_n}{\alpha_n^2 + (\varepsilon/E)^2 \beta_n^2} g_n u_n \ . \tag{5.36}$$

Another solution of the problem is obtained as follows. Let us assume for simplicity the the β_n^2 form a nondecreasing sequence. Then let us denote by N the greatest integer such that $\alpha_n \geq (\varepsilon/E)\beta_n$ (recall that the α_n form a nonincreasing sequence). Truncating the series (5.36) at $n = N$ and neglecting $\varepsilon\beta_n/E$ in comparison with α_n for $n \leq N$, we obtain

$$\tilde{f} = \sum_{n=0}^{N} \alpha_n^{-1} g_n u_n \ . \tag{5.37}$$

It can be proved that $\tilde{f}$ belongs to K_1 [5.16] and therefore is also an approximate solution. We recover here, with a prescribed cutoff, the well-known truncation method which is in use for eliminating the noise amplification due to eigenvalues very close to zero and which is usually called *numerical filtering* [5.29].

The preceding methods have the disadvantage that both the error bound ε and the constraint E have to be known. However, often in practice, only ε is known. Then it is possible to elaborate procedures which require the knowledge of only one element of the couple (ε, E). For instance, let us suppose that we know a good upper bound $\bar{\varepsilon}$ for the data accuracy but none for E.

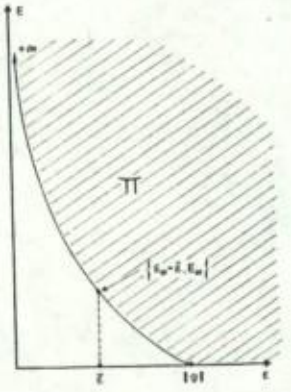

Fig. 5.3. The region Π of permissible couples (ε, E)

Nevertheless, we can obtain an estimate of the solution taking as our approximation that element of the space F which minimizes $\|Bf\|_F$ with respect to the constraint $\|Af - g\|_G = \bar{\varepsilon}$. In other words we will take as our approximation the function $\bar{f}_\alpha$ with α determined by the condition $\varepsilon_\alpha = \bar{\varepsilon}$, as shown in Fig.5.3. This method coincides with that proposed by MOROZOV [5.30] and called the "residue method". A complete discussion of this point can be found in [5.16].

Up to now we have considered an operator A such that A*A has only a discrete spectrum. By means of a simple example, we illustrate how the regularization procedure works when the spectrum is continuous. To this purpose we consider the case of a convolution operator

$$(Af)(x) = \int_{-\infty}^{+\infty} K(x - y)f(y)dy \quad . \tag{5.38}$$

If the function K(x) is integrable over (-∞, +∞), then its Fourier transform

$$\hat{K}(\nu) = \int_{-\infty}^{+\infty} K(x)\, e^{-2\pi i\nu x}dx \tag{5.39}$$

is a bounded continuous function on (-∞, +∞), such that $|\hat{K}(\nu)| \to 0$ when $|\nu| \to +\infty$ (Riemann-Lebesgue theorem). If we take as solution and data space the space of square integrable functions, i.e., $F = G = L^2(-\infty, +\infty)$, then $A : F \to G$ is a continuous operator. Besides A^{-1} (which exists if $\hat{K}(\nu)$ does not vanish over some interval) is not continuous since $\hat{K}(\nu)$ tends to zero at infinity.

Now we can take as a constraint operator B

$$(Bf)(x) = \int_{-\infty}^{+\infty} \hat{\beta}(\nu)\hat{f}(\nu)\, e^{2\pi i x\nu}d\nu \quad . \tag{5.40}$$

In such a form we can write, for instance, a differential operator of order n with constant coefficients. Then $\hat{\beta}(\nu)$ is a polynomial of order n, $\hat{\beta}(\nu) = a_0 + a_1\nu + \ldots + a_n\nu^n$, and the domain of B is the set of the functions f such that $\hat{\beta}(\nu)\hat{f}(\nu)$ is square integrable over (-∞, +∞). Besides, let us remark that B has a bounded inverse if and only if $\hat{\beta}(\nu)$ has no zeros on (-∞, +∞).

Since the operators A*A and B*B commute, we are in a situation completely analogous to that illustrated in the preceding example. Therefore, from (5.27) we get

$$\tilde{f}(x) = \int_{-\infty}^{+\infty} \frac{\hat{K}^*(\nu)}{|\hat{K}(\nu)|^2 + (\varepsilon/E)^2|\hat{\beta}(\nu)|^2}\, \hat{g}(\nu)\, e^{2\pi i x\nu}d\nu \quad . \tag{5.41}$$

This formula corresponds to the expansion (5.36). Of course, also in this case, we can obtain a second approximation corresponding to the truncated solution (5.37). Indeed, let us denote by Λ the set of the values of ν such that $|\hat{K}(\nu)| \geq (\varepsilon/E)|\hat{\beta}(\nu)|$; since $\hat{K}(\nu)$ tends to zero at infinity while $\hat{\beta}(\nu)$ is never zero, Λ is a bounded set of (-∞, +∞). Then the second solution is given by

$$\tilde{\tilde{f}}(x) = \int_{\Lambda} \frac{\hat{g}(\nu)}{\hat{K}(\nu)}\, e^{2\pi i\nu x}d\nu \quad . \tag{5.42}$$

5.2.3 Tikhonov Regularization Method

A general method for solving Fredholm equations of the first kind was proposed by TIKHONOV [5.13]. The method was successively developed and generalized by Tikhonov himself and by many Russian mathematicians. A copious list of references of the Russian school can be found in [5.4].

Let us consider the integral operator

$$(Af)(x) = \int_a^b K(x, y)f(y)dy \quad , \quad c \leq x \leq d \tag{5.43}$$

whose inverse is assumed to exist. The solution space F is the space of the continuous functions over [a, b], normed with the uniform norm (5.15). The data space G is the space $L^2(a, b)$. The purpose is to construct a uniform approximation to the solution of (5.5).

The basic idea is to restrict the class of admissible solutions to a compact subset of F. Then a general theorem of functional analysis, (due to Tikhonov himself) assures the continuity of the inverse mapping [Ref.5.4, Chap.1, Sect.1]. The restriction to a compact subset is achieved by means of a "regularizing functional" Ω(f): the compact subsets are defined by the condition $\Omega(f) \leq E^2$, where E^2 is a given arbitrary constant.

The functional proposed by Tikhonov for Fredholm equations of the first kind is

$$\Omega(f) = \int_a^b [p(x)|f'(x)|^2 + q(x)|f(x)|^2]dx \quad , \tag{5.44}$$

where the weight functions p(x) and q(x) are strictly positive. It is possible to prove, by means of the Ascoli-Arzelà theorem [Ref.5.10, Chap.1], that the set $\Omega(f) \leq E^2$ is a compact subset of F. Next a *"regularized family of approximate solutions"* $\{\tilde{f}_\alpha\}$, $\alpha > 0$, is defined as the set of functions minimizing the functionals

$$\Phi(\alpha; f) = \int_c^d |(Af)(x) - g(x)|^2dx + \alpha\Omega(f) \quad , \tag{5.45}$$

where α is a free parameter.

The functions $\tilde{f}_\alpha$ are the solutions of the Euler equation for $\Phi(\alpha; f)$

$$-\alpha\{[p(x)\tilde{f}_\alpha'(x)]' + q(x)\tilde{f}_\alpha(x)\} + \int_a^b \bar{K}(x, y)\tilde{f}_\alpha(y)dy = \bar{b}(x) \quad ,$$
$$\tilde{f}_\alpha'(a) = \tilde{f}_\alpha'(b) = 0 \quad , \tag{5.46}$$

where

$$\bar{K}(x, y) = \int_c^d K^*(s, x)K(s, y)ds \quad , \quad \bar{b}(x) = \int_c^d K^*(s, x)g(s)ds \quad . \tag{5.47}$$

Denote by $\{g_\varepsilon\}$, $\varepsilon > 0$, a family of data converging to the error-free datum $\bar{g}$ when $\varepsilon \to 0$ and let $\varepsilon = \| g_\varepsilon - \bar{g} \|_G$. Besides, let $\bar{\alpha}(\varepsilon)$ be a value of the parameter α such that $c_1\varepsilon^2 \leq \bar{\alpha}(\varepsilon) \leq c_2\varepsilon^2$ (where c_1, c_2 are given arbitrary constant); i.e., $\bar{\alpha}(\varepsilon) \sim \varepsilon^2$ when $\varepsilon \to 0$. Finally, denote by $\tilde{f}_\varepsilon$ the solution of (5.46) with $g = g_\varepsilon$ and $\alpha = \bar{\alpha}(\varepsilon)$. Then $\tilde{f}_\varepsilon \to \bar{f}$ in the uniform norm, when $\varepsilon \to 0$ ($\bar{f}$ is the unique solution corresponding to the error-free datum $\bar{g}$) [5.13]. A simple proof of this result can be found in [5.31].

In order to compare Tikhonov's method with the method outlined in Sect.5.2.1, we can take as solution space $L^2(a, b)$. The space of continuous functions is a subspace of $L^2(a, b)$ and the set $\Omega(f) \leq E^2$ is a compact subset of $L^2(a, b)$. Then we remark that the condition $\Omega(f) \leq E^2$ can be written in the form (5.22) if the operator B is such that

$$(B^*Bf)(x) = -[p(x)f'(x)]' + q(x)f(x) \; ; \quad f'(a) = f'(b) = 0 \quad . \tag{5.48}$$

Indeed, by means of a partial integration, it is easy to see that, for any f in the domain of B*B we have: $(B^*Bf, f)_F = \Omega(f)$. We want also to remark that the operator B*B has a discrete spectrum with eigenvalues accumulating to infinity. Indeed, the eigenvalue equation for B*B is nothing else but a Sturm-Liouville problem over [a, b]. It should also be remarked that the functional (5.45) coincides now with (5.23). Besides the solution $\tilde{f}$ of (5.46) can be formally written as in (5.27), with $(\varepsilon/E)^2$ replaced by α in (5.26). Clearly the method outlined in Sect.5.2.1 is essentially a generalization of Tikhonov's method to the case where F and G are Hilbert spaces and $A : F \to G$ an arbitrary linear continuous operator. Besides the parameter α is explicitly taken proportional to ε^2, the constant being given by the prescribed bound (5.22).

Numerical computations on Fredholm equations of the first kind have been done by many authors using the method described above [5.32-34]. When uniqueness does not hold, generalized inverses must be used. A simple discussion of existence and uniqueness theorems, using singular function expansions can be found in [5.35]. The clear result is that generalized inverses are not continuous. Their regularization has been analyzed using TIKHONOV's method [5.12].

5.2.4 Stability Estimates

Here we come back to the method of Sect.5.2.1. Indeed, if we want to estimate the error on the approximate solution (5.27) it is necessary to know both constants, ε and E. Of course, it is meaningless to speak about errors in the solution without specifiying how the accuracy of the solution is defined. When F is a Hilbert space, without further specifications, then there are two natural choices. The first is to measure errors in the solutions by means of the distance $\| f^{(1)} - f^{(2)} \|_F$. The second choice is to measure errors by means of the quantity (seminorm, in mathematical language) $|(f^{(1)} - f^{(2)}, w)_F|$, where w is a suitable function of F. As we shall see, this choice is convenient for the analysis of "blurred solutions". We do not consider here the case where errors are defined in terms of the uniform norm (5.15) (for a discussion of this problem see [5.22,36]).

In the first case the error may be defined as the maximum value of $\| f - \tilde{f} \|_F$ where f is any function of the set K. Recall that $\tilde{f}$ is given by (5.27) and that K is the set of the functions satisfying (5.21,22). We shall write

$$\delta(\varepsilon, E; g) = \max_{f \in K} \| f - \tilde{f} \|_F \quad . \tag{5.49}$$

In Fig.5.2 $\delta(\varepsilon, E; g)$ is the maximum distance between any point of K and $\tilde{f}$; it is clear that this maximum value is attained at the boundary of K.

Let us now denote by $\delta_0(\varepsilon, E; g)$ and $\delta_1(\varepsilon, E; g)$ the maximum length of the semi-axes of K_0 and K_1, respectively. Then, looking at Fig.5.2, we see

$$\delta_0(\varepsilon, E; g) \leq \delta(\varepsilon, E; g) \leq \delta_1(\varepsilon, E; g) \quad . \tag{5.50}$$

The quantities $\delta_i(\varepsilon, E; g)$, $i = 0,1$, may be easily related to the data g and to the spectrum of the operator C. Assume, for simplicity, that C has a discrete spectrum and let $\gamma^2(\varepsilon/E)$ be the smallest eigenvalue of C. Then, from (5.30), it follows that

$$\delta_i(\varepsilon, E; g) = \frac{1}{\gamma(\varepsilon/E)} \left[\varepsilon_i^2 - (\| g \|_G^2 - (g, A\tilde{f})_G) \right]^{\frac{1}{2}} \quad , \quad i = 0,1 \quad , \tag{5.51}$$

where $\varepsilon_0^2 = \varepsilon^2$ and $\varepsilon_1^2 = 2\varepsilon^2$. When C has a continuous spectrum, (5.51) is still true, $\gamma^2(\varepsilon/E)$ being the infimum of the spectrum of C. Therefore $\delta_0(\varepsilon, E; g)$ and $\delta_1(\varepsilon, E; g)$ can be computed in practical cases, and they give, respectively, a lower and an upper bound for $\delta(\varepsilon, E; g)$.

Inequality (5.29) shows that there exists an upper bound for $\delta(\varepsilon, E, g)$ independent of the data function g. More precisely: $\delta(\varepsilon, E, g) \leq \delta(\varepsilon, E)$, where

$$\delta(\varepsilon, E) = \sqrt{2}\, \frac{\varepsilon}{\gamma(\varepsilon/E)} \quad . \tag{5.52}$$

The quantity $\delta(\varepsilon, E)$ is called by MILLER [5.16] *stability estimate*. Indeed, if $\delta(\varepsilon, E) \to 0$ when $\varepsilon \to 0$, for fixed E, then the error in the solution of the inverse problem also tends to zero. In other words, if the noise tends to zero and therefore $g \to \bar{g}$, where $\bar{g} = A\bar{f}$ represents noiseless data, then the estimated solution $\tilde{f}$ tends to the exact solution $\bar{f}$. This corresponds to the collapse of the ellipses K_0 and K_1 of Fig.5.2 into a point. A general condition for ensuring this is that *B has a compact inverse*. This is just a reformulation of the general result of Tikhonov (see Sect.5.2.3).

We can now state more precisely what is meant by *Hölder continuity* and *logarithmic continuity* (Sect.5.1.5). The first case corresponds to $\delta(\varepsilon, E) \sim \varepsilon^\alpha$, $0 < \alpha < 1$ (for fixed E) and the second to $\delta(\varepsilon, E) \sim |\ln \varepsilon|^{-\beta}$, $\beta > 0$. An elementary discussion of the relationship between these properties of the stability estimate and the properties of the couple of operators A, B can be done when assuming that the operators A*A and B*B commute.

Like in Sect.5.2.2, consider firstly the case where A is a compact operator. Then, from (5.52), recalling that the eigenvalues of C are given by $\gamma_n^2 = \alpha_n^2 + (\varepsilon/E)^2\beta_n^2$, it follows that

$$\delta(\varepsilon, E) = \sqrt{2}\varepsilon \sup_n \left[\alpha_n^2 + (\varepsilon/E)^2\beta_n^2\right]^{-1/2} . \tag{5.53}$$

First of all, let us observe that we cannot restore the continuity by choosing bounded β_n. Indeed if we take, for instance, $\beta_n = 1$ in (5.53) then we have, for any ε, $\delta(\varepsilon, E) = \sqrt{2}\,E$. In fact $\delta(\varepsilon, E) \to 0$ for $\varepsilon \to 0$ and fixed E, if B satisfies the following conditions: I) each eigenvalue of B*B has finite multiplicity; II) the β_n grow to infinity for $n \to +\infty$ [5.26]. These assumptions are equivalent to require that B^{-1} is a compact operator.

More precise results can be obtained if stronger assumptions are imposed on the β_n. For instance, if we assume that for $n \to +\infty$, $\beta_n \sim \alpha_n^{-\mu}$, $\mu > 0$, then, by computing the minimum of the function $\varphi(s) = s + (\varepsilon/E)^2 s^{-\mu}$, $s > 0$, it is easy to show that $\delta(\varepsilon, E) \sim E(\varepsilon/E)^{\alpha}$ where $\alpha = \mu(\mu+1)^{-1}$. However, the condition $\beta_n \sim \alpha_n^{-\mu}$ is too restrictive when the α_n tend to zero very rapidly. This occurs, for instance, for integral operators having an analytic kernel, since their singular values have an exponential fall-off [5.24]. In such a case, a more reasonable condition [5.22,26] is to take β_n growing as a power of n, i.e., $\beta_n \sim n^{\mu}$, $\mu > 0$. But, observing that the function $\varphi(s) = \exp[-as] + (\varepsilon/E)^2 s^{2\mu}$, $s > 0$, has a minimum for $s_0 \sim |\ln(\varepsilon/E)|$ when $\varepsilon \to 0$, it follows that $\delta(\varepsilon, E) \sim E|\ln(\varepsilon/E)|^{-\mu}$. In most cases the condition $\beta_n \sim n^{\mu}$ implies a constraint upon a finite number of derivatives of the admissible solutions (see Sect.5.4).

As a second example, we consider the case of the convolution operator (5.38), the constraint operator B having the form (5.40). Noting that the infimum of the spectrum of the operator C is given by

$$\gamma^2(\varepsilon/E) = \inf_{\nu} [|\hat{K}(\nu)|^2 + (\varepsilon/E)^2 |\hat{\beta}(\nu)|^2] ,$$

from (5.52) we get

$$\delta(\varepsilon, E) = \sqrt{2}\,\varepsilon \sup_{\nu} [|\hat{K}(\nu)|^2 + (\varepsilon/E)^2 |\hat{\beta}(\nu)|^2]^{-1/2} . \tag{5.54}$$

In this case, we obtain results very similar to the previous ones. Let us suppose, for instance, that $\hat{K}(\nu)$ is a rational function (this may happen in the case of an electrical network), without zeros on the real axis, whose asymptotic behavior for $|\nu| \to +\infty$ is given by $\hat{K}(\nu) \sim \nu^{-m}$ (m is a positive integer). Then we can take as constraint operator B, a differential operator of order n, with constant coefficients; in such a case $\hat{\beta}(\nu)$ is a polynomial of order n, and therefore $\hat{\beta}(\nu) \sim \nu^n$ for $|\nu| \to +\infty$. Then, asymptotically, we have $\hat{\beta}(\nu) \sim [\hat{K}(\nu)]^{-\mu}$, $\mu = n/m$, and hence we have Hölder continuity. More precisely, $\delta(\varepsilon, E) \sim E(\varepsilon/E)^{\alpha}$ with $\alpha = \mu(\mu+1)^{-1} = n(m+n)^{-1} < 1$. Observe that $\alpha \sim 1$ when $n \gg m$, i.e., when the admissible solutions are very smooth. On the other hand, if $\hat{K}(\nu)$ decreases exponentially for $|\nu| \to +\infty$ while $\hat{\beta}(\nu)$ increases as a power, then the restored continuity is only logarithmic. As is well known, the asymptotic behaviour of $\hat{K}(\nu)$ for $|\nu| \to +\infty$ is strictly related to smoothness properties of K(x). In particular, $\hat{K}(\nu)$ decreases exponentially when K(x) is an analytic function. A list of various problems which can be reduced to the solution of integral equations of convolution type can be found, for instance, in [Ref.5.4, Chap.4, Sect.3].

As pointed out in Sect.5.1.5, when the restored continuity is of logarithmic type, it is convenient to consider the reconstruction of "blurred solutions". For the sake of simplicity, we analyze essentially the case of the convolution operator (5.38). Then we can proceed as follows. Let $w_D(x)$ be a positive, even function such that

$$\int_{-\infty}^{+\infty} w_D(x)dx = 1 , \quad \int_{-\infty}^{+\infty} x^2 w_D(x)dx = D^2 \tag{5.55}$$

and let

$$f_D(x_0) = \int_{-\infty}^{+\infty} w_D(x_0 - x) f(x)dx . \tag{5.56}$$

The "blurred solution" $f_D(x_0)$ is a local weighted average of f over the distance D, and an estimate $\tilde{f}_D(x_0)$ for it is obtained by replacing f with $\tilde{f}$ in (5.56). The error in $\tilde{f}_D(x_0)$ is the maximum value of $|f_D(x_0) - \tilde{f}_D(x_0)|$ where $f_D(x_0)$ corresponds to an arbitrary function of the set K. Since we consider a convolution operator commuting with the translation operators, it is quite clear that the error in $\tilde{f}_D(x_0)$ is independent of x_0, and hence it is enough to evaluate the error in $\tilde{f}_D(0) = (\tilde{f}, w_D)_F$.

For a fixed D, we define the error by

$$\delta(\varepsilon, E; g, w_D) = \sup_{f \in K} |(f - \tilde{f}, w_D)_F| . \tag{5.57}$$

In other words, it is the maximum value of the component of $f - \tilde{f}$ along the direction of the vector w_D. If we look at Fig.5.2, we clearly understand that

$$\delta_0(\varepsilon, E; g, w_D) \leq \delta(\varepsilon, E; g, w_D) \leq \delta_1(\varepsilon, E; g, w_D) , \tag{5.58}$$

where $\delta_0(\varepsilon, E; g, w_D)$ and $\delta_1(\varepsilon, E; g, w_D)$ are quantities analogous to (5.57), the supremum being taken over the sets K_0 and K_1, respectively. Again $\delta_0(\varepsilon, E; g, w_D)$ and $\delta_1(\varepsilon, E; g, w_D)$ can be easily computed. Indeed, if we use Fig.5.2 as a schematic representation of the infinite dimensional problem, we see that the component along w_D of a vector $u = f - \tilde{f}$ of K_0 is maximal when u coincides with that point u_0 of the boundary of K_0 such that the tangent to the ellipse at u_0 is orthogonal to w_D. Now, if we write the equation of the ellipse as $(Cu, u)_F = b^2$, then the equation of the tangent in u_0 is given by $(Cu_0, u)_F = b^2$ and therefore the tangent is orthogonal to the vector Cu_0. This vector is parallel to w_D if $u_0 = aC^{-1}w_D$, where a is a constant which can be determined by requiring that u_0 belongs to the boundary of K_0. It follows that $a = b(C^{-1}w_D, w_D)_F^{-1/2}$, so that $\delta_0(\varepsilon, E; g, w_D) = |(u_0, w_D)_F| = b(C^{-1}w_D, w_D)_F^{1/2}$. Using a similar argument for K_1 and recalling (5.30), we have

$$\delta_1(\varepsilon, E; g, w_D) = [\varepsilon_1^2 - (\|g\|_G^2 - (g, A\tilde{f})_G)]^{1/2}(C^{-1}w_D, w_D)_F^{1/2} , \tag{5.59}$$

where $\varepsilon_0^2 = \varepsilon^2$ and $\varepsilon_1^2 = 2\varepsilon^2$. For infinite dimensional "ellipsoids" the previous argument can be made completely rigorous using the Schwarz inequality.

Again we can find an upper bound on the error, which is independent of g, i.e., $\delta(\varepsilon, E; g, w_D) \leq \delta(\varepsilon, E; w_D)$ where [5.16]

$$\delta(\varepsilon, E; w_D) = \sqrt{2}\,\varepsilon (C^{-1}w_D, w_D)_F^{1/2} . \tag{5.60}$$

It is possible to prove that $\delta(\varepsilon, E; w_D)$ tends to zero when $\varepsilon \to 0$, provided that the constraint operator B would have a bounded inverse. Hence in this case, we are allowed to take B = I. This type of continuity can be called *weak continuity*.

More generally, the problem of restoring "blurred solutions" can just be formulated as the problem of restoring a family of linear functionals like $(f, w_\lambda)_F$ where λ is some suitable parameter (like the center of the averaging interval). Such a point of view is usually adopted for inverse problems in geophysics [5.37,38].

5.3 Optimum Filtering

When statistical properties of data and solutions are available, filtering methods provide an alternative way for regularizing ill-posed problems.

5.3.1 Random Variables in a Hilbert Space

Since we will use a perhaps unfamiliar description of stochastic processes, we begin with a brief sketch of the working frame. For further details the reader may consult standard books on random processes like [5.39-41] or, for the more specific questions concerning Hilbert space valued random variables, [Ref.5.7, Chap.6] and the review article [5.42].

Let us first recall that a *random variable* (in short: r.v.) X is an application of some set Ω (the set of the outcomes ω of an experiment) on the set of the real numbers, i.e., $X(\omega)$ is a real number called a value of the r.v. X. A probability measure P is defined on the subsets of Ω, called events, and X is described by the distribution function

$$F_X(x) = P[X \leq x] = P\{\omega | X(\omega) \leq x\} \quad . \tag{5.61}$$

where one must read the r.h.s. as the "probability of the event containing all outcomes ω such that $X(\omega) \leq x$". The *mean value* of X will be denoted by $m_X = E[X]$, where E means *mathematical expectation*, and its *variance* by $\sigma_X^2 = E[(X - m_X)^2]$. Two (or more) r.v. are said to be jointly distributed if they are defined on the same space Ω; then they may be described by the joint distribution function

$$F_{XY}(x, y) = P[X \leq x, Y \leq y] = P\{\omega | X(\omega) \leq x, Y(\omega) \leq y\} \quad . \tag{5.62}$$

Their *covariance coefficient* is $\mu_{XY} = E[(X - m_X)(Y - m_Y)]$. The two r.v. are *uncorrelated* if $\mu_{XY} = 0$. A complex r.v. $Z = X + iY$ is viewed as a pair of jointly distributed real r.v. X, Y.

A *Hilbert space valued random variable* ξ is an application of Ω on a Hilbert space F, i.e., $\xi(\omega)$ is an element of F. When F is a space of functions, then ξ is a stochastic process. We prefer, however, to use the previous appellation, since stochastic processes are not necessarily defined in a Hilbert space. It is quite obvious that, if F is a real (complex) Hilbert space, then, for any w in F, $\xi_w = (\xi, w)_F$ is a real (complex) r.v. Accordingly, the Fourier components of ξ in a basis (u_n) of F are an infinite set of jointly distributed r.v. $\xi_n = (\xi, u_n)_F$. Simplifying somehow, we can define ξ by requiring that, for any sequence (a_n) of real numbers, the following probability makes sense:

$$P[\xi_1 \leq a_1, \xi_2 \leq a_2, \cdots, \xi_n \leq a_n, \cdots] \quad . \tag{5.63}$$

However, in order to include processes like white noise, one has to consider *weak random variables* (in short: w.r.v.). In this case it is required that (5.63) be defined only for any sequence having a finite number of elements different from $+\infty$. Such a probability measure on F is also called a *cylinder set measure*, because it is only defined on the "cylinders" of F, i.e., the sets which are bounded only along a finite number of directions [5.7,42]. Indeed, as it is the case for white noise, the probability that ξ takes values in a bounded set is not necessarily defined. Let us remark that, in writing (5.63), we have implicitly assumed that F was a real Hilbert space; the extension to complex Hilbert spaces and to corresponding w.r.v. is quite obvious. Next, let us only mention that an alternative way to the introduction of w.r.v. is to define like GEL'FAND and VILENKIN [5.43] generalized processes, i.e., applications of Ω into a space of distributions.

In the following we will assume, for simplicity, that all w.r.v. ξ have zero mean; in other words, for any w in F, $E[(\xi, w)_F] = 0$. This is not a restrictive hypothesis since, when the mean is not zero, one can always consider, instead of ξ, the reduced w.r.v. $\xi' = \xi - E[\xi]$. At this point we still have to introduce the concept of covariance operator $R_{\xi\xi}$ of the w.r.v. ξ. Such an operator is strictly related to the so-called autocovariance function of a stochastic process (for zero mean processes autocorrelation and autocovariance functions coincide). Indeed, let us assume for a moment that F is $L^2(a, b)$ and that it is possible to define in some way the complex r.v. $\xi(x)$, where x is a point of [a, b]. Then the *autocovariance function* of ξ is given by

$$R_{\xi\xi}(x, y) = E[\xi(x)\xi^*(y)] \tag{5.64}$$

and we call *covariance operator* $R_{\xi\xi}$ the integral operator whose kernel is (5.64)

$$(R_{\xi\xi} f)(x) = \int_a^b R_{\xi\xi}(x, y) f(y) dy \quad . \tag{5.65}$$

"White noise" is by definition a Gaussian process ξ for which, formally, $R_{\xi\xi}(x, y) = \varepsilon^2 \delta(x - y)$, and hence $R_{\xi\xi} = \varepsilon^2 I$, where I is the identity operator in F. From (5.64,65) and the definition of the scalar product in $L^2(a, b)$ it is easy to check that, for any f, w in F

$$(R_{\xi\xi} f, w)_F = E[(f, \xi)_F (\xi, w)_F] \quad . \tag{5.66}$$

In the theory of w.r.v. (5.66) is adopted as a definition of $R_{\xi\xi}$, a definition which

remains valid in any Hilbert space F. Indeed, we will always assume that the w.r.v. has a finite second moment, i.e., we require that $E[|(\xi, f)_F|^2]$ is finite for any f in F, and is a continuous function of f. Then the r.h.s. of (5.66) is a continuous bilinear form over F and hence there exists a bounded, linear, self-adjoint, non-negative operator $R_{\xi\xi}$ fulfilling (5.66) (see, e.g. [5.7]).

For two stochastic processes ξ, η the *cross-covariance function* is defined by

$$R_{\xi\eta}(x, y) = E[\xi(x)\eta^*(y)] \tag{5.67}$$

and the cross-covariance operator $R_{\xi\eta}$ is the integral operator whose kernel is (5.67). If ξ takes values in the Hilbert space F and η in the Hilbert space G, then $R_{\xi\eta} : G \to F$, and it is easy to check that

$$(R_{\xi\eta}g, f)_F = E[(g, \eta)_G(\xi, f)_F] \quad . \tag{5.68}$$

Equation (5.68) can be taken as a definition of the cross-covariance operator for processes having a finite second moment. Besides the following relation holds: $R_{\xi\eta} = R^*_{\eta\xi}$.

5.3.2 Best Linear Estimates

With the previous background, let us turn back to our linear inverse problem. The basic equation is (5.17) and the functions f, g, h will be considered as values of jointly distributed w.r.v., respectively ξ, η, ζ. The w.r.v. ξ takes values in the Hilbert space F, while η and ζ take values in the Hilbert space G. The w.r.v. ξ, η, ζ are assumed to satisfy the following equation

$$A\xi + \zeta = \eta \ , \tag{5.69}$$

where the linear operator $A : F \to G$ is continuous, and its inverse A^{-1} is supposed to exist. The inverse problem consists in estimating a value of ξ, given an observed value g of η. Prior knowledge would be knowledge of the joint distribution of the w.r.v. ξ and ζ (solution and noise). This is usually too much for linear estimations. It is enough to assume a knowledge of the mean values of the w.r.v. ξ, ζ and of the appropriate covariance and cross-covariance operators. The following assumptions are usually introduced:

I) ξ *and* ζ *have zero mean*;
II) ξ *and* ζ *are uncorrelated, i.e.,* $R_{\xi\zeta} = 0$;
III) $R^{-1}_{\zeta\zeta}$ *exists.*

The third assumption is the mathematical formulation of the fact that all components of the data function are affected by noise, or in other words that no component of the noise is equal to zero with probability one. Thanks to the assumptions I), II)

the covariance operator of η is given by (see, e.g. [5.20])

$$R_{\eta\eta} = AR_{\xi\xi}A^* + R_{\zeta\zeta} \tag{5.70}$$

and the cross-covariance operator $R_{\xi\eta}$ is

$$R_{\xi\eta} = R_{\xi\xi}A^* \quad . \tag{5.71}$$

We will also assume that $R_{\zeta\zeta}$ contains a parameter ε, which tends to zero when the noise vanishes, i.e.,

$$R_{\zeta\zeta} = \varepsilon^2 N \ , \tag{5.72}$$

where N is a given operator (for white noise N = I).

The classical procedure of linear mean-square estimation can now be formulated as follows. A *linear estimator* of ξ will be any w.r.v. $\tilde{\xi}_L = L\eta$ where $L : G \to F$ is an arbitrary linear continuous operator. From a value g of η one obtains then a *linear estimate* of the possible values of ξ, $\tilde{f}_L = Lg$. Now we have to find some way of evaluating the validity of such an estimator. For instance, we can measure its validity in estimating the scalar r.v. $(\xi, w)_F$ (for any given element w in F) by the mean-square error

$$\delta^2(\varepsilon; w; L) = E[|(\xi - L\eta, w)_F|^2] \quad . \tag{5.73}$$

It is then natural to ask whether there exists an operator L_0 minimizing the error (5.73). If the covariance operator $R_{\zeta\zeta}$ has a bounded inverse, L_0 exists and is unique for any w in F. It is given by

$$L_0 = R_{\xi\eta}R^{-1}_{\eta\eta} = R_{\xi\xi}A^*[AR_{\xi\xi}A^* + R_{\zeta\zeta}]^{-1} \quad . \tag{5.74}$$

The w.r.v. $\tilde{\xi} = L_0\eta$ is called the *best linear estimator* of ξ and, given a value g of η, the *best linear estimate* $\tilde{f}$ for the value of ξ is

$$\tilde{f} = R_{\xi\xi}A^*[AR_{\xi\xi}A^* + R_{\zeta\zeta}]^{-1}g \quad . \tag{5.75}$$

Let us just sketch the proof of this result (see, e.g. [Ref.5.7, Chap.6] or [5.20]). Since $R_{\zeta\zeta}$ has a bounded inverse, $R_{\eta\eta}$ has also a bounded inverse and $L_0 = R_{\xi\eta}R^{-1}_{\eta\eta}$ is a linear continuous operator from G into F. On the other hand, using (5.66,68), one can write

$$\begin{aligned} E[|(\xi - L\eta, w)_F|^2] &= ([R_{\xi\xi} - R_{\xi\eta}L^* - LR^*_{\xi\eta} + LR_{\eta\eta}L^*]w, w)_F \\ &= ([L - L_0]R_{\eta\eta}[L^* - L_0^*]w, w)_F + ([R_{\xi\xi} - L_0R_{\eta\eta}L_0^*]w, w)_F \end{aligned} \tag{5.76}$$

and, since the operator $(L - L_0)R_{\eta\eta}(L^* - L_0^*)$ is positive definite when $L \neq L_0$, the minimum is attained if and only if $L = L_0$. Let us still remark that the previous result can be extended to the case where $R_{\zeta\zeta}$ is not bounded: there exists a unique continuous operator L_0 minimizing (5.73) if and only if the operator $R_{\xi\eta}R^{-1}_{\eta\eta}$ is bounded on its domain [5.21].

We consider also the case of a w.r.v. ξ with a finite *variance* defined by

$$E\{\|\xi\|_F^2\} = E\{\sum_{n=0}^{+\infty} (u_n, \xi)_F (\xi, u_n)_F\} < +\infty \quad , \tag{5.77}$$

where $\{u_n\}$ is a basis in F [note that (5.77) does not depend upon the choice of a particular basis]. Let us remark that (5.77) can also be written with the help of the covariance operator $R_{\xi\xi}$, using (5.66)

$$E\{\|\xi\|_F^2\} = \sum_{n=0}^{+\infty} (R_{\xi\xi} u_n, u_n) = \text{Trace}\ (R_{\xi\xi}) \quad . \tag{5.78}$$

Hence we see that ξ has finite variance if and only if $R_{\xi\xi}$ has a finite trace (one says then that $R_{\xi\xi}$ is a nuclear or trace class operator). When ξ has finite variance, we may define the following "global" mean-square error (for the estimator $\hat{\xi}_L = L\eta$)

$$\delta^2(\epsilon; L) = E\{\|\xi - L\eta\|_F^2\} \quad , \tag{5.79}$$

which will be finite if and only if $L\eta$ has also a finite variance. When it exists, the operator L_0, which minimizes (5.73), minimizes also (5.79) if and only if $L_0\eta$ has a finite variance. When $R_{\xi\xi}$ has a bounded inverse, the previous condition is satisfied if the operator $L_0 = R_{\xi\eta} R_{\eta\eta}^{-1}$ is of the Schmidt class, i.e., it satisfies the condition trace $(L_0^* L_0) < +\infty$.

Now, as in Sect.5.2.3, let us briefly discuss the situations where (5.75) can be conveniently represented by means of eigenfunction (singular function) expansions. We consider first the case of a compact operator A, using the same notations as in Sect.5.2.3. We expand ξ and ζ in terms of the eigenfunctions of the operators A*A and AA*, respectively; their Fourier components are the random variables $\xi_n = (\xi, u_n)_F$, $\zeta_m = (\zeta, v_m)_G$. Then we assume [in addition to I) - III)] that

IV) *the Fourier components of ξ are mutually uncorrelated as well as the Fourier components of ζ.*

Equivalently, the following representations for $R_{\xi\xi}$ and $R_{\zeta\zeta}$ are valid:

$$R_{\xi\xi} f = \sum_{n=0}^{+\infty} \rho_n^2 f_n u_n \quad , \quad R_{\zeta\zeta} g = \epsilon^2 \sum_{m=0}^{+\infty} \nu_m^2 g_m v_m \quad , \tag{5.80}$$

where $f_n = (f, u_n)_F$, $g_m = (g, v_m)_G$, ρ_n^2 is the variance of ξ_n and $\epsilon^2\nu_m^2$ the variance of ζ_m [recall (5.72)].

Then the best linear estimate (5.75) becomes

$$\hat{f} = \sum_{n=0}^{+\infty} \frac{\alpha_n \rho_n^2}{\alpha_n^2 \rho_n^2 + \epsilon^2 \nu_n^2} g_n u_n \tag{5.81}$$

and it results that the operator L_0 is bounded if and only if $\sup(\alpha_n \rho_n^2 \nu_n^{-2}) < +\infty$ [5.21,22].

Equation (5.81) can also be written as follows:

$$\hat{f} = \sum_{n=0}^{+\infty} [1 - \exp(-2J_n)] \frac{g_n}{\alpha_n} u_n \quad , \tag{5.82}$$

where

$$J_n = 1/2 \ln\left(1 + \frac{\alpha_n^2 \rho_n^2}{\epsilon^2 \nu_n^2}\right) \quad . \tag{5.83}$$

This form is interesting because, in the case of Gaussian processes, we recognize in J_n the average mutual information contained in the scalar random variables $\xi_n = (\xi, u_n)_F$ and $\eta_n = (\eta, v_n)_G$ (see, e.g. [5.44]). Indeed, we have

$$J_n = -1/2 \ln(1 - r_n^2) \quad , \tag{5.84}$$

where r_n, given by

$$r_n^2 = \frac{\alpha_n^2 \rho_n^2}{\alpha_n^2 \rho_n^2 + \epsilon^2 \nu_n^2} = \frac{|E[\xi_n \eta_n^*]|^2}{E[|\xi_n|^2] E[|\eta_n|^2]} \tag{5.85}$$

is precisely the correlation coefficient of ξ_n and η_n. The best linear estimate $\hat{f}$ hence appears as a penalized version of the unstable formal solution $A^{-1}g = \sum_n \alpha_n^{-1} g_n u_n$, where the penalized components are those components g_n containing too little information about the components f_n.

A truncated solution, similar to (5.37), can be obtained by introducing the set $I(\epsilon)$ of the values of the index n such that $\alpha_n \rho_n \geq \epsilon \nu_n$. This condition is equivalent to require $r_n^2 \geq 1/2$ or $J_n \geq (\ln 2)/2$. If $r_n \to 0$ when $n \to +\infty$ [this condition is assured by the condition $\sup(\alpha_n \rho_n^2 / \nu_n) < +\infty$ which implies the convergence of (5.81) for any g in G], then the set $I(\epsilon)$ is finite and we can consider the finite sum

$$\sum_{n \in I(\epsilon)} \frac{g_n}{\alpha_n} u_n \quad . \tag{5.86}$$

It can be proved [5.22] that $\tilde{L}_0$ minimizes (5.73) when we consider linear estimators with only a finite number of components different from zero.

As a second example, let us consider the case where A is a convolution operator – see (5.38). Besides we assume that both ξ and ζ are *stationary processes* with auto-covariance functions $R_{\xi\xi}(x-y)$ and $R_{\zeta\zeta}(x-y)$, respectively. Let $S_{\xi\xi}(\nu)$ and $S_{\zeta\zeta}(\nu)$ be the *power spectra* (spectral density) of ξ and ζ, respectively; then (5.75) takes the usual form of a *Wiener filter* [5.40,41]

$$\hat{f}(x) = \int_{-\infty}^{+\infty} \frac{\hat{K}^*(\nu) S_{\xi\xi}(\nu)}{|\hat{K}(\nu)|^2 S_{\xi\xi}(\nu) + S_{\zeta\zeta}(\nu)} \hat{g}(\nu)\, e^{2\pi i x \nu} d\nu \quad . \tag{5.87}$$

5.3.3 Mean-Square Errors

The r.v. $(\xi - L_0 n, w)_F$ is the error we commit when taking $(L_0 n, w)_F$ as an estimator of $(\xi, w)_F$. Its variance is

$$\delta^2(\varepsilon; w) = E[|(\xi - L_0 n, w)_F|^2] \tag{5.88}$$

and therefore $\delta(\varepsilon; w)$ is the *mean-square error* in the estimation of $(\xi, w)_F$. The parameter ε is defined by (5.72). For simplicity let us consider only the case where $R_{\xi\xi}$ has a bounded inverse. Then the optimum filter L_0 certainly exists and, from (5.74,76) it follows that

$$\delta(\varepsilon; w) = ([R_{\xi\xi} - L_0 R_{nn} L_0^*]w, w)_F^{1/2} \quad . \tag{5.89}$$

Furthermore, it is possible to prove that, when both inverse operators $R_{\xi\xi}^{-1}$ and A^{-1} exist, then $\delta(\varepsilon; w) \to 0$ when $\varepsilon \to 0$, for any w in F [5.21].

It is now natural to define a *relative mean-square error* as being the ratio between the variance of the error and the variance of the estimated r.v. $(\xi, w)_F$ [5.20]. Since $E[|(\xi, w)_F|^2] = (R_{\xi\xi} w, w)_F$, we get from (5.89)

$$\delta_{rel}(\varepsilon; w) = \frac{([R_{\xi\xi} - L_0 R_{nn} L_0^*]w, w)_F^{1/2}}{(R_{\xi\xi} w, w)_F^{1/2}} \quad . \tag{5.90}$$

This quantity gives a precise measure of the *reliability* of the estimate. It is interesting to remark [5.20] that, when A^{-1} is not continuous, one can find a sequence $\{w^{(n)}\}$ such that $\delta_{rel}(\varepsilon, w^{(n)}) \to 1$ when $n \to +\infty$, for fixed ε. In other words, in the case of an ill-posed problem, for any value of $\varepsilon > 0$, there will be vectors w in F for which the r.v. $(\xi, w)_F$ cannot be reliably estimated.

When $R_{\xi\xi}$ is of the trace class, i.e., ξ has a finite variance - see (5.78) - and $R_{\zeta\zeta}$ has a bounded inverse, then the optimum filter L_0 is of the Schmidt class and one can define a "global" mean-square error as $\delta(\varepsilon) = \delta(\varepsilon; L_0)$ - see (5.79). An expression for $\delta(\varepsilon)$, similar to (5.89), can be derived remarking that $\delta^2(\varepsilon)$ is the trace of the covariance operator of $\xi - L_0 n$

$$\delta(\varepsilon) = [\mathrm{Trace}(R_{\xi\xi} - L_0 R_{nn} L_0^*)]^{1/2} \quad . \tag{5.91}$$

When the inverse operators $R_{\xi\xi}^{-1}$ and A^{-1} both exist, then one can prove that $\delta(\varepsilon) \to 0$, when $\varepsilon \to 0$ [5.21].

In the case of a compact operator A, when assumption IV) of Sect.5.3.2 is satisfied, (5.89) becomes

$$\delta(\varepsilon; w) = \varepsilon \left(\sum_{n=0}^{+\infty} \frac{\rho_n^2 \nu_n^2}{\alpha_n^2 \rho_n^2 + \varepsilon^2 \nu_n^2} |w_n|^2 \right)^{1/2} , \tag{5.92}$$

where $w_n = (w, u_n)_F$. It is quite easy to show that $\delta(\varepsilon; w) \to 0$ when $\varepsilon \to 0$ [5.21], under the sole condition that the operators $R_{\xi\xi}^{-1}$ and $R_{\zeta\zeta}^{-1}$ exist [5.21,22].

As regards the "global" mean-square error (5.91), it becomes

$$\delta(\varepsilon) = \varepsilon \left(\sum_{n=0}^{+\infty} \frac{\rho_n^2 \nu_n^2}{\alpha_n^2 \rho_n^2 + \varepsilon^2 \nu_n^2} \right)^{1/2} \quad . \tag{5.93}$$

It is also easy to show that $\delta(\varepsilon) \to 0$, when $\varepsilon \to 0$, provided that ξ has a finite variance, i.e., $\sum_n \rho_n^2 < +\infty$ [5.21].

In the case of a convolution operator A and of stationary processes ξ, ζ, one can only define the mean-square error (5.89). Indeed, the covariance operator of a stationary process is never of the trace class. An expression for $\delta(\varepsilon; w)$ can be easily derived using (5.87,89).

5.3.4 Comparison with Miller's Regularization Method

In Miller's method, the estimates of the solution of the problem (5.17) have to belong to the set K defined by (5.21,22) and this is a "rigid" condition, in the sense that all the functions outside K are rejected as meaningless. In probabilistic methods, the restrictions are less categorical since one considers the probability distributions of the solutions and of the errors. In fact, the knowledge of $R_{\zeta\zeta}$ corresponds to the bound (5.21) for the error, while the knowledge of $R_{\xi\xi}$ corresponds to the bound (5.22) for the solution. Moreover, both Miller's method and optimum filtering are least square methods. Hence, it is not surprising to find similarities between the solutions provided by the two methods. Indeed, thanks to the following identity, valid when $R_{\xi\xi}^{-1}$ and $R_{\zeta\zeta}^{-1}$ exist

$$(A^* R_{\zeta\zeta}^{-1} A + R_{\xi\xi}^{-1}) R_{\xi\xi} A^* = A^* R_{\zeta\zeta}^{-1} (A R_{\xi\xi} A^* + R_{\zeta\zeta}) \tag{5.94}$$

it is easy to show that (5.27) and (5.75) coincide formally when putting

$$R_{\zeta\zeta} = \varepsilon^2 I \quad , \quad R_{\xi\xi} = E^2 (B^* B)^{-1} \quad . \tag{5.95}$$

In fact, the condition introduced by Miller in order to restore continuous dependence on the data (i.e., that the constraint operator B should have a bounded inverse) corresponds to the condition that the w.r.v. ξ has finite second moment, i.e., there exists a bounded operator $R_{\xi\xi}$ defined by (5.66). Moreover, the mean-square error (5.89) can be considered as the analogue of the stability estimate (5.60) for the restoration of "blurred solutions". Thanks to the identification (5.95), they coincide up to a factor $\sqrt{2}$. Looking at the relative error (5.90), we are tempted to define its fellow in regularization theory by

$$\delta_{rel}(\varepsilon, E; w) = \frac{\varepsilon}{E} \frac{(C^{-1}w,w)_F^{\frac{1}{2}}}{([B^*B]^{-1}w,w)_F^{\frac{1}{2}}} \quad . \tag{5.96}$$

In fact, it is also possible to derive this formula in an intrinsic way, without reference to its probabilistic analogue [5.26]. In spite of different starting points, the similarities of the solutions and of the error estimates are very interesting and therefore the conjunction of both points of view can provide complementary insights on the regularization of linear inverse problems.

For the reader's convenience we summarize the main analogies between Miller's regularization method and optimum filtering in the following scheme.

	Regularization method
Data	the function g; $g = Af + h$
Prior knowledge	$\|h\|_G = \|Af - g\|_G \leq \varepsilon$, $\|Bf\|_F \leq E$; knowledge of ε, E and of the operator B
Requirement	an estimate of the function f
Solution	$\bar{f} = [A^*A + (\frac{\varepsilon}{E})^2 B^*B]^{-1} A^* g$

	Optimum filtering
Data	a value g of the r.v. $\eta = A\xi + \zeta$
Prior knowledge	ξ, ζ are zero mean, uncorrelated r.v.; knowledge of the covariance operators $R_{\xi\xi}$, $R_{\zeta\zeta}$
Requirement	an estimate of a value of ξ
Solution	$\bar{f} = R_{\xi\xi}A^*[AR_{\xi\xi}A^* + R_{\zeta\zeta}]^{-1} g$

5.4 Linear Inverse Problems in Optics

Surveys of inverse problems in optics and electromagnetics can be found in [5.9,45]. Due to the rapid growth of research in this field, we do not attempt a complete review. Our aim is to focus on stability problems and therefore we select only a few examples, using for simplicity the scalar theory of light. The harmonic time dependence $\exp(-i\omega t)$ is assumed, and by wave functions we mean scalar complex amplitudes.

5.4.1 Inverse Problems in Fourier Optics

The problem of restoring data that have been degraded by a linear bandlimited system has for long received much attention both in optics [5.46] and in radio astronomy [5.47,48]. For simplicity we will consider only a one-dimensional system. Then, in the absence of noise, such a system is described by a linear equation like

$$\int_{-X/2}^{X/2} S(x - y)\bar{f}(y)dy = \bar{g}(x) \ , \tag{5.97}$$

where S(x), the *point spread function*, has a Fourier transform which vanishes outside a finite interval $[-\Omega/2, \Omega/2]$, $\bar{f}$ is the wave function in the object plane and $\bar{g}$ the wave function in the image plane. We call $\bar{f}$ the object and $\bar{g}$ the noiseless image.

Since $\bar{f}$ is zero outside the interval $[-X/2, X/2]$, its Fourier transform is an entire analytic function. So it was observed [5.46,48] that analytic continuation in the frequency domain will in principle allow for restoration of unlimited details of $\bar{f}$. As remarked by many authors [5.49,50], this result seems to be in contradiction with the concept of *number of degrees of freedom of an image* [5.51-53], which essentially means that the image never contains enough information to reconstruct the object unambiguously. The contradiction disappears if one takes into account the noise and *logarithmic continuity*, which arise for object restoration.

Prolate Spheroidal Wave Functions (PSWF)

We summarize here the main properties of the prolate spheroidal wave functions $\psi_n(c, x)$ [5.54-56], which are a fundamental tool for the analysis of bandlimited systems. The $\psi_n(c, x)$ can be defined as the continuous solutions, on the closed interval [-1, 1], of the differential equation

$$-[(1 - x^2)\psi'(x)]' + c^2x^2\psi(x) = \chi\psi(x) \quad . \tag{5.98}$$

Continuous solutions exist only for certain discrete positive values χ_n of the parameter $\chi : 0 < \chi_0 < \chi_1 < \ldots$. Then $\psi_n(c, x)$ is just the solution of (5.98) corresponding to the eigenvalue χ_n. The behavior of χ_n when $n \to +\infty$ is [5.57]

$$\chi_n = n(n + 1) + \frac{1}{2}c^2 + O\left(\frac{1}{n^2}\right) \quad . \tag{5.99}$$

The $\psi_n(c, x)$ can be uniquely extended to entire analytic functions, and they will be normalized as follows:

$$\int_{-\infty}^{+\infty} |\psi_n(c, x)|^2 dx = 1 \ ; \quad n = 0,1,2, \ldots \quad . \tag{5.100}$$

The PSWF are also solutions of the eigenvalue equation

$$\int_{-1}^{1} \frac{\sin[c(x-y)]}{\pi(x-y)} \psi_n(c, y)dy = \lambda_n \psi_n(c, x) \quad . \tag{5.101}$$

The eigenvalues λ_n form a decreasing sequence: $1 > \lambda_0 > \lambda_1 > \ldots > 0$ and have a step behavior: they are approximately equal to one for values of the index less than $N_0 = 2c/\pi$ and then fall off to zero exponentially. More precisely, their behavior for $n \to +\infty$ is [5.58]

$$\lambda_n = 0(\frac{1}{n} \exp[-2n \ln(\frac{n}{ec})]) \quad . \tag{5.102}$$

The eigenvalues λ_n are also the normalization constants of the PSWF on the interval [-1, 1]

$$\int_{-1}^{1} |\psi_n (c, x)|^2 dx = \lambda_n \quad . \tag{5.103}$$

The fundamental properties of the PSWF are:

a) The $\psi_n(c, x)$ are bandlimited functions; their Fourier transforms vanish outside the interval $[-c/2\pi, c/2\pi]$

$$\int_{-\infty}^{+\infty} e^{-2\pi i \nu x} \psi_n(c, x)dx = (-i)^n \sqrt{\frac{2\pi}{c\lambda_n}} \psi_n(c, \frac{2\pi\nu}{c}) 0(\frac{2\pi\nu}{c}) \quad , \tag{5.104}$$

where $0(s) = 1$ for $|s| < 1$ and $0(s) = 0$ for $|s| > 1$ (see, e.g. [5.59]).

b) The $\psi_n(c, x)$ are a basis in the space of the square-integrable bandlimited functions.

c) The functions $u_n(x) = \lambda_n^{-\frac{1}{2}} \psi_n(c, x)$ are a basis in $L^2(-1, 1)$.

Statements b) and c) exhibit a remarkable property of the PSWF: they are orthogonal over two different intervals. This property is fundamental for the extrapolation of bandlimited functions.

Perfect Lowpass Filter

We consider first (5.97) with the point spread function $S(x) = (\pi x)^{-1} \sin(\pi\Omega x)$ (perfect lowpass filter through the band $[-\Omega/2, \Omega/2]$). The connection with the general formulation of a linear inverse problem, given in Sect.5.1.3, is as follows. Since the object radiates a finite power, we can take $L^2(-X/2, X/2)$ as solution space F. Assuming that the noisy image g is known only on the interval $[-X/2, X/2]$, we can also take $L^2(-X/2, X/2)$ as data space G. Then, object restoration consists in inverting the integral operator

$$(Af)(x) = \int_{-X/2}^{X/2} \frac{\sin[\pi\Omega(x-y)]}{\pi(x-y)} f(y)dy \quad . \tag{5.105}$$

The operator A is self-adjoint, nonnegative and compact. The quantity $R = \Omega^{-1}$ is the *Rayleigh resolution distance* and $N_0 = \Omega X$ is the *number of degrees of freedom* of the image. Observe that N_0 is the number of eigenvalues of A which are approximately equal to one and also that $N_0 = \text{Trace}(A)$ [5.60]. In fact the eigenvalues of A are the λ_n associated to the PSWF with $c = \pi\Omega X/2$ and the corresponding eigenfunctions are

$$u_n(x) = (\frac{2}{X\lambda_n})^{\frac{1}{2}} \psi_n(c, \frac{2x}{X}) \quad , \quad c = \pi\Omega X/2 \quad . \tag{5.106}$$

In order to apply to this problem the general results of Sect.5.2.2, we need a constraint operator B commuting with A. This requirement is satisfied by the differential operator

$$(B^*Bf)(x) = -[(\tfrac{1}{4}X^2 - x^2)f'(x)]' + c^2x^2f(x) \tag{5.107}$$

since, from the definition of the PSWF, it follows that the u_n, defined by (5.106), are the eigenfunctions of B^*B and the χ_n are the corresponding eigenvalues. Furthermore, by means of a partial integration one gets

$$(B^*Bf, f)_F = \int_{-X/2}^{X/2} (\tfrac{1}{4}X^2 - x^2)|f'(x)|^2dx + c^2 \int_{-X/2}^{X/2} x^2|f(x)|^2dx \tag{5.108}$$

so that condition (5.22) is a constraint on the first derivative of f. Note that (5.108) has the same form as (5.44); however, the functions p(x) and q(x) in (5.108) are not strictly positive. Hence the set defined by (5.22,108) is not compact with respect to the uniform norm (5.15) (see [Ref.5.36, p.195]) but it is compact with respect to the L^2-norm.

Now the restored object is given by (5.36) with $\alpha_n = \lambda_n$, $\beta_n = \sqrt{\chi_n}$ and

$$g_n = \int_{-X/2}^{X/2} g(x)u_n(x)dx \tag{5.109}$$

or by the truncated solution (5.37). It is interesting to remark that, since the λ_n decrease exponentially fast when $n > N_0$, while $\beta_n \sim n$, the number of terms N in (5.37) is equal to N_0 (the number of degrees of freedom) plus a number of terms which is roughly proportional to $|\ln \varepsilon|$, which number is therefore rather insensitive to the noise. This fact is strictly related to logarithmic continuity [5.61]. Indeed, if we consider the stability estimate (5.53), with $\alpha_n = \lambda_n$ and $\beta_n = \sqrt{\chi_n}$, from the behavior (5.102) and (5.99), we conclude that $\delta(\varepsilon, E) \sim E|\ln(\varepsilon/E)|^{-1}$. This result can be extended to the case where a finite number of derivatives of f are bounded [5.22, 36]. The following statements are justified.

I) The error on the restored object tends to zero when $\varepsilon \to 0$ and therefore, *in principle*, unlimited resolution of details is possible (at least in the framework

of classical optics, since the previous analysis does not take into account the quantum-mechanical limitations on measurement of the light field [5.62]).

II) When the object is not extremely smooth (what is equivalent to say that its Fourier transform is not negligible outside the band $[-\Omega/2, \Omega/2]$ – see [5.22,61]), the error on the restored object tends to zero so slowly that, *in practice*, resolution beyond the limit corresponding to the number of degrees of freedom becomes impossible. This conclusion agrees with the results of earlier analysis of object restoration [5.63,64].

More precise results about resolution can be obtained by considering the restoration of "blurred solutions" as sketched in Sect.5.2.4. The "blurred object" is given by (5.56) (the integration ranging now over $[-X/2, X/2]$) at least when $D \ll X$ and x_0 is sufficiently far from the borders. Then the error in the restoration of $f_D(0) = (f, w_D)_F$ is an estimate of the error we commit in the restoration of details whose size is D. One should expect a trade-off between resolution and error [5.63]: the restoration error has to be greater for smaller values of D.

In order to analyze the effect on resolution of different types of noise [5.63], we focus on optimum filtering methods. The relative error in the restoration of $f_D(0)$ is given by (5.90) with $w = w_D$. Let us assume, for simplicity, that the stochastic processes describing object and noise satisfy the assumptions I) - IV) of Sect.5.3.2. Besides we assume that the variances of the Fourier components of the object [with respect to the basis (5.106)] are constant, i.e., $\rho_n^2 = E^2$. This assumption is reasonable if the correlation distance δ for the stochastic process representing the object (δ gives the size of the finest details that should be resolved) is much smaller than the Rayleigh distance R. Then, from (5.90,92) we get

$$\delta_{rel}(\varepsilon/E; w_D) = \left(1 - \frac{1}{\|w_D\|^2} \sum_{n=0}^{+\infty} \frac{\lambda_n^2}{\lambda_n^2 + (\varepsilon/E)^2 \nu_n^2} |w_{D,n}|^2\right)^{1/2}, \qquad (5.110)$$

where $\|w_D\|$ is the norm of w_D in $L^2(-X/2, X/2)$ and the $w_{D,n}$ are the Fourier coefficients of w_D in the basis (5.106). We consider two types of noise in the image plane [5.63]: *white measurement noise*, i.e., $\nu_n^2 = 1$, and *band-limited measurement noise*, i.e., $\nu_n^2 = \lambda_n$. In both cases it is easy to show [5.26] that $\delta_{rel}(\varepsilon/E; w_D) \to 1$ (100% error) when $D \to 0$, i.e., when w_D tends to the Dirac delta measure.

In Fig.5.4 we give the results of numerical computations (the numerical method is described in [5.26]) for $c = 10$, $X = 2$ and $w_D(x) = N\theta(x/d)\,\mathrm{sinc}^2(x/d)$ [N is a normalization constant, $\theta(s)$ is the characteristic function of the interval $[-1, 1]$ and d is a parameter related to D through (5.55)]. In the case of white measurement noise, the curves are rapidly decreasing up to a value of D/R of about 0.5 and then become rather flat. Besides a lowering of ε from 10^{-2} to 10^{-6} does not modify the situation in a significant way. If we accept only an error of a few percent, then it is difficul to get a resolution better than R. One expects that superresolution should become even more difficult for greater values of c [Ref.5.65, p.470]. Figure 5.4 also shows that a smoothing of the noise (band-limited measurement noise) [5.63] is equivalent to a lowering of the white noise from 10^{-2} to 10^{-6}.

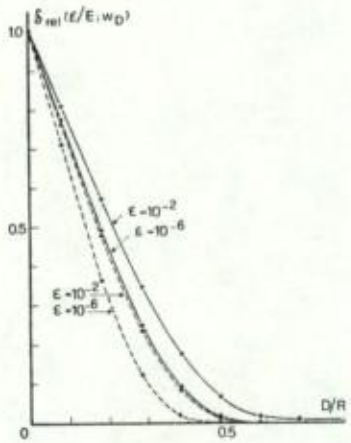

Fig. 5.4. Relative errors vs the resolution parameter D/R, where $R = \pi X/2c$ ($X = 2$, $c = 10$). Undotted curves correspond to white measurement noise and dotted curves to band-limited measurement noise. In both cases $E = 1$

The case of *incoherent illumination* in the object plane [point spread function $S(x) = \Omega(\pi\Omega x)^{-2} \sin^2(\pi\Omega x)$, frequency band $[-\Omega, \Omega]$] has been analyzed by many authors [5.66-68]. In this case, as well as for any bandlimited system, the restored continuity is also logarithmic. Indeed, thanks to the Paley-Wiener theorem [5.69], the point spread function of a bandlimited system, with bandwidth Ω, is an entire analytic function of order ρ, $0 < \rho \leq 1$, and type $\tau \leq \pi\Omega$. Then, it follows from general results of HILLE and TAMARKIN [5.24] that the singular values of the integral operator defined by (5.97) have the behavior $\alpha_n \sim C \exp(-Dn \ln n)$, $n \to +\infty$, where C, D are suitable constants. As we remarked in Sect. 5.2.4, this behavior implies logarithmic continuity. Therefore it should be possible to define, for any bandlimited system, a resolution limit and a number of degrees of freedom, practically noise independent. For instance, in the case of incoherent illumination, this number has been estimated to be of the order of $2\Omega X$ (the bandwidth is 2Ω), corresponding to a resolution limit of about R/2 [5.68]. This result agrees with an analysis of the restoration of "blurred objects" [5.26]. However it must be remarked that, in the case of incoherent illumination the solutions have to satisfy a positivity constraint which could improve the resolution limit (for a brief discussion of this point see Sect.5.4.6).

Bandwidth Extrapolation

A method which has been proposed for the analysis of an arbitrary bandlimited system is the following [5.46,48,70-72]: by Fourier transforming the image one can obtain the Fourier transform of the object over the band $[-\Omega/2, \Omega/2]$ of the system; since this Fourier transform is an entire analytic function, by analytic continuation one could restore its values for all frequencies and therefore also restore the object. We remark that this problem is mathematically equivalent to the extrapolation of optical images beyond borders [5.59,73].

Let f be the object and $\hat{f}$ its Fourier transform

$$\hat{f}(\nu) = \int_{-X/2}^{X/2} e^{-2\pi i\nu x} f(x)dx \quad . \tag{5.111}$$

The data $\hat{g}$ are the values (affected by errors of $\hat{f}$ on the band $[-\Omega/2, \Omega/2]$ and we can take $L^2(-\Omega/2, \Omega/2)$ as data space G. The solution space F is the set of the functions $\hat{f}$ like (5.111), normed with the norm of $L^2(-\infty, +\infty)$. Then F is a Hilbert space of analytic functions. The direct problem is merely the restriction of $\hat{f}(\nu)$ to $[-\Omega/2, \Omega/2]$, i.e., $(A\hat{f})(\nu) = \hat{f}(\nu)$ when $|\nu| < \Omega/2$, $(A\hat{f})(\nu) = 0$ elsewhere. Then $A^* : G \to F$ is the integral operator [5.22]

$$(A^*\hat{g})(\nu) = \int_{-\Omega/2}^{\Omega/2} \frac{\sin[\pi X(\nu-\nu')]}{\pi(\nu-\nu')} \hat{g}(\nu')d\nu' \quad . \tag{5.112}$$

Observing that AA^* coincides with the integral operator (5.105), except for an exchange of Ω and X, one can write for these operators A, A^* equations like (5.32) where $\alpha_n = \sqrt{\lambda_n}$ (the eigenvalues of the PSWF) and u_n, v_n are replaced by ($c = \pi\Omega X/2$)

$$\hat{u}_n(\nu) = \sqrt{\frac{2}{\Omega}}\, \psi_n(c, \tfrac{2\nu}{\Omega}) \quad , \quad \hat{v}_n(\nu) = \sqrt{\frac{2}{\Omega\lambda_n}}\, \psi_n(c, \tfrac{2\nu}{\Omega})\theta(\tfrac{2\nu}{\Omega}) \quad . \tag{5.113}$$

From properties II), III) of the PSWF it follows that $(\hat{u}_n)$ is a basis in F while $(\hat{v}_n)$ is a basis in G. Remark that the expansion of $\hat{f}(\nu)$ as a series of the $\hat{u}_n$ is equivalent to the expansion of f(x), in (5.111), as a series of the u_n given in (5.106). Indeed $i^n u_n$ and $\hat{u}_n$ are related by (5.104).

BUCK and GUSTINCIC [5.70] assumed that the stochastic processes, representing the object and the noise, satisfy conditions I) - IV) of Sect. 5.3.2 and that $\rho_n^2 = E^2$ (where the ρ_n^2 are the variances of the components of the object in the basis $\{u_n\}$). In the case of white noise, their solution to the problem of analytic continuation is given by (5.85) with $\alpha_n = \sqrt{\lambda_n}$, $\rho_n^2 = E^2$, $\nu_n^2 = 1$, $g_n = (\hat{g}, \hat{v}_n)_G$ and u_n replaced by $\hat{u}_n$ (we denote this estimate by $\tilde{f}$). It has been remarked by these authors that an increase by a factor of 10 in the signal-to-noise ratio E/ε adds only one more significant term in the series (5.85) for $\tilde{f}$ (and therefore, for large apertures, the improvement in resolution is negligible). The same estimate f was obtained by VIANO [5.73] in the framework of regularization theory, considering the constraint operator B = I [compare with (5.36) where $\alpha_n = \sqrt{\lambda_n}$, $\beta_n = 1$, $g_n = (\hat{g}, \hat{v}_n)_G$ and u_n replaced by $\hat{u}_n$]. This assumption is equivalent to requiring that the object radiates a finite power. VIANO [5.73] proved that such an estimate converges to the "true solution", when $\varepsilon \to 0$, uniformly over any finite interval containing the band $[-\Omega/2, \Omega/2]$. In order to have stability with respect to the norm of F, stronger conditions on the objects are required. If we introduce, for instance, the constraint operator (5.107), then it is easy to prove (as in the case of the perfect lowpass filter) that the analytic continuation of the Fourier transform of the object outside the band $[-\Omega/2, \Omega/2]$ is stable with respect to the norm of F, but that we get only logarithmic continuity. This result agrees with the conclusions of [5.70].

5.4.2 Inverse Diffraction

According to SHERMAN [5.76] and SHEWELL and WOLF [5.74], inverse diffraction can be defined as the problem of determining the field distribution on a boundary surface from the knowledge of the distribution on a surface situated within the domain where the wave propagates. An extensive treatment of uniqueness in inverse diffraction is given by HOENDERS [5.75] both in the scalar and in the vector case.

The fundamental reason of the instability of inverse diffraction is that space acts like a filter for the higher modes. For instance, in the scattering of a plane wave, with wave number k, by a body whose largest dimension is R, only kR modes are propagated up to the far zone, while the others are attenuated.

Inverse Diffraction from Plane to Plane

The direct problem is to determine a wave function u, solution of Helmholtz equation in the half-space $z \geq z_0$

$$\nabla^2 u + k^2 u = 0 \tag{5.114}$$

satisfying Sommerfeld's condition at infinity and the condition $u = u_0$ (u_0 being a given function) on the plane $z = z_0$. This solution can be most conveniently expressed in terms of Fourier transforms [5.74,76]. If we write

$$\tilde{u}(p, q; z) = \int_{-\infty}^{+\infty}\int e^{-ik(px + qy)} u(x, y, z)dx\, dy \quad , \tag{5.115}$$

then

$$\tilde{u}(p, q; z) = \exp[ikm(z - z_0)]\tilde{u}(p, q; z_0) \quad , \tag{5.116}$$

where

$$m = (1 - p^2 - q^2)^{1/2} \quad , \quad \mathrm{Im}(m) \geq 0 \quad . \tag{5.117}$$

The inverse problem is the following: given the values g (affected by errors) of the wave function u on the plane $z = z_1 > z_0$, estimate the values of u on any z plane between z_0 and z_1. If the radiating power is finite, then u is square integrable over any z plane and therefore we can take $L^2(R^2)$ both as solution and as data space. Writing $\bar{f}(x, y) = u(x, y, z)$, $\bar{g}(x, y) = u(x, y, z_1)$ ($\bar{g}$ is the noiseless wave function), from (5.116) we derive that $\bar{\tilde{g}} = A\tilde{\bar{f}}$, where

$$(A\bar{f})\tilde{}(p, q) = \exp[+ im(z_1 - z)]\tilde{f}(p, q) \quad . \tag{5.118}$$

Instability is due to the effect of inhomogeneous (evanescent) waves ($p^2 + q^2 > 1$). Assuming that f is generated by a field distribution on the plane $z = z_0$ (with an L^2-norm bounded by E^2), we get a class of admissible solutions defined by the constraint operator

$$(Bf)\tilde{}(p, q) = \exp[-im(z - z_0)]\tilde{f}(p, q) \quad . \tag{5.119}$$

The operators A, B have the form (5.38) and (5.40), respectively, with $\tilde{K}(p, q) = \exp[im(z_1 - z)]$ and $\tilde{s}(p, q) = \exp[-im(z - z_0)]$ so that the regularized solution is given by (5.41) or (5.42). Besides, observing that $\tilde{s}(p, q) = [\tilde{K}(p, q)]^{-\mu}$, $\mu = (z - z_0)/(z_1 - z)$, from (5.54) $\delta(\varepsilon, E) \sim E(\varepsilon/E)^{\alpha}$ follows, where $\alpha = (z - z_0)/(z_1 - z_0)$, $0 < \alpha < 1$.

This result is very similar to "three line theorems" derived by MILLER [5.77] in the case of the backward heat equation and the Cauchy problem for the Laplace equation. As we see, we get Hölder continuity if $z > z_0$, while we do not have stability for $z = z_0$ (in this case B = I). To restore the stability even there, we have to take stronger constraints. If we assume, for instance, that the wave function on the plane $z = z_0$ has also square integrable first derivatives, then we have stability up to the plane $z = z_0$. In this case, however, the restored continuity is only logarithmic (see Sect.5.2.4). Finally, if the wave function on the plane $z = z_0$ is assumed to contain only spatial frequencies below the wave number k, i.e., $\hat{u}(p, q; z_0) = 0$ if $p^2 + q^2 > 1$, then a well-behaved inversion formula can be derived [5.74]. In other words, inverse diffraction can be formulated as well-posed problem when the effect of evanescent waves can be disregarded. It has recently been shown [5.117-122] that the total field due to the inhomogeneous waves does not decay exponentially with distance z, but much slower ($z^{-3/2}$ or z^{-2}). In view of these results, the inverse diffraction problem seems to deserve reconsideration.

Inverse Diffraction for Cylindrical Waves

We consider a wave function $u = u(\rho, \varphi)$ (ρ, φ are circular cylinder coordinates), solution of (5.114), satisfying Sommerfeld's radiation condition at infinity and the condition $u = u_0$ on the circular cylinder of radius ρ_0. The solution of this problem (direct problem) is represented by the Fourier series

$$u(\rho, \varphi) = \sum_{n=-\infty}^{+\infty} \frac{H_n^{(1)}(k\rho)}{H_n^{(1)}(k\rho_0)} c_n e^{in\varphi} , \tag{5.120}$$

where the $H_n^{(1)}$ are the Hankel functions of the first kind and the c_n are the Fourier coefficients of $u_0(\varphi) = u(\rho_0, \varphi)$

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} u(\rho_0, \varphi) e^{-in\varphi} d\varphi . \tag{5.121}$$

The inverse problem is as follows: given the values g (affected by errors) of u on the cylinder of radius $\rho_1 > \rho_0$, estimate the wave function over any cylinder of radius ρ, $\rho_0 \leq \rho \leq \rho_1$. If we denote by $\bar{g}(\varphi) = u(\rho_1, \varphi)$ the noiseless data and by $\bar{f}(\varphi) = u(\rho, \varphi)$ ($\rho < \rho_1$) the unknown solution, then, from (5.120), $\bar{g} = A\bar{f}$ where

$$(A\bar{f})(\varphi) = \sum_{n=-\infty}^{+\infty} \frac{H_n^{(1)}(k\rho_1)}{H_n^{(1)}(k\rho)} \bar{f}_n e^{in\varphi} , \tag{5.122}$$

and the $\bar{f}_n = c_n H_n^{(1)}(k\rho)/H_n^{(1)}(k\rho_0)$ are the Fourier coefficients of $\bar{f}$. We take $L^2(-\pi, \pi)$ both as solution and data space. Then, if $\bar{f}(\varphi) = u(\rho, \varphi)$ is generated by a wave function on the cylinder of radius ρ_0, the class of admissible solutions on the cylinder of radius ρ is characterized by the constraint operator

$$(B\bar{f})(\varphi) = \sum_{n=-\infty}^{+\infty} \frac{H_n^{(1)}(k\rho_0)}{H_n^{(1)}(k\rho)} \bar{f}_n e^{in\varphi} , \tag{5.123}$$

which is derived from (5.120) and the equality $(B\bar{f})(\varphi) = u(\rho_0, \varphi)$. If we write $\alpha_n = |H_n^{(1)}(k\rho_1)/H_n^{(1)}(k\rho)|$, $\beta_n = |H_n^{(1)}(k\rho_0)/H_n^{(1)}(k\rho)|$, then the estimated wave function is given by (5.36) or (5.37) (remark that now the index n takes values from $-\infty$ to $+\infty$). From the behavior of Hankel functions when $|n| \to +\infty$, it follows that $\alpha_n \sim \exp[-|n| \ln(\rho_1/\rho)]$ while $\beta_n \sim \exp[|n| \ln(\rho/\rho_0)]$. These behaviors imply $\beta_n \sim \alpha_n^{-\mu}$ with $\mu = \ln(\rho/\rho_0)/\ln(\rho_1/\rho)$ and from (5.53) one has $\delta(\varepsilon, E) \sim E(\varepsilon/E)^{\alpha}$ with $\alpha = \ln(\rho/\rho_0)/\ln(\rho_1/\rho_0)$ (for a more precise estimate see [Ref.5.78, Chap.3]). It is interesting to compare this result with the stability estimate for analytic continuation implied by Hadamard's "three circle theorem" (Sect.5.1.5). A similar result holds for harmonic continuation in a disc [5.77].

The constraint operator (5.123) does not imply stability in the L^2-norm when $\rho = \rho_0$ (in this case B = I). Then we have stability if, for instance,

$$\|B\bar{f}\|^2 = \left\|\frac{\partial \bar{f}}{\partial \varphi}\right\|^2 + \|\bar{f}\|^2 = \sum_{n=-\infty}^{+\infty} (n^2 + 1)|\bar{f}_n|^2 \leq E^2 . \tag{5.124}$$

Let us remark that, if the wave function u represents a z-polarized electric field, i.e., $E_\rho = E_\varphi = 0$, $E_z = u$, then the components H_ρ, H_φ of the magnetic field are proportional to $\rho^{-1}\partial u/\partial \varphi$ and $\partial u/\partial \rho$, respectively. Therefore, in this case, (5.124) implies a bound on both E_z and H_ρ. However this constraint gives only logarithmic continuity, since $\beta_n \sim n$ while α_n tends to zero exponentially (Sect.5.2.4).

The previous results can be easily extended to the case of spherical surfaces, using the expansion of the wave function as a series of spherical harmonics $Y_n^m(\theta, \varphi)$.

Inverse Diffraction from Far-Field Data

For simplicity we consider again cylindrical waves. Then the wave function (5.120) has the behavior

$$u(\rho, \varphi) \sim \sqrt{\frac{2}{i\pi k\rho}} e^{ik\rho} \bar{g}(\varphi), \quad \rho \to +\infty , \tag{5.125}$$

where $\bar{g}$ is the *scattering amplitude* (or *pattern function*)

$$\bar{g}(\varphi) = \sum_{n=-\infty}^{+\infty} \frac{(-i)^n}{H_n^{(1)}(k\rho_0)} c_n e^{in\varphi} . \tag{5.126}$$

The direct problem is to compute $\bar{g}$, given the wave function u_0 on the cylinder of radius ρ_0. The inverse problem is to estimate u on any cylinder of radius ρ, $\rho_0 \leq \rho < \infty$, given a noisy scattering amplitude g. If we write $\bar{f}(\varphi) = u(\rho, \varphi)$, then from (5.120,126) we get $\bar{g} = A\bar{f}$ where

$$(A\tilde{f})(\varphi) = \sum_{n=-\infty}^{+\infty} \frac{(-i)^n}{H_n^{(1)}(k\rho)} \tilde{f}_n \, e^{in\varphi} \quad . \tag{5.127}$$

We can use again (5.123) as a constraint operator. Then we have $\alpha_n = |H_n^{(1)}(k\rho)|^{-1}$, $\beta_n = |H_n^{(1)}(k\rho_0)/H_n^{(1)}(k\rho)|$ and it is easy to see (using the results of Sect.5.2.4) that the regularized solution (5.36) or (5.37) is stable with respect to the L^2-norm when $\rho > \rho$. For $\rho = \rho_0$ the constraint (5.124) implies at most logarithmic continuity. It is interesting to understand in which cases one can get Hölder continuity. Let us consider, for instance, a perfectly conducting circular cylinder of radius ρ_0, illuminated by a plane wave. Then, at the surface of the cylinder the scattered wave takes the values $u_0(\varphi) = -\exp(ik\rho_0 \cos\varphi) = u(\rho_0, \varphi)$. If we write $\tilde{f}(\varphi) = u(\rho_0, \varphi)$, we have the Fourier expansion

$$\tilde{f}(\varphi) = -\sum_{n=-\infty}^{+\infty} i^n J_n(k\rho_0) \, e^{in\varphi} \quad . \tag{5.128}$$

Since $|H_n^{(1)}(k\rho_0)J_n(k\rho_0)| \sim (n|n|)^{-1}$, $\tilde{f}$ satisfies the condition

$$\| B\tilde{f} \|^2 = \sum_{n=-\infty}^{+\infty} |H_n^{(1)}(k\rho_0)|^2 |\tilde{f}_n|^2 \leq E^2 \quad , \tag{5.129}$$

where E is a suitable constant. Equation (5.129) implies $\beta_n = \alpha_n^{-1}$ and from (5.53) we get $\delta(\varepsilon, E) \sim E(\varepsilon/E)^{\frac{1}{2}}$, i.e., a rather good Hölder continuity. This arises also for scatterers with very smooth shape.

Finally, we discuss the angular resolution which can be obtained in the restoration of the wave function on the cylinder of radius ρ_0. We consider the restoration of a "blurred wave function" (Sect.5.2.4) and we use the constraint (5.123) with $\rho = \rho_0$ (i.e., B = I). The "blurring function" is $w_D(\varphi) = N\theta(\varphi/d)\,\mathrm{sinc}^2(\varphi/d)$. The constants N, D are given by (5.55) (where the integration ranges over [-d, d]). Then the relative error, for the restoration of the "blurred wave function" at $\varphi = 0$, can be computed by means of (5.96) (with B = I). The numerical method is described in [Ref.5.78, Chap.3]. In Fig.5.5 we give the values of the relative error as a function of the parameter D/D_0, where $D_0 = \lambda/(2\pi\rho_0) = (k\rho_0)^{-1}$ [Ref.5.78, Chap.3], for $\lambda = 2\rho_0$ and $\lambda = \rho_0/2$. As we see superresolution, i.e., restoration of details of the order of a wavelength and below, is easier when the wavelength is greater than the radius ρ_0 of the cylinder.

5.4.3 An Inverse Scattering Problem for Perfectly Conducting Bodies

An interesting combination of analytical and numerical techniques, involving the solution of linear problems, has been proposed by IMBRIALE and MITTRA [5.79] in the case of the inverse scattering problem for perfectly conducting bodies, and applied to the restoration of circular and elliptic cylinders.

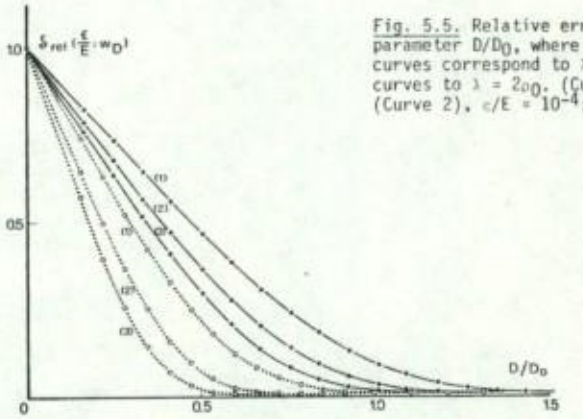

Fig. 5.5. Relative errors vs the resolution parameter D/D_0, where $D_0 = \lambda/2\pi\rho_0$. Undotted curves correspond to $\lambda = \rho_0/2$ and dotted curves to $\lambda = 2\rho_0$. (Curve 1), $\varepsilon/E = 10^{-2}$; (Curve 2), $\varepsilon/E = 10^{-4}$; (Curve 3), $\varepsilon/E = 10^{-6}$

We consider only plane wave incidence and we assume that the plane wave is z polarized. The perfectly conducting surfaces are assumed to be parallel to the z axis, so that the electromagnetic field may be derived from the single quantity $E_z = u$. The incident wave function is given by $u_0(\rho, \varphi) = \exp(ik\rho_0 \cos\varphi)$ and the associated scattered wave function u_s has the asymptotic behavior (5.125). The total field $u = u_0 + u_s$ satisfies the Helmholtz equation (5.114) in the free region and is subjected to the boundary condition $u = 0$ on the surfaces of the bodies.

The datum of the problem is the noisy scattering amplitude g and the main idea of the method is the following: reconstruct the wave function near the obstacle (from the knowledge of g) and locate points where the total wave function is zero, in order to identify points of the surface of the scatterer. This program can be accomplished in two steps. The *first step* is essentially the problem of inverse diffraction from far-field data discussed in Sect.5.4.2. Indeed, if ρ_0 is the radius of the circle tangent to the surface of the body (see Fig.5.6), at the exterior of this circle the field can be represented by the series (5.120). However, the radius ρ_0 is not known and must be determined: one must solve the inverse diffraction problem for various values of ρ_0 and choose the value for which the restored field has a zero. This zero gives a point of the surface of the scatterer. Of course, the accuracy in the determination of the zero depends on the accuracy in the restoration of the near field. As we have remarked in Sect.5.4.2, if the scatterer is a circular cylinder the accuracy can be very good (Hölder continuity). One can conjecture that, generally, the accuracy in the restoration of the near field is good when the surface of the scatterer is very smooth and poor when the surface of the scatterer is rough. The *second step* is the analytic continuation of the wave function into the region of nonconvergence of the series (5.120), i.e., $\rho < \rho_0$. If we know that the

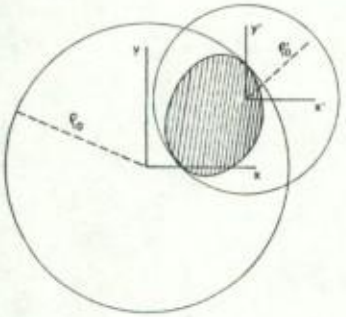

Fig. 5.6. Geometry for analytic continuation into region $\rho < \rho_0$ in the case of a convex body

scatterer is a *convex* body, then the analytic continuation can be accomplished by a simple technique of shifting the origin of the coordinate system [5.79] (see Fig. 5.6). Since the new coordinate system is obtained by translating the original one, then the scattering amplitude in the new system can be easily obtained from g [5.79]. The solution of the problem of inverse diffraction, with the new scattering amplitude as data, gives another point of the surface of the body. The procedure can be repeated and, since the body is convex, a few points can be sufficient in order to characterize the shape of the scatterer. An accurate analysis of this method in the case of a perfectly conducting *circular cylinder* has been done by CABAYAN et al. [5.80], using a stability criterion due to TWOMEY [5.29], which is equivalent to use the truncated solution (5.37) in the special case B = I. They show that even a "coarse" near-field map can give some information on the size and center position of the scatterer. Of course, the results of the method are very good because in this case, as shown in Sect.5.4.2, we have Hölder continuity.

When the body is *not convex*, analytic continuation of the wave function can be done as follows [5.79]. Once the total field, outside the minimum circle enclosing the scatterer, has been restored, then one can take a point in the exterior region as the origin of a new coordinate system ρ', φ' and represent the total field in the neighborhood of this point as a series of Bessel functions

$$u(\rho', \varphi') = \sum_{n=-\infty}^{+\infty} \frac{J_n(k\rho')}{J_n(k\rho_0')} c_n' e^{in\varphi'} \quad . \tag{5.130}$$

The circle $\rho' = \rho_0'$ is interior to the region $\rho > \rho_0$ (see Fig.5.7) and the c_n' are the Fourier coefficients of $u(\rho_0', \varphi')$. Of course, the summation of this series is an ill-posed problem if $\rho' > \rho_0'$, since $|J_n(k\rho')/J_n(k\rho_0')| \sim \exp[|n|\ln(\rho'/\rho_0')]$ when $|n| \to +\infty$. A discussion of the regularization of (5.130) can be found in [Ref.5.78, Chap.3]. The results are very similar to those for inverse diffraction. Anyway one can estimate the series (5.130) which, in principle, converges in the interior of the circle $\rho_i' > \rho_0'$, tangent to the body surface. Since ρ_i' is not known, one should estimate the series (5.130) for various values of $\rho' > \rho_0'$ and choose the smallest value of ρ' for which the restored field has a zero. So a new point of the scatterer has been determined. By a series of overlapping circles it is then possible, in principle, to obtain values of u in all the points of the space external to the scatterer. Of course error propagation can prevent having a satisfactory estimate of the shape of the scatterer.

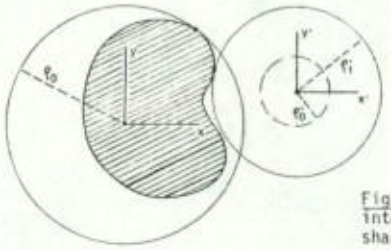

Fig. 5.7. Geometry for analytic continuation into region $\rho < \rho_0$ in the case of an arbitrarily shaped body

The previous method of analytic continuation has also been used by AHLUWALIA and BOERNER [5.81] for recovering the electrical size, the surface locus and the averaged local surface impedance in the case of circular cylindrical monobody and twobody shapes. The same authors have also extended the method to the case of spherical surfaces [5.82].

5.4.4 Inverse Scattering Problems in the Born Approximation

We consider two problems: the determination of the shape of a perfectly conducting body and the reconstruction of the refractive index of a semi-transparent object.

When a *perfectly conducting object* is illuminated by an electromagnetic wave, the scattered field can be determined using the Born approximation (also known as Kirchhoff's or physical optics approximation) if the wavelengths λ are small compared with the characteristic dimensions of the scatterer. Assuming that the measured data are the values of the backscattered far field, then the determination of the shape of the body is a linear inverse problem [5.83,84]. Consider a plane wave, with electric field $\underline{E}_i(\underline{r}) = \underline{E}_0 \exp(ik\underline{s}_0 \cdot \underline{r})$, scattered by a smooth, convex and bounded target τ. The backscattered field $\underline{E}_b(\underline{r})$, i.e., the field observed in the direction $\underline{s}_b = -\underline{s}_0$, is given by $\underline{E}_b(\underline{r}) \sim \rho(\underline{k})(2\sqrt{\pi}r)^{-1} \exp(ikr)\underline{E}_0$, $r \to +\infty$, $\rho(\underline{k})$ being proportional to the backward scattering amplitude. Then one can prove that, in the Born approximation

$$r(\underline{k}) = \frac{2\sqrt{\pi}}{k^2} [\rho(\underline{k}) + \rho^*(-\underline{k})] = \int \gamma(\underline{r}) \exp(2i\underline{k} \cdot \underline{r}) d\underline{r} \quad , \tag{5.131}$$

where $\underline{k} = k\underline{s}_0$ and $\gamma(\underline{r})$ is the characteristic function of the target, i.e., $\gamma(\underline{r}) = 1$ when $\underline{r}$ is in τ, $\gamma(\underline{r}) = 0$ otherwise [5.83,84] (see also [Ref.5.75, Sect.3.2.5]). If the backscattered field could be measured for all frequencies and for all directions of incidence, then the Fourier transform of $\gamma(\underline{r})$ would be known and $\gamma(\underline{r})$ could be

determined. In practice, $\Gamma(\underline{k})$ is measurable only in a restricted domain of the $\underline{k}$ space. This does not seem to be a serious restriction, as regards the possibility of a unique reconstruction of $\gamma(\underline{r})$, since $\Gamma(\underline{k})$ is an analytic function of $\underline{k}$ [Ref. 5.75, Sect.3.2.5]. However, this reconstruction, which is equivalent from the mathematical point of view to bandwidth extrapolation (Sect.5.4.1), is not stable.

In radar applications, $\Gamma(\underline{k})$ is measured only in a set of points interior to the annular region $m \leq |\underline{k}| \leq M$, where $k_1 = m/2$ and $k_2 = M/2$ correspond to minimum and maximum values of the usable frequency band. It must also be remarked that the lower bound has to be large enough in order to assure the validity of (5.131): the largest wavelength in the incident field, $\lambda_1 = 4\pi/m$, must be short compared to the target shape. The inaccessibility of low frequency information is essentially related to the intrinsic limitations of the Born approximation. The instability due to this fact has been investigated by many authors [5.85-88]. PERRY [5.86], for instance, applied Tikhonov's regularization method to the one-dimensional case, assuming that $\Gamma(\underline{k})$ is known for $|\underline{k}| \geq m$ (perfect high-pass filter). In other words, he did not care about the limitations due to the lack of information at high frequencies. If the object is located within the interval $[-X/2, X/2]$, then $\gamma(x)$ can be determined as the solution of the Fredholm integral equation of the second kind

$$\gamma(x) - \int_{-X/2}^{X/2} \frac{\sin[m(x-y)]}{\pi(x-y)} \gamma(y)dy = g(x) \quad , \tag{5.132}$$

where $g(x)$ is the inverse Fourier transform of the available values of $\Gamma(\underline{k})$. Equation (5.132) has the form $(1-A)\gamma = g$ where A is the integral operator (5.105) with $\pi\Omega = m$. Therefore solving (5.132) is not an ill-posed problem in the strict sense but, when m is large, it is an ill-conditioned problem. Indeed, the condition number of $1-A$ is $\alpha = (1-\lambda_0)^{-1}$, where λ_0 is the largest eigenvalue of A. From the behavior of λ_0 for large m [5.89] it follows: $\alpha \sim (2\pi/mX)^{\frac{1}{2}} \exp(mX)/4$. In radar applications mX is relatively large and therefore α can be rather large. After discretization, an ill-conditioned problem shows the same features as an ill-posed problem and therefore regularization methods can be useful.

In order to circumvent the lack of information at low frequencies, another technique has been first suggested by BOJARSKI [5.83], and further developed by MAGER and BLEISTEIN [5.88]. The essence of the method is to examine not the characteristic function of the target, but rather the directional derivative of this function. Indeed, if $\underline{s} \cdot \underline{\nabla}\gamma$ is the derivative of γ in the direction of the unit vector $\underline{s}$, then its Fourier transform is the product of $\Gamma(\underline{k})$ by the factor $\underline{s} \cdot \underline{k}$. In this way, one simultaneously attenuates low-frequency data while enhancing the effect of high-frequency data. One expects that the limitations of the method are essentially due to the lack of information at high frequencies. The function $\underline{s} \cdot \underline{\nabla}\gamma$ is highly singular; more precisely $\underline{s} \cdot \underline{\nabla}\gamma = \underline{s} \cdot \underline{n}\delta$ where $\underline{n}$ is the unit outward normal to the surface of the body and δ is a Dirac delta measure concentrated on the surface of the body. As proved by MAGER and BLEISTEIN [5.88], similar features are shown by the function

$$h(\underline{r}, \underline{s}) = \frac{1}{(2\pi)^2} \int i\underline{s} \cdot \underline{k} a(\underline{k}) \Gamma(\underline{k}) \, e^{-i\underline{k} \cdot \underline{r}} d\underline{k} \quad . \tag{5.133}$$

Here $a(\underline{k})$ is the characteristic function of the domain in $\underline{k}$-space (interior to the annular region $m \leq |\underline{k}| \leq M$) where the backscattered field is measured. In the high-

frequency limit ($m \gg 1$), if $\underline{r}_0$ is a point on the surface of the scatterer, then

$$\mathrm{Re}(h(\underline{r}, \underline{s})) \sim (\text{constant}) |\underline{r} - \underline{r}_0|^{-1} [\sin(M|\underline{r} - \underline{r}_0|) - \sin(m|\underline{r} - \underline{r}_0|)] \tag{5.134}$$

provided that the vector $\underline{r} - \underline{r}_0$ is orthogonal to the surface at $\underline{r}_0$ and its direction coincides with a direction of incidence. Therefore the function $\mathrm{Re}(h(\underline{r}, \underline{s}))$ has a central lobe which peaks on the target surface in regions with surface normals parallel to directions of incidence. The height of the central lobe is proportional to $M-m$ and its width is approximately equal to $2\pi/M$ if $M \gg m$. Remark that $2\pi/M = \lambda_2/2$ where λ_2 is the smallest wavelength in the incident field. It must also be observed that this result has been derived by MAGER and BLEISTEIN [5.88] for noiseless data. Now, the factor $\underline{s} \cdot \underline{k}$ does not only enhance the effect of high-frequency data but also the effect of the noise on these data. Error propagation in the determination of $\mathrm{Re}(h(\underline{r}, \underline{s}))$ is controlled by the quantity (condition number) $\alpha = M/m$. When this parameter is large, the resolution limit is certainly worse than $\lambda_2/2$. For the numerical examples presented by MAGER and BLEISTEIN [5.88], α is of the order of 2 and therefore the results are quite good.

Finally we want to remark that an improvement of the resolution limit intrinsic to $\mathrm{Re}(h(\underline{r}, \underline{s}))$ would require an analytic continuation of the backscattered field in the region $|\underline{k}| > M$. From the analysis of bandwidth extrapolation done in Sect. 5.4.1, it clearly appears that this problem is affected by logarithmic continuity and therefore an improvement of the resolution limit $\lambda_2/2$ is practically impossible.

The reconstruction of the *refractive index* of weakly scattering semi-transparent objects, using the Born approximation, has been widely discussed [5.90-94], with special attention to the problem of uniqueness of the solution [Ref.5.75, Sect. 3.4.3]. As shown by WOLF [5.91,95], modulus and phase of the scattering amplitude can be derived, using holographic data, from the homogeneous part of the angular spectrum of the scattered field. DANDLIKER and WEISS [5.92] stressed that appropriate variation of the direction of the incident wave is crucial for holographic 3D reconstruction.

The wave function u satisfies the equation

$$\nabla^2 u + k_0^2 n^2(\underline{r}) u = 0 \quad , \tag{5.135}$$

where $n(\underline{r})$ is the (possibly complex) refractive index at the point $\underline{r}$. If the object is situated in free space, then $n(\underline{r}) = 1$ outside the object. Equation (5.135) can be recasted in the following form

$$\nabla^2 u + k_0^2 u = F(\underline{r}) u \quad , \tag{5.136}$$

where

$$F(\underline{r}) = -k_0^2[n^2(\underline{r}) - 1] \quad . \tag{5.137}$$

The function $F(\underline{r})$ is called the *scattering potential* and it is evidently zero at all points outside the object. Consider an incident plane wave $u_i(\underline{r}) = \exp(ik_0\underline{s}_0 \cdot \underline{r})$; then the Born approximation can be used for the determination of the scattered wave function $u_s(\underline{r})$, if the object scatters weakly, i.e., if $|u_s| << |u_i|$. When this condition is satisfied, in the far zone we have

$$u_s(\underline{r}, k_0\underline{s}_0) \sim -\frac{\exp(ik_0 r)}{4\pi r} A_B(k_0\underline{s}_0, k_0\underline{s}) \quad , \tag{5.138}$$

where

$$A_B(k_0\underline{s}_0, k_0\underline{s}) = \int F(\underline{r}') \exp[-ik_0(\underline{s} - \underline{s}_0) \cdot \underline{r}']d\underline{r}' \quad . \tag{5.139}$$

Therefore the Born approximation to the scattering amplitude is essentially given by the Fourier transform $\tilde{F}(\underline{k})$ of the scattering potential $F(\underline{r})$. Inspection of (5.139) shows that, for a fixed direction of incidence $\underline{s}_0$, A_B gives those Fourier components of $F(\underline{r})$ which correspond to points on the surface of the sphere with center $k_0\underline{s}_0$ and radius k_0. By varying the direction of incidence $\underline{s}_0$, a (theoretically infinite) number of experiments would allow one to determine $\tilde{F}(\underline{k})$ for all values of $\underline{k}$ lying within the sphere of radius $2k_0$. Then a bandlimited approximation $F_{b\ell}(\underline{r})$ to the scattering potential is given by

$$F_{b\ell}(\underline{r}) = \frac{1}{(2\pi)^3} \int_{|\underline{k}| \leq 2k_0} \tilde{F}(\underline{k})\, e^{i\underline{k} \cdot \underline{r}}\, d\underline{k} \quad . \tag{5.140}$$

A rough measure of the limit of resolution of WOLF's approach [5.91], intrinsic to (5.140), is given by $\lambda_0/2 = \pi/k_0$ (when the scattered field is determined by side-band holography the limit of resolution is about $9\lambda_0$ [5.95]). An improvement beyond these limits is, in principle, possible since $F(\underline{k})$ is an analytic function when the object is localized within a finite volume τ. We encounter, once more, a problem which is equivalent, from the mathematical point of view, to bandwidth extrapolation. Therefore a significant improvement of the resolution limit seems to be, in practice, impossible. Besides, the effect of the noise can be very important (it is necessary to detect a weak scattered field in the presence of a strong unscattered field), so that even the theoretical limit of resolution cannot be reached. A new approach to the optical inverse scattering problem, based on interference with three variations of a spherical reference wave, has been proposed by LAM et al. [5.123].

Experiments have been undertaken in order to investigate the use of the technique suggested by Wolf's theory [5.91,95] and computational reconstruction of objects from holograms has been attempted [5.96-98]. Very simple objects have been considered, i.e., rectangular (homogeneous and inhomogeneous) and cylindrical bars. In these cases, because of the symmetry of the objects only one hologram is needed. The reported numerical results show spurious oscillations which probably can be smoothed by a filtering of the Fourier transform of $F(\underline{r})$. The 3D scattering potential of microscopic objects (40 μm diameter) has recently been reconstructed by FERCHER et al. [5.124].

5.4.5 Object Reconstruction from Projections and Abel Equation

Object reconstruction from projections and Abel equations are two examples of inverse problems which arise when the variations of the dynamical functions over a given wavelength are so small that diffraction can be neglected. For both problems there exists an enormous amount of literature. Our purpose is only to point out that, as a consequence of the physical approximation intrinsic to these problems, the restored continuity is quite good (Hölder continuity).

Assuming straight line ray propagation with the amplitude (or the phase) of the ray controlled by the line integral of a density function, a projection of the object onto a plane is measured. Typical examples are X-ray shadowgraphs. In two dimensions the mathematical formulation of the problem is as follows. Let $\bar{f}(\underline{r})$ be a density function which has support in the circle $|\underline{r}| \leq R$ and let L be the straight line defined by $\underline{s} \cdot \underline{r} = p$. Here $\underline{r} = (x, y)$ is a point of the plane, $\underline{s} = (\cos\varphi, \sin\varphi)$, $0 \leq \varphi < \pi$, is the unit vector orthogonal to L and p is the distance of L from the origin, $-\infty < p < +\infty$. If $\underline{t} = (-\sin\varphi, \cos\varphi)$ is the unit vector of the direction of L, consider the line integral

$$\bar{g}(p, \varphi) = \int_{-\infty}^{+\infty} \bar{f}(p\underline{s} + q\underline{t})dq \quad . \tag{5.141}$$

Obviously $\bar{g}(p, \varphi) = 0$ when $|p| \geq R$. The function $\bar{g}(p, \varphi)$ is known as the *Radon transform* of $\bar{f}(\underline{r})$ and, for fixed φ, it gives the projection of $\bar{f}(\underline{r})$ onto a straight line parallel to $\underline{s}$. Thus, in two dimensions, object reconstruction from projections is exactly the inversion of the Radon transform. The solution of this problem was given by RADON [5.99] in 1917. Nowadays there are many fields of application: it is sufficient to mention computerized tomography (see, e.g. [5.100,101]), radio astronomy [5.102], electron microscopy [5.103], radar target shape estimation [5.104] and so on (see also [Ref.5.45, Sect.2.3]). Solving (5.141) is an ill-posed problem. This fact clearly appears from the Radon inversion formula [5.100] since it contains the derivative of the noiseless data $\bar{g}$. In order to investigate error propagation, let us take as solution space F the space of the square integrable functions, which have support in the circle $|\underline{r}| \leq R$, and as data space G the space of the square integrable functions over the rectangle $0 \leq \varphi < \pi$, $|p| \leq R$. Then the linear operator A defined by (5.141) is a continuous operator form F into G [Ref.5.100, Sect.12].

Using Parseval's equality for Fourier transform, the norm of f can be written as follows:

$$\| f \|_F^2 = \int_{|\underline{r}| \leq R} |f(\underline{r})|^2 d\underline{r} = \frac{1}{(2\pi)^2} \int_0^{\pi} d\varphi \int_{-\infty}^{+\infty} |h(\nu, \varphi)|^2 d\nu \ , \tag{5.142}$$

where $h(\nu, \varphi) = |\nu|^{\frac{1}{2}} \hat{f}(\nu \underline{s})$ and $\underline{s} \equiv (\cos\varphi, \sin\varphi)$. Then, from the "projection slice theorem" (see, e.g. [5.105] or [Ref.5.45, Sect.2.3.4])

$$(Af)(p, \varphi) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ip\nu} \hat{f}(\nu \underline{s}) d\nu \tag{5.143}$$

and from Parseval's equality it follows

$$\| Af \|_G^2 = \int_0^{\pi} d\varphi \int_{-R}^{R} |(Af)(p, \varphi)|^2 dp = \frac{1}{2\pi} \int_0^{\pi} d\varphi \int_{-\infty}^{+\infty} \frac{|h(\nu, \varphi)|^2}{|\nu|} d\nu \ . \tag{5.144}$$

Finally we can require, as a constraint, a bound on the first derivatives of $f(\underline{r})$. This bound is not very realistic in many applications of the Radon transform, since one should also reconstruct discontinuous functions. However our purpose is only to discuss, in the simplest way, the stability of the inversion procedure. Then, using again Parseval's equality we have

$$\| Bf \|_F^2 = \int_{|\underline{r}| \leq R} \left(\left| \frac{\partial f}{\partial x} \right|^2 + \left| \frac{\partial f}{\partial y} \right|^2 \right) d\underline{r} = \frac{1}{(2\pi)^2} \int_0^{\pi} d\varphi \int_{-\infty}^{+\infty} |\nu|^2 |h(\nu, \varphi)|^2 d\nu \ . \tag{5.145}$$

It is now easy to recognize that the stability estimate for the Radon inverse problems is given by (5.54) if we put $|\hat{K}(\nu)| = |\nu|^{-\frac{1}{2}}$ and $|\hat{\beta}(\nu)| = |\nu|$. Therefore $|\hat{\beta}(\nu)| = |\hat{K}(\nu)|^{-\mu}$, $\mu = 2$ and $\delta(\varepsilon, E) \sim E(\varepsilon/E)^{\alpha}$ with $\alpha = 2/3$. The restored stability is quite good and this result should be related to the fact that, in this mathematical model, diffraction has been neglected.

When the object has circular symmetry, then $\bar{g}(p, \varphi)$, given by (5.141), is the same for all φ. Put $\varphi = 0$ in (5.141), so that $p = x$, $q = y$. In terms of the variable $\rho = (x^2 + y^2)^{\frac{1}{2}}$ we have ($x, y > 0$)

$$\bar{g}(x) = 2 \int_x^{+\infty} \frac{\rho \bar{f}(\rho)}{(\rho^2 - x^2)^{\frac{1}{2}}} d\rho \ . \tag{5.146}$$

This is a form of the Abel integral equation which is also fundamental whenever Fermat's principle can be used for the calculation of the rays. Many applications of the Abel equation are discussed in [5.106]. From the previous remark we expect that the same kind of stability holds for object reconstruction from projection and the Abel equation. However it is interesting to derive directly this result. To this purpose let us consider the inversion of the following integral operator (the various forms of the Abel equation can be treated in a similar way):

$$(Af)(x) = \int_0^x \frac{f(y)}{\sqrt{\pi x(x-y)}} dy \ , \tag{5.147}$$

taking $L^2(0, +\infty)$ both as solution and data space. Then $A : F \to G$ is a linear continuous operator. Introducing the Mellin transform of f,

$$\tilde{f}_M(\nu) = \int_0^{+\infty} x^{-\frac{1}{2}+i\nu} f(x) dx \ , \tag{5.148}$$

it is easy to show that the Mellin transform of Af is given by $[\Gamma(1/2 - i\nu)/\Gamma(1 - i\nu)]$ $\tilde{f}_M(\nu)$. Using Parseval's equality for Mellin transform [Ref.5.107, pp.94-95] we get

$$\| Af \|^2 = \int_0^{+\infty} |(Af)(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\tanh(\pi\nu)}{\nu} |\tilde{f}_M(\nu)|^2 d\nu \ . \tag{5.149}$$

Take the constraint operator defined by $(Bf)(x) = xf'(x)$. Observing that the Mellin transform of Bf is given by $-(1/2 + i\nu)\tilde{f}_M(\nu)$ and using again Parseval's equality we get

$$\| Bf \|^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (\nu^2 + \tfrac{1}{4}) |\tilde{f}_M(\nu)|^2 d\nu \ . \tag{5.150}$$

From (5.149,150) it is easy to recognize that the stability estimate $\delta(\varepsilon, E)$ for the Abel equation is given by (5.54) with $|\hat{K}(\nu)|^2 = \nu^{-1} \tanh(\pi\nu) \sim |\nu|^{-1}$, $|\nu| \to +\infty$, and $|\hat{\beta}(\nu)|^2 = \nu^2 + 1/4 \sim \nu^2$, $|\nu| \to +\infty$. Since $|\hat{\beta}(\nu)| \sim |\hat{K}(\nu)|^{-\mu}$, $|\nu| \to +\infty$, with $\mu = 2$, it follows $\delta(\varepsilon, E) \sim E(\varepsilon/E)^{2/3}$, i.e., the same result as for the inversion of the Radon transform.

5.4.6 Concluding Remarks and Open Problems

In this review of a continuously expanding field, we restricted ourselves to some typical linear inverse problems, and so we omitted many important topics where regularization methods apply as well. Let us mention for instance, polarization utilization in electromagnetic inverse scattering [5.45, and Chapt.7 of this volume] and laser anemometry data analysis [5.108,109]. Regularization methods can also be useful for synthesis problems [5.110]. The main difference between inverse and synthesis problems can be easily understood in the case of an antenna: the inverse problem is the identification of an actual antenna from measurement of its radiation pattern, while the synthesis problem is the design of an antenna producing a given radiation pattern. In the latter case, rather than in stability, one is interested in "sensitivity": if the computed antenna is not exactly realized, how much will its pattern function be modified?

Since the mathematical pathology of all those problems is quite similar whatever the particular field one considers, we think that regularization theory should

provide a unified framework for treating linear inverse problems and for investigating thoroughly their stability. However, regularization methods are by no means a cure-all. Indeed, as we have seen, the faster the decay of the eigenvalues (singular values) of the operator A, the greater the loss of information due to the smoothing effect of A. Hence it is expected that in some cases, the available data are truly insufficient and no meaningful prior knowledge can provide a satisfactory solution. Very important in this connection is the precise valuation of error propagation in the regularized inversion procedure. This enables us to estimate in practical cases, the accuracy of the solution for a given noise level and a given prior knowledge. We also showed that this error analysis enlightens theoretical questions like the problem of superresolution. Summarizing the results of Sect.5.4.1, we can say that superresolution appears practically impossible for imaging systems with a large aperture, because it would require unrealistically high signal-to-noise ratios. This is due to a very fast increase of the relative error on blurred solutions beyond the Rayleigh limit. A similar feature arises in near-field reconstruction from the scattering amplitudes (see Sect.5.4.2): the reconstruction of source details of the order of a wavelength and below (for a review on this subject see [5.111]) appears very difficult when the wavelength λ is much smaller than the characteristic dimension ℓ of the source. Superresolution becomes however easier when λ is of the same order as ℓ.

More fundamental questions are still open in the field of ill-posed problems. The first point, very important, is to develop a sound theory when data are given only at a finite number of points (in regularization theory, as described in this chapter, one assumes that the data are functions defined everywhere). Some results in this direction have been obtained for the problems of analytic continuation [5.112,113] harmonic continuation [5.114] and numerical differentiation [5.115]. In these cases, the main ideas of regularization theory (prior knowledge, least-square methods, stability estimates) have been maintained.

A second point would be to extend regularization theory beyond its actual frame: linear problems and prior knowledge expressed in the form (5.22). For instance, a positivity constraint, which appears naturally in some problems, cannot be expressed in this way. For particular ill-posed problems (harmonic continuation, backward heat equation), it is known that positive solutions are necessarily stable [5.116]. However, the requirement of positivity alone is not sufficient for stabilizing Fredholm integral equations of the first kind. Some algorithms, reviewed in [5.72], have been developed for introducing the positivity constraint in the analysis of imaging systems but, to our knowledge, no theoretical analysis of the solution accuracy has been done. It is clear that a supplementary constraint of positivity improves the solution, but the quantitative estimation of this improvement is still an open question.

Acknowledgements

We are deeply indebted to Dr. McWhirter who has kindly supplied us with a long list of references. We also want to thank Prof. G. Talenti for many long and informative discussions concerning the mathematics of ill-posed problems. One of us (C.D.M.) is indebted to Prof. J. Reignier for critical and helpful remarks. Last but not least, it is a pleasure for the authors to thank Miss B. Basiglio for the preparation of the manuscript and careful typing.

References

5.1 J. Hadamard: *Lectures on the Cauchy Problem in Linear Partial Differential Equations* (Yale University Press, New Haven 1923)
5.2 R. Courant, D. Hilbert: *Methods of Mathematical Physics*, Vol. 2 (Interscience, New York 1962)
5.3 M.N. Lavrentiev: *Some Improperly Posed Problems of Mathematical Physics*, Springer Tracts in Natural Philosophy, Vol. 11 (Springer, Berlin, Heidelberg, New York 1967)
5.4 A. Tikhonov, V. Arsenine: *Méthodes de Résolution de Problèmes Mal Posés* (Mir, Moscow 1976)
5.5 L.E. Payne: *Improperly Posed Problems in Partial Differential Equations* (SIAM, Philadelphia 1975)
5.6 S.G. Mikhlin: *Integral Equations* (Pergamon, London 1957)
5.7 A.V. Balakrishnan: *Applied Functional Analysis*, Applications of Mathematics, Vol.3 (Springer, Berlin, Heidelberg, New York 1976)
5.8 K. Chadan, P.C. Sabatier: *Inverse Problems in Quantum Scattering Theory* (Springer Berlin, Heidelberg, New York 1977)
5.9 H.P. Baltes (ed.): *Inverse Source Problems in Optics*, Topics in Current Physics, Vol. 9 (Springer, Berlin, Heidelberg, New York 1978)
5.10 M. Reed, B. Simon: *Methods of Modern Mathematical Physics* (Academic Press, New York 1972)
5.11 P.C. Sabatier (ed.): *Applied Inverse Problems*, Lecture Notes in Physics, Vol. 85 (Springer, Berlin, Heidelberg, New York 1978)
5.12 M.Z. Nashed (ed.): *Generalized Inverses and Applications* (Academic Press, New York 1976)
5.13 A.N. Tikhonov: Sov. Math. Dokl. *4*, 1035-1038 (1963)
5.14 F. John: Commun. Pure Appl. Math. *13*, 551-585 (1960)
5.15 C. Pucci: Atti Accad. Naz. Lincei *18*, 473-477 (1955)
5.16 K. Miller: SIAM J. Math. Anal. *1*, 52-74 (1970)
5.17 K. Miller, G.A. Viano: J. Math. Phys. *14*, 1037-1048 (1973)
5.18 V.F. Turchin, V.P. Kozlov, M.S. Malkevich: Sov. Phys.-Usp. *13*, 681-703 (1971)
5.19 V.A. Morozov: U.S.S.R. Comp. Math. and Math. Phys. *10*, N.4, 10-25 (1970)
5.20 J.N. Franklin: J. Math. Anal. Appl. *31*, 682-716 (1970)
5.21 M. Bertero, G.A. Viano: Bollettino U.M.I. *15-B*, 483-508 (1978)
5.22 M. Bertero, C. De Mol, G.A. Viano: J. Math. Phys. *20*, 509-521 (1979)
5.23 E.C. Titchmarsh: *The Theory of Functions* (Oxford University Press, Oxford 1939)
5.24 E. Hille, J.D. Tamarkin: Acta Math. *57*, 1-76 (1931)
5.25 G. Talenti: Bollettino U.M.I. *15-A*, 1-29 (1978)
5.26 M. Bertero, C. De Mol, G.A. Viano: Opt. Acta *27*, 307-320 (1980)
5.27 E. Picard: R.C. Mat. Palermo *29*, 615-619 (1910)
5.28 T. Kato: *Perturbation Theory for Linear Operators* (Springer, Berlin, Heidelberg, New York 1966)
5.29 S. Twomey: J. Franklin Inst. *279*, 95-109 (1965)
5.30 V.A. Morozov: Sov. Math. Dokl. *8*, 1000-1003 (1967)
5.31 J.N. Franklin: Math. Comp. *28*, 889-907 (1974)
5.32 D.L. Phillips: J. ACM *9*, 84-96 (1962)
5.33 S. Twomey: J. ACM *10*, 97-101 (1963)
5.34 L. Eldén: Ph. D. Thesis, Linköping University (1977)

5.35 G.F. Miller: "Fredholm Equations of the First Kind", in *Numerical Solution of Integral Equations*, ed. by L.M. Delves, J. Walsh (Claredon Press, Oxford 1974) pp.175-188
5.36 M. Bertero, C. De Mol, G.A. Viano: "On the Regularization of Linear Inverse Problems in Optics", in *Applied Inverse Problems*, ed. by P.C. Sabatier, Lecture Notes in Physics, Vol. 85 (Springer, Berlin, Heidelberg, New York 1978) pp.180-199
5.37 G. Backus, F. Gilbert: Geophys. J. R. Astron. Soc. *16*, 169-205 (1968)
5.38 G. Backus, F. Gilbert: Phil. Trans. Roy. Soc. A-*266*, 123-192 (1970)
5.39 J.L. Doob: *Stochastic Processes* (Wiley, New York 1953)
5.40 L.E. Franks: *Signal Theory* (Prentice-Hall, Englewood Cliffs 1969)
5.41 A. Papoulis: *Probability, Random Variables and Stochastic Processes* (McGraw-Hill, New York 1965)
5.42 G.E. Backus: "Inference from Inadequate and Inaccurate Data", in *Mathematical Problems in the Geophysical Sciences*, ed. by W.H. Reid, Lectures in Applied Math., Vol 14 (AMS, Providence 1971)
5.43 I.M. Gel'fand, N.Y. Vilenkin: *Generalized Functions*, Vol. 4 (Academic Press, New York 1964)
5.44 I.M. Gel'fand, A.M. Yaglom: Am. Math. Soc. Trans. *12*, 199-246 (1959)
5.45 W.M. Boerner: "Polarization Utilization in Electromagnetic Inverse Scattering"; Communications Laboratory Rpt. 78-3, University of Illinois at Chicago Circle (1978)
5.46 H. Wolter: "On Basic Analogies and Principal Differences between Optical and Electronic Information", in *Progress in Optics*, Vol. I, ed. by E. Wolf (North Holland, Amsterdam 1961) pp.155-210
5.47 R.N. Bracewell, J.A. Roberts: Aust. J. Phys. *7*, 615-640 (1954)
5.48 Y.T. Lo: J. Appl. Phys. *32*, 2052-2054 (1961)
5.49 H. Wolter: Physica *24*, 457-475 (1958)
5.50 B.J. Hoenders, H.A. Ferwerda: Optik *37*, 542-556 (1973)
5.51 C. Shannon: *The Mathematical Theory of Communication* (University of Illinois Press, Urbana 1949)
5.52 D. Gabor: "Light and Information", in *Astronomical Optics and Related Subjects*, ed. by Z. Kopal (North-Holland, Amsterdam 1956) pp.17-30
5.53 G. Toraldo di Francia: J. Opt. Soc. Am. *59*, 799-804 (1969)
5.54 D. Slepian, H.O. Pollack: Bell System Tech. J. *40*, 43-63 (1961)
5.55 D. Slepian: J. Math. & Phys. *44*, 99-140 (1965)
5.56 D. Slepian, E. Sonnenblick: Bell Syst. Tech. J. *44*, 1745-1759 (1965)
5.57 C. Flammer: *Spheroidal Wave Functions* (Stanford University Press, Stanford 1957)
5.58 H.J. Landau: Trans. Am. Math. Soc. *115*, 242-256 (1965)
5.59 B.R. Frieden: "Evaluation, Design and Extrapolation Methods for Optical Signals, Based on Use of the Prolate Functions", in *Progress in Optics*, Vol. 9, ed. by E. Wolf (North-Holland, Amsterdam 1971) pp.311-407
5.60 F. Gori, G. Guattari: Opt. Commun. *7*, 163-165 (1973)
5.61 M. Bertero, C. De Mol, G.A. Viano: Opt. Lett. *3*, 51-53 (1978)
5.62 C.W. Helstrom: J. Opt. Soc. Am. *67*, 833-838 (1977)
5.63 C.K. Rushforth, R.W. Harris: J. Opt. Soc. Am. *58*, 539-545 (1968)
5.64 J.W. Goodman: "Synthetic Aperture Optics", in *Progress in Optics*, Vol. 8, ed. by E. Wolf (North-Holland, Amsterdam 1970) pp. 1-50
5.65 G. Toraldo di Francia: Riv. Nuovo Cimento *1* (Numero Speciale), 460-484 (1969)
5.66 C.W. Helstrom: J. Opt. Soc. Am. *57*, 297-303 (1967)
5.67 C.L. Rino: J. Opt. Soc. Am. *59*, 547-553 (1969)
5.68 M. Bendinelli, A. Consortini, L. Ronchi, B.R. Frieden: J. Opt. Soc. Am. *64*, 1498-1502 (1974)
5.69 H. Dym, H.P. McKean: *Fourier Series and Integrals* (Academic Press, New York 1972)
5.70 G.J. Buck, J.J. Gustincic: IEEE Trans. AP-*15*, 376-381 (1967)
5.71 B.R. Frieden: J. Opt. Soc. Am. *57*, 1013-1019 (1967)
5.72 B.R. Frieden: "Image Enhancement and Restoration", in *Picture Processing and Digital Filtering*, 2nd ed., ed. by T.S. Huang, Topics in Applied Physics, Vol. 6 (Springer, Berlin, Heidelberg, New York 1979) pp.177-248

5.73 G.A. Viano: J. Math. Phys. *17*, 1160-1165 (1976)
5.74 J.R. Shewell, E. Wolf: J. Opt. Soc. Am. *58*, 1596-1603 (1968)
5.75 B.J. Hoenders: "The Uniqueness of Inverse Problems", in *Inverse Source Problems in Optics*, ed. by H.P. Baltes, Topics in Current Physics, Vol. 9 (Springer, Berlin, Heidelberg, New York 1978) pp.41-82
5.76 G.C. Sherman: J. Opt. Soc. Am. *57*, 1490-1498 (1967)
5.77 K. Miller: Arch. Rational Mech. Anal. *16*, 126-154 (1964)
5.78 C. De Mol: "Sur la Regularisation des Problèmes Inverses Linèaires"; Ph.D. Thesis, Université Libre de Bruxelles (1979)
5.79 W.A. Imbriale, R. Mittra: IEEE Trans. AP-*18*, 633-642 (1970)
5.80 H.S. Cabayan, R.C. Murphy, T.J.F. Pavlàsek: IEEE Trans. AP-*21*, 346-351 (1973)
5.81 H.P.S. Ahluwalia, W.M. Boerner: IEEE Trans. AP-*21*, 663-672 (1973)
5.82 H.P.S. Ahluwalia, W.M. Boerner: IEEE Trans. AP-*22*, 673-682 (1974)
5.83 N.N. Bojarski: "A Survey of Electromagnetic Inverse Scattering"; Syracuse Univ. Res. Corp., Special Projects Lab. Rpt., DDC #AD-813-851 (1966)
5.84 R.M. Lewis: IEEE Trans. AP-*17*, 308-314 (1969)
5.85 W. Tabbara: IEEE Trans. AP-*21*, 245-247 (1973)
5.86 W.L. Perry: IEEE Trans. AP-*22*, 826-829 (1974)
5.87 W. Tabbara: IEEE Trans. AP-*23*, 446-448 (1975)
5.88 R.D. Mager, N. Bleistein: IEEE Trans. AP-*26*, 695-699 (1978)
5.89 W.H.J. Fuchs: J. Math. Anal. Appl. *9*, 317-330 (1964)
5.90 R.W. Hart, E.P. Gray: J. Appl. Phys. *35*, 1408-1415 (1964)
5.91 E. Wolf: Opt. Commun. *1*, 153-156 (1969)
5.92 R. Dändliker, K. Weiss: Opt. Commun. *1*, 323-328 (1969)
5.93 H.A. Ferwerda, B.J. Hoenders: Optik *39*, 317-326 (1974)
5.94 A.J. Devaney: J. Math. Phys. *19*, 1526-1531 (1978)
5.95 E. Wolf: J. Opt. Soc. Am. *60*, 18-20 (1970)
5.96 W.H. Carter: J. Opt. Soc. Am. *60*, 306-314 (1970)
5.97 W.H. Carter, P.C. Ho: Appl. Opt. *13*, 162-172 (1974)
5.98 P.C. Ho, W.H. Carter: Appl. Opt.*15*, 313-314 (1976)
5.99 J. Radon: Ber. Verh. Sächs. Akad. Wiss. Leipzig, Math. Phys. Kl. *69*, 262-271 (1917)
5.100 K.T. Smith, D.C. Solomon, S.L. Wagner: Bull. AMS *83*, 1227-1270 (1977)
5.101 L.A. Shepp, J.B. Kruskal: Am. Math. Monthly *85*, 420-439 (1978)
5.102 R.N. Bracewell, A.C. Riddle: The Astrophys. J. *150*, 427-434 (1967)
5.103 R.A. Crowther, D.J. Rosier, A. Klug: Proc. Roy. Soc. London A-*317*, 319-340 (1970)
5.104 Y. Das, W.M. Boerner: IEEE Trans. AP-*26*, 274-279 (1978)
5.105 M.V. Berry, D.F. Gibbs: Proc. Roy. Soc. London A-*314*, 143-152 (1970)
5.106 L. Colin (ed): "Mathematics of Profile Inversion", NASA TM-X-62.150 (1972)
5.107 E.C. Titchmarsh: *Introduction to the Theory of Fourier Integrals*, 2nd ed. (Oxford Universtiy Press, Oxford 1948)
5.108 J.G. McWhirter, E.R. Pike: J. Phys. A-*11*, 1729-1745 (1978)
5.109 J.G. McWhirter, E.R. Pike: Phys. Scr. *19*, 417-425 (1979)
5.110 G.A. Deschamps, H.S. Cabayan: IEEE Trans. AP-*20*, 268-274 (1972)
5.111 H.G. Schmidt-Weinmar: "Spatial Resolution of Subwavelength Sources from Optical Far-Zone Data", in *Inverse Source Problems in Optics*, ed. by H.P. Baltes, Topics in Current Physics, Vol. 9 (Springer, Berlin, Heidelberg, New York 1978) pp. 83-118
5.112 J.R. Cannon, K. Miller: J. SIAM Numer. Anal. B-*2*, 87-98 (1965)
5.113 K. Miller, G.A. Viano: Nucl. Phys. B-*35*, 460-470 (1971)
5.114 G. Alessandrini: "An Extrapolation Problem for Harmonic Functions", Bollettino U.M.I. (to be published)
5.115 G. Alessandrini: "On Differentiation of Approximately Given Functions", preprint Istituto Matematico U. Dini, Firenze (1979)
5.116 F. John: Ann. Mat. Pura Appl. *40*, 129-142 (1955)
5.117 G.C. Sherman, J.J. Stamnes, E. Lalor: J. Math. Phys. *17*, 760-776 (1976)
5.118 H.P. Baltes, H.G. Schmidt-Weinmar: Phys. Lett. *60A*, 275-277 (1977)
5.119 H.G. Schmidt-Weinmar, W.B. Ramsay: Appl. Phys. *14*, 175-181 (1977)
5.120 H.G. Schmidt-Weinmar: Can. J. Phys. *55*, 1102-1114 (1977)
5.121 J.T. Foley, E. Wolf: J. Opt. Soc. Am. *69*, 761-764 (1979)

altes, B. Steinle: J. Opt. Soc. Am. *69*, 910 (1979)
am, H.G. Schmidt-Weinmar, A. Wouk: Can. J. Phys. *54*, 1925-1936 (1976)
ercher, H. Bartelt, H. Becker, E. Wiltschko: Appl. Opt. *18*, 2427-2439