

Linear inverse problems with discrete data: II. Stability and regularisation

M Bertero[†], C De Mol[‡] and E R Pike[§]

[†] Dipartimento di Fisica dell'Università and Istituto Nazionale di Fisica Nucleare, I-16146 Genova, Italy

[‡] Département de Mathématique, Université Libre de Bruxelles, B-1050 Bruxelles, Belgium

[§] Department of Physics, King's College, London and Centre for Theoretical Studies, RSRE, Malvern, WR14 3PS, England

Received 30 October 1987

Abstract. In the first part of this work a general definition of an inverse problem with discrete data has been given and an analysis in terms of singular systems has been performed. The problem of the numerical stability of the solution, which in that paper was only briefly discussed, is the main topic of this second part. When the condition number of the problem is too large, a small error on the data can produce an extremely large error on the generalised solution, which therefore has no physical meaning. We review most of the methods which have been developed for overcoming this difficulty, including numerical filtering, Tikhonov regularisation, iterative methods, the Backus–Gilbert method and so on. Regularisation methods for the stable approximation of generalised solutions obtained through minimisation of suitable seminorms (*C*-generalised solutions), such as the method of Phillips, are also considered.

1. Introduction

In the first part of this work [1, hereafter referred to as I] we have given the physical motivations of the following definition of a linear inverse problem with discrete data: *given the Hilbert space X , given the functions $\varphi_n \in X$, $n = 1, \dots, N$, and given the (real or complex) numbers g_n , $n = 1, \dots, N$, find a function $f \in X$ such that*

$$(f, \varphi_n)_X = g_n \quad n = 1, \dots, N. \quad (1.1)$$

The data g_n can be viewed as the components of a vector \mathbf{g} which belongs to an N -dimensional euclidean space Y whose scalar product can in general be defined as follows (see I)

$$(\mathbf{g}, \mathbf{h})_Y = \sum_{n, m=1}^N W_{nm} g_n h_m^* \quad (1.2)$$

where the star denotes complex conjugation and the weights W_{nm} are the matrix elements of a given positive definite matrix \mathbf{W} . Then, if we introduce the following operator L

$$(Lf)_n = (f, \varphi_n)_X \quad n = 1, \dots, N \quad (1.3)$$

which transforms a function of the Hilbert space X into a vector of Y , equation (1.1) takes the form

$$Lf = g. \quad (1.4)$$

The solution of equation (1.4) may not exist if the φ_n are not linearly independent. Furthermore the solution of (1.4) is not unique because the null space of L , $\mathcal{N}(L)$, is the infinite-dimensional orthogonal complement of the subspace spanned by the functions φ_n . We denote this subspace by $X_{N'}$ where $N' \leq N$ is the number of linearly independent φ_n .

The lack of existence and uniqueness can always be remedied by looking for a least-squares solution of minimal norm. This solution, denoted by f^+ and called the *generalised solution* of the problem, always exists and is unique. As shown in I, it can be expressed in terms of the *singular system* of the operator L , which is defined as the set of the triples $\{\alpha_k, u_k, v_k\}$, $k=1, \dots, N'$, which solve the following coupled equations

$$Lu_k = \alpha_k v_k \quad L^* v_k = \alpha_k u_k \quad (1.5)$$

where $L^*: Y \rightarrow X$ is the adjoint operator of L . The positive numbers α_k are called the singular values of the problem and are ordered in a decreasing sequence. For simplicity, we will assume throughout this paper that $N' = N$, i.e. that the functions φ_n are linearly independent. The case $N' < N$ can be treated straightforwardly and all formulae relative to this case can be easily deduced from the corresponding formulae relative to the case $N' = N$. With this assumption, the functions u_k , called the singular functions, form an orthonormal basis in X_N . Analogously the singular vectors v_k form an orthonormal basis in Y . It follows that the representation of f^+ in terms of the singular system is

$$f^+ = \sum_{k=1}^N \alpha_k^{-1} (g, v_k)_Y u_k. \quad (1.6)$$

Notice that a mapping $L^+: Y \rightarrow X$ is then defined by the relation

$$f^+ = L^+ g. \quad (1.7)$$

The operator L^+ is called the *generalised inverse* of L .

If the quantity

$$C(L) = \|L\| \|L^+\| = \alpha_1 / \alpha_N \quad (1.8)$$

the *condition number* of the problem, is large, then the generalised solution f^+ is affected by numerical instability. In practical inverse problems, the data are measured and hence affected by noise. In an ill-conditioned situation, even small errors on the data vector g can produce a completely different and unphysical generalised solution. Similar considerations apply to the so-called C -generalised solution (see I, § 4) which are least-squares solutions minimising some suitable seminorm.

In § 2 we give the general definition of a regularisation method for the approximate and stable computation of the generalised solution of a linear inverse problem with discrete data. In § 3 we introduce a wide class of regularisers based on the use of spectral windows or, in other words, based on a filtering of the representation of the generalised solution in terms of the singular system of the problem. In § 4 we discuss in detail a particular regularisation algorithm in this class, namely the Tikhonov regulariser, while in § 5 we investigate the regularising properties of several iterative

methods. In § 6 a short account is given of some general methods for the choice of the regularisation parameter. Section 7 is devoted to the regularisation of C -generalised solutions and finally the Backus–Gilbert method is described in § 8.

2. Regularisation algorithms

In §§2–6 we only consider generalised solutions. The case of C -generalised solutions is deferred to §7.

Let us point out first that the usual definition of a regularisation algorithm for an ill-posed problem [2–4] needs some modification in the case of an inverse problem with discrete data. The reason is that the usual definition applies to the case of a linear operator whose inverse or generalised inverse is not continuous. Then a family of regularising operators is essentially a family of bounded operators which approximate, in some suitable sense, an operator which is not bounded. On the other hand, in the case of a problem with discrete data, the generalised inverse is always continuous since it is a linear operator on a finite dimensional space. The need to regularise the problem is felt when the norm of L^+ is much greater than $1/\|L\|$, i.e. when the problem is ill-conditioned. It follows that a regularisation algorithm must provide an approximation of L^+ which has a smaller norm than L^+ .

Guided by these considerations, we give the following definition of a regularisation algorithm (or regulariser): we say that a one-parameter family of linear operators $\{R_\mu\}_{\mu>0}$, $R_\mu: Y \rightarrow X$, is a *regularisation algorithm* for the approximate computation of the generalised inverse L^+ , if the following conditions are satisfied:

(i) for any $\mu > 0$, the range of R_μ is contained in X_N , the subspace spanned by the functions φ_n ;

(ii) for any $\mu > 0$, the norm of R_μ is smaller than the norm of L^+ , i.e.

$$\|R_\mu\| \leq \|L^+\| = \alpha_N^{-1}; \tag{2.1}$$

(iii) the following property holds true

$$\lim_{\mu \rightarrow 0} R_\mu = L^+. \tag{2.2}$$

As in the case of regularisation methods for ill-posed problems, the variable μ will be called the *regularisation parameter* and one of the main problems in regularisation theory will be the choice of an appropriate value of μ . In general the chosen value results from some suitable compromise between the approximation error and the data error propagation. We will now clarify this point.

A noisy data vector g_t can always be represented in the following way

$$g_t = Lf^+ + h_t \tag{2.3}$$

where the first term of the RHS represents the noise-free signal (f^+ is the true generalised solution) and the second term is a small vector representing the effect of the noise. The quantity ε will denote an estimate of the norm of h_t in Y , so that

$$\|Lf^+ - g_t\|_Y \leq \varepsilon. \tag{2.4}$$

Notice that in writing equation (2.3), we do not assume the noise to be additive. The noise term h_t can be signal dependent. We only require that the inequality (2.4) holds true.

In a concrete problem the decomposition (2.3) is not known, but one can always assume that such a decomposition exists. Then, for a given choice of the regularisation algorithm and for a given value of the regularisation parameter μ , the approximation of the generalised solution f^+ provided by R_μ is $f_{\mu,\epsilon} = R_\mu g_\epsilon$. If we compare this approximation with f^+ , we obtain

$$f_{\mu,\epsilon} - f^+ = (R_\mu L f^+ - f^+) + R_\mu h_\epsilon. \tag{2.5}$$

Here the first term of the rhs represents the approximation error introduced by the use of the algorithm R_μ (with $\mu \neq 0$) instead of L^+ . This approximation error tends to zero when $\mu \rightarrow 0$, as follows from equation (2.2). On the other hand the second term is the error induced by the noise and it grows to unacceptable values when $\mu \rightarrow 0$. Therefore one must look for a compromise between the approximation error and the error propagation from the data to the solution. This problem will be discussed in detail in §6.

We notice that any regularisation algorithm defines, for a given μ , a mapping A_μ in X given by

$$A_\mu = R_\mu L. \tag{2.6}$$

The function $A_\mu f^+$ is the approximation of the generalised solution f^+ which can be obtained, in the absence of noise, by means of the regularisation operator R_μ . In the limit $\mu \rightarrow 0$, A_μ converges to the operator A defined in I (§2) which is given in terms of the generalised inverse by

$$A = L^+ L. \tag{2.7}$$

Clearly A is a projection operator, namely, the orthogonal projection onto the subspace X_N spanned by the functions φ_n .

It was shown in I that, when X is a weighted L^2 space with a scalar product defined by

$$(f, h)_X = \int w(x) f(x) h^*(x) dx \tag{2.8}$$

then A is an integral operator whose kernel can be used in order to define the resolution achievable by means of the available data. Similar considerations apply to the operator A_μ . Assume indeed that R_μ is a linear operator. In order to satisfy condition (i), it must be of the following form

$$(R_\mu g)(x) = \sum_{n,m=1}^N \rho_\mu^{nm} g_n \varphi_m(x). \tag{2.9}$$

Furthermore condition (iii) is satisfied if and only if

$$\lim_{\mu \rightarrow 0} \rho_\mu^{nm} = G^{nm} \tag{2.10}$$

where the G^{nm} are the matrix elements of the inverse of the Gram matrix of elements $G_{mn} = (\varphi_m, \varphi_n)_X$ (see I, § 2). From equations (2.9) and (2.8) we have

$$\begin{aligned} (R_\mu L f)(x) &= \sum_{n,m=1}^N \rho_\mu^{nm} (f, \varphi_n)_X \varphi_m(x) \\ &= \int w(x') \left(\sum_{n,m=1}^N \rho_\mu^{nm} \varphi_m(x) \varphi_n^*(x') \right) f(x') dx' \end{aligned} \tag{2.11}$$

and therefore A_μ is an integral operator with kernel

$$A_\mu(x, x') = \sum_{n,m=1}^N \rho_\mu^{nm} \varphi_m(x) w(x') \varphi_n^*(x'). \tag{2.12}$$

We conclude that, at least under the circumstances specified above, one can associate an averaging kernel with a regularisation algorithm. When, for a given x , this kernel has a central lobe concentrated around x and flanked by decreasing side-lobes, the half-width of the central lobe can be used as a measure of the resolution achievable at point x by means of the regularisation operator R_μ . The averaging kernel $A_\mu(x, x')$ is a sort of *impulse response* describing the global effect of both the transmission by the instrument and the subsequent recovery procedure by means of the algorithm R_μ . In fact the action of the operator L in (2.6) represents the error-free measurement of the values of the functionals associated with the functions φ_n . Then the action of the second operator R_μ represents the inversion of this data (neglecting the effect of round-off errors) at the given parameter value μ .

3. Spectral windows and numerical filtering

The spectral representation of L^+ , in terms of the singular system of L , is given by equation (1.6). In this section we consider a class of regularisation algorithms which can be viewed as a windowing of this representation. Such algorithms can be generally expressed as following

$$R_\mu \mathbf{g} = \sum_{k=1}^N W_{\mu,k} \alpha_k^{-1} (\mathbf{g}, \mathbf{v}_k)_Y u_k \tag{3.1}$$

where $\{W_{\mu,k}\}$ is a set of windowing coefficients which depend on a positive parameter μ and satisfy the following conditions:

- (i') $0 \leq W_{\mu,k} \leq 1$, for any $\mu > 0$ and any $k = 1, \dots, N$;
- (ii') for any $k = 1, \dots, N$, the following limit holds true

$$\lim_{\mu \rightarrow 0} W_{\mu,k} = 1.$$

Noticing that condition (i) of § 2 is satisfied by definition, that condition (i') implies condition (ii) as follows from the inequality

$$\|R_\mu\| = \max_k (W_{\mu,k} / \alpha_k) \leq \alpha_N^{-1} \tag{3.2}$$

and finally that condition (ii') implies condition (iii), we conclude that equation (3.1) defines a regularisation algorithm.

The simplest example of a regularisation algorithm belonging to this class is provided by the rectangular (or top-hat) window

$$W_{\mu,k} = 1, \text{ if } k \leq [1/\mu] \quad W_{\mu,k} = 0, \text{ if } k > [1/\mu] \quad (3.3)$$

where $[1/\mu]$ denotes the maximum integer contained in $1/\mu$. It is obvious that conditions (i') and (ii') are satisfied by the coefficients (3.3) and that this window amounts to taking in equation (1.6) a number of terms smaller than N when $\mu > N^{-1}$, while $R_\mu = L^+$ when $\mu \leq N^{-1}$. This method is one of the most frequently used in practice and it is also known as *numerical filtering* [5]. When combined with a suitable method for the choice of the number of terms, this method has an interesting physical interpretation [6–8] which will be discussed in § 6.

Another interesting example is provided by the triangular window

$$W_{\mu,k} = 1 - (k - 1)/[1/\mu], \text{ if } k \leq [1/\mu] \quad W_{\mu,k} = 0, \text{ if } k > [1/\mu]. \quad (3.4)$$

It is easy to verify that conditions (i') and (ii') are satisfied also in this case.

Let us stress an important property of this window in the particular case where the functionals (1.1) are the Fourier coefficients of f . It is well known that such a window appears when Fourier series are treated using Cesàro summation by arithmetic means. This is the Fejèr theory of Fourier series [9]. As in I, let us consider the case where the functionals are given by

$$(f, \varphi_n)_X = (2\pi)^{-1} \int_{-\pi}^{\pi} f(x) \exp(-inx) dx \quad n = 0, \pm 1, \dots, \pm N \quad (3.5)$$

and where $X = L^2(-\pi, \pi)$. Then the averaging kernel associated with the rectangular window, in the case $\mu = N^{-1}$, is just the Dirichlet kernel

$$A(x, x') = (2\pi)^{-1} \sin[(N + \frac{1}{2})(x - x')]/\sin[\frac{1}{2}(x - x')] \quad (3.6)$$

while the averaging kernel associated with the triangular window, also in the case $\mu = N^{-1}$, is the Fejèr kernel

$$A(x, x') = [2\pi(N + 1)]^{-1} \sin^2[\frac{1}{2}(N + 1)(x - x')]/\sin^2[\frac{1}{2}(x - x')]. \quad (3.7)$$

The latter kernel is positive and therefore, in the absence of noise, it provides a positive approximation of any positive function [10]. The width of its central peak, however, is twice the width of the central peak of the kernel (3.6) and therefore positivity is apparently obtained at the price of a loss in resolution although in practice this is not necessarily the case since more terms may be used in the presence of a given level of noise.

4. The Tikhonov regularisation algorithm

A special kind of regularisation algorithm was introduced by Tikhonov [11] for the approximate solution of first-kind Fredholm integral equations. A similar method which will be discussed in § 7, was also proposed by Phillips [12].

In the case where the object f is a square integrable function defined on some interval $[a, b]$, the algorithm of Tikhonov can be obtained by minimising, for each μ , the following functional

$$\Phi_\mu[f] = \|Lf - \mathbf{g}\|_Y^2 + \mu\Omega[f] \tag{4.1}$$

where μ is the regularisation parameter and where the penalty functional $\Omega[f]$, also called the regularising functional, is given by

$$\Omega[f] = \sum_{i=0}^M \int_a^b p_i(x) |f^{(i)}(x)|^2 dx. \tag{4.2}$$

The p_i are positive continuous functions and $f^{(i)} = d^i f/dx^i$.

Functionals of the type (4.2) are particular cases of more general functionals of the form

$$\Omega[f] = \|Cf\|_Z^2 \tag{4.3}$$

where C is a linear operator transforming a function f of the Hilbert space X into a function of a new Hilbert space Z , which can be called the *constraint space*. When the operator C is closed and has a bounded inverse, i.e. when the equation $Cf=0$ has the unique solution $f=0$ and furthermore the range of C is closed, we can consider a new object space X which is defined as the domain of C equipped with the norm: $\|f\|_X = \|Cf\|_Z$. Then it is easy to see that, thanks to the assumptions on C , the new object space is also a Hilbert space. The previous conditions are satisfied by the functional (4.2) when $p_0(x) \neq 0$ everywhere. They are not satisfied in the case where the operator C has a nontrivial null space, i.e. the equation $Cf=0$ has nontrivial solutions. In such a case the functional (4.3) defines a seminorm and not a norm. This case, which is related to the method of Phillips [12], will be considered in § 7.

The previous remarks, combined with the fact that we can modify the choice of the object space X , as discussed in I, show that we can restrict the analysis to the following case

$$\Omega[f] = \|f\|_X^2 \tag{4.4}$$

so that the functional (4.1) becomes

$$\Phi_\mu[f] = \|Lf - \mathbf{g}\|_Y^2 + \mu\|f\|_X^2. \tag{4.5}$$

Since $\{u_k\}$ is an orthogonal basis in X_N while $\{v_k\}$ is an orthogonal basis in Y , we see that

$$f = \sum_{k=1}^N (f, u_k)_X u_k + f^{(0)} \tag{4.6}$$

where $f^{(0)}$ is the component of f which is orthogonal to the subspace X_N spanned by the functions φ_n , and also that

$$Lf - \mathbf{g} = \sum_{k=1}^N [\alpha_k(f, u_k)_X - (\mathbf{g}, \mathbf{v}_k)_Y] \mathbf{v}_k. \quad (4.7)$$

It follows that the functional (4.5) can be written in the following form

$$\Phi_\mu[f] = \sum_{k=1}^N (|\alpha_k(f, u_k)_X - (\mathbf{g}, \mathbf{v}_k)_Y|^2 + \mu |(f, u_k)_X|^2) + \mu \|f^{(0)}\|_X^2 \quad (4.8)$$

and therefore the element minimising $\Phi_\mu[f]$ belongs to X_N and is given by

$$f_\mu = R_\mu \mathbf{g} = \sum_{k=1}^N \alpha_k (\alpha_k^2 + \mu)^{-1} (\mathbf{g}, \mathbf{v}_k)_Y u_k. \quad (4.9)$$

This equation defines a regularisation algorithm which has the form (3.1) with

$$W_{\mu,k} = \alpha_k^2 (\alpha_k^2 + \mu)^{-1}. \quad (4.10)$$

This set of windowing coefficients can be called the Tikhonov window and it is easy to see that they satisfy the conditions (i') and (ii') of § 2. Therefore equation (4.9) defines a regularisation algorithm.

It is interesting to notice that, in the case of a problem with discrete data, the computation of the Tikhonov regulariser can be reduced to a matrix inversion and therefore that the knowledge of the singular system of the problem is not required. By annihilating the first variation of the functional (4.5), one finds indeed that the regularising operator R_μ defined by equation (4.9) is also given by

$$R_\mu = (L^*L + \mu I)^{-1} L^* \quad (4.11)$$

where the operator L^*L is an operator in X . Then, using the following obvious relation

$$(L^*L + \mu I)L^* = L^*(LL^* + \mu I) \quad (4.12)$$

and observing that the operator $(LL^* + \mu I)$ is always invertible, one finds that the operator (4.11) can also be written in the following form

$$R_\mu = L^*(LL^* + \mu I)^{-1}. \quad (4.13)$$

Since the operator $(LL^* + \mu I)$ is an operator in Y , it is associated with an $N \times N$ matrix and therefore the computation of the operator R_μ implies a matrix inversion, followed by the application of the operator L^* .

The problem of the choice of the regularisation parameter μ in the case of the Tikhonov regulariser has been the object of many studies. Here we limit ourselves to a summary of the main methods. We also point out that some of these methods are not specifically designed for the determination of the regularisation parameter in the case of the Tikhonov regulariser but may be applied as well to more general situations.

4.1. Constrained least-squares solutions

We will not present here the general formulation of the method of constrained least-squares solutions introduced by Ivanov [13]. We just sketch the idea on which this

method relies, with reference to the problems we are considering in this paper. As discussed in I, a linear inverse problem with discrete data has an infinite-dimensional set of least-squares solutions. Only one of them has minimal norm and this is the generalised solution. We assume now that the problem is badly ill-conditioned and we consider the set of the generalised solutions corresponding to all data vectors g_ϵ lying inside a sphere of radius ϵ whose centre is the noise-free data vector g . As a consequence of the ill-conditioning of the problem, this set is a very prolonged ellipsoid. In fact the minimum and maximum semi-axes of this ellipsoid are proportional to α_1^{-1} and α_N^{-1} respectively. It follows that most of these generalised solutions are not good approximations of the generalised solution corresponding to the noise-free data g . Usually they are wildly oscillating functions which take completely unphysical values. Therefore the basic idea is to restrict the class of admissible solutions by looking for approximate solutions in some prescribed set. For example, if the square of the unknown function f has the physical meaning of an energy density, the integral of f^2 is the total energy and in some instances one knows beforehand that the energy must be smaller than some prescribed value, say E^2 .

Let H be the set of the functions f satisfying the physical constraints. It may happen that there does not exist any solution of equation (1.4) in the set H . Then it is quite natural to look for the functions $f \in H$ such that the distance between Lf and g is minimal, i.e. to solve the problem

$$\|Lf - g\|_Y = \text{minimum} \quad f \in H. \tag{4.14}$$

This is a least-squares problem with constraints. Its solution, if it exists and is unique, can be called the H -constrained least-squares solution of the problem (1.4).

The simplest case occurs when the set H is a sphere in X with a prescribed radius E . Then the problem (4.14) becomes

$$\|Lf - g\|_Y = \text{minimum} \quad \|f\|_X \leq E. \tag{4.15}$$

The solution of this problem is not unique if the generalised solution f^+ satisfies the constraint, i.e. $\|f^+\|_X \leq E$. In such a case the set of the H -constrained least-squares solutions is the intersection of the set of the unconstrained least-squares solutions—see I—with the sphere H and there exists a unique H -constrained least squares solution of minimal norm which coincides with f^+ .

The previous situation, however, is quite unlikely in the case of noisy data. When the generalised solution has a norm greater than E , i.e. $\|f^+\|_X > E$, then the intersection of the sphere of radius E with the set of the solutions, or least-squares solutions, of equation (1.4) is empty. Under these circumstances, the minimum points of $\|Lf - g\|_Y$ cannot be interior to the sphere of radius E but they must lie on the surface of this sphere, i.e. they must satisfy the condition $\|f\|_X = E$. It follows that one can use the method of the Lagrange multipliers for determining the solution of problem (4.15). But this method is just equivalent to minimising the functional (4.5) for any μ and then looking for a minimum point f_μ satisfying the condition

$$\|f_\mu\|_X = E. \tag{4.16}$$

If there exists a unique value of μ which solves this equation, this means that there exists a unique solution of the original constrained least-squares problem (4.15). Furthermore, by comparison with the previous discussion of the Tikhonov regulariser, it clearly appears that equation (4.16) can also be interpreted as a rule for selecting the value of the regularisation parameter: among all the regularised solutions satisfying a

prescribed bound, choose the regularised solution that provides the best approximation of the experimental data.

In order to prove that there exists a unique value of μ satisfying equation (4.16), we notice that from equation (4.9) we obtain the following expression for the norm $\nu(\mu)$ of the minimum point of the functional (4.5)

$$\nu^2(\mu) = \|f_\mu\|_X^2 = \sum_{k=1}^N \alpha_k^2 (\alpha_k^2 + \mu)^{-2} |(\mathbf{g}, \mathbf{v}_k)_Y|^2. \quad (4.17)$$

It follows that $\nu(\mu)$ is a strictly decreasing function of μ and that its value at $\mu=0$ is $\nu(0) = \|f^+\|_X > E$ while its value at $\mu=\infty$ is zero. As a consequence there exists a unique value of μ , say μ_0 , satisfying equation (4.16). The corresponding constrained least-squares solution will be denoted by f_0 .

Assume now that we have a sequence of noisy data \mathbf{g}_ε , given by equation (2.3), and let f_ε^+ , $f_{0,\varepsilon}$ be respectively the corresponding generalised solutions and constrained least-squares solutions. Then, if the error ε tends to zero and if the generalised solution f^+ corresponding to the noise-free data satisfies the constraint (i.e. if the choice of the upper bound E is physically correct), since f_ε^+ converges to f^+ we must have, for a certain value of ε , $\|f_\varepsilon^+\|_X = E$ and therefore $f_\varepsilon^+ = f_{0,\varepsilon}$. It is not restrictive to assume that, for smaller values of ε , the condition $\|f_\varepsilon^+\|_X \leq E$ is always satisfied. Then the solution of the problem (4.15) is not unique and the solution of minimal norm is just f_ε^+ . For all these values of ε we can redefine $f_{0,\varepsilon}$ by putting $f_{0,\varepsilon} = f_\varepsilon^+$ and then we may conclude that the constrained least-squares solution $f_{0,\varepsilon}$ converges, when ε tends to zero, to the generalised solution f^+ corresponding to noise-free data.

However, if the constraint is not physically correct or more precisely if the constant E has been underestimated so that the generalised solution corresponding to noise-free data satisfies the condition $\|f^+\|_X > E$, then $f_{0,\varepsilon}$ does not converge to f^+ when $\varepsilon \rightarrow 0$. The limit will be the constrained least-squares solution corresponding to noise-free data.

4.2. Regularised solutions of minimal norm and the discrepancy principle

The unpleasant feature of the previous method is that it requires the knowledge of a reasonable upper bound on the solution of the problem. In many practical circumstances, however, this upper bound is not available while it is possible to estimate an upper bound on the experimental errors. This remark leads us to consider a problem which is the dual of the problem considered in the previous subsection and which can be formulated as follows [14]. Assume that an upper bound ε on the Y norm of the error is known and denote by $J_\varepsilon(\mathbf{g})$ the set of all the elements of X which are compatible with the data \mathbf{g} up to accuracy ε , i.e. the set of all functions f satisfying the condition

$$\|Lf - \mathbf{g}\|_Y \leq \varepsilon. \quad (4.18)$$

This set is not bounded, since the null space of L is infinite dimensional, but its projection on X_N is bounded. The important property of $J_\varepsilon(\mathbf{g})$ is that it is closed and convex and therefore, thanks to a well known theorem of functional analysis (see, for example, [15], p 9) there exists in $J_\varepsilon(\mathbf{g})$ a unique element of minimal norm. In other

words, the problem

$$\|f\|_X = \text{minimum} \quad \|Lf - g\|_Y \leq \varepsilon \tag{4.19}$$

which is just the dual of problem (4.15), has always a unique solution. This solution is not the null element of X , provided that the data vector satisfies the condition $\|g\|_Y > \varepsilon$. This inequality is quite reasonable, since it means that the data are greater than the noise. If it is not satisfied, the data contain only noise and therefore they do not contain any information about f . In such a case the solution of problem (4.19) is just $f=0$.

In order to obtain explicitly the unique solution of problem (4.19) when $\|g\|_Y > \varepsilon$, let us first observe that the solution must necessarily belong to X_N since it is a solution of minimal norm (the component belonging to the null space of L must be zero). Then it is quite easy to see that the solution of minimal norm must lie on the boundary of the bounded set which is the intersection of X_N with $J_\varepsilon(g)$. As a consequence, the solution we are looking for satisfies the condition $\|Lf - g\|_Y = \varepsilon$ and one can use the method of Lagrange multipliers for determining the solution of problem (4.19). Again we must minimise the functional (4.5) for any μ and then look for a minimum point f_μ satisfying the condition

$$\|Lf_\mu - g\|_Y = \varepsilon. \tag{4.20}$$

Thanks to the previous remarks we already know that there must exist a unique value of μ allowing (4.20) to be satisfied. It is useful, however, to check directly this result.

From the representation (4.9) of the regularised solution f_μ , we obtain

$$\rho^2(\mu) = \|Lf_\mu - g\|_Y^2 = \mu^2 \sum_{k=1}^N (\alpha_k^2 + \mu)^{-2} (g, v_k)_Y^2 \tag{4.21}$$

and this is a strictly increasing function of μ , which is zero at $\mu=0$ and tends to $\|g\|_Y^2$ when $\mu \rightarrow \infty$. Therefore, when the condition $\|g\|_Y > \varepsilon$ is satisfied, there exists a unique value of μ , say μ_1 , which solves equation (4.20). The corresponding regularised solution will be denoted by f_1 .

If we consider now the case of a sequence of noisy data g_ε , it is obvious that, when the error ε tends to zero, the corresponding value of the regularisation parameter $\mu_1 = \mu_1(\varepsilon)$ also tends to zero. As a consequence, the sequence of regularised solutions, $f_{1,\varepsilon}$, obtained by means of this method always converges to the generalised solution corresponding to noise-free data.

The method considered in this subsection is a particular case of a general method for the choice of the regularisation parameter which is known as the *discrepancy principle* [16]. The function $\rho(\mu)$, defined by (4.21) and representing the norm of the difference between the measured data and the data which can be computed from the regularised function f_μ , is usually called the *discrepancy function*.

4.3. Regularised solutions satisfying prescribed bounds

The method of § 4.1 requires a prescribed bound on the solution while the method of § 4.2 requires a prescribed bound on the data error. In a paper by Miller [17], the case where both bounds are known is also considered. Let us assume that two constants ε ,

E are given and that one looks for a function f such that

$$\|Lf - g\|_Y \leq \varepsilon \quad \|f\|_X \leq E. \tag{4.22}$$

Any function satisfying these conditions can be called an *admissible solution* of the problem. Therefore the conditions (4.22) define a set \mathbf{K} of admissible solutions and the first problem is to find a condition ensuring that this set is not empty.

For this purpose we introduce the functional

$$\Phi[f] = \|Lf - g\|_Y^2 + (\varepsilon/E)^2 \|f\|_X^2 \tag{4.23}$$

which is just the functional (4.5) with $\mu = (\varepsilon/E)^2$, and we consider the set \mathbf{K}_0 which is the set of all the functions f satisfying the condition

$$\Phi[f] \leq \varepsilon^2. \tag{4.24}$$

Since any function satisfying this condition also satisfies the conditions (4.22), we have the inclusion $\mathbf{K}_0 \subset \mathbf{K}$ and the set \mathbf{K} is not empty if \mathbf{K}_0 is not empty. This is true if and only if $\Phi[f_2] \leq \varepsilon^2$, where f_2 is the function which minimises the functional (4.23) and is given by equation (4.9) with $\mu = (\varepsilon/E)^2$. In such a case $f_2 \in \mathbf{K}$. It follows that the regularised solution corresponding to this choice of the regularisation parameter is an admissible solution in the sense specified above.

4.4. Generalised cross-validation

The previous methods for the choice of the regularisation parameter cannot be used if no upper bound on the error or on the solution is known. A method which does not require such information has been proposed and is currently used in applications of the Tikhonov regulariser to inverse problems with discrete data and in particular to smoothing problems. This method, which is known as *generalised cross-validation* [18, 19], is based on the idea of letting the data themselves choose the value of the regularisation parameter. It is required that a good value of the regularisation parameter should allow the prediction of missing data values. In such a way no *a priori* information about the solution and/or the noise is required.

Let $f_{\mu,k}$ be the minimiser of the functional

$$\Phi_{\mu,k}[f] = N^{-1} \sum_{n \neq k} |(Lf)_n - g_n|^2 + \mu \|f\|_X^2 \tag{4.25}$$

which is just the functional (4.5) with the k th data value missing. Here we assume that the Y norm is the usual euclidean norm, but the extension to the case of a norm with weights is easy. Then the *cross-validation function* $V_0(\mu)$ is defined by

$$V_0(\mu) = N^{-1} \sum_{k=1}^N |(Lf_{\mu,k}) - g_k|^2 \tag{4.26}$$

and the cross-validation method consists in determining the unique value of μ which minimises $V_0(\mu)$. The computation of the minimum is based on the relation [19]

$$V_0(\mu) = N^{-1} \sum_{k=1}^N |1 - \tilde{A}_{kk}(\mu)|^{-2} |(Lf_{\mu}) - g_k|^2 \tag{4.27}$$

where f_μ is the minimiser of the functional

$$\Phi_\mu[f] = N^{-1} \sum_{n=1}^N |(Lf)_n - g_n|^2 + \mu \|f\|_X^2 \tag{4.28}$$

which is a special case of the functional (4.5), and $\tilde{A}_{kk}(\mu)$ is the kk entry of the $N \times N$ matrix

$$[\tilde{A}(\mu)] = [LL^*][LL^* + \mu I]^{-1}. \tag{4.29}$$

Notice that $[LL^*]$ is essentially the Gram matrix of the functions φ_n since, in this formulation, $[W] = N^{-1}[I]$.

It has been shown [19] that, from the point of view of minimising the predictive mean-square error, the minimisation of $V_0(\mu)$ must be replaced by the minimisation of the *generalised cross-validation function* defined by

$$V(\mu) = (N^{-1} \text{Tr}[I - \tilde{A}(\mu)])^{-2} (N^{-1} \|[I - \tilde{A}(\mu)]\mathbf{g}\|^2) \tag{4.30}$$

where $\|\cdot\|$ denotes the usual euclidean norm. An important property of $V(\mu)$ is its invariance with respect to permutations, and more generally with respect to rotations, of the data values.

The determination of the value of μ which minimises $V(\mu)$ might seem a quite heavy computational problem. However very efficient and fast algorithms have been developed for the solution of this problem [19].

5. Iterative methods

It is known that several iterative methods, such as successive approximations [20, 21], steepest descent [22] and conjugate gradient [23], define regularising algorithms, in the sense of Tikhonov [3], for the approximate solution of ill-posed linear operator equations. These methods can be applied, in particular, to the solution of first kind Fredholm integral equations. Their formulation in the present case is rather natural, since they are methods for computing the minimum of the quadratic form $\|Lf - \mathbf{g}\|_Y$. The Euler equation associated with this minimum problem is

$$L^*Lf = L^*\mathbf{g}. \tag{5.1}$$

When the functions φ_n are linearly independent, the minimum value of the functional $\|Lf - \mathbf{g}\|_Y$ is zero and the normal solution of equation (5.1) coincides with the normal solution of equation (1.4).

Since the operator L^*L is a finite rank operator in X , the iterative methods mentioned above converge to the normal solution of equation (5.1) for any data vector \mathbf{g} , provided that the initial function of the iterative scheme is properly chosen. Equation (5.1), however, is a functional equation and it is convenient to transform it into an equivalent matrix equation. We give here the basic formulae for the three methods.

If f_n denotes the approximation of the normal solution f^+ provided by the n th step of the iteration scheme and if r_n denotes the residual

$$r_n = L^*Lf_n - L^*\mathbf{g} \tag{5.2}$$

then we put

$$f_n = L^* f_n \quad (5.3)$$

and also

$$r_n = L^* r_n \quad r_n = \hat{L} f_n - g \quad (5.4)$$

where f_n and r_n are N -dimensional vectors. As in I, $\hat{L} = LL^*$ and in the case where Y is the usual euclidean space, \hat{L} is the Gram matrix of the functions φ_n .

The algorithms mentioned above can now be formulated as follows.

5.1. Successive approximations

$$f_0 = 0 \quad f_{n+1} = f_n - \tau r_n \quad (5.5)$$

where τ , the relaxation parameter, can be any number satisfying the conditions

$$0 < \tau < 2/\alpha_1^2. \quad (5.6)$$

5.2. Steepest descent

$$f_0 = 0 \quad f_{n+1} = f_n - \tau_n r_n \quad (5.7)$$

where

$$\tau_n = (\hat{L} r_n, r_n)_Y / (\hat{L}^2 r_n, r_n)_Y. \quad (5.8)$$

5.3. Conjugate gradient

$$f_0 = 0 \quad f_{n+1} = f_n - \tau_n p_n \quad (5.9)$$

where

$$p_0 = r_0 = -g \quad p_n = r_n + \sigma_{n-1} p_{n-1} \quad (5.10)$$

and

$$\tau_n = (\hat{L} r_n, p_n)_Y / (\hat{L}^2 p_n, p_n)_Y \quad (5.11)$$

$$\sigma_{n-1} = -(\hat{L}^2 r_n, p_{n-1})_Y / (\hat{L}^2 p_{n-1}, p_{n-1})_Y. \quad (5.12)$$

It is not difficult to show that these three iterative algorithms have the following properties:

(i'') the n th approximation f_n , given by equation (5.3), belongs to X_N and converges to f^+ when $n \rightarrow \infty$; furthermore the following inequalities hold true

$$\|r_{n+1}\|_X \leq \|r_n\|_X \quad (5.13)$$

$$\|f_{n+1} - f^+\|_X \leq \|f_n - f^+\|_X; \quad (5.14)$$

(ii'') for any n , $\|f_n\|_X \leq \|f_{n+1}\|_X$ and therefore $\|f_n\|_X \leq \|f^+\|_X$.

It follows then that the family of operators $\{R_n\}$ defined by

$$f_n = R_n g \tag{5.15}$$

satisfies the assumptions (i)–(iii) of §2 and therefore defines a regularisation algorithm. The inverse of the number of iterations plays here the role of the regularisation parameter. Notice that the operators R_n are linear in the case of successive approximations while they are nonlinear in the case of steepest descent and conjugate gradient.

The method of successive approximations defines a regularising algorithm which belongs to the class introduced in §3. From equations (5.3)–(5.5) one can derive, indeed, that

$$f_n = R_n g = \tau \sum_{m=0}^{n-1} (1 - \tau L^* L)^m L^* g \tag{5.16}$$

and, by introducing the singular system of the operator L , one obtains

$$f_n = \sum_{k=1}^N \alpha_k^{-1} [1 - (1 - \tau \alpha_k^2)^n] (g, v_k) v u_k. \tag{5.17}$$

This equation has the same structure as equation (3.1), with the following choice of the windowing coefficients

$$W_{n,k} = 1 - (1 - \tau \alpha_k^2)^n. \tag{5.18}$$

As explained above, the regularisation parameter is the inverse of the number of iterations.

We conclude this section by recalling that a well known method for band-limited extrapolation, the so-called Gerchberg–Papoulis algorithm [24, 25], can be considered as a special case of the method of successive approximations. This result was essentially contained in a paper of De Santis and Gori [26] and more explicitly shown by Maitre [27]. In the latter paper it was also proved, by means of numerical simulations, that the conjugate gradient method requires a much smaller number of iterations in order to produce the same accuracy as the Gerchberg–Papoulis algorithm. Similar results were obtained in the case of the Laplace transform inversion and in this problem, which is extremely ill-conditioned, a simple relationship between conjugate gradient and singular function expansions was also found [28]. An argument justifying this result is given by Natterer [29].

6. Choice of the regularisation parameter

In §4 we have already presented several methods for the choice of the regularisation parameter in the case of the Tikhonov regulariser. While the methods of §§4.3 and 4.4 are *ad hoc* methods for this algorithm, the methods of §§4.1 and 4.2 can also be applied to other regularisers. For this purpose we introduce, in the general case, two functions of the regularisation parameter which have been already introduced in §4 in the case of the Tikhonov regulariser. These are the norm of the regularised solution

$$v(\mu) = \|R_\mu g\|_X \tag{6.1}$$

and the discrepancy function

$$\rho(\mu) = \|LR_\mu \mathbf{g} - \mathbf{g}\|_Y. \quad (6.2)$$

We will assume that these functions satisfy the following conditions

(a) $\nu(\mu)$ is a strictly decreasing function of μ ; its value at $\mu = 0$ is determined by the condition (iii) of §2, i.e. $\nu(0) = \|f^-\|_X$;

(b) $\rho(\mu)$ is a strictly increasing function of μ ; its value at $\mu = 0$ is determined again by the condition (iii) of §2, i.e. $\rho(0) = 0$.

These conditions are satisfied by the Tikhonov regulariser, by the examples of regularisers given in §3 and by the iterative algorithms of §5.

Let us assume, as in §4.1, that, on the basis of physical considerations, we know an upper bound on the norm of the unknown solution f

$$\|f\|_X \leq E. \quad (6.3)$$

If the prescribed constant E is smaller than the norm of the generalised solution f^+ , then it follows from property (a) that there exists a unique value of μ , say μ_E , such that $\nu(\mu_E) = E$ and, for $\mu > \mu_E$, $\nu(\mu) < E$. Furthermore property (b) implies that, for $\mu > \mu_E$, we have $\rho(\mu) > \rho(\mu_E)$. We conclude that $\mu = \mu_E$ is the value of the regularisation parameter which gives an approximate solution compatible with the constraint (6.3) and minimises the discrepancy between the computed and the measured data.

On the other hand, let us now assume, as in §4.2, that the unknown function f satisfies the inequality

$$\|Lf - \mathbf{g}\|_Y \leq \varepsilon. \quad (6.4)$$

If the prescribed constant ε is smaller than the norm of \mathbf{g} , then property (b) implies that there exists a unique value of μ , say μ_ε , such that $\rho(\mu_\varepsilon) = \varepsilon$ and $\rho(\mu) < \varepsilon$ for $\mu < \mu_\varepsilon$. On the other hand, it follows from property (a) that, for $\mu < \mu_\varepsilon$, we have $\nu(\mu) > \nu(\mu_\varepsilon)$ and therefore the value $\mu = \mu_\varepsilon$ provides a regularised solution which is compatible with the data up to accuracy ε and has minimal norm. This second method, as already mentioned in §4.2, is usually called the *discrepancy principle*. In the case of ill-posed problems it has been shown [30] that this principle (more precisely a slightly modified form of it, which we do not need to mention here), when applied to the most important regularisation algorithms, always provides a regularised solution which converges to the true generalised solution of the problem as the errors on the data tend to zero. In the case of problems with discrete data, this result is almost obvious since property (b) implies then that $\mu_\varepsilon \rightarrow 0$ when $\varepsilon \rightarrow 0$.

We conclude this section by mentioning a method which applies to the case of the truncated singular function expansions given by (3.1) and (3.3), and is the analogue of the method discussed in §4.3. Assume that the unknown solution satisfies the constraints (4.22). Then, as shown by Miller [17], if we keep in the expansion (1.6) only those terms which correspond to singular values fulfilling the condition

$$\alpha_k \geq \varepsilon/E \quad (6.5)$$

the resulting truncated expansion satisfies the constraints (4.22) except for a factor of $\sqrt{2}$. Notice that the quantity controlling the truncation of the expansion is a kind of signal-to-noise ratio and therefore we have here an extension of the usual method of numerical filtering [5].

The result mentioned above has also an interesting interpretation from the point of view of information theory [7, 8] since it provides a generalisation of the concept of the ‘information content’ of a band-limited signal as introduced by Shannon. The number of singular values satisfying condition (6.5) is a generalisation of the Shannon number and, using a term introduced in optics, can be called the *number of degrees of freedom* (NDF) of the problem (corresponding to the given signal-to-noise ratio). The NDF is a measure of the number of independent pieces of information which can be extracted from the data and is a useful measure of information in many inversion problems, including image processing.

Let us mention that there are still many other methods for choosing the regularisation parameter, which have been proposed in the literature. Reviewing them all and comparing their performances in practical inversions is beyond the scope of this paper.

7. Regularisation of C -generalised solutions

In this section we investigate the regularisation of C -generalised solutions, i.e. of solutions which minimise a seminorm instead of a norm. Since this type of solution was only shortly discussed in I, we first reconsider the problem in more detail. We analyse then the generalisation of the Tikhonov regulariser to the case of C -generalised solutions and finally we discuss a few examples.

Assume again, for simplicity, that the functions φ_n are linearly independent so that the range of the operator L , defined by equation (1.3), is the whole space Y . Moreover, let C be an operator from the Hilbert space X into the Hilbert space Z (the constraint space) satisfying the following conditions, already stated in I:

(i) $C: X \rightarrow Z$ is a linear closed operator, with domain $\mathcal{D}(C)$ dense in X and $\mathcal{R}(C)$ closed in Z ;

(ii) the null space $\mathcal{N}(C)$ of C is nontrivial but $\mathcal{N}(L) \cap \mathcal{N}(C) = \{0\}$, i.e. the unique solution of the problem

$$Lf = 0 \quad Cf = 0 \tag{7.1}$$

is $f = 0$.

Then the C -generalised solution, f_C^\dagger , is the solution of equation (1.4) which minimises the functional (4.3). Notice that, since $\mathcal{N}(C)$ is nontrivial, the functional $p(f) = \|Cf\|_Z$ is not a norm but a seminorm.

As stated in I, under the previous conditions on the constraint operator C , there exists a unique C -generalised solution for any $g \in Y$. By applying to our special case a method developed in a more general framework [31] it is possible to characterise f_C^\dagger as the generalised solution of a modified problem.

The product operator $\{L, C\}: X \rightarrow Y \oplus Z$, with domain $\mathcal{D}(C)$, is linear, closed and invertible. Moreover its range is $Y \oplus \mathcal{R}(C)$ and hence it is closed in $Y \oplus Z$. It follows that the inverse operator is continuous and this implies that there exists a constant $\beta > 0$ such that

$$\|Lf\|_Y^2 + \|Cf\|_Z^2 \geq \beta \|f\|_X^2 \tag{7.2}$$

for any $f \in \mathcal{D}(C)$. By means of this inequality it is easy to show that $\mathcal{D}(C)$, endowed with the scalar product

$$(f, h)_C = (Lf, Lh)_Y + (Cf, Ch)_Z \tag{7.3}$$

is a Hilbert space, which we denote by X_C . We also denote by L_C the restriction of L to X_C . It is now obvious that the problem

$$Lf = g \quad \|Cf\|_Z = \min \tag{7.4}$$

is equivalent to the following one

$$L_C f = g \quad \|f\|_C = \min \tag{7.5}$$

which always has a solution since the range of L_C is Y . We have

$$f_C^+ = L_C^+ g \tag{7.6}$$

where L_C^+ is the generalised inverse of L_C . Since L_C is a finite rank operator, L_C^+ is continuous. However it can be ill-conditioned and in such a case regularisation methods are required. The analysis runs parallel to that given in §2.

We say that a one-parameter family of linear operators $\{R_{C,\mu}\}_{\mu>0}$ is a *regularisation algorithm* for the approximate computation of the C -generalised inverse L_C^+ if the following conditions are satisfied:

(i') for any $\mu > 0$, the range of $R_{C,\mu}$ is contained in the orthogonal complement of the null space of L_C or equivalently in $\mathcal{R}(L_C^+)$;

(ii') for any $\mu > 0$, the norm of $R_{C,\mu}$ is smaller than the norm of L_C^-

$$\|R_{C,\mu}\| \leq \|L_C^-\| = 1/\alpha_{C,N} \tag{7.7}$$

where $\alpha_{C,N}$ is the smallest singular value of L_C ;

(iii') the following property holds true

$$\lim_{\mu \rightarrow 0} R_{C,\mu} = L_C^+ \tag{7.8}$$

A regularisation algorithm analogous to the Tikhonov regulariser is obtained by considering the minimiser of the following functional

$$\Phi_\mu[f] = \|Lf - g\|_Y^2 + \mu \|Cf\|_X^2 \tag{7.9}$$

which is given by

$$f_\mu = R_{C,\mu} g = (L^*L + \mu C^*C)^{-1} L^* g \tag{7.10}$$

If we notice now that the adjoint of L_C has the following expression

$$L_C^* = (L^*L + C^*C)^{-1} L^* \tag{7.11}$$

we obtain from equation (7.10) that

$$R_{C,\mu} = (1 - \mu)^{-1} (L_C^* L_C + \mu' I)^{-1} L_C^* \tag{7.12}$$

where $\mu' = \mu/(1 - \mu)$. In such a way we have reduced $R_{C,\mu}$ to a standard Tikhonov regulariser and the results given at the beginning of §2 can be used in order to show that $R_{C,\mu}$ is a regularisation algorithm for L_C^+ . Notice that condition (ii') is certainly satisfied for any μ since all the singular values of L_C are smaller than 1. Equation (7.12) implies indeed that $R_{C,\mu}$, written in terms of the singular values $\alpha_{C,k}$ of the operator L_C , has the form (3.1) where now

$$W_{\mu,k} = \alpha_{C,k}^2 [\alpha_{C,k}^2 + \mu(1 - \alpha_{C,k}^2)]^{-1} \tag{7.13}$$

We conclude this section by giving a few examples to which the previous scheme can be applied. The first is the *interpolation* problem when the interpolating function must minimise a functional $\|Cf\|_Z$ which is the L^2 norm of its derivative of order k . The

corresponding C -generalised solution is a natural spline of degree $m = 2k - 1$. Then the regularised solution provided by equation (7.10) is just the solution of the *smoothing* problem investigated by Reinsch [32] and is again a spline function. The choice of the regularisation parameter is based on the method that we have called the discrepancy principle.

The second example is the method of Phillips [12] for the solution of first-kind Fredholm integral equations and in such a case the functional $\|Cf\|_Z$ is the L^2 norm of the second derivative of the unknown function. In that paper only regularised solutions are considered and again the discrepancy principle is used for the choice of the regularisation parameter.

The scheme presented in this section also applies to a recent investigation of the moment problem [33] where the constraint was the L^2 norm of the first derivative of the solution. A stability theorem was also proved in this paper and the dependence of the solution on the number of moments was quantified for a given noise level.

8. The Backus–Gilbert method

As shown in § 2, in the case where X is a space of square integrable functions, any regularised solution can be written, in the absence of noise, as an average of the true solution f

$$f_{\text{reg}}(x) = \int A(x, x')f(x') \, dx' \tag{8.1}$$

the averaging kernel $A(x, x')$ being given, in general, by equation (2.12). The same remark applies as well to the generalised solution—see I.

The relation (8.1) is the starting point of the Backus–Gilbert method [34, 35] which applies to the case where X is a space of square integrable functions. As in equation (2.12), it is assumed that the averaging kernel $A(x, x')$ is a linear combination of the functions $\varphi_n^*(x')$

$$A(x, x') = \sum_{n=1}^N a_n(x)\varphi_n^*(x') \tag{8.2}$$

but the coefficients $a_n(x)$ are not assumed to be linear combinations of the functions $\varphi_m(x)$. They are determined by introducing a variational principle for the averaging kernel, namely by requiring that they give the sharpest averaging kernel in a sense to be specified. For this purpose Backus and Gilbert introduce a function $J(x, x')$ which vanishes when $x = x'$ and increases monotonically as x goes away from x' . An example of such a function is

$$J(x, x') = (x - x')^2. \tag{8.3}$$

Then the averaging kernel is determined by solving, for any x in the domain of the unknown function $f(x)$, the following constrained minimisation problem

$$\delta^2(x) = \int J(x, x')|A(x, x')|^2 \, dx' = \min \tag{8.4}$$

$$\int A(x, x') \, dx' = 1. \tag{8.5}$$

If we introduce the quantities

$$b_n = \int \varphi_n(x) dx \quad S_{nm}(x) = \int J(x, x') \varphi_n(x') \varphi_m^*(x') dx' \quad (8.6)$$

then we find that $\delta^2(x)$ is a quadratic functional of the functions $a_n(x)$ and that the problem (8.4)–(8.5) is a standard minimisation problem

$$\delta^2(x) = \sum_{n, m=1}^N S_{nm}(x) a_n(x) a_m^*(x) = \min \quad (8.7)$$

$$\sum_{n=1}^N b_n^* a_n(x) = 1 \quad (8.8)$$

which can be solved by means of the method of Lagrange multipliers. If for any x the matrix $[S(x)]$ is nonsingular and if we denote by $S^{nm}(x)$ the elements of the inverse matrix, then the solution of the previous problem is

$$a_n(x) = \lambda(x) \sum_{m=1}^N S^{nm}(x) b_m \quad (8.9)$$

the Lagrange multiplier $\lambda(x)$ being given by

$$\lambda(x) = \left(\sum_{n, m=1}^N S^{nm}(x) b_m b_n^* \right)^{-1}. \quad (8.10)$$

Using equation (1.1) we find that the solution provided by the Backus–Gilbert method is

$$f_{BG}(x) = \int A(x, x') f(x') dx' = \sum_{n=1}^N a_n(x) \int f(x') \varphi_n^*(x') dx' = \sum_{n=1}^N g_n a_n(x) \quad (8.11)$$

and therefore it belongs to the subspace spanned by the functions $a_n(x)$. This solution is a type of generalised solution even if it is different from $f^+(x)$ since it belongs to the subspace spanned by the functions $a_n(x)$ while f^+ belongs to the subspace spanned by the functions $\varphi_n(x)$. It is also obvious that the functions $a_n(x)$ depend on the function $J(x, x')$ which has been chosen for characterising the sharpness of the averaging kernel. As already remarked by Backus and Gilbert, their solution can also be affected by numerical instability. In order to overcome this difficulty they proceed as follows [35].

If the covariance matrix $[C]$ of the noise is known, then from equation (8.11) one easily derives that, at any given point x , the variance of the error induced by the noise on $f_{BG}(x)$ is

$$\sigma^2(x) = \sum_{n, m=1}^N C_{nm} a_m(x) a_n^*(x). \quad (8.12)$$

Then Backus and Gilbert consider two methods:

(1) minimise the functional $\sigma^2(x)$ with the constraint (8.8) and also the constraint $\delta^2(x) \leq E^2$, which prescribes an upper limit on the desired resolution distance;

(2) minimise the functional $\delta^2(x)$ with the constraint (8.8) and also the constraint $\sigma^2(x) \leq \varepsilon^2$, which prescribes an upper limit on the error affecting the reconstructed solution.

Both problems can be solved again by means of the method of Lagrange multipliers and we do not give the details here. Let us just remark that these two methods are similar to the methods described respectively in §§ 4.1 and 4.2 for the case of the Tikhonov regulariser. They amount in fact to a form of regularisation of the Backus–Gilbert solution. Applications of these methods and numerical algorithms for implementing them have been considered by several authors [36–38]. In the case where the functionals are the Fourier coefficients of $f(x)$ as given by equation (3.5), a similarity between the Backus–Gilbert method and the Fejèr method has also been noticed [39].

Acknowledgments

This research has been partly supported by NATO Grant No 463/84 and EEC Contract No ST29-0089-3. C De Mol is ‘Chercheur qualifié’ with the Belgian National Fund for Scientific Research (FNRS).

References

- [1] Bertero M, De Mol C and Pike E R 1985 Linear inverse problems with discrete data. I: General formulation and singular system analysis *Inverse Problems* **1** 301–30
- [2] Lavrentiev M M 1967 *Some Improperly Posed Problems of Mathematical Physics* (Berlin: Springer)
- [3] Tikhonov A N and Arsenjine V Y 1977 *Solutions of Ill-Posed Problems* (Washington, DC: Winston/Wiley)
- [4] Groetsch C W 1984 *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind* (Research Notes in Mathematics vol. 105) (Boston: Pitman)
- [5] Twomey S 1985 The application of numerical filtering to the solution of integral equations encountered in indirect sensing measurements *J. Franklin Inst.* **279** 95–109
- [6] Twomey S 1974 Information content in remote sensing *Appl. Opt.* **13** 942–5
- [7] Bertero M and De Mol C 1981 Ill-posedness, regularization and number of degrees of freedom *Atti Fond. G. Ronchi* **36** 619–32
- [8] Pike E R, McWhirter J G, Bertero M and De Mol C 1984 Generalised information theory for inverse problems in signal processing *IEEE Proc.* **131** 660–7
- [9] Titchmarsh E C 1948 *Introduction to the Theory of Fourier Integrals* (London: Oxford University Press)
- [10] Bertero M, Brianzi P, De Mol C and Pike E R 1986 Positive regularized solutions in electromagnetic inverse scattering *Proc. URSI Int. Symp. on Electromagnetic Theory, Budapest Part A* (Budapest: Akademiai Kiado) pp 150–2
- [11] Tikhonov A N 1963 Solution of incorrectly formulated problems and the regularization method *Sov. Phys.–Math. Dokl.* **4** 1035
- [12] Phillips D L 1962 A technique for the numerical solution of certain integral equations of the first kind *J. Assn Comput. Mach.* **9** 84
- [13] Ivanov V K 1962 On linear problems which are not well-posed *Sov. Phys.–Math. Dokl.* **3** 981
- [14] Ivanov V K 1966 The approximate solution of operator equations of the first kind *USSR Comp. Math. Math. Phys.* **6** 6 197
- [15] Balakrishnan A V 1976 *Applied Functional Analysis* (Berlin: Springer)
- [16] Morozov V A 1968 The error principle in the solution of operational equations by the regularization method *USSR Comp. Math. Math. Phys.* **8** 2 63
- [17] Miller K 1970 Least squares methods for ill-posed problems with a prescribed bound *SIAM J. Math. Anal.* **1** 52
- [18] Golub G H, Heath M and Wahba G 1979 Generalized cross-validation as a method for choosing a good ridge parameter *Technometrics* **21** 215

- [19] Craven P and Wahba G 1979 Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation *Numer. Math.* **31** 377
- [20] Landweber L 1951 An iteration formula for Fredholm integral equations of the first kind *Am. J. Math.* **73** 615
- [21] Bialy H 1959 Iterative behandlung linearer funktionalgleichungen *Arch. Rat. Mech. Anal* **4** 166
- [22] Kammerer W J and Nashed M Z 1971 Steepest descent for singular linear operators with nonclosed range *Applicable Anal.* **1** 143
- [23] Kammerer W J and Nashed M Z 1972 On the convergence of the conjugate gradient method for singular linear operator equations *SIAM J. Numer. Anal.* **9** 165
- [24] Gerchberg R W 1974 Super-resolution through error energy reduction *Opt. Acta* **21** 709
- [25] Papoulis A 1975 A new algorithm in spectral analysis and band-limited extrapolation *IEEE Trans. Circuits Syst.* **CAS-22** 735
- [26] De Santis P and Gori F 1975 On an iterative method for super-resolution *Opt. Acta* **22** 691
- [27] Maitre H 1981 Iterative superresolution. Some new fast methods *Opt. Acta* **28** 973
- [28] Bertero M, Brianzi P, Defrise M and De Mol C 1986 Iterative inversion of experimental data in weighted spaces *Proc. URSI Int. Symp. on Electromagnetic Theory, Budapest Part A* p 315
- [29] Natterer F 1986 Numerical treatment of ill-posed problems *Inverse Problems* (ed. G Talenti) (Lect. Notes in Math., vol 1225) (Berlin: Springer)
- [30] Vainikko G M 1982 The discrepancy principle for a class of regularization methods *USSR Comp. Math. Math. Phys.* **22** 3 1
- [31] Groetsch C W 1986 Regularization with linear equality constraints *Inverse Problems* (ed. G Talenti) (Lect. Notes in Math., vol 1225) (Berlin: Springer)
- [32] Reinsch C H 1967 Smoothing by spline functions *Numer. Math.* **10** 177
- [33] Talenti G 1987 Recovering a function from a finite number of moments *Inverse Problems* **3** 501
- [34] Backus G and Gilbert F 1968 The resolving power of gross Earth data *Geophys. J. R. Astron. Soc.* **16** 169
- [35] Backus G and Gilbert F 1970 Uniqueness in the inversion of inaccurate gross Earth data *Phil. Trans. R. Soc.* **266** 123
- [36] Oldenburg D W 1976 Calculation of Fourier transforms by the Backus–Gilbert method *Geophys. J. R. Astron. Soc.* **44** 413
- [37] Colton D 1984 The inverse scattering problem for time-harmonic acoustic waves *SIAM Rev.* **26** 323
- [38] Haario H and Somersalo E 1985 On the numerical implementation of the Backus–Gilbert method *Cahiers Math. de Montpellier* **32** 107
- [39] Bertero M, Brianzi P, Pike E R and Rebolia L 1988 Linear regularizing algorithms for positive solutions of linear inverse problems *Proc. R. Soc. A* **415** 257–75