

## Linear inverse problems with discrete data. I: General formulation and singular system analysis

M Bertero†, C De Mol‡ and E R Pike§

† Dipartimento di Fisica dell'Università and Istituto Nazionale di Fisica Nucleare, Genova, Italy

‡ Département de Mathématique, Université Libre de Bruxelles, Bruxelles, Belgium

§ Royal Signals and Radar Establishment, St Andrews Road, Great Malvern, Worcestershire WR14 3PS, UK

Received 26 July 1985

**Abstract.** This paper is the first part of a work which is concerned with linear methods for the solution of linear inverse problems with discrete data. Such problems occur frequently in instrumental science, for example tomography, radar, sonar, optical imaging, particle sizing and so on. We give a general formulation of the problem by extending the approach of Backus and Gilbert and by defining a mapping from an infinite-dimensional function space into a finite-dimensional vector space. The singular system of this mapping is introduced and used to define natural bases both in the solution and in the data space. We analyse in this context normal solutions, least-squares solutions and generalised inverses. We illustrate the wide applicability of the singular system technique by discussing several examples in detail. Particular attention is devoted to showing the many connections between this method and techniques developed in other topics like the extrapolation of band-limited signals and the interpolation of functions specified on a finite set of points. For example, orthogonal polynomials for least-squares approximation, spline functions and discrete prolate spheroidal functions are particular cases of the singular functions introduced in this paper.

The problem of numerical stability is briefly discussed but the investigation of the methods developed for overcoming this difficulty, like truncated expansions in the singular bases, regularised solutions, iterative methods and so on, is deferred to a second part of this work.

### 1. Introduction

In the last two decades, the theory and practice of inverse problems have been developed in many scientific domains. The fundamental reason is that, in much of experimental science, a common method for determining certain characteristics of a physical sample is to observe the interaction between the sample and the radiation emitted by a known source. One or several detectors measure variations in space and/or time of the scattered radiation, as scattering amplitudes, field fluctuations, absorption coefficients and so on. The output of the detectors can be digitised and the final result of the experiment is a set of (real or complex) numbers  $\{g_n\}_{n=1}^N$  stored in a computer. The interpretation of the data requires the solution of a mathematical problem, i.e. the restoration of the unknown characteristic of the sample from the set of numbers  $g_n$ .

The experimental data  $g_n$  can be viewed as the components of a vector which will be called the *data vector* and denoted by  $\mathbf{g}$ . The unknown characteristic of the sample,

denoted by  $f$ , will be called the *object*, irrespective of its physical nature. The object is supposed to be a function (or a set of functions) of a variable  $x$ , whose physical interpretation is also left unspecified and which can also be multi-dimensional.

We restrict our investigation to those cases where, to a good approximation, the relation between the object and the data vector is linear. This property holds true when the following conditions are satisfied. Firstly the field quantity  $g$  measured by the detectors is linearly related to the object  $f$ :

$$g(y) = \int K(y, x)f(x) dx. \quad (1.1)$$

Here  $y$  denotes a space and/or time variable and the kernel  $K(y, x)$  describes the interaction between the incident radiation and the sample. In some cases, like x-ray tomography,  $K(y, x)$  is not a function but a distribution. We assume that this kernel is known or, in other words, we assume that we can solve the direct problem, which consists of computing the scattered radiation  $g(y)$  on the basis of a known sample,  $f(x)$ .

Secondly a linear response of the detectors is also required, namely any  $g_n$  must be proportional to the value of  $g$  at some point  $y_n$ :

$$g_n = \int K(y_n, x)f(x) dx \quad n = 1, \dots, N. \quad (1.2)$$

More generally, since any detector integrates over some space-time region, equation (1.2) must be replaced by

$$g_n = \int P(y_n - y) \left( \int K(y, x)f(x) dx \right) dy \quad n = 1, \dots, N \quad (1.3)$$

where the averaging function  $P$ , which typically has a peak centred near  $y = y_n$ , describes the integration effect of the detector (in the case of equation (1.2),  $P(y_n - y) = \delta(y - y_n)$ ). Here we have assumed, for simplicity, that all the detectors are identical and time invariant. The generalisation to the case where these conditions are not fulfilled is straightforward.

Equation (1.3) is still an idealisation since it does not take into account the effect of noise. Let us forget for a moment this very important point which will be discussed in the following. Then equation (1.2) or (1.3) shows that the output of one detector is the value of a *linear functional* of the object  $f$ . As usual, we call a functional any rule for assigning a real (or complex) number to each member of a certain class of functions.

We can summarise the previous analysis as follows. Any given experiment for the determination of the object  $f$  specifies a set of functionals which we denote by  $\{F_n\}_{n=1}^N$ . The outcome of the experiment is a set of measured values of these functionals. Then, ignoring the errors on the data, a preliminary formulation of a *linear inverse problem with discrete data* is the following: *given a class  $X$  of functions, given a set  $\{F_n\}_{n=1}^N$  of linear functionals defined on  $X$  and given a set  $\{g_n\}_{n=1}^N$  of values of such functionals, find in  $X$  an object  $f$  such that*

$$F_n(f) = g_n \quad n = 1, \dots, N. \quad (1.4)$$

A similar formulation of an inverse problem was first given by Backus and Gilbert [1, 2] in the case of the inversion of gross Earth data. In their case the object  $f$  is the Earth model and examples of functionals are provided by the total mass, the total moment of

inertia, the frequency of oscillation of a particular normal mode, etc. Since most of these functionals are not linear they are usually linearised in a neighbourhood of a reasonable Earth model. Other important examples of problems included in the previous general formulation are provided by the moment problem, the numerical interpolation or the numerical differentiation of a function given on a discrete set of points, the reconstruction of a function from the knowledge of a finite set of values of its Fourier or Laplace transform and so on. Some of these problems will be discussed in detail in the following.

The previous formulation needs some further specification. First of all we require the class  $X$  of the objects to be a *Hilbert space* which will be called the *object space*. This assumption seems rather restrictive but it is quite satisfactory in most practical cases and obviously it makes the mathematics much simpler.

A second important assumption concerns the set of linear functionals  $\{F_n\}_{n=1}^N$ . We will require that they are continuous so that from the Riesz representation theorem [3] it follows that there exists in  $X$  a set of functions  $\{\phi_n\}_{n=1}^N$  such that

$$F_n(f) = (f, \phi_n)_X \quad n = 1, \dots, N \tag{1.5}$$

where  $(f, \phi)_X$  denotes the scalar (inner) product of two functions  $f, \phi \in X$ . While the functionals  $F_n$  are uniquely specified by the experiment, the functions  $\phi_n$  are specified both by the experiment and by the choice of the Hilbert space  $X$ , which is largely at our disposal. This problem will be discussed in § 3.

A more precise formulation of the inverse problem can now be stated as follows: *given the Hilbert space  $X$ , given the set of functions  $\{\phi_n\}_{n=1}^N \subset X$  and given the set of numbers  $\{g_n\}_{n=1}^N$ , find a function  $f \in X$  such that*

$$(f, \phi_n)_X = g_n \quad n = 1, \dots, N. \tag{1.6}$$

Clearly the information about  $f$  contained in equation (1.6) is incomplete since from (1.6) we can derive only a finite number of components of  $f$ . However, the lack of uniqueness of the solution is not the main problem. When the number of data is large a lack of numerical stability is also possible. In such a case it may be necessary to reformulate the problem and to renounce the requirement of strict equality in (1.6).

The lack of numerical stability derives from the fact that in most cases the problem (1.6) is the projection on a finite-dimensional space of an *ill-posed problem* (IPP): for example, a Fredholm integral equation of the first kind when the functionals  $F_n$  are given by equation (1.2). Now the numerical analysis of IPP is usually developed along two routes [4]: the first is to apply to the IPP, formulated in an infinite-dimensional space, some methods of solution, such as regularisation and only at the last stage to use numerical methods to approximate the solution; the second route is first to project the IPP on a finite-dimensional space and then treat the corresponding ill-conditioned linear system by means of suitable numerical methods [5].

Here we follow an alternative approach which is more powerful in practice; we consider discrete data but we maintain reconstructed solutions in infinite-dimensional spaces and only at the last stage perform a fine discretisation of the solution in order to produce a numerical or graphical result. In practical inverse problems, there are indeed found two different types of discretisation. The first is discretisation of the data, which is related to the design of the experiment and therefore can be submitted to rather strong restrictions (for instance, only a small number of detectors can be used). The second is discretisation of the solution, which can be very accurate and is essentially related to the power of the computer used for the numerical analysis.

**2. The normal solution**

We assume in this section that the functions  $\phi_n$  are linearly independent (but not necessarily orthogonal) and we denote by  $X_N$  the subspace of  $X$ , with dimension  $N$ , spanned by the  $\phi_n$ . In this case the set of functionals (1.5) defines a mapping of  $X$  onto  $\mathbb{R}^N$  (or  $\mathbb{C}^N$ ). The proof is very simple. If the mapping is not onto, i.e. its range is a subspace of  $\mathbb{R}^N$  (or  $\mathbb{C}^N$ ), there exists a vector  $c$  such that  $c_1(f, \phi_1)_X + c_2(f, \phi_2)_X + \dots + c_N(f, \phi_N)_X = 0$ , for any  $f \in X$ , contradicting the hypothesis that the  $\phi_n$  are linearly independent.

The previous remark implies that a solution of the inverse problem (1.6) always exists. However, it is not unique. Any function  $f_1$  orthogonal to  $X_N$  is such that  $(f_1, \phi_n)_X = 0$  for  $n = 1, \dots, N$ . Therefore the set  $S$  of the solutions is the infinite-dimensional affine subspace of all functions  $f = f_0 + f_1$ , where  $f_0$  is a given solution and  $f_1$  is any function orthogonal to  $X_N$ , i.e.  $S$  is a translation of  $X_N^\perp$ , the orthogonal complement of  $X_N$ .

Clearly  $S$  is a closed and convex set of  $X$  and therefore there exists a unique element of  $S$  which has minimal distance from the null element of  $X$  ([3], p 9). This solution of minimal norm is called the *normal solution* of problem (1.6) and it will be denoted by  $f^\dagger$  [6].

The previous results can be summarised as follows.

*Theorem 2.1.* If the functionals  $F_n$  are linearly independent, then there always exists a unique solution of minimal norm  $f^\dagger$  of problem (1.6);  $f^\dagger$  is the unique solution orthogonal to the subspace  $X_N$  of the functions which annihilate all the functionals  $F_n$ .

The interpretation of the previous proposition is obvious: the finite set of data allows only the determination of the projection of the unknown object on the subspace  $X_N$ . The component of  $f$  orthogonal to  $X_N$  is ‘invisible’. The prescription of taking the normal solution is equivalent to setting this ‘invisible’ component equal to zero. Notice however that the complete characterisation of the set of ‘invisible’ components depends not only on the functionals  $F_n$  but also on the choice of the Hilbert space  $X$ : only when  $X$  has been specified does there exist a unique decomposition of an arbitrary object into a ‘visible’ and an ‘invisible’ component.

An explicit expression for  $f^\dagger$  can be obtained by means of standard computations. Since  $f^\dagger \in X_N$ , then

$$f^\dagger = \sum_{n=1}^N a_n \phi_n \tag{2.1}$$

and the coefficients  $a_n$  solve the linear system

$$\sum_{m=1}^N G_{mn} a_m = g_n \quad n = 1, \dots, N \tag{2.2}$$

where the  $G_{mn} = (\phi_m, \phi_n)_X = G_{nm}^*$  (the asterisk denotes complex conjugation) are matrix elements of the *Gram matrix*. If we denote by  $G^{nm}$  the matrix elements of the inverse of the Gram matrix and we also introduce in  $X_N$  the dual basis

$$\phi^n = \sum_{m=1}^N G^{nm} \phi_m \quad n = 1, \dots, N, \tag{2.3}$$

namely the set of functions satisfying the conditions  $(\phi^n, \phi_m)_X = \delta_{nm}$ , it follows that

$$f^\dagger = \sum_{n=1}^N g_n \phi^n. \tag{2.4}$$

This representation clearly shows that  $f^\dagger$  depends continuously on the data: if  $\delta g$  denotes a variation of  $g$  and  $\delta f^\dagger$  is the corresponding variation of  $f^\dagger$ , then  $\|\delta f^\dagger\|_X$  tends to zero when  $\delta g$  tends to zero. It follows that the problem of the computation of  $f^\dagger$  is always *well posed* in the strict mathematical sense. However, it can be strongly ill-conditioned and therefore exhibit numerical instability. As is well known ([7], p 35), the determinant of the Gram matrix is equal to the square of the volume of the  $N$ -dimensional polyhedron defined by the functions  $\phi_n$  and represents a measure of the ‘effective’ linear independence of these functions: the ‘flatter’ the polyhedron, the smaller is the Gram determinant. In such an ill-conditioned situation, as a consequence of the errors on the data (noise), not all the functionals can be considered as effectively linearly independent and, for a given noise level, we have the problem of estimating how many effectively linearly independent functionals we have. This number can be much smaller than the number  $N$  of data. This point will be discussed in the next section.

As a concluding remark we point out that, in the absence of errors on the data, the normal solution defines a mapping  $A$  of  $X$  into itself,  $f^\dagger = Af$ , as one can see by combining equations (2.4) and (1.6). It has been remarked [8] that this mapping is the only one satisfying the conditions of linearity, continuity, computability, consistency and optimality. It is easy to recognise that  $Af$  is just the orthogonal projection of  $f$  onto  $X_N$ .

In the special case where  $X$  is a (weighted)  $L^2$  space:

$$(f, h)_X = \int w(x)f(x)h^*(x) dx \tag{2.5}$$

( $w(x) > 0$  on the integration domain), the mapping  $A$  is an integral operator whose kernel is given by

$$\begin{aligned} A(x, x') &= \sum_{n=1}^N \phi^n(x)w(x')\phi_n^*(x') \\ &= \sum_{n,m=1}^N G^{nm} \phi_m(x)w(x')\phi_n^*(x'). \end{aligned} \tag{2.6}$$

In the case  $w(x) = 1$ , this kernel is called the Dirichlet kernel by Backus and Gilbert [1] since in the case where the functionals  $F_n$  are the Fourier coefficients of  $f$ , it coincides with the well known Dirichlet kernel of the Fourier series theory. Indeed, if

$$F_n(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-inx}f(x) dx \quad n = 0, \pm 1, \dots, \pm N \tag{2.7}$$

and  $X = L^2(-\pi, \pi)$ , then

$$A(x, x') = \frac{1}{2\pi} \frac{\sin(N + \frac{1}{2})(x - x')}{\sin \frac{1}{2}(x - x')}. \tag{2.8}$$

For fixed  $x$ ,  $A(x, x')$  is an ‘approximation’ of the Dirac delta distribution: the integral of  $A(x, x')$  is one and  $A(x, x')$  has a peak at  $x' = x$ , with maximum value equal to  $(2N + 1)/2\pi$  and half width to its first zero equal to  $2\pi/(2N + 1)$ . Therefore, in the absence of noise the

normal solution is approximately an average of the 'true' solution  $f$  over this distance, which can be taken as a measure of the 'resolution limit' due to the incomplete knowledge of  $f$  provided by the functionals (2.7).

### 3. Numerical stability

In the introduction we have already pointed out that in most relevant cases the problem (1.6) is the projection onto a finite-dimensional space of an IPP. Now, the main feature of IPP is the lack of continuous dependence of the solution on the data [9, 10]: even if the error on the data tends to zero, the induced error in the solution may tend to infinity. Clearly this concept applies only to problems formulated in infinite-dimensional spaces: as shown in § 2, for an inverse problem with discrete data the computation of the normal solution is always a well posed problem in the strict mathematical sense. However, as an effect of the ill-posedness of the related infinite-dimensional problem, the discrete problem can be extremely ill-conditioned, especially when the number of data is large, since in that case the finite-dimensional problem is 'closer' to the infinite-dimensional one. As a consequence the Gram matrix has many small eigenvalues and therefore a very small determinant. By adding more and more experimental points we add smaller and smaller eigenvalues and therefore we get Gram matrices with smaller and smaller determinants. We do not, therefore, add effectively independent information about the object, at least for a given level of noise. In other words, even in the ideal case of an infinite-dimensional problem, it is only possible to derive from the data a finite 'amount' of information on the object. In optics and communication theory, this is usually quantified in terms of a 'resolution limit' or of a 'number of degrees of freedom' [11]. Such a finite amount of information has to be obtained with a suitable finite number of data values. This 'optimum' number is the number of effectively linearly independent functionals and therefore it is expected to be approximately equal to the 'number of degrees of freedom'. The 'resolution limit' and the optimum number of data points will depend both on the noise and on the structure of the Hilbert space  $X$ .

For all these reasons the effect of the noise, which was ignored in the previous sections, is in fact a crucial one. We will now discuss a number of topics which are relevant from this point of view.

#### 3.1. The object space $X$

In many cases the most obvious choice is to take a space  $X$  of square-integrable functions [1, 2]. However, it is known from the theory of IPP that an important tool for restoring the solution is the use of *a priori* knowledge of properties of the object  $f$  [12, 13]. For instance, regularity properties of  $f$  (existence of derivatives up to a certain order) can be used in order to produce stable, approximate solutions (regularised solutions) for large classes of IPP [14]. Furthermore it has been shown in some relevant problems [15, 16] that *a priori* knowledge about the 'localisation' of the object, namely the knowledge that the object is significantly different from zero only inside some 'small' region, can be used in order to improve resolution limits. Such kinds of information in some cases can be derived from the experimental data and in others can be obtained from independent experiments.

A way to use *a priori* knowledge is to incorporate it in the structure of the object space  $X$ . In the case of regularity properties, Sobolev spaces are appropriate, while in the case of

localisation properties, weighted  $L^2$  spaces can be used. Weighted Sobolev spaces are convenient if one wishes to take into account both regularity and localisation.

There are other reasons for considering a space  $X$  with some particular structure. For instance in the case of x-ray tomography (see § 9), the use of Sobolev spaces is appropriate [17] if we require that the functionals  $F_n$  are continuous. For the same reasons, in the problem of numerical interpolation and derivation it may be convenient to use reproducing kernel Hilbert spaces (RKHS), namely Hilbert spaces of continuous functions where the evaluation functionals are continuous (see § 6).

All the previous remarks indicate that an accurate choice of the object space  $X$  is an important point in the definition of the problem.

### 3.2. The data space $Y$

The definition of the normal solution (§ 2) only requires that the set of all the possible data vectors  $\mathbf{g}$  has the structure of a linear space. This minimal assumption is not sufficient when we wish to quantify the ill-conditioning of the problem. In that case we need a measure of the errors or, in other words, we must introduce a metric in the data space. Such a structure is also necessary for considering least-squares problems.

We will assume that the *data space*, denoted by  $Y$ , has a euclidean structure, with the following scalar product:

$$(\mathbf{g}, \mathbf{h})_Y = \sum_{n,m=1}^N W_{nm} g_n h_m^* \tag{3.1}$$

the weights  $W_{nm}$  being the matrix elements of a given *positive definite* matrix  $\mathbf{W}$ .

The simplest choice of the weights is obviously  $W_{nm} = \delta_{nm}$ . However, other choices can be convenient. For instance, if we wish to study the discrete problem as a discrete projection of an ill-posed problem, then it may be convenient to choose the weights according to some quadrature formula in order to approximate by means of equation (3.1) the scalar product in the function space of the data. If the latter is an  $L^2$  space, then diagonal weights can be suitable:  $W_{nm} = w_n \delta_{nm}$ .

On the other hand, in the case of least-squares problems a natural choice is to relate the weight matrix  $\mathbf{W}$  to the covariance matrix  $\mathbf{C}$  of the errors  $\delta g_n$  on the components of the data vector  $\mathbf{g}$ . In linear regression theory the relation is

$$\mathbf{W} = \mathbf{C}^{-1}. \tag{3.2}$$

Then, as is well known, assuming that the average value of the random vector  $\delta \mathbf{g}$  is zero, it follows that the average value of  $\|\delta \mathbf{g}\|_Y = (\delta \mathbf{g}, \delta \mathbf{g})_Y$  is just  $N$ , the number of experimental data values. Furthermore, if we project  $\delta \mathbf{g}$  onto any basis which is orthonormal with respect to the scalar product (3.1), (3.2), then the components of  $\delta \mathbf{g}$  with respect to this basis are uncorrelated and their variance is one.

### 3.3. Optimum choice of the experiment

In some problems, like for example the one investigated by Backus and Gilbert [1, 2], the choice of the functionals is not free. However, in the type of experiments described in the introduction, the number and the positions of the detectors are at the disposal of the experimentalist. Then the problem arises of an 'optimum' (in some sense) choice of the experiment. The optimality criterion can also include some economical parameter (the cost

of the detectors, for instance). In any case the first important mathematical fact is that already discussed at the beginning of this section, namely, the existence of an 'optimum' number of data points. Here we want to elucidate this fact by discussing an elementary example along lines developed in [18]. A more general approach for a broader class of ill-posed problems is given in [19].

Consider the problem of computing the derivative  $f(x)$  of a function  $g(x)$ , which is defined on some interval  $[a, b]$  and whose values  $\{g_n\}_{n=1}^N$ , affected by errors, are given at  $N$  equidistant points:  $x_1 = a, x_2, \dots, x_N = b$ . The distance between adjacent points is  $h = (b - a)/(N - 1)$ . Using a linear interpolation, the estimate of the derivative in the interval  $(x_n, x_{n+1})$  is:  $\tilde{f}_n = (g_{n+1} - g_n)/h$  ( $n = 1, \dots, N - 1$ ). Assume that the unknown derivative  $f$  belongs to a class of functions satisfying the condition

$$|f''(x)| \leq E \quad \forall x \in (a, b) \quad (3.3)$$

and that the maximum error on the values of  $g(x)$  does not exceed  $\varepsilon$ , so that

$$|g(x_n) - g_n| \leq \varepsilon \quad n = 1, \dots, N. \quad (3.4)$$

Then, if we take  $\tilde{f}_n$  as an estimate of any  $f = g'$ , satisfying conditions (3.3) and (3.4), at the point  $\tilde{x}_n = \frac{1}{2}(x_n + x_{n+1})$  ( $n = 1, \dots, N - 1$ ), the error is given by

$$|\tilde{f}_n - f(\tilde{x}_n)| \leq |\tilde{f}_n - (g(x_{n+1}) - g(x_n))/h| + |(g(x_{n+1}) - g(x_n))/h - f(\tilde{x}_n)|. \quad (3.5)$$

From condition (3.4) it follows that the first term on the RHS is bounded by  $2\varepsilon/h$ , while, using a Taylor expansion for  $g(x)$  and condition (3.3), it follows that the second term is bounded by  $h^2 E/24$ , so that

$$|\tilde{f}_n - f(\tilde{x}_n)| \leq \frac{2\varepsilon}{h} + \frac{h^2}{24} E. \quad (3.6)$$

The first term represents the error on the restored solution due to the error on the data and it increases with the number of points as a consequence of the ill-posedness of the problem. The second term is the error due to the discretisation of the data and it tends to zero when the number of points tends to infinity. As a consequence the error has a minimum at  $h = (24\varepsilon/E)^{1/3}$  and the corresponding minimum error is bounded by  $1.04 E(\varepsilon/E)^{2/3}$ . Therefore  $N \simeq (b - a)(E/24\varepsilon)^{1/3}$  is the 'optimum' number of data points.

The previous example is elementary but contains the essential features of the problem. The optimum number of points is obtained by minimising some reconstruction error and depends both on the noise and on the class of the solutions through the 'signal-to-noise ratio'  $E/\varepsilon$ . Furthermore it is not difficult to show that, treating the problem in infinite-dimensional space by means of truncated singular function expansions [13], the number of terms in the truncated expansion is of the same order as the number of points estimated before, namely it grows like  $(E/\varepsilon)^{1/3}$  when  $E/\varepsilon \rightarrow \infty$ . Similar results are known in communication theory and are related to the so called *2wT* theorem [20].

A further refinement can be obtained when the 'optimum' number  $N$  of data points has been fixed. Then one can look for 'optimum' positions of such points (see, for example, results derived by Luttrell [21]).

### 3.4. Resolution limits

In optics it is well known that the number of degrees of freedom is related to a resolution limit [11]. Such a result has been generalised to other important problems [22]. Obviously,

as for the other parameters, the resolution limit depends both on the noise and on the class of objects.

The meaning of a resolution limit from the point of view of numerical stability is that, if we want to recover details of the object which are smaller than such a limit, then we will necessarily find instabilities.

A route for a rigorous determination of resolution limits can be that indicated at the end of § 2. When the number of data points is optimum, the corresponding normal solution is numerically stable. Then one can look at the corresponding ‘averaging’ kernel  $A(x, x')$ , equation (2.6), and take the width of the main peak at  $x' = x$  as a measure of the resolution limit in the neighbourhood of  $x$ . By analogy with the case of Fourier series, equation (2.8), one expects an improvement in resolution if one can increase the number of data points. But if we require optimality this is possible only by a reduction of the noise on the data.

It is important to notice that the previous definition of a resolution limit can be extended also to the case where the choice of the data points is not optimum. The second part of this work will be devoted, indeed, to the investigation of the methods, for example regularisation, which provide stable solutions also when optimality of the data points does not hold. Any such method of solution can be described in terms of a mapping, say  $R$ , from the data space  $Y$  into the object space  $X$ . If we consider the successive application first of the mapping  $L$  (see § 4), which transforms an object into a noise-free data vector, and then of the mapping  $R$ , we get a mapping  $A$  of  $X$  into itself. The product  $A = RL$  represents the ‘approximation’ in the restoration of an arbitrary object  $f$  using the algorithm  $R$  in the absence of noise. If  $A$  is an integral operator and if its kernel  $A(x, x')$  shows the features described at the end of § 2, then the width of the main peak can be taken again as a measure of the resolution achievable with the algorithm  $R$ .

#### 4. The generalised solution or normal pseudosolution

If the functions  $\phi_n$  are not linearly independent and the data  $g_n$  are affected by experimental errors, then, in general, the normal solution does not exist, because the data do not satisfy the same linear relationships as the  $\phi_n$ . Obviously, if one is able to show that some of the functions  $\phi_n$  are linear combinations of the others, these functions can be discarded and one may take the normal solution associated with the remaining linearly independent functionals. However, in most cases, the linear independence of the functions  $\phi_n$  cannot be checked in practice and therefore it is convenient to have a procedure which produces a ‘solution’ in any case. The most popular is the *least-squares method*. For this purpose we need a metric in the data space and we assume that this metric is generated by a scalar product as given in equation (3.1).

Now the mapping  $X \ni f \rightarrow g \in Y$ , defined by equations (1.4) and (1.5), can be described in terms of a linear operator  $L: X \rightarrow Y$ , which transforms a function in  $X$  into a vector of  $Y$  according to the rule

$$(Lf)_n = (f, \phi_n)_X \quad n = 1, \dots, N. \tag{4.1}$$

The mapping is onto when the  $\phi_n$  are linearly independent; otherwise the range of  $L$  is a subspace  $Y_{N'}$  of  $Y$ , with dimension  $N' < N$ . Notice that the subspace in  $X$  spanned by the  $\phi_n$  also has dimension  $N'$ ; we will denote this subspace by  $X_{N'}$ . The orthogonal complement  $X_{N'}^\perp$ , i.e. the subspace of the ‘invisible’ components, is just the null space, denoted by  $\mathcal{N}(L)$ , of the linear operator  $L$ .

The adjoint operator  $L^*$ , defined by

$$(Lf, \mathbf{g})_Y = (f, L^*\mathbf{g})_X \tag{4.2}$$

transforms a data vector  $\mathbf{g}$  into an element of  $X$ , more precisely of  $X_{N'}$ , since from (4.2) it follows that, if  $f \in X_{N'}^\perp$  (i.e.  $Lf=0$ ), then  $(f, L^*\mathbf{g})=0$ .

From equations (4.2), (3.1) and (4.1) it is easy to derive an explicit expression for  $L^*$ :

$$L^*\mathbf{g} = \sum_{n=1}^N \left( \sum_{m=1}^N W_{nm} g_m \right) \phi_n. \tag{4.3}$$

After these preliminary considerations, we notice that the problem (1.4), (1.5) can now be stated in the following concise form: *given  $\mathbf{g} \in Y$ , find an  $f \in X$  such that*

$$Lf = \mathbf{g}. \tag{4.4}$$

When the  $\phi_n$  are not linearly independent, this problem has a solution iff  $\mathbf{g} \in Y_{N'}$ . If  $\mathbf{g} \notin Y_{N'}$ , one can define a *least-squares solution* or *pseudosolution* of equation (4.4) as any  $\tilde{f} \in X$  such that

$$\|L\tilde{f} - \mathbf{g}\|_Y = \text{minimum}. \tag{4.5}$$

The transition from the solution to the pseudosolution implies a first change in the concept of ‘solution’ of the problem (4.4): we do not require strict equality as in equation (4.4) but we only require that the ‘distance’ between  $Lf$  and  $\mathbf{g}$  is of the order of the experimental errors. Indeed, as we will see in a moment (equation (4.7)), the minimum value of  $\|Lf - \mathbf{g}\|_Y$  is just the norm of the component of  $\mathbf{g}$  orthogonal to  $Y_{N'}$  and this component is purely an effect of the experimental errors.

The following propositions can be easily proved:

(i) if  $\mathbf{g} \in Y_{N'}$ , i.e. if the data satisfy the compatibility conditions required by the linear dependence of the  $\phi_n$ , then the set of the pseudosolutions coincides with the set of the solutions of equation (4.4);

(ii) the set of the pseudosolutions coincides with the set of the solutions of the Euler equation associated with the variational problem (4.5), i.e.

$$L^*L\tilde{f} = L^*\mathbf{g} \tag{4.6}$$

or also with the set of the solutions of the equation

$$L\tilde{f} = P\mathbf{g} \tag{4.7}$$

where  $P\mathbf{g}$  denotes the orthogonal projection of  $\mathbf{g}$  onto  $Y_{N'}$ .

The set  $\tilde{S}$  of the pseudosolutions is again an infinite-dimensional affine subspace, containing all the functions  $\tilde{f} = \tilde{f}_0 + \tilde{f}_1$ , where  $\tilde{f}_0$  is a given pseudosolution and  $\tilde{f}_1$  is any function orthogonal to  $X_{N'}$ . It follows that there exists a unique pseudosolution of minimal norm. It is called the normal pseudosolution or *generalised solution* and denoted by  $f^\dagger$ , as the normal solution, since it coincides with the latter when the functionals are linearly independent. These results can be collected in the following proposition.

*Theorem 4.1.* For any  $\mathbf{g} \in Y$  there exists a unique generalised solution  $f^\dagger$  of the problem; this solution depends continuously on the data in the sense that if  $\delta\mathbf{g}$  is the error in the data and  $\delta f^\dagger$  the induced error in  $f^\dagger$ , then  $\|\delta f^\dagger\|_X \rightarrow 0$  when  $\|\delta\mathbf{g}\|_Y \rightarrow 0$ .

As already remarked, when the  $\phi_n$  are linearly independent the generalised solution coincides with the normal solution. Indeed the operator  $L^*$  is invertible, i.e. the equation  $L^*g=0$  has only the trivial solution  $g=0$ , and therefore equation (4.6) is equivalent to equation (4.4). In this case the generalised solution  $f^\dagger$  is independent of the structure (scalar product) of the space  $Y$ . On the other hand, the choice of the scalar product of  $Y$  is essential when the  $\phi_n$  are not linearly independent, because the projection  $P$  on  $Y_{N'}$ , equation (4.6), depends on the structure of  $Y$ .

The structure of  $Y$  is also important in another fundamental problem, i.e. the measure of the numerical stability of the normal or the generalised solution.

To show this, let us remark that the mapping  $Y \ni g \rightarrow f^\dagger \in X$  defines an operator  $L^\dagger: Y \rightarrow X$ , the so-called *generalised inverse* (or Moore–Penrose inverse) [6] of  $L$ :

$$f^\dagger = L^\dagger g. \tag{4.8}$$

The operator  $L^\dagger$  is continuous, as stated in the previous proposition. Then, it is easy to derive that

$$\frac{\|\delta f^\dagger\|_X}{\|f^\dagger\|_X} \leq C(L) \frac{\|\delta g\|_Y}{\|g\|_Y} \tag{4.9}$$

where

$$C(L) = \|L^\dagger\| \|L\| \tag{4.10}$$

is the so called *condition number*. Clearly the value of  $C(L)$  is not independent of the choice of the scalar product of  $Y$ , since both the norm of  $L$  and the norm of  $L^\dagger$  depend on such a choice.

An extension of the generalised solution is sometimes introduced (see, for instance, § 6) by looking for a pseudosolution which minimises a suitable seminorm instead of the norm. We can formulate the method in the abstract as follows. Let  $C$  be an operator from the Hilbert space  $X$  into a new Hilbert space  $Z$  (which could be called the ‘constraint’ space) satisfying the following conditions:

- (i)  $C: X \rightarrow Z$  is a closed linear operator with domain  $\mathcal{D}(C)$  dense in  $X$ ;
- (ii) the range of  $C$ ,  $\mathcal{R}(C)$ , is closed in  $Z$ , so that  $C^\dagger$ , the generalised inverse of  $C$ , is bounded [6];
- (iii) the problem: find  $f \in X$  such that

$$Lf=0 \quad Cf=0 \tag{4.11}$$

has only the trivial solution  $f=0$ , i.e. the intersection of the null space of  $L$ ,  $\mathcal{N}(L)$ , and the null space of  $C$ ,  $\mathcal{N}(C)$ , contains only the null element of  $X$ .

Let us consider the seminorm, defined on  $\mathcal{D}(C)$ ,

$$p(f) = \|Cf\|_Z. \tag{4.12}$$

Then, by means of minor modifications of the proof of the classical result that a convex and closed set has only one element of minimal norm ([3], p 9) one can derive the following proposition.

*Theorem 4.2.* For each  $g \in Y$  there exists in  $\mathcal{D}(C)$  one and only one solution of equation (4.6) which minimises the seminorm (4.12). This solution depends continuously on the data vector  $g$ .

We will call this pseudosolution the *C-normal pseudosolution* or *C-generalised solution* and denote it by  $f_C^\dagger$ . It defines a *C-generalised inverse*  $L_C^\dagger$ . Notice that  $f_C^\dagger$  does not necessarily belong to  $X_N$  and therefore it can contain ‘invisible’ components. This remark makes clear that the use of the seminorm (4.12) must be suggested by sound physical considerations.

**5. The singular system**

The generalised solution  $f^\dagger$  is a function in  $X_N$  (or, more precisely, in  $X_{N'}$  if some of the  $\phi_n$  are not linearly independent) and therefore it may be expressed in terms of any orthogonal basis of  $X_N$ . Among the possible bases one is the most natural for the problem, namely the set of orthogonal functions  $u_1, u_2, \dots, u_N$  which are transformed by the operator  $L$  into a set of orthogonal vectors  $v_1, v_2, \dots, v_N$ . It is quite easy to recognise that this requirement implies that the functions  $u_k$  must be eigenfunctions of the operator  $L^*L$ . The two bases  $\{u_k\}, \{v_k\}$ , the first in the object space and the second in the data space, can be obtained by means of an extension of a procedure which is known in linear algebra as the singular value decomposition (SVD) of a matrix.

We denote by  $\tilde{L} = L^*L$  the left iterate of  $L$  and by  $\hat{L} = LL^*$  its right iterate. The operator  $\tilde{L}$  is a finite-rank, self-adjoint, non-negative definite operator in  $X$ , whose explicit expression is obtained by combining equations (4.2) and (4.3):

$$\tilde{L}f = \sum_{n,m=1}^N W_{nm}(f, \phi_m)_X \phi_n. \tag{5.1}$$

For instance, in the case where  $X$  is a weighted  $L^2$  space, equation (2.5),  $\tilde{L}$  is an integral operator whose kernel is given by

$$\tilde{T}(x, x') = \sum_{n,m=1}^N W_{nm} \phi_n(x) w(x') \phi_m^*(x'). \tag{5.2}$$

Analogously  $\hat{L}$  is a self-adjoint, non-negative definite (positive definite, when the  $\phi_n$  are linearly independent) operator in the finite-dimensional space  $Y$ . Therefore it can be characterised by means of a matrix. If we put

$$(\hat{L}g)_n = (L^*g, \phi_n) = \sum_{m=1}^N \hat{\mathbf{L}}_{nm} g_m \tag{5.3}$$

then, by combining again equations (4.2) and (4.3), it follows that

$$\hat{\mathbf{L}} = \mathbf{G}^T \mathbf{W}$$

where  $\mathbf{G}^T$  denotes the transpose of the Gram matrix. Since, by assumption,  $\mathbf{W}$  is positive definite, the rank of  $\hat{\mathbf{L}}$  is equal to the rank of  $\mathbf{G}$  and therefore coincides with the number  $N' \leq N$  of linearly independent functionals. We conclude that the matrix  $\hat{\mathbf{L}}$  has  $N'$  positive eigenvalues.

For simplicity we assume  $N' = N$  and we count each eigenvalue as many times as required by its multiplicity. We denote these eigenvalues by  $\alpha_k^2$  and we assume that they are ordered in such a way as to form a non-decreasing sequence

$$\alpha_1^2 \geq \alpha_2^2 \geq \dots \geq \alpha_N^2. \tag{5.5}$$

We denote by  $v_k$  the eigenvector of  $\hat{L}$  associated with  $\alpha_k^2$ . The  $v_k$  form an orthonormal basis in  $Y$ .

The operator  $\tilde{L}$  has the same positive eigenvalues as the operator  $\hat{L}$  with the same multiplicity. We denote by  $u_k$  the eigenfunction of  $\tilde{L}$  associated with  $\alpha_k^2$ . The  $u_k$  form an orthonormal basis in  $X_N$ .

As is known, it is always possible to choose the functions  $u_k$  and the vectors  $v_k$  in such a way as to solve the so called *shifted eigenvalue problem* ([23], ch. 3):

$$Lu_k = \alpha_k v_k \quad L^*v_k = \alpha_k u_k. \tag{5.6}$$

As we see,  $L$  transforms the basis  $\{u_k\}$  into the basis  $\{v_k\}$  except for the scaling factors  $\alpha_k$  and  $L^*$  transforms the basis  $\{v_k\}$  back into the basis  $\{u_k\}$  again except for the same scaling factors  $\alpha_k$ .

We call the set of triples  $\{\alpha_k; u_k, v_k\}_{k=1}^N$  the *singular system* of the operator  $L$ . The numbers  $\alpha_k$  are the *singular values*, the  $u_k$  are the *singular functions* and the  $v_k$  are the *singular vectors*.

The computation of the singular system for a given inverse problem with discrete data requires the solution of standard numerical problems. The singular values  $\alpha_k$  and singular vectors  $v_k$  can be obtained by diagonalising the matrix  $\hat{L}$ . When this matrix has many very small eigenvalues, methods like the well known power method can be suitable for computing the largest eigenvalues and the corresponding eigenvectors. Once  $\alpha_k$  and  $v_k$  have been computed, the corresponding singular function  $u_k$  can be obtained by means of the second equation (5.6) which, written explicitly, gives

$$u_k = \frac{1}{\alpha_k} \sum_{n=1}^N \left( \sum_{m=1}^N W_{nm}(v_k)_m \right) \phi_n. \tag{5.7}$$

It is quite easy to obtain an explicit representation of the generalised solution  $f^\dagger$ . Since  $f^\dagger$  can be represented as a linear combination of the singular functions  $u_k$  and must satisfy equation (4.6), by representing  $g$  as a linear combination of the singular vectors  $v_k$ , we obtain

$$f^\dagger = \sum_{k=1}^N \frac{1}{\alpha_k} (g, v_k)_Y u_k. \tag{5.8}$$

Therefore the algorithm for computing  $f^\dagger$  is very simple. We must compute the components of  $g$  with respect to the basis  $\{v_k\}_{k=1}^N$ , divide these components by the corresponding singular values  $\alpha_k$  and the results are the components of  $f^\dagger$  with respect to the basis  $\{u_k\}_{k=1}^N$ . When the number of linearly independent  $\phi_n$  is  $N'$ , then we have only  $N'$  terms in equations (5.8).

It is also easy to recognise that  $\|L\| = \alpha_1$  and  $\|L^\dagger\| = 1/\alpha_N$ , so that the condition number  $C(L)$  is given by

$$C(L) = \alpha_1/\alpha_N. \tag{5.9}$$

If  $C(L) \gg 1$ , i.e.  $\alpha_N \ll \alpha_1$ , the problem of computing  $f^\dagger$  is *ill-conditioned* and this ill-conditioning is related to the ill-conditioning of the Gram matrix, namely to the approximate linear dependence of the functions  $\phi_n$ . Notice that when we take in  $Y$  the usual euclidean norm ( $W_{nm} = \delta_{nm}$ ) the condition number is just the square root of the ratio between the largest and the smallest eigenvalues of the Gram matrix.

In the case where  $X$  is a space of square-integrable functions and  $Y$  a vector space with the usual euclidean norm, the singular system introduced in this section was essentially

proposed by Twomey [24] (see also Wahba [25]), using somewhat different terms. For instance the Gram matrix is called the ‘covariance matrix’ and the condition number  $C(L)$  is called the ‘worst relative error magnification’. Twomey computes also an ‘average relative error magnification’ in the case where the errors on the data values are random variables. If we denote by  $D(L)$  such an ‘average magnification’, then in our notation it is given by

$$D(L) = \frac{1}{N} \left( \sum_{k=1}^N \alpha_k^{-2} \right)^{1/2} \left( \sum_{k=1}^N \alpha_k^2 \right)^{1/2}. \tag{5.10}$$

One easily verifies that  $D(L)$  is always less than  $C(L)$  and greater than  $C(L)^{-1}$ . Expression (5.10) was computed by Twomey, assuming that the noise random variables are uncorrelated and have the same variance. However, equation (5.10) holds also in the general case if the scalar product in  $Y$  is defined by equations (3.1) and (3.2).

### 6. Interpolation and differentiation

The classical interpolation problem, namely the problem of determining a function  $f$  on an interval  $[a, b]$ , given its values  $g_n$  on  $N$  points  $a \leq x_1 < x_2 < \dots < x_N \leq b$ :

$$f(x_n) = g_n \quad n = 1, \dots, N \tag{6.1}$$

can be considered perhaps as the simplest example of an inverse problem with discrete data. Interpolation may be formulated in the form (1.6) if  $f$  belongs to a Hilbert space  $X$  of continuous functions such that all the evaluation functionals (namely functionals of the form  $F(f) = f(x)$  where  $x$  is a given point of  $[a, b]$ ) are continuous. Then it follows from the Riesz theorem that there exists a function  $Q_x \in X$  such that

$$f(x) = (f, Q_x)_X. \tag{6.2}$$

The function

$$Q(x, x') = (Q_x, Q_{x'})_X = Q_x(x') = Q_{x'}(x) \tag{6.3}$$

is called the *reproducing kernel* of  $X$  and the space  $X$  is said to be a *reproducing kernel Hilbert space* (RKHS) [26].

We will consider two examples: firstly the interpolation problem for band-limited functions and secondly the interpolation and differentiation of ‘smooth’ functions.

#### 6.1. Interpolation of band-limited functions

Let  $X$  be the Paley–Wiener space  $\mathcal{B}_\Omega$  of entire functions of exponential type  $\leq \Omega$ . As is well known, these functions are band-limited with bandwidth  $\Omega$ , namely the Fourier transform of  $f$ ,  $\hat{f}$ , is zero outside the interval  $[-\Omega, \Omega]$ .  $\mathcal{B}_\Omega$  is a Hilbert space with respect to the scalar product of  $L^2(-\infty, +\infty)$  and it is also well known that any function  $f \in \mathcal{B}_\Omega$  satisfies the relation

$$f(x) = \int_{-\infty}^{+\infty} \frac{\sin \Omega(x-x')}{\pi(x-x')} f(x') dx'. \tag{6.4}$$

By comparing this relation with equation (6.2), we see that  $\mathcal{B}_\Omega$  is a RKHS, with a

reproducing kernel given by

$$Q(x, x') = \frac{\sin \Omega(x - x')}{\pi(x - x')} \tag{6.5}$$

Assume now that we wish to find a function  $f \in \mathcal{B}_\Omega$  which interpolates the values given at the points  $x_1, x_2, \dots, x_N$ . Then the problem can be put in the form (1.6), the functions  $\phi_n(x)$  being given by

$$\phi_n(x) = \frac{\sin \Omega(x - x_n)}{\pi(x - x_n)} \quad n = 1, \dots, N. \tag{6.6}$$

The corresponding Gram matrix is

$$G_{nm} = \frac{\sin \Omega(x_n - x_m)}{\pi(x_n - x_m)} \quad n, m = 1, \dots, N. \tag{6.7}$$

In the case of equidistant points sampled at the Nyquist rate,  $x_{n+1} - x_n = \pi/\Omega$ , the Gram matrix is diagonal,  $G_{nm} = (\Omega/\pi)\delta_{nm}$ , and the normal solution defined in § 2 is the interpolating function obtained by truncating the well known Whittaker–Shannon expansion

$$f^\dagger(x) = \sum_{n=1}^N g_n \frac{\sin \Omega(x - x_n)}{\pi(x - x_n)}. \tag{6.8}$$

If the points  $x_n$  are still equidistant, but the distance between adjacent points is less than  $\pi/\Omega$ ,

$$x_n = x_1 + n\sigma \quad n = 0, 1, \dots, N-1 \quad W = \sigma\Omega/2\pi < \frac{1}{2}; \tag{6.9}$$

the Gram matrix multiplied by  $\sigma$  takes the form

$$G_{nm} = \frac{\sin 2\pi W(n - m)}{\pi(n - m)} \quad n = 0, 1, \dots, N-1. \tag{6.10}$$

This matrix has been studied by Slepian [27] who denotes its eigenvalues by  $\lambda_k(N, W)$

$$\lambda_0(N, W) > \lambda_1(N, W) > \dots > \lambda_{N-1}(N, W) \tag{6.11}$$

and defines the *discrete prolate spheroidal sequences* (DPSS) as the real solutions, for  $k = 0, 1, \dots, N-1$ , of the system of equations

$$\sum_{m=0}^{N-1} G_{nm} v_m^{(k)}(N, W) = \lambda_k(N, W) v_n^{(k)}(N, W) \quad n = 0, \pm 1, \pm 2, \dots \tag{6.12}$$

Therefore the eigenvectors of  $\mathbf{G}$  are obtained by index-limiting the DPSS to  $(0, N-1)$ .

By comparing with § 5, we see that the singular system of this interpolation problem is obtained as follows: the singular values  $\alpha_k$  are the square roots of the  $\lambda_k(N, W)$ ,

$$\alpha_k = \lambda_k^{1/2}(N, W) \quad k = 0, \dots, N-1; \tag{6.13}$$

the singular vectors  $v_k$ , normalised with respect to the usual euclidean norm, are obtained by index-limiting the DPSS

$$(v_k)_n = v_n^{(k)}(N, W) \quad k, n = 0, 1, \dots, N-1; \tag{6.14}$$

and finally the singular functions  $u_k$ , normalised with respect to the norm of

$L^2(-\infty, +\infty)$ , are obtained by ‘interpolating’ the  $v_k$ ,

$$u_k(x) = \left( \frac{\sigma}{\lambda_k(N, W)} \right)^{1/2} \sum_{n=0}^{N-1} v_n^{(k)}(N, W) \frac{\sin \Omega(x - x_n)}{\pi(x - x_n)}. \tag{6.15}$$

If we denote by  $X_0$  the length of the interval where the sampling points are given and if we increase the number of sampling points with  $X_0$  fixed, the interpolation problem becomes extremely ill-conditioned. Indeed, from the relation ([27], p 1374)

$$\lambda_{N-1}(N, W) = 1 - \lambda_0(N, \frac{1}{2} - W) \tag{6.16}$$

and from the asymptotic behaviour of  $\lambda_0(N, \frac{1}{2} - W)$  for large  $N$  ([27], p 1387),

$$\lambda_0(N, \frac{1}{2} - W) \simeq 1 - \sqrt{\pi} 2^{9/4} \beta^{1/4} (2 - \beta)^{-1/2} N^{1/2} e^{-\gamma N} \tag{6.17}$$

$$\beta = 1 - \cos(2\pi W) \quad \gamma = \ln \left( 1 + \frac{2\sqrt{\beta}}{2 - \sqrt{\beta}} \right),$$

it follows that

$$C(L) = \left( \frac{\lambda_0(N, W)}{\lambda_{N-1}(N, W)} \right)^{1/2} \simeq \frac{\delta}{N^{1/4}} e^{\gamma N/2} \tag{6.18}$$

where  $\delta$  and  $\gamma$  depend on  $W$ . The number of singular values significantly different from zero is approximately given by  $N_0 = 2WN = (\Omega/\pi)X_0$  and therefore coincides with the number of data points when the function is sampled at the Nyquist rate.

### 6.2. Interpolation of smooth functions

By a ‘smooth’ function we mean a function which has derivatives up to order  $k \geq 1$ . Consider now the class of all smooth functions of order  $k \leq N$ , defined on the interval  $[0, 1]$  and having square-integrable derivatives. We can give to such a class a structure of Hilbert space by introducing the scalar product

$$(f, h)_X = \sum_{j=0}^{k-1} f^{(j)}(0)h^{(j)}(0) + \int_0^1 f^{(k)}(x)h^{(k)}(x) dx \tag{6.19}$$

where  $f^{(j)} = djf/dx^j$ . The space  $X$  is a RKHS and using the Taylor formula

$$f(x) = \sum_{j=0}^{k-1} \frac{f^{(j)}(0)}{j!} x^j + \frac{1}{(k-1)!} \int_0^x (x-x')^{k-1} f^{(k)}(x') dx' \tag{6.20}$$

one can easily derive that the reproducing kernel of  $X$  is

$$Q(x, x') = \sum_{j=0}^{k-1} \frac{x^j x'^j}{(j!)^2} + \sum_{j=0}^{k-1} \frac{(-1)^{k+j+1} x^{2k-j-1} x'^j}{(2k-j-1)! j!} + \frac{(-1)^k}{(2k-1)!} (x-x')_+^{2k-1} \tag{6.21}$$

where  $x_+$  stands for the function which is equal to  $x$  when  $x > 0$  and equal to zero when  $x < 0$ . Therefore, for functions  $f \in X$ , the interpolation problem (6.1), with  $x_n \in [0, 1]$ , can be put again into the form (1.6) with

$$\phi_n(x) = Q(x, x_n) \tag{6.22}$$

and the Gram matrix given by

$$G_{nm} = Q(x_n, x_m) = G_{mn}. \tag{6.23}$$

The normal solution of the problem

$$f^\dagger(x) = \sum_{n=1}^N a_n Q(x, x_n) \tag{6.24}$$

is a *spline of degree*  $m = 2k - 1$ . In the case  $k = 1$ , equation (6.24) gives the usual linear interpolation of the data.

Notice that the  $f^\dagger(x)$  is not a *natural spline* (see [28], p 3). Interpolation by natural splines is obtained when it is required to minimise the seminorm

$$p(f) = \int_0^1 |f^{(k)}(x)|^2 dx \tag{6.25}$$

(see [28], p 6). This seminorm is of the type (4.11), with  $Cf = f^k$ ,  $C: X \rightarrow Z$ ,  $Z = L^2(0, 1)$ . The operator  $C$  is bounded ( $\|Cf\|_Z \leq \|f\|_X$ ) and its null space  $\mathcal{N}(C)$  is the set of all the polynomials of degree  $k - 1$ . On the other hand, the null space  $\mathcal{N}(L)$  of the linear operator  $L$  defined by equation (6.1) is the set of all the functions which are zero at the knots  $x_1, x_2, \dots, x_N$  and since  $k \leq N$ , the intersection of  $\mathcal{N}(C)$  and  $\mathcal{N}(L)$  contains only the function which is zero everywhere. The conditions of theorem 4.2, are fulfilled and therefore there exists a unique  $C$ -normal solution. This solution is just a natural spline.

### 6.3. Numerical differentiation

We consider now the problem of numerical differentiation, namely the problem of computing the derivative  $f'(x)$  of a function  $g(x)$  which is given only at the points  $x_1, x_2, \dots, x_N$ . This problem has many important applications not only in numerical analysis but also in many practical problems, as for example in artificial vision [29]. It is the projection on a finite-dimensional space of a classical example of an  $\text{IPP}$  (see [30, 18]).

Here we assume that  $f$  belongs to the space  $X$  introduced in § 6.2; furthermore we assume that the data have been shifted in such a way that  $g(0) = 0$ . Then the problem can be put in the form (1.4) with

$$F_n(f) = \int_0^{x_n} f(x) dx \quad n = 1, \dots, N. \tag{6.26}$$

These functionals are linear and continuous in  $X$  and from equation (6.20) we have

$$F_n(f) = \sum_{j=0}^{k-1} f^{(j)}(0) \frac{x_n^{j+1}}{(j+1)!} + \frac{1}{k!} \int_0^1 f^{(k)}(x') (x_n - x')_+^k dx'. \tag{6.27}$$

Therefore equation (1.5) holds with

$$\phi_n(x) = \sum_{j=0}^{k-1} \frac{x_n^{j+1} x^j}{(j+1)! j!} + \sum_{j=0}^{k-1} (-1)^{k+j+1} \frac{x_n^{2k-j} x^j}{(2k+j)! j!} + \frac{(-1)^k}{(2k)!} (x_n - x)_+^{2k}. \tag{6.28}$$

It follows that the normal solution is the derivative of a spline of degree  $m' = 2k + 1$ . In other words the procedure is equivalent to interpolating the data  $g_n$  by a spline of degree  $m' = 2k + 1$  and then differentiating the result.

Notice that the functionals (6.26) are continuous also in  $L^2(0, 1)$ . If we take  $X = L^2(0, 1)$  then the normal solution is the step function which is obtained by differentiating the linear interpolation of the data.

## 7. Moment problem and Laplace transform inversion

According to the definition of Stieltjes, the moment problem is to find a 'distribution of positive mass on the interval  $[0, +\infty)$ , given the moments of order  $n$  ( $n=0, 1, 2, \dots$ ) of the distribution' [31]. Here we consider the case where the function to be recovered is not necessarily positive and the number of given moments is finite. As is known, a problem strictly related to the moment problem is the inversion of the Laplace transformation when the Laplace transform is given at equidistant points [32].

### 7.1. Hausdorff moment problem

In this case the unknown function  $f(x)$  is assumed to be supported on a bounded interval which, without loss of generality, can be assumed to be the interval  $[0, 1]$ . Then the problem takes the form (1.4) with

$$F_n(f) = \int_0^1 x^{n-1} f(x) dx \quad n = 1, \dots, N. \quad (7.1)$$

The solution of this problem is unique in the case  $N = \infty$ . Furthermore it is possible to give a complete characterisation of the sequences  $\{g_n\}_{n=1}^\infty$  which are the moments of a function  $f \in L^2(0, 1)$  ([32], p 109), i.e. a solubility criterion.

Assume  $N < \infty$ , take  $X = L^2(0, 1)$  and  $Y$  an  $N$ -dimensional vector space with the usual euclidean norm. Then the moment problem has the form (1.6) with

$$\phi_n(x) = x^{n-1} \quad n = 1, \dots, N. \quad (7.2)$$

The adjoint operator  $L^*$ , equation (4.3), is

$$(L^*g)(x) = \sum_{n=1}^N g_n x^{n-1} \quad (7.3)$$

and therefore the range of  $L^*$ ,  $\mathcal{R}(L^*) = X_N$ , is the subspace of the polynomials of degree  $N-1$ . It follows that the normal solution, defined in § 2, is also a polynomial of degree  $N-1$ .

A standard procedure for computing the normal solution (see, for instance, [33, 34]) is to represent  $f^\dagger(x)$  in terms of shifted Legendre polynomials

$$f^\dagger(x) = \sum_{m=1}^N c_m P_{m-1}(2x-1) \quad (7.4)$$

and to use their orthogonality properties in order to find the relation

$$g_n = \sum_{m=1}^n \beta_{nm} c_m \quad n = 1, \dots, N \quad (7.5)$$

where  $g_n$  is the given value of the moment of order  $n$  and

$$\beta_{nm} = \int_0^1 x^{n-1} P_{m-1}(2x-1) dx. \quad (7.6)$$

Since the coefficients  $\beta_{nm}$  can be computed, it is possible to obtain recursively the coefficients  $c_n$  from the moments  $g_n$ . The important feature is that the coefficient  $c_n$  depends only on the moments of order  $\leq n$  and therefore it is not necessary to change the algorithm when the number  $N$  of given moments is changed.

The drawback of the previous method is that it masks the severe ill-conditioning of the problem. If we use the method of § 5, we find that the operator  $\tilde{L} = L^*L$  is the integral operator

$$(\tilde{L}f)(x) = \int_0^1 \frac{1 - (xx')^N}{1 - xx'} f(x') dx' \tag{7.7}$$

and that the matrix  $\tilde{L}$  is just the Hilbert matrix

$$\tilde{L}_{nm} = [H_N(-1)]_{nm} = 1/(n + m - 1). \tag{7.8}$$

As usual we define  $[H_N(\theta)]_{nm} = (n + m + \theta)^{-1}$ ;  $n, m = 1, 2, \dots, N$ .

The Hilbert matrix is a classical example of an ill-conditioned matrix and is often used for testing numerical algorithms [35]. However, the eigenvectors of  $[H_N(\theta)]$  can be easily computed since they are also eigenvectors of a tridiagonal matrix  $[T_N(\theta)]$  [36].

Now the singular values of the moment problem are just the square roots of the eigenvalues of  $[H_N(-1)]$ , the singular vectors are the eigenvectors of  $[H_N(-1)]$  and the singular functions form a set of orthogonal polynomials of degree  $N - 1$ . We just mention that these polynomials are the orthogonal polynomials which appear in the problem of least-squares approximation of a function on  $[0, 1]$  by means of a polynomial of degree  $N - 1$ . Indeed such a problem is equivalent to the computation of the normal solution of the moment problem considered in this section.

In order to give an idea of the ill-conditioning of the moment problem, we give the condition numbers, derived from the eigenvalues of  $[H_N(-1)]$  quoted in [35], in the cases  $N = 3, 4, 5, 6$ :

$$C(L) = 22.9, 124.6, 690.6, 3865. \tag{7.9}$$

For large  $N$  the asymptotic estimate of  $C(L)$  is [35]

$$C(L) = e^{1.75N}. \tag{7.10}$$

Therefore if we want a normal solution stable with respect to errors of 0.1% on the moments, we can give at most five moments of the function  $f$ . However, when  $N$  increases, using the singular system analysis one can verify that the information on the function  $f$  increases, even if rather slowly. For instance in the case  $N = 100$ , we have estimated that one can extract nine components from the data since  $\alpha_1/\alpha_9 = 505.8$  while  $\alpha_1/\alpha_{10} = 1319$ .

This extreme ill-conditioning of the moment problem should make it clear that one must be very prudent in the use of moments for pattern recognition [34]. An open question is the investigation of the singular systems for moment problems in two and three dimensions.

### 7.2. Stieltjes and Hamburger moment problems

The problem of Stieltjes is to determine a function  $f$  defined on  $[0, \infty)$  from the knowledge of its moments, while the problem of Hamburger is to determine a function on  $(-\infty, +\infty)$  ([32], p 125). In the first case the functionals  $F_n$  are given by

$$F_n(f) = \int_0^{+\infty} x^{n-1} f(x) dx \tag{7.11}$$

while in the second case the functionals are

$$F_n(f) = \int_{-\infty}^{+\infty} x^{n-1} f(x) dx. \quad (7.12)$$

It is known that the uniqueness result, which holds for the Hausdorff problem, does not hold for these problems. Examples of functions, defined on  $[0, \infty)$  and such that all their moments are zero, were already given by Stieltjes ([32], p 126).

The functionals (7.11) or (7.12) are not continuous in  $L^2(0, +\infty)$  and  $L^2(-\infty, +\infty)$  respectively. However, they are continuous in suitable weighted spaces. If we consider a scalar product of the form (2.5) with  $w(x) \neq 0$  everywhere, then the functionals (7.11) or (7.12) are continuous for any  $n$  provided that all the moments of the function  $1/w(x)$  exist, i.e. in the case (7.11)

$$\int_0^{+\infty} \frac{x^{n-1}}{w(x)} dx < +\infty \quad n = 1, 2, \dots \quad (7.13)$$

Then the functionals take the form (1.5) with

$$\phi_n(x) = \frac{x^{n-1}}{w(x)} \quad n = 1, 2, \dots, \quad (7.14)$$

and the Gram matrix is given by

$$G_{nm} = (\phi_n, \phi_m)_X = \int_0^{+\infty} \frac{x^{n+m-2}}{w(x)} dx. \quad (7.15)$$

In the case where  $N$  moments are specified, the singular functions  $u_k(x)$  are polynomials of degree  $N-1$ , multiplied by  $1/w(x)$ .

### 7.3. Laplace transform inversion

As usual we define the Laplace transform of a function  $f(t)$  defined on  $[0, +\infty)$  by

$$g(p) = \int_0^{+\infty} e^{-pt} f(t) dt. \quad (7.16)$$

The problem of Laplace transform inversion is closely related to the moment problem. If we make the change of variable  $t = -\ln x$  in equation (7.16), we obtain

$$g(p) = \int_0^1 x^{p-1} f(-\ln x) dx. \quad (7.17)$$

Therefore the Hausdorff moment problem is just equivalent to the inversion of the Laplace transform when the latter is given at the points  $p_n = n$  ( $n = 1, 2, 3, \dots$ ).

More generally, consider a finite set of equidistant points

$$p_n = c + d(n-1) \quad n = 1, \dots, N \quad (7.18)$$

with  $c \geq 0$  and a space of functions with the norm

$$\|f\|_X^2 = \int_0^{+\infty} e^{2at} |f(t)|^2 dt \quad (a \geq 0). \quad (7.19)$$

Then the problem of estimating  $f$ , given  $g$  at the points  $p_n$ , takes the form (1.6) with

$$\phi_n(t) = \exp\{-[2a + c + d(n - 1)]t\}. \tag{7.20}$$

Notice that it is not possible to have  $a = 0$  and  $c = 0$  simultaneously, because in such a case  $\phi_1(t) = 1$  and therefore  $\phi_1$  does not belong to  $X$ .

If we introduce in  $Y$  a scalar product given by (3.1) with  $W_{nm} = c\delta_{nm}$ , then it is easy to verify that the matrix  $\hat{L}$  is again a Hilbert matrix

$$\hat{L}_{nm} = [H_N(\theta)]_{nm} \tag{7.21}$$

with

$$\theta = 2\left(\frac{a + c}{d} - 1\right). \tag{7.22}$$

The method described in this paper was applied by two of the authors to the problem of Laplace transform inversion both in the case where the function  $f$  is supported in a bounded interval  $[a, b]$  ( $0 < a < b < \infty$ ) [37] and in the case where it belongs to a weighted  $L^2$  space [38]. Two distributions of data points were considered: equidistant points, as in equation (7.18), and points forming a geometric progression

$$p_n = c\Delta^{n-1} \quad n = 1, \dots, N. \tag{7.23}$$

In the latter case the scalar product in the data space was defined by (3.1) with  $W_{nm} = (n\Delta)p_n\delta_{nm}$  and this choice was justified by an extension of the sampling theorem [37]. ‘Optimality’ criteria for the choice of the data points were introduced and it was found that a small number of geometrically spaced points can give better results than a much larger set of equidistant points [37, 38].

### 8. Fourier transform inversion

A basic problem in many domains of applied science, for instance spectrum estimation, is the determination of an unknown function from sampled values of its Fourier transform. In the simplest case the unknown function  $f$  has a bounded support, say  $[-\Omega, \Omega]$ .

Assume that its Fourier transform

$$g(t) = \int_{-\Omega}^{\Omega} f(\omega)e^{-it\omega} d\omega \tag{8.1}$$

is given at the points

$$t_n = n\sigma \quad n = 0, 1, \dots, N - 1 \tag{8.2}$$

with  $\sigma \leq \pi/\Omega$ , the Nyquist sampling distance. Introduce the new variable

$$x = (\sigma/2\pi)\omega \tag{8.3}$$

and define  $W$  as in equation (6.9) (here  $W \leq \frac{1}{2}$ ). Then the problem takes the form (1.6) with

$$F_n(f) = \int_{-W}^W f(x)e^{-i2\pi nx} dx \quad n = 0, 1, \dots, N - 1 \tag{8.4}$$

and these functionals are continuous in  $L^2(-W, W)$ . When  $W = \frac{1}{2}$ , the functions

$$\phi_n(x) = e^{2i\pi nx} \quad n = 0, 1, \dots, N - 1 \tag{8.5}$$

are orthonormal. This is just the case where the data values  $g_n = g(t_n)/2\pi$  are the Fourier coefficients of  $f$  and the normal solution is a truncated Fourier series expansion

$$f^\dagger(x) = \sum_{n=0}^{N-1} g_n e^{i2\pi nx}. \quad (8.6)$$

In the case  $W < \frac{1}{2}$ , i.e. in the case of oversampled data, the problem becomes ill-conditioned and the normal solution can be expressed in terms of the *discrete prolate spheroidal functions* (DPSF) of Slepian [27]. Indeed, if we take the usual euclidean norm in the data space  $Y$ , the operator  $\tilde{L} = L^*L$  is an integral operator given by

$$(\tilde{L}f)(x) = \int_{-W}^W \tilde{T}(x-x')f(x') dx' \quad (8.7)$$

with

$$\tilde{T}(x) = e^{i\pi(N-1)x} \frac{\sin(\pi Nx)}{\sin(\pi x)} \quad (8.8)$$

and the matrix  $\hat{L}$ , i.e. the Gram matrix of the functions (8.5), coincides with the matrix (6.10). Therefore the singular values of the problem are given again by equation (6.13) and the singular vectors by equation (6.14). It is easy to recognise that the singular functions  $u_k(x)$  are just the DPSF, except for a normalisation factor. Indeed the relation given by Slepian ([27], p 1380) between DPSF and DPSS is essentially the second equation of (5.6). The exact relation, taking into account normalisation, is

$$u_k(x) = (\alpha_k \varepsilon_k)^{-1} e^{i\pi(N-1)x} U_k(N, W; x) \quad (8.9)$$

where  $\varepsilon_k = 1$  when  $k$  is even and  $\varepsilon_k = i$  when  $k$  is odd. In the limit  $W \rightarrow 0$ ,  $N \rightarrow \infty$ ,  $\pi NW \rightarrow c$ , the functions  $\sqrt{W}U_k(N, W; Wx)$  converge to the usual prolate spheroidal functions  $\psi_k(c; x)$ .

The problem of Fourier transform inversion has also been considered in the case where  $X$  is a weighted  $L^2$  space [39] in order to take into account *a priori* knowledge of shape and extent of  $f(x)$ . If the scalar product is defined as in equation (2.5) and we assume that  $w(x) > 0$  everywhere on  $(-\infty, +\infty)$ , then the matrix  $\hat{L}$  is given by

$$\hat{L}_{nm} = \int_{-\infty}^{+\infty} \frac{e^{-i2\pi(n-m)x}}{w(x)} dx. \quad (8.10)$$

The singular system of this problem may be computed for specific cases of the weighting function.

## 9. Radon and Abel transform inversion

According to John [40], the term *Radon transform* indicates the representation of an arbitrary function  $f$ , defined on  $\mathbb{R}^n$ , by means of its integrals over hyperplanes of dimension  $(n-1)$ . Nowadays this term is used for indicating the inverse of this mapping, and therefore the representation, called the Radon transform by John, becomes the inversion formula of the Radon transform.

The inversion of the Radon transform with discrete data has many important applications, especially in x-ray tomography, but also in radioastronomy, radar target estimation, combustion diagnostics and so on. Presumably the analysis in terms of singular systems, proposed in this paper, is not feasible in practice at least in the case of a large number of data. However, it should be of some interest in so called 'limited angle' tomography or in other cases of limited experimental data.

9.1. Radon transform inversion

For simplicity we consider the case of functions of two variables. Then the Radon transform is the mapping  $f \rightarrow P_\theta f$  defined by

$$(P_\theta f)(p) = \int_{-\infty}^{+\infty} f(pe + qe') \, dq \quad p \in \mathbb{R} \tag{9.1}$$

where  $e = \{\cos \theta, \sin \theta\}$ ,  $e' = \{-\sin \theta, \cos \theta\}$ . It is clear that  $(P_\theta f)(p)$  is the integral of  $f$  over the straight line perpendicular to  $e$  and with signed distance  $p$  from the origin. For fixed  $\theta$  and  $p$  this integral is a functional of  $f$  but it is not continuous in  $L^2(\mathbb{R}^2)$ . Continuous functionals can be obtained if we take into account the integration effect of the detectors discussed in the introduction. Another way is to assume that  $f$  belongs to a suitable Sobolev space [17]. More precisely we can assume that  $f$  belongs to a weighted Sobolev space so that the weight can incorporate *a priori* knowledge about the localisation of the solution [8].

Thus we put

$$f(\mathbf{x}) = P(\mathbf{x})\phi(\mathbf{x}). \tag{9.2}$$

Assuming that  $\phi$  is square integrable is equivalent to taking for  $f$  a norm as in equation (2.5) with a weight function defined by

$$w(\mathbf{x}) = \frac{1}{P^2(\mathbf{x})}. \tag{9.3}$$

We can call  $P(\mathbf{x})$  the 'profile function'. We assume now that  $\phi$  belongs to the Sobolev space  $H^\alpha(\mathbb{R}^2)$ ,  $\alpha > 0$ . If  $\hat{\phi}$  is the Fourier transform of  $\phi$ , then the norm of  $H^\alpha(\mathbb{R}^2)$  is defined by

$$\|\phi\|_\alpha = \frac{1}{2\pi} \left( \int_{\mathbb{R}^2} (1 + |\xi|^2)^\alpha |\hat{\phi}(\xi)|^2 \, d\xi \right)^{1/2}. \tag{9.4}$$

In the case  $\alpha = 0$ ,  $H^0(\mathbb{R}^2) = L^2(\mathbb{R}^2)$ . In the case  $\alpha = 1$ , the norm (9.4) can be written as

$$\|\phi\|_1^2 = \int_{\mathbb{R}^2} |\phi(\mathbf{x})|^2 \, d\mathbf{x} + \int_{\mathbb{R}^2} |\nabla\phi(\mathbf{x})|^2 \, d\mathbf{x} \tag{9.5}$$

where  $|\nabla\phi(\mathbf{x})|^2$  is the modulus squared of the gradient of  $\phi$ .

If the Radon transform of  $f$  is given at the points  $\{\theta_1, p_1\}, \{\theta_2, p_2\}, \dots, \{\theta_N, p_N\}$ , then we have to investigate the functionals

$$F_n(f) = \int_{-\infty}^{+\infty} f(p_n e_n + qe'_n) \, dq \quad n = 1, \dots, N. \tag{9.6}$$

If we take, for simplicity, a profile function with circular symmetry, i.e.  $P(\mathbf{x}) = P(\rho)$  with  $\rho = (x_1^2 + x_2^2)^{1/2}$ , then the functionals (9.6) are continuous in  $H^\alpha(\mathbb{R}^2)$  provided that the

functions  $\phi_n(\mathbf{x})$ , whose Fourier transforms are given by

$$\hat{\phi}_n(\xi) = (1 + |\xi|^2)^{-\alpha} \exp(-ip_n(e_n, \xi)) \int_{-\infty}^{+\infty} P[(p_n^2 + q^2)^{1/2}] \exp(-iq(e'_n, \xi)) dq \tag{9.7}$$

belong to  $H^\alpha(\mathbb{R}^2)$  for  $n = 1, \dots, N$ .

The proof follows from the relation

$$\begin{aligned} F_n(f) &= \int_{-\infty}^{+\infty} P[(p_n^2 + q^2)^{1/2}] \phi(p_n e_n + q e'_n) dq \tag{9.8} \\ &= \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} P[(p_n^2 + q^2)^{1/2}] \left( \int_{\mathbb{R}^2} \hat{\phi}(\xi) \exp(ip_n(e_n, \xi) + iq(e'_n, \xi)) d\xi \right) dq \\ &= \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} (1 + |\xi|^2)^\alpha \hat{\phi}(\xi) \hat{\phi}_n^*(\xi) d\xi, \end{aligned}$$

$\hat{\phi}_n(\xi)$  being defined in equation (9.7).

In the case of a gaussian profile, i.e.

$$P(\rho) = \exp(-\rho^2/2) \tag{9.9}$$

we have

$$\hat{\phi}_n(\xi) = \sqrt{2\pi} (1 + |\xi|^2)^{-\alpha} \exp(-\frac{1}{2} p_n^2 - ip_n(e_n, \xi) - \frac{1}{2} |(e'_n, \xi)|^2) \tag{9.10}$$

and it is easy to prove that  $\phi_n \in H^\alpha(\mathbb{R}^2)$  for any  $\alpha > 0$ .

Given the data points  $\{\theta_n, p_n\}$ , the profile function  $P(\mathbf{x})$  and the Sobolev space  $H^\alpha(\mathbb{R}^2)$ , one can compute the functions  $\phi_n$  and the corresponding Gram matrix and apply the general formalism described in the previous sections.

### 9.2. Abel transform inversion

When the function  $f$  has circular symmetry, i.e.  $f$  is a function of  $\rho = (x_1^2 + x_2^2)^{1/2}$ , then all the ‘projections’  $P_\theta f$  coincide. If we put  $\theta = 0$  in equation (9.1), so that  $p = x_1$  and  $q = x_2$ , in terms of the variable  $\rho$  we have

$$(P_0 f)(p) = 2 \int_p^{+\infty} \rho (\rho^2 - p^2)^{-1/2} f(\rho) d\rho. \tag{9.11}$$

If the projection  $g(p) = (P_0 f)(p)$  is given for any  $p$ , then the problem is equivalent to the inversion of the Abel integral equation and it is related to the computation of fractional derivatives of a function  $f$  ([41], p 115).

First assume that the function  $f$  has a bounded support, say  $[0, 1]$ . Since, for fixed  $p$ , equation (9.11) does not define a continuous functional in  $L^2(0, 1)$ , the previous remark suggests to use the space  $X$  introduced in the problem of numerical differentiation, § 6.3. Consider the simplest case, namely the case where the scalar product is defined by equation (6.19) with  $k = 1$ , corresponding to the reproducing kernel

$$Q(\rho, \rho') = 1 + \rho - (\rho - \rho')_+ = 1 + \min\{\rho, \rho'\}. \tag{9.12}$$

If the data values are given at the points  $p_1, p_2, \dots, p_N$ , each data point corresponds to the

functional

$$F_n(u) = 2 \int_{p_n}^1 \rho(\rho^2 - p_n^2)^{-1/2} f(\rho) \, d\rho \quad n = 1, \dots, N. \tag{9.13}$$

We can prove that, for each  $n$ ,  $F_n$  is a continuous functional in the RKHS with reproducing kernel (9.12).

By means of a partial integration we get

$$\begin{aligned} F_n(f) &= 2(1 - p_n^2)f(1) - 2 \int_{p_n}^1 (\rho^2 - p_n^2)^{1/2} f'(\rho) \, d\rho \\ &= 2(1 - p_n^2)^{1/2} f(0) + 2 \int_0^1 [(1 - p_n^2)^{1/2} - (\rho^2 - p_n^2)^{1/2}] f'(\rho) \, d\rho \end{aligned} \tag{9.14}$$

and therefore  $F_n(f) = (f, \phi_n)_X$ , where

$$\phi_n(\rho) = \int_0^1 Q(\rho, \rho') \psi_n(\rho') \, d\rho' \tag{9.15}$$

with

$$\psi_n(\rho) = \begin{cases} 0 & 0 \leq \rho < p_n \\ 2(\rho^2 - p_n^2)^{-1/2} & p_n < \rho \leq 1. \end{cases} \tag{9.16}$$

One can verify that

$$\phi_n(\rho) = \begin{cases} 2(1 + \rho)(1 - p_n^2)^{1/2} & 0 \leq \rho < p_n \\ 2(1 + \rho)(1 - p_n^2)^{1/2} - 2 \int_{p_n}^{\rho} (\rho'^2 - p_n^2)^{1/2} \, d\rho' & p_n \leq \rho \leq 1. \end{cases} \tag{9.17}$$

The functions  $\phi_n$  can be easily computed and the corresponding Gram matrix is given by

$$G_{nm} = (\phi_n, \phi_m)_X = (\phi_n, \psi_m)_{L^2(0, 1)}. \tag{9.18}$$

In the case where the support of  $f$  is not bounded, one can use again Sobolev spaces, defined now in terms of the Mellin transform of  $f$ . We notice that the functional  $F_n(f)$  can be written in the form

$$F_n(f) = \int_0^{+\infty} K(p_n/\rho) \phi(\rho) \, d\rho/\rho, \tag{9.19}$$

i.e. as a value of a convolution integral (with respect to the dilation group). In equation (9.19),  $K(x) = 2(1 - x^2)^{-1/2}$  for  $0 \leq x < 1$ ,  $K(x) = 0$  for  $x > 1$  and  $\phi(\rho) = \rho f(\rho)$ . Then, introducing the Mellin transform of  $\phi$  [42]

$$(\mathcal{M}\phi)(\xi) = \int_0^{+\infty} \phi(\rho) \rho^{-\xi + i\zeta} \, d\xi \tag{9.20}$$

and using well known transforming properties of the Mellin transform with respect to the convolution product and also the Mellin transform inversion formula [42], we get

$$F_n(\phi) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (\mathcal{M}K)(\xi) (\mathcal{M}\phi)(\xi) p_n^{-(\xi + i\zeta)} \, d\xi \tag{9.21}$$

where

$$(\mathcal{MK})(\xi) = \sqrt{\pi} \frac{\Gamma(\frac{1}{4} + \frac{1}{2}i\xi)}{\Gamma(\frac{3}{4} + \frac{1}{2}i\xi)} \tag{9.22}$$

If we introduce now the Sobolev space  $H^\alpha(\mathbb{R}_+)$  defined by the norm

$$\|\phi\|_\alpha^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (1 + \xi^2)^\alpha |(\mathcal{M}\phi)(\xi)|^2 d\xi \tag{9.23}$$

and use the fact that when  $|\xi| \rightarrow \infty$

$$|(\mathcal{MK})(\xi)| \simeq (2\pi/|\xi|)^{1/2} \tag{9.24}$$

it is easy to show that, except in the case  $p_n=0$ , the functionals  $F_n$  can be written in the form  $F_n(\phi) = (\phi, \phi_n)_\alpha$  with

$$(\mathcal{M}\phi_n)(\xi) = \frac{(\mathcal{MK})^*(\xi)}{(1 + \xi^2)^\alpha} p_n^{-\frac{1}{2} + i\xi} \tag{9.25}$$

and that  $\phi_n \in H^\alpha(\mathbb{R}_+)$  for any  $\alpha > 0$ .

**10. Fredholm equations of the first kind**

A Fredholm integral equation of the first kind is an equation of the form

$$\int_a^b K(y, x)f(x) dx = g(y) \quad c \leq y \leq d \tag{10.1}$$

where  $g(y)$  and  $K(y, x)$  are given functions and  $f(x)$  is the function to be determined. Particular cases of Fredholm equations have already been considered in the previous sections since Laplace, Fourier and Abel transform inversions are just problems of the type (10.1).

In the case of square-integrable kernels, conditions for existence and uniqueness of the solution of equation (10.1) were given by Picard [43]. This analysis, which is based on the singular system associated with the integral operator defined by the kernel  $K(y, x)$ , can be extended to the case of an arbitrary compact operator. The application of Tikhonov regularisation to Fredholm equations of the first kind is given in [44] in the case of a compact integral operator.

Here we consider essentially the case where the intervals  $[a, b]$ ,  $[c, d]$  are bounded, and the kernel  $K(y, x)$  is a bounded continuous function. We introduce the operator

$$(L_0 f)(y) = \int_a^b K(y, x)f(x) dx \quad c \leq y \leq d \tag{10.2}$$

and we consider  $L_0$  as an operator from  $X = L^2(a, b)$  into  $Y_0 = L^2(c, d)$ . Then  $L_0$  is compact and we denote its singular system by  $\{\alpha_{0,k}; u_{0,k}, v_{0,k}\}_{k=1}^\infty$ .

Assume now that  $y_1, \dots, y_N$  and  $w_1, \dots, w_N$  are respectively the ‘abscissas’ and the ‘weights’ of some gaussian quadrature formula, so that, by using a sufficiently high  $N$ , one can obtain any desired degree of accuracy in the computation of the integral over  $[c, d]$  of a continuous function. Accordingly we assume that the data function  $g(y)$  is given at the

points  $y_n$  and we introduce in the data space  $Y$  the scalar product

$$(\mathbf{g}, \mathbf{h}) = \sum_{n=1}^N w_n g_n h_n^* \tag{10.3}$$

i.e. the scalar product (3.1) with  $W_{nm} = w_n \delta_{nm}$ . If we take again  $X=L^2(a, b)$ , then the problem of solving equation (10.1) in the case of discrete data has the form (1.6) with

$$\phi_n(x) = K(y_n, x) \quad n = 1, \dots, N. \tag{10.4}$$

We investigate the relationship between the singular system of the operator  $L$  corresponding to discrete data and the singular system of the operator  $L_0$ , equation (10.2).

The singular functions  $u_{0,k}$  are eigenfunctions of the integral operator  $\tilde{L}_0 = L_0^* L_0$ , associated with the eigenvalues  $\alpha_{0,k}^2$ . The kernel of  $\tilde{L}_0$  is

$$\tilde{T}_0(x, x') = \int_c^d K^*(y, x) K(y, x') dy. \tag{10.5}$$

Analogously, the singular functions  $u_k$  are eigenfunctions of  $\tilde{L} = L^* L$ , associated with the eigenvalues  $\alpha_k^2$  and, from equation (5.2), it follows that the kernel of  $\tilde{L}$  is given by

$$\tilde{T}(x, x') = \sum_{n=1}^N w_n K^*(y_n, x) K(y_n, x'). \tag{10.6}$$

Clearly  $\tilde{T}(x, x')$  is just the approximation of  $\tilde{T}_0(x, x')$  provided by the gaussian quadrature formula we have considered.

*Theorem 10.1.* When  $N \rightarrow \infty$  the singular values  $\alpha_k$  converge to the corresponding singular values  $\alpha_{0,k}$  and the singular functions  $u_k$  converge, in the norm of  $L^2(a, b)$ , to the corresponding singular functions  $u_{0,k}$ .

*Proof.* Consider the kernel  $\tilde{R}(x, x') = \tilde{T}_0(x, x') - \tilde{T}(x, x')$ . Since  $K(y, x)$  is continuous,  $\tilde{R}(x, x')$  tends to zero for any  $x, x'$  when  $N \rightarrow \infty$ . Furthermore  $\tilde{R}(x, x')$  is bounded by a constant independent of  $N$  (by assumption,  $K(y, x)$  is bounded and the sum of the weights is just the length of the interval,  $b - a$ ). Therefore, from the dominated convergence theorem it follows that

$$\lim_{N \rightarrow \infty} \int_a^b \left( \int_a^b |\tilde{R}(x, x')|^2 dx' \right) dx = 0. \tag{10.7}$$

The Weyl–Courant lemma [45] implies the inequality

$$|\alpha_k^2 - \alpha_{0,k}^2| \leq \|\tilde{R}\| \tag{10.8}$$

where  $\|\tilde{R}\|$  denotes the norm of the integral operator associated with the kernel  $\tilde{R}(x, x')$ . From equation (10.7) we get  $\|\tilde{R}\| \rightarrow 0$  and therefore the first part of the theorem is proved. The second part can be proved as in [46] by investigating the eigenprojections over the subspaces spanned by the eigenfunctions.

In the case of truncated convolution operators with a band-limited kernel, one can also obtain convergence without decreasing the distance between the data points to zero just by increasing to infinity the number of points taken at a fixed distance. A paradigm for these

equations is provided by the integral operator

$$(L_0 f)(y) = \int_{-1}^1 \frac{\sin c(y-x)}{\pi(y-x)} f(x) dx \tag{10.9}$$

with  $y$  ranging from  $-\infty$  to  $+\infty$ .

This case has been investigated in [15, 46]. In [46] it was assumed that data values are given at the points

$$y_n = n\sigma \quad n = 0, \pm 1, \dots, \pm N \quad \sigma \leq \pi/c \tag{10.10}$$

where  $\pi/c$  is the usual Nyquist distance. Furthermore  $X$  was  $L^2(-1, 1)$  and the scalar product in  $Y$  was defined by (3.1) with  $W_{nm} = \sigma \delta_{nm}$ .

When  $N \rightarrow \infty$  with fixed  $\sigma$ , the singular values of the problem with discrete data converge to the square roots of the eigenvalues of the prolate spheroidal functions  $\psi_k(c, x)$  and the singular functions  $u_k(x)$  converge to  $\psi_k(c, x)$ .

We conclude this section by giving some properties of the singular system of this remarkable problem.

If we denote by  $v_k(n\sigma)$  the  $n$ th component of the singular vector  $v_k$ , then the equations (5.6) imply

$$u_k(x) = \frac{\sigma}{\alpha_k} \sum_{n=-N}^N \frac{\sin c(x-n\sigma)}{\pi(x-n\sigma)} v_k(n\sigma) \tag{10.11}$$

$$v_k(n\sigma) = \frac{1}{\alpha_k} \int_{-1}^1 \frac{\sin c(x-n\sigma)}{\pi(x-n\sigma)} u_k(x) dx. \tag{10.12}$$

We notice that each singular vector  $v_k$  can be extended into a sequence  $\{v_k(n\sigma)\}_{n=-\infty}^{+\infty}$  just by allowing  $n$  to range from  $-\infty$  to  $+\infty$  in equation (10.12).

When  $\sigma = \pi/c$ , i.e. data are sampled at the Nyquist rate, from (10.11) it follows that  $\alpha_k u_k(n\sigma) = v_k(n\sigma)$  for  $n = 0, \pm 1, \dots, \pm N$ , while the relation does not hold for  $|n| > N$  since, in such a case,  $u_k(n\sigma) = 0$ . Furthermore the singular functions  $u_k(x)$  are also orthogonal over  $(-\infty, +\infty)$ , as follows from the relation

$$\int_{-\infty}^{+\infty} u_k(x) u_j(x) dx = \frac{\sigma}{\alpha_k \alpha_j} \sum_{n=-N}^N v_k(n\sigma) v_j(n\sigma) = \frac{1}{\alpha_k^2} \delta_{kj}. \tag{10.13}$$

In the case  $\sigma < \pi/c$ , i.e. in the case of oversampled data, both equation (10.13) and the relation between the values of  $u_k(x)$  and the components of  $v_k$  do not hold. However, it is interesting to notice a band-limiting property of the sequences  $\{v_k(n\sigma)\}_{n=-\infty}^{+\infty}$ . Define the periodic function

$$V_k(\omega) = \sigma \sum_{n=-\infty}^{+\infty} v_k(n\sigma) e^{in\sigma\omega} \tag{10.14}$$

and consider the fundamental period  $[-b, b]$ , where  $b = \pi/\sigma > c$ . Then  $V_k(\omega)$  is band-limited to the interval  $[-c, c]$ . Indeed, one can easily verify that equation (10.12) can be written in the form

$$v_k(n\sigma) = \frac{1}{2\pi} \int_{-c}^c \left( \frac{1}{\alpha_k} \int_{-1}^1 u_k(x) e^{i\omega x} dx \right) e^{-in\sigma\omega} d\omega \tag{10.15}$$

and therefore

$$V_k(\omega) = \begin{cases} \frac{1}{\alpha_k} \int_{-1}^1 u_k(x) e^{i\omega x} dx & |\omega| < c \\ 0 & c < |\omega| < b \end{cases} \quad (10.16)$$

as follows from the uniqueness of the Fourier coefficients.

### Acknowledgment

This work was partially supported by NATO grant no 463/84. Dr C De Mol is 'Chercheur qualifié' of the Belgian National Fund for Scientific Research.

### References

- [1] Backus G and Gilbert F 1968 The resolving power of gross Earth data *Geophys. J. R. Astron. Soc.* **16** 169–205
- [2] Backus G and Gilbert F 1970 Uniqueness in the inversion of inaccurate gross Earth data *Phil. Trans. R. Soc.* **266** 123–92
- [3] Balakrishnan A V 1976 *Applied Functional Analysis* (Berlin: Springer)
- [4] Nashed M Z 1981 Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory *IEEE Trans. Antennas Propagat.* **AP-29** 220–31
- [5] Rust B W and Burrus W R 1972 *Mathematical Programming and the Solution of Linear Equations* (New York: Elsevier)
- [6] Groetsch C W 1977 *Generalized Inverses of Linear Operators* (New York: Dekker)
- [7] Courant R and Hilbert D 1953 *Methods of Mathematical Physics* vol. 1 (New York: Interscience)
- [8] Byrne C L and Fitzgerald R M 1982 Reconstruction from partial information with applications to tomography *SIAM J. Appl. Math.* **42** 933–40
- [9] Hadamard J 1923 *Lectures on the Cauchy Problem in Linear Partial Differential Equations* (New York: Yale University Press)
- [10] Courant R and Hilbert D 1962 *Methods of Mathematical Physics* vol. 2 (New York: Interscience)
- [11] Toraldo di Francia G 1969 Degrees of freedom of an image *J. Opt. Soc. Am.* **59** 799–804
- [12] Tikhonov A N and Arsenin V Y 1977 *Solutions of Ill-Posed Problems* (Washington: Winston/Wiley)
- [13] Bertero M, De Mol C and Viano G A 1980 The stability of inverse problems *Inverse Scattering Problems in Optics* ed. H P Baltes *Topics in Current Physics* vol. 20 (Berlin: Springer)
- [14] Tikhonov A N 1963 Solution of incorrectly formulated problems and the regularization method *Sov. Math.—Dokl.* **4** 1035–8
- [15] Bertero M and Pike E R 1982 Resolution in diffraction limited imaging, a singular value analysis I. The case of coherent illumination *Opt. Acta* **29** 727–46
- [16] Bertero M, Boccacci P and Pike E R 1982 On the recovery and resolution of exponential relaxation rates from experimental data: a singular value analysis of the Laplace transform inversion in the presence of noise *Proc. R. Soc. A* **383** 15–29
- [17] Louis A K and Natterer F 1983 Mathematical problems of computerized tomography *Proc. IEEE* **71** 379–89
- [18] Franklin J N 1974 On Tikhonov's method for ill-posed problems *Math. Comp.* **28** 889–907
- [19] Natterer F 1983 Discretizing ill-posed problems *Problemi non ben posti ed inversi, N.9, Pubblicazioni dell'Istituto di Analisi Globale e Applicazioni, Florence*
- [20] Slepian D 1976 On bandwidth *Proc. IEEE* **64** 292–300
- [21] Luttrell S P 1985 The use of transinformation in the design of data sampling schemes for inverse problems *Inverse Problems* **1** 199–218
- [22] McWhirter J G and Pike E R 1978 On the numerical inversion of the Laplace transform and similar Fredholm integral equations of the first kind *J. Phys. A: Math. Gen.* **11** 1729–45
- [23] Lanzaos C 1961 *Linear Differential Operators* (London: Van Nostrand)

- [24] Twomey S 1974 Information content in remote sensing *Appl. Opt.* **13** 942–5
- [25] Wahba G 1979 Ill-posed problems: numerical and statistical methods for mildly, moderately and severely ill-posed problems with noisy data *Ill-posed Problems: Theory and Practice* ed. M Z Nashed (Dordrecht: Reidel) to be published
- [26] Nashed M Z and Wahba G 1974 Generalized inverses in reproducing kernel spaces: an approach to regularization of linear operator equations *SIAM J. Math. Anal.* **5** 974–87
- [27] Slepian D 1978 Prolate spheroidal wave functions, Fourier analysis and uncertainty—V: the discrete case *Bell Syst. Tech. J.* **57** 1371–430
- [28] Greville T N E (ed.) 1969 *Theory and Applications of Spline Functions* (New York: Academic)
- [29] Torre V and Poggio T 1985 On edge detection *Preprint A. I. Memo 768 MIT, 1984 IEEE Trans. PAMI* in press
- [30] Cullum J 1971 Numerical differentiation and regularization *SIAM J. Numer. Anal.* **8** 254–65
- [31] Akhiezer N I 1965 *The Classical Moment Problem and Some Related Questions in Analysis* (Edinburgh: Oliver & Boyd)
- [32] Widder D V 1946 *The Laplace Transform* (Princeton: University Press)
- [33] Papoulis A 1956 A new method of inversion of the Laplace transform *Q. Appl. Math.* **14** 405–14
- [34] Teague M R 1980 Image analysis via the general theory of moments *J. Opt. Soc. Am.* **70** 920–30
- [35] Gregory R T and Karney D L 1969 *A Collection of Matrices for Testing Computational Algorithms* (New York: Wiley-Interscience)
- [36] Grünbaum F A 1982 A remark on Hilbert's matrix *Linear Algebra Appl.* **43** 119–24
- [37] Bertero M, Brianzi P and Pike E R 1985 The effect of sampling and truncation of data on the Laplace transform inversion *Proc. R. Soc. A* **398** 23–44
- [38] Bertero M, Brianzi P and Pike E R 1985 On the recovery and resolution of exponential relaxation rates from experimental data: Laplace transform inversions in weighted spaces *Inverse Problems* **1** 1–15
- [39] Byrne C L and Fitzgerald R M 1984 Spectral estimators that extend the maximum entropy and maximum likelihood methods *SIAM J. Appl. Math.* **44** 425–42
- [40] John F 1955 *Plane Waves and Spherical Means applied to Partial Differential Equations* (London: Interscience)
- [41] Gelfand I M and Shilov G E 1964 *Generalized Functions* vol. 1 (New York: Academic)
- [42] Titchmarsh E C 1948 *Introduction to the Theory of Fourier Integrals* (Oxford: Clarendon)
- [43] Picard E 1910 Sur un théorème général relatif aux équations intégrales de première espèce et sur quelques problèmes de physique mathématique *R. C. Mat. Palermo* **29** 615–9
- [44] Groetsch C W 1984 *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind, Research Notes in Mathematics* vol. 105 (Boston: Pitman)
- [45] Riesz F and Nagy B Sz 1955 *Functional Analysis* (New York: Ungar)
- [46] Bertero M, Brianzi P, Parker P and Pike E R 1984 Resolution in diffraction-limited imaging, a singular value analysis III. The effect of sampling and truncation of the data *Opt. Acta* **31** 181–201