Generalised information theory for inverse problems in signal processing*

E.R. Pike, J.G. McWhirter, M. Bertero and C. de Mol

Indexing term: *Signal processing*

Abstract: A powerful technique for the solution of a number of experimental inverse problems, described by an underlying first-kind Fredholm equation is presented. Such problems include, for example, diffraction-limited imaging and the analysis of laser light scattering data. The technique requires the construction of the 'singular system' of the problem which then provides exact orthonormal bases both for the description of sampled and truncated measured data and for the reconstructed continuous object 'solution' of the inversion. The singular-system approach may be regarded as a theory of information which generalises in several directions the well known classical concepts of Shannon and Nyquist.

1 Introduction

Inverting integral equations of the first kind is extremely important in a wide range of signal- and data-processing situations. Such problems are characterised by a Fredholm equation of the first kind:

$$g(p) = \int_a^b dtK(p, t)f(t) \quad c \leqslant p \leqslant d \tag{1}$$

which can be written in the general form

$$g = Kf \tag{2}$$

where K is an operator which relates the function space containing f to that containing g. In the case of eqn. 1 the operator K has the form

$$(Kf)(p) = \int_a^b K(p, t)f(t)\, dt \quad c \leqslant p \leqslant d \tag{3}$$

The functions f, g may be continuous or discrete; the operator K takes the form of a matrix in the latter case. We shall also be concerned with the case where g is discrete and f continuous.

The inverse problem amounts to finding the function f assuming that the function g is known. In signal processing, for example, the function could represent a message to be transmitted, K could denote the impulse response of the communication channel and g would then describe the received signal from which it is necessary to derive as much information as possible about the original message.

Unfortunately the problem of inverting operator equations of the first kind is generally ill conditioned. This situation arises when the function g is not sensitive to certain perturbations in the function f, i.e. there exist functions δf such that

$$Kf \simeq K(f + \delta f) \tag{4}$$

Under such circumstances the solution cannot be sharply defined and may not even be unique. For example, if the kernel K is integrable then it follows that (Riemann-Lebesgue theorem)

$$\int dtK(p, t) \sin \omega t \rightarrow 0 \quad \text{as } \omega \rightarrow \infty \tag{5}$$

and so the function g is not sensitive to the addition of arbitrary high-frequency components to the solution f. In other words, K acts as a smoothing operator which filters out high-frequency components and once these have been lost they cannot be recovered in the solution. Any attempt to do so numerically would lead to an entirely arbitrary result. It follows from the discussion above that, in general, equations of the form of eqn. 1 cannot be inverted exactly. To extract the maximum amount of reliable information about the solution without introducing arbitrary components it is necessary to use either explicitly or implicitly the method of eigenvalue decomposition or, if f and g lie in different functional spaces, singular-value decomposition.

The purpose of this paper is to describe some recent developments in these techniques by considering two particularly interesting examples. The first of these concerns the extraction of information from images which are band-limited in spatial frequency. It serves to highlight the fundamental nature of the Shannon number or Nyquist sampling limit in classical information theory and suggests how these concepts may be generalised to apply to a much wider range of problems. The second example concerns the analysis of data obtained from light scattering experiments to obtain information about the distribution of particle sizes in a polydisperse solution. The kernel of the Fredholm integral equation is quite different from the previous example; it is that of the Laplace transform and it is particularly suitable for demonstrating how the generalised information theory applies over a wide range of problems.

2 Eigenfunction analysis

2.1 Numerical filtering

Many inverse problems in electromagnetic imaging can be reduced to the solution of a Fredholm integral equation of the first kind like eqn. 1. When the kernel satisfies suitable conditions (for instance it is an integrable function) then the solution is a classical example of an ill-posed problem.

The mathematical analysis is greatly simplified if we assume that both the 'object' f and the 'image' g are elements of L^2 spaces, respectively denoted by X and Y.

The solution of eqn. 1 can be easily investigated when one can introduce eigenvalues and eigenfunctions of the integral operator (eqn. 3). If the supports of the 'object' f and of the 'image' g coincide (namely, $a = c$ and $b = d$) and if the kernel $K(p, t)$ is square integrable and satisfies the condition: $K(p, t) = K^*(t, p)$ (here the star denotes complex conjugation), then K is a compact, self-adjoint operator in $L^2(a, b)$ and the set $\{\phi_k\}_{k=0}^{+\infty}$ of the eigenfunctions associated with the eigenvalues λ_k is an orthogonal basis of the orthogonal complement of the null space

* This is an updated version of a paper presented at the NATO Advanced Research Workshop in Bad Winsheim, September 1983

Paper 3358F (E4, E10), first received 23rd May 1984 and in revised form 26th July 1984

Dr. Pike and Dr. McWhirter are with the Royal Signals & Radar Establishment, St. Andrews Road, Gt. Malvern, Worcs WR14 3PS, England, Prof. Bertero is with the Instituto di Scienze Fisiche, Universita di Genova, viale Benedetto, XV5 CAP16132 Genova, Italy and Dr. de Mol is with the University of Bruxelles, Campus Plaine, CP217 Boulevard de Trioriphe, B1050 Bruxelles, Belgium

of K and therefore an orthogonal basis of $L^2(a, b)$ when K is injective (we will consider only this case). By expanding both f and g as series of the ϕ_k we obtain the solution of eqn. 1 in the absence of noise:

$$f(t) = \sum_{k=0}^{+\infty} \frac{g_k}{\lambda_k} \phi_k(t) \tag{6}$$

where

$$g_k = \int_a^b g(p)\phi_k^*(p)\, dp \tag{7}$$

Since $\lambda_k \to 0$ when $k \to +\infty$ (this is a general property of the eigenvalues of compact operators) the series in eqn. 6 generally is not convergent when the 'image' g is affected by noise or experimental errors. Then the most simple technique for obtaining stable approximate solutions is to truncate this series at a suitable value of k. This technique, often called numerical filtering [1], can be justified both from the point of view of regularisation methods and from the point of view of Wiener filter methods [2]. If we assume, as usual, that the eigenvalues are ordered in a decreasing sequence, $\lambda_0 \geqslant \lambda_1 \geqslant \lambda_2 \geqslant \ldots$, then we take, in eqn. 6, only the eigenvalues which are greater than a prescribed constant C and if $M = \max\{k \mid \lambda_k \geqslant C\}$, then we get the filtered (regularised) solution:

$$\tilde{f}(t) = \sum_{k=0}^{M} \frac{g_k}{\lambda_k} \phi_k(t) \tag{8}$$

The quantity $1/C$ can often be interpreted as a 'signal/noise ratio' and the number M as the 'number of degrees of freedom' of the 'image' g.

2.2 *Band-limited inversion problems: Shannon theory of information*

One of the most well studied inverse problems is that of the temporal frequency band-limited communication channel or, equivalently, the spatial frequency band-limited imaging system. The mathematical description of this problem is easily stated as follows, where we use the latter system in one dimension for illustration and notation. An 'object', $O(t)$, $-1 \leqslant t \leqslant +1$, is resolved into its Fourier components $\tilde{O}(k)$ (by straightforward generalisations t and k may be multidimensional)

$$\tilde{O}(k) = \int_{-1}^{+1} O(t) e^{-ikt}\, dt \tag{9}$$

and a 'geometric' image, $I(p)$, is formed in the interval $-1 \leqslant p \leqslant +1$ by resumming only those spatial frequencies lying below a 'band-limit' Ω

$$I(p) = \frac{1}{2\pi} \int_{-\Omega}^{+\Omega} \tilde{O}(k) e^{ikp}\, dk \tag{10}$$

Thus

$$I(p) = \frac{1}{2\pi} \int_{-1}^{+1} \left(\int_{-\Omega}^{+\Omega} e^{-ik(t-p)}\, dk \right) O(t)\, dt$$

$$= \int_{-1}^{+1} \frac{\sin[\Omega(p-t)]}{\pi(p-t)} O(t)\, dt \tag{11}$$

This has the form of a general Fredholm equation of the first kind (eqn. 1) with the 'kernel' $K(p, t)$ of the integral operator K in this case a sinc function.

The remarkable solution of eqn. 11 in terms of its orthonormal eigenfunctions, namely the prolate spheroidal wave functions [3] helps to understand the classical theory of information and is the source of the accepted theory of resolution in diffraction-limited imaging problems. We recall that the eigenfunctions ϕ_n are the solutions of the homogeneous equation

$$\lambda_n \phi_n(p) = \int_{-1}^{+1} \frac{\sin[\Omega(p-t)]}{\pi(p-t)} \phi_n(t)\, dt \tag{12}$$

$$-1 \leqslant p \leqslant +1$$

so that the object may be reconstructed as a series

$$O(t) = \sum_{n=0}^{\infty} \frac{c_n}{\lambda_n} \phi_n(t) \tag{13}$$

where

$$c_n = \int_{-1}^{+1} I(p)\phi_n(p)\, dp \tag{14}$$

For the prolate spheroidal functions the eigenvalue series λ_n drops sharply from approximately unity to nearly zero at a value of n known as the Shannon number $(2\Omega/\pi)$ which means that only this number of components may be recovered almost independently of the actual noise level in any particular situation. One may thus, with this problem, speak of a number of degrees of freedom, number of 'bits' of information, Nyquist sampling limit and other concepts related only to the band limit and not to the noise. These concepts are used in what we may call the classical Shannon theory of information.

A nice property of the eigenfunctions of many integral operators encountered in experimental inverse problems is that the eigenfunction ϕ_k has exactly k zeros inside the interval $[a, b]$. Then if the distribution of the zeros is approximately uniform one can define a resolution distance, associated with eqn. 8, as follows:

$$d = (b-a)/M \tag{15}$$

This is the case for the prolate spheroidal functions and this resolution distance gives the Nyquist sampling rate in the limit of infinite object support. If the zeros are not uniformly distributed then they provide an indication of local resolution proportional to their density. 'Point-spread' functions of the form

$$S(x) = \sum_{n=0}^{M} (\phi_n \otimes \phi_n)(x) \tag{16}$$

achieved in reconstruction with a finite number N of terms of eqn. 13 are discussed in Reference 4.

2.3 *Laplace transform inversion*

First-kind Fredholm equations appear in many more physical inverse problems, of course, than that with the sinc kernel discussed above. We shall consider in this Section the case where the kernel is an exponential function, giving rise to the problem of the inversion of the Laplace transform

$$g(p) = \int_0^{\infty} e^{-pt} f(t)\, dt \quad 0 \leqslant p < +\infty \tag{17}$$

This problem arises in diverse fields of science, our particular interest being its application to the determination of molecular size distributions from laser light scattering data [5]. Many *ad hoc* methods of numerical inversion of eqn. 17 have been proposed but it was not until five years ago that the problem was approached in the spirit of the information theory sketched above when McWhirter and Pike

[6] proposed its solution in terms of generalised eigenfunctions, ϕ_ω (where ω is now a continuous variable), given by

$$\phi_\omega(t) = t^{-1/2-i\omega} \frac{\mathrm{Re}\,[\Gamma(\frac{1}{2}+i\omega)]^{1/2}}{|\Gamma(\frac{1}{2}+i\omega)|^{1/2}} \quad \omega \geqslant 0 \tag{18}$$

$$\phi_\omega(t) = t^{-1/2+i\omega} \frac{\mathrm{Im}\,[\Gamma(\frac{1}{2}+i\omega)]^{1/2}}{|\Gamma(\frac{1}{2}+i\omega)|^{1/2}} \quad \omega < 0 \tag{19}$$

Furthermore the corresponding continuous spectrum of the operator was shown to have the simple form

$$\lambda_\omega = \frac{\omega}{|\omega|} \sqrt{\frac{\pi}{\cosh(\pi\omega)}} \tag{20}$$

This spectrum shows an unfortunate exponential fall-off causing a severe ill-posedness and a strong limitation, in the presence of noise, on the band of frequencies which may be recovered. It must be noted that in this case, contrary to the sinc-kernel case, the resolution limit depends on the noise level. The reason is that the spectrum in eqn. 20 no longer has a plateau before a sharp fall as observed in the previous case.

In this solution the zeros of the eigenfunctions are distributed in a geometrical fashion along the axis and this fact defines a resolution ratio $\delta = \exp(\pi/\omega_{\max})$ where $\omega_{\max}$ is the maximum recoverable frequency. By sampling the solution at the exponentially varying rate appropriate to the problem, the reconstruction procedure is reduced to the inversion of a well conditioned matrix. This is known in the light scattering experiment above as the exponential-sampling method and is discussed in the next Section.

3 Singular function analysis

3.1 Finite Laplace transform

In order to use the McWhirter-Pike results for numerical inversion Ostrowsky *et al.* [5] found that, provided a sufficiently low number of sample points of the solution were demanded, a discrete matrix inversion based on

$$g(p_i) = \sum_{n=1}^{S} a_n e^{-p_i t_n} \quad i = 1, \ldots, N \tag{21}$$

could be used. As mentioned above a geometric ratio δ the 'resolution ratio', was used for the spacing of the points p_i. The somewhat surprising fact emerged that the number S of sample points which could be recovered was somewhat larger than the number of degrees of freedom predicted by the McWhirter-Pike theory.

Following the work of Ostrowsky *et al.* [5], Bertero, Boccacci and Pike [7] traced the increase of information capacity found by these authors to the implicit restriction of object support in the matrix inversion procedure used. The effect of a known finite support of the object in the Laplace transform was given a thorough investigation using an approach based on singular functions (see for example Reference 8.)

Bertero *et al.* [7] considered the problem of the finite Laplace transform

$$(Kf)(p) = \int_a^b e^{-pt} f(t)\, dt \quad 0 \leqslant p < +\infty \tag{22}$$

restricting the transformation to square integrable object functions on the interval (a, b). It is easy to verify that the kernel $K(p, t) = \exp(-pt)$ is square integrable and therefore K is compact.

The previous eigenfunction analysis does not apply to the case where the supports of the 'object' f and of the 'image' g do not coincide. But if the functions f and g have support $[a, b]$ and $[c, d]$, respectively, and the operator K satisfies the condition

$$\int_c^d dp \int_a^b dt\, |K(p, t)|^2 < +\infty \tag{23}$$

(more generally, if K is compact), a powerful generalisation of the eigenfunction method is provided by the singular function method [9]. The singular system $\{\alpha_k; u_k, v_k\}_{k=0}^{+\infty}$ of the operator K is the set of the solutions of the coupled equations

$$Ku_k = \alpha_k v_k \qquad K^* v_k = \alpha_k u_k \tag{24}$$

where K^* is the adjoint of K, eqn. 3

$$(K^* g)(t) = \int_c^d K^*(p, t) g(p)\, dp \quad a \leqslant t \leqslant b \tag{25}$$

The numbers α_k are called the singular values of K (usually they are ordered in a decreasing sequence) and the functions u_k, v_k are called the singular functions of K associated with α_k. The singular functions u_k are also eigenfunctions of the operator K^*K associated with the eigenvalues α_k^2 and they form an orthogonal basis of the orthogonal complement of the null space of K (of $L^2(a, b)$ if K is injective). Analogously the singular functions v_k are eigenfunctions of KK^*, still associated with the eigenvalues α_k^2, and they form an orthogonal basis of the orthogonal complement of the null space of K^* (of $L^2(c, d)$ if K^* is injective).

In the absence of noise the solution of eqn. 2 with eqn. 22 is now given by

$$f(t) = \sum_{k=0}^{+\infty} \frac{g_k}{\alpha_k} u_k(t) \tag{26}$$

where

$$g_k = \int_0^{+\infty} g(p) v_k^*(p)\, dp \tag{27}$$

This solution has a structure very similar to that of the solution eqn. 8, which applies to the case of a self-adjoint operator, except for the fact that now two orthogonal systems, instead of one, are used. Again one can introduce a filtered solution as follows:

$$\tilde{f}(t) = \sum_{k=0}^{M} \frac{g_k}{\alpha_k} u_k(t) \tag{28}$$

and therefore it is clear that the concepts of 'number of degrees of freedom' and of resolution distance (or resolution ratio) can be easily extended to the present case.

As a consequence of the scaling properties of the Laplace transform the singular values of the operator K depend on the upper bound b and on the lower bound a of the support of f through the ratio

$$\gamma = b/a \tag{29}$$

Therefore we find it convenient to take as a support of f the interval $[1, \gamma]$ and to define the operator K as follows:

$$(Kf)(p) = \int_1^{\gamma} e^{-pt} f(t)\, dt \tag{30}$$

The singular values α_k and the singular functions $u_k(t)$ are given by the solutions of the eigenvalue problem

$$\int_1^{\gamma} T(t+s) u_k(s)\, ds = \alpha_k^2 u_k(t) \tag{31}$$

where

$$T(t) = 1/t \tag{32}$$

Then the singular functions $v_k(p)$ are given by the finite Laplace transforms of the singular functions $u_k(t)$:

$$v_k(p) = \frac{1}{\alpha_k} \int_1^\gamma e^{-pt} u_k(t)\, dt \tag{33}$$

All the eigenvalues α_k^2 have multiplicity one [7] and therefore they can be ordered in a strictly decreasing sequence. They tend to zero exponentially fast as follows from general results on the asymptotic behaviour of the eigenvalues of integral operators [10]. As a consequence, for moderate values of γ, very few singular values are significant. For instance, in the case $\gamma = 5$, we have [7]:

$$\begin{aligned} &\alpha_0 = 8.751 \times 10^{-1}, \quad &&\alpha_1 = 1.935 \times 10^{-1}, \\ &\alpha_2 = 3.827 \times 10^{-2}, \quad &&\alpha_3 = 7.434 \times 10^{-3}, \\ &\alpha_4 = 1.435 \times 10^{-3}, \quad &&\alpha_5 = 2.765 \times 10^{-4}, \ldots \end{aligned} \tag{34}$$

Accurate computations of the singular values α_k and of the singular functions u_k can be performed along the following lines [7]:

(i) approximation of the kernel $T(t + s)$ by means of a tensor product of splines [11]; the approximating kernel is of finite rank and its eigenvalues coincide with the eigenvalues of a nonsymmetric matrix Q

(ii) the largest eigenvalues of Q are computed by a combination of the power method and of the deflation method [12]; this procedure works very efficiently when the eigenvalues are well separated as in our case

(iii) once the eigenvalues have been computed, the corresponding eigenvectors of Q are determined by means of the inverse power method [13]

(iv) finally the eigenfunctions $u_k(t)$ are computed as linear combinations of splines, the coefficients being provided by the components of the eigenvectors of Q.

As follows from these computations, the singular functions $u_k(t)$ have the property mentioned at the end of Section 2.2 namely $u_k(t)$ has exactly k zeros in the interval $[1, \gamma]$. Furthermore, the distribution of the zeros is approximately geometric and, therefore, one can define a resolution ratio δ as

$$\delta^M = \gamma \tag{35}$$

where M is the maximum index for eqn. 28.

In Reference 7 it was shown that the singular value spectrum of operator eqn. 30 has a slower decrease rate than the corresponding spectrum for the problem where $f(t)$ is assumed to have an infinite support. As a consequence, the knowledge that $f(t)$ is *a priori* restricted to a finite known support allows the recovery of more information from the data and this is found quantitatively by Ostrowsky *et al.* [5].

3.2 Diffraction-limited imaging

The work of Bertero *et al.* [7] described above gave rise to the suggestion that the increase in information transfer obtained in the Laplace inversion by using a known restricted support for the solution might have an analogy in the diffraction-limited imaging problem with coherent illumination discussed in Section 1, allowing an improvement in resolution over that defined by the Shannon number to be obtained. This possibility was investigated and proved to be correct in Reference 14. In this paper it was shown that the restriction of the image to a support equal to that of the geometrical image, which is implicit in the eigenvalue analysis, causes a loss of information which can be regained by the use of the total diffraction image. It is then required to solve a first kind Fredholm equation where the integral operator K is now defined as follows

$$(Kf)(p) = \int_{-1}^{+1} \frac{\sin[\Omega(p - t)]}{\pi(p - t)} f(t)\, dt \qquad -\infty < p < +\infty \tag{36}$$

The singular values of this operator can be shown to be $\sqrt{\lambda_n}$ and fall off more slowly than the eigenvalues λ_n defined by eqn. 12. The image singular functions are orthogonal to out of band noise allowing more coefficients to be calculated in the reconstruction and hence a higher resolution to be obtained. The two-dimensional case with square and circular apertures was also considered in Reference 7 and in a later paper [15] the analysis was also carried out for incoherent illumination. Furthermore it has been remarked in these problems that resolution beyond the classical diffraction limit (i.e. super-resolution) is much easier to achieve when the space-bandwidth product is small (i.e. the constant Ω in eqn. 36). A similar effect has also been demonstrated for the problem of inverse diffraction from plane to plane, when it is known *a priori* that the source is restricted to a small support with linear dimensions comparable to the wavelength. An example is given in Fig. 1 (private communication*) where information beyond the geometrical image allowed a significant super-resolution to be obtained.

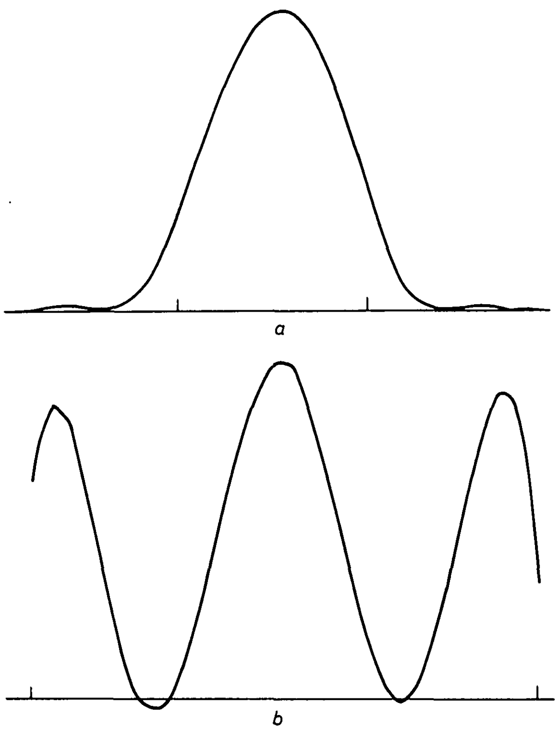

Fig. 1 *Reconstruction of three-peaked object using the first seven singular functions*

The Shannon number was 2.17 and the scale markers represent the boundaries of the geometrical image

a experimental image *b* object reconstruction

4 Case of discrete data

4.1 Finite sampled Laplace transform

In the preceding two Sections we discussed the improvement in resolution in the Laplace transform and

* Abbiss, J.B., Greenaway, A.H., Pike, E.R., Walker, J., and de Villiers, G.: (in preparation)

diffraction-limited imaging cases obtained when using *a priori* knowledge of the support of the object function. The singular systems used were constructed in bases of functions of continuous variables. In real experiments, however, the data are measured over truncated regions of finite extent and at finite numbers of sampled points. A possible route is then to discretise the unknown function.

Standard singular value decomposition techniques can then be used as has been done, for example, in the case of the Laplace transform inversion by Bertero, Boccacci and Pike [15] and in the case of band-limited imaging by Zhou, Rushforth and Frost [17].

In the discretisation methods the object is constructed only on a set of sample points and, usually, an interpolation scheme is used between them.

However, the singular value method offers a very interesting alternative since the types of functions, u describing the object and v describing the data, may be different. In particular, we may regard the data as a vector but yet reconstruct the object in a basis of square integrable functions so that the interpolation step and other distortions due to sampling of the object are completely avoided.

Although in our earlier work we have discussed discrete object models for Laplace inversion we have bypassed this stage in our work on diffraction-limited imaging in favour of reconstructions in a basis of continuous functions.

Let us assume that the 'image' $g(p)$ is measured only on a finite set of points, $p_1, p_2, \ldots, p_N$, so that eqn. 29 is replaced by

$$g(p_n) = \int_a^b K(p_n, t)f(t)\,dt : n = 1, \ldots, N \tag{37}$$

Now the data space Y is an N-dimensional vector space equipped with the scalar product

$$(g, h)_Y = \sum_{n=1}^{N} w_n g(p_n)h^*(p_n) \tag{38}$$

where the weights w_n must be introduced in order to take into account a nonuniform distribution of the data points p_n or, alternatively, their statistical properties.

Let K_N be the operator, defined by eqn. 37, which transforms an 'object' of $L^2(a, b)$ into a vector of Y. The adjoint K_N^* transforms a vector of Y into a function of $L^2(a, b)$ and it is given by

$$(K_N^* g)(t) = \sum_{n=1}^{N} w_n g(p_n)K^*(p_n, t) \tag{39}$$

Then $K_N^* K_N$ is a self-adjoint integral operator of finite rank in $L^2(a, b)$, whose kernel is given by

$$T_N(t, s) = \sum_{n=1}^{N} w_n K^*(p_n, t)K(p_n, s) \tag{40}$$

while the operator $K_N K_N^*$ is an $N \times N$ matrix whose matrix elements are given by

$$t_{nm}^{(N)} = w_m \int_a^b K(p_n, t)K^*(p_m, t)\,dt \tag{41}$$

We denote by $\{\alpha_{N,k}; u_{N,k}, v_{N,k}\}_{k=0}^{N-1}$ the singular system of K_N, the $u_{N,k}$ being the eigenfunctions of $K_N^* K_N$ and the $v_{N,k}$ being the eigenvectors of the matrix $K_N K_N^*$.

The singular value $\alpha_{N,k}$ provides a good approximation of the corresponding singular value α_k of the problem with continuous data, if the integral operator $K_N^* K_N$ is a good approximation of the integral operator K^*K. Indeed, let $T(t, s)$ be the kernel of K^*K:

$$T(t, s) = \int_c^d K^*(p, t)K(p, s)\,dp \tag{42}$$

and let us write

$$\varepsilon_N(t, s) = T(t, s) - T_N(t, s) \tag{43}$$

$T_N(t, s)$ being defined in eqn. 40. Clearly $K_N^* K_N$ is a good approximation of K^*K if $\varepsilon_N(t, s)$ is small and by the Weyl-Courant lemma [18] we have

$$|\alpha_{N,k}^2 - \alpha_k^2| \leqslant \left(\int_a^b dt \int_a^b ds |\varepsilon_N(t, s)|^2 \right)^{1/2} \tag{44}$$

Therefore if M is the number of degrees of freedom in the problem with continuous data, we can look for a distribution of the data points which allows a good approximation of the singular values α_k up to $k = M$. If M is small, it is reasonable to argue that, using a small number of data points, conveniently placed, it is possible to obtain such a good approximation. Then the truncated solution (eqn. 28) can be approximated by

$$\tilde{f}_N(t) = \sum_{k=0}^{M} \frac{g_{N,k}}{\alpha_{N,k}} u_{N,k}(t) \tag{45}$$

where

$$g_{N,k} = \sum_{n=1}^{N} w_n g(p_n)v_{N,k}^*(p_n) \tag{46}$$

In the last equation $v_{N,k}(p_n)$ denotes the nth component of the singular vector $v_{N,k}$.

The above analysis can be immediately applied to the problem of the finite Laplace transform inversion and, due to the simplicity of the kernel, some computations can be done analytically. For instance the matrix $t_{nm}^{(N)}$ of eqn. 41 takes the very simple form

$$t_{nm}^{(N)} = \frac{w_m}{p_n + p_m} \{\exp[-(p_n + p_m)] - \exp[-\gamma(p_n + p_m)]\} \tag{47}$$

In the case of equidistant data points, $p_n = (n - 1)\sigma$, $(n = 1, \ldots, N)$, we can take $w_n = \sigma$ so that the matrix 47 becomes

$$t_{nm}^{(N)} = \frac{1}{n + m - 2} \{\exp[-\sigma(n + m - 2)] - \exp[-\sigma\gamma(n + m - 2)]\} \tag{48}$$

(note that $t_{11}^{(N)} = \sigma(\gamma - 1)$). The greatest eigenvalues and the corresponding eigenvectors of this matrix can be easily computed by means of a combination of the power method and the Hotelling deflation method [12], since the matrix is symmetric. When we have the eigenvalues $\alpha_{N,k}$ and the corresponding singular vectors $v_{N,k}$, the singular functions $u_{N,k}(t)$ can be obtained by means of the relation

$$u_{N,k}(t) = \frac{\sigma}{\alpha_{N,k}} \sum_{n=1}^{N} v_{N,k}(p_n)e^{-p_n t} \tag{49}$$

where $v_{N,k}(p_n)$ denotes the nth component of $v_{N,k}$.

The singular values $\alpha_{N,k}$ can provide a good approximation of the singular values α_k of the problem with continuous data. In the case of equidistant data points the kernel of eqn. 43 is given by $\varepsilon_N(t, s) = \varepsilon_N(t + s)$ where

$$\varepsilon_N(t) = \frac{1}{t} - \sigma \frac{1 - \exp(-N\sigma t)}{1 - \exp(-\sigma t)} \tag{50}$$

and this function tends to zero uniformly over the interval $2 \leq t \leq 2\gamma$ when $\sigma \to 0$ and $N \to +\infty$. Therefore from eqn. 44 it is expected that $\alpha_{N,k} \sim \alpha_k$ if one takes a sufficiently large number of data points at a sufficiently small distance σ. The following singular values

$$\alpha_{N,0} = 9.075 \times 10^{-1}, \qquad \alpha_{N,1} = 1.983 \times 10^{-1},$$
$$\alpha_{N,2} = 3.555 \times 10^{-2} \tag{51}$$

have been obtained in the case $\gamma = 5$ using $N = 64$ and $\sigma = 0.03$ [19]. If one compares with the singular values of the continuous case, eqn. 29, one finds an average error of about 4%. This is the best approximation which can be obtained using 64 linearly spaced data points.

In the case of Laplace inversion the use of equidistant points is not the best one. Such a choice is only convenient for deconvolution problems, i.e. for the solution of problems which have simple transformation properties with respect to the translation group. In the case of Laplace transform inversion the underlying group is the dilation group and for this reason a geometric distribution of data points suitably placed seems to be more natural, at least from the theoretical point of view.

Therefore we consider now the following data-point distribution:

$$p_n = p_1 \Delta^{n-1} \qquad n = 1, \ldots, N \tag{52}$$

where p_1 and Δ can be considered as free parameters. Concerning the choice of the weights w_n, the following result is proved in Reference 20: if the p_n are as in eqn. 52 and if

$$w_n = (\ln \Delta) p_n \tag{53}$$

then, given ε, it is possible to choose Δ and N so that, for any $t \in [2, 2\gamma]$

$$|\varepsilon_N(t)| = \left| \frac{1}{t} - \sum_{n=1}^{N} w_n e^{-p_n t} \right| \leq \varepsilon \tag{54}$$

This result, combined with eqns. 40 and 41 suggests the introduction of the weights (eqn. 53) into the scalar product (eqn. 38).

With reference to the computation of the eigenvalues and eigenvectors of the matrix (4.11) we remark that it is not symmetric in the usual sense ($t_{nm}^{(N)} \neq t_{mn}^{(N)}$) since it is self-adjoint with respect to the weighted scalar product (4.2). Then, for computational convenience, we transform the matrix (4.11) into a symmetric one having the same eigenvalues

$$t_{nm}^{(N)} = \frac{\sqrt{w_n w_m}}{p_n + p_m} \{\exp[-(p_n + p_m)] - \exp[-\gamma(p_n + p_m)]\} \tag{55}$$

We also remark that if $\bar{v}_{N,k}$ is an eigenvector of this matrix, the corresponding eigenvector of the matrix in eqn. 47 is given by

$$v_{N,k}(p_n) = w_n^{-1/2} \bar{v}_{N,k}(p_n) \tag{56}$$

Therefore singular values and singular vectors can be computed again by a combination of the power method and of the Hotelling deflation method.

The fundamental property of geometric sampling of data is that, using a very small number of data points, it is possible to obtain extraordinarily good approximations of the singular values of the problem with continuous data (and also of the singular functions). For instance, in the case $\gamma = 5$, using 5 data points, $\Delta = 5.5$ and $p_1 = 1.366 \times 10^{-3}$ we have always obtained the following singular values [19].

$$\alpha_{N,0} = 8.742 \times 10^{-1}, \qquad \alpha_{N,1} = 1.913 \times 10^{-1},$$
$$\alpha_{N,2} = 3.827 \times 10^{-2} \tag{57}$$

If we compare with the singular values of the problem with continuous data, we find an average error of about 0.1%. Also the singular functions $u_{N,k}(t)$ ($k = 0, 1, 2$) provide an excellent approximation of the singular functions $u_k(t)$ ($k = 0, 1, 2$) corresponding to the problem with continuous data.

Now, if we notice that, when the 'signal/noise ratio' is equal to 100, the 'number of degrees of freedom' in the case $\gamma = 5$ is just 3, we can conclude that using 5 data points optimally placed, the filtered solution (eqn. 45) practically coincides within the error on the data with the filtered solution (eqn. 28). Therefore 5 data points allow almost the same resolution as continuous data.

We believe that the previous result can be extended to any ill-posed problem: a number of data points, optimally placed, approximately equal to the 'number of degrees of freedom' is equivalent, as explained above, to continuous data in the common case where the covariance matrix of the data does not favour averaging of closely spaced samples.

4.2 *Finite sampled diffraction-limited imaging*

In the diffraction-limited imaging problem, just as in the Laplace problem discussed in the last subsection, the same idea of using a singular vector basis for the data and a singular function basis for the object can be applied. Successful numerical computation of such bases have been performed in Genova and at Malvern [21] and are currently in the process of being applied to practical problems in microscopic and other imaging. The theory looks particularly promising for application, for example, in electromagnetic imaging problems in the microwave region.

An attractive feature of the method is the speed at which the reconstruction can be affected, all that is required is to form a scalar product of the data set with each of the basis data vectors in turn and these scalar products are used as the coefficients, divided by the corresponding singular value, of the singular function expansion of the object. Both the basis vector set and the basis function set are, of course, stored numerically once and for all for a given problem.

5 Extension to other problems

So far in this paper we have only mentioned two cases of the kernel $K(p, t)$ of the integral equation, namely, the sinc kernel and the Laplace exponential kernel. There are, however, many other problems of a similar type and in this last Section we will mention briefly some problems which are under study using the singular system technique.

5.1 *Fraunhofer diffraction*

In this problem the kernel has the form $K(p, t) = J_1^2(pt)/(pt)^2$. This represents the form of the intensity scattered in the far field by a spherical particle illuminated by a coherent laser beam and the solution of the problem will be the distribution of sizes of the scattering particles. The problem is of great importance in many industrial and scientific applications and the technique of particle sizing by Fraunhofer diffraction is quite widely used.

As a first attempt to quantify the information content in a typical data set in such an experiment we have used the results of Reference 6 and found the following spectrum for the above kernel [21]:

$$\lambda_\omega^\pm = \pm \frac{2^{-5/2}\Gamma\left(\frac{5}{2} - i\omega\right)\Gamma\left(\frac{1}{4} + \frac{i\omega}{2}\right)}{\Gamma^2\left(\frac{7}{4} - \frac{i\omega}{2}\right)\Gamma\left(\frac{11}{4} - \frac{i\omega}{2}\right)} \tag{58}$$

This spectrum can be used to calculate the number of resolution cells in particle size, which can be found in a certain level of experimental noise. For example we can ask how many size fractions can be resolved in the range 2 to 100 microns when the signal/noise ratio is 10^2. The result is that 7 or 8 fractions on a geometrical scale may be resolved in this range. Further work is nearing completion [22] on this problem extending the calculations to the use of singular functions and vectors, which will provide a very practical approach to inversion for particle size distributions in this range.

5.2 Laser anemometry

Yet another example of this sort of inversion problem is provided by laser anemometry. In this type of experiment a laser beam scattered off a particle in uniform motion at velocity v, gives rise to a measured data set with the form of an exponentially damped cosine wave. In turbulent flows or nonuniform flows the data represent the motion of particles with a probability distribution of velocities which one needs to recover as the solution. The kernel in this case has the form

$$K(v, t) = e^{-(vt)^2/2\sigma^2}(1 + a \cos (bvt)) \tag{59}$$

where a, b and σ are constants. A certain amount of work with this kernel has been done using the concepts of information theory discussed above and one finds, in this case, that the fall-off of the eigenvalue or singular value spectrum is a function of the value of σ, characterising the damping of the cosine signal. In the limit of small damping a large amount of information is available from the data, while if the damping is more severe there is much less velocity information to be found. Again one can introduce a concept of resolution associated with a number of components of the solution which may be transmitted above the noise level. The pressure to push the analysis of this particular kernel to the same degree as has been done for the Laplace problem above is not so strong since in many cases the damping term may be arranged to be small. However in critical circumstances, as for example are met in very high speed supersonic flows, or where very high spatial resolution is required it will be important to extract the maximum information from data of this type and further work along the lines described above will allow this to be done.

5.3 Uncertain localisation or nonuniform illumination

In this last Section we come back to the diffraction-limited imaging problem but consider the case where the object is illuminated by a coherent laser beam of infinite spatial extent but with a known amplitude profile. This is not the place to report detailed results of this calculation which can also be interpreted in terms of an *a priori* estimate of the reconstruction, but it will suffice to say that singular value spectra may be calculated numerically [4] and these allow a discussion of such an imaging problem in terms of the concepts of number of degrees of freedom as a function of noise and the implications of this in terms of the limits of resolution which may be achieved. These results allowed theoretical point-spread functions as in eqn. 16 to be calculated for various situations in scanning microscopy. Our results indicate that the singular value spectrum for this problem falls off more slowly than for the associated diffraction-limited imaging problem in which the illumination is assumed to have a 'top hat' form (sinc kernel) of the same total power and the same width. However, the zeros of the singular functions do not increase in density at the same rate as for the 'top hat' case and therefore the overall resolution limits turn out always to be worse.

The basic equation is

$$g(p) = \int_{-\infty}^{\infty} \frac{\sin \Omega(p - t)}{\pi(p - t)} P(t)f(t)\, dt \tag{60}$$

where $P(t)$ is a 'profile' function which may be interpreted in either of the two ways mentioned above. The singular function approach is now generalised to take a weighted L^2 space for the space of solutions.

This method has also recently been applied to the finite sampled Laplace inversion with interesting theoretical and experimental results [23].

The parameters of the profile functions in this case are chosen to match the cumulants of the solution which may be calculated without performing the inversion. An example is given in Fig. 2.

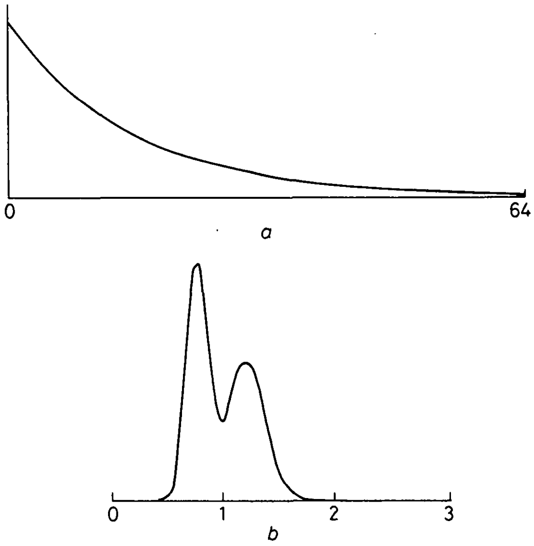

Fig. 2 *Reconstruction of two delta functions of equal weight at 0.75 and 1.25 from their Laplace transform using the first five singular functions*
The profile function was a gamma distribution with the same mean as the object
a Laplace transform *b* Reconstruction

6 Conclusions

Various applications of singular system theory have been described and shown to be a generalisation of classical eigenfunction methods in signal processing. The theory allows *a priori* information on the support or localisation of a solution of an inverse problem to be used to improve the resolution in a manner amenable to fast computation. Continuous functions may be reconstructed from finite and sampled data.

7 References

1 TWOMEY, S.: *J. Franklin Inst.*, 1965, **279**, pp. 95–109
2 BERTERO, M., DE MOL, C., and VIANO, G.A.: 'The stability of inverse problems' *in* BALTES, H.P. (Ed.): 'Inverse scattering problems in optics–Topics in current Physics—Vol. 20' (Springer, 1980), pp. 161–214
3 SLEPIAN, D., and POLLAK, H.O.: 'Prolate spheroidal wave functions, Fourier analysis and uncertainty—I', *Bell Syst. Tech. J.*, 1961, **40**, pp. 43–63

4 BERTERO, M., DE MOL, C., PIKE, E.R., and WALKER, J.: 'Resolution in diffraction-limited imaging, a singular value analysis IV', *Opt. Acta*, (in press)
5 OSTROWSKY, N., SORNETTE, D., PARKER, P., and PIKE, E.R.: 'Exponential sampling method for light scattering polydispersity analysis', *ibid.*, 1981, **28**, pp. 1059–1070
6 MCWHIRTER, J., and PIKE, E.R.: 'On the numerical inversion of the Laplace transform and similar Fredholm integral equations of the first kind', *J. Phys. A*, 1978, **II**, pp. 1729–1745
7 BERTERO, M., BOCCACCI, P., and PIKE, E.R.: '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. London Ser. A*, 1982, **A383**, pp. 15–29
8 SMITHIES, F.: 'Integral equations' (Cambridge University Press, 1958)
9 PICARD, E., MAT PALERMO, R.C.: 1910, **29**, pp. 615–619
10 HILLE, E., and TAMARKIN, J.D.: *Acta Math*, 1931, **57**, pp. 1–76
11 HAMMERLIN, G., and SCHUMAKER, L.L.: 'Procedures for kernel approximation and solution of Fredholm integral equations of the second kind', *Numer. Math.*, 1980, **34**, pp. 125–141
12 RALSTON, A.: 'A first course in numerical analysis' (McGraw Hill, 1965)
13 WILKINSON, J.H.: 'The algebraic eigenvalue problem' (Clarendon Press, Oxford, 1965)
14 BERTERO, M., and PIKE, E.R.: 'Resolution in diffraction-limited imaging, a singular value analysis. I: The case of coherent illumination', *Opt. Acta.*, 1982, **29**, pp. 727–746
15 BERTERO, M., BOCCACCI, P., and PIKE, E.R.: 'Resolution in diffraction-limited imaging, a singular value analysis II. The case of incoherent illumination', *ibid.*, 1982, **29**, pp. 1599–1611
16 BERTERO, M., BOCCACCI, P., and PIKE, E.R.: 'On the recovery and resolution of exponential relaxation rates from experimental data II', *Proc. R. Soc. London*, 1984, **A393**, pp. 51–65
17 ZHOU, Y., RUSHFORD, C.K., and FROST, R.L.: 'Singular value decomposition, singular vectors and the discrete prolate spheroidal sequences', *Proc. ICASSP*, 1984, pp. 37.6.1–37.6.4
18 RIESZ, F., and NAGY, B.: 'Functional analysis' (New York: Ungar, 1955)
19 BERTERO, M., BRIANZI, P., and PIKE, E.R.: 'On the recovery and resolution of exponential relaxation rates from experimental data III', submitted to *Proc. R. Soc.* A
20 BERTERO, M., BRIANZI, P., PARKER, P., and PIKE, E.R.: 'Resolution in diffraction-limited imaging, a singular value analysis III. The effect of staggering and truncation of the data', *Opt. Acta.*, 1984, **31**, pp. 181–201
21 BERTERO, M., and PIKE, E.R.: 'Particle size distributions from Fraunhofer diffraction I', *ibid.*, 1983, **30**, pp. 1043–1049
22 BERTERO, M., and PIKE, E.R.: 'Particle size distributions for Fraunhofer diffraction II—the singular value spectrum', *ibid.*, (to be published)
23 BERTERO, M., BRIANZI, P., PIKE, E.R., DE VILLIERS, G.D., OSTROWSKY, N., and LAN, K.: 'Light scattering polydispersity analysis of molecular diffusion by Laplace transform inversion in weighted spaces', *J. Chem. Phys. Lett.* (to be published)

IEE Colloquium report

IEE Colloquium on the Current status of systolic array research in the UK

Savoy Place, London
1st February 1984

Systolic arrays (SA) have been of interest to the signal-processing community, at large, since the appearance in 1980 of the seminal chapter by Kung and Leiserson [1]. Current interest in the UK can be gauged by the fact that over 100 people attended this colloquium.

Although researchers at Naval Ocean Systems Centre (NOSC) in San Diego had previously been experimenting with the concept of arrays of identical inner-product type processors (multiplier-accumulators, or MACs), it was Kung who did the groundwork on systolic arrays, producing papers on architectures for pattern matching, matrix multiplication, convolution etc. Among advantages claimed for VLSI SAs are: modularity, regularity, functional parallelism, minimal processor interconnection, short data propagation, high packing density, pipelining, simple control and expandability.

Significantly, perhaps, it has been the mathematics and computer science communities who have been publishing, in the main, since 1980. Common to all SA Papers are the interconnected grids of boxes, through which data and results flow in several directions. However the proposed contents of these boxes give some cause for concern to electronics engineers.

For 'order N & N squared' tasks, such as convolution, DFT, matrix-vector multiplication etc., linear or square arrays of SAs may be constructed, with MACs in the boxes. However, conventional solutions also exist for these tasks. The area where SAs show most promise is in 'order N cubed' tasks such as linear systems solution, singular value decomposition, recursive-least-squares minimisation etc. Much work has been done by the maths/computer-science community on Fortran/Algol solutions to these problems on mainframe computers (for example, see Reference 2). Unfortunately the grids of little squares in the SA literature now contain little circles as well (boundary processors). Elsewhere in the diagrams are details of the contents of the little circles—operations such as division and square-root. Hence the reticence shown by the electronic engineering community in realising these processors in hardware. However, these problems will no doubt be solved in the future, and we can look forward to the appearance of SAs which process a signal matrix to extract, for instance, its eigenvalues.

Most of the UK SA work is being conducted at RSRE, GEC and STL. John McWhirter of RSRE is the 'father-figure' of the community. He favours a bit-level approach to SAs, which makes for very small array elements (gated full adders as opposed to MACs). GEC are fabricating such a correlator in bulk CMOS for RSRE [3]. McWhirter proposes to achieve yield enhancement by row-bypassing.

GEC also have their own SA designs in fabrication. A. McCabe described a 16-point 4 × 1-bit correlator in CMOS-SOS, containing 9 × 16 bit-level cells for 9-bit output as 30 MHz. Disadvantages of this chip, which is near to production, are difficulty in cascading, only 50% hardware utilisation, no 2's complement arithmetic, no time wedging, and no fault tolerance. Another design is a 64 point 4-bit by 1-bit cascadable correlator chip, whose projected performance betters that of TRWs 'state of the