On the numerical inversion of the Laplace transform and similar Fredholm integral equations of the first kind

1978 J. Phys. A: Math. Gen. 11 1729

(http://iopscience.iop.org/0305-4470/11/9/007)

J. Phys. A: Math. Gen., Vol. 11, No. 9, 1978. Printed in Great Britain.

On the numerical inversion of the Laplace transform and similar Fredholm integral equations of the first kind

J G McWhirter and E R Pike

The Royal Signals and Radar Establishment, St Andrews Road, Great Malvern, Worcs WR14 3PS, UK

Received 27 February 1978

Abstract. Analytic expressions are derived for the eigenfunctions and eigenvalues of the Laplace transform and similar dilationally invariant integral equations of the first kind. Some generalised concepts of information theory are introduced to show how the use of these eigenfunctions enables the maximum possible amount of information to be obtained when solving the inverse problem numerically. These concepts also explain how the amount of information available depends on the level of noise in the calculation and on the structure of the particular integral kernel. Some numerical examples which illustrate these points are presented.

1. Introduction

In this paper we study the problem of obtaining the numerical solution of integral equations of the type

$$g(\tau)=\int_0^\infty K(v\tau)p(v)\,dv \qquad 0\leqslant\tau<\infty \tag{1.1}$$

where the kernel K depends only on the product of v and τ and has the property that

$$\int_0^\infty |K(x)|x^{-1/2}\,dx$$

are also finite.

This general class of equation includes the Laplace transform, the Fourier sine and cosine transforms and many other integral equations of importance in physics. Such equations often arise in the context of physical experiments where the function $p(v)$ about which information is required is related to the measured quantity $g(\tau)$ by means of a Mellin convolution with the instrumental function K. A particular example which occurs in laser anemometry will be discussed in detail in a separate publication in which experimental factors such as sampled data and truncation are considered. The purpose of the present paper is to study the problem quite generally, to gain a clear insight to the difficulties involved and to develop the mathematical formulae needed to tackle such problems.

Equation (1.1) has the form of a Fredholm integral equation of the first kind and it is well known that the problem of solving such equations is basically ill-conditioned. Many physicists have discovered, after much wasted effort, that it is essential to understand this feature before attempting to compute solutions.

0305-4770/78/0009-1729$03.00

In § 2 we discuss the ill-conditioning and fundamental difficulties associated with it. Then, by means of an example from classical optics, we show how it may be understood physically in terms of limited information capacity and introduce the concept of a generalised Shannon number. In this context we review, briefly, the established regularisation techniques of smoothing and truncation which are often used to avoid the inherent difficulties. In § 3 we derive analytic expressions for the eigenfunctions and eigenvalues of integral equations in the class described by equation (1.1). The eigenfunctions are shown to be mutually orthogonal and complete and in § 4 they are used to derive a formal solution to the integral equation. The solution is, of course, inherently ill-conditioned. However, when it is derived in terms of eigenfunctions it is easy to apply the information theory discussed in § 2 and extract the maximum amount of available information from the equation.

In § 4 the analytic expression for the eigenvalues is used to study the concept of channel capacity in the context of equation (1.1). In this way it is possible to describe how the amount of information which can be obtained from a given equation depends on the kernel and it is easy to show for example, that a Fourier transform contains more information than the equivalent Laplace transform—a fact which is clearly reflected in the comparative ease with which it can be inverted.

Finally, in § 5, we include some specific numerical examples related to the solution of equation (1.1). These serve to illustrate the points discussed previously and demonstrate clearly the need to consider information content in order to avoid obtaining meaningless results. In particular the theory is used to tackle the Laplace transform inversion in a well-conditioned manner. This difficult numerical problem, which is frequently encountered by physicists and engineers, is still the subject of much attention in the literature.

2. Ill-conditioning and information theory

The general Fredholm integral equation of the first kind takes the form

$$h(x)=\int_a^b K(x, y)q(y)\,\mathrm{d}y. \tag{2.1}$$

The problem of solving such equations has been studied by many authors (e.g. Phillips 1962, Tikonov 1963, Baker *et al* 1964) and a useful review is given by Miller (1974). Their ill-conditioned nature may be illustrated in a rather simple manner as follows.

In equation (2.1) the addition to $q(y)$ of an oscillatory function such as

$$\delta q_\omega(y)=\sin(\omega y) \tag{2.2}$$

causes a change in the function $h(x)$ given by

$$\delta h_\omega(x)=\int_a^b K(x, y)\sin(\omega y)\,\mathrm{d}y. \tag{2.3}$$

Now for any integrable kernel

$$\int_a^b K(x, y)\sin(\omega y)\,\mathrm{d}y \xrightarrow[\omega\to\infty]{} 0 \tag{2.4}$$

and so it is possible to make δh_ω arbitrarily small by choosing a sufficiently large value of ω. Therefore, since any numerical calculation must be subject to some error,

however small, there will be values of ω for which it is not possible to distinguish the appropriate change δh_ω when computing the solution to equation (2.1) and any attempt to determine the corresponding component $\delta q_\omega(y)$ in the solution will lead to entirely arbitrary results. In such circumstances it is not possible to obtain a complete and unique result. In view of the form of equation (2.1) it is convenient to restrict our discussion to equations having a non-degenerate, symmetric kernel.

The following specific example (figure 1) from classical optics and communication theory has this form and serves to illustrate further the problem of ill-conditioning by indicating the fundamental cause and showing how it may be understood physically.

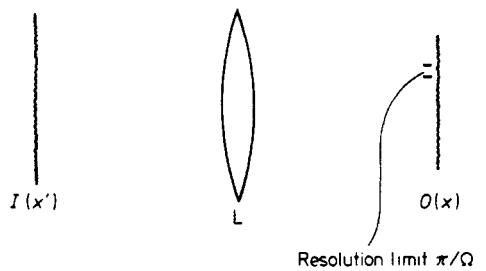

Figure 1. Optical analogy.

Light from a one-dimensional space-limited object described by the function $O(x)(|x|<X/2)$ passes through a lens L of finite aperture to form a band-limited object $I(x')$ which is given by the equation

$$I(x')=\frac{1}{2\pi}\int_{-\Omega}^{\Omega} d\omega\, e^{-i\omega x'}\int_{-X/2}^{X/2} e^{i\omega x}O(x)\,dx \tag{2.5}$$

i.e.

$$I(x')=\int_{-X/2}^{X/2}\frac{\sin[\Omega(x-x')]}{\pi(x-x')}O(x)\,dx \tag{2.6}$$

where Ω is the highest spatial frequency transmitted by the lens. In principle, given any image function $I(x')$ it is possible to determine mathematically the exact object $O(x)$ from which it was obtained. However it is well known physically that a lens with finite spatial frequency bandwidth Ω has an associated resolution limit π/Ω and so the image formed by an object of spatial extent X can contain only a finite number $S=X\Omega/\pi$ of independent components or degrees of freedom, S being the Shannon number of information theory. The apparent contradiction between these two points of view is easily resolved when the mathematical solution is derived in the following manner (di Francia 1969). Since the kernel in equation (2.6) is symmetric in x and x' it follows from Hilbert–Schmidt theory that the eigenfunctions $\phi_n(x')$ which satisfy the equation

$$\int_{-X/2}^{X/2}\frac{\sin[\Omega(x-x')]}{\pi(x-x')}\phi_n(x')\,dx'=\lambda_n\phi_n(x) \tag{2.7}$$

form a complete orthonormal basis over $[-X/2, X/2]$ (see, e.g., Courant and Hilbert 1953) and that the eigenvalues λ_n are real. It is possible, therefore, to expand the

image function $I(x')$ in terms of these eigenfunctions:

$$I(x') = \sum_{n=0}^{\infty} b_n \phi_n(x') \tag{2.8}$$

where

$$b_n = \int_{-X/2}^{X/2} I(x')\phi_n(x')\,dx'. \tag{2.9}$$

$O(x)$ may also be expanded in this way:

$$O(x) = \sum_{n=0}^{\infty} a_n \phi_n(x) \tag{2.10}$$

where

$$a_n = \int_{-X/2}^{X/2} O(x)\phi_n(x)\,dx \tag{2.11}$$

and by substituting this expression into equation (2.6) it is easy to show that

$$b_n = \lambda_n a_n. \tag{2.12}$$

Equation (2.10) may therefore be written in the form

$$O(x) = \sum_{n=0}^{\infty} \frac{b_n}{\lambda_n} \phi_n(x) \tag{2.13}$$

which expresses the object function $O(x)$ in terms of the image function $I(x')$. This expansion expresses the complete mathematical solution to equation (2.6) but in practice it is impossible to evaluate entirely, due to the behaviour of the eigenvalues λ_n which appear in the denominators. Slepian and Pollack (1961) showed that the eigenfunctions of equation (2.6) are the prolate spheroidal functions and that the eigenvalues are essentially unity for values of n up to $X\Omega/\pi$ but fall off to zero extremely rapidly as n is increased further. Figure 2 illustrates this behaviour for $X\Omega/\pi = 11{\cdot}5$. The extremely sharp cut-off means that those terms of equation (2.13) for which $n > X\Omega/\pi$ must be divided by an extremely small number and so noise or error on the value of a_n will cause these terms to diverge dramatically. Such components, in effect, cannot be determined and must be omitted from the result if it

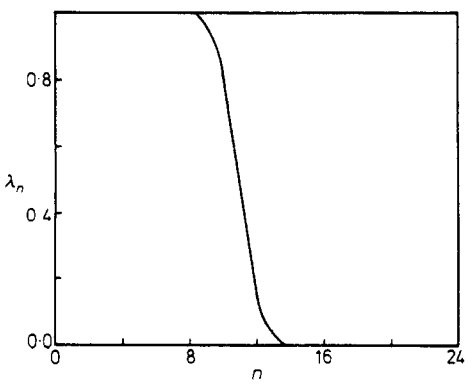

Figure 2. Eigenvalues λ_n of equation (2.7) as a function of n.

is not to be completely corrupted. In practice only $X\Omega/\pi$ terms of the expansion in equation (2.13) can be calculated and so the calculation is limited to the same number of degrees of freedom (the Shannon number) as one finds physically due to the finite resolution limit of the lens.

It is clear from the above example that the eigenvalue structure associated with problems of this kind is of fundamental importance. The eigenfunctions may be regarded as basic elements of information which retain their identity under the action of the integral operator

$$\int_a^b K(x, y)\phi_n(y)\,dy = \lambda_n\phi_n(x) \tag{2.14}$$

but are scaled in magnitude by the eigenvalue λ_n. The larger λ_n the more efficient is the transmission of the corresponding information element through the integral. Elements for which λ_n is sufficiently small are transmitted so weakly that they become lost in noise.

In general the behaviour of λ_n with n will not be as dramatic as that found by Slepian and Pollack but it is always possible to order the eigenfunctions and eigenvalues so that $\lambda_n \geqslant \lambda_m$ for $n < m$ and then the best conditioned approach towards the solution is to evaluate the lowest-order components first and progressively add as many higher-order terms as possible. Unless λ_n is asymptotically constant (a very special situation pertaining to the Fourier transforms as shown in § 4) the calculations will eventually become noise limited and so the series must be terminated. The number of terms which can be included in the series may be regarded as a generalised Shannon number which obviously depends on the size of the errors which occur in the calculations and measurements. It also depends on the actual kernel according to its eigenvalue structure. The faster λ_n decays to any given noise level the less information can be extracted from the equation. Each integral operator may be thought to have a finite channel capacity which is related to the width of its eigenvalue 'spectrum'.

It is important to realise that in such circumstances one can never obtain the solution to equation (2.1). An infinite amount of information is not available and the solution must be written in the form

$$q(y) = \sum_{n=0}^{N} a_n\phi_n(y) + \sum_{n=N+1}^{\infty} \theta_n\phi_n(y) \tag{2.15}$$

where the θ_n are unknown parameters. If the kernel describes a physical experiment then the corresponding information cannot be obtained unless the experiment is suitably changed. In the optics example it is not possible to achieve higher resolution without using a larger aperture lens. On the other hand, there may be situations in which the information available is sufficient for a particular experiment. For example, if it is known that an object under investigation contains no very fine detail then the image formed by a lens with the appropriate resolution limit can contain all the information necessary to specify the object completely. In this case it is known, *a priori*, that the coefficients of all terms in equation (2.13) for which $n > X\Omega/\pi$ must be zero. More generally, there must be sufficient information to specify the parameters of a suitable physical model for the solution from which the missing components may then be determined. A complete solution can only be obtained when some sort of model is employed.

Current mathematical literature on the numerical solution of Fredholm integral equations of the first kind contains two main methods of regularisation to avoid the problems of ill-conditioning. The first, which was considered by Tikonov (1963) and by Baker *et al* (1964), involves the use of an eigenfunction expansion as described above. The higher-order terms which cannot be evaluated are assumed to be zero and so the series is simply truncated. In effect the assumption that higher-order terms can be ignored is a simple form of model based on the fact that one is usually concerned with fairly smooth solutions which can be represented quite accurately in terms of lower-order eigenfunctions. The second method, which was introduced by Phillips (1962), involves the use of smoothing constraints to weight against highly oscillatory components which might otherwise appear with arbitrarily large magnitude in the solution. This approach is also based on an assumption that the solution is fairly smooth and can be modelled adequately using only slowly varying components.

The two methods clearly have much in common and Miller (1974) shows that the smoothing technique is, in fact, equivalent to the eigenfunction approach. The series is not truncated sharply but the higher-order components are progressively weighted out. Lewis (1975) compares the two methods at some length and treats some specific examples. Whichever approach is taken it is necessary to choose a value for the relevant regularisation parameter—the number N of eigenfunctions which may be used, or the smoothing parameter α—and in general this can only be decided by trial and error against some criterion of expected results.

In view of the discussion earlier in this section and the information concepts introduced, the expansion in terms of eigenfunctions can be seen to have fundamental advantages over other approaches. Those components of the solution which can be determined in practice are optimally compacted into the lower-order terms and may be evaluated without any influence from the weaker ones. Moreover the transition to noise will be as sharp as possible under these conditions—an important factor in choosing the regularisation parameter. In order to implement this approach it is necessary to obtain the eigenfunctions and eigenvalues of the particular integral equation under consideration. However, the numerical evaluation of these functions presents, in itself, a problem of noise and information similar to that associated with the original inversion. Indeed there are situations in which conventional matrix methods may prove quite futile for this purpose. For example, the eigenfunctions of equation (1.1) (derived in § 3) have zero-crossings which are exponentially spaced and so they cannot be represented effectively in terms of uniformly sampled values. Clearly it is of the utmost value to have available analytic expressions for the appropriate eigenfunctions and eigenvalues.

3. Derivation of eigenfunctions and eigenvalues

This section is devoted to deriving expressions for the eigenfunctions $\phi_\omega(v)$ and eigenvalues λ_ω which satisfy

$$\int_0^\infty K(v\tau)\phi_\omega(v)\,\mathrm{d}v = \lambda_\omega\phi_\omega(\tau) \tag{3.1}$$

in order that the theory discussed in § 2 may be applied to equations in the class described by equation (1.1).

Consider the function

$$\phi_s(v) = Av^{-s} + Bv^{s-1} \tag{3.2}$$

where A, B and s are complex numbers. Substituting $\phi_s(v)$ into equation (1.1) leads to a function of the form

$$g_s(\tau) = \int_0^\infty K(v\tau)(Av^{-s} + Bv^{s-1})\,dv \tag{3.3}$$

which exists and is finite provided that the integral converges. On making the substitution $z = v\tau$ we obtain

$$g_s(\tau) = \int_0^\infty K(z)\left[A\left(\frac{z}{\tau}\right)^{-s} + B\left(\frac{z}{\tau}\right)^{s-1}\right]\frac{dz}{\tau} \tag{3.4}$$

i.e.

$$g_s(\tau) = A\tilde{K}(1-s)\tau^{s-1} + B\tilde{K}(s)\tau^{-s} \tag{3.5}$$

where $\tilde{K}(s)$ is the Mellin transform of K, defined by

$$\tilde{K}(s) = \int_0^\infty x^{s-1}K(x)\,dx. \tag{3.6}$$

Assuming that $\tilde{K}(s)$ exists for $\alpha < \mathrm{Re}(s) < \beta$, then it follows that $\tilde{K}(1-s)$ exists for $1-\beta < \mathrm{Re}(s) < 1-\alpha$ and so equation (3.5) is properly defined provided $\alpha \leqslant \frac{1}{2} \leqslant \beta$. Clearly if we choose

$$A = \sqrt{(\tilde{K}(s))}; \qquad B = \pm\sqrt{(\tilde{K}(1-s))} \tag{3.7}$$

then

$$g_s(\tau) = \pm\sqrt{(\tilde{K}(s)\tilde{K}(1-s))}\phi_s(\tau) \tag{3.8}$$

and so the functions

$$\phi_s^{\pm}(v) = \sqrt{(\tilde{K}(s))}v^{-s} \pm \sqrt{(K(1-s))}v^{s-1} \tag{3.9}$$

are eigenfunctions satisfying equation (3.1) with corresponding eigenvalues

$$\lambda_s^{\pm} = \pm\sqrt{(\tilde{K}(s)\tilde{K}(1-s))}. \tag{3.10}$$

Setting $s = \frac{1}{2} + i\omega$ where ω is real and unbounded and introducing a convenient multiplicative factor yields a continuum of real eigenfunctions given by

$$\psi_\omega^+(v) = \frac{\sqrt{(\tilde{K}(\frac{1}{2}+i\omega))}v^{-\frac{1}{2}-i\omega} + \sqrt{(\tilde{K}(\frac{1}{2}-i\omega))}v^{-\frac{1}{2}+i\omega}}{2\sqrt{(\pi|\tilde{K}(\frac{1}{2}+i\omega)|)}} = \frac{\mathrm{Re}[\sqrt{(\tilde{K}(\frac{1}{2}+i\omega))}v^{-\frac{1}{2}-i\omega}]}{\sqrt{(\pi|\tilde{K}(\frac{1}{2}+i\omega)|)}} \tag{3.11}$$

and

$$\psi_\omega^-(v) = \frac{\sqrt{(\tilde{K}(\frac{1}{2}+i\omega))}v^{-\frac{1}{2}-i\omega} - \sqrt{(\tilde{K}(\frac{1}{2}-i\omega))}v^{-\frac{1}{2}+i\omega}}{2i\sqrt{(\pi|\tilde{K}(\frac{1}{2}+i\omega)|)}} = \frac{\mathrm{Im}[\sqrt{(\tilde{K}(\frac{1}{2}+i\omega))}v^{-\frac{1}{2}-i\omega}]}{\sqrt{(\pi|\tilde{K}(\frac{1}{2}+i\omega)|)}} \tag{3.12}$$

with real eigenvalues

$$\lambda_\omega^{\pm} = \pm|\tilde{K}(\tfrac{1}{2}+i\omega)|. \tag{3.13}$$

These functions are well defined provided that the transform $\tilde{K}(\frac{1}{2}+i\omega)$ exists and a sufficient requirement for this to be true is that $\int_0^\infty |K(x)|x^{-1/2}\,dx$ is finite. In view of the symmetry relationship

$$\psi_\omega^\pm(v) = \pm\psi_{-\omega}^\pm(v) \tag{3.14}$$

it is sufficient to consider $\omega \geqslant 0$.

The eigenfunctions are not normalisable but they may be shown to be mutually orthogonal. By making the substitution $v = e^x$ it is easy to deduce that

$$\int_0^\infty v^{-\frac{1}{2}+i\omega} v^{-\frac{1}{2}-i\omega'}\,dv = \int_{-\infty}^\infty e^{i\omega x} e^{-i\omega' x}\,dx = 2\pi\delta(\omega-\omega') \tag{3.15}$$

and from this relationship it follows that

$$\begin{aligned}
&\int_0^\infty \psi_\omega^+(v)\psi_{\omega'}^+(v)\,dv \\
&\quad = \frac{1}{4\pi|\tilde{K}(\frac{1}{2}+i\omega)|}\int_0^\infty [\sqrt{(\tilde{K}(\tfrac{1}{2}+i\omega))}v^{-\frac{1}{2}-i\omega} + \sqrt{(\tilde{K}(\tfrac{1}{2}-i\omega))}v^{-\frac{1}{2}+i\omega}] \\
&\quad\quad \times[\sqrt{(\tilde{K}(\tfrac{1}{2}+i\omega'))}v^{-\frac{1}{2}-i\omega'} + \sqrt{(\tilde{K}(\tfrac{1}{2}+i\omega'))}v^{-\frac{1}{2}+i\omega'}]\,dv \\
&\quad = \begin{cases} \delta(\omega-\omega') & \text{if } \omega \neq 0 \\ 2\delta(\omega) & \text{if } \omega = 0. \end{cases}
\end{aligned} \tag{3.16}$$

Similarly it may be shown that

$$\int_0^\infty \psi_\omega^-(v)\psi_{\omega'}^-(v)\,dv = \begin{cases} \delta(\omega-\omega') & \text{if } \omega \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.17}$$

and

$$\int_0^\infty \psi_\omega^+(v)\psi_{\omega'}^-(v)\,dv = 0. \tag{3.18}$$

In terms of real quantities the eigenfunctions take the form

$$\psi_\omega^+(v) = \frac{1}{\sqrt{\pi}}\{\cos(\theta/2)v^{-1/2}\cos[\omega\ln(v)] + \sin(\theta/2)v^{-1/2}\sin[\omega\ln(v)]\} \tag{3.19}$$

and

$$\psi_\omega^-(v) = \frac{1}{\sqrt{\pi}}\{\sin(\theta/2)v^{-1/2}\cos[\omega\ln(v)] - \cos(\theta/2)v^{-1/2}\sin[\omega\ln(v)]\} \tag{3.20}$$

and the eigenvalues are given by

$$\lambda_\omega^\pm = \pm\sqrt{(a^2+b^2)} \tag{3.21}$$

where

$$a = \mathrm{Re}(\tilde{K}(\tfrac{1}{2}+i\omega)); \qquad b = \mathrm{Im}(\tilde{K}(\tfrac{1}{2}+i\omega)) \tag{3.22}$$

and

$$\theta = \tan^{-1}(b/a). \tag{3.23}$$

If K is a real kernel, a and b take the forms

$$a = \int_0^\infty K(z) z^{-1/2} \cos[\omega \ln(z)]\, dz \tag{3.24}$$

$$b = \int_0^\infty K(z) z^{-1/2} \sin[\omega \ln(z)]\, dz. \tag{3.25}$$

Any piecewise continuous function $p(v)$ for which $\int_0^\infty |p(v)| v^{-1/2}\, dv$ exists can be expanded in terms of these eigenfunctions, i.e. we may write

$$p(v) = \int_0^\infty a_\omega^+ \psi_\omega^+(v)\, d\omega + \int_0^\infty a_\omega^- \psi_\omega^-(v)\, d\omega \tag{3.26}$$

where

$$a_\omega^\pm = \int_0^\infty p(v) \psi_\omega^\pm(v)\, dv \tag{3.27}$$

by virtue of the orthogonality of the $\psi_\omega(v)$. This may be shown quite simply as follows. Equation (3.26) may be written in the equivalent form

$$p(v) = \int_0^\infty C_\omega v^{-1/2} \cos[\omega \ln(v)]\, d\omega + \int_0^\infty S_\omega v^{-1/2} \sin[\omega \ln(v)]\, d\omega \tag{3.28}$$

where

$$C_\omega = \frac{1}{\sqrt{\pi}} [a_\omega^+ \cos(\theta/2) + a_\omega^- \sin(\theta/2)] \tag{3.29}$$

and

$$S_\omega = \frac{1}{\sqrt{\pi}} [a_\omega^+ \sin(\theta/2) - a_\omega^- \cos(\theta/2)] \tag{3.30}$$

and on making the substitution $v = e^y$ this becomes

$$e^{y/2} p(e^y) = \int_0^\infty C_\omega \cos(\omega y)\, d\omega + \int_0^\infty S_\omega \sin(\omega y)\, d\omega. \tag{3.31}$$

Equation (3.31) simply expresses the function $e^{y/2} p(e^y)$ (which must also be piecewise continuous) in terms of its Fourier sine and cosine transform components C_ω and S_ω and from the theory of Fourier transforms this is known to be valid provided $\int_{-\infty}^\infty e^{y/2} |p(e^y)|\, dy$ exists.

It follows that equation (3.26) is also valid provided that $\int_0^\infty |p(v)| v^{-1/2}\, dv$ exists.

4. Inversion formulae and information capacity

Using the eigenfunctions and eigenvalues derived in § 3 it is easy to derive the solution to equation (1.1). By substituting the expansion (3.26) for $p(v)$ into the equation we obtain

$$g(\tau) = \int_0^\infty a_\omega^+ \lambda_\omega^+ \psi_\omega(\tau)\, d\omega + \int_0^\infty a_\omega^- \lambda_\omega^- \psi_\omega^-(\tau)\, d\omega \tag{4.1}$$

and from the orthogonality relationships (3.16), (3.17), (3.18), it follows that

$$a_\omega^\pm = \frac{1}{\lambda_\omega^\pm}\int_0^\infty g(\tau)\psi_\omega^\pm(\tau)\,d\tau \tag{4.2}$$

so that $p(v)$ may be written

$$p(v)=\int_0^\infty d\omega\,\psi_\omega^+(v)\frac{1}{\lambda_\omega^+}\int_0^\infty d\tau\,\psi_\omega^+(\tau)g(\tau)+\int_0^\infty d\omega\,\psi_\omega^-(v)\frac{1}{\lambda_\omega^-}\int_0^\infty d\tau\,\psi_\omega^-(\tau)g(\tau). \tag{4.3}$$

The form of this solution is clearly analogous to that given in equation (2.13) the discrete summation being replaced by an integral since equation (1.1) possesses an infinite continuum of eigenvalues. From the discussion in § 2 it is clear that, in general, equation (4.3) is quite ill-conditioned. In practice, it is impossible to gain any information about those components of $p(v)$ for which $\omega > \omega_{max}$, a number whose value depends on the errors during the calculation or in measuring $g(\tau)$; the solution must therefore be written in the form

$$p(v)=\int_0^{\omega_{max}} d\omega\,\psi_\omega^+(v)\frac{1}{\lambda_\omega^+}\int_0^\infty d\tau\,\psi_\omega^+(\tau)g(\tau)+\int_0^{\omega_{max}} d\omega\,\psi_\omega^-(v)\frac{1}{\lambda_\omega^-}\int_0^\infty d\tau\,\psi_\omega^-(\tau)g(\tau)$$
$$+\int_{\omega_{max}}^\infty d\omega\,\alpha_\omega^+\psi_\omega^+(v)+\int_{\omega_{max}}^\infty d\omega\,\alpha_\omega^-\psi_\omega^-(v) \tag{4.4}$$

where the coefficients $\alpha_\omega^\pm$ cannot be evaluated independently and must be determined by means of some *a priori* knowledge about the result.

The solution to equation (1.1) may be derived more directly by taking the Mellin transform of both sides. Since the integral has the form of a Mellin convolution this yields

$$\tilde{g}(s)=\tilde{K}(s)\tilde{p}(1-s) \tag{4.5}$$

and so

$$\tilde{p}(s)=\tilde{g}(1-s)/\tilde{K}(1-s). \tag{4.6}$$

The transform of $p(v)$ is analytic for $\mathrm{Re}(s)=\frac{1}{2}$ and so it follows from the formula for the inverse Mellin transform that

$$p(v)=\frac{1}{2\pi i}\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty} v^{-s}\frac{\tilde{g}(1-s)}{\tilde{K}(1-s)}\,ds. \tag{4.7}$$

The equivalence of equation (4.7) and equation (4.3) is readily shown by making the substitution $s=\frac{1}{2}+i\omega$ and writing equation (4.7) in the form

$$p(v)=\frac{1}{2\pi}\int_{-\infty}^\infty d\omega\frac{v^{-\frac{1}{2}-i\omega}\sqrt{(\tilde{K}(\frac{1}{2}+i\omega))}}{|\tilde{K}(\frac{1}{2}-i\omega)|^2}\int_0^\infty dz\,g(z)z^{-\frac{1}{2}-i\omega}\sqrt{(\tilde{K}(\tfrac{1}{2}+i\omega))}. \tag{4.8}$$

Expressed in terms of eigenfunctions this becomes

$$p(v)=\frac{1}{2}\int_{-\infty}^\infty d\omega\frac{1}{\lambda_\omega}(\psi_\omega^+(v)+i\psi_\omega^-(v))\int_0^\infty g(z)(\psi_\omega^+(z)+i\psi_\omega^-(z))\,dz \tag{4.9}$$

and by applying the symmetry properties of $\psi_\omega^\pm(v)$ equation (4.3) is immediately recovered.

An alternative approach worth mentioning involves making the substitution $v = e^y$ and $\tau = e^{-x}$. Equation (1.1) then has the form of a Fourier convolution and so it may be solved by taking the Fourier transform of both sides. The form of solution which results is identical to that in equation (4.7), the Mellin transforms being replaced by equivalent Fourier transforms. This approach was used by Gardner *et al* (1959) in an effort to evaluate inverse Laplace transforms using the Fourier transform algorithm. Their calculations, of course, were subject to the usual problems of ill-conditioning encountered when inverting Laplace transforms.

When the solution to equation (1.1) is expressed in the form of equation (4.7) or the equivalent form used by Gardener *et al*, it is not clear how to separate it into well-conditioned components which can be evaluated and ill-conditioned ones which cannot. The derivation in terms of eigenfunctions, however, enables us to consider the information aspect and deduce immediately equation (4.4).

In § 2 we pointed out that the value of ω_{max} depends on the level of noise in the calculation and also on the eigenvalue structure of the kernel. It is interesting to compare the information capacity of some familiar integral equations, in this way. Consider for example, the equation

$$g(\tau) = \int_0^\infty e^{-\alpha v \tau} \cos(\beta v \tau) p(v)\, dv \tag{4.10}$$

where α and β are real. In the limit $\alpha \to 0$ this takes the form of a Fourier cosine transform while in the limit $\beta \to 0$ it becomes a Laplace transform. The rate at which the eigenvalues decay to zero is easily deduced from equation (3.10). The eigenvalue 'spectrum' takes the form

$$|\lambda_\omega^{\pm}|^2 = |\tilde{K}(\tfrac{1}{2} + i\omega)|^2 \tag{4.11}$$

where

$$\begin{aligned}\tilde{K}(\tfrac{1}{2} + i\omega) &= \int_0^\infty z^{-\frac{1}{2}+i\omega} e^{-\alpha z} \cos(\beta z)\, dz \\ &= \frac{\Gamma(\frac{1}{2} + i\omega) \cos[(\frac{1}{2} + i\omega) \tan^{-1}(\beta/\alpha)]}{(\alpha^2 + \beta^2)^{\frac{1}{4}+i\omega/2}}.\end{aligned} \tag{4.12}$$

In the limit $\beta \to 0$ this becomes

$$\tilde{K}(\tfrac{1}{2} + i\omega) = \alpha^{-\frac{1}{2}-i\omega} \Gamma(\tfrac{1}{2} + i\omega) \tag{4.13}$$

and so the eigenvalue 'spectrum' for the Laplace transform is given by

$$|\lambda_\omega^{\pm}|^2 = \frac{1}{\alpha} |\Gamma(\tfrac{1}{2} + i\omega)|^2 \equiv \frac{\pi}{\alpha \cosh(\pi\omega)} \tag{4.14}$$

which decays as $(\pi/\alpha) e^{-\pi\omega}$ for large ω. When $\alpha = 0$ we have

$$\tilde{K}(\tfrac{1}{2} + i\omega) = \beta^{-\frac{1}{2}-i\omega} \Gamma(\tfrac{1}{2} + i\omega) \cos(\tfrac{1}{4}\pi + \tfrac{1}{2} i\omega\pi)$$

and so the eigenvalue spectrum for the Fourier transform takes the form

$$|\lambda_\omega^{\pm}|^2 = \frac{1}{2\beta} |\Gamma(\tfrac{1}{2} + i\omega)|^2 \left[\cosh^2\left(\frac{\omega\pi}{2}\right) + \sinh^2\left(\frac{\omega\pi}{2}\right) \right] \equiv \frac{\pi}{2\beta} \tag{4.15}$$

which never decays to zero.

From the above analysis it is clear that the Fourier transform has a much greater information capacity than the Laplace transform and this feature becomes evident in practice in terms of the comparative ease with which it can be inverted. In fact the Fourier transform has an infinite information capacity by virtue of which its analytic inversion formula can be used to obtain a complete solution. This desirable property is not shared by the Laplace transform. Although there is also a well known formula for inverting these transforms it cannot be used in practice to obtain a complete result and the solution must be expressed in the form of equation (4.4) with an infinite amount of information undetermined.

The more general limit $\alpha/\beta \ll 1$ is also interesting. In this limit

$$\tilde{K}(\tfrac{1}{2}+i\omega) \simeq \frac{1}{\sqrt{2}}(\alpha^2+\beta^2)^{-\frac{1}{4}-i\omega/2}\Gamma(\tfrac{1}{2}+i\omega)\left\{\cosh\left[\omega\left(\frac{\pi}{2}-\frac{\alpha}{\beta}\right)\right]-i\sinh\left[\omega\left(\frac{\pi}{2}-\frac{\alpha}{\beta}\right)\right]\right\} \tag{4.16}$$

and so

$$|\lambda_\omega^{\pm}|^2 = \frac{\pi}{2(\alpha^2+\beta^2)^{1/2}}\frac{\cosh[\omega(\pi-2\alpha/\beta)]}{\cosh(\pi\omega)}, \tag{4.17}$$

which decays as

$$\frac{\pi}{2(\alpha^2+\beta^2)^{1/2}}e^{-2\omega\alpha/\beta} \qquad \text{for large } \omega.$$

Clearly the width of the eigenvalue spectrum is proportional to β/α and so the information capacity of equation (4.4) can be seen to depend directly on the number of cosine cycles which exist within the exponential half-width.

Any equation of type (1.1) may be studied in this way although few lend themselves to such convenient analytic treatment as that presented in the above example. In all cases, however, such information features become apparent when the solution is approached by means of equation (4.4).

5. Numerical examples

We now illustrate the general theory discussed above by applying it together with the eigenfunction expressions (3.11) and (3.12), to the numerical solution of the equation

$$g(\tau) = \frac{(1+\tau)^2-\beta^2\tau^2}{[(1+\tau)^2+\beta^2\tau^2]^2} = \int_0^\infty e^{-v\tau}\cos(\beta v\tau)p(v)\,dv. \tag{5.1}$$

This equation has the form (4.10) which was discussed in detail in §4, the particular form of $g(\tau)$ having been chosen to correspond to the analytic solution

$$p(v) = v\,e^{-v}. \tag{5.2}$$

The results of our computations are compared with this known solution but no knowledge of it is assumed during the calculations. These involve the use of equation (4.4) and demonstrate clearly the importance of the finite upper limit ω_{max} which appears in the integrals. We examine the results obtained by varying ω_{max} under varying conditions of noise and information capacity, the latter being determined by the particular value of β. Throughout the calculations the undetermined parameters α_ω have been set arbitrarily equal to zero as in the truncation method of regularisation and we will see that as the noise is increased the maximum value of ω for which

meaningful coefficients a_ω can be obtained decreases and α_ω's of more and more significance are lost until this incomplete inversion becomes incapable of reconstructing adequately the $p(v)$.

In order to evaluate (4.4) numerically it is necessary to set a finite upper limit L_2 and also a positive, non-zero lower limit L_1 (since $\psi^{\pm}_{\omega}(\tau)$ is not defined at $\tau = 0$) on the eigenfunction projection integrals (4.2), but the magnitude of the error caused by introducing these limits may be made arbitrarily small by choosing L_2 sufficiently large and L_1 sufficiently small. In the numerical work described in this section these integrals were evaluated to an accuracy of 10^{-5} using a standard computer algorithm and the limits

$$L_1 = 10^{-15}, \qquad L_2 = 10^5$$

were chosen so that the truncation errors given by the expressions

$$\left| \int_0^{L_1} g(\tau)\psi^{\pm}_{\omega}(\tau)\,\mathrm{d}\tau \right| \leqslant \frac{1}{\sqrt{\pi}} \int_0^{L_1} |g(\tau)|\tau^{-1/2}\,\mathrm{d}\tau \sim L_1^{1/2}$$

and

$$\left| \int_{L_2}^{\infty} g(\tau)\psi^{\pm}_{\omega}(\tau)\,\mathrm{d}\tau \right| \leqslant \frac{1}{\sqrt{\pi}} \int_{L_2}^{\infty} |g(\tau)|\tau^{-1/2}\,\mathrm{d}\tau \sim L_2^{-3/2}$$

were even smaller ($\sim 10^{-7{\cdot}5}$).

It is easy to show that the discrete subset of eigenfunctions

$$\psi^{\pm}_{n\Delta\omega}(\tau) \qquad n = 0, 1, 2, \ldots$$

where

$$\Delta\omega = 2\pi/[\ln(L_2) - \ln(L_1)] \tag{5.3}$$

form a complete orthogonal set on the interval $[L_1, L_2]$ and may be normalised by the factor $\Delta\omega^{1/2}$. For this reason our calculations reduced to the evaluation of the finite discrete summations

$$p_N(v) = \sum_{n=0}^{N} \frac{c_n^+}{\lambda^+_{n\Delta\omega}} \psi^+_{n\Delta\omega}(v) + \sum_{n=1}^{N} \frac{c_n^-}{\lambda^-_{n\Delta\omega}} \psi^-_{n\Delta\omega}(v) \tag{5.4}$$

where

$$c_n^{\pm} = \begin{cases} \Delta\omega \displaystyle\int_{L_1}^{L_2} g(\tau)\psi^{\pm}_{n\Delta\omega}(\tau)\,\mathrm{d}\tau & n \neq 0 \\ \dfrac{\Delta\omega}{2} \displaystyle\int_{L_1}^{L_2} g(\tau)\psi_0(\tau)\,\mathrm{d}\tau & n = 0 \end{cases} \tag{5.5}$$

with

$$\Delta\omega \simeq 0{\cdot}136 \qquad \text{and} \qquad N\Delta\omega = \omega_{\max}. \tag{5.6}$$

A uniform discretisation of this kind was particularly convenient for these examples since the value of $\omega_{\max}$ in which we are interested is simply proportional to the number N of terms retained in the series. In general it may be possible to evaluate the integral with respect to ω more efficiently using other techniques such as Gaussian quadrature.

The eigenfunction projections (5.5) were evaluated by making a substitution of the form $\tau = \mathrm{e}^x$ (analagous to that used in § 3) in order to obtain better conditioning of the

integrals which were calculated numerically using a four-point Gauss quadrature algorithm.

When $v = 0$, $\psi^{\pm}_{n\Delta\omega}(v)$ is singular and so it is not possible to evaluate $p_N(v)$.

5.1. Laplace transform

We first set $\beta = 0$ in equation (5.1) and so the problem was reduced to one of inverting a Laplace transform. In the first calculations the eigenfunction components of $g(\tau)$ were not subject to any source of noise apart from the small error ($\sim 10^{-5}$) which was incurred during the integration process. The results $p_N(v)$ which were obtained in this case are plotted in figure 3. The full curve is a plot of the analytic solution $v\,e^{-v}$ from which $g(\tau)$ was generated. The broken curve shows the result obtained for $p_N(v)$ by including twenty terms in equation (5.4) and although its shape is not dissimilar to that of the true solution, it does deviate noticeably from it due to the loss of higher components. When forty terms were included in equation (5.4) the result was found to differ very little from the known solution and on the scale of figure 3, the plot of $p_N(v)$ could not be distinguished from the full curve. Clearly the particular $p(v)$ chosen here does not contain significant components higher than these. However, when sixty terms were included in equation (5.4) the result deviated considerably as shown in the chain curve and as the value of N was increased further the result rapidly became wildly oscillatory and bore no resemblance whatever to the full curve. Thus, when $N \sim 60$ or greater the value of $\lambda_{N\Delta\omega}$ is so small that even the minute errors which were tolerated in the numerical evaluation of the eigenfunction components c_n were sufficient to dominate $p_N(v)$.

The next set of computations which we carried out was identical to the first one except for the fact that the calculated eigenfunction components were subjected to the addition of Gaussian random errors with standard deviation 10^{-3}. The results obtained are shown in figure 4 where the full curve is again a plot of the known

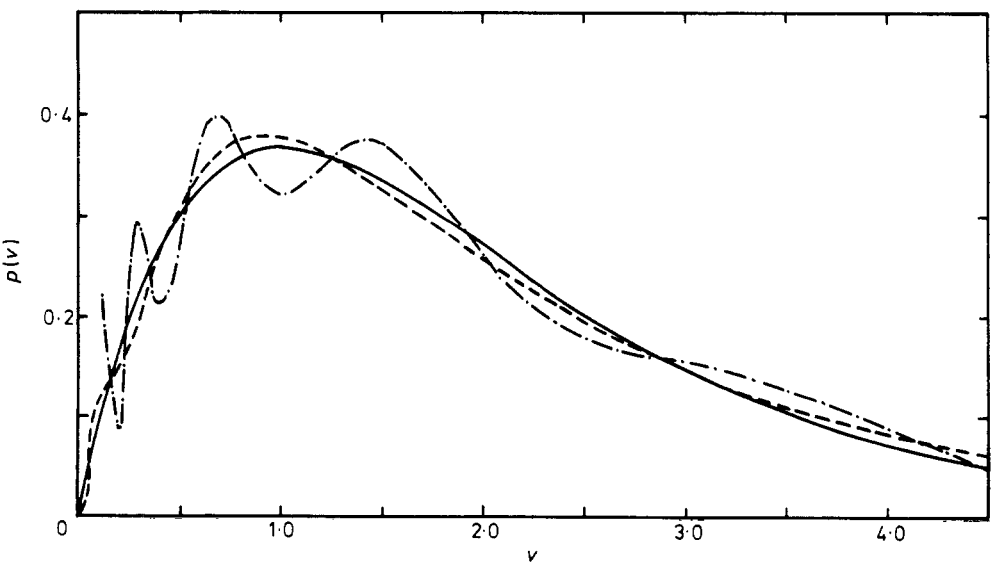

Figure 3. $p_N(v)$ as a function of v. $\beta = 0$ and noise $\sim 10^{-5}$: broken curve, $N = 20$; chain curve, $N = 60$; full curve, actual solution $p(v) = v\,e^{-v}$ (and $N = 40$).

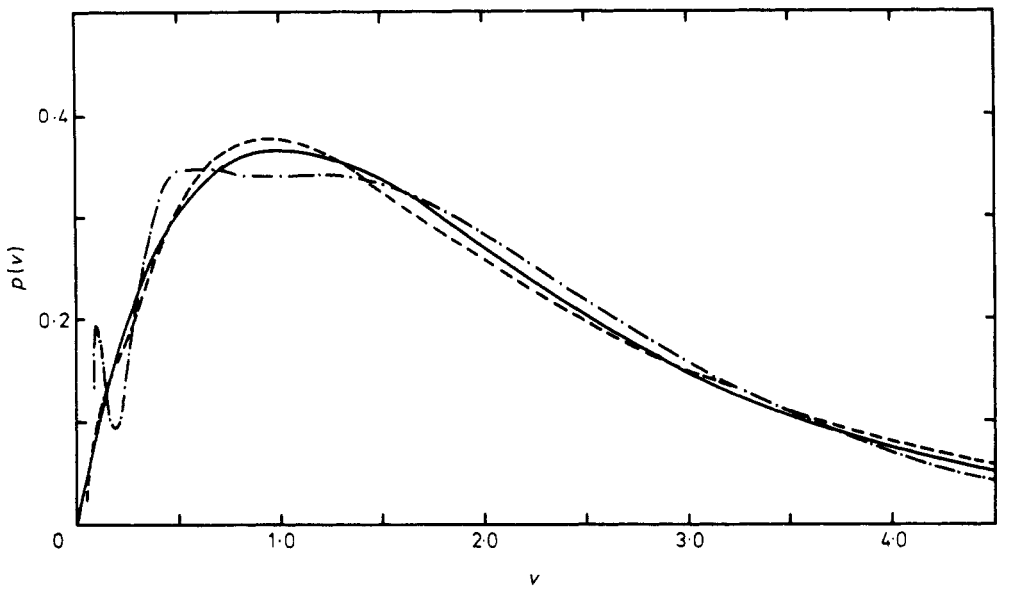

Figure 4. $p_N(v)$ as a function of v. $\beta = 0$ and noise $\sim 10^{-3}$: broken curve, $N = 20$; chain curve, $N = 30$; full curve, actual solution $p(v) = v\,e^{-v}$.

solution $v\,e^{-v}$. The result of including twenty terms in equation (5.4) is again plotted by means of the broken curve which is similar to that obtained in the previous case. When more terms were included in the summation, however, the results grew steadily worse. Even at $N = 30$ the behaviour was quite erratic as shown in the chain curve. There was no value of N for which $p_N(v)$ converged any closer to the known solution than at $N = 20$. The effect of the increased noise level in this calculation was to reduce the amount of information which could be extracted from equation (5.1) to the extent that $p(v)$ could not be represented accurately in terms of it. In other words the value of ω_{max} required to reconstruct $p(v)$ adequately ($\sim 40\,\Delta\omega$) was greater than the generalised Shannon number ($\sim 20\,\Delta\omega$) in this situation. However, if the true solution $p(v)$ had contained no eigenfunction components beyond $N = 20$ this level of noise would not have been significant.

5.2. Oscillatory kernel

In the final set of calculations β was given the value 4π and the computed eigenfunction components were again subjected to the addition of Gaussian random errors with standard deviation 10^{-3}. The results obtained are shown in figure 5 and, as before, the full curve is a plot of the known solution. The effect of including twenty terms in equation (5.4) is again illustrated in the broken curve which clearly deviates from the full curve in the same manner as before due to the incomplete basis. When forty terms were included in the summation, the function $p_N(v)$ which resulted could not be distinguished from the full curve in our figure and even when N was increased to 60 this was still true. Only when N was increased even further did the results begin to differ considerably from the actual solution. On comparing the results of this set of calculations with those shown in figure 4, it is clear that the effect of increasing β from 0 to 4π was to improve significantly the amount of information which could be

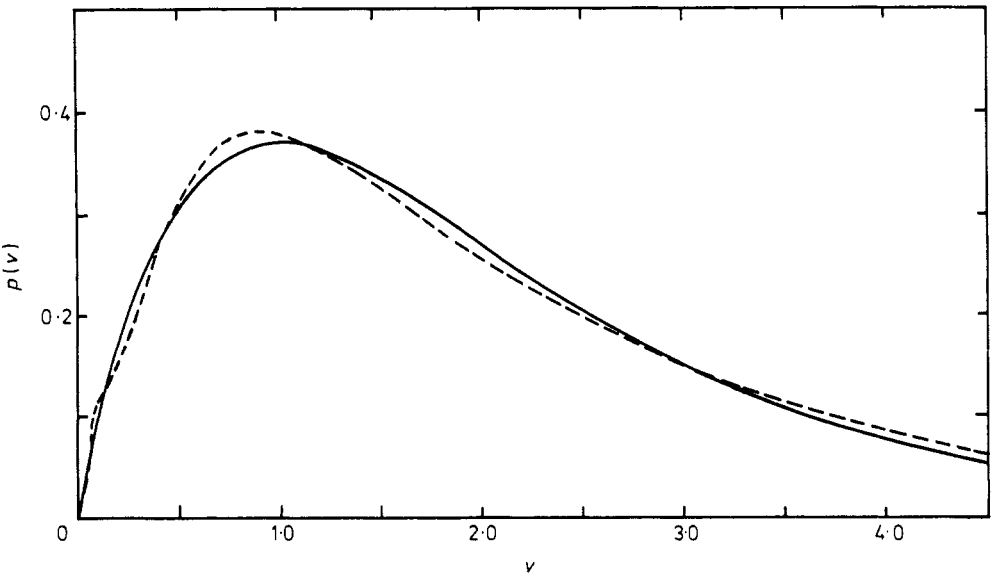

Figure 5. $p_N(v)$ as a function of v. $\beta = 4\pi$ and noise $\sim 10^{-3}$: broken curve, $N = 20$; full curve, actual solution $p(v) = v\,e^{-v}$ (and $N = 40$ and $N = 60$).

extracted from equation (5.1) in the same noise situation, i.e. to increase the generalised Shannon number from the region of 40 $\Delta\omega$ to approximately 60 $\Delta\omega$.

Although it is never possible to obtain a complete numerical solution to this sort of problem it is clear that in the first and third sets of calculations it was possible to extract sufficient information to represent the solution quite accurately. This would not have been so, of course, if the actual solution contained finer detail which could not be resolved using the number of eigenfunction components available 'above the noise'. From standard sampling theory it is easy to show that the function $p_N(v)$ of equation (5.4) cannot be resolved at the points v_1 and v_2 unless

$$v_2 < v_1\, e^{1/2N\Delta\omega} \tag{5.7}$$

Given the number N of eigenfunction coefficients which can be determined in any situation, equation (5.7) expresses the degree of resolution which may be achieved. In the third set of calculations it would have been possible to represent a solution $p(v)$ with finer detail than that chosen, using the higher number of components which could be recovered in this case.

It is important to note also, that the only grounds for dismissing the more oscillatory results, which were obtained for larger values of N in the calculations above, would be a good estimate of the generalised Shannon number which is a function only of the noise level and the eigenvalue 'spectrum'. Even with this value given, we have seen that it may not be possible to reconstruct the solution and it is essential to have *a priori* knowledge either of the absence of higher eigenvalue components than those available on information grounds or, more realistically, of a parametric form for the solution whereby the undetermined coefficients may be mapped onto the known ones. This approach raises some further interesting problems which are not dealt with here but which we hope to discuss in a future publication.

Acknowledgments

We are indebted to Mr J Abbiss of the Royal Aircraft Establishment, Farnborough for many stimulating and constructive discussions during the course of this work and we would also like to thank Mr J M Priday, a student at Kings College, London, for his valuable assistance with the computation.

References

Baker C T H, Fox L, Mayers D F and Wright K 1964 *Comp. J.* **7** 141–7
Courant R and Hilbert D 1953 *Methods of Mathematical Physics* (New York: Interscience)
di Francia T 1969 *J. Opt. Soc. Am* **59** 799
Gardner D G, Gardner J C, Laush G and Meinke W W 1959 *J. Chem. Phys.* **31** 978–86
Lewis B A 1975 *J. Inst. Math. Applic.* **16** 207–20
Miller G F 1974 *Numerical Solution of Integral Equations* eds L M Delves and J Walsh (Oxford: Clarendon)
Phillips D L 1962 *J. Assoc. Comp. Mach.* **9** 84–96
Slepian D and Pollack H O 1961 *Bell. Syst. Tech. J* **40** 43–63
Tikonov A N 1963, *Sov. Math.* **4** 1035–8