JOURNAL OF COMPUTATIONAL PHYSICS 7, 219–238 (1971)

A Smoothing Solution (Unfolding) of a Two Dimensional Density Function from its Measured Spectrum*

I. K. ABU-SHUMAYS AND L. D. MARINELLI

Argonne National Laboratory, Argonne, Illinois 60439

Received May 13, 1970

Density functions such as the density of radioactivity *in vivo* and those arising in spectral theory, potential theory and other branches of physics, are related to their measured spectra by Fredholm equations of the first kind. The problem of solving a Fredholm equation of the first kind is known to be mathematically "ill-posed." The experimental and statistical errors in a measurement may often be such that the corresponding solution for the density function is physically meaningless in the absence of additional constraints.

This paper is devoted to the solution of two-dimensional Fredholm equations of the first kind where the physical kernel of the equation can be approximated by a separable kernel. It introduces a smoothness constraint on the solution, which results essentially in smoothness by convolution. Two numerical experiments are given.

I. INTRODUCTION

In recent years it became increasingly evident that reliable methods of solving Fredholm integral equations of the first kind [5, 8]

$$\int_D K(\mathbf{x}, \mathbf{x}') f(\mathbf{x}')\, d\mathbf{x}' = g(\mathbf{x}), \qquad \mathbf{x} \in \mathscr{R}, \tag{1}$$

would have many useful applications in spectroscopy, potential theory, optimal programming, and other significant problems [2, 3, 5, 7]. During the past two years, jointly with others [1, 4], we have examined the various available numerical methods for solving the one dimensional form of Eq. (1) with the kernel $K(x, x')$ fairly accurately known and the spectra $g(x)$ experimentally measured with modest accuracy. Equation (1) can be regarded as a linear operator, operating on a density function $f(\mathbf{x})$ in the domain D to produce a spectrum $g(\mathbf{x})$ in the range $\mathscr{R}$. It is well

* Work performed under the auspices of the U. S. Atomic Energy Commission.

known [8] that this operator does *not have a bounded inverse* (we assume here that the kernel is nonsingular and that the operator has an inverse): an *infinitesimal change in g may cause a finite change in f.* Thus, the problem is mathematically "ill-posed."

We have found [1, 4] that for the one-dimensional model, the smoothing technique,[1] initiated by D. L. Phillips [8] in 1962, for solving Eq. (1), is very suitable for determining the unknown concentration of radioactivity per unit length of a linear source. This method is based on the physically meaningful assumption that (except possibly at a few irregular points) the magnitude of the density function (concentration of radioactivity, for example) cannot change abruptly between one point and adjacent points. Thus, the method is to seek that solution of Eq. (1) which satisfies appropriate smoothness requirements [1, 4, 5, 8].

Our initial investigations encourage us to generalize the smoothing techniques to solve two-dimensional Fredholm equations of the first kind. It is evident that a two- or three-dimensional description of the concentration of small amounts of radioactivity *in vivo* would have many applications, ranging from localization of internal or external contamination in research and industrial personnel, to the distribution of radioactivated natural element contents (^{24}Na, ^{49}Ca, etc.) in men occupationally exposed to tolerable doses of neutrons.

The Fredholm equation under consideration here, takes the form

$$\int_a^b dx' \int_c^d dy' \, K(x, x', y, y') f(x', y') = g(x, y) + \epsilon(x, y). \tag{2}$$

The function $f(x', y')$ is the *unknown density* of radioactivity, $g(x, y)$ is the *measured spectrum* subject to both statistical and experimental error, and $\epsilon(x, y)$ is the error. The measurement g is relatively insensitive to fictitious (positive-negative) density functions f' which satisfy the inequality

$$\int_a^b dx' \int_c^d dy' \, K(x, x', y, y') f'(x', y') \leqslant \epsilon'(x, y),$$

where ϵ' is a distribution whose amplitude is of the order of the statistical error in the measurement of $g(x, y)$. In other words, small errors in $g(x, y)$ may be amplified to such an extent that the exact solution of the Fredholm equation (2) [deleting the unknown error term $\epsilon(x, y)$] is physically meaningless. In addition, the effect of an error in g on a calculated value of f depends on the smoothness of the kernel. Thus,

[1] We have previously [1, 4, 6] referred to this technique as "regularization unfolding technique," especially when smoothing is combined with an iterative technique (see Also Ref. 7). The term "unfolding" denotes the inverse problem of finding a density function from its measured spectrum.

the success of solving Eq. (2) approximately by any method depends to a large extent on the accuracy of g and the shape of the kernel K.

In practice, if the range of $g(x, y)$ happens to be larger than the domain of $f(x', y')$ (for example, when measuring radioactivity in the human body, it is advantageous to take counts over a range including the body and extending beyong it, thus increasing the information content), then the domain is extended to coincide with the range, with the added restriction that the density function $f(x', y')$ should vanish beyond the original region.

II. Method of Solution

A straightforward extension of the smoothing technique to two (or more) dimensions can be achieved with the aid of tensor product analysis [2]. However, in arriving at the modified smoothing method presented here, we were seeking both a simplified form and a reduction of the time required for computations. We achieve this by the use of a separable kernel and by the adoption of smoothness criteria which result essentially in smoothness by convolution.

The first approximation we make is that the kernel $K(x, x', y, y')$ is separable, i.e.,

$$K(x, x', y, y') = K^{(1)}(x, x')\, K^{(2)}(y', y). \tag{3}$$

We expect kernels for physical problems to depend upon the distance between the point at which a measurement is made and the point measured as follows:

$$K(x, x', y, y') = K([x - x']^2 + [y - y']^2). \tag{4}$$

For the one-dimensional distribution of radioactivity *in vivo*, which we have reported elsewhere [1, 4], our experimental results indicate that the kernel can often be approximated by a Gaussian distribution. For two dimensions, such a kernel typically takes the form

$$K(x, x', y, y') = A \exp\{-[(x - x')^2 + (y - y')^2]/4\sigma^2\}, \tag{5}$$

and is indeed separable as in Eq. (3).

For the separable Kernel Eq. (3), the Fredholm equation (2) can be approximated by a matrix equation. We assume that the domain of $f(x', y')$ (the region $[a, b] \times [c, d]$) and the range of $g(x, y)$ coincide and thus use (x', y') and (x, y) interchangeably (see last paragraph of the introduction). The interval $[a, b]$ is subdivided into M parts $a = x_1 < x_2 < \cdots < x_M = b$ and we associate with these divisions appropriate quadrature weights $v_1, v_2, \ldots, v_M$. Similarly, the interval

$[c, d]$ is subdivided into N parts $c = y_1 < y_2 < \cdots < y_N = d$, and we associate with these divisions appropriate quadrature weights $u_1, u_2, \ldots, u_N$. Using Eq. (3), Eq. (2) can now be approximated as follows:

$$g(x_k, y_\ell) + \epsilon(x_k, y_\ell) = \int_a^b dx' \, K^{(1)}(x_k, x') \int_c^d f(x', y') \, K^{(2)}(y', y_\ell) \, dy'$$

$$\simeq \int_a^b dx' \, K^{(1)}(x_k, x') \sum_{j=1}^{N} f(x', y_j') \, u_j K^{(2)}(y_j', y_\ell)$$

$$\simeq \sum_{i=1}^{M} \sum_{j=1}^{N} K^{(1)}(x_k, x_i') \, v_i f(x_i', y_j') \, u_j K^{(2)}(y_j', y_\ell),$$

$$(k = 1, 2, \ldots, M), \quad (\ell = 1, 2, \ldots, N). \tag{6}$$

Using the abbreviated notation $g_{k\ell}$ for $g(x_k, y_\ell)$, etc., we replace the approximate Eq. (6) by

$$g_{k\ell} + \epsilon_{k\ell} = \sum_{i=1}^{M} \sum_{j=1}^{N} K_{ki}^{(1)} v_i f_{ij} u_j K_{j\ell}^{(2)}. \tag{7}$$

If the quadrature formula is sufficiently accurate, we expect the discrete solution f_{ij} of Eq. (7) to approximate the solution $f(x_i', y_j')$ of the original Eq. (2). Assuming that $\epsilon_{k\ell}$ is negligible compared to $g_{k\ell}$ and introducing the matrix notation $G = (g_{k\ell})$, $F = (f_{ij})$, $A = (K_{ki}^{(1)} v_i)$ and $B = (u_j K_{j\ell}^{(2)})$, Eq. (7) becomes formally

$$G = AFB. \tag{8}$$

Equation (8) as such incorporates the statistical and experimental errors in evaluating the elements of G and the quadrature error in the integrations leading to the new kernels A and B. Thus, from the considerations given above we expect in general that the exact solution $F = A^{-1}GB^{-1}$ of Eq. (8) will be meaningless, and hence we need additional constraints in order to get a physically meaningful solution. We propose smoothness constraints on the solution F as is indicated below. We also assume that the expected errors in G for a given problem are known and thus we require an acceptable solution F to be such that the corresponding spectrum G computed on the basis of Eq. (8) deviates from the original measured spectrum by an order of magnitude comparable with the statistical and experimental errors (see the discussion of Eqs. (10) and (14) and the examples below).

The smoothing technique is based on the physically meaningful assumption that (except possibly at a few irregular locations) the magnitude of the density function $f(x', y')$ cannot change abruptly: (a) between one point and its immediate neighbors on the grid $i = 1, \ldots, M$; $j = 1, \ldots, N$, and (b) between adjacent grids; in other

words, we require the surface $f(x', y')$ of the distribution to be sufficiently smooth. Here we seek a solution of Eqs. (6), (7), and (8) which is piecewise smooth and such that the results of its integration over x alone and over y alone as in Eq. (6) are also piecewise smooth. For example, typical smoothness constraints on the result of integrating over x alone,

$$h(x, y) = \int_a^b K^{(1)}(x, x') f(x', y)\, dx',$$

are to minimize the sum of the squares of second differences, i.e., to require each three adjacent points (say, $h_{i,j-1}$, h_{ij}, $h_{i,j+1}$) to be on or very close to a straight line. For convenience in deriving Eq. (12) below, we shall assume that the rectangular grid $\{x_i, y_i\}$ is uniform. The second difference of $h(x, y)$ for fixed x then becomes $(h_{i,j-1} - 2h_{ij} + h_{i,j+1})/(\Delta y)^2$ as is implicit in Eqs. (9) and (13). Equation (12) holds for a nonuniform grid provided that the smoothing matrices S_1 and S_2 given below are modified to properly account for the second difference, etc.

A suitable formulation of the problem is then to find the solutions $f_{k\ell}$ which minimize the function

$$\begin{aligned} H(f_{11}, f_{12}, \ldots, f_{MN}\,;\, \gamma_1, \gamma_2) &= \sum_{k,\ell} P_{k\ell}\epsilon_{k\ell}^2 + \gamma_2 \sum_{k,\ell} p_k[(AF)_{k,\ell-1} - 2(AF)_{k\ell} + (AF)_{k,\ell+1}]^2 \\ &+ \gamma_1 \sum_{k,\ell} q_\ell[(FB)_{k-1,\ell} - 2(FB)_{k\ell} + (FB)_{k+1,\ell}]^2 \\ &+ \gamma_1\gamma_2 \sum_{k,\ell} [f_{k-1,\ell-1} - 2f_{k-1,\ell} + f_{k-1,\ell+1} - 2f_{k,\ell-1} + 4f_{k\ell} - 2f_{k,\ell+1} \\ &+ f_{k+1,\ell-1} - 2f_{k+1,\ell} + f_{k+1,\ell+1}]^2 + \text{Boundary Terms.} \end{aligned} \tag{9}$$

The first term on the rightside of Eq. (9) is a measure of the error in solving Eq. (8). This term in detail [using Eq. (7)] is

$$e^2 = \sum_{k=1}^{M} \sum_{\ell=1}^{N} P_{k\ell} \left(g_{k\ell} - \sum_{i=1}^{M} \sum_{j=1}^{N} A_{ki} f_{ij} B_{j\ell} \right)^2, \tag{10}$$

where $P_{k\ell} = p_k q_\ell$ are appropriate weight factors. These weight factors need not be displayed if the matrix notation given above is modified so that $G = (P_{k\ell} g_{k\ell})$, $A = (p_k K_{k\ell}^{(1)} v_\ell)$ and $B = (u_k K_{k\ell}^{(2)} q_\ell)$. If there is no *a priori* evidence that some of the $g_{k\ell}$'s are more accurate than others, then one may set $P_{k\ell} = 1$, $p_k = 1$, $q_k = 1$

for all k,ℓ. Minimizing this measure of the error e^2, yields the least square solution[2]

$$F = [A^T A]^{-1} A^T G B^T [B B^T]^{-1}. \tag{11}$$

This solution is equivalent to the exact solution of Eq. (8), and is likewise expected to be meaningless whenever G has appreciable (statistical or experimental) error. Moreover,[3] these solutions are inappropriate if A and/or B is "ill conditioned" (nearly singular) and a better way to minimize Eq. (10) would be to apply Householder's transformations as indicated in Ref. [10].

The coefficients γ_1 and γ_2 of the smoothing terms in Eq. (9) denote optional weights that control the degree of smoothing along the x and y directions, respectively. The second and third terms on the rightside of Eq. (9) have the effect of bringing the elements of each row or column close to the elements of the two neighboring rows or columns. The fourth term corrects for expected random fluctuations at each point of the solution of the density function by weighting each net of nine neighboring points as indicated in Fig. 1.

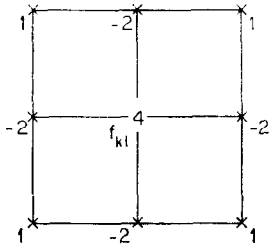


FIG. 1. A net of nine points and their corresponding weights. These weights appearing in the fourth term on the right side of Eq. (9) have the effect of minimizing the derivative $\partial^4 f/\partial x^2\,\partial y^2$ and reducing unnecessary fluctuations.

Finally, the last terms in Eq. (9) control the value of the density functions at the boundaries, as will be indicated below in connection with the smoothing matrices.

The algebra of minimizing H of (9) while holding γ_1 and γ_2 constant leads to the result

$$[A^T A + \gamma_1 S_1{}^T S_1]\, F [B B^T + \gamma_2 S_2 S_2{}^T] = A^T G B^T.$$

Hence, the desired solution for the density function $f(x, y)$ is

$$F = [A^T A + \gamma_1 S_1{}^T S_1]^{-1} A^T G B^T [B B^T + \gamma_2 S_2 S_2{}^T]^{-1}, \tag{12}$$

[2] The superscript T denotes transpose.

[3] The authors are obliged to the referee for pointing this out and recomending Householder's transformations.

where S_1 and S_2 are the following square matrices:

$$S_1 = \begin{bmatrix} S_{11}^1 & S_{12}^1 & & & & & & & 0 \\ 1 & -2 & 1 & & & & & & \\ & 1 & -2 & 1 & & & & & \\ & & \cdot & \vdots & \vdots & & & & \\ & & & \vdots & \vdots & \cdot & & & \\ & & & & & & 1 & -2 & 1 \\ 0 & & & & & & & S_{M,M-1}^1 & S_{MM}^1 \end{bmatrix},$$

(13a,b)

$$S_2 = \begin{bmatrix} S_{11}^2 & 1 & & & & & 0 \\ S_{21}^2 & -2 & 1 & & & & \\ & 1 & -2 & & & & \\ & & 1 & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & 1 & \\ & & & & & -2 & S_{N-1,N}^2 \\ 0 & & & & & 1 & S_{NN}^2 \end{bmatrix}$$

The matrices S_1 and S_2 are defined as the smoothing matrices for the x-variables and the y-variables respectively. The matrices A and/or B corresponding to a physical problem can be "ill-conditioned" (nearly singular) and thus in the absence of smoothing ($\gamma_1 = \gamma_2 = 0$) or for very weak smoothing (γ_1 and/or γ_2 very small) one expects the numerical results based on Eq. (12) to be rather sensitive to the choice of the matrix inversion method (see footnote 3). However, one advantage of smoothing is that in general the matrices $[A^TA + \gamma_1 S_1{}^T S_1]$ and $[BB^T + \gamma_2 S_2 S_2{}^T]$ are not ill-conditioned and furthermore are real, symmetric and positive definite. This enables us to utilize the well-developed and relatively fast matrix inversion subroutines based on Gaussian elimination technique, especially designed for such matrices [9]. However, if ill-conditioning is a problem, Eq. (9) can be rewritten in a form in which the minimum is obtainable by Householder transformations, as was mentioned for Eq. (10), above.

The unspecified terms in the matrices S_1 and S_2 of Eq. (13) control the conditions of the first and last rows and the first and last columns of the solution matrix F. For example, if the first row of F is known to be zero (if the distribution $f(x, y)$ is known to be zero at the edge $x = a$), one writes $S_{12}^1 = 0$ and $S_{11}^1 = 1$, or better, $S_{11}^1 =$ large number (the dominant term for $f(x, y)$ in this case can be shown to be $f(a, y) = \text{const}/S_{11}^1$); if the first row is known to be almost the same as the second row (if for y fixed $f(x, y)$ is known to approach a constant as x approaches the

edge $x = a$) one puts $S^1_{11} = -S^1_{12} = 1$, and if no condition is to be given at the first row (at the end point $x = a$) one sets $S^1_{11} = S^1_{12} = 0$. The equivalent conditions hold for the last row and the first and last columns of F.

We have so far specified only one option of smoothing by the matrices given in Eq. (13). One may choose to smooth by minimizing the sum of the squares of the first or fourth differences or any suitable combination thereof, rather than the second [5]. In fact, *the choice of smoothing matrices S_1 and S_2 is optional and may depend on the problem considered and the amount of initial information available.* If a given row (column) of F, say the k-th row (column) is known to vanish [for example, when the range of $g(x, y)$ exceeds the actual domain of $f(x, y)$] then, as in the treatment of boundary terms above, the element $S^1_{kk}(S^2_{kk})$ of the matrix $S_1(S_2)$ is set equal to a large number, while the rest of the elements in the row and column of $S_1(S_2)$ containing the element $S^1_{kk}(S^2_{kk})$ are set equal to zero.

As mentioned above, γ_1 and γ_2 are weights that control the degree of smoothness of the solution Eq. (12). Usually it suffices to compute the rightside of Eq. (12) for a few values of γ_1 and γ_2 in order to estimate the values of these weights γ_1 and γ_2 which would lead to an acceptable solution F. The acceptable values of γ_1 and γ_2 are those which would lead to a new spectrum G' [computed from the solution F by Eq. (8)] which deviates from G by an amount comparable to the (given) expected statistical and experimental error in the measured spectrum $g(x, y)$. We have already specified one measure of the error, e^2 given by Eq. (10); another useful measure of the error is

$$\bar{\epsilon} \equiv \sum_{k=1}^{M} \sum_{\ell=1}^{N} \bar{P}_{k\ell} \mid \epsilon_{k\ell} \mid /MN, \tag{14}$$

where $\mid \epsilon_{k\ell} \mid$ is the absolute value of the difference between the spectrum $g_{k\ell}$ at (x_k, y_ℓ) and its value computed on the basis of the solution Eq. (12); more precisely

$$\bar{\epsilon} \equiv \sum_{k=1}^{M} \sum_{\ell=1}^{N} \bar{P}_{k\ell} \left| g_{k\ell} - \sum_{i=1}^{M} \sum_{j=1}^{N} A_{ki} f_{ij} B_{j\ell} \right| /MN. \tag{15}$$

The factors $\bar{P}_{k\ell}$ are appropriate weights which again may be set equal to one if there is no *a priori* evidence that some of the $g_{k\ell}$'s are more accurate than others.

If necessary, the physical constraint that the density function F should be nonnegative can be imposed by an addition of a one-step iteration to the solution Eq. (12). The details are similar to what we have reported in Ref. [4] for one dimension and will be omitted here. However, from our experience in one dimension, we have reason to believe that the correct choice of γ_1 and γ_2 will invariably lead to an essentially nonnegative solution (i.e., $f \geqslant -\epsilon$, $\epsilon \to 0$) whenever such a solution is required by the physics of the problem.

III. Preliminary Applications

A computer program for the IBM 360 has been written by A. Meyer to solve Eqs. (9) and (12). The program has a large number of options and for details we refer to the write-up Ref. [6]. Among the options are various ways to specify the kernel, the smoothing matrices and the weights for quadrature and for different regions of the measured spectrum $g(x, y)$. The program allows numerical and graphical print-outs.

In order to investigate the applicability of the method described above, we have selected to treat the two mathematical models presented below. We also selected uniform mesh spacing and Simpson's rule quadrature weights for simplicity.

Example 1. *Single Peak Distribution*

In this and in the following example, simple surface distributions and separable Kernels [as in Eq. (3)] are chosen. Here we select the elliptic paraboloid density function

$$z = f(x, y) = \alpha \left\{2 - \left(\frac{x}{x_0} - 1\right)^2 - \left(\frac{y}{y_0} - 1\right)^2\right\}. \tag{16}$$

This surface is symmetric with respect to the planes $x = x_0$ and $y = y_0$. It is concave downwards and limited to the region $z \leqslant 2\alpha$. Each plane $z = \text{const}$ $(z < 2\alpha)$ perpendicular to the z-axis cuts the surface in an ellipse and each plane $x = \text{const}$ or $y = \text{const}$ cuts the surface in a parabola. We confine our attention to the region above the xy-plane with a base restricted to the interval $x \epsilon [0, 2x_0]$, $y \epsilon [0, 2y_0]$.

The kernel is selected to be the Gaussian

$$K(x, x', y, y') = a^2 \exp[-\{(x - x')^2 + (y - y')^2\}/4\sigma^2]. \tag{17}$$

The spectrum $g(x, y)$ is then given by

$$g(x, y) = \int_0^{2x_0} dx' \int_0^{2y_0} dy'\, a^2 \exp[-\{(x - x')^2 + (y - y')^2\}/4\sigma^2]$$

$$\times \alpha \left\{2 - \left(\frac{x'}{x_0} - 1\right)^2 - \left(\frac{y'}{y_0} - 1\right)^2\right\}. \tag{18}$$

Carrying out the integration, we get

$$
\begin{aligned}
g(x, y) &= \alpha a^2 \left[\sigma^2 \pi \left\{\Phi\left(\frac{2y_0 - y}{2\sigma}\right) + \Phi\left(\frac{y}{2\sigma}\right)\right\}\left\{\Phi\left(\frac{2x_0 - x}{2\sigma}\right) + \Phi\left(\frac{x}{2\sigma}\right)\right\}\right. \\
&\quad \times \left\{2 - \left(\frac{x}{x_0} - 1\right)^2 - \left(\frac{y}{y_0} - 1\right)^2 - \frac{2\sigma^2}{y_0^2} - \frac{2\sigma^2}{x_0^2}\right\} \\
&\quad + \frac{2\sigma^3 \sqrt{\pi}}{y_0^2}\left\{\Phi\left(\frac{2x_0 - x}{2\sigma}\right) + \Phi\left(\frac{x}{2\sigma}\right)\right\}\{ye^{-(2y_0 - y)^2/4\sigma^2} + (2y_0 - y)\, e^{-y^2/4\sigma^2}\} \\
&\quad \left. + \frac{2\sigma^3 \sqrt{\pi}}{x_0^2}\left\{\Phi\left(\frac{2y_0 - y}{2\sigma}\right) + \Phi\left(\frac{y}{2\sigma}\right)\right\}\{xe^{-(2x_0 - x)^2/4\sigma^2} + (2x_0 - x)\, e^{-x^2/4\sigma^2}\}\right],
\end{aligned}
\tag{19}
$$

where $\Phi(t)$ is the error function

$$
\Phi(t) = \operatorname{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-t^2}\, dt. \tag{20}
$$

For the purpose of numerical computations we select $\alpha = 1$, $a = 10$, $x_0 = 5$ and $y_0 = 10$. We also select three values of σ, $\sigma = 0.5$, 1 and 2, and the mesh size $M \times N = 11 \times 21$.[4]

The process of solving a Fredholm equation of the first kind [Eq. (2)] by approximating it with a matrix equation [Eq. (8)] involves quadrature error which depends on the following variables: the kernel shape and size, the order of the matrices (the grid density or mesh size, $M \times N$) and the particular density function studied. For the present problem, we have computed the matrix G in two ways, (i) from Eq. (19) for the analytical representation of the spectra, and (ii) directly from the matrix Eq. (8) with the matrix F for the density function, and the response

numerical results are not published, but indicate, as we would expect, that the response kernel is better represented, and hence the quadrature error becomes smaller, as the kernel becomes wider and/or grid lines become denser (i.e., as the order of the matrices F, G, and hence the order of A and B increases).

In the present example, we will eliminate the quadrature error by starting with the spectra G computed on the basis of the matrix Eq. (8) rather than the integral (Fredholm) Eq. (2). Depending on the shape of the response kernel K, the matrices

[4] Results of additional computations will be presented in an Argonne National Laboratory Report.

A^TA and BB^T may have an eigenvalue very close to zero and hence may be almost singular (or singular due to machine limits on the accuracy of matrix computations).

We have applied the method described above to the spectra G of size $M \times N = 11 \times 21$. Table I summarizes the main results when $G = AFB$ is used and also when an error is introduced in G by reducing its elements to 4 S.F. (significant figures), 3 S.F., and 2 S.F. The results for γ_1 and $\gamma_2 = 0$ represent the usual (unsmoothed) solutions; other values of γ_1 and γ_2 represent different degrees of

TABLE I

Evaluation of the Computed Density Matrix F for the Single Peak Distribution relative to its Exact Value; Summary of Numerical Results for $M \times N = 11 \times 21$

	Narrow Kernel $\sigma = 0.5$ $G = xx. - xxx.$[a]			Medium Kernel $\sigma = 1.0$ $G = xxx. - xxxx.$[a]			Wide Kernel $\sigma = 2.0$ $G = xxxx.$[a]		
	$\gamma_1 = \gamma_2$	$\bar{\epsilon}$[b]	Computed F[c]	$\gamma_1 = \gamma_2$	$\bar{\epsilon}$	Computed F	$\gamma_1 = \gamma_2$	$\bar{\epsilon}$	Computed F
$G = AFB$	0.0		A^+	0.0		D	0.0	No Solution	
	10^{-4}	1.3×10^{-5}	A^+	10^{-4}	5.5×10^{-4}	A	10^{-4}	5.6×10^{-3}	A
	10^{-2}	1.3×10^{-3}	A	10^{-2}	1.5×10^{-2}	B	10^{-3}	8.9×10^{-3}	B
							10^{-2}	2.4×10^{-2}	B
Comments	Problem correctly posed			A, B ill conditioned			A^TA and BB^T singular		
G	0.0		A	0.0		E	0.0	No Solution	
(Reduced	10^{-4}	1.3×10^{-5}	A	0.01	2.4×10^{-2}	B	0.1	0.20	C
to 4 S.F.	10^{-2}	1.3×10^{-3}	B	0.1	7.3×10^{-2}	A	0.3	0.41	B
				0.5	0.23	B	0.5	0.51	C
G	0.0		B	0.0		E	0.0	No solution	
(Reduced	0.1	1.2×10^{-2}	B	0.1	0.21	C	0.1	0.27	C
to 3 S.F.)	0.5	4.0×10^{-2}	A	0.5	0.29	B	0.5	0.54	B
	1.0	6.2×10^{-2}	B	1.0	0.35	C	0.7	0.60	C
G	0.0		C	0.0		E	0.0	No solution	
(Reduced	0.5	8.9×10^{-2}	C	5.0	2.1	C	1.0	2.6	C
to 2 S.F.)	1.0	0.13	B	10.0	2.3	B	5.0	2.6	B
	5.0	0.28	C	20.0	2.5	C	10.0	2.6	C

[a] Each x stands for a digit between 0 and 9 in the position it occupies relative to the decimal point indicated—$xx.$, $xxx.$, give the order of magnitude of the value under consideration.

[b] $\bar{\epsilon}$ stands for the average error in G corresponding to the computed F.

[c] The ratings indicate that the agreement of the computed F with its known value over most of the domain $[a, b] \times [c, d]$ is accurate at least to within $= A = 0.01\,\%$, $B = 0.1\,\%$, $C = 1\,\%$, $D = 20\,\%$ and E indicates no agreement (see Figs. 2–4).

smoothing. A sample of the graphical results is given in Figs. 2–4. These figures show contours for the spectrum G and the corresponding exact density matrix and/or solution matrix F.

We observe that for the narrow response kernel ($\sigma = 0.5$) the matrices A^TA and BB^T are positive definite and well-conditioned (possess well-defined inverses). As the response kernel becomes wider, these matrices become more and more ill-conditioned (their smallest eigenvalues become closer and closer to zero). In fact, for $\sigma = 2.0$ and $M \times N = 11 \times 21$, the computer program indicated that these matrices are singular to within double precision (16 digit) accuracy. We were able to satisfy our curiosity, however, and get inverses for A^TA and BB^T by rescaling these matrices. The details will be omitted, but the corresponding solutions of the matrix equation $G = AFB$ were highly oscillatory and physically meaningless.

The above justifies the results given in Table I and shown in Figs. 2–4. For $\sigma = 0.5$, $M \times N = 11 \times 21$, although the shape of the spectrum is distorted due to quadrature error in evaluating G by AF(exact)B [Eq. (8) instead of Eq. (19)]; the solution given by Eq. (11) is stable as shown in Fig. 2. In this case smoothing becomes necessary only if very large errors are introduced in the spectra. Smoothing

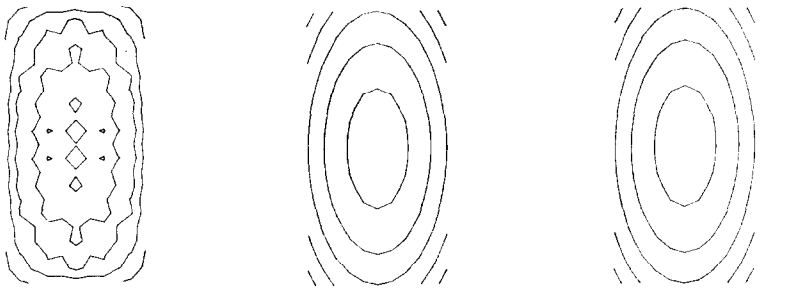


FIG. 2. Narrow Kernel: $\sigma = 0.5$, $M \times N = 11 \times 21$.

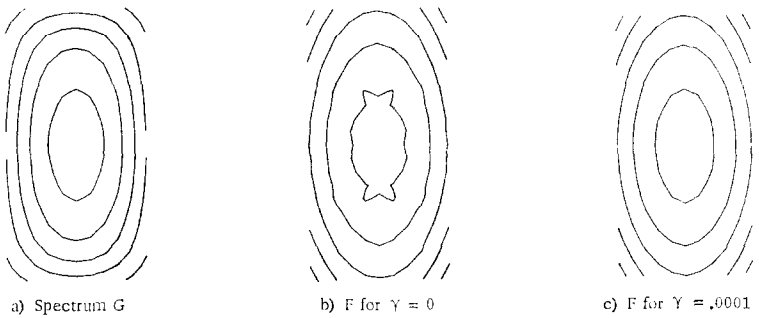


FIG. 3. Average Kernel: $\sigma = 1.0$, $M \times N = 11 \times 21$.

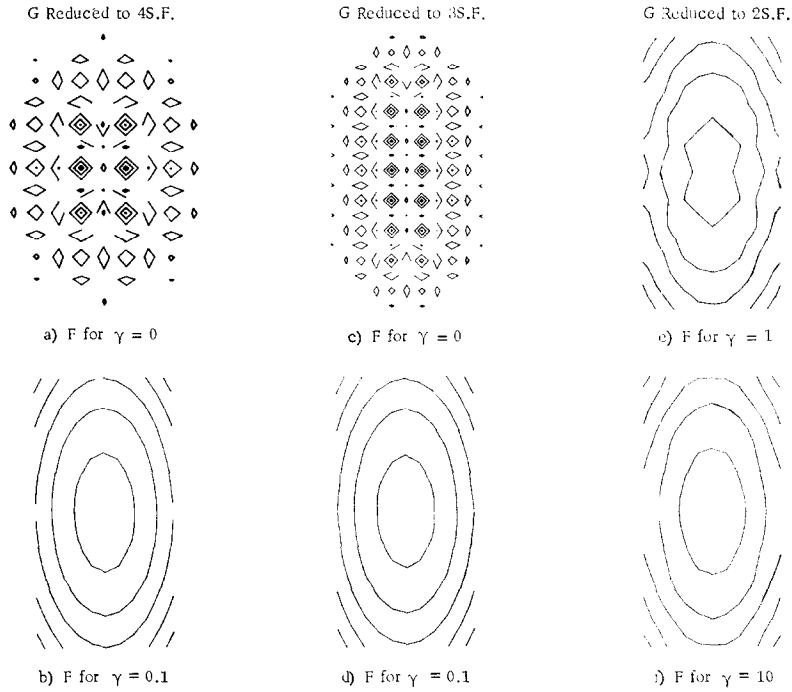


FIG. 4. Effect of errors introduced in the spectrum G of Fig. 3a.

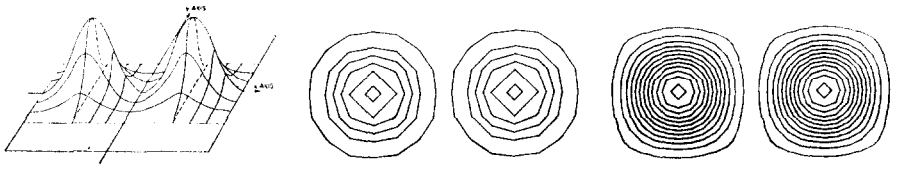
is necessary for $\sigma = 1.0$ and $\sigma = 2.0$. We note that the values of $\gamma_1 = \gamma_2$ (for an explanation of how to choose γ_1 and γ_2, see the main text above and Example 2 below) and hence the degree of smoothness needed to yield an optimum solution for the density matrix F increases as the error in G increases and, in general, as the response matrices become more and more ill-conditioned.

Figures 3–4 illustrate the situation for $\sigma = 1.0$. Figure 3 shows that the unsmoothed solution ($\gamma_1 = \gamma_2 = 0$) is only slightly distorted and the distortion is removed by smoothing with $\gamma_1 = \gamma_2 = 0.0001$. Figure 4a shows that an error of less than .1 % introduced by reducing G to 4 S.F. (retaining only the largest 4 S.F. for each value of G) leads to an unacceptable solution. The desirable solution of Fig. 4b is achieved by smoothing with $\gamma_1 = \gamma_2 = 0.1$. The rest of Fig. 4 illustrates the results when G is reduced to 3 S.F. and to 2 S.F.

EXAMPLE 2. *Double Peak Distribution.*

We consider here the following analogue of Example 1 of Phillips, Ref. [8, p. 87] given in Fig. 5:

$$f(x, y) = f_1(x) f_2(y), \tag{21a}$$



a) Actual Shape b) Contours for M x N = 13 x 13 c) Contours for M x N = 25 x 25

FIG. 5. Exact density function.

where

$$f_1(x) = \left(1 - \cos\frac{\pi x}{3}\right) \qquad \text{for} \quad x \in [-6, 0]$$
$$= A\left(1 - \cos\frac{\pi x}{3}\right) \qquad \text{for} \quad x \in [0, 6], \tag{21b}$$

and

$$f_2(y) = B\left(1 + \cos\frac{\pi y}{3}\right) \qquad \text{for} \quad y \in [-3, 3]$$
$$= 0 \qquad \text{for} \quad |y| \geqslant 3. \tag{21c}$$

Note that the parameter B controls the height of the peaks shown in Fig. 5a while the parameter A allows the right hand peak ($x \geqslant 0$) to be equal to, smaller than or larger than the left hand ($x \leqslant 0$) peak. In addition, note that the region considered is $x\epsilon[-6,6]$, $y\epsilon[-6,6]$ while the positive part of the density of interest is restricted to only half of this region. The density function vanishes for $y \geqslant 3$.

The kernel for this problem has the same shape as one of the peaks in Fig. 5a. It is given by

$$K(x, x', y, y') = K(|x - x'|)\, K(|y - y'|), \tag{22a}$$

where

$$K(t) = 1 + \cos\frac{\pi t}{3} \qquad \text{for} \quad |t| \leqslant 3$$
$$= 0 \qquad \text{for} \quad |t| \geqslant 3. \tag{22b}$$

The spectrum $g(x, y)$,

$$g(x, y) = \int_{-6}^{6} dx' \int_{-6}^{6} dy'\, K(|x - x'|)\, K(|y - y'|) f(x', y')$$

is then given by

$$g(x, y) = B g_1(x)\, g_2(y), \tag{23}$$

where as in Ref. [8] (omitting algebraic details),

$$g_2(y) = (6 - |y|)\left(1 + \frac{1}{2}\cos\frac{\pi|y|}{3}\right) + \frac{9}{2\pi}\sin\frac{\pi|y|}{3}, \qquad y \in [-6, 6] \tag{24}$$

and $g_1(x)$ is given in terms of g_2 as

$$\begin{aligned} g_1 &= g_2(x+3), && \text{for } x \in [-6, -3] \\ &= g_2(x+3) + Ag_2(x-3), && \text{for } x \in [-3, 3] \\ &= Ag_2(x-3), && \text{for } x \in [3, 6]. \end{aligned} \tag{25}$$

For numerical computation we select $A = B = 1$ (the two peaks are of the same height for the results reported here). Furthermore, we compute the elements $G_{ij} = g(x_i, y_j)$ from the analytic expressions of Eqs. (23)–(25) and thus, contrary to what was done in Example 1, we do not eliminate the quadrature error here. Table II summarizes the computational and graphical results corresponding to the grid density $M \times N = 13 \times 13$. For this case, the quadrature error is bounded by 0.4, and seems to dominate the errors artificially introduced by reducing g to 4 S.F., 3 S.F., and 2 S.F. We observe from Table II[5] that, throughout the computations, the optimum results for the density matrix F were those for which the average error $\bar{\epsilon}$ is in the expected range $\bar{\epsilon} = 0.2$–0.4. Table III summarizes the computational and graphical results which correspond to the grid density $M \times N = 25 \times 25$. Here, the quadrature error is bounded by $(0.0x)$ and the average quadrature error is 0.008 ± 0.001. We observe from Table III that the best choices of γ corresponding to $G = AFB$ and $G = AFB$ reduced to 4 S.F., are those for which $\bar{\epsilon}$ is in the range 0.008 ± 0.001 of the average quadrature error. For $G =$ (AFB reduced to 3 S.F.) and (AFB reduced to 2 S.F.), Table III indicates that the best choices of $\gamma_1 = \gamma_2$ are those which lead to $\bar{\epsilon}$ in the range of the error introduced in G.

The symbols $SM1$ to $SM5$ in Tables II and III correspond to smoothing with the following types of boundary conditions (see Fig. 5 for the exact density matrix F):

SM1. No boundary conditions imposed on the solution matrix F.

SM2. The first and last columns and rows of F are required to be relatively small

corresponds to the fact that the actual distribution (see Fig. 5) vanishes at the extreme boundaries.

SM4. The solution is required to vanish outside the actual nonnegative part of the distribution (outside the range of the two peaks, see Fig. 5).

SM5. The same conditions as for SM4 plus the requirement that values of the solution corresponding to the dividing line between the two peaks (see Fig. 5) should vanish.

[5] A slight difference in the optimum choice of γ often still leads to reasonably accurate solutions.

TABLE II[a]

Evaluation of the Computed Density Matrix F for the Double Peak Distribution relative to its Exact Value; Summary of numerical Results for $M \times N = 13 \times 13$

	$G: G_{ij} \leqslant xx.$[b] $\epsilon_{ij} \leqslant .x$			G (reduced to 4 S.F.)			G (reduced to 3 S.F.)			G (reduced to 2 S.F.)		
	$\gamma_1 = \gamma_2$	$\bar{\epsilon}$	Computed F	$\gamma_1 = \gamma_2$	$\bar{\epsilon}$	Computed F	$\gamma_1 = \gamma_2$	$\bar{\epsilon}$	Computed F	$\gamma_1 = \gamma_2$	$\bar{\epsilon}$	Computed F
	0.0		D	0.0		D	0.0		E	0.0		E
SM1	.02	.19	C	.02	.19	C	.02	.19	C	.02	.24	D
	.03	.21	B	.03	.21	B	.03	.22	B	.04	.28	C
	.04	.24	C	.04	.24	C	.04	.24	C	.06	.32	D
SM2	.02	.19	C	.02	.19	C	.02	.19	C	.02	.24	D
	.03	.21	B	.03	.21	B	.03	.22	C	.04	.29	C
	.04	.24	C	.04	.24	C	.04	.24	B	.06	.32	D
	Slightly better than SM1											
SM3	.03	.21	B	.03	.21	B	.04	.24	B	.04	.30	C
	.04	.24	A	.04	.24	A	.06	.28	A	.06	.33	C
	.05	.26	B	.05	.26	B	.08	.31	B	.08	.36	B
	Noticeably better than SM2											
SM4	.04	.27	A	.04	.27	A	.04	.28	A	.06	.36	C
	.05	.29	A	.05	.29	A	.06	.31	B	.08	.39	B
	.06	.31	B	.06	.31	B	.08	.35	B	.10	.42	C
	Slightly better than SM3											
SM5	.04	.27	B	.04	.27	A	.04	.28	A	.06	.36	C
	.05	.29	A	.05	.29	A	.06	.32	B	.08	.40	B
	.06	.31	B	.06	.31	B	.08	.35	B	.10	.43	C
	Hardly better than SM4											

[a] See footnotes to Table I.

[b] The magnitude of each element of G is $(G)_{ij} \leqslant 81$; ϵ_{ij} stands for the quadrature error.

A sample of the results summarized in Tables II and III are presented in Figs. 6–8. Figs. 6a, b show the solution, Eq. (12) for $M \times N = 13 \times 13$, in the absence of smoothing. These figures complement each other in the sense that while Fig. 6a shows the contours corresponding to the solution, Fig. 6b (which shows the cross-sections of the solution corresponding to one of the two peaks of the density function) indicates the relative magnitudes of the solution and the known exact value of the density function. In particular, Fig. 6b indicates that the third and

TABLE III[a]

Evaluation of the Computed Density Matrix F for the Double Peak Distribution relative to its Exact Value; Summary of Numerical Results for $M \times N = 25 \times 25$

	$G: G_{ij} \leqslant xx.$[b] $\epsilon_{ij} \leqslant .0x$ $\langle \epsilon_{ij} \rangle = .008 \pm .001$			G (reduced to 4 S.F.) error introduced $\leqslant .00x$			G (reduced to 3 S.F.) error introduced $\leqslant .0x$			G (reduced to 2 S.F.) error introduced $\leqslant 0.x$		
	$\gamma_1 = \gamma_2$	$\bar{\epsilon}$	Computed F	$\gamma_1 = \gamma_2$	$\bar{\epsilon}$	Computed F	$\gamma_1 = \gamma_2$	$\bar{\epsilon}$	Computed F	$\gamma_1 = \gamma_2$	$\bar{\epsilon}$	Computed F
	0.0		D	0.0		E	0.0		E	0.0		E
SM1	.002	8.2×10^{-3}	C	.001	7.8×10^{-3}	C	.04	.025	C	0.3	.13	D
	.003	8.8×10^{-3}	C	.002	8.4×10^{-3}	B	.06	.027	B	0.4	.14	C
	.004	9.4×10^{-3}	B	.003	9.0×10^{-3}	C	.08	.029	C	0.5	.15	D
SM2	.002	7.7×10^{-3}	C	.001	7.6×10^{-3}	C	.04	.021	C	0.3	.13	D
	.003	8.2×10^{-3}	B	.002	8.0×10^{-3}	B	.06	.023	B	0.4	.14	C
	.004	8.5×10^{-3}	C	.003	8.4×10^{-3}	C	.08	.026	C	0.5	.15	D
					Slightly better than SM1							
SM3	.001	7.3×10^{-3}	A	.001	7.6×10^{-3}	B	.02	.018	B	0.1	.12	C
	.002	7.5×10^{-3}	B	.002	7.7×10^{-3}	A	.04	.019	A	0.2	.13	B
	.003	7.7×10^{-3}	C	.003	7.9×10^{-3}	B	.06	.022	B	0.4	.15	C
					Noticeably better than SM2							
SM4	.0005	7.5×10^{-3}	B	.003	8.1×10^{-3}	B	.02	.020	A	0.1	.11	C
	.001	7.6×10^{-3}	A	.004	8.2×10^{-3}	A	.04	.024	B	0.2	.12	B
	.00	7.7×10^{-3}	B	.005	8.2×10^{-3}	B	.06	.027	B	0.4	.13	C
					Slightly better than SM3							
SM5	.0005	7.5×10^{-3}	A	.002	8.1×10^{-2}	B	.02	.021	A	0.1	.12	C
	.001	7.6×10^{-3}	B	.003	8.2×10^{-2}	A	.04	.024	B	0.2	.12	B
	.00	7.7×10^{-3}	B	.004	8.3×10^{-2}	B	.06	.027	B	0.4		C
					Hardly better than SM4							

[a] See footnotes to Tables I and II.

[b] The magnitude of each element of G is ($G_{ij} \leqslant 81$; ϵ_{ij} is the quadrature error and $\langle \epsilon_{ij} \rangle$ is the average quadrature error.

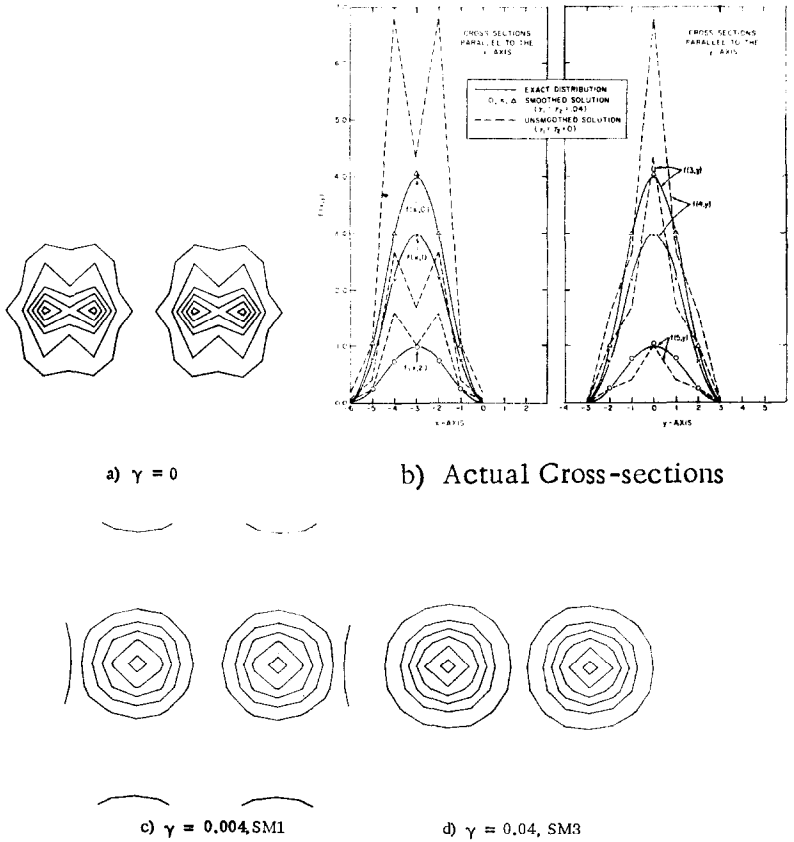


FIG. 6. Solution F corresponding to G computed from Eqs. (23)–(25) for $M \times N = 13 \times 13$.

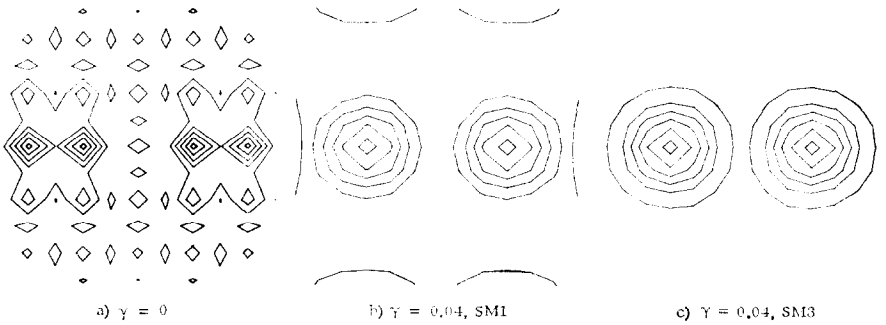
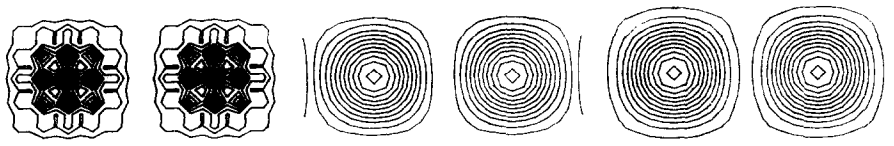


FIG. 7. Solution F corresponding to G reduced to 4 *S.F.*, $M \times N = 13 \times 13$.



a) $\gamma = 0$ b) $\gamma = 0.004$, $SM1$ c) $\gamma = 0.001$, $SM3$

FIG. 8. Solution F corresponding to G computed from Eqs. (23)–(25), for $M \times N = 25 \times 25$.

fourth contours circumscribing each part of Fig. 6a, correspond to values of density higher than both the innermost and outermost contours. While the solution given in Fig. 6a is unacceptable, the smoothed solutions given in Figs. 6b, c, d are very close to the exact density function. Figure 6c corresponds to $SM1$ and shows that a slight error is introduced when the boundary conditions are not known. Fig. 6d corresponds to $SM3$ with known boundary conditions (see above). Thus, as we would expect, the computational and graphical results show clearly that advance knowledge of useful information about a distribution may be utilized to improve the solution.

Fig. 7a, b, c corresponds respectively to Fig. 6a, c, d when G is reduced to 4 S.F. Fig. 7a indicates the wide error introduced in the unsmoothed solution when an error of less than 0.1 % is introduced in the spectrum. Finally, Fig. 8a, b, c corresponds to Fig. 6a, c, d when the mesh size is 25×25 instead of 13×13. Here again the unsmooth solution, though its contours are very appealing to the eye, is an unacceptable representation of the actual density; the smoothed solutions Fig. 8b, c demonstrate the effectiveness of the method given above for calculating a density function from its measured spectra.

ACKNOWLEDGMENTS

The authors are grateful to D. Phillips and G. Birkhoff for many critical comments, to R. Wessel especially for supplying us with "contour plot" subroutines and other help and to A. B. Meyer for her efficient programming. They also thank E. H. Bareiss, R. Buchal, V. Klema, J. H. Wilkinson, and G. Zgrablich for their help and encouragement.

REFERENCES

1. G. F. CLEMENTE, L. D. MARINELLI, AND I. K. ABU-SHUMAYS, "Regularization Unfolding in Low γ-Activity Measurements: I. Evaluation of One-Dimensional Scanning," Argonne National Laboratory Report ANL-7615, (1969), 87–95.
2. R. E. LYNCH AND J. R. RICE, Convergence of ADI Methods with smooth initial error, *Math. Comp.* **22** (1968), 311–335.
3. M. S. LYNN AND W. P. TIMLAKE, The numerical solution of singular integral equations in potential theory, *Numer. Math.* **11** (1968), 77–98.

4. L. D. MARINELLI, G. F. CLEMENTE, I. K. ABU-SHUMAYS, AND O. J. STEINGRABER, "Localization of Radioactivity *In Vivo* by Photon TIME-OF-FLIGHT Techniques," Argonne National Laboratory Report ANL-7489, (1968) 1–12, See also *Radiology* **92** (1969), 167.
5. N. H. MARSHALL, "Numerical Techniques for Solving Fredholm's Integral Equations of the First Kind," *U. S. At. Energy Comm.* Report IDO-17175, May, (1966).
6. A. B. MEYER AND I. K. ABU-SHUMAYS, "A Special Regularization-Unfolding Technique for the Numerical Solution of Two-Dimensional Fredholm Equations of the First Kind," Applied Mathematics Division Program A1T8286 (IBM360/50/75), Argonne National Laboratory, 1969.
7. V. A. MOROZOV, Regularization of incorrectly posed problems and the choice of regularization parameters, *U.S.S.R. Comput. Math. Math. Phys.* **6** (1966), 242–251.
8. D. L. PHILLIPS, A Technique for the numerical solution of certain integral equations of the first kind. *J. Ass. Comput. Mach.* **9** (1962), 84–97.
9. B. S. GARBOW, "ANL F152S, Symmetric Matrix Inversion with Accompanying Solution of Linear Equations," Argonne National Laboratory AMD System/360 Library Subroutine, 1968.
10. G. H. GOLUB, Numerical methods for solving linear least squares problems, *Numer. Math.* **7** (1965), 206–216.