Comput. & Elect. Engng, Vol. 2, pp. 95–103. Pergamon Press, 1975. Printed in Great Britain.

AN ANALOG COMPUTER METHOD TO SOLVE FREDHOLM INTEGRAL EQUATIONS OF THE FIRST KIND

V. VEMURI

School of Advanced Technology, State University of New York at Binghamton, Binghamton, New York 13901, U.S.A.

(*Received* 29 *April* 1974)

Abstract—Many problems of contemporary interest are characterized by Fredholm type integral equations of the first kind. These equations are inherently ill-posed and difficult to solve. It is customary to convert the equation into a set of m algebraic equations $Af = g$ in n unknowns with m not necessarily equal to n. Then one can solve these m equations in a least square sense. Among the class of vectors f that minimize the Euclidean norm of the error, there exists a unique vector A^+g which is of least norm where A^+ is the generalized inverse of A. One method of finding the generalized inverse of A is to reformulate the problem into an equivalent system of first order ordinary differential equations with specified initial conditions. The steady state solution of this system is A^+g, the required value of f. This procedure was implemented on an analog computer and the results presented.

INTRODUCTION

The Fredholm integral equation of the first kind

$$\int_a^b K(x, y) f(x)\mathrm{d}x = g(y); a \leq y \leq b \tag{1}$$

occurs frequently in many branches of physical and biological sciences. The hapten binding equation of immunology[1, 2], for example, is an equation of this kind. The problem of locating tumors in a body using radiographic techniques[3] can also be formulated as an integral equation of the first kind. The problem of deducing the structure of a planetary atmosphere from satellite observations[4, 5] can be reduced to a mathematical problem similar to equation (1). Indeed equation (1), and variations thereof play a central role in many experimental sciences whenever physical data are gathered by indirect sensing devices as in the case with many types of remote sensing experiments[6–8]. The convolution type integral equation

$$\int_{-\infty}^{\infty} h(t - \tau) f(\tau)\mathrm{d}\tau = g(t) \tag{2}$$

which is a special case of equation (1) plays a central role in many problems of electrical engineering.

Usually, the situation can be briefly described as follows:

(a) A function $g(y)$, which is usually obtained experimentally, is given.
(b) The kernel $K(x, y)$, which is a function of two variables, is usually known.
(c) We wish to find or compute a function $f(x)$ which satisfies equation (1) or, equivalently, equation (2).

For certain types of kernels, equation (1) is ill-posed (not ill-formulated) from analytical as well as numerical points of view. That is to say, even if the integral equation is derived from sound physical principles, the problem is complicated primarily because the solution $f(x)$ is extremely sensitive to the presence of noise such as measurement errors or rounding errors. Nevertheless the practical importance of this problem prompted many to try various approaches in order to obtain approximate solutions which are acceptable from a physical point of view. For instance, Bellman *et al.*[9] treated this problem using dynamic programming in conjunction with a successive approximation technique. While agreeing that no one technique can satisfactorily resolve the fundamental problem of obtaining sensible solutions from ill-conditioned systems, the authors strived to point out that a reasonable compromise is to combine various techniques such as dynamic programming, successive approximations, extrapolation and smoothing. In a discussion of the difficulties involved in solving equation (1), Phillips[10] demonstrated that the existence of errors converts the problem from one with a unique solution to one with infinitely many solutions. To pick one out of many solutions, it is necessary to impose constraints on the problem. One suggested constraint has been to pick a solution that exhibits a minimum second difference over the family of solutions. Following this lead, Twomey[11] formulated the problem as a constrained optimization problem using Lagrange multipliers. One common feature of both the above methods is that they require a matrix inversion which is generally a time consuming operation. To overcome the storage and time restrictions, Hunt[12] used fast Fourier transforms (FFTs) to solve the convolution type integral equation. As pointed out by Hunt himself, the transform method lacks the generality of the method suggested by Phillips and Twomey.

THE ANALOG COMPUTER PROCEDURE

The purpose of this paper is to present an analog computer method, or an initial-value method, of solving Fredholm integral equations of the first kind.

A first step in the new procedure is to replace, as is usually done in conventional methods, the continuous variable y with a finite set of mesh points $y_1, y_2, \ldots, y_n$,

$$a \leq y_1 < y_2 < \ldots < y_n \leq b$$

and to write

$$\begin{aligned} g(y_1) &= \int_a^b K(x, y_1) f(x) \mathrm{d}x \\ &\vdots \\ g(y_m) &= \int_a^b K(x, y_m) f(x) \mathrm{d}x. \end{aligned} \tag{3}$$

In experimental sciences $g(y_i) \triangleq g_i$, $i = 1, 2, \ldots, m$, generally correspond to m experimentally observed data points.

The next step of the discretization process is to replace the continuous variable x by a finite set of mesh points $x_1, x_2, \ldots, x_n$ such that

$$a \leq x_1 < x_2 < \ldots < x_n \leq b.$$

To carry out this step, each of the integrals in equation (3) is replaced by a suitably chosen

numerical quadrature formula to yield

$$\begin{aligned} g(y_1) &\approx \sum_{j=1}^{n} w_j K(x_j, y_1) f(x_j) \\ &\vdots \qquad\qquad \vdots \\ g(y_m) &\approx \sum_{j=1}^{n} w_j K(x_j, y_m) f(x_j) \end{aligned} \tag{4}$$

where $w_1, w_2, \ldots, w_n$ are the weights associated with the quadrature formula. Equation (4) in vector-matrix notation is

$$g \approx Af \tag{5}$$

where $g = [g(y_1), g(y_2), \ldots, g(y_m)]^T$, $f = [f(x_1), f(x_2), \ldots, f(x_n)]^T$ in which the superscript T denotes the transpose and A is an $(m \times n)$ matrix given by

$$A = \begin{bmatrix} w_1 K(x_1, y_1) & w_2 K(x_2, y_1) \cdots w_n K(x_n, y_1) \\ w_1 K(x_1, y_2) & w_2 K(x_2, y_2) \cdots w_n K(x_n, y_2) \\ \vdots & \\ w_1 K(x_1, y_m) & w_2 K(x_2, y_m) \cdots w_n K(x_n, y_m) \end{bmatrix}.$$

Solution of equation (5) is beset with difficulties primarily because of the poor conditioning of matrix A. While equation (1) is ill-posed, the matrix A in equation (5) is ill-conditioned. Equation (1) is sensitive to measurement errors and sampling errors. If one attempts to choose m and n larger and larger in an effort to avoid discretization or sampling errors, the condition of A will correspondingly deteriorate. It can be shown that as $m \to \infty$ and $n \to \infty$, then A indeed approaches the Hilbert matrix, which is often cited as a classical example of an ill-conditioned matrix. That is to say that large n tends to make A more and more ill-conditioned. Making n too small is not desirable either as it tends to make the approximation very crude.

While solving equation (5), the system may exhibit one of three characteristics: (a) the system may be infeasible (i.e. no solution may exist), (b) the system may have an infinite number of solutions, or (c) the system may have a unique solution. Regardless of the nature of the system, a reasonable approach is to minimize the error $(Af - g)$ in a least square sense. Among the class of vectors f that minimize the Euclidean norm of the error, there exists a unique vector $A^+ g$ which is of least norm where A^+ is the generalized inverse of A[13, 14]. Thus the problem of solving the integral equation (1) is equivalent to solving equation (5), which in turn is tantamount to finding the generalized inverse of A. This observation is the point of departure between the conventional methods and the new method.

There exist many elegant iterative methods of determining the generalized inverse of an arbitrary rectangular matrix[15]. However, for an initial-value formulation, the task of finding A^+ is made possible by the following two theorems[16, 17].

Theorem 1

For any arbitrary, real $(m \times n)$ matrix A the generalized inverse is given by

$$A^+ = \lim_{\lambda \to 0} [(\lambda I + A^T A)^{-1} A^T]. \tag{6}$$

Proof of this theorem can be found elsewhere[17].

Theorem 2

For any real, finite $(m \times n)$ matrix A and column vector g, the system

$$\frac{df(t)}{dt} = -A^T A f(t) + A^T g$$
$$f(0) = 0 \tag{7}$$

has a unique asymptotic solution given by

$$\lim_{t \to \infty} f(t) = A^+ g. \tag{8}$$

Theorem 2 can be readily proved by taking Laplace transforms on both sides of equation (7) and applying the Final Value Theorem:

$$F(s) = \frac{1}{s}(sI + A^T A)^{-1} A^T g \tag{9}$$

$$\lim_{t \to \infty} f(t) = \lim_{s \to 0} sF(s) = \lim_{s \to 0} (sI + A^T A)^{-1} A^T g. \tag{10}$$

Use of equations (10) and (6) give the desired result in equation (8).

Thus, if equation (7) is programmed, say, on an analog computer, and if the computer is allowed to reach steady state (i.e. until the integrator outputs become sufficiently constant), the value of $f(t)$ obtained at the output of each integrator constitutes a solution of the given integral equation.

The result shown in equation (10) can also be obtained by a different route. Note that $A^T A$ is at best positive semidefinite. Therefore an unforced version of equation (7) is not necessarily stable. The forced system, however, is stable due to the form of the forcing function $A^T g$. Since $A^T g$ cannot be programmed exactly anyway (due to several factors, not excluding the measurement errors in g), equation (7) can be slightly modified to ensure stability.

$$\frac{df}{dt} = -(sI + A^T A) + A^T g$$
$$f(0) = 0 \tag{11}$$

where s is a small positive scalar. If s is small compared to the nonzero eigenvalues of $A^T A$, then the error introduced is small. The limiting process indicated by the right side of equation (10) also suggests that s be small. In practice, a nominally small value of s can be chosen first and as the solution reaches its asymptotic value, the value of s can be reduced in stages until no further improvement in the solution is apparent.

EXPERIMENTAL RESULTS

Example 1

Consider

$$\int_0^1 (x^2 + y^2)^{1/2} f(x) dx = \tfrac{1}{3}[(1 + y^2)^{3/2} - y^3]; 0 \le y \le 1. \tag{12}$$

This problem was first discussed by Fox and Goodwin[18]. Solution of this equation is readily seen to be $f(x) = x$ by direct integration of the left hand side. Fox and Goodwin discretized this problem into an $n \times n$ (for two cases with $n = 5$ and 9) linear system

$$Ax = b$$

using both Simpson's and Gauss's quadrature formulas. The results obtained from the more powerful Gaussian quadrature are shown in Fig. 1. Note that the solution corresponding to the smaller discretization interval (i.e. $n = 9$) tends to oscillate more vigorously supporting the earlier contention that larger n tends to deteriorate the condition of A.

This problem is solved once again by implementing equation (7) with $n = 5$ in conjunction with Simpson's quadrature and with $n = 6$ in conjunction with Newton–Cotes formula. Some analog integrator outputs are displayed in Fig. 2. These results are replotted, $f(x)$ vs x, in Fig. 3. As can be seen from Fig. 2, some of the analog integrator outputs displayed an apparent steady state for a while and began to diverge slowly. This behavior is due to terms like $c_i \exp(+\alpha_i t)$, $\alpha_i > 0$ in the solution. These terms are due to small errors such as truncation errors and errors arising out of approximating the kernel $K(x, y)$.

This problem is solved for the third time by implementing equation (11) with $n = 11$ and for various values of s. One particular result is shown in Fig. 4.

Example 2

As a second example, consider

$$\int_{0\cdot5}^{5\cdot5} e^{-(x-y)^2}(\tfrac{1}{3}x^3 - 3x^2 + 8x + 1)dx = g(y); 3 \le y \le 7. \tag{13}$$

This problem is not given in the same form as the previous one. If one performs the integration, one readily gets

$$g(y) = P_0(y)e^{-(5\cdot5-y)^2} + P_1(y)e^{-(0\cdot5-y)^2} + P_2(y)[I(5\cdot5 - y) - I(0\cdot5 - y)] \tag{14}$$

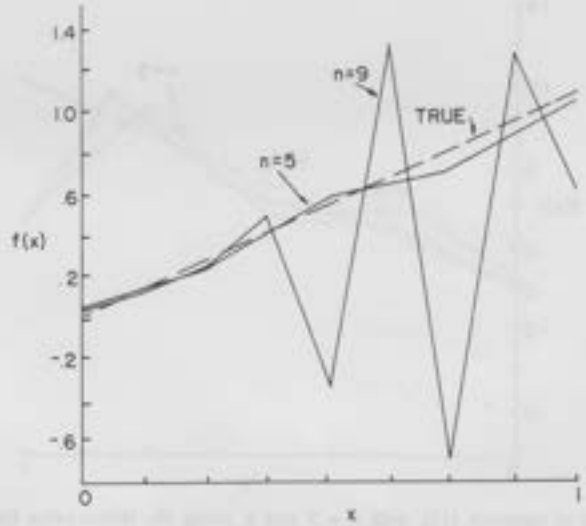

Fig. 1. Solution of equation (12), for $n = 5$ and 9, using Gaussian elimination.

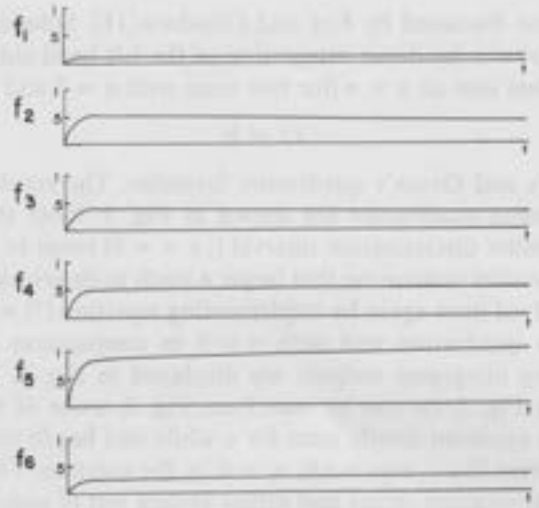

Fig. 2. Typical outputs of some integrators while solving equation (12), with $n = 6$, on an analog computer using initial-value formulation. Each integrator output at steady state corresponds to $f_i = \sum_j a_{ij}^* g_j$ where f_i, a_{ij} and g_j are defined in equation (5).

where $I(x)$, the error function, is defined by

$$I(x) = \int_0^x e^{-u^2}\, du$$

and P_0, P_1 and P_2 are certain polynomials in y. The values of $g(y)$ in equation (14) are first evaluated for various values of y. This information is considered as measurement data on equation (13) and $f(x)$ is recovered.

This problem is solved with $m = 10$, $n = 11$ and with $s = 0$ and $s = 0{\cdot}05$. The exact solution $f(x)$ and the computed solutions are shown in Figs. 5–7.

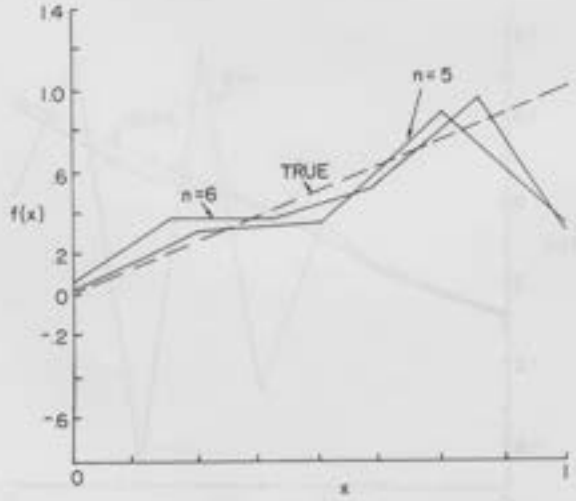

Fig. 3. Solution of equation (12), with $n = 5$ and 6, using the initial-value formulation, i.e. by implementing equation (7).

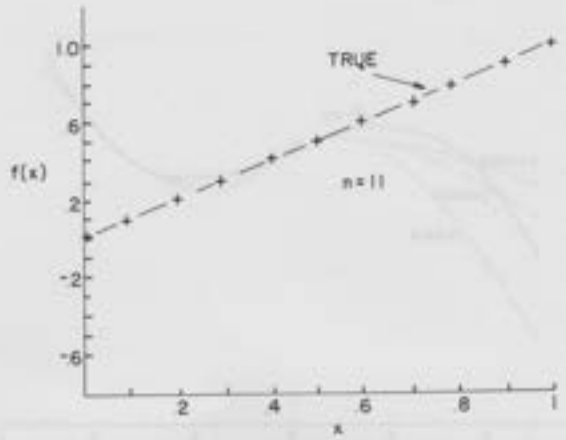

Fig. 4. Solution of equation (12), for $n = 11$, using a smoothing technique, i.e. by implementing equation (11).

DISCUSSION

This paper demonstrated the feasibility of solving some types of Fredholm integral equations of the first kind on an analog computer. As analog computers are ideally suited for solving differential equations with specified initial conditions, one approach is to convert the given integral eguation to an equivalent set of ordinary differential equations. Towards this goal, the integral is first replaced by a summation to get $Af = g$ where A is, in general, a rectangular matrix. The next stage is to solve for f as $f = A^+ g$, where A^+ is the generalized inverse of A. Instead of finding A^+ by iterative methods, the entries of $A^+ g$ are then obtained using an analog computer implementation. Some of the results are presented. Some more results appear elsewhere[19].

The above type of reformulation allows one to use many of the powerful techniques associated with ordinary differential equations. Just as A (or $A^T A$) is often ill-conditioned,

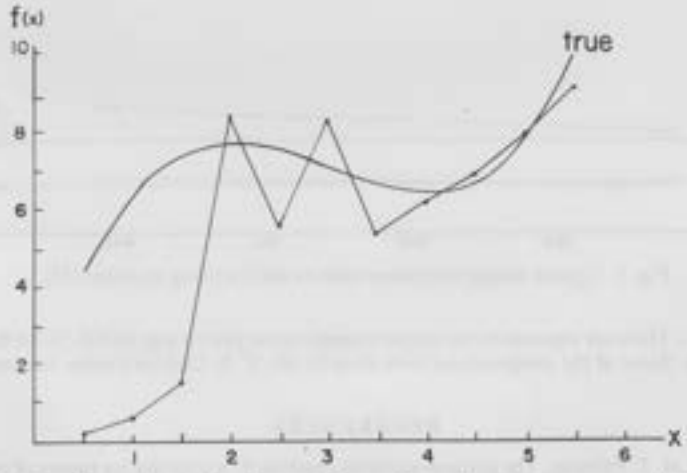

Fig. 5. Solution of equation (13) with $m = 10$, $n = 11$ and $s = 0$ (i.e. no smoothing) using the initial value method.

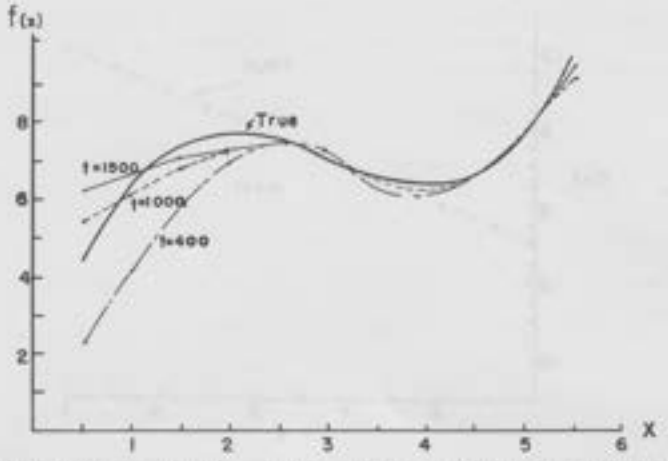

Fig. 6. Solution of equation (13) with $m = 10$, $n = 11$ and $\varepsilon = 0{\cdot}05$ using the initial value method.

the ordinary differential equations obtained after the reformulations tend to be unstable. At this point, the well-developed theory of ordinary differential equations can be readily used to tackle the stability problem.

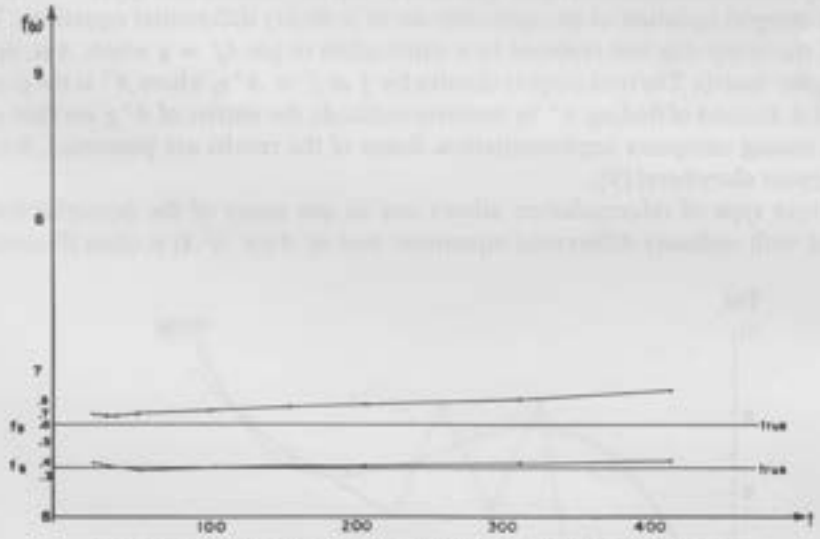

Fig. 7. Typical analog integrator outputs while solving equation (13).

Acknowledgements—The work reported in this paper is supported in part by a grant GK 31786 from the National Science Foundation. Some of the computations were done by Mr. F. P. Chen of Purdue University.

REFERENCES

1. F. Aladjem and M. T. Palmiter, The antigen–antibody reaction V: a quantitative theory of antigen–antibody reactions which allows for heterogeneity of antibodies, *J. Theor. Biol.* **8**, 8 (1961).
2. R. J. Hanson, Integral equations of immunology, *Commun. of the ACM* **15**, 883 (1972).

3. B. Hunt, The inverse problem of radiography, *Intern. J. Math. BioSci.* **8**, 161 (1970).
4. J. C. Gille, Inversion of radiometric measurements, *Bull. Am. Meteorol. Soc.* **49**, 903 (1968).
5. M. T. Chahine, Inverse problems in radiative transfer: determination of atmospheric parameters, *J. Atmos. Sci.* **27** (6), 960 (1970).
6. E. T. Florance, Numerical study of nonlinear inversion, *Proc. NASA–MSFC Conf. on Molecular Radiation* (1968).
7. J. I. King, Inversion by slabs of varying thickness, *J. Atmos. Sci.* **21** (3), 324 (1964).
8. S. Twomey, Indirect measurements of atmospheric temperature profiles from satellites II: mathematical aspects of the inversion problem, *Monthly Weather Rev.* **94** (6), (1966).
9. R. Bellman, R. Kalaba and J. Lockett, Dynamic programming and ill-conditioned linear systems, *J. Math. Anal. Appl.* **10**, 206 (1965).
10. D. L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, *J. ACM* **9**, 84 (1962).
11. S. Twomey, On the numerical solution of Fredholm integral equation of the first kind by the inversion of the linear system produced by quadrature, *J. ACM* **10**, 97 (1963).
12. B. Hunt, An improved technique for using the fast Fourier transform to solve convolution type integral equations, Los Alamos Scientific Lab. Rep. LA-4515-MS, Los Alamos, N. M. (1970).
13. R. Penrose, On best approximate solutions of linear algebraic equations, *Proc. Cambridge Phil. Soc.* **52**, 17 (1956).
14. R. Penrose, A generalized inverse of matrices, *Proc. Cambridge Phil. Soc.* **51**, 406 (1955).
15. C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications*. Wiley, New York (1971).
16. C. A. Den Broeder, Jr. and A. Charnes, Contributions to the theory of generalized inverses of matrices, Purdue University, Lafayette, Indiana (1957).
17. R. K. Manherz, B. W. Jordon and S. L. Hakimi, Analog methods for computation of the generalized inverse, *IEEE Trans. Automatic Control* **AC-13** (5), 582 (1968).
18. L. Fox and E. T. Goodwin, The numerical solution of nonsingular linear integral equations, *Phil. Trans. Roy. Soc. London* **A 245**, 501 (1953).
19. V. Vemuri and F. P. Chen, An initial value method to solve Fredholm integral equations of the first kind, *J. of Franklin Institute* **297** (3), 187 (1974).