THE APPLICATION OF THE CONJUGATE GRADIENT METHOD TO THE SOLUTION OF OPERATOR EQUATIONS - AN UNCONVENTIONAL PERSPECTIVE

Tapan K. Sarkar
Ercument Arvas
Department of Electrical Engineering
Syracuse University
Syracuse, New York 13244-1240

ABSTRACT: This narrative presents an alternate philosophy for the accurate solution of operator equations, you might say "both singular and nonsingular" in general. In this approach, we try to solve the exact operator equation in an approximate way, quite differently from the matrix methods which try to solve the approximate operator equation in an exact fashion. The advantage of this new philosophy is that convergence is assured and a priori error estimates are available. The conjugate gradient methods are numerical methods which provide a means to reach this new goal, as opposed to an efficient means of just solving matrix equations, which some researchers have assumed them to be. We thereby take the position that there is a heaven-and-hell difference between the application of the conjugate gradient method to solve an operator equation and its application to the solution of matrix equations.

1. THE BASIC PHILOSOPHY: The objective is to solve the operator equation $AX = Y$, where A is the known integro-differential operator and X is the unknown to be solved for the known excitation Y. The actual problem setting is in an infinite dimensional space, which in simple terms means that we have an infinite number of unknowns to be solved for. Historically, the matrix methods, starting with Method of Moments, have first projected the original problem posed in an infinite dimensional space to a finite dimensional space (described by the moment matrix) and then have tried to solve the approximate finite dimensional problem exactly using Gaussian elimination and, in recent times, with the iterative methods, particularly the conjugate gradient method. Unfortunately, this basic philosophy lacks mathematical rigor. The area in which this manifests itself is a complete lack of theoretical convergence analysis of the sequence of solutions for an arbitrary operator equation. Whatever convergence analysis exists for matrix methods is generated from numerical experimentation of a particular problem. Hence, there is no guarantee that as the number of unknowns is increased, there is a monotonic convergence of the sequence of approximate solutions [1-2].

What we have tried to do over the years is to usher in a new concept and also point out the deficiencies of the conventional matrix methods. The approach taken by us and Van den Berg [3] are philosophically the same and similar to the work of Hayes [4]. The basic philosophy is simple: Let us not discretize the problem right from the beginning or assume a set of known expansion functions by projecting the operator to a finite dimensional space. Let us see if we can develop a theoretical solution symbolically in an exact fashion. It is at this stage, that our philosophies differ radically from the conventional matrix methods viewpoint. First let us see if we can find a solution to the exact operator equation - let it be in a symbolic fashion. By developing the solution in this way, we have an absolute guarantee to begin with, namely that as the degree of approximation is increased, we indeed have

a monotonic convergence of the solution and that in the limit our solution converges to the exact solution. So in our method, we start with the "blessings" of convergence and, unlike matrix methods, we do not have to "tweak" the expansion functions sometimes in midstream to generate meaningful results. Now we observe that the computer cannot generate the exact solution or, for that matter, follow the exact recipe to reach the solution as it cannot handle an infinite number of unknowns. Therefore, we try to approximate the exact solution.

In summary, the matrix methods first approximate the operator equation and then seek to solve it exactly, whereas in our approach we try to solve exactly the operator equation by utilizing an iterative method, say one of the conjugate gradient methods [5-7] (there are various versions of the conjugate gradient method) and then approximate the exact recipe numerically, yielding an approximate solution. The reward of following the latter procedure is that there is an unconditional guarantee of monotonic convergence to the true solution, as the number of unknowns is increased without "tweaking" any expansion or weighting functions. No such statements can be made for matrix methods, indicating that there are some fundamental differences, in reality, between these two procedures - differences which are not tautological.

In the next section we show how to utilize this new operator form to generate solutions.

2. THE ACT:

Consider the following integral equation:

$$\int_0^1 f(x')\cos\,\pi(x-x')\,dx' = \sin\,\pi x \;;\; 0\le x\le 1 \qquad (1).$$

The objective is to solve for f(x). Before we start number crunching let us take a few moments to "meditate" over the problem. The first question that is raised is: does a solution to this problem exist? The existence of the solution of an operator equation is given by the Fredholm Alternative Theorem, which states that a solution to AX = Y exists, iff Y is orthogonal to every non-trivial solution of the homogeneous adjoint equation $A^*u = 0$, where A^* is the adjoint operator. Hence for a solution to exist all u must be orthogonal to Y. If this condition is violated then a solution to the problem does not exist. In this example, we have a self-adjoint operator, since

$$\langle Au;v\rangle = \int_0^1 dx\; v(x) \int_0^1 dx'\; u(x')\cos\,\pi(x-x') = \langle u;A^*v\rangle;\ \text{so } A=A^*. \qquad (2)$$

By expanding the kernel

$$\cos\,\pi(x-x') = \cos\,\pi x\cos\,\pi x' + \sin\,\pi x\sin\,\pi x'$$

it is seen that there is an infinite set of nontrivial solutions to the adjoint homogeneous equation. Hence, unless Y is orthogonal to all such solutions u, we are just wasting our time trying to solve this problem. It is seen that sin πx is orthogonal to all such solutions (sin $m\pi x$ and cos $m\pi x$ for

$m > 1$ and m odd) of the homogeneous equation and hence the solution to the problem exists. However, the solution is not unique, as a solution to the homogeneous equation can be added to any solution creating a different solution.

But, what has "existence" got to do with electromagnetics? All electromagnetics problems do not have solutions! Consider the problem of electromagnetic scattering from a closed conducting structure at a frequency corresponding to the internal resonant frequency of the same structure. This problem has been recently addressed quite exhaustively!!!. Now the simple truth is that the above problem, when represented by an electric field integral equation, has for the homogeneous equation a nontrivial solution, and unless the excitation is orthogonal to every solution of the homogeneous equation, a solution to the problem does not exist according to the Fredholm alternative. Therefore, instead of trying to solve a problem which is not solvable mathematically, we think we ought to pose the problem in a different way. Yet, methods are still being researched as how to solve this unsolveable problem! An interested reader should look at the development of the modified Green's function as discussed on pp.215-218 of Stakgold[8].

Next, questions about uniqueness, ill-conditioning and the like are addressed. The operator in (1) has a nontrivial solution to the homogeneous equation and it is a positive semidefinite operator. Hence, any matrix methods utilized to solve this equation will fail as the matrix is singular. The strength of the conjugate gradient method lies in the fact that it can solve singular operator equations and the user does not have to worry about the nature of the equation. But, now comes the question: what is the meaning of the solution if the operator is singular? It turns out that the conjugate gradient method will yield the minimum norm solution, if the iteration was started with a zero initial guess. The minimum norm solution implies that of all the possible solutions of this equation, the conjugate gradient method will yield a solution which has the least energy. The solution procedure for a positive semidefinite operator will start with $x_o = 0$ and residual $r_o = Y - AX = \sin \pi x$. Since the operator is self-adjoint, $P_o = r_o = \sin \pi x$.

We update $x_1 = x_o + a_o p_o$, where $a_o = \|r_o\|^2 / \langle Ap_o ; p_o \rangle = 2$ and $x_1 = 2 \sin \pi x$ and $r_1 = 0$ and hence $2 \sin \pi x$ is the minimum norm solution. It can be shown that another solution $q = (-\pi^3/4)x(x-1)$ also satisfies (1). However,

$$\|x_1\|^2 = \int_0^1 |x_1|^2 \, dx > \int_0^1 |q|^2 \, dx$$

and the second solution is not minimum norm. So if we have an ill-conditioned problem, in this case perfectly singular, we can find the minimum norm solution through the use of iterative methods. Direct methods do not work well for ill-conditioned, singular problems. Observe that we have utilized the conjugate gradient method to solve the operator equation directly as first suggested by Hayes [4].

In electromagnetics problems, for example, evaluation of Ap_o and $\|x_1\|^2$ cannot be done analytically. Hence, we have to evaluate these quantities numerically. It is at this point that we introduce numerical

approximations. An additional advantage of handling it in this way is that one can have a grasp on the numerical value of the discretization error. The discretization error in the evaluation of Ap_0 and $\|x_1\|$ can be minimized by simply taking more samples of the functions of interests. For such situations, the residual AX_n -Y will never go to zero as $n \to \infty$. Whatever is left will be the discretization error.

3. EPILOGUE: For illustrative purposes, it is educational to look into the philosophical differences of first discretizing the operator equation and then finding an exact solution to the problem, as opposed to first finding s symbolically exact solution and then finding an approximation to that. In the conventional matrix methods, let us assume that the elements of the matrix have been integrated with sufficient degree of accuracy (even if one chooses a Galerkin procedure) and the final error is always zero as the matrix equation has been solved to the machine precision using either Gaussian elimination or conjugate gradient or by any other method.

Now in the conjugate gradient solution of the operator equation, there are two errors. First the error in the generation of the sequence of the approximation, i.e. $\|X_{exact} - X_n\|$ after m iterations and, secondly, the discretization error made in the evaluation of AXn . If we perform a large number of iterations, presumably $\|X_{exact} - X_n\| \to 0$, whereas the operator $(AX_n - Y)$ would not be zero due to discretization error. So by applying the conjugate gradient method directly to the solution of the operator equation, it is seen that the final error may never become zero, unlike that of matrix methods. The global residual error provides an estimate of the discretization error (i.e. we have obtained X_{exact} subject to the stated discretization error). If this error is large, finer discretization may be preferred. Also no "tweaking" of the expansion functions is involved when one applies the conjugate gradient method directly to the solution of the operator equation. This is the sme philosophy in Van den Berg's approach.

Another point to make: What is the difference between applying the iterative method to the solution of the matrix equation, where each element of the matrix is evaluated at each iteration and the storage decreases from N^2 to 6N, as opposed to applying the conjugate gradient method directly to the solution of the operator equation? It is interesting to note that the application of the conjugate gradient method directly to the solution of an operator equation may sometimes even be computationally more efficient than computing the matrix elements once and using them at each iteration, particularly, when the scatterer geometry fits into an FFT (Fast Fourier Transform) grid [6-7]. However, for an arbitrarily shaped structure, it may not be efficient in some instances to use FFT to perform the evaluation of the convolution. In that case, application of an iterative method directly to the solution of an operator-application of an iterative method directly to the solution of an operator equation may be rather time consuming. However, in spite of this disadvantage, the reward of applying the conjugate gradient method directly to the solution of the operator equation lies in the fact that not only does one have a handle on the discretization error, but also he can solve a problem to a "global" prespecified degree of accuracy.

CONCLUSION: An alternate philosophy is presented for solving operator equations. In this new philosophy the exact system is solved in an approximate numerical fashion as opposed to solving an approximate matrix

equation in an exact way. The advantage of this new philosophy is that convergence to the exact solution is guaranteed and a priori error estimates are available. The conjugate gradient method therefore just turns out to be a method which accomplishes our desired objective of formulating and evaluating a symbolic exact solution of the problem. The use of the conjugate gradient method is distinctly different from its use in solving moment-method matrix equations, sometimes in an efficient way. The basic difference between these two philosophies is the stage at which numerical discretization is made. Our claim is that the new philosophy just presented not only guarantees absolute convergence but also an estimate of the numerical discretization error incurred in the actual solution of the problem.

REFERENCES:

[1] T. K. Sarkar, "The conjugate gradient method as applied to electromagnetic field problems", IEEE Trans. Antennas and Propagation Newsletter, Aug. 1986, pp. 5-14.

[2] T. K. Sarkar, "Reply to comments on "Application of FFT and conjugate gradient method for the solution of electromagnetic scattering from electrically large and small conducting bodies", IEEE Trans. Antennas and Propagation", vol. AP-35, No. 5, May 1987, pp. 608-609.

[3] P. M. Van den Berg, "Iterative computational techniques in scattering based upon the integrated square error criterion", IEEE Trans. Antennas and Propagation", vol. AP-32, No.10, October 1984, pp. 1063-1071.

[4] R. M. Hayes, "Iterative methods of solving linear problems in Hilbert space", in contributions to the solution of systems of linear equations and the determination of eigenvalues", O. Taussky, ed. NBS Appl. Math. Ser., vol. 39, pp. 71-104, 1954.

[5] T. K. Sarkar and E. Arvas, "On a class of finite step iterative methods (conjugate directions) for the solution of an operator equation arising in electromagnetics", IEEE Trans. Antennas and Propagation, vol. AP-33, No.10, October 1985, pp. 1058-1066.

[6] T. K. Sarkar, "On the application of the Generalized BiConjugate gradient method", Journal of Electromagnetic Waves and Applications, Vol. 1, No.3, July 1987, pp. 223-242.

[7] T. K. Sarkar, E. Arvas and S. M. Rao, "Application of the fast fourier transform and conjugate gradient method for the solution of electromagnetic scattering from electrically large and small conduction bodies", IEEE Trans. Antennas and Propagation, vol. AP-34, No.5, May 1986, pp. 635-640.

[8] I. Stakgold, "Green's Functions and Boundary Value Problems", J. Wiley & Sons, New York, 1973.