On the Choice of Expansion and Weighting Functions in the Numerical Solution of Operator Equations

TAPAN K. SARKAR, SENIOR MEMBER, IEEE, ANTONIJE R. DJORDJEVIĆ, AND ERCUMENT ARVAS

Abstract—**One of the objectives of this paper is to discuss the mathematical requirements that the expansion functions must satisfy in the method of moments (MM) solution of an operator equation. A simple differential equation is solved to demonstrate these requirements. The second objective is to study the numerical stability of point matching method, Galerkin's method, and the method of least squares. Pocklington's integral equation is considered and numerical results are presented to illustrate the effect of various choices of weighting functions on the rate of convergence. Finally, it is shown that certain choices of expansion and weighting functions yield numerically acceptable results even though they are not admissible from a strictly mathematical point of view. The reason for this paradox is outlined.**

I. Introduction

THE METHOD OF moments (MM) is a generic name for a whole variety of techniques which solve a linear operator equation by converting it to a matrix equation. In this paper we first discuss some of the requirements on the expansion functions used in MM. To illustrate these requirements, a simple second order differential equation is solved. Then the properties of the point-matching method, Galerkin's method, and the method of least squares are discussed and the influence of the weighting functions on the solution procedure is observed.

Finally, we address the question of numerical stability for the solution of electromagnetic scattering from thin wires due to the choice of various solution techniques. Interestingly, this problem has remained with numerical methods in electromagnetics since its conception. We consider the age-old problem of scattering from wire antennas because this is the first test example for most numero-analytical techniques. We also discuss the paradox resulting from the application of Galerkin's method to the scattering problem. In an earlier paper [1], it has been shown that from a strict mathematical point of view, Galerkin's method should not to be applied to thin wire scatterer problem. However, it is known that the result obtained by this method agrees with experimental results. The reason for this paradox is outlined.

Manuscript received July 23, 1984; revised February 3, 1985.

T. K. Sakar was with the Department of Electrical Engineering, Rochester Institute of Technology, Rochester, NY. He is now with the Department of Electrical Engineering, Syracuse University, Syracuse, NY 13210.

A. R. Djordjević is with the Department of Electrical Engineering, Rochester Institute of Technology, Rochester, NY 14623, on leave from the Department of Electrical Engineering, University of Belgrade, Belgrade, Yugoslavia.

E. Arvas is with the Department of Electrical Engineering, Rochester Institute of Technology, Rochester, NY 14623.

II. Method Of Moments

The method of moments is well-known [2]–[4] and the salient features are summarized here for the sake of completeness.

The objective is to solve a linear operator equation which can symbolically be written as

$$AI = Y \tag{1}$$

where A denotes the operator, and I is the unknown to be solved for a given excitation Y. In MM one usually starts with expanding I in terms of a set of known expansion functions x_i with unknown coefficients α_i, i.e.,

$$I \simeq I_N = \sum_{i=1}^{N} \alpha_i x_i. \tag{2}$$

When this approximation is substituted in (1), one obtains

$$AI \simeq AI_N = \sum_{i=1}^{N} \alpha_i A x_i = Y_N = P_N(Y). \tag{3}$$

Here P_N is defined as the projection operator. This operator projects the excitation Y into the space spanned by the basis functions Ax_i. Thus $P_N(Y)$ represents an element in the range space of the operator whose domain is spanned by x_i.

The next step in MM is to form the residual R_N as follows:

$$R_N = AI_N - Y = P_N(Y) - Y. \tag{4}$$

Finally, the residual is weighted to zero with respect to certain weighting functions W_j such that

$$\langle R_N,\ W_j \rangle = 0, \qquad j = 1,\ 2,\ \cdots,\ N. \tag{5}$$

The criterion under which the weighting functions should be selected is discussed in [1], and the requirements on the expansion functions are summarized in the next section.

The inner product in (5) is defined as

$$\langle C,\ D \rangle = \int_0^L C(z)\bar{D}(z)\ dz \tag{6}$$

where the overbar denotes the complex conjugate.

The unknown α_i can be obtained from the solution of (5) which can be written in the form of a matrix equation as

$$\sum_{i=1}^{N} \alpha_i \langle Ax_i,\ W_j \rangle = \langle Y,\ W_j \rangle, \qquad j = 1,\ 2,\ \cdots,\ N. \tag{7}$$

0018-926X/85/0900-0988$01.00

It is clear that if the approximate solution I_N were to converge to the exact solution I, or equivalently if the residual were to approach zero as $N \to \infty$, then $P_N(Y) \to Y$ as $N \to \infty$. Otherwise, i.e., if $P_N(Y) \not\to Y$ as $N \to \infty$, then the sequence of approximate solutions may not converge to any meaningful result [5], [6].

III. On The Choice Of Expansion Functions

The expansion functions chosen for a particular problem have to satisfy the following criteria.

1) The expansion functions should be in the domain of the operator in some sense, i.e., they should satisfy the differentiability criterion and they must satisfy the boundary conditions for an integrodifferential operator [4], [7], [8]. It is not necessary for each expansion function to satisfy exactly the boundary conditions. What is required is that the total solution must satisfy the boundary conditions at least in some distributional sense. The same holds for the differentiability conditions. When the boundary and the differentiability conditions are satisfied exactly, we have a classical solution and when the above conditions are satisfied in a distributional sense, we have a distributional solution.

2) The expansion functions must be such that Ax_i form a complete set for the range of the operator. It really does not matter whether the expansion functions are complete in the domain of the operator. What is important is that x_i must be chosen in such a way that Ax_i is complete. This will be demonstrated by an example. It is interesting to note that when A is a differential operator, then x_i does not have to be linearly independent for Ax_i to form a complete set. This is illustrated by the following example.

Consider the solution to the differential equation

$$\frac{-d^2I}{dz^2} = 2 + \sin(z), \qquad 0 \le z \le 2\pi \tag{8}$$

with the boundary conditions

$$I(z=0) = I(z=2\pi) = 0. \tag{9}$$

As pointed out earlier, the first step in MM is to choose the expansion functions. A careless choice will be

$$x_i = \sin(iz), \qquad i = 1, 2, \cdots. \tag{10}$$

Then

$$I \simeq I_N = \sum_{i=1}^{N} a_i \sin(iz). \tag{11}$$

The expansion functions satisfy both the differentiability conditions and the boundary conditions. Also, since the operator in (8) is self-adjoint, both Galerkin's method and the method of least squares would yield identical solutions. (The point-matching technique is not considered for this example.)

The above choice of expansion functions leads (for any $N \ge 1$) to the solution

$$I_N = \sin(z) \tag{12}$$

It is quite clear that this does not satisfy (8), and hence is not the solution. Where is the problem? Perhaps the set $\{\sin(iz)\}$ does not form a complete set, even though they are orthogonal in the interval $[0, 2\pi]$. Therefore, in addition to the sin terms, we add the constant and the cos terms in (11). This results in

$$I_N = a_0 + \sum_{i=1}^{N} [a_i \sin(iz) + b_i \cos(iz)] \tag{13}$$

where a_0, a_i, and b_i are constants to be solved for. Now the total solution of (13) has to satisfy the boundary conditions described by (9). Observe that (as $N \to \infty$) (13) is the classical Fourier series solution and the expansion functions used in (13) form a complete set in the interval $[0, 2\pi]$. Now if we solve the problem again by Galerkin's method or by the method of least squares, we still obtain the solution as

$$I_N = \sin(z) \tag{14}$$

which is the same as (12) and is known to be the incorrect solution.

The problem is that even though $x_i\{1, \sin(iz), \cos(iz)\}$ form a complete set, Ax_i do not. This is because Ax_i are merely $\{\cos(iz), \sin(iz)\}$. The constant term is missing from Ax_i. Hence the representation in (13) is not proper. To have the constant term in Ax_i, the representation of I_N must be of the form

$$I_N = \sum_{i=1}^{N} [a_i \sin(iz) + b_i \cos(iz)] + a_0 + cz + dz^2. \tag{15}$$

Observe that the expansion functions $\{1, z, z^2, \sin(iz), \cos(iz)\}$ in the interval $[0, 2\pi]$ form a linearly dependent set. This is because any function such as z or z^2 can be represented in the interval $[0, 2\pi]$ by the set $\{1, \sin(iz), \cos(iz)\}$. The final solution is obtained by using (15) as

$$I = I_N = \sin(z) + z(2\pi - z) \tag{16}$$

which is the exact solution.

This simple example illustrates that it is just not enough to choose the expansion functions to be a set of complete functions in the domain of the operator. In addition, certain completeness conditions have to be satisfied for Ax_i.

IV. Choice Of Weighting Functions

The general conditions on the choice of the weighting functions is addressed in [1]. We here look at some of the properties of the point-matching method, Galerkin's method and the method of least squares and observe how the solution procedure is influenced by the choice of weighting functions, as discussed by Linz [10].

A. Point Matching Method [10]

For this method we have

$$W_j = \delta(z - z_j) \tag{17}$$

and

$$P_N(Y) = Y(z_j) \tag{18}$$

where δ denotes the delta function and z_j are distant (matching) points where Y is evaluated. This method is often suggested

for its simplicity, because the elements in (7) are simply

$$\langle Ax_i,\ W_j\rangle = Ax_i\ |_{z=z_j}. \tag{19}$$

There are, however, difficulties with this method. In general, for a given operator, it is not possible to determine *a priori* which matching points would be suitable. For example, assume A to be the identity operator and choose expansion functions to be polynomials, i.e.,

$$x_i = z^i, \qquad i=0,\ 1,\ 2,\ \cdots,\ N \tag{20}$$

Then the operator equation defined by $AI = Y$ reduces to the approximation of a function Y by polynomial interpolation. This approximation, however, does not always converge. As an example, consider the continuous function

$$I(z) = \frac{1}{1+z^2} = Y(z) \tag{21}$$

on the closed interval $[-a;\ +a]$. It is well-known that for this problem the Lagrange interpolation using equally spaced points z_j converges to $I(z)$ only in the interval $-3.63 \le z \le 3.63$. Outside this interval, it diverges. The rate may be quite slow, even if one does have convergence. For example, consider the following approximation problem:

$$E = \max_{\text{for } -2\le z\le 2} \left| I_N(z) - \frac{1}{1+z^2} \right| \tag{22}$$

with

$$I_N(z) = \sum_{i=0}^{N} a_i z^i \triangleq P_N(I). \tag{23}$$

The error E for different order N is

N	7	11	15
E	.060	.044	.038

The problem in this case is that the norm of the projection operator P_N of (23) is unbounded. It is pointed out in [11] that the projection operator for the general equidistant polynomial interpolation is unbounded and can be approximated by

$$\|P_N\| \triangleq \max_{\text{for } Y\neq 0} \frac{\|P_N(Y)\|}{\|Y\|} \simeq \theta(2^N) \tag{24}$$

where θ represents "of the order of" and the norm is defined as

$$\|C\| = \langle C,\ C\rangle^{1/2}. \tag{25}$$

The order relation in (24) can be improved if, for example, the matching points are at the zeros of the Nth degree Chebyshev polynomial, i.e.,

$$z_j = \cos\frac{(2j-1)\pi}{2N} \quad j=1,\ 2,\ \cdots,\ N \tag{26}$$

shifted to the interval $[a, b]$, or if z_j are given by

$$z_j = \cos\frac{(j-1)\pi}{N-1}, \qquad \text{for } j=1,\ 2,\ 3,\ \cdots,\ N \tag{27}$$

shifted to $[a, b]$, then as $N \to \infty$

$$\|P_N\| \cong \theta\ (\ln N). \tag{28}$$

Observe that for this choice of matching points given by the zeros of the Nth order Chebyshev polynomial, the projection operator diverges less rapidly and it can be shown that this order relationship in (28) cannot be improved upon for any other choice of matching points. However, it is not possible to determine *a priori* for a particular operator equation what collocation points would be suitable.

The general situation is very complicated, but we can get some insight into the difficulties from the following two theorems. For the purpose of the theorems, let Δ be a triangular array of points

$$\begin{array}{llll} z_{00} & & & \\ z_{10} & z_{11} & & \\ z_{20} & z_{21} & z_{22} & \\ \vdots & & & \\ z_{N0} & z_{N1} & \cdots & z_{NN} \end{array} \tag{29}$$

and let $\{I_N(z)\}$ be a sequence of polynomials generated by Lagrange interpolation at the points $a \le z_{N0} \le z_{N1} \cdots \le z_{NN} \le b$.

Theorem 1: For any given triangular set Δ, there exists a continuous function $I(z)$ such that the sequence $\{I_N(z)\}$ does not converge uniformly to $I(z)$, where

$$I_N(z) = \sum_{j=0}^{N} I(z_j)\, C_j(z) \tag{30}$$

and

$$C_j(z) = \frac{(z-z_0)(z-z_1)\cdots(z-z_{j-1})(z-z_{j+1})\cdots(z-z_N)}{(z_j-z_0)(z_j-z_1)\cdots(z_j-z_{j-1})(z_j-z_{j+1})\cdots(z_j-z_N)} \tag{31}$$

for $j = 0,\ 1,\ 2,\ \cdots,\ N$.

Theorem 2: Given any function $I(z)$ which is continuous in the interval $[a, b]$, there exists a Δ such that the sequence $\{I_N(z)\}$ converges to $I(z)$ uniformly.

The proof of these theorems may be found in Rivlin [12]. The first theorem indicates that there is no universally effective interpolation schemes involving function values only, while the second theorem represents a more positive result. However, it is generally not known what the appropriate Δ is [10]–[15].

The objective so far has been to demonstrate that there exist certain difficulties that are associated with the point-matching procedure and with polynomial basis functions. The problem of entire domain polynomial interpolation with equidistant matching points can be alleviated by splitting the given interval into small subintervals and using polynomial of relatively low degree over each subinterval.

B. Galerkin's Method

For Galerkin's method the weighting functions are not delta functions. However, the operator must be of a certain specific type before Galerkin's method is applicable. Often for

Galerkin's method, the weighting functions are chosen as piecewise continuous functions defined over a subinterval. The mathematical justification is that we can get an arbitrarily good accuracy by making the subintervals sufficiently small. Of course the underlying assumption is that Galerkin's method is applicable to that particular operator equation.

Let $P_N(Y)$ be a piecewise linear continuous function which coincides with Y at points z_j and which is linear in the subinterval $[z_j, z_{j+1}]$, $j = 1, 2, 3, \cdots, N$. Then

$$P_N(Y)=\sum_{j=1}^{N} Y(z_j)e_j(z) \tag{32}$$

where $e_j(z)$ is a piecewise linear function (triangular for example) taking the value 1 at z_j and 0 at all other nodes of z_k, with $k \neq j$. $Y(z_j)$ is the value of the function corresponding to the node z_j so that P_N is a linear projection of unity norm from the space of continuous functions in the interval $[a, b]$ to the subspace S_N spanned by $e_1, e_2, \cdots, e_N$.

If we assume that

$$s_N=\max_{1\le j\le N} |z_{j-1}-z_j|\to 0 \tag{33}$$

then it is clear that the following convergence is pointwise, i.e.,

$$\max\ |P_N(Y)-Y|\to 0, \qquad \text{as } N\to\infty \tag{34}$$

for all z.

However, if Y is not zero at the end points a, b, then the above convergence occurs only in a mean square sense. For Galerkin's method to be applicable, the expansion functions x_i must span both the domain and the range of the operator.

C. Method of Least Squares

For this method, the weighting functions are already preselected and defined by Ax_i. It is necessary that Ax_i form a complete set. It is important to point out that the assumption is that Y must be in the domain of the adjoint operator. If this condition is not satisfied, a classical least squares solution is not defined because A^*Y is undefined. The solution computed by the method of least squares is obtained from the solution of

$$A^*Ax=A^*Y. \tag{35}$$

However, numerically one can always define a least squares solution, since for that case we are not solving the problem in an infinite dimensional space, but solving for the unknown constants α_i for a given x_i, in a least squares sense, i.e., we are minimizing the following norm:

$$\left\| \sum_{i=1}^{N} \alpha_i Ax_i - Y \right\|^2 . \tag{36}$$

The least squares technique mathematically and numerically is one of the safest techniques to utilize when very little is known about the nature of the operator and the exact solution.

V. Numerical Example

In this section we apply all the three methods; namely, the point-matching method, Galerkin's method, and the method of least squares to the solution of the current distribution on a wire antenna irradiated from the broadside direction by an incident field of 1v/m. In our numerical procedure, we choose the antenna to be oriented along the z-axis and the origin to be located at the center of the antenna. The antenna is assumed to be 0.5 λ long ($L = 0.5\ \lambda$) and of radius $a = 0.007\ \lambda$. Since the antenna is thin ($a \ll L$ and $a \ll \lambda$), we assume that the current induced on this hollow, thin-walled structure will be z-directed and ϕ-independent. Then the current distribution $I(z')$ is the solution of Pocklington's equation, which is given by

$$\frac{1}{j\omega 4\pi\epsilon_0}\left[k^2\int_{-L/2}^{L/2} I(z')G(z,\ z')\ dz' + \frac{\partial^2}{\partial z^2}\int_{-L/2}^{L/2} I(z')G(z,\ z')\ dz'\right] = -E_{\text{inc}} \tag{37}$$

where E^{inc} is the incident electric field, and

$$k=\frac{2\pi}{\lambda}=\omega\sqrt{\mu_0\epsilon_0} \tag{38}$$

and the kernel $G(z, z')$ is given by

$$G_e(z,\ z')=\frac{1}{2\pi}\int_0^{2\pi}\frac{\exp\ (-jkR)}{R}\ d\phi \tag{39}$$

where

$$R=\sqrt{(z-z')^2+4a^2\ \sin^2\frac{\phi}{2}}\ . \tag{40}$$

The subscript e denotes that the kernel is exact. Often an approximation is used for G_e [9], which is denoted by G_a and is given by

$$G_a(z,\ z')=\exp\ (-jkr)/r \tag{41}$$

where

$$r=\sqrt{(z-z')^2+a^2}\ . \tag{42}$$

The kernel in (41) is usually referred to as the thin-wire kernel. We utilize both G_e and G_a in the numerical computations and illustrate how the rate of convergence is affected by the particular choice of the kernel.

This problem has been solved many times utilizing different numerical approaches. In our procedure, we assume the expansion functions consist of a constant and cos $(nkz)/2$ for $n = 1, 2, \cdots, N$. These functions form a complete set. (Even though cos $knz/2$ is complete for n odd with respect to the L_2 norm (mean square sense), it is not complete with respect to the L_1 norm (point-wise convergence). Therefore, since we are using the point matching procedure, which converges in L_1 norm, it is necessary to choose the above expansion functions). The boundary conditions on the total solution is enforced by multiplying each expansion functions by the term $\sqrt{(L/2)^2 - z^2}$. This function also takes into account the proper edge condition for the current and produces a well-behaved

field near the edges. So by the terms of the problem

$$I_N(z)=\sqrt{(L/2)^2-z^2}\left\{\alpha_0+\sum_{n=1}^{N}\alpha_n\cos\frac{knz}{2}\right\} \tag{43}$$

where $k = 2\pi/\lambda$.

Observe that we have avoided utilizing the conventional piecewise constant (pulse), linear (triangle) or piecewise sinusoidal functions as expansion functions for the current. This is because these expansion functions are discontinuous or have discontinuous first derivatives and hence produce infinite field at certain points along the antenna. This is admissible if one is interested in finding an approximate solution. However, in our case we would like to perform a careful convergence study and, hence, we utilize entire domain expansion functions with continuous first derivatives and with the edge conditions incorporated so that numerical convergence is enhanced.

We now utilize the same expansion functions for all three techniques, the point-matching method, Galerkin's method, and the method of least squares.

In this numerical procedure we compute the left side of (37) by performing an integral of the product of the current given by (43) and the kernel. In this paper we have utilized both the exact kernel given by (39) and the approximate kernel given by (41). The integration is achieved by first extracting out the $1/R$ singularity from the kernel and integrating it analytically around the wire circumference. The remaining regular part is integrated by breaking the z' integral up into three intervals $[-L/2, 0]$, $[0, z]$, $[z, L/2]$ where $z > 0$. Each part is then integrated utilizing a 32-point Gaussian quadrature formula. The integral is calculated only for $z > 0$. The electric field is now obtained from the potential integral (integral of the current with the kernel) by making a finite difference approximation of the double derivative. The $z > 0$ part of the antenna is subdivided into M segments (we vary M from 20 to 90) and the potential integral is computed at the center point of each subinterval. Once these values are known, the field is computed by the finite difference approximation.

Our results compare three parameters of interest for all three methods. The first parameter is the value of the magnitude of the current at the center of the antenna as a function of the number of expansion functions N and also as a function of the number of subdivisions M on the $z > 0$ part of the antenna.

The second parameter of interest is the radar cross section σ in the broadside direction which is given by

$$\begin{aligned}\sigma&=4\pi r^2\frac{|E^s|^2}{|E^{\text{inc}}|^2}\{\text{since } E^{\text{inc}}=1\}\\&=4\pi r^2\left|\frac{-j\omega\mu}{4\pi}\frac{e^{-jkr}}{r}\int_{-L/2}^{L/2}I(z')\,dz'\right|^2\\&=14400\pi^3\left|\int_{-L/2}^{L/2}I(z')\,dz'\right|^2.\end{aligned} \tag{44}$$

The radar cross section is computed for both the exact and the approximate kernel, for various values of the number of expansion functions N and the number of subdivisions M.

The final parameter considered is the residual or the squared magnitude of the residual error of the tangential electric field on the antenna surface, i.e.,

$$\|AI_N-Y\|=\left\{\int_0^L(AI_N-1)^2\,dz\right\}^{1/2}. \tag{45}$$

In Table I, we present the radar cross section σ in the broadside direction for all three different techniques—the point-matching method, Galerkin's method, and the method of least squares. For all three techniques the radar cross section does not increase or decrease monotonically as one increases the number of expansion functions. All three techniques finally settle down to a value which is close to $0.6\,\lambda^2$. The line E represents the computation with the exact kernel G_e defined by (39), whereas the line A corresponds to the approximate kernel defined by (41). For this example there seems to be a difference of about 3 percent between the radar cross section computed by the exact kernel and the approximate kernel. All the computations have been done for $M = 40$. The current at the center of the dipole settles down to 3.417 mA as the number of expansion functions is increased. For all three methods, the current does not converge monotonically as the number of expansion functions is increased.

It is interesting to observe from Table I that the least squares solution has a slower rate of convergence than Galerkin's method as predicted [7]. The point matching procedure takes the least amount of central processing unit (CPU) time and for approximate computational results yields reasonable solutions.

In Table II the residual error is presented as a function of the number of expansion functions. It is seen that even though the least squares solution has the fastest convergence, the difference in the residual for all three techniques is small. Also, the errors given by the least squares solution are consistently lower than with the other solutions. The errors for the approximate kernel appears to be larger than for the exact kernel. By comparing Tables I and II, it is clear that the radar cross section and the current at the center of the dipole are quite insensitive to the quality of the solution. (The quality of the solution is expressed by the residual, i.e., how well the approximate solution satisfies the equation.) Also, the point-matching technique does not produce monotonically convergent sequence of solutions I_N even though the radar cross section appears to be reasonable.

Since we compute the fields by a finite difference approximation of the potential, it would be interesting to observe how the results vary as the subdivisions are made much smaller for the same expansion functions. In Table III the radar cross section and the current at the center of the antenna are presented as a function of the number of divisions M made on the half-length of the antenna. The number of expansion functions remains the same, i.e., $N = 10$. The blank entries for the point matching procedure imply that for these cases the point matching procedure broke down. The reason for this is that the matching points were placed at the locations of comparatively small magnitude fields. For the point matching procedure the matching points have been evenly distributed on the antenna structure, which may not be the best way to do it.

TABLE I
RADAR CROSS SECTION IN THE BROADSIDE DIRECTION $|\sigma|$ AND THE CURRENT AT THE CENTER OF THE DIPOLE (I) AGAINST NUMBER OF EXPANSION FUNCTIONS (N) FOR A FIXED NUMBER OF SUBDIVISIONS
$M = 40$

Number of Expansion Functions (N)		σ_P/λ^2	σ_G/λ^2	σ_L/λ^2	$\lvert I_P\rvert$	$\lvert I_G\rvert$	$\lvert I_L\rvert$
1	E	.4787	.8541	6.5×10^{-3}	2.637	3.533	.3072
	A	.4799	.8332	8.8×10^{-3}	2.640	3.479	.3581
2	E	.5737	.5942	.4367	3.314	3.435	2.986
	A	.5743	.5851	.4496	3.316	3.394	3.001
3	E	.5911	.6018	.5852	3.382	3.417	3.367
	A	.5915	.5910	.5005	3.384	3.381	3.120
4	E	.5997	.6001	.5984	3.416	3.420	3.416
	A	.5996	.5909	.5027	3.417	3.381	3.120
5	E	.6022	.6008	.5993	3.426	3.421	3.417
	A	.6015	.5881	.5156	3.424	3.377	3.170
6	E	.6027	.6009	.5999	3.428	3.421	3.427
	A	.6018	.5899	.5403	3.426	3.376	3.233
7	E	.6031	.6003	.5995	3.430	3.420	3.428
	A	.6009	.5857	.5547	3.422	3.368	3.281
8	E	.6033	.6004	.6002	3.430	3.419	3.418
	A	.6009	.5871	.5706	3.422	3.366	3.319
9	E	.6034	.5999	.5997	3.431	3.418	3.417
	A	.6009	.5836	.5749	3.422	3.358	3.334
10	E	.6018	.5999	.5999	3.425	3.417	3.417
	A	.5957	.5843	.5806	3.402	3.357	3.346

E: exact kernal, A: approximate kernel, P: point matching method, G: Galerkin's method, L: least squares method.

TABLE II
RESIDUAL ERROR FOR ALL THREE METHODS AS A FUNCTION OF THE NUMBER OF EXPANSION FUNCTIONS (N) FOR A FIXED NUMBER OF SUBDIVISIONS $M = 40$

Number of Expansion Functions (N)		$\|\|Ax - Y\|\|_P$	$\|\|Ax - Y\|\|_G$	$\|\|Ax - Y\|\|_L$
1	E	1.399	1.759	.4817
	A	1.254	1.551	.4793
2	E	.3491	.2282	.1835
	A	.2625	.1960	.1751
3	E	.1652	.1005	.0971
	A	.1567	.1557	.1504
4	E	.0912	.0828	.0828
	A	.1791	.1561	.1494
5	E	.0828	.0826	.0825
	A	.2032	.1436	.1383
6	E	.0853	.0823	.0822
	A	.2129	.1264	.1235
7	E	.0887	.0817	.0817
	A	.2130	.1099	.1085
8	E	.0899	.0814	.0813
	A	.2135	.0977	.0971
9	E	.0904	.0812	.0812
	A	.2117	.0896	.0894
10	E	.0848	.0811	.0811
	A	.1725	.0851	.0850

Therefore, if point matching is to yield good solutions, then it is imperative that matching points should be located away from regions where the field is small.

In Table IV, the residuals for all three different approaches are presented as a function of the number of subsections M, for a fixed number of expansion functions $N = 10$. It appears that the method of least squares with the exact kernel settles down much faster than the other two approaches. Also note that the residuals are quite larger for the approximate kernel than for the exact kernel, even though the approximate kernel yields reasonable estimate for the current at the center of the dipole and the radar cross section in the broadside direction. Blank entries represent that data for that particular choice of M are not available due to numerical instabilities.

It appears from the above computer simulations that the radar cross section σ and the current at the center of the antenna are quite insensitive to the quality of the solution I_N. Also the point-matching procedure seems to yield reasonable results taking the least CPU time. Galerkin's method and the method of least squares take the same CPU time as computed by our procedure. However, the least squares method ensures monotonic convergence of the residual whereas Galerkin's method does not.

VI. CONCLUSION

Some of the mathematical restrictions on the expansion functions are discussed. An example is given to illustrate that it is just not enough for the expansion functions x_i to be in the domain of the operator A. In addition, it is required that Ax_i form a complete set.

Secondly, some of the properties of the point-matching method, Galerkin's method and the method of least squares are discussed and careful convergence studies have been performed for the analysis of a cylindrical dipole for point matching method, Galerkin's method and the method of least squares. It appears for this type of problem that the point matching procedure yields a good approximate solution in the least amount of time and the procedure is straightforward. However, one has to be very careful in selecting the location of matching points, away from the regions of zero fields produced by the expansion functions. It is interesting to point out that point matching procedure gives a least squares

TABLE III
RADAR CROSS SECTION $|\sigma|$ AND CURRENT AT THE CENTER OF THE DIPOLE $|I|$ AS A FUNCTION OF SUBSECTIONS (M) ON THE ANTENNA FOR A FIXED NUMBER OF EXPANSION FUNCTIONS $N = 10$

M		σ_P/λ^2	σ_G/λ^2	σ_L/λ^2	$\|I_P\|$	$\|I_G\|$	$\|I_L\|$
20	E	x	.6003	.6003	x	3.420	3.420
	A	.5941	.5927	.5927	3.397	3.392	3.392
30	E	x	.6000	.6000	x	3.418	3.418
	A	.5993	.5878	.5875	3.417	3.371	3.370
40	E	.6018	.5999	.5999	3.425	3.417	3.417
	A	.5957	.5843	.5806	3.402	3.357	3.346
50	E	.6023	.5998	.5998	3.427	3.417	3.417
	A	.5983	.5818	.5649	3.412	3.345	3.298
60	E	.6018	.5998	.5997	3.425	3.417	3.416
	A	.5954	.5798	.5370	3.401	3.337	3.215
70	E	.6024	.5998	.5997	3.427	3.417	3.416
	A	.5974	.5782	.4995	3.409	3.330	3.102
80	E	.6018	.5998	.5997	3.424	3.417	3.416
	A	.5951	.5769	.4583	3.400	3.324	2.973
90	E	.6022	.5999	.5997	3.426	3.417	3.416
	A	.5968	.5758	.4180	3.406	3.320	2.843

TABLE IV
RESIDUALS FOR THREE DIFFERENT APPROACHES AS A FUNCTION OF NUMBER OF SUBSECTIONS M FOR A FIXED NUMBER OF EXPANSION FUNCTIONS $N = 10$

M		$\|\|Ax - Y\|\|_P$	$\|\|Ax - Y\|\|_G$	$\|\|Ax - Y\|\|_L$
20	E	x	.0715	.0713
	A	.4744	.4379	.4372
30	E	x	.0588	.0584
	A	.5098	.3560	.3539
40	E	.0853	.0554	.0551
	A	.4785	.2931	.2898
50	E	.0878	.0544	.0541
	A	.4981	.2483	.2435
60	E	.0841	.0539	.0538
	A	.4750	.2194	.2122
70	E	.0895	.0538	.0536
	A	.4906	.2035	.1931
80	E	.0831	.0537	.0536
	A	.4720	.1973	.1836
90	E	.0875	.0536	.0536
	A	.4852	.1977	.1810

solution as the number of matching points approaches infinity [10]. Also the expansion functions must be chosen in such a way that they have a continuous first derivative, otherwise the field would be infinity at certain points. Least squares method is more sensitive to the proper choice of expansion functions (the continuity of the first derivative and the inclusion of the edge condition) than the other two techniques. However, as the reward, such a careful choice always leads to a monotonic convergence of the residual. Finally the antenna problem, even though the application of Galerkin's method is not mathematically justified, it yields reasonable results. The reason for this is that the improper choice of weighting functions might cause large residual fields near the antenna ends. This in turn could be interpreted as a large excitation near the end. Since a large excitation of relatively short spatial extension at the end does not significantly change the current distribution on the open antenna structure, Galerkin's method yields reasonable results, even though its application cannot be justified from a strictly mathematical point of view.

References

[1] T. K. Sarkar, "A note on the choice of weighting functions in the method of moments," *IEEE Trans. Antennas Propagat.*, vol. AP-33, pp. 436–441, Apr. 1985.

[2] R. F. Harrington, "Matrix methods for field problems," *Proc. IEEE*, vol. 55, pp. 136–149, Feb. 1967.

[3] ——, *Field Computation by Moment Methods*. New York: MacMillan, 1968.

[4] R. F. Harrington and T. K. Sarkar, "Boundary elements and the method of moments," in *Boundary Elements—Proc. Fifth Int. Conf.*, C. A. Brebbia, T. Futugami, and M. Tanaka, Eds., Nov. 1983, pp. 31–40.

[5] T. K. Sarkar, D. D. Weiner, and V. K. Jain, "Some mathematical considerations in dealing with inverse problems," *IEEE Trans. Antennas Propagat.*, vol. AP-31, pp. 373–379, Mar. 1981.

[6] S. G. Miklin, *The Numerical Performance of Variational Methods*. Groningen: Walters-Noordhoff, 1971.

[7] T. K. Sarkar, "A note on the variational method, (Rayleigh–Ritz), Galerkin's method, and the method of least squares," *Radio Sci.*, vol. 18, pp. 1207–1224, Nov. 1983.

[8] ——, "A study of the various methods for computing electromagnetic field utilizing thin wire integral equations," *Radio Sci.*, vol. 18, pp. 29–38, Jan. 1983.

[9] B. D. Popović, M. B. Dragović, and A. R. Djordjević, *Analysis and Synthesis of Wire Antennas*. New York: Wiley, 1982.

[10] P. Linz, *Theoretical Numerical Analysis*. New York: Wiley, 1979.

[11] M. A. Kransnoselskii, G. M. Vainikko, P. P. Zabreiko, Ya B. Rutitskii, and V. Ya Stetsenko, *Approximate Solution of Operator Equations*. Groningen: Walters-Noordhoff, 1972.

[12] T. J. Rivlin, *An Introduction to the Approximation of Functions*. Waltham, MA: Blaisdell, 1969.

[13] J. L. Phillips, "The use of collocation as a projection method for solving linear operator equations," *SIAM J. Numer. Anal.*, vol. 9, pp. 14–28, 1972.

[14] B. Noble, "Error analysis of collocation methods for solving Fredholm integral equations," in *Topics in Numerical Analysis*, J. J. Miller, Ed. New York, London: Academic, 1973, pp. 211–232.

[15] S. G. Mikhlin, *The Numerical Performance of Variational Methods* (translated from Russian by R. S. Anderssen). Groningen: Walters-Noordhoff, 1971.

Tapan K. Sarkar (S'69–M'76–SM'81), for a photograph and biography please see page 539 of the May 1985 issue of this TRANSACTIONS.

Antonije R. Djordjević, for a photograph and biography please see page 539 of the May 1985 issue of this TRANSACTIONS.

Ercument Arvas, for a photograph and biography please see page 724 of the September 1983 issue of this TRANSACTIONS.