CHAPTER

1

Deterministic Problems

1-1. Introduction

The use of high-speed digital computers not only allows more computations to be made than ever before, it makes practicable methods of solution too repetitious for hand calculation. In the past much effort was expended to analytically manipulate solutions into forms which minimized the computational effort. It is now often more convenient to use computer time to reduce the analytical effort. Approximation techniques, once considered a last resort, can be carried to such high orders on computers that they are for most purposes as good as exact answers. They also permit treatment of problems not solvable by exact methods.

This text has been written to provide a unified treatment of matrix methods for computing the solutions to field problems. The basic idea is to reduce a functional equation to a matrix equation, and then solve the matrix equation by known techniques. These concepts are best expressed in the language of linear spaces and operators. However, it is not necessary that the reader have prior knowledge of this theory, because we shall define and illustrate the concepts as they are introduced. A brief summary of linear spaces and operators is given in Appendix A. Detailed expositions may be found in many textbooks [1–3].[1]

In this chapter we consider equations of the inhomogeneous type

$$L(f) = g \tag{1-1}$$

[1] Bracketed numbers refer to the References at the end of each chapter.

where L is an *operator*, g is the *source* or *excitation* (known function), and f is the *field* or *response* (unknown function to be determined). By the term *deterministic* we mean that the solution to (1-1) is unique; that is, only one f is associated with a given g. A problem of *analysis* involves the determination of f when L and g are given. A problem of *synthesis* involves a determination of L when f and g are specified. In this text we consider only the analysis problem.

This chapter presents the basic mathematical techniques for reducing functional equations to matrix equations. A unifying principle for such techniques is found in the general *method of moments*, in terms of which most specific solutions can be interpreted. We shall consider a deterministic problem solved once it is reduced to a suitable matrix equation, since the solution is then given by matrix inversion. Most computers have subroutines available for matrix inversion, which is a relatively simple operation. For reference, the widely used Gauss-Jordan method is given in Appendix B.

The examples of this chapter are simple, chosen to illustrate the theory without clouding the picture with physical concepts or complicated mathematics. However, when these methods are applied to problems of practical interest the procedures are not so simple. The details vary according to the type of problem, and can be illustrated only by treating a variety of problems. For this reason we treat many specific problems in the subsequent chapters. It is hoped that these examples will not only allow the reader to solve similar problems, but will suggest extensions and modifications to treat other types. Although most of the examples are taken from electromagnetic theory, the procedures are general and apply to field problems of any kind.

1-2. Formulation of Problems

The general methods of solution will be discussed in the notation of linear spaces and operators, and hence specific problems should be put into this notation. Given a deterministic problem of the form $L(f) = g$, we must identify the operator L, its domain (the functions f on which it operates), and its range (the functions g resulting from the operation). Furthermore, we usually need an *inner product* $\langle f, g\rangle$, which is a scalar defined to satisfy[2]

$$\langle f, g\rangle = \langle g, f\rangle \tag{1-2}$$

$$\langle \alpha f + \beta g, h\rangle = \alpha\langle f, h\rangle + \beta\langle g, h\rangle \tag{1-3}$$

$$\begin{aligned} \langle f^*, f\rangle &> 0 \qquad \text{if } f \neq 0 \\ &= 0 \qquad \text{if } f = 0 \end{aligned} \tag{1-4}$$

[2] The usual definition of inner product in Hilbert space corresponds to $\langle f^*, g\rangle$ in our notation. For this text it is more convenient to show the conjugate operation explicity wherever it occurs, and to define the adjoint operator without conjugation.

where α and β are scalars and * denotes a complex conjugate. We sometimes need the *adjoint operator* L^a and its domain, defined by

$$\langle Lf, g\rangle = \langle f, L^a g\rangle \tag{1-5}$$

for all f in the domain of L. An operator is *self-adjoint* if $L^a = L$ and the domain of L^a is that of L.

Properties of the solution depend upon properties of the operator. An operator is *real* if Lf is real whenever f is real. An operator is *positive definite* if

$$\langle f^*, Lf\rangle > 0 \tag{1-6}$$

for all $f \neq 0$ in its domain. It is *positive semidefinite* if $>$ is replaced by $\geq$ in (1-6), *negative definite* if $>$ is replaced by $<$ in (1-6), etc. We shall identify other properties of operators as they are needed.

If the solution to $L(f) = g$ exists and is unique for all g, then the *inverse operator* L^{-1} exists such that

$$f = L^{-1}(g) \tag{1-7}$$

If g is known, then (1-7) represents the solution to the original problem. However, (1-7) is itself an inhomogeneous equation for g if f is known, and its solution is $L(f) = g$. Hence L and L^{-1} form a pair of operators, each of which is the inverse of the other.

Facility in formulating problems using the concepts of linear spaces comes only with practice, which will be provided by the many examples in the following chapters. For the present, let us consider a simple abstract example so that mathematical concepts may be illustrated without bringing physical concepts into the picture.

Example. Given $g(x)$, find $f(x)$ in the interval $0 \leq x \leq 1$ satisfying

$$-\frac{d^2f}{dx^2} = g(x) \tag{1-8}$$

$$f(0) = f(1) = 0 \tag{1-9}$$

This is a boundary-value problem for which

$$L = -\frac{d^2}{dx^2} \tag{1-10}$$

The range of L is the space of all functions g in the interval $0 \leq x \leq 1$ that we wish to consider. The domain of L is the space of those functions f in the interval

$0 \leq x \leq 1$, satisfying the boundary conditions (1-9), and having second derivatives in the range of L. The solution to (1-8) is not unique unless appropriate boundary conditions are included. In other words, both the differential operator and its domain are required to define the operator.

A suitable inner product for this problem is

$$\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx \tag{1-11}$$

It is easily shown that (1-11) satisfies the postulates (1-2) to (1-4), as required. Note that the definition (1-11) is not unique. For example,

$$\int_0^1 w(x)f(x)g(x)\,dx \tag{1-12}$$

where $w(x) > 0$ is an arbitrary weighting function, is also an acceptable inner product. However, the adjoint operator depends on the inner product, which can often be chosen to make the operator self-adjoint.

To find the adjoint of a differential operator, we form the left side of (1-5), and integrate by parts to obtain the right side. For the present problem

$$\begin{aligned}\langle Lf, g\rangle &= \int_0^1 \left(-\frac{d^2f}{dx^2}\right) g\,dx \\ &= \int_0^1 \frac{df}{dx}\frac{dg}{dx}\,dx - \left[\frac{df}{dx}\,g\right]_0^1 \\ &= \int_0^1 f\left(-\frac{d^2g}{dx^2}\right)dx + \left[f\frac{dg}{dx} - g\frac{df}{dx}\right]_0^1 \end{aligned} \tag{1-13}$$

The last terms are boundary terms, and the domain of L^a may be chosen so that these vanish. The first boundary terms vanish by (1-9), and the second vanish if

$$g(0) = g(1) = 0 \tag{1-14}$$

It is then evident that the adjoint operator to (1-10) for the inner product (1-11) is

$$L^a = L = -\frac{d^2}{dx^2} \tag{1-15}$$

Since $L^a = L$ and the domain of L^a is the same as that of L, the operator is self-adjoint.

It is also evident that L is a real operator, since Lf is real when f is real. That L is a positive definite operator is shown from (1-6) as follows:

$$\begin{aligned}\langle f^*, Lf\rangle &= \int_0^1 f^*\left(-\frac{d^2f}{dx^2}\right)dx\\ &= \int_0^1 \frac{df^*}{dx}\frac{df}{dx}\,dx - \left[f^*\frac{df}{dx}\right]_0^1\\ &= \int_0^1 \left|\frac{df}{dx}\right|^2 dx \qquad (1\text{-}16)\end{aligned}$$

Note that L is a positive definite operator even if f is complex.

The inverse operator to L can be obtained by standard Green's function techniques.[3] It is

$$L^{-1}(g) = \int_0^1 G(x, x')g(x')\,dx' \qquad (1\text{-}17)$$

where G is the Green's function

$$G(x, x') = \begin{cases} x(1 - x') & x < x' \\ (1 - x)x' & x > x' \end{cases} \qquad (1\text{-}18)$$

We can verify that (1-17) is the inverse operator by forming $f = L^{-1}(g)$, differentiating twice, and obtaining (1-8). Note that no boundary conditions are needed on the domain of L^{-1}, which is characteristic of most integral operators. That L^{-1} is self-adjoint follows from the proof that L is self-adjoint, since

$$\langle Lf_1, f_2\rangle = \langle g_1, L^{-1}g_2\rangle \qquad (1\text{-}19)$$

Of course, the self-adjointness of L^{-1} can also be proved directly. It similarly follows that L^{-1} is positive definite whenever L is positive definite, and vice versa.

1-3. Method of Moments

We now discuss a general procedure for solving linear equations, called the *method of moments* [4,5]. Consider the inhomogeneous equation

$$L(f) = g \qquad (1\text{-}20)$$

[3] See, for example, reference [2], Chapter 3.

where L is a linear operator, g is known, and f is to be determined. Let f be expanded in a series of functions $f_1, f_2, f_3, \ldots$ in the domain of L, as

$$f = \sum_n \alpha_n f_n \tag{1-21}$$

where the α_n are constants. We shall call the f_n *expansion functions* or *basis functions*. For exact solutions, (1-21) is usually an infinite summation and the f_n form a complete set of basis functions. For approximate solutions, (1-21) is usually a finite summation. Substituting (1-21) in (1-20), and using the linearity of L, we have

$$\sum_n \alpha_n L(f_n) = g \tag{1-22}$$

It is assumed that a suitable inner product $\langle f, g \rangle$ has been determined for the problem. Now define a set of *weighting functions*, or *testing functions*, $w_1, w_2, w_3, \ldots$ in the range of L, and take the inner product of (1-22) with each w_m. The result is

$$\sum_n \alpha_n \langle w_m, Lf_n \rangle = \langle w_m, g \rangle \tag{1-23}$$

$m = 1, 2, 3, \ldots$. This set of equations can be written in matrix form as

$$[l_{mn}][\alpha_n] = [g_m] \tag{1-24}$$

where

$$[l_{mn}] = \begin{bmatrix} \langle w_1, Lf_1 \rangle & \langle w_1, Lf_2 \rangle & \cdots \\ \langle w_2, Lf_1 \rangle & \langle w_2, Lf_2 \rangle & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} \tag{1-25}$$

$$[\alpha_n] = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \end{bmatrix} \qquad [g_m] = \begin{bmatrix} \langle w_1, g \rangle \\ \langle w_2, g \rangle \\ \vdots \end{bmatrix} \tag{1-26}$$

If the matrix $[l]$ is nonsingular its inverse $[l^{-1}]$ exists. The α_n are then given by

$$[\alpha_n] = [l_{nm}^{-1}][g_m] \tag{1-27}$$

and the solution for f is given by (1-21). For concise expression of this result, define the matrix of functions

$$[\tilde{f}_n] = [f_1 \quad f_2 \quad f_3 \quad \ldots] \tag{1-28}$$

and write

$$f = [\tilde{f}_n][\alpha_n] = [\tilde{f}_n][l_{mn}^{-1}][g_m] \tag{1-29}$$

This solution may be exact or approximate, depending upon the choice of the f_n and w_n. The particular choice $w_n = f_n$ is known as *Galerkin's method* [6,7].

If the matrix $[l]$ is of infinite order, it can be inverted only in special cases, for example, if it is diagonal. The classical eigenfunction method leads to a diagonal matrix, and can be thought of as a special case of the method of moments. If the sets f_n and w_n are finite, the matrix is of finite order, and can be inverted by known methods (Appendix B).

One of the main tasks in any particular problem is the choice of the f_n and w_n. The f_n should be linearly independent and chosen so that some superposition (1-21) can approximate f reasonably well. The w_n should also be linearly independent and chosen so that the products $\langle w_n, g \rangle$ depend on relatively independent properties of g. Some additional factors which affect the choice of f_n and w_n are (1) the accuracy of solution desired, (2) the ease of evaluation of the matrix elements, (3) the size of the matrix that can be inverted, and (4) the realization of a well-conditioned matrix $[l]$.

Example. Consider the same equation as in the example of Section 1-2, but with the specific source $g = 1 + 4x^2$. Hence our problem is

$$-\frac{d^2 f}{dx^2} = 1 + 4x^2 \tag{1-30}$$

$$f(0) = f(1) = 0 \tag{1-31}$$

This is, of course, a simple boundary-value problem with solution

$$f(x) = \frac{5x}{6} - \frac{x^2}{2} - \frac{x^4}{3} \tag{1-32}$$

To illustrate the procedure, the problem will be reconsidered by the method of moments.

For a power-series solution, let us choose

$$f_n = x - x^{n+1} \tag{1-33}$$

$n = 1, 2, 3, \ldots, N$, so that the series (1-21) is

$$f = \sum_{n=1}^{N} \alpha_n (x - x^{n+1}) \tag{1-34}$$

Note that the term x is needed in (1-33), else the f_n will not be in the domain of L; that is, the boundary conditions will not be satisfied. For testing functions, choose

$$w_n = f_n = x - x^{n+1} \tag{1-35}$$

in which case the method is that of Galerkin. In Section 1-8 it is shown that the w_n should be in the domain of the adjoint operator. Since L is self-adjoint for this problem, the w_n should be in the domain of L, as are those of (1-35).

Evaluation of the matrices (1-25) and (1-26) for the inner product (1-11) and $L = -d^2/dx^2$ is straightforward, and results in

$$l_{mn} = \langle w_m, Lf_n \rangle = \frac{mn}{m+n+1} \tag{1-36}$$

$$g_m = \langle w_m, g \rangle = \frac{m(3m+8)}{2(m+2)(m+4)} \tag{1-37}$$

For any fixed N (number of expansion functions), the α_n are given by (1-27) and the approximation to f by (1-34).

To illustrate convergence, let us consider successive approximations as N is increased. For $N = 1$, we have $l_{11} = 1/3$, $g_1 = 11/30$, and hence from (1-24) $\alpha_1 = 11/10$. For $N = 2$, the matrix equation (1-24) becomes

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \frac{11}{30} \\ \frac{7}{12} \end{bmatrix}, \tag{1-38}$$

from which the α's are found as

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{10} \\ \frac{2}{3} \end{bmatrix} \tag{1-39}$$

For $N = 3$, the matrix equation (1-24) becomes

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{2} & \frac{3}{5} \\ \frac{1}{2} & \frac{4}{5} & 1 \\ \frac{3}{5} & 1 & \frac{9}{7} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} \frac{11}{30} \\ \frac{7}{12} \\ \frac{51}{70} \end{bmatrix} \tag{1-40}$$

from which the α's are found as

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{3} \end{bmatrix} \tag{1-41}$$

Note that this third-order solution is the exact solution, (1-32). For $N = 4$ we

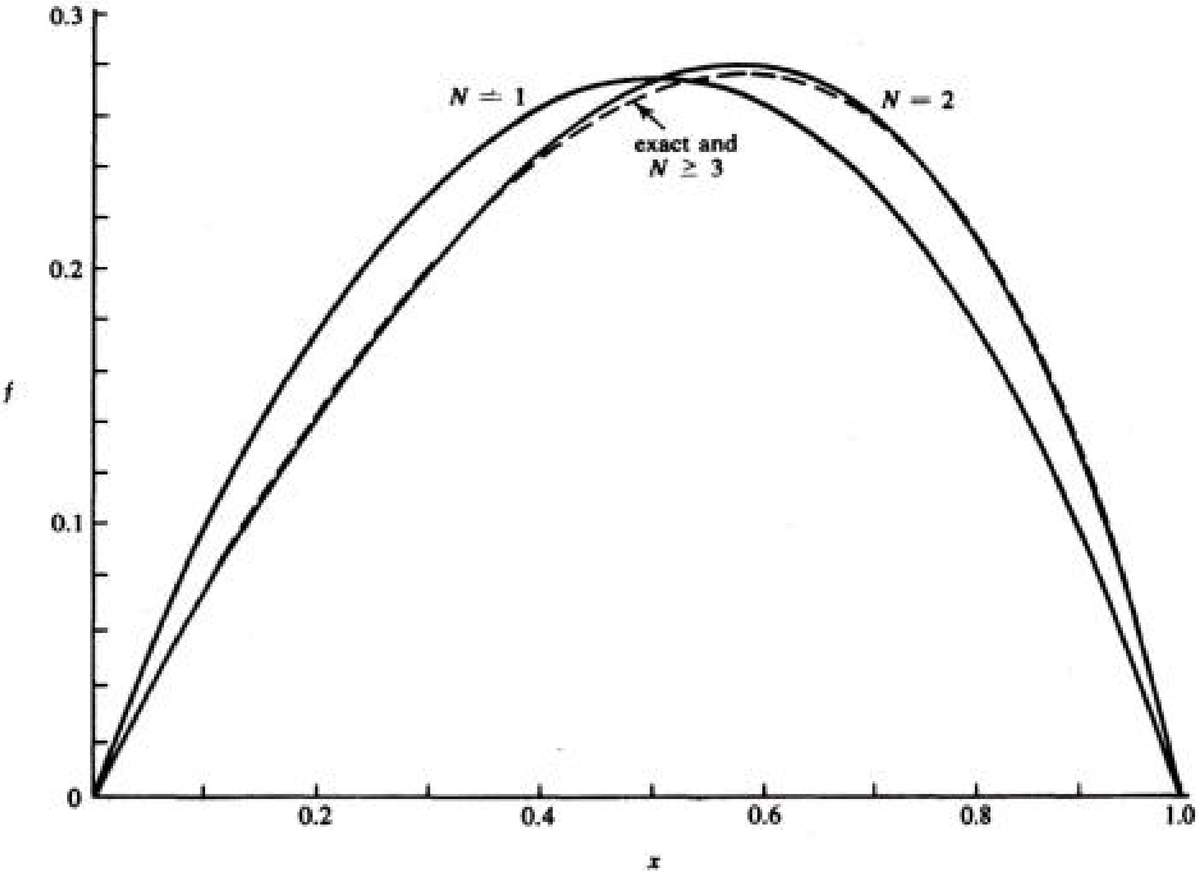


Figure 1-1. **Solutions using $f_n = x - x^{n+1}$ and Galerkin's method.**

again obtain the exact solution, and so on for higher N. Plots of the various solutions are shown in Fig. 1-1.

The reason an exact solution is obtained for this problem is that some combination of the f_n can exactly represent the solution, and any N linearly independent tests must correctly determine the coefficients. If the solution cannot be expressed as a finite series of the f_n, then we continue to obtain approximate solutions converging to the exact solution in the sense of projections, as discussed in Section 1-8.

More important than solving any particular equation, the inverse matrix $[l^{-1}]$ gives a representation of the inverse operator L^{-1}. Hence we have a solution (usually approximate) to $Lf = g$ for *any* g. In physical problems, L represents the system, g the excitation, and f the response. A determination of the $[l^{-1}]$ matrix therefore gives us a general solution for the system, that is, the response f for arbitrary excitation g, assuming that g is reasonably well behaved.

1-4. Point Matching

The integration involved in evaluating the $l_{mn} = \langle w_m, Lf_n\rangle$ of (1-25) is often difficult to perform in problems of practical interest. A simple way to obtain approximate solutions is to require that equation (1-22) be satisfied at discrete points in the region of interest. This procedure is called a *point-matching method.* In terms of the method of moments, it is equivalent to using Dirac delta functions as testing functions. The following example illustrates this in the one-dimensional case.

Example. Reconsider the problem of Section 1-3, stated by (1-30) and (1-31). Again we choose expansion functions (1-33), so that (1-22) becomes

$$\sum_{n=1}^{N} \alpha_n \left[-\frac{d^2}{dx^2} (x - x^{n+1}) \right] = 1 + 4x^2 \tag{1-42}$$

For a point-matching solution, let us take the points

$$x_m = \frac{m}{N+1} \qquad m = 1, 2, \ldots, N \tag{1-43}$$

which are equispaced in the interval $0 \le x \le 1$. Requiring (1-42) to be satisfied at each x_m gives us the matrix equation (1-24), with elements

$$l_{mn} = n(n+1)\left(\frac{m}{N+1}\right)^{n-1} \tag{1-44}$$

$$g_m = 1 + 4\left(\frac{m}{N+1}\right)^2 \tag{1-45}$$

Note that this result is identical to choosing weighting functions

$$w_m = \delta(x - x_m) \tag{1-46}$$

where $\delta(x)$ is the Dirac delta function, and applying the method of moments with inner product (1-11).

To illustrate some numerical results, consider the solution as N is increased. For $N = 1$, we have $l_{11} = 2$, $g_1 = 2$, and from (1-27) $\alpha_1 = 1$. For $N = 2$, the matrix equation is

$$\begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \frac{13}{9} \\ \frac{25}{9} \end{bmatrix} \tag{1-47}$$

from which the α's are found as

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{18} \\ \frac{2}{3} \end{bmatrix} \tag{1-48}$$

For $N = 3$, the exact solution (1-41) must again be obtained, since the exact solution is a linear combination of the f_n's and we are applying N independent tests. Similarly, for $N > 3$ we continue to obtain the exact answer for the same reason. Plots of these solutions differ to some extent from those of Fig. 1-1 but are qualitatively similar. The point-matching solutions in this case are actually less accurate than the corresponding Galerkin approximations, but for low orders of solution they are usually sensitive to the particular points of match. For high-order solutions the use of equispaced points normally gives excellent results.

Note that even though the $[l]$ matrices of (1-36) and (1-44) are quite different in form, they give similar results. There are infinitely many possible sets of basis functions and of testing functions. Some sets may give faster convergence than others, or give matrices easier to evaluate, or give acceptable results with smaller matrices, etc. For any particular problem one of our tasks is to choose sets well suited to the problem.

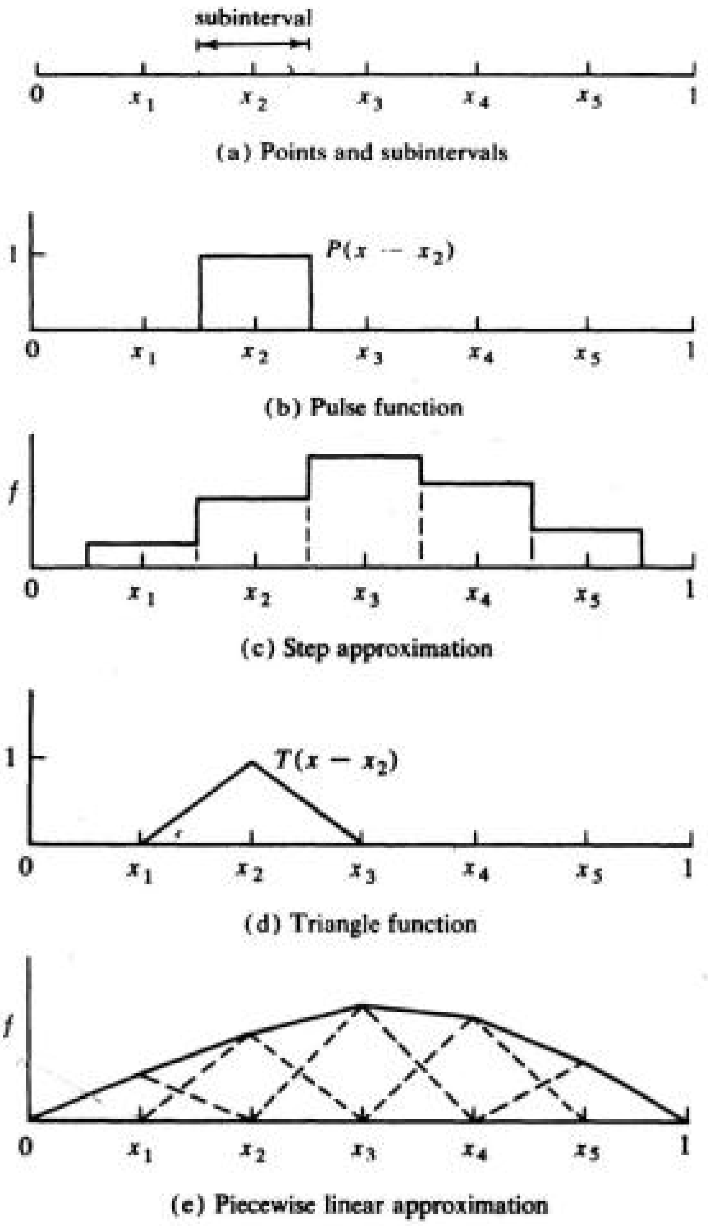
1-5. Subsectional Bases

Another approximation useful for practical problems is the *method of subsections*. This involves the use of basis functions f_n each of which exists only over subsections of the domain of f. Then each α_n of the expansion (1-21) affects the approximation of f only over a subsection of the region of interest. This procedure often simplifies the evaluation and/or the form of the matrix $[l]$. Sometimes it is convenient to use the point-matching method of Section 1-4 in conjunction with the subsectional method.

Example. Again consider the problem of Section 1-3, stated by (1-30) and (1-31). N equispaced points on the interval $0 \le x \le 1$ are defined by the x_m of (1-43). A subinterval is defined to be of width $1/(N + 1)$ centered on the x_m. This is shown for case $N = 5$ in Fig. 1-2(a). A function which exists over only one subinterval is the *pulse function*

$$P(x) = \begin{cases} 1 & |x| < \dfrac{1}{2(N+1)} \\ 0 & |x| > \dfrac{1}{2(N+1)} \end{cases} \tag{1-49}$$

For $N = 5$, the function $P(x - x_2)$ is shown in Fig. 1-2(b). A linear combination of $f_n = P(x - x_n)$ according to (1-21) gives a *step approximation* to f, as represented by Fig. 1-2(c). However, for $L = -d^2/dx^2$, the operation LP does not yield a function in the range of L. Hence the pulse functions cannot be used as basis functions unless we extend the operator (Section 1-7) or use an approximate operator (Section 1-6).



(a) Points and subintervals

(b) Pulse function

(c) Step approximation

(d) Triangle function

(e) Piecewise linear approximation

***Figure 1-2.* Subsectional bases and functional approximations.**

A better-behaved function is the *triangle function,* defined as

$$T(x) = \begin{cases} 1 - |x|(N+1) & |x| < \dfrac{1}{N+1} \\ \\ 0 & |x| > \dfrac{1}{N+1} \end{cases} \tag{1-50}$$

For the case $N = 5$ the function $T(x - x_2')$ is shown in Fig. 1-2(d). A linear combination of triangle functions of the form

$$f = \sum_{n=1}^{N} \alpha_n T(x - x_n) \tag{1-51}$$

gives a *piecewise linear approximation* to f, as represented by Fig. 1-2(e). For $L = -d^2/dx^2$, the operation LT gives the symbolic function

$$LT(x - x_n) = (N + 1)[-\delta(x - x_{n-1}) + 2\delta(x - x_n) - \delta(x - x_{n+1})] \tag{1-52}$$

where $\delta(x)$ is the Dirac delta function. We can use this result in the method of moments as long as the w_n are not also symbolic functions. We cannot use a point-matching procedure in this case.

To follow through the method of moments, let $f_n = T(x - x_n)$, that is, use the expansion (1-51). As testing functions, choose $w_m = P(x - x_m)$. For inner product (1-11), the matrix elements of (1-25) and (1-26) are easily evaluated as

$$l_{mn} = \begin{cases} 2(N+1) & m = n \\ -(N+1) & |m - n| = 1 \\ 0 & |m - n| > 1 \end{cases} \tag{1-53}$$

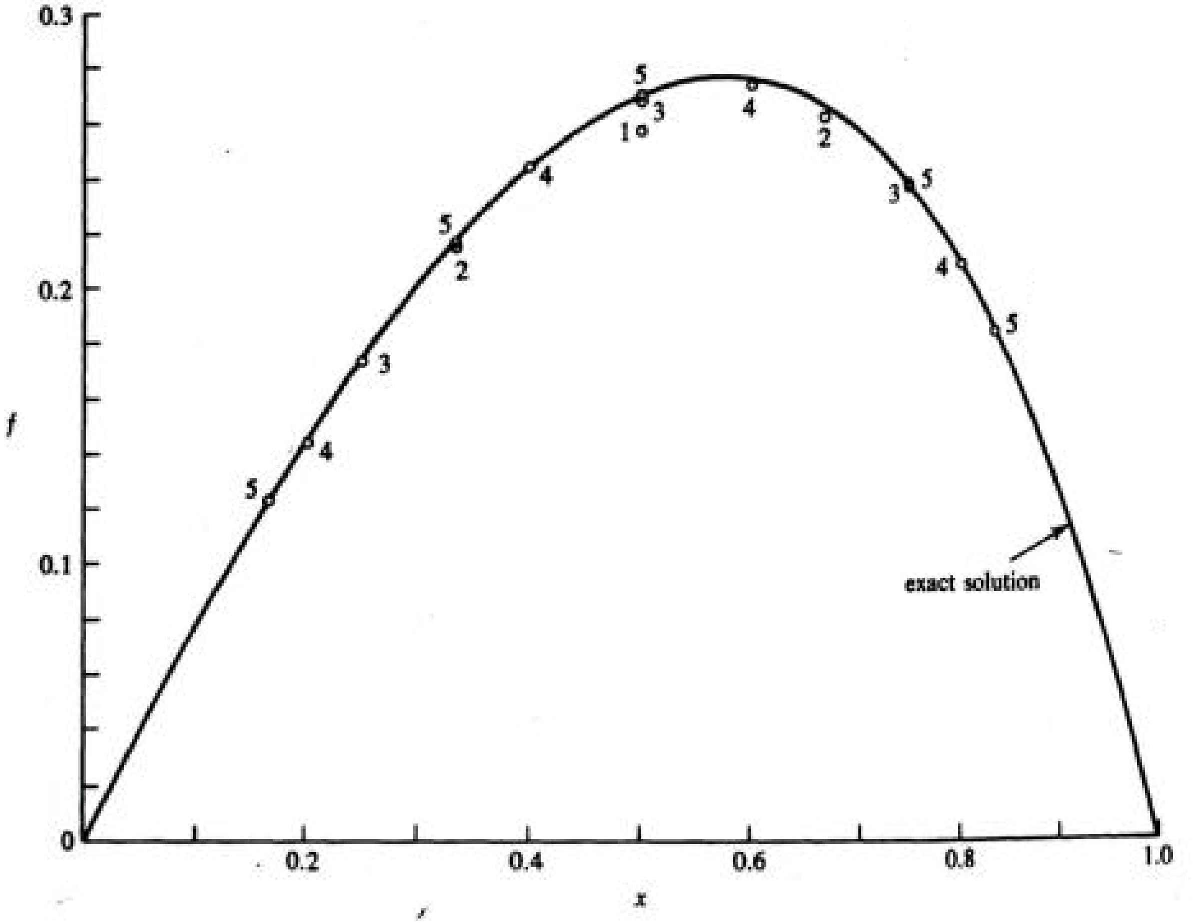


Figure 1-3. **Moment solutions using triangles for expansion and pulses for testing. Numbers adjacent to points denote order of solution.**

$$g_m = \frac{1}{N+1}\left[1 + \frac{4m^2 + (1/3)}{(N+1)^2}\right] \tag{1-54}$$

Note the particularly simple form of [*l*]. We shall encounter this form again in connection with difference equations (Section 1-6).

Figure 1-3 illustrates the convergence of the above solution as N (number of subsections) is increased. Only the break points of the piecewise linear solution are shown; the functional approximation is given by straight lines joining these points. The break points are, of course, also the α_n, since they are the peaks of the triangle-function components.

1-6. Approximate Operators

In complex problems it is sometimes convenient to approximate the operator to obtain approximate solutions. For differential operators, the finite-difference approximation has been widely used [8]. For integral operators, an approximate operator can be obtained by approximating the kernel of the integral operator [6]. Any method whereby a functional equation is reduced to a matrix equation can be interpreted in terms of the method of moments. Hence for any matrix solution using approximation of the operator there will be a corresponding moment solution using approximation of the function.

Example. Let us consider the problem (1-30) and (1-31) by a finite-difference approximation. This involves replacing all derivatives by finite differences; that is, for a given Δx,

مشتقات

$$\frac{df}{dx} \approx \frac{1}{\Delta x}\left[f\left(x + \frac{\Delta x}{2}\right) - f\left(x - \frac{\Delta x}{2}\right)\right]$$

$$\frac{d^2f}{dx^2} \approx \frac{1}{\Delta x}\left[f'\left(x + \frac{\Delta x}{2}\right) - f'\left(x - \frac{\Delta x}{2}\right)\right] \tag{1-55}$$

$$\approx \frac{1}{(\Delta x)^2}\left[f(x - \Delta x) - 2f(x) + f(x + \Delta x)\right]$$

For our present problem, consider the interval $0 \le x \le 1$ divided into $N + 1$ segments, with end points x_n, as depicted in Fig. 1-2(a). For Δx equal to one segment, $\Delta x = 1/(N + 1)$, and a finite-difference approximation to $L = -d^2/dx^2$ is

$$L^d f = (N+1)^2\left[-f\left(x - \frac{1}{N+1}\right) + 2f(x) - f\left(x + \frac{1}{N+1}\right)\right] \tag{1-56}$$

Note that $L^d \to L$ as $N \to \infty$ for all f in the domain of L.

We can now apply the method of moments to the approximate equation

$$L^d f = 1 + 4x^2 \tag{1-57}$$

subject to boundary conditions $f(0) = f(1) = 0$. Most commonly this is done by a point-matching procedure at the x_m. The result is a matrix equation of the form (1-24), where the α_n correspond to $f(x_n)$,

$$l_{mn} = \begin{cases} 2(N+1)^2 & m = n \\ -(N+1)^2 & |m-n| = 1 \\ 0 & |m-n| > 1 \end{cases} \tag{1-58}$$

$$g_m = 1 + 4\left(\frac{m}{N+1}\right)^2 \tag{1-59}$$

Note that the $[l]$ matrix of (1-58) is the same form as that of (1-53) obtained from a subsectional basis. [The trivial difference in the position of $N + 1$ can be taken care of by choosing $w_m = (N + 1)\,P(x - x_m)$ in the solution of Section 1-5.] The g_m of (1-59) and (1-54) are slightly different, and hence the two solutions will be slightly different. However, as N becomes larger the two g_m approach one another, so the rates of convergence of the two solutions are about the same.

Numerical results for the above solution are similar to those of Fig. 1-3. Iterative procedures are sometimes used to solve the matrix equations obtained by difference approximations [9]. However, iterative procedures usually converge slowly, and with high-speed large-memory computers it is often simpler to invert the matrix. Because of the tridiagonal form of $[l]$, special techniques can be used to invert it [10].

1-7. Extended Operators

As noted earlier, an operator is defined by an operation (for example, $L = -d^2/dx^2$) plus a domain (space of functions to which the operation may be applied). We can *extend the domain* of an operator by redefining the operation to apply to new functions (not in the original domain) as long as this extended operation does not change the original operation in its domain. If the original operator is self-adjoint, it is desirable to make the extended operator self-adjoint also. By this procedure we can use a wider class of functions for solution by the method of moments. This becomes particularly important in multivariable problems (fields in multidimensional space), where it is not always easy to find simple functions in the domain of the original operator.

Example A. Suppose we wish to use pulse functions for an expansion of f in a moment solution for the operator $L = -d^2/dx^2$. As noted in Section 1-5, these are not in the original domain of L. However, for any functions w and f in the original domain,

$$\langle w, Lf \rangle = \int_0^1 \frac{dw}{dx}\frac{df}{dx}\,dx \tag{1-60}$$

obtained from (1-11) by integration by parts. If Lf does not exist, but df/dx does exist, (1-60) can be used to define an extended operator. This extends the domain of L to include functions f whose second derivatives do not exist, but whose first derivatives do exist. It is still assumed that $f(0) = f(1) = 0$. Actually, the type of extension represented here is precisely that which gives rise to the theory of symbolic functions. By using Dirac delta functions in earlier sections we anticipated this concept of extending the domain of a differential operator.

To apply the method of moments using pulse functions and the extended operator, let

$$f = \sum_{n=1}^{N} \alpha_n P(x - x_n) \tag{1-61}$$

where P are the pulse functions defined by (1-49). For testing functions, let $w_m = T(x - x_m)$, where T are the triangle functions defined by (1-50). The elements of the $[l]$ matrix are found using (1-60) as

$$l_{mn} = \langle w_m, Lf_n \rangle = \begin{cases} 2(N+1) & m = n \\ -(N+1) & |m-n| = 1 \\ 0 & |m-n| > 1 \end{cases} \tag{1-62}$$

Note that these are identical to the elements (1-53), which were for f_n and w_m reversed from those of the present solution. We could have anticipated this result because L is self-adjoint. The elements of the $[g]$ matrix are now given by

$$g_m = \int_0^1 T(x - x_m)(1 + 4x^2)\,dx \tag{1-63}$$

which yields a result slightly different from (1-54). However, the two g_m approach each other as N becomes large, and the convergence of the two solutions is about the same.

Numerical results for the above example are similar to those of Fig. 1-3 for various N. However, the functional approximation in this case is a step approximation; that is, the points are midpoints of steps, instead of break points of a piecewise linear approximation as in Fig. 1-3.

Example B. As a second example, let us extend the original domain of $L = -d^2/dx^2$ to apply to functions not satisfying the boundary conditions $f(0) = f(1) = 0$. Referring to (1-13), we note that boundary terms appear if the functions do not obey the given boundary conditions. However, if an extended operator L^e is defined by

$$\langle w, L^e f\rangle = \int_0^1 wLf\,dx - \left[f\frac{dw}{dx}\right]_0^1 \tag{1-64}$$

we have $\langle w, L^e f\rangle = \langle f, L^e w\rangle$ even if the original boundary conditions are not met. Hence the extended operator is self-adjoint regardless of boundary conditions. A method-of-moments solution therefore proceeds in this extended domain in the same manner as for the original domain, except that the expansion and testing functions need not satisfy boundary conditions.

To illustrate the procedure, consider the choice

$$f_n = w_n = x^n \qquad n = 1, 2, \ldots, N \tag{1-65}$$

For $N \geq 4$ these functions form a basis for the exact solution (1-32), and hence the exact solution should be obtained. Evaluating the matrices in the usual way, using the extended operator for $l_{mn} = \langle w_m, L^e f_n\rangle$, for $N = 4$ we obtain the matrix equation

$$\begin{bmatrix} -1 & -2 & -3 & -4 \\ -2 & -\frac{8}{3} & -\frac{7}{2} & -\frac{22}{5} \\ -3 & -\frac{7}{2} & -\frac{21}{5} & -5 \\ -4 & -\frac{22}{5} & -5 & -\frac{40}{7} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ \frac{17}{15} \\ \frac{11}{12} \\ \frac{27}{35} \end{bmatrix} \tag{1-66}$$

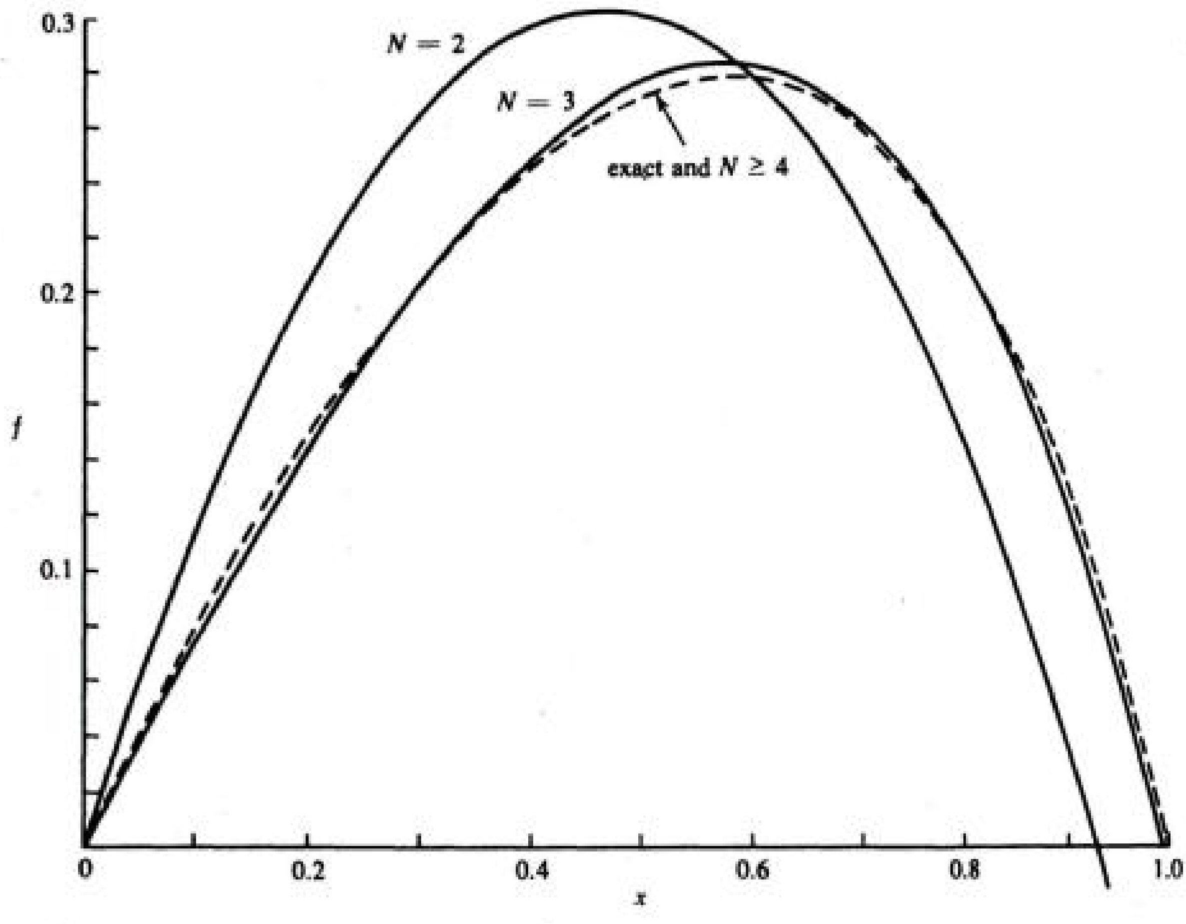


Figure 1-4. **Extended operator moment solutions using powers of *x* for expansion and testing.**

This may be solved for the α's to obtain

$$\begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\\ \alpha_4\end{bmatrix} = \begin{bmatrix}\frac{5}{6}\\ -\frac{1}{2}\\ 0\\ -\frac{1}{3}\end{bmatrix} \tag{1-67}$$

which is indeed the exact solution. Note that if (1-65) are used with the original operator $L = -d^2/dx^2$ a singular $[l]$ matrix results, and hence no solution is obtained. To illustrate convergence using the extended operator, Fig. 1-4 shows plots of the cases $N = 2$ and $N = 3$, plus the exact solution ($N \geq 4$).

1-8. Variational Interpretation

It is well known that Galerkin's method ($w_n = f_n$) is equivalent to the Rayleigh-Ritz variational method [6,7]. That the general method of moments is also a variational method is usually not noted, but the proof is essentially the same as for Galerkin's method [7].

Let us first interpret the method of moments according to the concepts of linear spaces. Let $\mathscr{S}(Lf)$ denote the range of L, $\mathscr{S}(Lf_n)$ denote the space spanned by the Lf_n, and $\mathscr{S}(w_n)$ denote the space spanned by the w_n. The method of moments (1-23) then equates the projection of Lf onto $\mathscr{S}(w_n)$ to the projection of the approximate Lf onto $\mathscr{S}(w_n)$. Figure 1-5 represents this pictorially. In the special case of Galerkin's method, $\mathscr{S}(w_n) = \mathscr{S}(f_n)$. Because the process of obtaining projections minimizes an error, the method of moments is an error-minimizing procedure. Because the error is orthogonal to the projections, it is of second order. This same conclusion is obtained from the calculus of variations [7]. The derivation of the variational results will not be given here, but we shall summarize the conclusions.

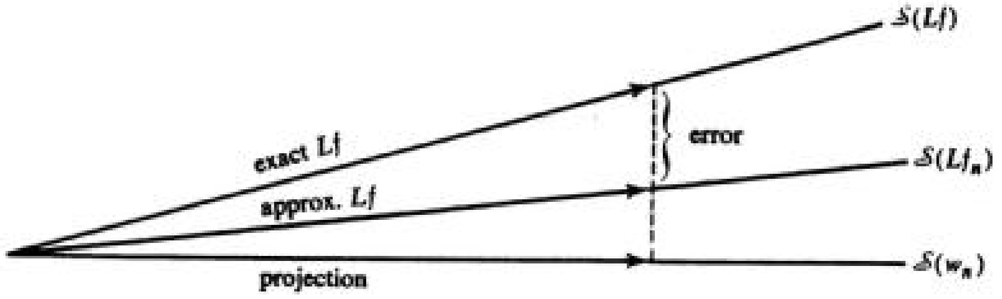


Figure 1-5. Pictorial representation of the method of moments in function space.

Given an operator equation $Lf = g$, it is desired to determine a functional of f (number depending on f)

$$\rho(f) = \langle f, h \rangle \tag{1-68}$$

where h is a given function. If h is a continuous function, then $\rho(f)$ is a *continuous linear functional*. Th• functional ρ may be f itself if h is an impulse function, but then ρ is no longer a continuous functional. Now let L^a be the adjoint operator to L, and define an adjoint function f^a (adjoint field) by

$$L^a f^a = h \tag{1-69}$$

By the calculus of variations, it can then be shown that [7]

$$\rho = \frac{\langle f, h \rangle \langle f^a, g \rangle}{\langle Lf, f^a \rangle} \tag{1-70}$$

is a variational formula for ρ with stationary point (1-68) when f is the solution to $Lf = g$ and f^a the solution to (1-69). For an approximate evaluation of ρ, let

$$f = \sum_n \alpha_n f_n \qquad f^a = \sum_m \beta_m w_m \tag{1-71}$$

Substitute these in (1-70), and apply the Rayleigh-Ritz conditions $\partial\rho/\partial\alpha_i = \partial\rho/\partial\beta_i = 0$ for all i. The result is that the necessary and sufficient conditions for ρ to be a stationary point are equations (1-23). Hence the method of moments is equivalent to the Rayleigh-Ritz variational method [7]. The method of moments is closely related to the direct methods of the calculus of variations, so called because they yield a solution to the variational problem without recourse to the associated differential equation.

The above variational interpretation can be used to give additional insight into how to choose the testing functions. It is evident from (1-69) and (1-71) that the w_n should be chosen so that some linear combination of them can closely represent the adjoint field f^a. When we calculate f itself by the method of moments, h of (1-68) is a Dirac delta function and f^a of (1-69) is a Green's function. This implies that some combination of the w_n should be able to approximate the Green's function. Since a Green's function is usually poorly behaved, we should expect computation of a field by the method of moments to converge less slowly than computation of a continuous linear functional. This is found actually to be the case.

1-9. Perturbation Solutions

Sometimes the problem under consideration is only slightly different (perturbed) from a problem which can be solved exactly (the unperturbed problem). A first-order solution to the perturbed problem can then be obtained by using the solution to the unperturbed problem as a basis for the method of moments. This procedure is called a *perturbation method*. Higher-order perturbation solutions can be obtained by using the unperturbed solution plus correction terms in the method of moments. Sometimes this is done as successive approximations by including one correction term at a time, but for machine computations it is usually easier to include all correction terms at once.

To express these concepts in equation form, let

$$L_0(f_0) = g \tag{1-72}$$

represent the unperturbed problem for which the solution f_0 is known. Let $M = L - L_0$ be the difference operator, and hence

$$L(f) = (L_0 + M)(f) = g \tag{1-73}$$

represents the perturbed problem for which the solution f is desired. For a first-order perturbation solution, let

$$f = \alpha f_0 \tag{1-74}$$

and apply the method of moments. If L is self-adjoint, the testing function $w = f_0$ may be chosen; otherwise we should choose $w = f_0^a$, the solution to the unperturbed adjoint problem. An application of the method of moments to this one-term expansion yields

$$(\langle f_0, L_0 f_0\rangle + \langle f_0, M f_0\rangle)\alpha = \langle f_0, g\rangle \tag{1-75}$$

Now, by (1-72), $\langle f_0, L_0 f_0\rangle = \langle f_0, g\rangle$, and the above equation can be written

$$\alpha = 1 - \frac{\langle f_0, M f_0\rangle}{\langle f_0, g\rangle + \langle f_0, M f_0\rangle} \tag{1-76}$$

If the perturbation is truly small, the second term in the denominator of (1-76) will be small compared to the first term, and from (1-74) and (1-76)

$$f \approx \left(1 - \frac{\langle f_0, M f_0\rangle}{\langle f_0, g\rangle}\right) f_0 \tag{1-77}$$

This is the first-order perturbation solution.

For higher-order solutions, we merely choose $f_1 = f_0$ in the general method of moments (Section 1-3) and $f_2, f_3, \ldots$ serve as correction terms. For self-adjoint operators, choose $w_1 = f_0$; otherwise choose $w_1 = f_0^a$. The advantage of a perturbation approach over other moment solutions rests primarily in the faster convergence of the perturbation solution.

References

[1] B. Z. Vulikh, *Introduction to Functional Analysis for Scientists and Technologists*, translated by I. N. Sneddon, Pergamon Press, Oxford, 1963.

[2] B. Friedman, *Principles and Techniques of Applied Mathematics*, John Wiley & Sons, Inc., New York, 1956.

[3] J. W. Dettman, *Mathematical Methods in Physics and Engineering*, McGraw-Hill Book Co., New York, 1962.

[4] L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces*, translated by D. E. Brown, Pergamon Press, Oxford, 1964, pp. 586–587.

[5] R. F. Harrington, "Matrix Methods for Field Problems," *Proc. IEEE*, Vol. **55**, No. 2, Feb. 1967, pp. 136–149.

[6] L. V. Kantorovich and V. I. Krylov, *Approximate Methods of Higher Analysis*, 4th ed., translated by C. D. Benster, John Wiley & Sons, Inc., New York, 1959, Chap. IV.

[7] D. S. Jones, "A Critique of the Variational Method in Scattering Problems," *IRE Trans.*, Vol. **AP-4**, No. 3, 1956, pp. 297–301.

[8] G. E. Forsythe and W. R. Wasow, *Finite-Difference Methods for Partial Differential Equations*, John Wiley & Sons, Inc., New York, 1960.

[9] R. V. Southwell, *Relaxation Methods in Theoretical Physics*, Vol. **1**, Oxford University Press, London, 1946.

[10] P. Henrici, *Discrete Variable Methods in Ordinary Differential Equations*, John Wiley & Sons, Inc., New York, 1962, pp. 350–355.

CHAPTER

2

Electrostatic Fields

2-1. Operator Formulation

The static electric intensity **E** is conveniently found from an electrostatic potential ϕ according to

$$\mathbf{E} = -\nabla\phi \tag{2-1}$$

where ∇ is the gradient operator. In a region of constant permittivity ε and volume charge density ρ, the electrostatic potential satisfies the *Poisson equation*

$$-\varepsilon\nabla^2\phi = \rho \tag{2-2}$$

where ∇^2 is the Laplacian operator. For unique solutions, boundary conditions on ϕ are needed. In other words, the domain of the operator must be specified.

For now, consider fields from charges in unbounded space, in which case

$$r\phi \to \text{constant as } r \to \infty \tag{2-3}$$

where r is the distance from the coordinate origin, for every ρ of finite extent. Now the differential operator formulation is

$$L\phi = \rho \tag{2-4}$$

where

$$L = -\varepsilon\nabla^2 \tag{2-5}$$

and the domain of L is those functions ϕ whose Laplacian exists and have $r\phi$ bounded at infinity according to (2-3). The well-known solution to this problem is

$$\phi(x, y, z) = \iiint \frac{\rho(x', y', z')}{4\pi\varepsilon R}\, dx'\, dy'\, dz' \tag{2-6}$$

where $R = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}$ is the distance from a source point (x', y', z') to a field point (x, y, z). Hence the inverse operator to L is

$$L^{-1} = \iiint dx'\, dy'\, dz' \frac{1}{4\pi\varepsilon R} \tag{2-7}$$

It is important to keep in mind that (2-7) is inverse to (2-5) only for the boundary conditions (2-3). If the boundary conditions are changed, L^{-1} changes. Also, the designation of (2-5) as L and (2-7) as L^{-1} is arbitrary, and we could reverse the notation if desired.

A suitable inner product for electrostatic problems (ε constant) is[1]

$$\langle \phi, \psi \rangle = \iiint \phi(x, y, z)\psi(x, y, z)\, dx\, dy\, dz \tag{2-8}$$

where the integration is over all space. That (2-8) satisfies the required postulates (1-2), (1-3), and (1-4) is easily verified. We now wish to show that L is self-adjoint for this inner product. For this, form the left side of (1-5),

$$\langle L\phi, \psi \rangle = \iiint (-\varepsilon\nabla^2\phi)\psi\, d\tau \tag{2-9}$$

where $d\tau = dx\, dy\, dz$. Green's identity is

$$\iiint\limits_V (\psi\nabla^2\phi - \phi\nabla^2\psi)\, d\tau = \oiint\limits_S \left(\psi\frac{\partial\phi}{\partial n} - \phi\frac{\partial\psi}{\partial n}\right) ds \tag{2-10}$$

where S is the surface bounding the volume V and n is the outward direction normal to S. Let S be a sphere of radius r, so that in the limit $r \to \infty$ the volume V includes all space. For ϕ and ψ satisfying boundary conditions (2-3), $\psi \to C_1/r$ and $\partial\phi/\partial n \to C_2/r^2$ as $r \to \infty$. Hence $\psi\, \partial\phi/\partial n \to C/r^3$ as $r \to \infty$, and similarly for $\phi\, \partial\psi/dn$. Since $ds = r^2 \sin\theta\, d\theta\, d\phi$ increases only as r^2, the right side of (2-10) vanishes as $r \to \infty$. Equation (2-10) then reduces to

$$\iiint \psi\nabla^2\phi\, d\tau = \iiint \phi\nabla^2\psi\, d\tau \tag{2-11}$$

[1] For ε a function of position, the differential operator (2-5) is changed to $-\nabla\cdot(\varepsilon\nabla)$, and ε should be included in (2-8) as a weight function to make this new operator self-adjoint.

from which it is evident that the adjoint operator L^a is

$$L^a = L = -\varepsilon\nabla^2 \tag{2-12}$$

Since the domain of L^a is that of L, the operator L is self-adjoint. The mathematical concept of self-adjointness in this case is related to the physical concept of reciprocity [1].

It is evident from (2-5) and (2-7) that L and L^{-1} are real operators. It will now be shown that they are also positive definite; that is, they satisfy (1-6). As discussed in Section 1-2, we need only show it for either L or L^{-1}. For L, form

$$\langle\phi^*, L\phi\rangle = \iiint \phi^*(-\varepsilon\nabla^2\phi)\, d\tau \tag{2-13}$$

and use the vector identity $\phi\nabla^2\phi = \nabla\cdot(\phi\nabla\phi) - \nabla\phi\cdot\nabla\phi$ plus the divergence theorem. The result is

$$\langle\phi^*, L\phi\rangle = \iiint_V \varepsilon\nabla\phi^*\cdot\nabla\phi\, d\tau - \oiint_S \varepsilon\phi^*\nabla\phi\cdot d\mathbf{s} \tag{2-14}$$

where S bounds V. Again take S a sphere of radius r. For ϕ satisfying (2-3), the last term of (2-14) vanishes as $r \to \infty$ for the same reasons as in (2-10). Then

$$\langle\phi^*, L\phi\rangle = \iiint \varepsilon\,|\nabla\phi|^2\, d\tau \tag{2-15}$$

and, for ε real and $\varepsilon > 0$, L is positive definite. In this case positive definiteness of L is related to the concept of electrostatic energy.

2-2. Charged Conducting Plate

Consider a square conducting plate $2a$ meters on a side and lying on the $z = 0$ plane with center at the origin, as shown in Fig. 2-1. Let $\sigma(x, y)$ represent the surface charge density on the plate, assumed to have zero thickness. The electrostatic potential at any point in space is

$$\phi(x, y, z) = \int_{-a}^{a} dx' \int_{-a}^{a} dy' \frac{\sigma(x', y')}{4\pi\varepsilon R} \tag{2-16}$$

where $R = \sqrt{(x - x')^2 + (y - y')^2 + z^2}$. The boundary condition is $\phi = V$ (constant) on the plate. The integral equation for the problem is therefore

$$V = \int_{-a}^{a} dx' \int_{-a}^{a} dy' \frac{\sigma(x', y')}{4\pi\varepsilon\sqrt{(x - x')^2 + (y - y')^2}} \tag{2-17}$$

where $|x| < a$, $|y| < a$. The unknown to be determined is the charge density $\sigma(x, y)$. A parameter of interest is the capacitance of the plate

$$C = \frac{q}{V} = \frac{1}{V} \int_{-a}^{a} dx \int_{-a}^{a} dy\, \sigma(x, y) \tag{2-18}$$

which is a continuous linear functional of σ.

Let us first go through a simple subsection and point-matching solution [2], and later interpret it in terms of more general concepts. Consider the plate divided into N square subsections, as shown in Fig. 2-1. Define functions

$$f_n = \begin{cases} 1 & \text{on } \Delta s_n \\ 0 & \text{on all other } \Delta s_m \end{cases} \tag{2-19}$$

and let the charge density be represented by

$$\sigma(x, y) \approx \sum_{n=1}^{N} \alpha_n f_n \tag{2-20}$$

Substituting (2-20) in (2-17), and satisfying the resultant equation at the midpoint (x_m, y_m) of each Δs_m, we obtain the set of equations

$$V = \sum_{n=1}^{N} l_{mn} \alpha_n \qquad m = 1, 2, \ldots, N \tag{2-21}$$

where

$$l_{mn} = \int_{\Delta x_n} dx' \int_{\Delta y_n} dy' \frac{1}{4\pi\varepsilon\sqrt{(x_m - x')^2 + (y_m - y')^2}} \tag{2-22}$$

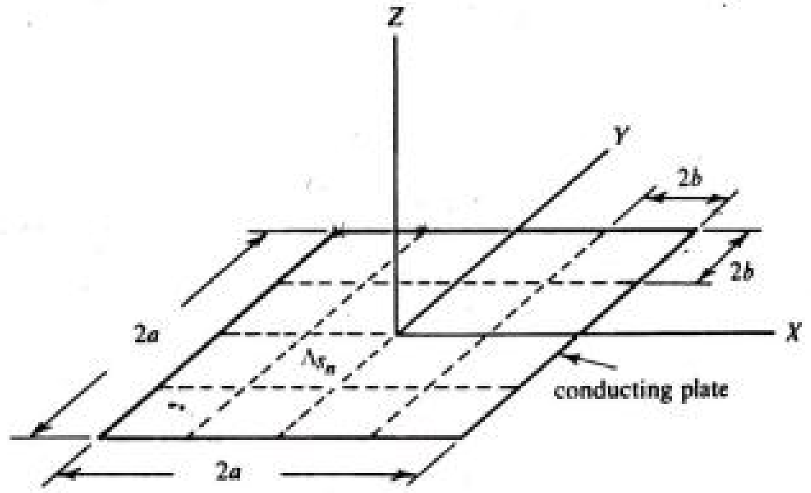


Figure 2-1. Square conducting plate and subsections.

Note that l_{mn} is the potential at the center of Δs_m due to a uniform charge density of unit ampiitude over Δs_n. A solution to the set (2-21) gives the α_m, in terms of which the charge density is approximated by (2-20). The corresponding capacitance of the plate, approximating (2-18), is

$$C \approx \frac{1}{V} \sum_{n=1}^{N} \alpha_n \, \Delta s_n = \sum_{mn} l_{mn}^{-1} \, \Delta s_n \tag{2-23}$$

This result can be interpreted as stating that the capacitance of an object is the sum of the capacitances of all its subsections plus the mutual capacitances between every pair of subsections.

To translate the above results into the language of linear spaces and the method of moments, let

$$f(x, y) = \sigma(x, y) \tag{2-24}$$

$$g(x, y) = V \qquad |x| < a, |y| < a \tag{2-25}$$

$$L(f) = \int_{-a}^{a} dx' \int_{-a}^{a} dy' \frac{f(x', y')}{4\pi\varepsilon\sqrt{(x - x')^2 + (y - y')^2}} \tag{2-26}$$

Then $L(f) = g$ is equivalent to (2-17). A suitable inner product, satisfying (1-2) to (1-4), for which L is self-adjoint, is

$$\langle f, g \rangle = \int_{-a}^{a} dx \int_{-a}^{a} dy \, f(x, y) g(x, y) \tag{2-27}$$

To apply the method of moments, we use the functions (2-19) as a subsectional basis, and define testing functions

$$w_m = \delta(x - x_m)\delta(y - y_m) \tag{2-28}$$

which is the two-dimensional Dirac delta function. Now the elements of the $[l]$ matrix (1-25) are those of (2-22), and the $[g]$ matrix of (1-26) is

$$[g_m] = \begin{bmatrix} V \\ V \\ \vdots \\ V \end{bmatrix} \tag{2-29}$$

The matrix equation (1-24) is, of course, identical to the set of equations (2-21). In terms of the inner product (2-27), the capacitance (2-18) can be written

$$C = \frac{\langle \sigma, \phi \rangle}{V^2} \tag{2-30}$$

since $\phi = V$ on the plate. Equation (2-30) is the conventional stationary formula for the capacitance of a conducting body [3].

For numerical results, the l_{mn} of (2-22) must be evaluated. Let $2b = 2a/\sqrt{N}$ denote the side length of each Δs_n. The potential at the center of Δs_n due to unit charge density over its own surface is

$$\begin{aligned} l_{nn} &= \int_{-b}^{b} dx \int_{-b}^{b} dy \frac{1}{4\pi\varepsilon\sqrt{x^2+y^2}} \\ &= \frac{2b}{\pi\varepsilon}\ln(1+\sqrt{2}) = \frac{2b}{\pi\varepsilon}(0.8814) \end{aligned} \tag{2-31}$$

This derivation uses Dwight [4], 200.01 and 731.2. The potential at the center of Δs_m due to unit charge over Δs_n can be similarly evaluated, but the formula is complicated. For most purposes it is sufficiently accurate to treat the charge on Δs_n as if it were a point charge, and use

$$l_{mn} \approx \frac{\Delta s_n}{4\pi\varepsilon R_{mn}} = \frac{b^2}{\pi\varepsilon\sqrt{(x_m - x_n)^2 + (y_m - y_n)^2}} \qquad m \neq n \tag{2-32}$$

This approximation is 3.8 per cent in error for adjacent subsections, and has less error for nonadjacent ones. Table 2-1 shows capacitance, calculated by (2-23) using the α's obtained from the solution of (2-21), for various numbers of subareas. The second column of Table 2-1 uses the approximation (2-32), the third

TABLE 2-1. Capacitance of a Unit Square Plate (picofarads/meter)

No. of subareas	$C/2a$ approx. l_{mn}	$C/2a$ exact l_{mn}
1	31.5	31.5
9	37.3	36.8
16	38.2	37.7
36	39.2	38.7
100	39.8	39.5

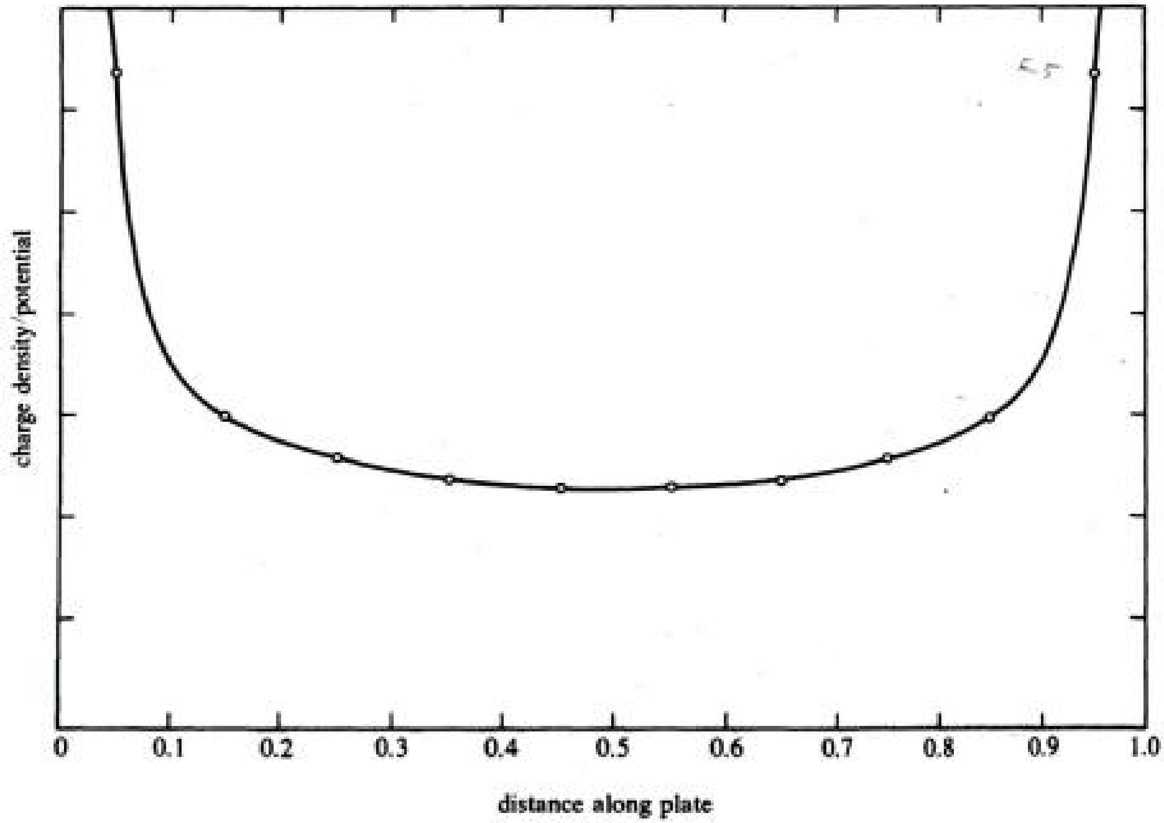


Figure 2-2. **Approximate charge density on subsections adjacent to the centerline of a square conducting plate.**

column uses an exact evaluation of the l_{mn}. A good estimate of the true capacitance is 40 picofarads. Figure 2-2 shows a plot of the approximate charge density along the subareas nearest the center line of the plate, for the case $N = 100$ subareas. Note that σ exhibits the well-known square root singularity at the edges of the plate.

Other geometries for which square subareas have been used to obtain numerical solutions are rectangular plates [2] and solid conducting cubes [5]. The related problem of a parallel-plate capacitor is treated in Section 2-4.

2-3. Conductors of Complex Shape

Often it is not possible to use square subareas for electrostatic problems. In this section we consider some simple approximations which enable almost any conducting body to be treated by subarea approximations.

First, consider the plane disk of radius r, with uniform charge density of unit amplitude. The electrostatic potential ϕ at its center is given by the simple integral

$$\phi = \int_0^{2\pi} d\theta \int_0^r \rho \, d\rho \frac{1}{4\pi\varepsilon\rho} = \frac{r}{2\varepsilon} \tag{2-33}$$

Let us compare this potential for a disk to that at the center of a square area with unit charge density and the same area A, given by (2-31). The result is

$$\phi_{\text{disk}} = \frac{\sqrt{A}}{\varepsilon}(0.2821)$$
$$\phi_{\text{square}} = \frac{\sqrt{A}}{\varepsilon}(0.2806) \qquad (2\text{-}34)$$

There is less than 0.54 per cent difference between the two. This is because the major contribution to ϕ is due to the charge in the immediate vicinity of the field point, and this is the same in each case. Hence if a subarea is not too narrow (has a reasonably large area/perimeter ratio), a good approximation to the diagonal elements of the $[l]$ matrix is

$$l_{nn} \approx \frac{0.282}{\varepsilon}\sqrt{A_n} \qquad (2\text{-}35)$$

where A_n is the area of the nth subarea. A useful approximation for the off-diagonal elements is the point-charge approximation of (2-32), which can be written in general as

$$l_{mn} \approx \frac{A_n}{4\pi\varepsilon R_{mn}} \qquad m \neq n \qquad (2\text{-}36)$$

where $R_{mn} = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2}$ is the distance between the centers of the mth and nth subareas. Approximation (2-36) cannot be used if the body has different areas very close together, as, for example, in the parallel-plate capacitor (see Section 2-4).

When the above approximations are not sufficiently accurate, the following procedure is convenient for calculating the l_{mn}. Figure 2-3 shows an elongated

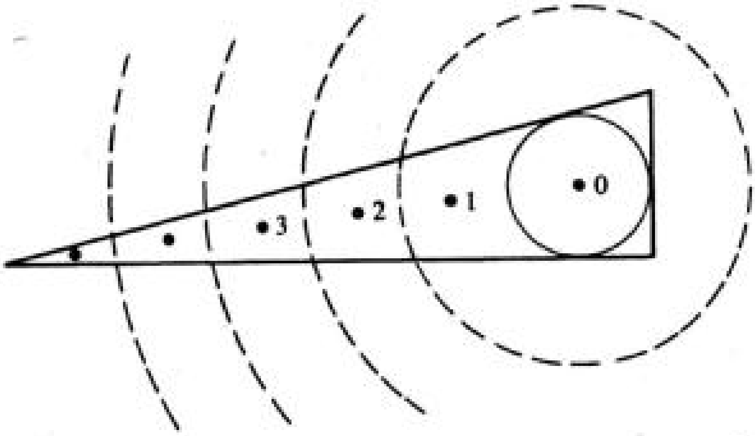


Figure 2-3. Numerical evaluation of l_{nn}.

triangular subarea. To evaluate l_{nn}, divide the area into a disk plus segments of circular annuli, as shown. Label these subsubareas 0, 1, 2,.... Then

$$l_{nn} = \frac{1}{\varepsilon}\left(0.282\sqrt{A_0} + \frac{1}{4\pi}\sum_i \frac{A_i}{R_{0i}}\right) \tag{2-37}$$

where A_0 is the area of the disk, the A_i ($i = 1, 2, \ldots$), are the areas of the annular segments, and R_{0i} is the distance from the center of the ith annulus to the center of the disk. Equation (2-37) is basically a numerical evaluation of the integral for l_{nn}. If the subarea is not planar, the subsubareas can be taken as those lying between concentric spheres. Evaluation of l_{mn} elements for very close subareas can be accomplished in a similar manner. For problems having rotational symmetry, it is sometimes convenient to take complete annular subareas, as demonstrated by the following example.

Example. Consider a hollow conducting tube of circular cross section and length L, as shown in Fig. 2-4. We wish to determine the electrostatic capacitance.

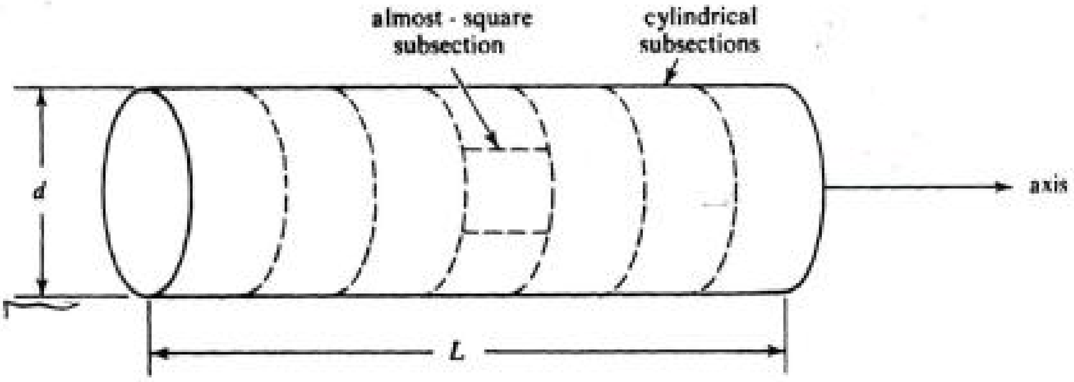


Figure 2-4. **Hollow conducting circular cylinder.**

The tube has rotational symmetry about its axis, and hence cylindrical subsections are convenient, as indicated on the figure. To evaluate the l_{mn}, each subcylinder can be further divided into smaller, almost square, subsections, as shown in Fig. 2-4. The l_{mn} for a point-matching solution are then evaluated by formulas similar to (2-37) as applied to the almost-square subsubsections. Note that for this problem all the l_{nn} are equal, and the l_{mn} depend only on $|m - n|$. Hence $[l]$ is a *Toeplitz matrix* [6].

Some numerical results are given in Table 2-2, calculated using 10 cylindrical subsections. The corresponding charge density was as expected, being almost uniform in the central region of a thin tube and singular at the ends. A similar problem, that of the capacitance of washer-type conducting plate, has been treated in the literature [7]. This latter problem was done using an analytical evaluation of the l_{mn} rather than a numerical one.

Table 2-2. Capacitance C(picofarads) for a Hollow Tube of Length 1 Meter, for Various Length/Diameter (L/d) Ratios

L/d	1	2	6	20	60
C	63	42	25	17	12

2-4. Arbitrary Excitation of Conductors

So far we have been considering only the specific problem of a charged conducting body. We now wish to take the more general viewpoint that the $[l]$ matrix characterizes the conducting body (or bodies) for any excitation. The excitation may be due to charge on the conductors or to external charges which produce an "impressed" field. The particular excitation enters only into the $[g]$ matrix of the method of moments, and hence $[l]$ depends only on the geometry of the conductors. Once the inverse matrix $[l^{-1}]$ is obtained, a specific solution is obtained by matrix multiplication according to (1-27).

To express these ideas in equation form, consider the general problem represented by Fig. 2-5. There are N conducting bodies, having net charges $q_1, q_2, \ldots, q_N$, and potentials $V_1, V_2, \ldots, V_N$. External to the conductors there may be additional sources which, in the absence of conductors, produce a potential ϕ^i (impressed field). The boundary condition is that ϕ^i plus the potential due to charges on the conductors must be constant on each conductor. In equation form, this is

$$\phi^i + \oiint_{\Sigma S_n} \frac{\sigma}{4\pi\varepsilon R}\, ds = \begin{cases} V_1 \text{ on } S_1 \\ V_2 \text{ on } S_2 \\ \vdots \\ V_N \text{ on } S_N \end{cases} \tag{2-38}$$

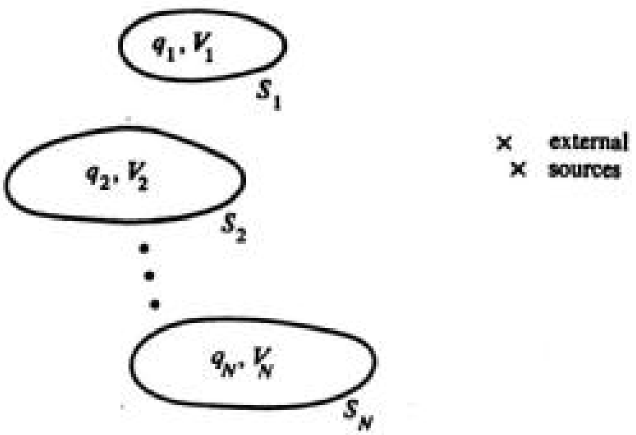


Figure 2-5. N charged conductors in the field of external sources.

where σ is the surface charge density on the conductors. The ϕ^i and V_n are assumed known, and (2-38) is an integral equation for σ. Equation (2-17) is the specialization of (2-38) to a charged conducting plate with no external sources. The total charge q_n instead of V_n may be specified on each conductor, in which case the V_n are treated as unknown constants in (2-38) to be obtained after σ is found.

Example. To illustrate these concepts, consider the two-body problem of parallel square conducting plates, as shown in Fig. 2-6. We here treat the case V_n specified on the plates but with no external sources ($\phi^i = 0$). The same plates in an impressed field are considered in Section 2-5.

Let both the top and bottom plates be divided into N square subsections, so that the total number of subsections is $2N$. The charge density is assumed constant on each subsection, and the total field is matched at the center of each subsection. The evaluation of the $[l]$ matrix follows the procedure of Section 2-2, and results in the following $2N$ by $2N$ matrix

$$[l] = \begin{bmatrix} [l^{tt}] & [l^{tb}] \\ [l^{bt}] & [l^{bb}] \end{bmatrix} \tag{2-39}$$

where t denotes "top plate" and b denotes "bottom plate." The N by N submatrices on the diagonal are single-plate matrices; hence

$$[l^{tt}] = [l^{bb}] = [l] \qquad \text{of Section 2-2} \tag{2-40}$$

The off-diagonal submatrices are the plate-to-plate matrices, which must be equal:

$$[l^{tb}] = [l^{bt}] \tag{2-41}$$

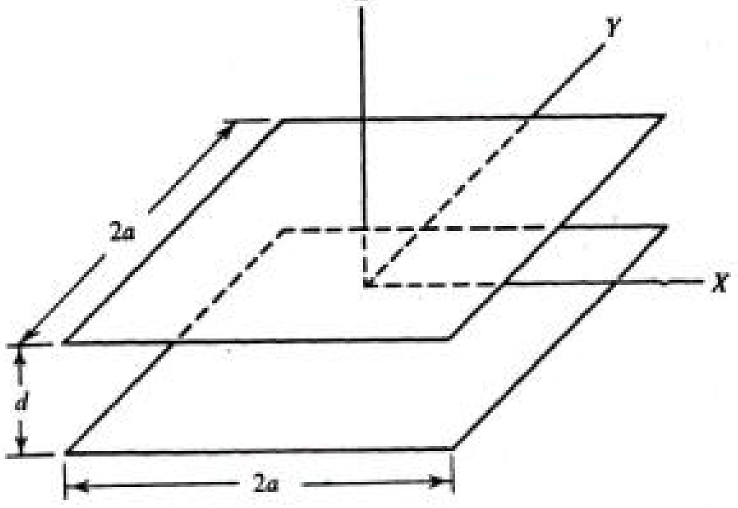


Figure 2-6. Parallel square conducting plates.

Let the elements l_{mn}^{tb} be ordered so that when $m = n$ the subareas are one on top of the other; that is, they coincide as $d \to 0$. Now if $m \neq n$, the point-charge approximation of (2-32) gives good results; that is,

$$l_{mn}^{tb} = \frac{b^2}{\pi\varepsilon\sqrt{(x_m - x_n)^2 + (y_m - y_n)^2 + d^2}} \tag{2-42}$$

When $m = n$, the square subsection can be approximated by a circular one of the same area, and the potential evaluated a distance d above it. The integration gives (2-33) with r replaced by $\sqrt{r^2 + d^2} - d$; hence the desired l_{nn}^{tb} is (2-35) modified by the ratio of these factors, or

$$l_{nn}^{tb} \approx \frac{0.282}{\varepsilon}(2b)\left[\sqrt{1 + \frac{\pi}{4}\left(\frac{d}{b}\right)^2} - \frac{\sqrt{\pi}\,d}{2b}\right] \tag{2-43}$$

6.37E10

This completes the evaluation of $[l]$.

Suppose we wish to evaluate the usual capacitance between the two plates. This corresponds to voltage $+V$ on the top plate and $-V$ on the bottom one. Hence the excitation matrix is

$$[g_m] = \begin{bmatrix} [g_m^t] \\ [g_m^b] \end{bmatrix} \tag{2-44}$$

where

$$[g_m^t] = -[g_m^b] = \begin{bmatrix} V \\ V \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \tag{2-45}$$

The α_n correspond to the charge densities on each subarea and are given by (1-27). However, for this problem, it is evident from symmetry that the charge density on the top plate is minus that on the bottom plate. Hence

$$[\alpha_n] = \begin{bmatrix} [\alpha_n^t] \\ [\alpha_n^b] \end{bmatrix} = \begin{bmatrix} [\alpha_n^t] \\ -[\alpha_n^t] \end{bmatrix} \tag{2-46}$$

and we can use this to reduce $[l][\alpha] = [g]$ to

$$[l_{mn}^{tt} - l_{mn}^{tb}][\alpha_n^t] = [g_m^t] \tag{2-47}$$

which is only an N by N matrix equation. The charge densities on the top plate are now found by inversion as

$$[\alpha_m^t] = [(l^{tt} - l^{tb})_{mn}^{-1}][g_n^t] \tag{2-48}$$

where $[g^t]$ is given by (2-45). The capacitance of the parallel-plate capacitor is

$$C = \frac{\text{charge on top plate}}{V}$$

$$= \frac{1}{V} \sum_{\text{top}} \alpha_n^t \, \Delta s_n \tag{2-49}$$

Since all the $\Delta s = 4b^2$ and all elements of $[g^t] = V$, this can be written

$$C = 4b^2 \sum_{mn} (l^{tt} - l^{tb})_{mn}^{-1} \tag{2-50}$$

which is simply $4b^2$ times the sum of all elements of $[(l^{tt} - l^{tb})^{-1}]$.

Computations for this case have been made and compared with other approximate solutions [8]. When fringing is neglected, the capacity is $C \approx \varepsilon A/d$. Figure 2-7 shows the results obtained from (2-50) for the case $N = 36$, normalized to $\varepsilon A/d$. It is interesting to note that, when d is as little as $0.05a$, neglecting fringing results in 6 per cent error. The error rapidly increases as d becomes larger, becoming 100 per cent as $d \to \infty$.

Now suppose we want the capacitance of the two plates when connected together. This is obtained by keeping both plates at the same potential V. Then, instead of (2-45), we have

$$[g_m^t] = [g_m^b] = \begin{bmatrix} V \\ V \\ \vdots \\ \vdots \end{bmatrix} \tag{2-51}$$

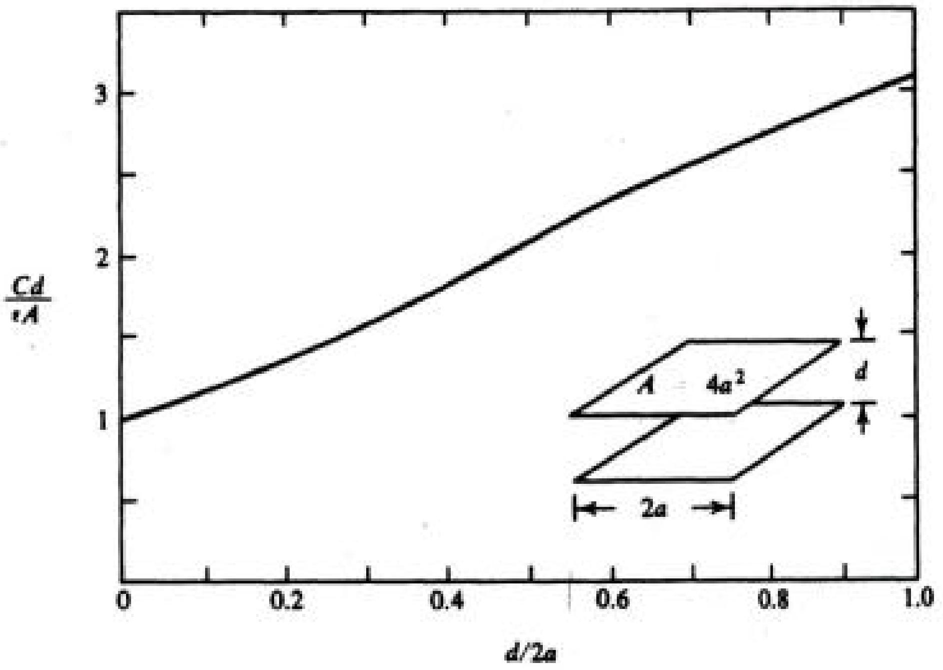


Figure 2-7. Capacitance of a square parallel-plate capacitor, normalized to $\varepsilon A/d$.

and, from symmetry, instead of (2-46),

$$[\alpha_n] = \begin{bmatrix} [\alpha_n^t] \\ [\alpha_n^t] \end{bmatrix} \tag{2-52}$$

Analogous to (2-47), the N by N matrix equation for the present excitation is

$$[l_{mn}^{tt} + l_{mn}^{tb}][\alpha_n^t] = [g_m^t] \tag{2-53}$$

and, analogous to (2-48), the solution is

$$[\alpha_m^t] = [(l^{tt} + l^{tb})_{mn}^{-1}][g_n^t] \tag{2-54}$$

The capacitance of the two plates connected together is then

$$C = \frac{\text{total charge}}{V}$$

$$= \frac{2}{V} \sum_{\text{top}} \alpha_n^t \, \Delta s_n \tag{2-55}$$

which can also be written in the form of (2-50) as

$$C = 8b^2 \sum_{mn} (l^{tt} + l^{tb})_{mn}^{-1} \tag{2-56}$$

Note that as $d \rightarrow 0$, $[l^{tt}] \rightarrow [l^{tb}]$ and C becomes the capacitance of a single plate (Section 2-2). As $d \rightarrow \infty$, $[l^{tb}] \rightarrow 0$, and C becomes twice the capacitance of a single plate.

2-5. Electric Polarizability

If a conducting body with no net charge is placed in a uniform electrostatic field, a net dipole moment **p** usually results. In general,

$$\mathbf{p} = \oint\!\!\oint_S \mathbf{r}\sigma \, ds \tag{2-57}$$

where $\mathbf{r} = \mathbf{u}_x x + \mathbf{u}_y y + \mathbf{u}_z z$ is the radius vector from the origin to a point on the surface S of the conductor, and $\sigma(x, y, z)$ is the surface charge density on S. The dipole moment is proportional to the impressed field $\mathbf{E}^i$ which produces σ; hence

$$\mathbf{p} = [\chi] \cdot \mathbf{E}^i \tag{2-58}$$

where $[\chi]$ is the *polarizability tensor*. Elements of $[\chi]$ may be found by applying a unit field **E** and evaluating components of **p**. For example, $\chi_{xy} = p_x$ for

$\mathbf{E} = \mathbf{u}_y$, and so on. The polarizability tensor is a useful quantity for the analysis of artificial dielectrics [9] and for scattering by small objects.

The appropriate integral equation is (2-38) specialized to a single conducting body S, which is

$$\oiint_S \frac{\sigma}{4\pi\varepsilon R}\, ds = V - \phi^i \tag{2-59}$$

where ϕ^i is a potential from which the electrostatic field is determined by $\mathbf{E}^i = -\nabla\phi^i$. The constant potential V must be obtained from the condition

$$\oiint_S \sigma\, ds = 0 \tag{2-60}$$

That is, the net charge on S is zero. Whenever $\mathbf{E}^i$ is perpendicular to a plane of reflection symmetry for the conductor, we can choose $\phi^i = 0$ on that plane and $V = 0$ in (2-59), which is equivalent to satisfying condition (2-60).

Example. Consider the parallel conducting plates of Fig. 2-6. We wish to determine the polarizability tensor when they are connected together, that is, maintained at the same potential. From symmetry considerations, it is apparent that an E_x will produce only a p_x, an E_y only a p_y, and an E_z only a p_z. Hence the polarizability tensor is diagonal:

$$[\chi] = \begin{bmatrix} \chi_{xx} & 0 & 0 \\ 0 & \chi_{yy} & 0 \\ 0 & 0 & \chi_{zz} \end{bmatrix} \tag{2-61}$$

and the x, y, and z axes are principal axes of $[\chi]$. Also, from symmetry, $\chi_{xx} = \chi_{yy}$ for square plates.

To evaluate χ_{zz}, take $\mathbf{E}^i = \mathbf{u}_z$ and $\phi^i = -z$. Note that $\phi^i = 0$ on $z = 0$, the plane of symmetry, and hence (2-60) will be satisfied. Now the integral equation (2-59) becomes

$$\oiint_S \frac{\sigma}{4\pi\varepsilon R}\, ds = \begin{cases} d/2 & \text{on top plate} \\ -d/2 & \text{on bottom plate} \end{cases} \tag{2-62}$$

This is the same integral equation as for the parallel-plate capacitor, except that V is replaced by $d/2$. Hence the charge distribution is given by (2-48), where

$$[g^t] = \begin{bmatrix} \frac{d}{2} \\ \frac{d}{2} \\ \vdots \end{bmatrix} \tag{2-63}$$

The polarizability is then found by approximating (2-57) by the summation

$$\chi_{zz} = p_z = \sum_{n=1}^{N} \left(\frac{d}{2} \alpha_n \, \Delta s_n + \frac{d}{2} \alpha_n \, \Delta s_n \right)$$

$$= \frac{d^2 \, \Delta s}{2} \sum_{mn} (l^{tt} - l^{tb})_{mn}^{-1} \tag{2-64}$$

where $\Delta s = 4b^2$. In terms of the capacitance between the parallel plates, (2-50),

$$\chi_{zz} = \tfrac{1}{2} d^2 C \tag{2-65}$$

This relationship between polarizability and capacitance results because of the parallel-plate nature of the problem, and does not result in general. As a check on (2-65), note that $q = CV = Cd/2$, and $p_z = qd = d^2C/2$. Note that when fringing can be neglected, (2-65) becomes

$$\chi_{zz} \approx \tfrac{1}{2} \varepsilon \, dA = \frac{\varepsilon}{2} (\text{volume}) \tag{2-66}$$

where A is the area of one plate and the volume is that between the plates.

For the other two elements $\chi_{xx} = \chi_{xy}$ of (2-61), let $\mathbf{E}^i = \mathbf{u}_x$ and $\phi^i = -x$. Again $\phi^i = 0$ on a plane of symmetry, whence (2-60) is satisfied. Now, instead of (2-62), the integral equation is

$$\oiint \frac{\sigma}{4\pi\varepsilon R} \, ds = -x \qquad \text{on the plates} \tag{2-67}$$

It is evident that the charge distribution is the same on both plates, and hence is given by (2-54) with

$$[g_n^t] = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} \tag{2-68}$$

where x_n is the x coordinate of the Δs_n subarea. The approximate evaluation of (2-57) then gives

$$\chi_{xx} = p_x = \sum_{n=1}^{N} x_n 2\alpha_n \, \Delta s_n$$

$$= 8b^2 \sum_{mn} x_m (l^{tt} + l^{tb})_{mn}^{-1} x_n \tag{2-69}$$

Another way of writing this result is

$$\chi_{xx} = 2b^2[\tilde{g}_m^t][(l^{tt} + l^{tb})_{mn}^{-1}][g_n^t] \tag{2-70}$$

where $\sim$ denotes transpose. This is a form that we shall encounter again in subsequent chapters.

2-6. Dielectric Bodies

The electrical state of a dielectric body in an electrostatic field is characterized by its polarization,

$$\mathbf{P} = \mathbf{D} - \varepsilon_0 \mathbf{E} = (\varepsilon - \varepsilon_0)\mathbf{E} \tag{2-71}$$

where ε is the capacitivity (permittivity) of the dielectric and ε_0 that of vacuum. The electric field due to the polarization is given by [10]

$$\mathbf{E}^p = \mathscr{E}(\mathbf{P}) = -\nabla\left(\iiint \frac{\mathbf{P}\cdot\mathbf{u}_R}{4\pi\varepsilon_0 R^2}\,d\tau\right) \tag{2-72}$$

where $\mathbf{u}_R$ is the unit vector pointing from the source point to the field point. Basically (2-72) is a superposition of the fields from all dipole elements $\mathbf{P}d\tau$ of source. The total field $\mathbf{E}^i + \mathbf{E}^p$ must satisfy (2-71) in the dielectric; hence an integral equation for $\mathbf{P}$ is

$$\mathbf{E}^i + \mathscr{E}(\mathbf{P}) = \frac{1}{\Delta\varepsilon}\mathbf{P} \tag{2-73}$$

where $\mathscr{E}$ is defined by (2-72) and $\Delta\varepsilon = \varepsilon - \varepsilon_0$.

A solution may be obtained by subsection and point-matching techniques. In canonical form (2-73) is

$$L(\mathbf{P}) = \mathscr{E}(\mathbf{P}) - \frac{1}{\Delta\varepsilon}\mathbf{P} = -\mathbf{E}^i \tag{2-74}$$

The functions in (2-74) are vectors, and require three numbers to represent them at a point. Following the method of moments, we use the following subsectional basis functions:

$$f_n = \begin{cases} (\mathbf{u}_x, \mathbf{u}_y, \mathbf{u}_z) & \text{in } \Delta\tau_n \\ (0, 0, 0) & \text{elsewhere} \end{cases} \tag{2-75}$$

where the **u**'s are coordinate unit vectors and $\Delta\tau_n$ is a representative volume element. The elements of the α_n coefficients of (1-21) can then be interpreted as the amplitude of the x, y, and z components of **P** in $\Delta\tau_n$; that is

$$\begin{aligned}\alpha_n &= (\alpha_{nx}, \alpha_{ny}, \alpha_{nz}) \\ &= \mathbf{P}(x_n, y_n, z_n) = \mathbf{P}_n \end{aligned} \tag{2-76}$$

where (x_n, y_n, z_n) are the coordinates of the center of $\Delta\tau_n$. Using the expansion (1-21) in (2-74), and matching the resultant equation at the centers of all $\Delta\tau_m$, we obtain the matrix equation

$$[l_{mn}][\mathbf{P}_n] = -[\mathbf{E}_m^i] \tag{2-77}$$

where $\mathbf{E}_m^i = \mathbf{E}^i(x_m, y_m, z_m)$. Each element of $[l]$ is a dyadic, of the form

$$l_{mn} = \begin{bmatrix} e_{mn}^{xx} - \dfrac{1}{\Delta\varepsilon} & e_{mn}^{xy} & e_{mn}^{xz} \\[2ex] e_{mn}^{yx} & e_{mn}^{yy} - \dfrac{1}{\Delta\varepsilon} & e_{mn}^{yz} \\[2ex] e_{mn}^{zx} & e_{mn}^{zy} & e_{mn}^{zz} - \dfrac{1}{\Delta\varepsilon} \end{bmatrix} \tag{2-78}$$

where the e_{mn} are derived from $\mathscr{E}$ in the same manner as the l_{mn} are derived from L. For a physical interpretation of the elements of (2-78), we note that e_{mn}^{xx} is the x component of **E** at (x_m, y_m, z_m) due to $\mathbf{P} = \mathbf{u}_x$ at (x_n, y_n, z_n), e_{mn}^{yx} is the y component of **E** due to the same **P**, etc. The solution to (2-77) is, of course, given by

$$[\mathbf{P}_m] = -[l_{mn}^{-1}][\mathbf{E}_n^i] \tag{2-79}$$

If m and n range from 1 to N, this is a $3N$ by $3N$ matrix equation due to the vector nature of **P** and **E**. Note that the e terms of (2-78) are independent of ε, which enters only into the $\Delta\varepsilon$ terms.

For crude solutions, the following approximations are often adequate. When $m \neq n$ each $\mathbf{P}_n\,\Delta\tau_n$ can be viewed as a point dipole, and **E** evaluated at $\Delta\tau_m$. The result is

$$e_{mn}^{ij} \approx \mathbf{u}_i \cdot e_{mn} \cdot \mathbf{u}_j \tag{2-80}$$

where i, j denote x, y, or z, and e_{mn} is the dyad

$$e_{mn} = -\frac{\Delta\tau_n}{4\pi\varepsilon_0}\,\nabla_m(\mathbf{R}/R^3) \tag{2-81}$$

where $\mathbf{R} = \mathbf{u}_x(x_m - x_n) + \mathbf{u}_y(y_m - y_n) + \mathbf{u}_z(z_m - z_n)$. When $m = n$ the field can be approximated by that at the center of a sphere having the same **P**. This results in

$$e_{nn}^{ii} \approx -\frac{1}{3\varepsilon_0} \tag{2-82}$$

and $e_{nn}^{ij} = 0$, $i \neq j$. For better results, the approximation (2-82) may be replaced by the field at the center of a spheroid or cylinder which approximates $\Delta\tau$. Still better results can be obtained by numerical integrations similar to those of Section 2-3.

The above solution remains valid for inhomogeneous dielectrics (ε a function of position), in which case the ε of each $\Delta\tau$ is taken to be that at its center. For *homogeneous* dielectrics, the problem can be formulated in terms of a surface distribution of bound charge [11], instead of a volume distribution of **P**. Since charge is a scalar quantity, this procedure materially reduces the number of unknowns in the matrix solution.

References

[1] R. F. Harrington, *Introduction to Electromagnetic Engineering*, McGraw-Hill Book Co., New York, 1958, pp. 117–120.

[2] D. Reitan and T. Higgins, "Accurate Determination of the Capacitance of a Thin Conducting Rectangular Plate," *A.I.E.E. Trans.*, Vol. **75**, Part I, Jan. 1957, pp. 761–766.

[3] J. Van Bladel, *Electromagnetic Fields*, McGraw-Hill Book Co., New York, 1964, p. 96.

[4] H. B. Dwight, *Tables of Integrals and Other Mathematical Data*, The Macmillan Co., New York, 1947.

[5] D. Reitan and T. Higgins, "Calculation of the Electrical Capacitance of a Cube," *Journ. Appl. Phys.*, Vol. **22**, No. 2, Feb. 1951, pp. 223–226.

[6] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*, University of California Press, Berkeley, 1958.

[7] T. Higgins and D. Reitan, "Calculation of the Capacitance of a Circular Annulus by the Method of Subareas," *A.I.E.E. Trans.*, Vol. **70**, 1951, pp. 926–933.

[8] D. K. Reitan, "Accurate Determination of the Capacitance of Rectangular Parallel-Plate Capacitors," *Journ. Appl. Phys.*, Vol. **30**, No. 2, Feb. 1959, pp. 172–176.

[9] R. E. Collin, *Field Theory of Guided Waves*, McGraw-Hill Book Co., New York, 1960, Chap. 12.

[10] R. Plonsey and R. E. Collin, *Principles and Applications of Electromagnetic Fields*, McGraw-Hill Book Co., New York, 1961, p. 75.

[11] Reference [3], pp. 73–77.

CHAPTER

3

Two-dimensional Electromagnetic Fields

3-1. Transverse Magnetic Fields

To avoid unnecessary details, we start our consideration of electromagnetic fields with two-dimensional problems. These can be thought of as three-dimensional problems for which there is no variation of field quantities with respect to one cartesian coordinate, taken to be the z coordinate. We postpone a general discussion of three-dimensional fields until Chapter 5, after we have treated a number of special cases.

An arbitrary electromagnetic field can be expressed as the sum of a *transverse magnetic* (TM) part and a *transverse electric* (TE) part. The TM part has only components of magnetic field **H** transverse to z, and the TE part has only components of **E** transverse to z. For two-dimensional fields in isotropic media, the TM part has only a z component of **E** and the TE part only a z component of **H**. In many cases the TM and TE parts can be treated separately, reducing the problem to a scalar problem. In this section we consider only TM fields, the TE case being considered in Section 3-4.

In general a time-harmonic electromagnetic field ($e^{j\omega t}$ time variation) satisfies the *Maxwell equations*

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} \tag{3-1}$$

$$\nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E} + \mathbf{J} \tag{3-2}$$

where **J** is the volume distribution of electric currents. For TM fields, assume that $\mathbf{E} = \mathbf{u}_z E_z(x, y)$, and similarly for $\mathbf{J}$. The Maxwell equations then lead to

$$\nabla^2 E_z + k^2 E_z = j\omega\mu J_z \tag{3-3}$$

where $k = \omega\sqrt{\varepsilon\mu} = 2\pi/\lambda$ is the wavenumber (λ = wavelength). Equation (3-3) is the two-dimensional *Helmholtz equation*. Solutions may be obtained by first finding the field from a two-dimensional point source, that is, a three-dimensional line source. The field at $\boldsymbol{\rho} = \mathbf{u}_x x + \mathbf{u}_y y$ due to a filament of current I at $\boldsymbol{\rho}' = \mathbf{u}_x x' + \mathbf{u}_y y'$ is [1]

$$E_z = \frac{-k\eta}{4} I H_0^{(2)}(k\,|\boldsymbol{\rho} - \boldsymbol{\rho}'|) \tag{3-4}$$

where $\eta = \sqrt{\mu/\varepsilon} \approx 120\pi$ is the intrinsic impedance of free space and $H_0^{(2)}$ is the Hankel function of the second kind, zero order. The E_z of (3-4) is the *Green's function* for the operator of (3-3). A general solution is then the superposition of E_z due to all elements of source $J_z\, ds$, or

$$E_z(\boldsymbol{\rho}) = \frac{-k\eta}{4} \iint J_z(\boldsymbol{\rho}') H_0^{(2)}(k\,|\boldsymbol{\rho} - \boldsymbol{\rho}'|)\, ds' \tag{3-5}$$

where the integration is over the cross section of the cylinder of currents J_z.

3-2. Conducting Cylinders, TM Case

Consider a perfectly conducting cylinder excited by an impressed electric field E_z^i, as represented by Fig. 3-1. The impressed field induces surface currents J_z on the conducting cylinder, which produce a scattered field E_z^s. The field due to J_z is given by (3-5) specialized to the cylinder surface C. The boundary condition is

$$E_z = E_z^i + E_z^s = 0 \qquad \text{on } C \tag{3-6}$$

that is, the tangential electric field vanishes on C. Hence, combining (3-5) and (3-6), we have the integral equation

$$E_z^i(\boldsymbol{\rho}) = \frac{k\eta}{4} \int_C J_z(\boldsymbol{\rho}') H_0^{(2)}(k\,|\boldsymbol{\rho} - \boldsymbol{\rho}'|)\, dl' \qquad \rho \text{ on } C \tag{3-7}$$

where $E_z^i(\boldsymbol{\rho})$ is known and J_z is the unknown to be determined.

The simplest numerical solution of (3-7) consists of using pulse functions for a basis and point matching for testing. To accomplish this, the scatterer contour C is divided into N segments ΔC_n and pulse functions defined as

$$f_n(\boldsymbol{\rho}) = \begin{cases} 1 & \text{on } \Delta C_n \\ 0 & \text{on all other } \Delta C_m \end{cases} \tag{3-8}$$

Letting $J_z = \sum \alpha_n f_n$, substituting in (3-7), and satisfying the resultant equation at the midpoint (x_m, y_m) of each ΔC_m, we obtain the matrix equation

$$[l_{mn}][\alpha_n] = [g_m] \tag{3-9}$$

where the elements of $[\alpha_n]$ are the α_n coefficients, the elements of $[g_m]$ are

$$g_m = E_z^i(x_m, y_m) \tag{3-10}$$

and the elements of $[l_{mn}]$ are

$$l_{mn} = \frac{k\eta}{4}\int_{\Delta C_n} H_0^{(2)}[k\sqrt{(x - x_m)^2 + (y - y_m)^2}]\, dl \tag{3-11}$$

A solution for the current is then given by $J_z = [\tilde{f}_n][l_{nm}^{-1}][g_m]$, as discussed in Section 1-3.

There is no simple analytic expression for the integral (3-11), but we can evaluate it by various approximations. The crudest approximation is to treat an element $J_z\, \Delta C_n$ as a filament of current when the field point is not on ΔC_n; that is,

$$l_{mn} \approx \frac{\eta}{4} k\, \Delta C_n H_0^{(2)}[k\sqrt{(x_n - x_m)^2 + (y_n - y_m)^2}] \tag{3-12}$$

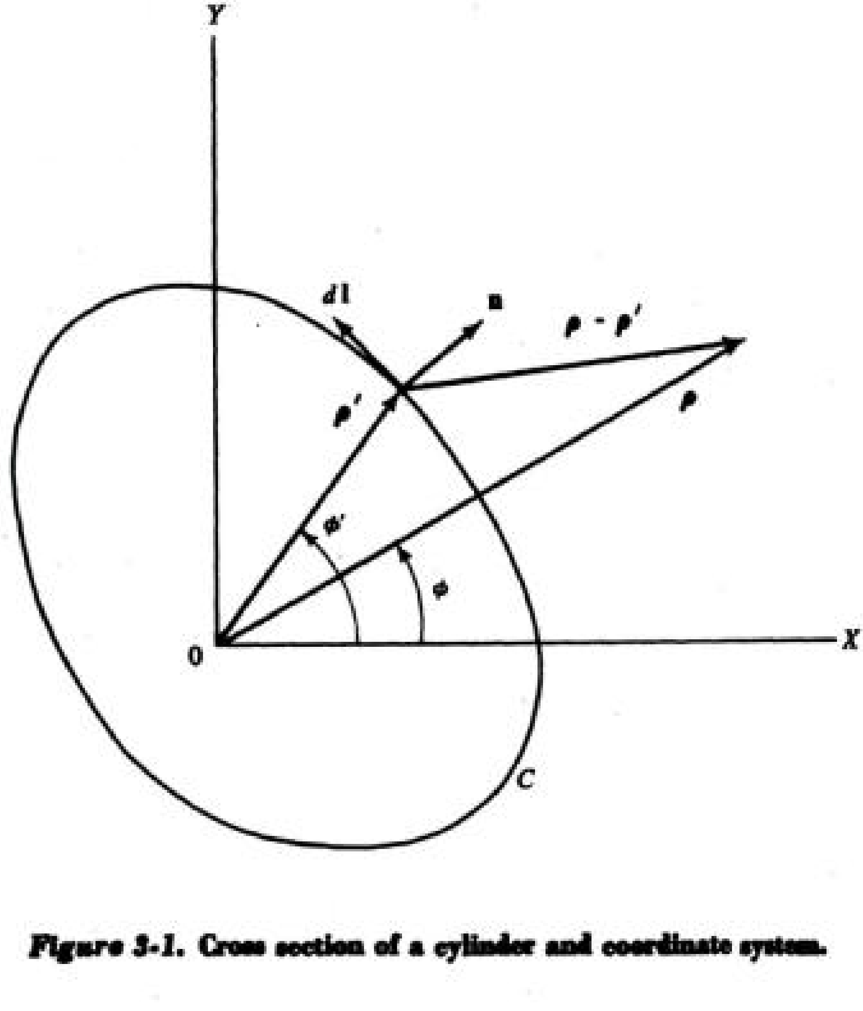


Figure 3-1. Cross section of a cylinder and coordinate system.

when $m \neq n$. For the diagonal elements l_{nn} the Hankel function has an integrable singularity, and the integral must be evaluated analytically. For this, we approximate ΔC_n by a straight line and use the small argument formula

$$H_0^{(2)}(z) \approx 1 - j\frac{2}{\pi}\log\left(\frac{\gamma z}{2}\right) \tag{3-13}$$

where $\gamma =$ 1.781 ... is Euler's constant. An evaluation of (3-11) then gives

$$l_{nn} \approx \frac{\eta}{4}k\,\Delta C_n\left[1 - j\frac{2}{\pi}\log\left(\frac{\gamma k\,\Delta C_n}{4e}\right)\right] \tag{3-14}$$

where $e = 2.718$ ... The approximations (3-12) and (3-14) are analogous to those used in Section 2-2 for electrostatic problems. Better approximations for the present problem will be discussed in Section 3-3.

Example. Consider TM plane-wave scattering by conducting cylinders [2,3]. In this case the impressed field is a uniform plane wave, which, if incident from the direction ϕ_i, is given by

$$E_z^i = e^{jk(x\cos\phi_i + y\sin\phi_i)} \tag{3-15}$$

This determines the excitation $[g_m]$ according to (3-10). An approximate evaluation of $[l_{mn}]$ is given by (3-12) and (3-14). The solution for J_z is then found by matrix inversion in the usual manner.

A parameter of interest is the *scattering cross section* σ, defined as the width (area in three-dimensional problems) for which the incident wave carries sufficient power to produce, by omnidirectional radiation, the same scattered power density in a given direction. In equation form, this is

$$\sigma(\phi) = 2\pi\rho\left|\frac{E^s(\phi)}{E^i}\right|^2 \tag{3-16}$$

where $E^s(\phi)$ is the distant field from J_z. It can be found by using the asymptotic expression for $H_0^{(2)}$ in (3-5). The result is [1]

$$E^s(\phi) = \eta k K \int_C J_z(x', y')e^{jk(x'\cos\phi + y'\sin\phi)}\,dl' \tag{3-17}$$

where

$$K(\rho) = \frac{1}{\sqrt{8\pi k\rho}}e^{-j(k\rho + 3\pi/4)} \tag{3-18}$$

Substituting (3-15) and (3-17) in (3-16), we obtain

$$\sigma(\phi) = \frac{k\eta^2}{4} \left| \int_C J_z(x', y') e^{jk(x' \cos\phi + y' \sin\phi)}\, dl' \right|^2 \tag{3-19}$$

This can be evaluated numerically once J_z is found.

A particularly descriptive form for the evaluation of (3-19) is obtained as follows. Let the integral be approximated by a sum over all ΔC_n, with $J_z = \alpha_n$, $x = x_n$, $y = y_n$ in the integrand for each ΔC_n. The result is

$$\sigma(\phi_i, \phi_s) = \frac{k\eta^2}{4} |[\tilde{V}^s_n][Z^{-1}_{nm}][V^i_m]|^2 \tag{3-20}$$

where $[V^i_m]$ is an "excitation" voltage matrix

$$[V^i_m] = [\Delta C_m e^{jk(x_m \cos\phi_i + y_m \sin\phi_i)}] \tag{3-21}$$

$[Z_{mn}]$ is a scatterer "impedance" matrix

$$[Z_{mn}] = [\Delta C_m l_{mn}] \tag{3-22}$$

and $[V^s_n]$ is a "measurement" voltage matrix.

$$[V^s_n] = [\Delta C_n e^{jk(x_n \cos\phi_s + y_n \sin\phi_s)}] \tag{3-23}$$

where $\phi = \phi_s$ is the angle at which σ is evaluated. We shall encounter this form again in Section 3-6 and subsequent chapters, it being a special case of the generalized network parameters discussed in Chapter 5. Note that (3-20) obeys the reciprocity relationship $\sigma(\phi_i, \phi_s) = \sigma(\phi_s, \phi_i)$; that is, the scattering cross section is unchanged if the transmitter and receiver are interchanged.

A number of computations have been made for rectangular conducting cylinders using approximations similar to those above [2]. A more accurate numerical evaluation of the integral equation was used by Andreasen to compute solutions for cylinders of other shapes [3]. It should be pointed out that the approximations made above will not converge to the exact solution as N is increased, because the l_{mn}, $m \neq n$, are not exact in the limit. The solution will converge to the exact solution if (3-12) is replaced by a more accurate approximation. To illustrate the accuracy that can be obtained using the simple approximations of this section, Fig. 3-2 shows the magnitude of the current on an ellipse as computed by Andreasen and by the formulas of this section. It is interesting to note that if the current is calculated by $\mathbf{n} \times \mathbf{H}$ on C instead of using the α_n, a better solution is obtained, as indicated in Fig. 3-2. The scattering cross section, as computed by Andreasen and by the above formulas, is illustrated by Fig. 3-3.

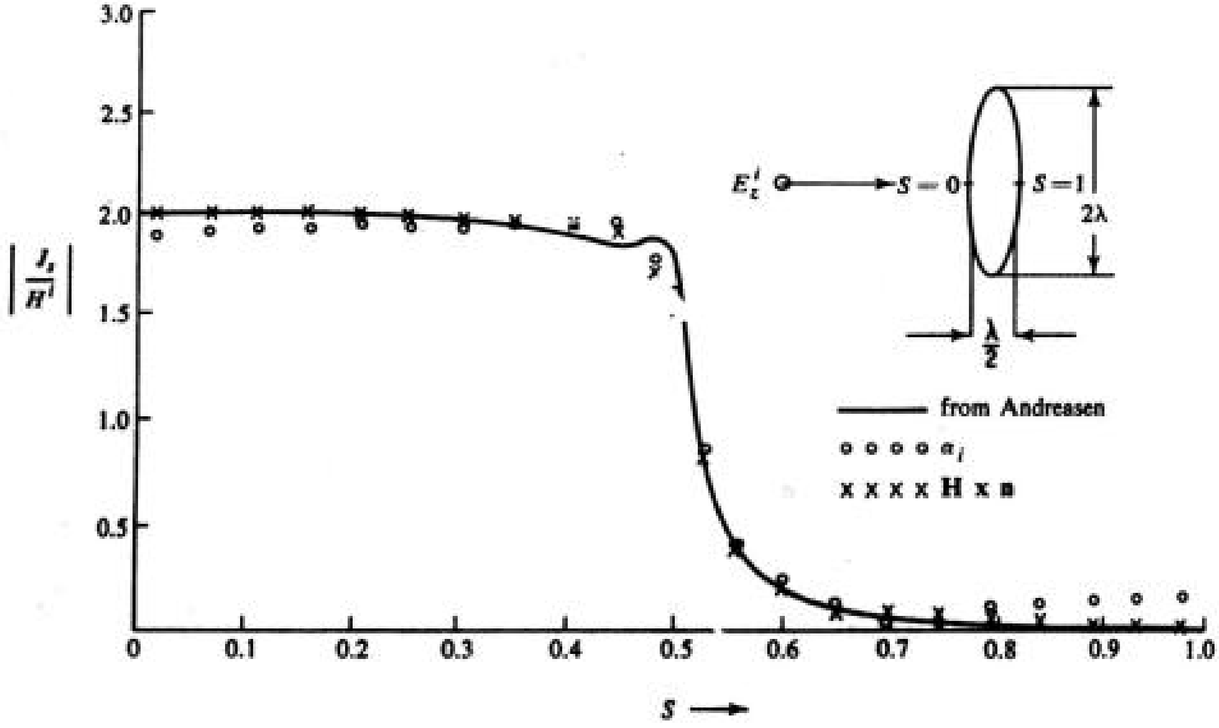


Figure 3-2. Current density on a conducting elliptic cylinder excited by a plate wave, TM case.

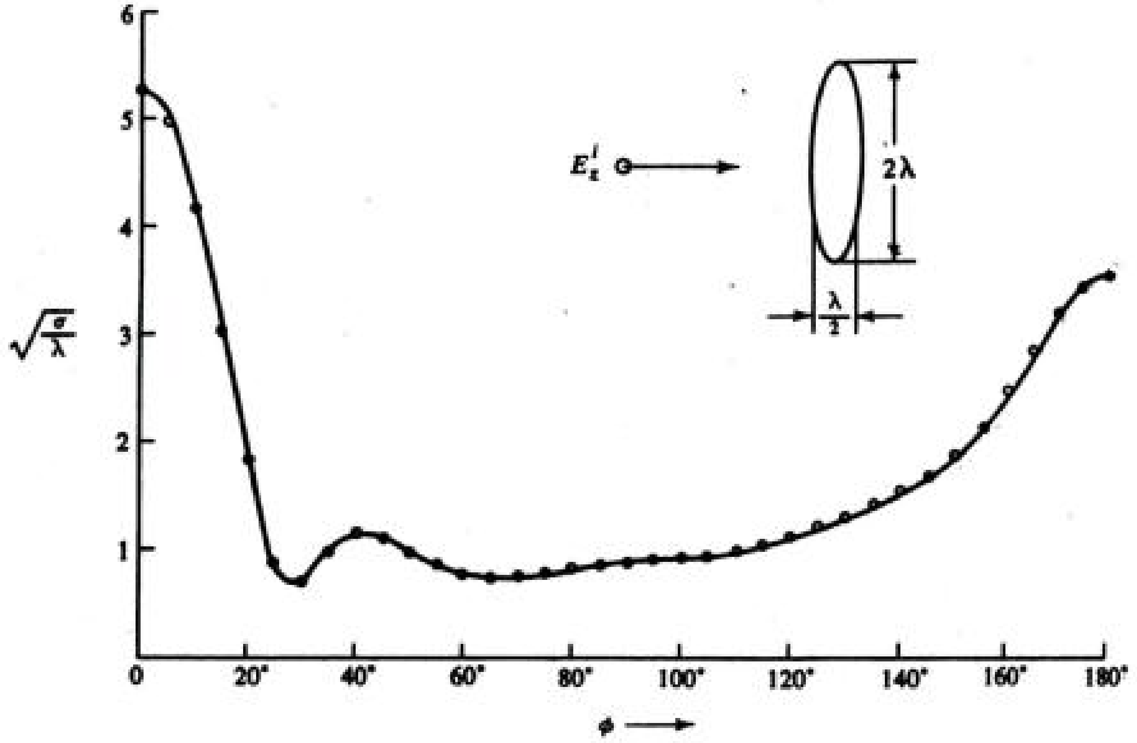


Figure 3-3. Scattered field pattern for a conducting elliptic cylinder excited by a plane wave, TM case.

Note that the two results are almost identical, even though the currents (Fig. 3-2), differ appreciably. This is because $\sqrt{\sigma}$ is a continuous linear functional of J, and hence is insensitive to small variations in J about its true value (Section 1-8).

3-3. Various Approximations

The accuracy of a solution and the rate of convergence depend upon the approximations made. The solution of Section 3-2 can be improved by more accurate evaluation of the l_{mn}, as follows. For the l_{nn}, additional terms can be included in (3-13), but this will not appreciably affect convergence, since (3-14) is exact in the limit $\Delta C_n \to 0$. For the l_{mn} terms, $m \neq n$, we can expand the integrand of (3-11) in a Taylor series about (x_n, y_n), and integrate the dominant terms analytically. This will give both improved accuracy and convergence to the exact solution as $N \to \infty$.

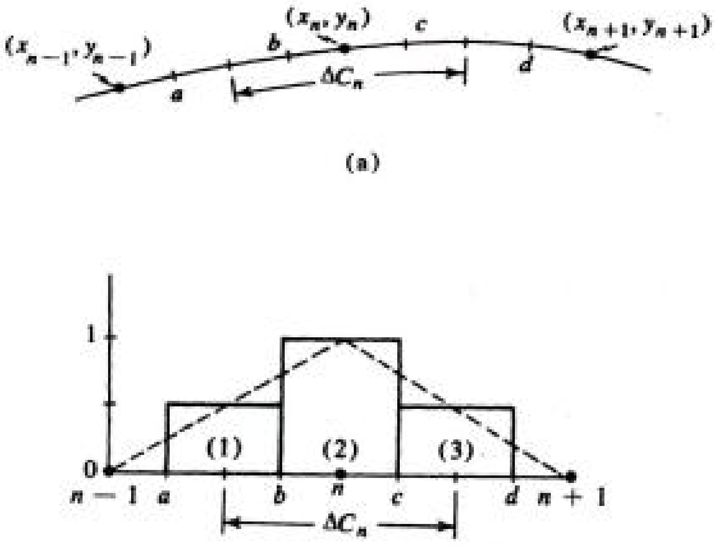
It has been found that the rate of convergence is almost twice as fast if a piecewise linear approximation to J_z is used instead of the step approximation. In other words, the Nth-order linear solution gives about the same accuracy as the $2N$th-order step solution. For a piecewise linear solution, instead of the steps of (3-8) we use the triangles of (1-50), as discussed in Section 1-5. The evaluation of the l_{mn} proceeds similarly to that for the pulse functions [4].

Solutions have also been obtained by Galerkin's method, using pulses for both expansion and testing functions. It was found that, for solutions of the subsectional-basis type, the accuracy and convergence of the Galerkin solution were about the same as for the point-matching solution. The Galerkin method apparently has its greatest utility in perturbational solutions, that is, when the solution is represented by only one expansion function, or by a few functions.

Perhaps the most convenient way of obtaining better approximations when using computers is to numerically evaluate the l_{mn}. For this, we divide each ΔC_n into smaller subintervals, and approximate the integral over each subinterval by (3-12) if nonsingular and by (3-14) if singular. To be explicit, let Fig. 3-4(a) represent a small section of the contour of a cylindrical conductor. Let the subintervals ΔC_{n-1}, ΔC_n, and ΔC_{n+1} be further subdivided as indicated by points a, b, c, and d. Figure 3-4(b) shows the same contour straightened out, and an expansion function constructed of three pulses. This three-stepped function approximates a triangle function, shown dashed. Now, remembering that each l_{mn} represents the field $-E_z$ at (x_m, y_m) due to expansion function f_n at (x_n, y_n), we can easily justify that, for $m = n$,

$$l_{mn} = (\tfrac{1}{2}l_{21} + l_{22} + \tfrac{1}{2}l_{23})_{mn} \tag{3-24}$$

where l_{21} and l_{23} are given by (3-12) with ΔC_n replaced by C_{ab} and C_{cd}, and l_{22} is given by (3-14) with ΔC_n replaced by C_{bc} (see Fig. 3-4). The factors 1/2 in the first and third terms of (3-24) arise from the fact that the pulses over C_{ab} and C_{cd} are one half the amplitude of the pulse over C_{bc}. For the l_{mn} elements, $m \neq n$, the



(a)

(b)

Figure 3-4. (*a*) Section of the contour. (*b*) Expansion function consisting of three constrained pulses.

procedure is the same, except that (3-12) is used for all l_{ij} since the field point never coincides with the source point.

To illustrate the accuracy obtainable with the above procedure, Fig. 3-5 shows the resultant current compared with Andreasen's results [3]. Note that we have taken smaller ΔC's in the region of rapid curvature on the ellipse for better accuracy. It was found that when point m was distant from point n, say

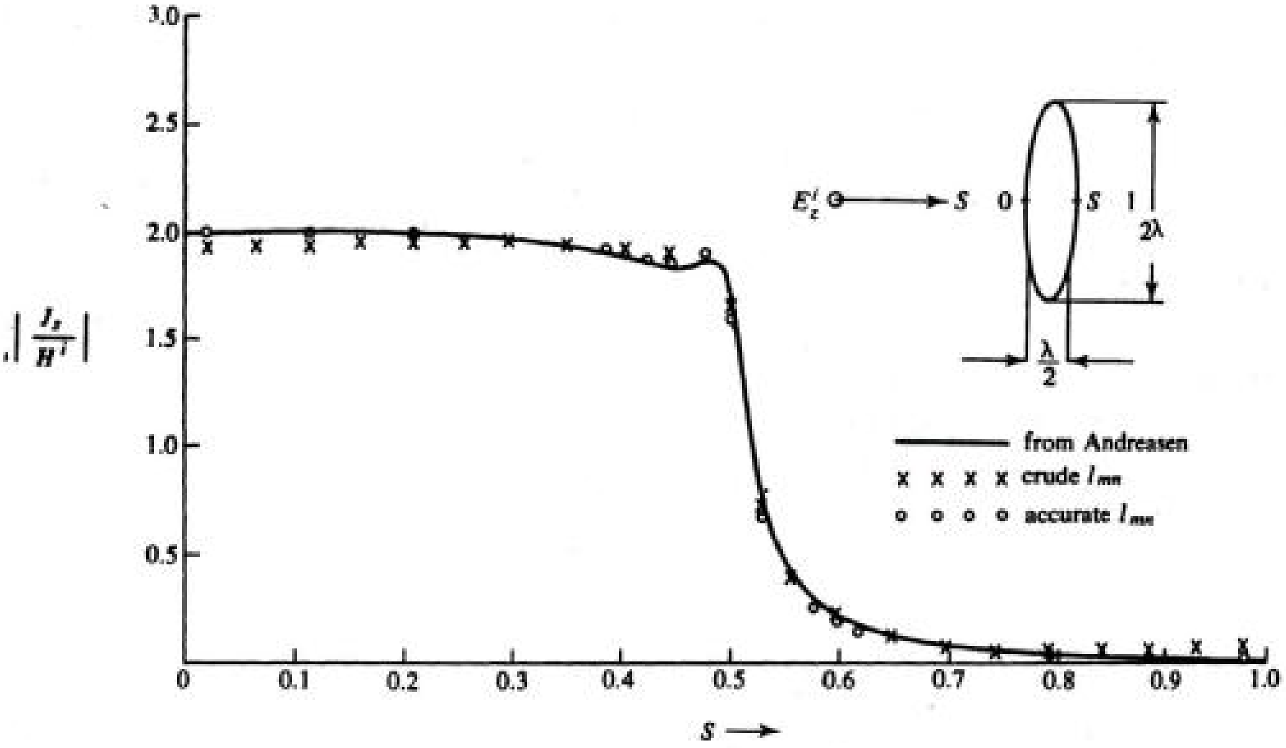


Figure 3-5. Current density on a conducting elliptic cylinder excited by a plane wave, using constrained pulses, TM case.

$|\rho_m - \rho_n| > \lambda/4$, we can use (3-12) instead of (3-24) with no appreciable loss in accuracy. In other words, it is more important to evaluate l_{mn} carefully for ΔC's close together than for distant ones. The use of expansion functions of the type shown in Fig. 3-4 is equivalent to dividing the conductor into $2N$ segments, and constraining every other α_n to be the average of its adjacent α_n's before inverting the $[l_{mn}]$ matrix. We can, of course, use more pulses to approximate a triangle function, but, judging from the accuracy of Fig. 3-5, this probably is unnecessary for most purposes.

If we wish an approximation to the Galerkin solution, instead of the point-matching solution, the functions of Fig. 3-4 can be used for both expansion and testing. However, instead of analytically evaluating the second integration, we can numerically evaluate it using approximations (3-12) and (3-14). The result is

$$l_{mn} = [\tfrac{1}{2}l_{22} + \tfrac{1}{4}(l_{12} + l_{21} + l_{23} + l_{32}) + \tfrac{1}{8}(l_{11} + l_{13} + l_{31} + l_{33})]_{mn} \qquad (3\text{-}25)$$

where the l_{ij} are the same l_{ij} that appear in (3-24). One factor of 1/2 comes from the fact that ΔC for each component pulse is 1/2 of ΔC_n, other factors of 1/2 come from the fact that the two end pulses are 1/2 the amplitude of the central pulse (Fig. 3-4). It is apparent from the forms of (3-24) and (3-25) that there will be little difference between the two l_{mn}, and hence between the two solutions. Of course, in the Galerkin solution the g_m of (3-10) should also be modified to represent a numerical integration of E_z^i with the testing function of Fig. 3-4.

If the conductor is symmetrical about some axis, as is the ellipse, the problem can be reduced to two matrices of order $N/2$, instead of a single matrix of order N. Since the time required to invert a matrix is proportional to N^3, this reduces the matrix inversion time to one fourth the original time. The procedure is discussed in the literature [3,4]. Finally, if the incident field E_z^i is also symmetrical about the same axis as is the conductor, only a single matrix of the order $N/2$ need be inverted.

3-4. Transverse Electric Fields

A two-dimensional TE field in isotropic media has no z component of **E** and only a z component of **H**. The most convenient general expression for the field is in terms of potentials[1]

$$\mathbf{H} = \frac{1}{\mu}\nabla \times \mathbf{A} \qquad (3\text{-}26)$$

$$\mathbf{E} = -j\omega\mathbf{A} - \nabla\Phi \qquad (3\text{-}27)$$

[1] In reference [1] the vector potential is defined so that $\mu\mathbf{A}$ replaces $\mathbf{A}$ in (3-26) to (3-28). We denote the scalar potential by Φ and the charge density by q to avoid confusion with the polar coordinates ρ and ϕ.

where the *magnetic vector potential* **A** and the *electric scalar potential* Φ satisfy

$$\nabla^2 \mathbf{A} + k^2 \mathbf{A} = -\mu \mathbf{J} \tag{3-28}$$

$$\nabla^2 \Phi + k^2 \Phi = -\frac{q}{\varepsilon} \tag{3-29}$$

The electric charge density q is related to **J** by the *equation of continuity*

$$\nabla \cdot \mathbf{J} = -j\omega q \tag{3-30}$$

Both (3-28) and (3-29) are Helmholtz equations, the same as (3-3), and hence solutions are of the form (3-5). Defining the two-dimensional Green's function

$$G(\boldsymbol{\rho}, \boldsymbol{\rho}') = \frac{1}{4j} H_0^{(2)}(k\,|\boldsymbol{\rho} - \boldsymbol{\rho}'|) \tag{3-31}$$

we can express solutions to (3-28) and (3-29) in unbounded two-dimensional space as [1]

$$\mathbf{A}(\boldsymbol{\rho}) = \mu \iint \mathbf{J}(\boldsymbol{\rho}') G(\boldsymbol{\rho}, \boldsymbol{\rho}')\, ds' \tag{3-32}$$

$$\Phi(\boldsymbol{\rho}) = \frac{1}{\varepsilon} \iint q(\boldsymbol{\rho}') G(\boldsymbol{\rho}, \boldsymbol{\rho}')\, ds' \tag{3-33}$$

where the integration is over a z = constant cross section of the cylinder. In evaluating the formulas of this section it should be remembered that all quantities are independent of z; hence all z derivatives are zero.

3-5. Conducting Cylinders, TE Case

Let the conducting cylinder of Fig. 3-1 be excited by an impressed TE field. We wish to determine the current on the cylinder and the field produced by this current. This problem can be solved by enforcing the condition tangential $E = 0$ on C, as shown in Section 3-6, but first we consider the H-field formulation used in the literature [2,3].

As discussed in Section 3-4, the TE field has only a z component of **H**, and a transverse component of **J**. The total magnetic field H_z at any point is the sum of the impressed field H_z^i plus the scattered field H_z^s due to **J** on C; that is,

$$H_z = H_z^i + H_z^s \tag{3-34}$$

The scattered field is related to its source **J** by (3-26) and (3-32), or

$$H_z^s = \mathbf{u}_z \cdot \nabla \times \int_C JG\, d\mathbf{l}' \tag{3-35}$$

where the vector $d\mathbf{l}'$ designates the reference direction of J. The field H_z is finite external to C, zero internal to C, and the discontinuity of H_z on C equals the current density. If the interior of C lies on the left side of $d\mathbf{l}$ (right-hand rule), then

$$J = -[H_z]_{C_+} \tag{3-36}$$

where the C_+ denotes that H_z is evaluated just external to C. Specializing (3-34) to C_+, we have

$$J = -\left[H_z^i + \mathbf{u}_z \cdot \nabla \times \int_C JG\, d\mathbf{l}'\right]_{C_+} \tag{3-37}$$

which is an equation for the unknown current J. Equation (3-37) differs from the classical integral equation in that a derivative operator as well as an integral operator is present.

Because of the discontinuity in H_z at C we have to be particularly careful in evaluating (3-37). The Green's function G is singular, and a simple interchange of differentiation and integration is not always possible [5]. Figure 3-6 shows an expanded view of the conductor boundary to help clarify these concepts. The contour C lies on the current sheet, C_+ lies just outside, and C_- just inside. At point a on C_+, $H_z = -J$, and at point b on C_-, $H_z = 0$. If the scatterer is a conducting sheet of infinitesimal thickness, it should be treated as the limit of one of finite thickness.

We can write (3-37) in general operator notation as

$$L(J) = -H_z^i \tag{3-38}$$

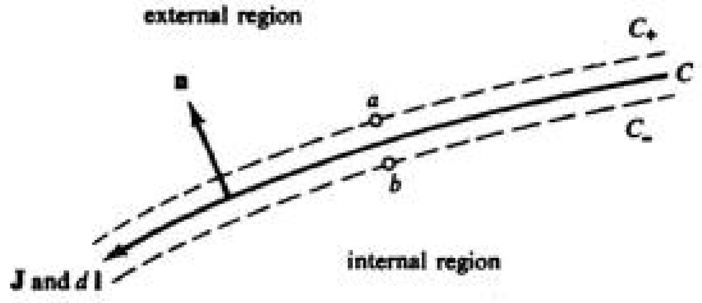


Figure 3-6. Section of cylindrical boundary.

where

$$L(J) = J + \left[\mathbf{u}_z \cdot \nabla \times \int_C JG\, dl'\right]_{C_+} \tag{3-39}$$

and proceed according to the method of moments. Again the simplest approximation is to use the pulses (3-8) as basis functions, and point matching for testing. The current is then given by $J = \sum \alpha_n f_n$, and the resulting matrix equation is (3-9) with

$$g_m = -H_z^i(x_m, y_m) \tag{3-40}$$

$$l_{mn} = \delta_{mn} + H_z(m, n) \tag{3-41}$$

where δ_{mn} is the Kronecker delta and $H_z(m, n)$ denotes H_z at (x_m, y_m) on C_+ due to unit current density on ΔC_n at (x_n, y_n). Figure 3-7 represents a typical current element $J\mathbf{l} = \Delta \mathbf{C}_n$ and local coordinates (x, y). From symmetry, and the fact that the discontinuity in H_z is J, we have

$$H_z\Big|_{\substack{x=0+\\ y=0}} = -H_z\Big|_{\substack{x=0-\\ y=0}} = -1/2 \tag{3-42}$$

and hence, by (3-41),

$$l_{nn} \approx 1/2 \tag{3-43}$$

If $\Delta C_n \ll \lambda$ and the field point (x, y) is distant from $J\mathbf{l} = \Delta \mathbf{C}_n$, then the source

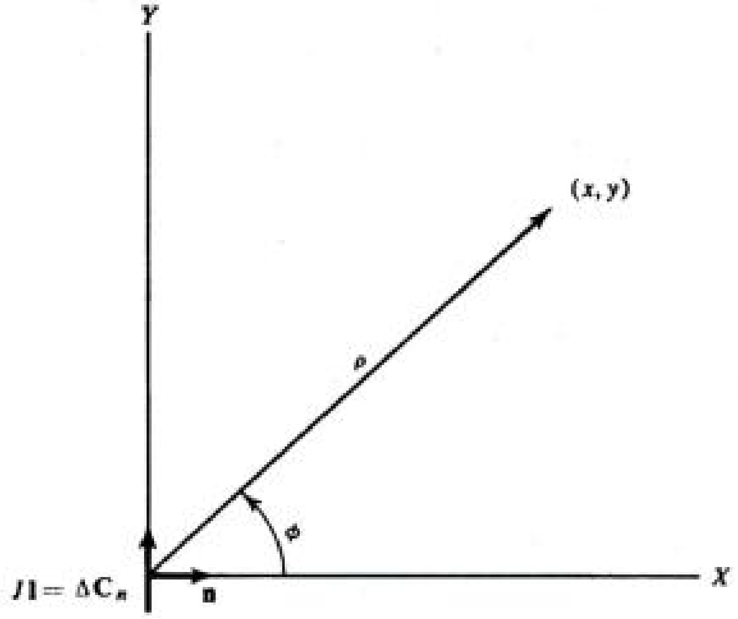


Figure 3-7. Element of current $J\mathbf{l}$ and local coordinates.

behaves as a point source. From (3-32)

$$A_y = \frac{\mu\,\Delta C_n}{4j} H_0^{(2)}(k\rho) \tag{3-44}$$

and from (3-26)

$$\begin{aligned} H_z &= \frac{\Delta C_n}{4j}\frac{\partial}{\partial x} H_0^{(2)}(k\rho) \\ &= \frac{j}{4} k\,\Delta C_n \cos\phi\, H_1^{(2)}(k\rho) \end{aligned} \tag{3-45}$$

where $H_1^{(2)}$ is the Hankel function of order 1. We can translate this to an arbitrary origin by replacing ρ by $|\rho_m - \rho_n|$ and $\cos\phi$ by $\mathbf{n}\cdot\mathbf{R}$, where

$$\mathbf{R} = \frac{\boldsymbol{\rho}_m - \boldsymbol{\rho}_n}{|\boldsymbol{\rho}_m - \boldsymbol{\rho}_n|} \tag{3-46}$$

is a unit vector from the source point (x_n, y_n) to the field point (x_m, y_m). This result can be used as an approximation for all $m \neq n$. Hence (3-41) becomes, for $m \neq n$,

$$l_{mn} \approx \frac{j}{4} k\,\Delta C_n(\mathbf{n}\cdot\mathbf{R}) H_1^{(2)}(k|\boldsymbol{\rho}_m - \boldsymbol{\rho}_n|) \tag{3-47}$$

The solution is then given by $J = [\tilde{J}_n][l_{nm}^{-1}][g_m]$, as discussed in Section 1-3.

For better approximations we can use the methods of Section 3-3 to obtain more accurate l_{mn}. For example, the pulse approximation to a triangle function, Fig. 3-4, can be used, with the new l_{mn} given by (3-24). Alternatively, the approximate triangle function can be used for both expansion and testing, giving the Galerkin result (3-25). Still more accurate evaluation of the l_{mn} may be required to treat thin conducting sheets when points m and n are close together.

Example. Consider TE plane-wave scattering by conducting cylinders. An impressed uniform plane wave incident from the direction ϕ_i is given by

$$H_z^i = e^{jk(x\cos\phi_i + y\sin\phi_i)} \tag{3-48}$$

The g_m are determined from this by (3-40), and the l_{mn} are given by (3-43) and (3-47) for a first-order solution. The current is then found by matrix inversion and multiplication in the usual manner.

Again the scattering cross section σ is of interest, given by

$$\sigma(\phi) = 2\pi\rho \left| \frac{H^s(\phi)}{H^i} \right|^2 \qquad (3\text{-}49)$$

analogous to (3-16). Here $H^s(\phi)$ is the distant field from J, obtainable by using the asymptotic formula for $H_1^{(2)}$ in (3-45), and summing over all elements of source. This gives [3]

$$H_z^s(\phi) = Kk \int_C J(x', y')\mathbf{n} \cdot \mathbf{R} e^{jk(x' \cos\phi + y' \sin\phi)}\, dl' \qquad (3\text{-}50)$$

where K is given by (3-18). Substituting (3-48) and (3-50) in (3-49), we obtain

$$\sigma(\phi) = \frac{k}{4} \left| \int_C J(x', y')\mathbf{n} \cdot \mathbf{R} e^{jk(x' \cos\phi + y' \sin\phi)}\, dl' \right|^2 \qquad (3\text{-}51)$$

which can be evaluated once J is found. The numerical evaluation of (3-51) can be put in a form similar to (3-20) for computational convenience.

To illustrate a typical result, Fig. 3-8 shows the TE solution for the current induced on the same elliptic cylinder as in Fig. 3-2 for the TM case. The computations are those of Andreasen [3], and correspond in accuracy to using

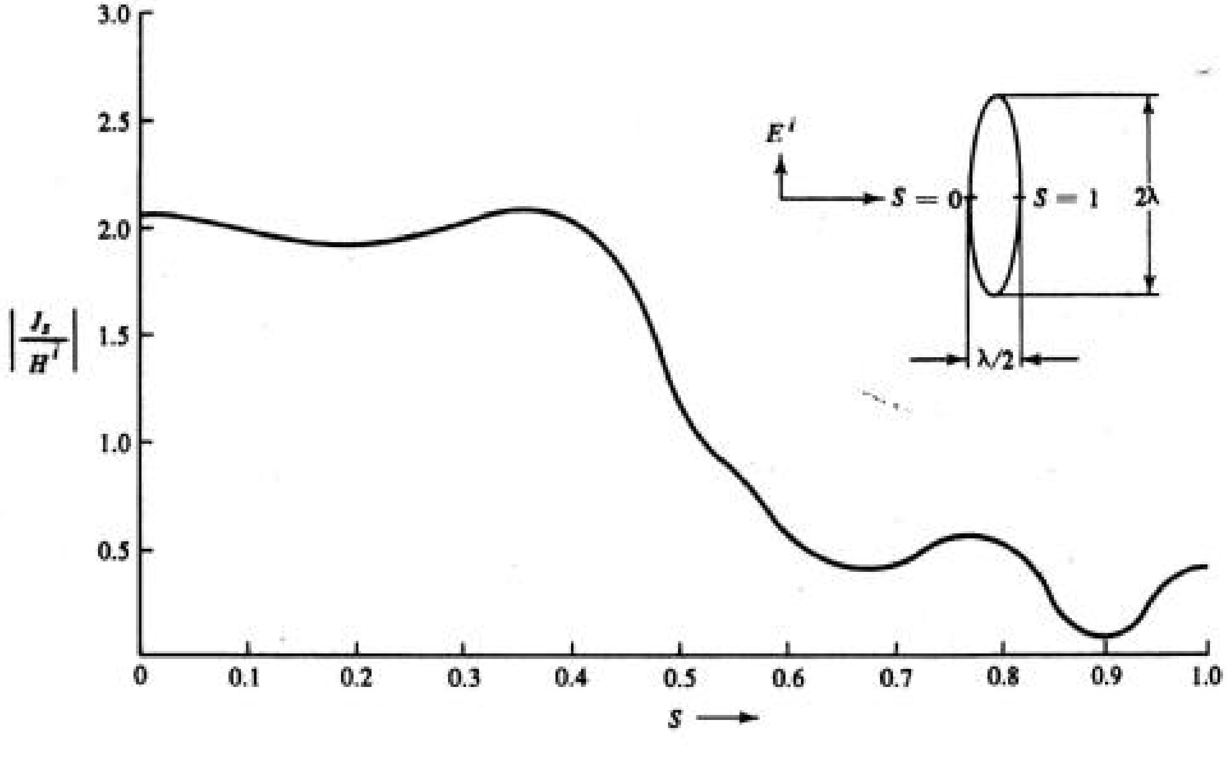


Figure 3-8. **Current density on a conducting elliptic cylinder excited by a plane wave, TE case (after Andreasen [3]).**

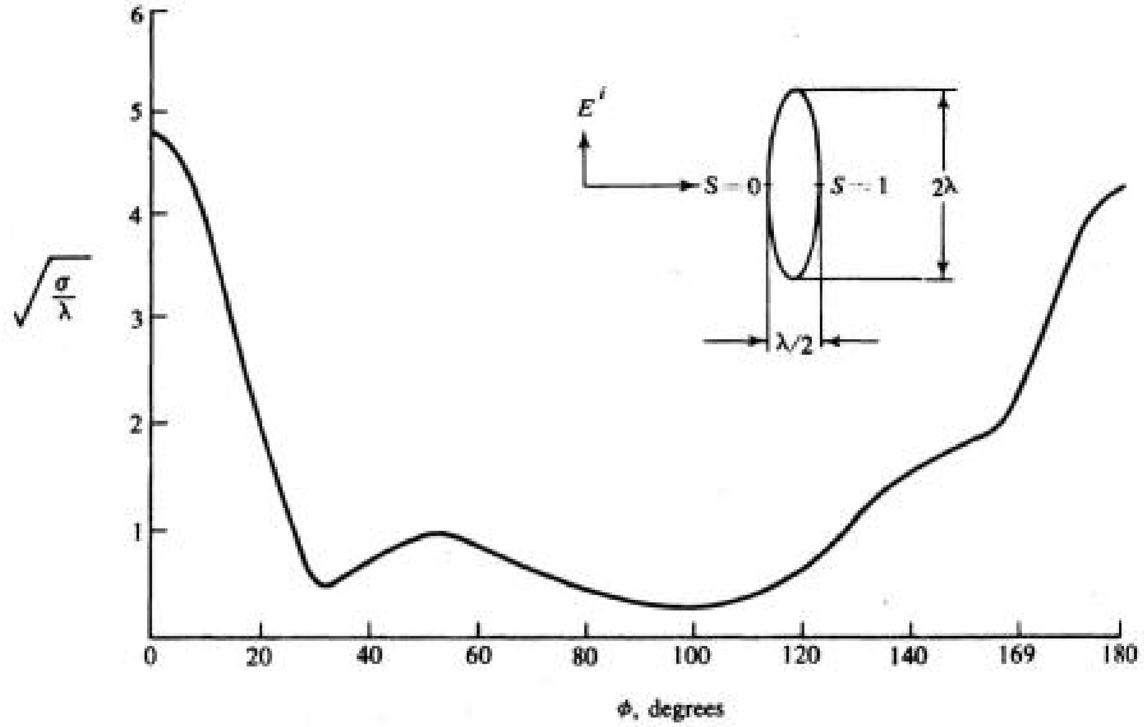


Figure 3-9. **Scattered field pattern for a conducting elliptic cylinder excited by a plane wave, TE case (after Andreasen [3]).**

approximations of the type illustrated by Fig. 3-4. Figure 3-9 shows the TE scattering pattern of the elliptic cylinder, which may be compared to the corresponding TM case of Fig. 3-3. Many other computations are available in the literature [2,3].

3-6. Alternative Formulation

The TM problem was treated by an E-field formulation in Section 3-2, and the TE problem was treated by an H-field formulation in Section 3-5. Actually, both cases can be treated either by an E-field method or an H-field method. To illustrate this, we reconsider the TE case by an E-field formulation.

Let Fig. 3-1 represent a conducting cylinder excited by an impressed TE field $\mathbf{E}^i$ transverse to z. The scattered field $\mathbf{E}^s$ is produced by transverse currents J on C according to the formulas of Section 3-4. For the present problem, these become

$$\mathbf{E}^s = -j\omega\mathbf{A} - \nabla\Phi \tag{3-52}$$

$$\mathbf{A}(\boldsymbol{\rho}) = \mu \oint_C J(\boldsymbol{\rho}')G(\boldsymbol{\rho}, \boldsymbol{\rho}')\, d\mathbf{l}' \tag{3-53}$$

$$\Phi(\boldsymbol{\rho}) = \frac{1}{\varepsilon} \oint_C \left(\frac{-1}{j\omega}\frac{dJ}{dl'}\right) G(\boldsymbol{\rho}, \boldsymbol{\rho}')\, dl' \tag{3-54}$$

where G is given by (3-31). The boundary condition is the tangential component of total **E** vanishes on the conductor; that is,

$$[E_l^i + E_l^s]_{\text{on } C} = 0 \tag{3-55}$$

Defining the operator

$$L(J) = -E_l^s|_{\text{on } C} = \left[j\omega A_l + \frac{\partial \Phi}{\partial l} \right]_{\text{on } C} \tag{3-56}$$

we can write (3-55) in operational notation as

$$L(J) = E_l^i|_{\text{on } C} \tag{3-57}$$

Note that the L of (3-56) contains derivatives, which require careful treatment.

If J is continuous and has a continuous derivative on C, we can solve (3-57) by the method of moments in a straightforward manner. However, this restriction on J is not convenient for cylinders of arbitrary shape. If J is expanded in terms of triangle functions, a point-matching solution works reasonably well unless the field is matched at the breakpoint of the triangles. If J is expanded in terms of pulse functions, dJ/dl gives impulse functions, and the point-matching solution becomes questionable. At any rate, it does not converge in the limit as the number of subsections become infinite. Perhaps the best procedure when using pulses is either to approximate the operator (Section 1-6), or to extend the operator (Section 1-7).

An approximate operator is obtained from (3-56) by replacing all derivatives by difference approximations. The procedure is identical to that given in Chapter 4 for three-dimensional wires, except that the Green's function is different. For a solution the approximate operator is used with pulse functions for expansion and point matching for testing. This procedure is presented in detail in Section 4-2, and we summarize only the results here. The transverse current J on C is represented as

$$J = \sum_n I_n P(l - l_n) \tag{3-58}$$

where $P(x)$ are the pulse functions of (1-49). The coefficients I_n are given by the matrix solution (4-21). The $[Z]$ matrix corresponds to the $[l]$ matrix in the general notation of Section 1-3. The elements Z_{mn} are given by (4-20) with $\Delta \mathbf{l}_n$ replaced by $\Delta \mathrm{C}_n$, and the ψ of (4-16) replaced by

$$\psi(n, m) = \frac{1}{4j\, \Delta C_n} \int_{\Delta C_n} H_0^{(2)}(k\rho_m)\, dl \tag{3-59}$$

where $\rho_m = \sqrt{(x - x_m)^2 + (y - y_m)^2}$. The excitation matrix is given by (4-14), with $\Delta \mathbf{l}_n$ replaced by $\Delta \mathbf{C}_n$. For a simple solution, we can use approximations similar to (3-12) and (3-14); that is,

$$\psi(n, m) \approx \begin{cases} \dfrac{1}{4j\,\Delta C_n} H_0^{(2)}(k\rho_{mn}) & m \neq n \\ \dfrac{1}{4j\,\Delta C_n}\left[1 - j\dfrac{2}{\pi}\log\left(\dfrac{\gamma k\,\Delta C_n}{4e}\right)\right] & m = n \end{cases} \tag{3-60}$$

where ρ_{mn} is the distance between the midpoints of ΔC_m and ΔC_n. For a higher-order solution, it is convenient to further subdivide C and use the methods of Section 3-3. For example, expansion functions of the type shown in Fig. 3-4 can be used, in which case the new Z_{mn} are given by (3-24) or (3-25) with the l_{ij} replaced by Z_{ij}.

Alternatively, we can extend the operator as follows. Define the inner product

$$\langle A, B\rangle = \oint_C A(\boldsymbol{\rho})B(\boldsymbol{\rho})\,dl \tag{3-61}$$

for which L is self-adjoint, and consider a Galerkin solution. If J_m and J_n are two expansion functions for J, the elements of $[l]$ are given by

$$l_{mn} = \langle J_m, LJ_n\rangle = \oint_C J_m(\boldsymbol{\rho})LJ_n(\boldsymbol{\rho})\,dl \tag{3-62}$$

Substituting from (3-56), we have

$$l_{mn} = \oint_C \left[j\omega J_m A_{ln} + J_m \frac{d\Phi_n}{dl}\right] dl \tag{3-63}$$

where the subscripts n on A and Φ denote that they are due to J_n. The first term in the brackets of (3-63) involves no derivatives, and gives no difficulty when pulse functions are used. The second term in the brackets may be integrated once by parts with respect to l. Boundary terms vanish if J_n is in the domain of L, and (3-63) reduces to

$$l_{mn} = \oint_C \left[j\omega J_m A_{ln} - \frac{dJ_m}{dl}\Phi_n\right] dl \tag{3-64}$$

An extended operator can now be defined by specifying that (3-64) apply even for J not in the original domain of L. This is permissible, because nothing is changed if J is in the original domain. Equation (3-64) gives convergent results if J is expanded in triangles and reasonably good results if pulses are used. However, in applying (3-64) to pulse functions, it is better to replace dJ/dl by a difference approximation, in which case convergence is obtained in the limit. It is of interest to note that the latter procedure leads to precisely the same formulas as does the extended operator formulation given earlier, if the same approximations are used for $H_0^{(2)}$.

3-7. Dielectric Cylinders

Consider a dielectric cylinder of cross section S in an impressed field $\mathbf{E}^i$. The dielectric permittivity ε may be a function of x and y, but not of the axial coordinate z. The impressed field excites polarization currents $\mathbf{J}$ in the cylinder, which produce a scattered field $\mathbf{E}^s$. Let L represent the operation relating $-\mathbf{E}^s$ to $\mathbf{J}$; that is,

$$-\mathbf{E}^s = L(\mathbf{J}) \tag{3-65}$$

The total field is $\mathbf{E}^i + \mathbf{E}^s$, and the polarization current is related to the total field by

$$\mathbf{J} = j\omega(\varepsilon - \varepsilon_0)(\mathbf{E}^i + \mathbf{E}^s) \tag{3-66}$$

where ε_0 is the permittivity of free space. Combining (3-65) and (3-66), we have

$$L(\mathbf{J}) + \frac{1}{j\omega(\varepsilon - \varepsilon_0)}\mathbf{J} = \mathbf{E}^i \tag{3-67}$$

within S. In this equation $\mathbf{E}^i$ is known, and $\mathbf{J}$ is the unknown to be determined.

For the case of TM fields, the $\mathbf{E}$ and $\mathbf{J}$ have only z components, and L is given by (3-5); that is,

$$L(J) = \frac{-k\eta}{4}\iint_S J_z(\boldsymbol{\rho}')H_0^{(2)}(k|\boldsymbol{\rho} - \boldsymbol{\rho}'|)\,ds' \tag{3-68}$$

This is an integral operator, and (3-67) can be solved by the method of moments in a straightforward manner. The simplest procedure is to expand J_z in terms of pulse functions and use a point-matching procedure for testing. The details can be found in the literature [6]. An evaluation of the l_{mn} is found to be insensitive to the shape of the subareas Δs_n into which S is divided. Hence the l_{mn} can be conveniently evaluated by treating the Δs_n as if they were of circular cross section, which gives a particularly simple solution of excellent accuracy. Figure 3-10 shows the scattering cross section of a cylindrical shell computed by this method, and compares it to the exact eigenfunction solution. A total of 36 subareas of equal size were used for the matrix solution.

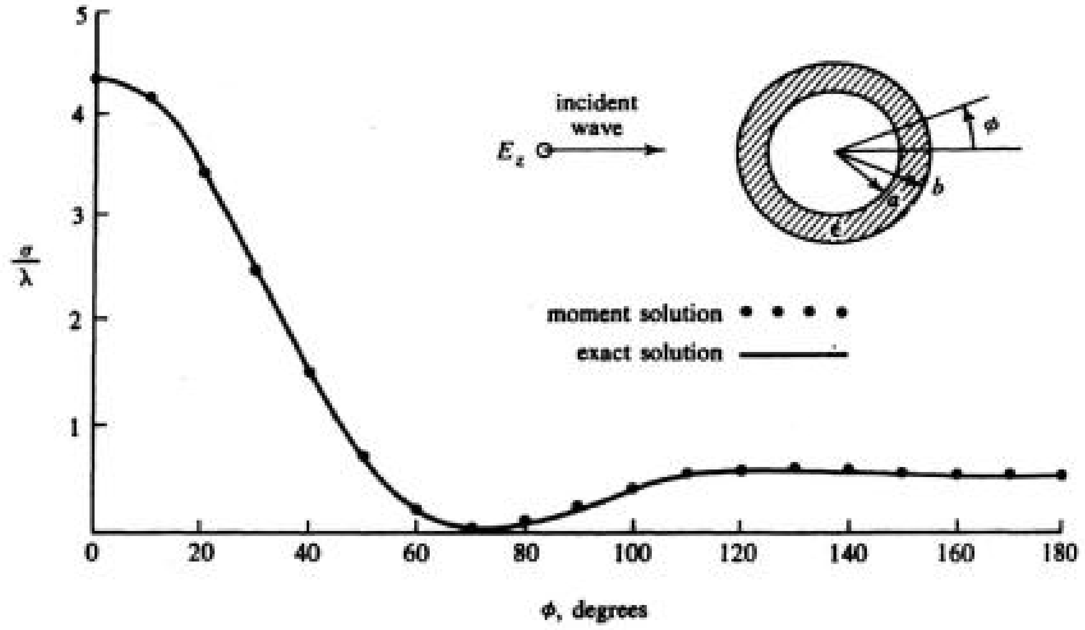


Figure 3-10. **Scattered power pattern for a circular dielectric tube, $a = 0.25\lambda$, $b = 0.30\lambda$, $\varepsilon_r = 4$, TM case (after Richmond [6]).**

In the TE case, L is the more complicated operator

$$L(\mathbf{J}) = j\omega\mathbf{A}(\mathbf{J}) + \nabla\Phi(\mathbf{J}) \tag{3-69}$$

where **A** and Φ are the potential integrals

$$4j\ \mathbf{A}(\mathbf{J}) = \mu \iint\limits_S \mathbf{J}(\boldsymbol{\rho}')H_0^{(2)}(k\,|\boldsymbol{\rho} - \boldsymbol{\rho}'|)\,ds' \tag{3-70}$$

$$4j\ \Phi(\mathbf{J}) = \frac{1}{\varepsilon} \iint\limits_S \left(-\frac{1}{j\omega}\nabla' \cdot \mathbf{J}\right)H_0^{(2)}(k\,|\boldsymbol{\rho} - \boldsymbol{\rho}'|)\,ds' \tag{3-71}$$

Because of the derivatives in (3-69) and (3-71), more care is necessary in applying the method of moments. Strictly speaking, pulse functions are not in the domain of L, and hence should not be used for expanding **J**. However, if they are used in conjunction with a point-matching procedure, usable results can be obtained [7]. Figure 3-11 shows the scattering cross section of the cylindrical shell computed by this procedure using 38 subareas, and compares it to the eigenfunction solution. Note that, because of the crude treatment of the problem, the error is appreciable. Since $\sqrt{\sigma}$ is a continuous linear functional of J, we should expect even more error in J itself. Furthermore, we should not expect the solution to converge to the exact solution as the number of subareas is increased. More accurate computations can be obtained by using expansion functions in the domain of L. Alternatively, we can continue to use pulse functions with either

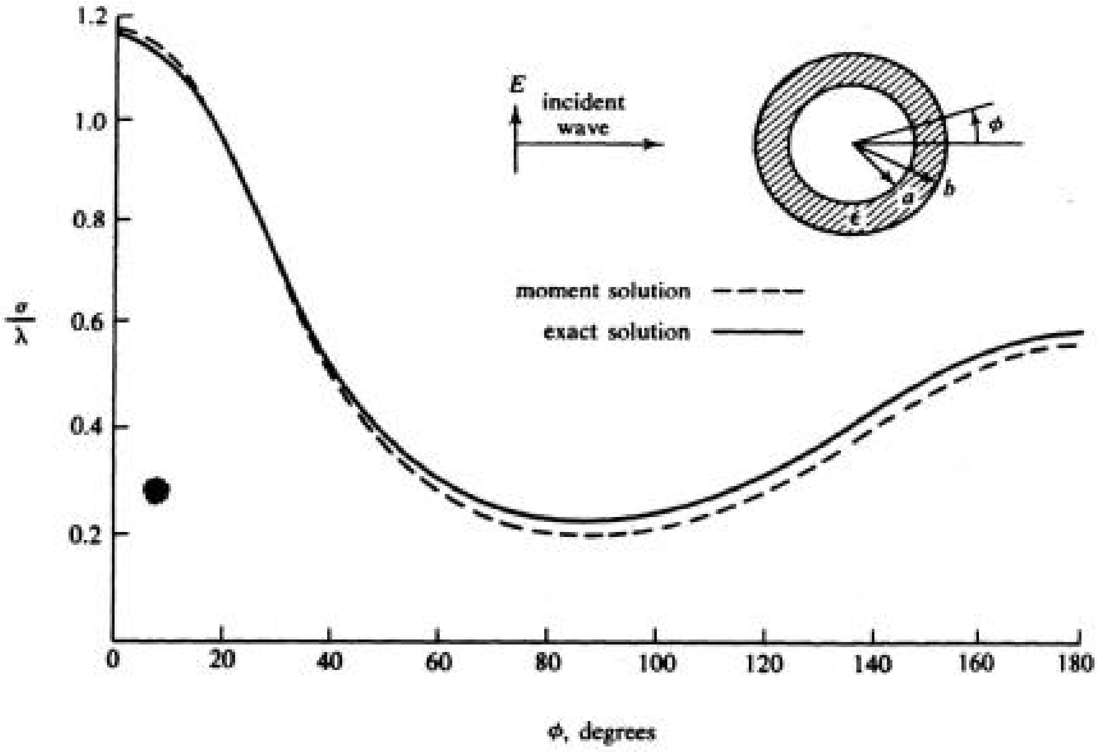


Figure 3-11. **Scattered power pattern for a circular dielectric tube,** $a = 0.25\lambda$, $b = 0.30\lambda$, $\varepsilon_r = 4$, **TE case (after Richmond [7]).**

an approximate L or an extended L, as discussed in Section 3-6. If properly done, TE solutions of accuracy comparable to that for TM solutions should be obtainable.

If the cylinder has a permeability μ different from μ_0 (that of free space), but $\varepsilon = \varepsilon_0$, the problem is dual to that just treated. The appropriate equation is dual to (3-67), that is, obtained from (3-67) by replacing ε by μ, **E** by **H**, and **J** by **M** (magnetic current). Solution proceeds in the same manner as for the dielectric case. If the cylinder has both μ different from μ_0 and ε different from ε_0, the problem is more difficult. It involves a combination of (3-67) and its dual equation. We shall discuss this further in Section 5-7.

If the cylinder is homogeneous in both ε and μ, the problem can be formulated in terms of **E** and **H** on the contour C which bounds the cylinder [4]. This has the advantage of reducing the problem from two dimensions to one dimension; hence fewer subsections are needed for a solution. However, the procedure cannot be applied to inhomogeneous cylinders.

References

[1] R. F. Harrington, *Time-Harmonic Electromagnetic Fields*, McGraw-Hill Book Co., New York, 1961, pp. 223–230.

[2] K. Mei and J. Van Bladel, "Scattering by Perfectly Conducting Rectangular Cylinders," *IEEE Trans.*, Vol. **AP-11**, No. 2, March 1963, pp. 185–192. See also comments on this paper, Vol. **AP-12**, No. 2, March 1964, pp. 235–236.

[3] M. G. Andreasen, "Scattering From Parallel Metallic Cylinders with Arbitrary Cross Sections," *IEEE Trans.*, Vol. **AP-12**, No. 6, Nov. 1964, pp. 746–754.

[4] R. F. Harrington et al., *Matrix Methods for Solving Field Problems*, final report for Contract AF30(602)-3724 with Rome Air Development Center, Griffiss Air Force Base, Rome, N.Y., DDC No. AD 639744, August 1966.

[5] J. G. Van Bladel, "Some Remarks on Green's Dyadic for Infinite Space," *IRE Trans.*, Vol. **AP-9**, No. 6, Nov. 1961, pp. 563–566.

[6] J. H. Richmond, "Scattering by a Dielectric Cylinder of Arbitrary Cross Section Shape," *IEEE Trans.*, Vol. **AP-13**, No. 3, May 1965, pp. 334–341.

[7] J. H. Richmond, "TE Wave Scattering by a Dielectric Cylinder of Arbitrary Cross Section Shape," *IEEE Trans.*, Vol. **AP-14**, No. 4, July 1966, pp. 460–464.