Chapter 11

Chebyshev and Spectral Methods for Partial Differential Equations

11.1 Introduction

Chebyshev polynomial applications to partial differential equations (PDEs)

$$Eu = 0 \text{ on a domain } S, \tag{11.1a}$$

subject to boundary conditions

$$Bu = 0 \text{ on } \partial S, \tag{11.1b}$$

where ∂S is the boundary of the domain S, are a natural progression of the work of Lanczos (1938) and Clenshaw (1957) on ordinary differential equations. However, the first formal publications in the topic of PDEs appear to be those of Elliott (1961), Mason (1965, 1967) and Fox & Parker (1968) in the 1960s, where some of the fundamental ideas for extending one-dimensional techniques to multi-dimensional forms and domains were first developed. Then in the 1970s, Kreiss & Oliger (1972) and Gottlieb & Orszag (1977) led the way to the strong development of so-called pseudo-spectral methods, which exploit the fast Fourier transform of Cooley & Tukey (1965), the intrinsic rapid convergence of Chebyshev methods, and the simplicity of differentiation matrices with nodal bases.

Another important early contribution was the expository paper of Finlayson & Scriven (1966), who set the new methods of the 1960s in the context of the established "method of weighted residuals" (MWR) and classified them formally into the categories of *Galerkin*, *collocation*, and *least squares* methods, as well as into the categories of *boundary*, *interior* and *mixed* methods.

Let us first clarify some of this nomenclature, as well as looking at early and basic approximation methods. We assume that the solution of (11.1a), (11.1b) is to be approximated in the form

$$u \simeq u_n = f(L_n) \tag{11.2}$$

where

$$L_n = \sum_{k=1}^{n} c_k \phi_k \tag{11.3}$$

is a linear combination of an appropriate basis of functions $\{\phi_k\}$ of the independent variables (x and y, say) of the problem and where f is a quasi-linear function

$$f(L) = A.L + B, \tag{11.4}$$

where A, B are specified functions (of x and y).

© 2003 by CRC Press LLC

11.2 Interior, boundary and mixed methods

11.2.1 Interior methods

An *interior method* is one in which the approximation (11.2) exactly satisfies the boundary conditions (11.1b) for all choices of coefficients $\{c_i\}$. This is typically achieved by choosing each basis function ϕ_i appropriately. If Bu in (11.1b) is identically u, so that we have the homogeneous Dirichlet condition

$$u = 0 \text{ on } \partial S, \tag{11.5}$$

then we might well use the identity function for f, and choose a basis for which every ϕ_i vanishes on ∂S. For example, if S is the square domain with boundary

$$\partial S : x = 0,\ x = 1,\ y = 0,\ y = 1. \tag{11.6}$$

then one possibility would be to choose

$$\phi_k = \Phi_{ij} = \sin i\pi x \, \sin j\pi x \tag{11.7}$$

with

$$k = i + n(j-1)$$

and

$$c_k = a_{ij},$$

say, so that the single index $k = 1, \ldots, n^2$ counts row by row through the array of n^2 basis functions corresponding to the indices $i = 1, \ldots, n$ and $j = 1, \ldots, n$. In practice we might in this case change notation from ϕ_k to Φ_{ij} and from u_{n^2}, L_{n^2} to u_{nn}, L_{nn}, setting

$$u \simeq u_{nn} = f(L_{nn})$$

where

$$L_{nn} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \Phi_{ij}(x, y). \tag{11.8}$$

It only remains to solve the interior problem (11.1a).

There is a generalisation of the above method, that is sometimes applicable to the general Dirichlet boundary conditions

$$u = B(x, y) \tag{11.9}$$

on the boundary

$$\Gamma : A(x, y) = 0,$$

where we know a formula $A = 0$ for the algebraic equation of Γ, as well as a formula $B = 0$ for the boundary data. Then we may choose

$$u \simeq u_{nn} = f(L_{nn}) = A(x, y) L_{nn} + B(x, y), \tag{11.10}$$

© 2003 by CRC Press LLC

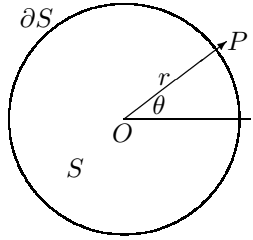


Figure 11.1:

which automatically satisfies (11.5), whatever we take for L_{nn}. See Mason (1967) for a successful application and early discussion of such techniques.

In the discussion that follows we assume unless otherwise stated that f is the identity, so that u_n and L_n are the same function.

11.2.2 Boundary methods

A *boundary method* is one in which the approximation (11.2) exactly satisfies the PDE (11.1a) for all choices of coefficients $\{c_i\}$. If the PDE is linear, for example, then this is achieved by ensuring that every basis function ϕ_k is a particular solution of (11.1a). This method is often termed the "method of particular solutions" and has a long history — see for example Vekua (1967) — and indeed the classical method of separation of variables for PDEs is typically of this nature. It remains to satisfy the boundary conditions approximately by suitable choice of coefficients $\{c_i\}$.

For example, consider Laplace's equation in (r, θ) coordinates:

$$\triangle u = r^2 \frac{\partial^2 u}{\partial r^2} + r \frac{\partial u}{\partial r} + \frac{\partial^2 u}{\partial \theta^2} = 0 \tag{11.11a}$$

in the disk $S : r \leq 1$, together with

$$u = g(\theta) \tag{11.11b}$$

on $\partial S : r = 1$, where g is a known 2π-periodic function of the orientation θ of a general point, P say, on the boundary (Figure 11.1).

Then

$$u \simeq u_n(r, \theta) = {\sum_{k=0}^{n}}' [a_k(r^k \cos(k\theta)) + b_k(r^k \sin(k\theta))] \tag{11.12}$$

is an exact solution of (11.11a) for all $\{a_k, b_k\}$, since $r^k \cos(k\theta)$ and $r^k \sin(k\theta)$ are particular solutions of (11.11a), which may readily be derived by separation of variables in (11.11a) (see Problem 1).

© 2003 by CRC Press LLC

Substituting (11.12) into (11.11b) gives

$$u = g(\theta) \simeq u_n(1,\theta) = \sum_{k=0}^{n}{}' [a_k \cos(k\theta) + b_k \sin(k\theta)]. \tag{11.13}$$

Clearly we require the latter trigonometric sum to approximate $g(\theta)$. This may theoretically be achieved by choosing a_k and b_k to be coefficients in the full Fourier series expansion of $g(\theta)$, namely

$$a_k = \pi^{-1} \int_0^{2\pi} g(\theta) \cos(k\theta)\, \mathrm{d}\theta, \quad b_k = \pi^{-1} \int_0^{2\pi} g(\theta) \sin(k\theta)\, \mathrm{d}\theta. \tag{11.14}$$

These integrals must be replaced by numerical approximations, which may be rapidly computed by the fast Fourier transform (FFT, see Section 4.7). The FFT computes an approximate integral transform, by "exactly" computing the discrete Fourier transform given by

$$a_k = n^{-1} \sum_{i=0}^{2n}{}'' g(\theta_i) \cos(k\theta_i), \quad b_k = n^{-1} \sum_{i=0}^{2n}{}'' g(\theta_i) \sin(k\theta_i), \tag{11.15}$$

where

$$\theta_i = i\pi/n \quad (i = 0, \ldots, 2n). \tag{11.16}$$

Here the periodic formulae (11.14) have been approximated by Filon's rule, namely the Trapezoidal rule for trigonometric functions, which is a very accurate substitute in this case.

Several examples of the method of particular solutions are given by Mason & Weber (1992), where it is shown that the method does not always converge! See also, however, Fox et al. (1967) and Mason (1969) where the "L-shaped membrane eigenvalue problem" is solved very rapidly and accurately by this method.

Boundary MWR methods are important because, when they are applicable, they effectively reduce the dimension of the problem by restricting it to the domain boundary. In consequence such methods can be very efficient indeed. Moreover, because they normally incorporate precise features of the solution behaviour, they are often very accurate too — see Mason (1969) where the first L-shaped membrane eigenvalue is computed correct to 13 significant figures for $(n =)24$ basis functions.

However, boundary MWR methods are not the only available techniques for in effect reducing the problem dimension. The method of fundamental solutions, which has been adopted prolifically by Fairweather & Karageorghis (1998), uses fundamental PDE solutions as a basis. These solutions typically have singularities at their centres, and so must be centred at points exterior to S. This method is closely related to the boundary integral equation (BIE) method and hence to the boundary element method (BEM) — for which

© 2003 by CRC Press LLC

there is a huge literature (Brebbia et al. 1984, for example), and indeed the boundary integral equation method adopts the same fundamental solutions, but as weight functions in integral equations. For example, functions behaving like $\log r$ occur in both the method of fundamental solutions and the boundary integral equation method for Laplace's equation in two dimensions.

Both the BIE method and the BEM convert a PDE on a domain into an integral equation over its boundary. They consequently have the possibility for considerable improvements in efficiency and accuracy over classical finite element methods for the original PDE, depending on the nature of the geometry and other factors.

11.2.3 Mixed methods

A *mixed method* is one in which both the PDE (11.1a) and its boundary conditions (11.1b) need to be approximated. In fact this is generally the case, since real-life problems are usually too complicated to be treated as boundary or interior problems alone. Examples of such problems will be given later in this chapter.

11.3 Differentiation matrices and nodal representation

An important development, which follows a contrasting procedure to that of the methods above, is to seek, as initial parameters, not the coefficients c_k in the approximation form L_n (11.3) but instead the values $u_n(x_i, y_j)$ of u_n at a suitable mesh of Chebyshev zeros. Derivatives can be expressed in terms of these u_n values also, and hence a system of (linear) algebraic equations can be formed for the required values of u_n. It is then possible, if required, to recover the coefficients c_k by a Chebyshev collocation procedure.

An example of the procedure was given in Chapter 10 (Section 10.5.1) for ordinary differential equations (ODEs). In the case of PDEs it should be noted that the procedure is primarily suited to rectangular regions.

11.4 Method of weighted residuals

11.4.1 Continuous MWR

The standard MWR, which we call the continuous MWR, seeks to solve an interior problem by finding an approximation of the form (11.2) which minimises, with respect to c_k $(k = 1, \ldots, n)$, the expression

$$\langle Eu_n \, , \, W_k \rangle^2 \equiv \left[\int_S (Eu_n).W_k \, \mathrm{d}S \right]^2 , \tag{11.17}$$

© 2003 by CRC Press LLC

where W_k is a suitable weight function (Finlayson & Scriven 1966). Here we assume that E is a linear partial differential operator. More specifically :

(i) MWR is a *least squares method* if

$$W_k \equiv w.Eu_n, \quad (k = 1, \ldots, n), \tag{11.18}$$

where w is a fixed non-negative weight function. Then, on differentiating (11.17) with respect to c_k, we obtain the simpler form

$$\langle Eu_n \, , \, w.E\phi_k \rangle = 0, \quad (k = 1, \ldots, n). \tag{11.19}$$

This comprises a linear system of n equations for c_k.

(ii) MWR is a *Galerkin method* if

$$W_k \equiv w.\phi_k. \tag{11.20}$$

Note that, in this case, we can give a zero (minimum) value to (11.17) by setting

$$\langle Eu_n \, , \, w.\phi_k \rangle = 0, \quad (k = 1, \ldots, n), \tag{11.21}$$

again a system of linear equations for c_k. It follows from (11.21) that

$$\langle Eu_n \, , \, w.u_n \rangle = 0. \tag{11.22}$$

More generally, we can if we wish replace ϕ_k in (11.21) by any set of test functions ψ_k, forming a basis for u_k and solve

$$\langle Eu_n \, , \, w.\psi_k \rangle = 0, \quad (k = 1, \ldots, n). \tag{11.23}$$

(iii) MWR is a *collocation method (interpolation method)* at the points $P_1, \ldots, P_n$ if

$$W_k \equiv \delta(P_k), \tag{11.24}$$

where $\delta(P)$ is the Dirac delta function (which is infinite at the point P, vanishes elsewhere and has the property that $\langle u \, , \, \delta(P) \rangle = u(P)$ for any well-behaved function u). Then Eu_n in (11.17) will be set to zero at P_k, for every k.

11.4.2 Discrete MWR — a new nomenclature

It is also possible to define a discrete MWR, for each of the three types of methods listed above, by using a discrete inner product in (11.17). Commonly we do not wish, or are unable, to evaluate and integrate $Eu_n.W_k$ over a continuum, in which case we may replace the integral in (11.17) by the sum

$$\sum_{S_n} (Eu_n).W_k, \tag{11.25}$$

© 2003 by CRC Press LLC

where S_n is a discrete point set representing S.

The discrete MWR, applied to an interior problem, is based on a discrete inner product. It seeks an approximation of the form (11.2) which solves

$$\min_{c_k} \left[\left(\sum_{j=1}^{p} Eu_n(\mathbf{x}_j) W_k(\mathbf{x}_j) \right)^2 \equiv (Eu_n, W_k)^2 \right], \tag{11.26}$$

where $\mathbf{x}_j$ $(j = 1, \ldots, p)$ are a discrete set of nodes in S, selected suitably from values of the vector $\mathbf{x}$ of independent variables, and W_k are appropriate weights.

(i) The discrete MWR is a *discrete least-squares method* if

$$W_k \equiv wEu_n. \tag{11.27}$$

This is commonly adopted in practice in place of (11.18) for convenience and to avoid integration.

(ii) The discrete MWR is a *discrete Galerkin method* if

$$W_k \equiv w\phi_k \tag{11.28}$$

or, equivalently,

$$(Eu_n, w\psi_k) = 0. \tag{11.29}$$

Note that the PDE operator Eu_n is directly orthogonal to every test function ψ_k, as well as to the approximation u_n, so that

$$(Eu_n, wu_n) = 0. \tag{11.30}$$

(iii) The discrete MWR is a *discrete collocation method* if (11.24) holds, where $\{P_k\}$ is contained within the discrete point set S_n.

11.5 Chebyshev series and Galerkin methods

The most basic idea in Chebyshev polynomial methods is that of expanding a solution in a (multiple) Chebyshev series expansion, and using the partial sum as an approximation. This type of approach is referred to as a *spectral method* by Gottlieb & Orszag (1977). This type of ODE/PDE method had previously, and still has, several other names, and it is known as (or is equivalent to) a Chebyshev series method, a Chebyshev–Galerkin method, and the tau method of Lanczos.

Before introducing PDE methods, we consider the Lanczos tau method: one of the earliest Chebyshev methods for solving a linear ODE

$$Ey = 0$$

© 2003 by CRC Press LLC

in the approximate form y_n.

Lanczos (1938) and Ortiz and co-workers (Ortiz 1969, Freilich & Ortiz 1982, and many other papers) observed that, if y_n is expressed in the power form

$$y_n = b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n, \tag{11.31}$$

then, for many important linear ODEs, Ey_n can be equated to a (finite) polynomial with relatively few terms, of the form

$$Ey_n = \tau_1 T_{q+1}(x) + \tau_2 T_{q+2}(x) + \cdots + \tau_s T_{q+s}(x), \tag{11.32}$$

where q and s are some integers dependent on E. The method involves substituting y_n (11.31) into the perturbed equation (11.32) and equating powers of x from x^0 to x^{q+s}. The t (say) boundary conditions are also applied to y_n, leading to a total of $q+s+t+1$ linear equations for $b_0, \ldots, b_n$, $\tau_1, \ldots, \tau_s$. We see that for the equations to have one and only one solution we must normally have

$$q + t = n. \tag{11.33}$$

The equations are solved by first expressing $b_0, \ldots, b_n$ in terms of $\tau_1, \ldots, \tau_s$, solving s equations for the τ values and hence determining the b values. Because of the structure of the resulting matrix and assuming s is small compared to n, the calculation can routinely reduce to one of $O(n)$ operations, and hence the method is an attractive one for suitable equations.

The above method is called the (Lanczos) *tau method* - with reference to the introduction by Lanczos (1938) of perturbation terms, with coefficients $\tau_1, \ldots, \tau_s$, on the right hand side of $Ey = 0$ to enable the ODE to be exactly solved in finite form. The nice feature of this approach is that the tau values give a measure of the sizes of the contributions that the perturbation terms make to the ODE — at worst,

$$|Ey_n| \le |\tau_1| + |\tau_2| + \cdots + |\tau_s|. \tag{11.34}$$

For some special cases, Lanczos (1957), Fox & Parker (1968), Mason (1965), Ortiz (1980, 1986, 1987), Khajah & Ortiz (1991) and many others were able to give quite useful error estimates based on the known form (11.32).

The tau method is also equivalent to a Galerkin method, since Ey_n is orthogonal with respect to $(1-x^2)^{-1/2}$ to all polynomials of degree up to q, as a consequence of (11.32). Note that the Galerkin method proper is more robust than equivalent tau or Chebyshev series methods, since, for example, it is unnecessary to introduce τ terms or to find and use the form (11.32). The Galerkin method directly sets up a linear system of equations for its coefficients. For example, if we wish to solve

$$u' - u = 0, \quad u(0) = 1 \tag{11.35}$$

© 2003 by CRC Press LLC

by a Galerkin procedure using Legendre polynomials $P_i^*(x)$ appropriate to $[0,1]$, namely

$$u \sim u_n = \sum_{i=0}^{n} c_i P_i^*(x), \tag{11.36}$$

then we solve

$$\int_0^1 (u_n' - u_n).P_i^*(x)\,\mathrm{d}x = 0 \quad (i = 0, 1, \ldots, n-1) \tag{11.37}$$

and

$$\sum_{i=0}^{n} c_i P_i^*(0) = 1. \tag{11.38}$$

Here (11.37) and (11.38) comprise $n+1$ equations for $c_0, \ldots, c_n$. Note that a snag in the Galerkin method is the need to evaluate the various integrals that occur, which are likely to require a numerical treatment except in simple problems such as (11.35).

It is worth remembering that Chebyshev series are also transformed Fourier series, and so Chebyshev methods may be based on known methods for generating Fourier partial sums or Fourier transforms, based on integrals and expansions.

11.6 Collocation/interpolation and related methods

We have seen, in Sections 5.5 and 6.5, that a Chebyshev series partial sum of degree n of a continuous function is a near-minimax approximation on $[-1,1]$ within a relative distance of order $4\pi^{-2}\log n$, whereas the polynomial of degree n interpolating (collocating) the function at the $n+1$ zeros of $T_{n+1}(x)$ is near-minimax within a slightly larger relative distance of order $2\pi^{-1}\log n$. Thus, we may anticipate an error that is $\pi/2$ times as large in Chebyshev interpolation compared with Chebyshev series expansion. In practice, however, this is a very small potential factor, and polynomial approximations from the two approaches are virtually indistinguishable. Indeed, since collocation methods are simpler, more flexible and much more generally applicable, they offer a powerful substitute for the somewhat more mathematically orthodox but restrictive series methods.

The title *pseudo-spectral method* was introduced by Gottlieb & Orszag (1977), in place of *Chebyshev collocation method*, to put across the role of this method as a robust substitute for the spectral method. Both series (spectral) and collocation (pseudo-spectral) methods were rigorously described by Mason (1970) as near-minimax. Note that minimax procedures generally involve infinite procedures and are not practicably feasible, while spectral, and more particularly pseudo-spectral, methods are typically linear and very close to

© 2003 by CRC Press LLC

minimax and therefore provide an excellent and relatively very inexpensive substitute for a minimax approximation method.

It has long been realised that collocation for differential equations is almost identical to series expansion. Lanczos (1957) noted that the ODE error form adopted in his tau method (11.32) could conveniently be replaced with nearly identical results (though different τ coefficients) by

$$Ey_n = T_{q+1}(x).(\tau_1 + \tau_2 x + \cdots + \tau_s x^{s-1}), \tag{11.39}$$

where $q + s$ is the degree of Ey_n. Note that the error in the ODE vanishes at the zeros of $T_{q+1}(x)$, and so the method is equivalent to a collocation method (in the ODE). Lanczos called this method the *selected points method*, where the zeros of T_{q+1} are the points selected in this case. Lanczos sometimes also selected Legendre polynomial zeros instead, since in practice they sometimes give superior results.

We have already shown that the Chebyshev collocation polynomial, $f_n(x)$ of degree n to a given $f(x)$, may be very efficiently computed by adopting a discrete orthogonalisation procedure

$$f(x) \simeq f_n(x) = {\sum_{i=0}^{n}}' c_i T_i(x), \tag{11.40}$$

where

$$c_i = \frac{2}{N} \sum_{k=0}^{N} f(x_k) T_i(x_k) = \frac{2}{N} \sum_{k=0}^{N} f(\cos(\theta_k)) \cos(i\theta_k), \tag{11.41}$$

with

$$x_k = \cos(\theta_k) = \cos\left(\frac{(2k+1)\pi}{2(N+1)}\right) \quad (k = 0, 1, \ldots, n). \tag{11.42}$$

For $N = n$, this yields the collocation polynomial, and this clearly mimics the Chebyshev series partial sum of order n, which has the form (11.40) with (11.41) replaced by

$$c_i = \frac{2}{\pi} \int_{-1}^{1} (1 - x^2)^{-1/2} f(x) T_i(x)\, \mathrm{d}x = \frac{2}{\pi} \int_0^{\pi} f(\cos(\theta)) \cos(i\theta)\, \mathrm{d}\theta. \tag{11.43}$$

with $x = \cos(\theta)$.

Note that the discrete Chebyshev transform in x and the discrete Fourier transform in θ, that appear in (11.41), also represent an excellent numerical method (Filon's rule for periodic integrands) for approximately computing the continuous Chebyshev transform and Fourier transform that appear in (11.43). The fast Fourier transform (FFT), which is of course a very efficient method of computing the Fourier transform, does in fact compute the discrete

© 2003 by CRC Press LLC

Fourier transform instead. However, (11.43) is typically replaced by (11.41) for a value of N very much larger than n, say $N = 1024$ for $n = 10$. So there are really two different discrete Fourier transforms, one for $N = n$ (collocation) and one for $N \gg n$ (approximate series expansion).

11.7 PDE methods

We note that, for simplicity, the above discussions of nomenclature have been based on ODEs, for which boundary conditions apply at just a few points, usually only one or two. Typically these boundary conditions are imposed exactly as additional constraints on the approximation, with only a little effect on the number of coefficients remaining. For example, in the tau method for

$$Eu \equiv u' - u = 0, \quad u(0) = 1 \quad \text{in } [0,1], \tag{11.44}$$

we determine

$$u \sim u_n = c_0 + c_1 x + \cdots + c_n x^n$$

by equating coefficients of $1, x, x^2, \ldots, x^n$ in

$$\begin{aligned} Eu_n \equiv (c_1 - c_0) + (2c_2 - c_1)x + (3c_3 - c_2)x^2 + \cdots + \\ + (nc_n - c_{n-1})x^{n-1} - c_n x^n = \tau T_n^*(x). \end{aligned} \tag{11.45}$$

This yields $n+1$ linear equations for $c_0, \ldots, c_n, \tau$, and an additional equation is obtained by setting $c_0 = 1$ to satisfy the boundary (initial) condition.

In spectral and pseudo-spectral methods for PDEs, the boundary conditions play a much more crucial role than for ODEs, and it becomes important to decide whether to satisfy the boundary conditions implicitly, in the form chosen for the basis functions, or to apply the boundary conditions as additional constraints. For this reason, Gottlieb & Orszag (1977) and Canuto et al. (1988) differentiate between Galerkin and tau methods primarily in terms of their treatment of boundary conditions — whereas we have above viewed these methods as equivalent, one based on the orthogonality requirement and the other based on the form of the ODE (perturbation) error. Canuto et al. (1988) view a Galerkin method as a series method in which the boundary conditions are included implicitly in the chosen solution form, whereas a tau method is seen by them as a series method for which the boundary conditions are applied explicitly through additional constraints.

The distinction made by Canuto et al. (1988) between Galerkin and tau methods has virtues. In particular the number of free approximation coefficients needed to satisfy boundary conditions can be very large, whereas this may be a peripheral effect if the boundary can be treated implicitly. So a distinction is needed. But the words, Galerkin and tau, do not conjure up boundary issues, but rather an orthogonality technique and tau perturbation

© 2003 by CRC Press LLC

terms. A better terminology, we would suggest, would be to refer to methods which include/exclude boundary conditions from the approximation form as implicit/explicit methods respectively. We could alternatively use the titles interior/mixed methods, as discussed for the MWR above.

Nomenclature and methods become more complicated for PDEs in higher dimensions. In the following sections we therefore give a number of examples of problems and methods to illustrate the formalisms that result from approaches of the Galerkin, tau, collocation, implicit, explicit (&c.) variety. We do not view spectral and pseudo-spectral methods, unlike Galerkin and tau methods, as specifically definable methods, but rather as generic titles for the two main branches of methods (series and collocation). A generalisation of the Lanczos tau method might thus be termed a spectral explicit/mixed tau method.

11.7.1 Error analysis

Canuto et al. (1988) and Mercier (1989), among others, give careful attention to error bounds and convergence results. In particular, Canuto et al. (1988) address standard problems such as the Poisson problem, as well as individually addressing a variety of harder problems. In practice, however, the main advantage of a spectral method lies in the rapid convergence of the Chebyshev series; this in many cases makes feasible an error estimate based on the sizes of Chebyshev coefficients, especially where convergence is exponential.

11.8 Some PDE problems and various methods

It is simplest to understand, develop and describe spectral and pseudo-spectral methods by working through a selection of problems of progressively increasing complexity. This corresponds quite closely to the historical order of development, which starts, from a numerical analysis perspective, with the novel contributions of the 1960s and is followed by the fast (FFT-based) spectral methods of the 1970s. Early work of the 1960s did establish fundamental techniques and compute novel approximations to parabolic and elliptic PDEs, based especially on the respective forms

$$u(x,t) \sim u_n(x,t) = \sum_{i=0}^{n} c_i f_i(t) T_i(x) \quad (-1 \le x \le 1;\ t \ge 0) \tag{11.46}$$

for parabolic equations, such as $u_{xx} = u_t$, and

$$u(x,y) \sim u_n(x,y) = \sum_{i=0}^{m} \sum_{j=0}^{n} c_{ij} T_i(x) T_j(y) \quad (-1 \le x, y \le 1) \tag{11.47}$$

for elliptic problems, such as $\Delta u \equiv u_{xx} + u_{yy} = f$.

© 2003 by CRC Press LLC

An early paper based on (11.46) was that of Elliott (1961), who determined $f_i(t)$ as approximate solutions of a system of ODEs, in the spirit of the "method of lines". Another early paper based on (11.47) was that of Mason (1967), which solves the membrane eigenvalue problem (see Section 11.8.2 below)

$$\Delta u + \lambda u = 0 \text{ in } S, \quad u = 0 \text{ on } \partial S, \tag{11.48}$$

for the classical problem of an L-shaped membrane (consisting of three squares co-joined), based on a preliminary conformal mapping of the domain and an approximation

$$u \simeq A(x,y).\phi_n(x,y), \tag{11.49}$$

where $A = 0$ is the algebraic equation of the mapped boundary. Mason (1969) also used an approximation of form (11.46) to solve a range of separable PDEs including (11.48). Indeed the leading eigenvalue of the L-membrane was computed to 13 significant figures by Mason (1969) ($\lambda = 9.639723844022$).

These early quoted papers are all essentially based on the collocation method for computing coefficients c_i or c_{ij}. It is also possible to develop tau/series methods for the form (11.46), based on the solution by the Lanczos tau method of the corresponding ODEs for $f_i(t)$; this has been carried out for very basic equations such as the heat equation and Poisson equation (Berzins & Dew 1987).

11.8.1 Power basis: collocation for Poisson problem

Consider the Poisson problem

$$\Delta u \equiv \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x,y) \text{ in } S, \quad u = 0 \text{ on } \partial S, \tag{11.50}$$

where S is the square with boundaries $x = \pm 1, y = \pm 1$. Suppose we approximate as

$$u \simeq u_{mn} = \phi(x,y) \sum_{i=0}^{m-2} \sum_{j=0}^{n-2} a_{ij} x^i y^j, \tag{11.51}$$

where we adopt the power basis $x^i y^j$ and include a multiplicative factor $\phi(x)$ such that $\phi = 0$ is the (combined) equation of the boundaries. In this case,

$$\phi(x,y) = (x^2 - 1)(y^2 - 1). \tag{11.52}$$

Then we may rewrite u_{mn} as

$$u_{mn} = \sum_{i=0}^{m-2} \sum_{j=0}^{n-2} a_{ij} (x^{i+2} - x^i)(y^{j+2} - y^j) \tag{11.53}$$

© 2003 by CRC Press LLC

and hence, applying Δ, obtain

$$\begin{aligned}\Delta u_{mn} &= \sum_{i=0}^{m-2}\sum_{j=0}^{n-2} a_{ij}\left([(i+2)(i+1)x^2 - i(i-1)]x^{i-2}y^j(y^2-1)\right.\\ &\quad \left. + [(j+2)(j+1)y^2 - j(j-1)]x^i(x^2-1)y^{j-2}\right). \qquad (11.54)\end{aligned}$$

Now set Δu_{mn} equal to $f(x,y)$ at the $(m-1)(n-1)$ points (x_k, y_l) $(k = 1,\ldots,m-1;\, l = 1,\ldots,n-1)$, where $\{x_k\}, \{y_l\}$ are the respective sets of zeros of $T_{m-1}(x)$, $T_{n-1}(y)$, respectively, namely the points

$$x_k = \cos\left(\frac{(2k-1)\pi}{2(m-1)}\right), \quad y_l = \cos\left(\frac{(2l-1)\pi}{2(n-1)}\right). \qquad (11.55)$$

This leads to a full linear algebraic system of $(m-1)(n-1)$ equations for a_{ij}. It is relatively straightforward to code a computer procedure for the above algorithm.

We also observe that the approximation u_{mn} adopted in (11.53) above could equally well be replaced by the equivalent form

$$u_{mn} = \sum_{i=2}^{m}\sum_{j=2}^{n} a_{ij}(x^i - x^{i \bmod 2})(y^j - y^{j \bmod 2}), \qquad (11.56)$$

where $(i \bmod 2)$ is 0 or 1 according as i is even or odd, since x^2-1 and y^2-1 are in every case factors of u_{mn}. This leads to a simplification in Δu_{mn} (as in (11.54)), namely

$$\begin{aligned}\Delta u_{mn} &= \sum_{i=2}^{m}\sum_{j=2}^{n} a_{ij}\left[i(i-1)x^{i-2}(y^j - y^{j \bmod 2})\right.\\ &\quad \left. + j(j-1)y^{j-2}(x^i - x^{i \bmod 2})\right]. \qquad (11.57)\end{aligned}$$

The method then proceeds as before. However, we note that (11.51) is a more robust form for more general boundary shapes ∂S and more general boundary conditions $Bu = 0$, since simplifications like (11.56) are not generally feasible.

The above methods, although rather simple, are not very efficient, since no account has been taken of special properties of Chebyshev polynomials, such as discrete orthogonality. Moreover, (11.53) and (11.56) use the basis of power functions $x^i y^j$ which, for m and n sufficiently large, can lead to significant loss of accuracy in the coefficients a_{ij}, due to rounding error and poor conditioning in the resulting linear algebraic system. We therefore plan to follow up this discussion in a later Section by considering more efficient and well conditioned procedures based on the direct use of a Chebyshev polynomial product as a basis, namely $T_i(x)T_j(y)$.

However, before we return to the Poisson problem, let us consider a more difficult problem, where the form (11.51) is very effective and where a power basis is adequate for achieving relatively high accuracy.

© 2003 by CRC Press LLC

11.8.2 Power basis: interior collocation for the L-membrane

Consider the eigenvalue problem

$$\Delta u + \lambda u = 0 \text{ in } S, \quad u = 0 \text{ on } \partial S, \tag{11.58}$$

where S is the L-shaped region shown (upside down for convenience) in Figure 11.2. It comprises three squares of unit sides placed together. To remove the re-entrant corner at O, we perform the mapping, adopted by Reid & Walsh (1965),

$$z' = z^{2/3} \quad (z' = x' + \mathrm{i}y', \quad z = x + \mathrm{i}y), \tag{11.59}$$

where x, y are coordinates in the original domain S (Figure 11.2) and x', y' are corresponding coordinates in the mapped domain S' (Figure 11.3).

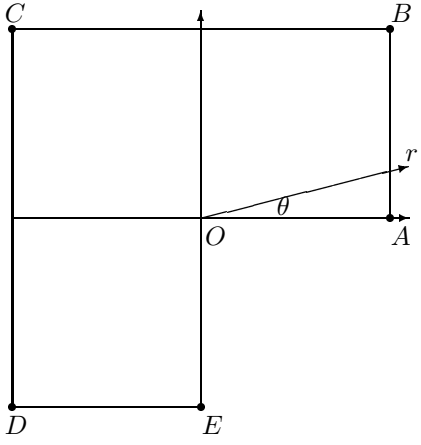


Figure 11.2: L-membrane

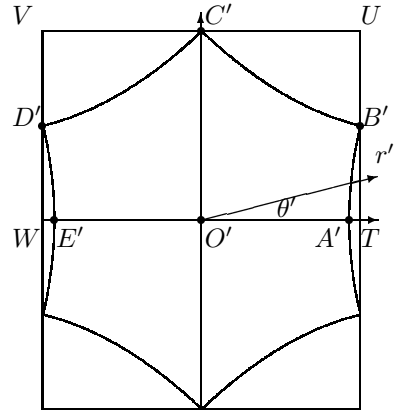


Figure 11.3: Mapped domain

Note that the domain S' is the upper half of a curved hexagon with corners of angle $\frac{\pi}{2}$ shown in Figure 11.3, where the vertices A', B', C', D', E' correspond to A, B, C, D, E in Figure 11.2. (The lower half of the hexagon does not concern us, but is included in Figure 11.3 to show the geometry and the symmetry.) Now, from the mapping,

$$r' = r^{2/3}, \quad \theta' = \tfrac{2}{3}\theta. \tag{11.60}$$

Then

$$\begin{aligned} \Delta u + \lambda u &\equiv r^{-2}\left[r^2\frac{\partial^2 u}{\partial r^2} + r\frac{\partial u}{\partial r} + \frac{\partial^2 u}{\partial \theta^2} + \lambda u\right] \\ &= r^{-2}\left[r\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \frac{\partial^2 u}{\partial \theta^2}\right] + \lambda u \\ &= (r')^{-3}\left[\tfrac{2}{3}\left(r'\frac{\partial}{\partial r'}\right)\tfrac{2}{3}\left(r'\frac{\partial u}{\partial r'}\right) + \tfrac{4}{9}\frac{\partial^2 u}{\partial \theta'^2}\right] + \lambda u. \end{aligned} \tag{11.61}$$

© 2003 by CRC Press LLC

Hence

$$\Delta u + \lambda u = \tfrac{4}{9}(r')^{-1}.(\Delta' u + \tfrac{9}{4} r' \lambda u) = 0, \tag{11.62}$$

where dashes on Δ', r' indicate that dashed coordinates are involved. Thus the problem (11.58) has transformed, now dropping dashes on r, u, to

$$\Delta u + \tfrac{9}{4} r \lambda u = 0 \text{ in } S', \quad u = 0 \text{ on } \partial S'. \tag{11.63}$$

Before proposing a numerical method, we need to find the algebraic equation of the boundary $O'A'B'C'D'E'(O')$ in Figure 11.3. This boundary has two parts: the straight line $E'O'A'$, namely $y' = 0$, and the set of four curves $A'B'C'D'E'$ which correspond to $(x^2-1)(y^2-1) = 0$ in S. Thus the boundary equation is

$$\begin{aligned} 0 &= A(x,y) = 4y'(x^2-1)(y^2-1) = 4y'(x^2y^2 - r^2 + 1) \\ &= y'(r^4 \sin^2(2\theta) - 4r^2 + 4) = y' \left[(r')^6 \sin^2(3\theta') - 4(r')^3 + 4 \right] \\ &= y' \left[(y')^2 \{3(x')^2 - (y')^2\}^2 - 4\{(x')^2 + (y')^2\}^{3/2} + 4 \right]. \end{aligned} \tag{11.64}$$

Dropping dashes again,

$$A(x,y) = y. \left[y^2(3x^2 - y^2)^2 - 4(x^2+y^2)^{3/2} + 4 \right] = 0. \tag{11.65}$$

We now adopt as our approximation to u, using (11.65) for ϕ:

$$u \simeq u_{mn} = A(x,y). \sum_{i=0}^{m} \sum_{j=0}^{n} c_{ij} x^{2i+t} y^j, \tag{11.66}$$

where $t = 0$ or 1, according as we seek a solution which is symmetric or anti-symmetric about OC. For the leading (i.e., largest) λ, we choose $t = 0$.

The rectangle $TUVW$, which encloses the mapped membrane, has sides $O'T$, TU in the x, y directions, respectively, of lengths a, b, say, given by

$$\begin{aligned} a &= O'T = O'B' \cos(\pi/6) = 2^{1/3} 3^{1/2}/2 = 2^{-2/3} 3^{1/2}, \\ b &= TU = O'C' = (2^{1/2})^{2/3} = 2^{1/3}. \end{aligned} \tag{11.67}$$

An appropriate 'interior' collocation method is now simply constructed. We specify that the form of approximation (11.66) should satisfy the PDE (11.63) at the tensor product of $(m+1)(n+1)$ positive zeros of $T^*_{m+1}(x/a) T^*_{n+1}(y/b)$, where a, b are given in (11.67), namely the points

$$\begin{aligned} \{x, y\} &= \left\{ a. \cos^2 \left(\frac{(2k-1)\pi}{4(m+1)} \right), \quad b. \cos^2 \left(\frac{(2l-1)\pi}{4(n+1)} \right) \right\} \\ &\qquad (k = 1, \ldots, m+1;\ l = 1, \ldots, n+1). \end{aligned} \tag{11.68}$$

© 2003 by CRC Press LLC

Table 11.1: Estimates of first 3 eigenvalues of L-membrane

$m = n$	λ	Rayleigh quotient
Functional form $A(x,y)\sum_0^m \sum_0^n x^{2i}y^j$		
4	9.6398	9.6723
6	9.6400	9.6397
8	9.6397	9.6397
Functional form $xA(x,y)\sum_0^m \sum_0^n x^{2i}y^{2j}$		
4	15.2159	
5	15.1978	15.1980
6	15.1974	15.1974
7	15.1974	15.1974
Functional form $A(x,y)\sum_0^m \sum_0^n x^{2i}y^{2j}$		
4	19.8054	
5	19.7394	
6	19.7392	19.7392
7	19.7392	19.7392

This leads to a homogeneous system of $(m+1)(n+1)$ linear equations, which we may write in matrix form as $\mathbf{A}.\mathbf{c} = 0$, for the determination of $\mathbf{c} = \{c_{ij}\}$, where $\mathbf{A}$ depends on λ. The determinant of $\mathbf{A}$ must vanish, thus defining eligible values of λ, corresponding to eigenvalues of the PDE. We have applied the secant method to find the λ nearest to a chosen guess. Results for the first three eigenvalues, taken from Mason (1965), are shown in Table 11.1 together with Rayleigh quotient estimates. Clearly the method is rather successful, and the application serves as an interesting model problem.

Strictly speaking, the collocation method above is not an interior method, since some collocation points are exterior to S although interior to the rectangle $TUVW$. However, the PDE solution does extend continuously across the problem boundaries to these exterior points.

In fairness we should point out that, although the above collocation method is probably at least as effective for this problem as the best finite difference method, such as that of Reid & Walsh (1965), it is not the best method of all. A better method for the present problem is the boundary method, based on separation of variables, due to Fox et al. (1967) and further extended by Mason (1969). This breaks down for regions with more than one re-entrant corner, however, on account of ill-conditioning; a better method is the finite-element/domain-decomposition method described by Driscoll (1997).

© 2003 by CRC Press LLC

11.8.3 Chebyshev basis and discrete orthogonalisation

In the remaining discussion, we concentrate on the use of a Chebyshev polynomial basis for approximation and exploit properties such as discrete orthogonality and the FFT for efficiency. However, it is first appropriate to remind the reader that the classical method of separation of variables provides both a fast boundary method for Laplace's equation and a superposition method, combining interior and boundary methods for the Poisson equation with non-zero boundary conditions.

Separation of variables: basic Dirichlet problem

Consider the basic Dirichlet problem for Laplace's equation on a square, namely

$$\Delta u = 0 \text{ in } S, \tag{11.69a}$$

$$u = g(x, y) \text{ on } \partial S, \tag{11.69b}$$

where ∂S is the square boundary formed by $x = \pm 1, y = \pm 1$, S is its interior and g is defined only on ∂S. Then we may solve (11.69a) analytically for the partial boundary conditions

$$u = g(-1, y) \text{ on } x = -1; \quad u = 0 \text{ on } x = +1, y = -1, y = +1, \tag{11.70}$$

in the form

$$u = \sum_{k=1}^{\infty} a_k \sinh \tfrac{1}{2}k\pi(1 - x) \, \sin \tfrac{1}{2}k\pi(1 - y), \tag{11.71}$$

where a_k are chosen to match the Fourier sine series of $g(-1, y)$ on $x = -1$. Specifically

$$g(-1, y) = \sum_{k=1}^{\infty} b_k \sin \tfrac{1}{2}k\pi(1 - y), \tag{11.72}$$

where

$$b_k = 2 \int_{-1}^{1} g(-1, y) \, \sin \tfrac{1}{2}k\pi(1 - y) \, \mathrm{d}y, \tag{11.73}$$

and hence a_k is given by

$$a_k = b_k [\sinh k\pi]^{-1}. \tag{11.74}$$

Clearly we can define three more solutions of (11.69a), analogous to (11.71), each of which matches $g(x, y)$ on one side of ∂S and is zero on the remainder of ∂S. If we sum these four solutions then we obtain the analytical solution of (11.69a) and (11.69b). For an efficient numerical solution, the Fourier series should be truncated and evaluated by using the FFT [see Section 4.7].

© 2003 by CRC Press LLC

Chebyshev basis: Poisson problem

The Poisson Problem can be posed in a slightly more general way than in Section 11.8.3, while still permitting efficient treatment. In particular we may introduce two general functions, f as the right-hand side of the PDE, and g as the boundary function, as follows.

$$\Delta u = f(x,y) \text{ in } S, \quad u = g(x,y) \text{ in } \partial S, \tag{11.75}$$

where S and ∂S denote the usual square $\{-1 \le x \le 1, -1 \le y \le 1\}$ and its boundary. Then we may eliminate g (and replace it by zero) in (11.75), by superposing two problem solutions

$$u = u_1 + u_2, \tag{11.76}$$

where u_1 is the solution of the Laplace problem ((11.75) with $f \equiv 0$) and u_2 is the solution of the simple Poisson problem ((11.75) with $g \equiv 0$), so that

$$\Delta u_1 = 0 \text{ in } S, \quad u_1 = g(x,y) \text{ on } \partial S, \tag{11.77a}$$

$$\Delta u_2 = f(x,y) \text{ in } S, \quad u_2 = 0 \text{ on } \partial S. \tag{11.77b}$$

We gave details above of the determination of u_1 from four Fourier sine series expansions based on separation of variables, as per (11.71) above. We may therefore restrict attention to the problem (11.77b) defining u_2, which we now rename u.

We now re-address (11.77b), which was discussed in Section 11.8.1 using a power basis, and, for greater efficiency and stability, we adopt a boundary method based on a Chebyshev polynomial representation

$$u_{mn} = (x^2-1)(y^2-1) \sum_{i=0}^{m-2}{}' \sum_{j=0}^{n-2}{}' c_{ij} T_i(x) T_j(y), \tag{11.78}$$

or equivalently, again to ensure that $u = 0$ on ∂S,

$$u_{mn} = \sum_{i=2}^{m} \sum_{j=2}^{n} a_{ij} [T_i(x) - T_{i \bmod 2}(x)] \, [T_j(y) - T_{j \bmod 2}(y)]. \tag{11.79}$$

Now, as in Problem 16 of Chapter 2,

$$\frac{\partial^2}{\partial x^2} T_i(x) = \sum_{\substack{r=0 \\ i-r \text{ even}}}^{i-2} (i-r) i (i+r) T_r(x), \tag{11.80}$$

and hence

$$\Delta u_{mn} = \sum_{i=2}^{m} \sum_{j=2}^{n} a_{ij} \left[\sum_{\substack{r=0 \\ i-r \text{ even}}}^{i-2} (i-r) i (i+r) T_r(x) \, (T_j(y) - T_{j \bmod 2}(y)) + \right.$$

© 2003 by CRC Press LLC

$$+ \sum_{\substack{s=0 \\ j-s \text{ even}}}^{j-2} (j-s)j(j+s)T_s(y)\,(T_i(x) - T_{i \bmod 2}(x)) \Bigg]$$
$$= f \tag{11.81}$$

Now define collocation points $\{x_k(k = 1, \ldots, m-1)\}$ and $\{y_l(l = 1, \ldots, n-1)\}$ to be, respectively, the zeros of $T_{m-1}(x)$ and $T_{n-1}(y)$. Then discrete orthogonality gives, for p, r less than $m-1$ and q, s less than $n-1$,

$$2(m+1)^{-1} \sum_{k=1}^{m-1} T_p(x_k)T_r(x_k) = \begin{cases} 2, & p = r = 0, \\ 1, & p = r \neq 0, \\ 0, & p \neq r, \end{cases} \tag{11.82a}$$

$$2(n+1)^{-1} \sum_{l=1}^{n-1} T_q(y_l)T_s(y_l) = \begin{cases} 2, & q = s = 0, \\ 1, & q = s \neq 0, \\ 0, & q \neq s. \end{cases} \tag{11.82b}$$

Evaluating (11.81) at (x_k, y_ℓ), multiplying by $4[(m-1)(n-1)]^{-1}$ and by $T_p(x_k)T_q(y_\ell)$ for $p = 0, \ldots, m-2; q = 0, \ldots, n-2$, summing over k, ℓ, and using discrete orthogonality, we obtain

$$A_{pq} + B_{pq} = 4[(m-1)(n-1)]^{-1} \sum_{k=1}^{m-1} \sum_{\ell=1}^{n-1} f(x_k, y_\ell)T_p(x_k)T_q(y_\ell), \tag{11.83}$$

where

$$A_{pq} = \begin{cases} \displaystyle\sum_{\substack{i=2 \\ i-p \text{ even}}}^{m} a_{iq}(i-p)i(i+p), & q \geq 2, \\ \displaystyle -\sum_{\substack{i=2 \\ i-p \text{ even}}}^{m} \sum_{\substack{j=3 \\ j \text{ odd}}}^{n} a_{ij}(i-p)i(i+p), & q = 1, \\ \displaystyle -2 \sum_{\substack{i=2 \\ i-p \text{ even}}}^{m} \sum_{\substack{j=2 \\ j \text{ even}}}^{n} a_{ij}(i-p)i(i+p), & q = 0, \end{cases}$$

$$B_{pq} = \begin{cases} \displaystyle\sum_{\substack{j=2 \\ j-q \text{ even}}}^{n} a_{pj}(j-q)j(j+q), & p \geq 2, \\ \displaystyle -\sum_{\substack{j=2 \\ j-q \text{ even}}}^{n} \sum_{\substack{i=3 \\ i \text{ odd}}}^{m} a_{ij}(j-q)j(j+q), & p = 1, \\ \displaystyle -2 \sum_{\substack{j=2 \\ j-q \text{ even}}}^{n} \sum_{\substack{i=2 \\ i \text{ even}}}^{m} a_{ij}(j-q)j(j+q), & p = 0. \end{cases} \tag{11.84}$$

© 2003 by CRC Press LLC

This system of linear equations for a_{ij} is very sparse, having between 2 and $(m+n-2)/2$ entries in each row of the matrix for $p, q \geq 2$. It is only the equations where "boundary effects" enter (for $p = 0, 1;\ q = 0, 1$), that fill out the matrix entries into alternate rows and/or columns. Note also that all right-hand sides are discrete Chebyshev transforms, which could be evaluated by adopting FFT techniques.

The border effects can be neatly avoided for this particular Poisson problem, by adopting instead a matrix method based on differentiation matrices, in which the unknowns of the problem become the solution values at Chebyshev nodes, rather than the solution coefficients. This approach was introduced in Section 10.5.3 of Chapter 10 for ODEs and is particularly convenient for some relatively simple problems. We now see its advantages for the present problem.

11.8.4 Differentiation matrix approach: Poisson problem

To illustrate this approach, we follow the treatment of Trefethen (2000), setting $m = n$ and adopting as collocation points the tensor product of the zeros of $(1-x^2)U_{n-1}(x)$ and the zeros of $(1-y^2)U_{n-1}(y)$. The reader is referred to Section 10.5.2 for a detailed definition of the $(n+1) \times (n+1)$ differentiation matrix $\mathbf{D} \equiv \mathbf{D}_n$, which transforms all u values at collocation points into approximate derivative values at the same points, by forming linear combinations of the u values. The problem is significantly simplified by noting that the border elements of $\mathbf{D}$, namely the first and last rows and columns of $\mathbf{D}_n$, correspond to zero boundary values and may therefore be deleted to give an active $(n-1) \times (n-1)$ matrix $\tilde{\mathbf{D}}_n$.

For the Poisson problem

$$\Delta u = f(x, y) \text{ in } S : \{-1 \leq x \leq 1, -1 \leq y \leq 1\}, \tag{11.85a}$$

$$u = 0 \text{ on } \partial S : \{x = \pm 1, y = \pm 1\}, \tag{11.85b}$$

the method determines a vector $\mathbf{u}$ of approximate solutions at the interior collocation points (compare (10.52) with $\mathbf{e}_0 = \mathbf{e}_n = \mathbf{0}$) by solving

$$\mathbf{E}_n \mathbf{u} = \mathbf{f}_n \tag{11.86}$$

where

$$\mathbf{E}_n = \mathbf{I} \otimes \tilde{\mathbf{D}}_n^2 + \tilde{\mathbf{D}}_n^2 \otimes \mathbf{I} \tag{11.87}$$

and $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker (tensor) product, illustrated by the example

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \otimes \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \left(\begin{array}{cc|cc} a\alpha & a\beta & b\alpha & b\beta \\ a\gamma & a\delta & b\gamma & b\delta \\ \hline c\alpha & c\beta & d\alpha & d\beta \\ c\gamma & c\delta & d\gamma & d\delta \end{array}\right).$$

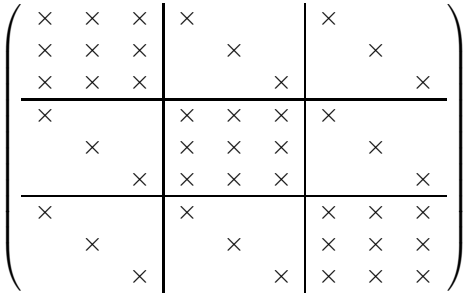
© 2003 by CRC Press LLC

For example, for $n = 3$, this gives (see (10.49) *et seq.*)

$$\begin{aligned}\mathbf{E}_3 &= \mathbf{I} \otimes \tilde{\mathbf{D}}_3^2 \;+\; \tilde{\mathbf{D}}_3^2 \otimes \mathbf{I} \\ &= \begin{pmatrix} -14 & 6 & -2 & & & & & & \\ 4 & -6 & 4 & & & & & & \\ -2 & 6 & -14 & & & & & & \\ & & & -14 & 6 & -2 & & & \\ & & & 4 & -6 & 4 & & & \\ & & & -2 & 6 & -14 & & & \\ & & & & & & -14 & 6 & -2 \\ & & & & & & 4 & -6 & 4 \\ & & & & & & -2 & 6 & -14 \end{pmatrix} + \\ &\quad + \begin{pmatrix} -14 & & & 6 & & & -2 & & \\ & -14 & & & 6 & & & -2 & \\ & & -14 & & & 6 & & & -2 \\ 4 & & & -6 & & & 4 & & \\ & 4 & & & -6 & & & 4 & \\ & & 4 & & & -6 & & & 4 \\ -2 & & & 6 & & & -14 & & \\ & -2 & & & 6 & & & -14 & \\ & & -2 & & & 6 & & & -14 \end{pmatrix}. \end{aligned} \tag{11.88}$$

Clearly the matrix $\mathbf{E}_n$ is sparse and easily calculable. Hence (11.86) may readily be solved by Gauss elimination, using for example Matlab's efficient system. The right-hand side of (11.86) is simply obtained by evaluating f at the interior collocation points. The shape (location of the non-zero terms) of

Figure 11.4: Shape of matrix $\mathbf{E}_n$ for $n = 3$



the matrix $\mathbf{E}_n$ is illustrated in Figure 11.4 and consists of a diagonal of square matrices, flanked by diagonal matrices. The matrix is very sparse in all rows, having $2n - 3$ non-zero entries out of $(n - 1)^2$.

© 2003 by CRC Press LLC

This Differentiation Matrix method is very attractive and efficient for this problem, and should always be given consideration in problems that respond to it. We now turn our attention to a more general problem, with non-zero boundary conditions.

11.8.5 Explicit collocation for the quasilinear Dirichlet problem: Chebyshev basis

We now continue to exploit the better conditioning of a Chebyshev polynomial basis, but we also consider the greater generality of a Dirichlet problem for a quasilinear elliptic equation on a square, namely

$$Lu \equiv a.u_{xx} + b.u_{xy} + c.u_{yy} + d.u_x + e.u_y = f \text{ in } D : |x| \le 1, |y| \le 1, \quad (11.89a)$$

$$u = g(x, y) \text{ on } \partial D : \{x = \pm 1, y = \pm 1\}, \quad (11.89b)$$

where a, b, c, d, e, f are functions of x and y defined in D, $g(x, y)$ is defined on ∂D only, and where, to ensure ellipticity,

$$a.c \ge b^2 \text{ for all } (x, y) \text{ in } D. \quad (11.90)$$

This is an extension of recent work by Mason & Crampton (2002).

For generality we do not attempt to include the boundary conditions (11.89b) implicitly in the form of approximation, but rather we represent them by a set of constraints at a discrete set of selected points, namely Chebyshev zeros on the boundary. Moreover we adopt a Chebyshev polynomial basis for u:

$$u \simeq u_{mn} = {\sum_{i=0}^{m}}' {\sum_{j=0}^{n}}' a_{ij} T_i(x) T_j(y). \quad (11.91)$$

As it happens, we find that the apparently most logical collocation procedure, similar to that of Section 11.8.1 above, for approximately solving (11.89a), (11.89b) in the form (11.91), leads to a singular matrix and requires modification. More details about this follow as the method develops. The fundamental idea that we use is that, since u_{mn}, given by (11.91), has $(m + 1)(n + 1)$ undetermined coefficients, we expect to be able to generate an appropriate set of equations for a_{ij} if we form $(m - 1)(n - 1)$ equations by collocating (11.89a) at a tensor product of Chebyshev zeros and a further $2m + 2n$ equations by collocating (11.89b) at Chebyshev zeros on the boundary. It is in the formation of the latter boundary equations that difficulties arise, and so we consider these equations first, noting that they are completely independent of the specification $Lu = f$ of the elliptic equation (11.89a).

To form the $2m + 2n$ boundary equations for a_{ij}, we set u_{mn} equal to g at the zeros, X_k $(k = 1, \ldots, m)$ and Y_ℓ $(\ell = 1, \ldots, n)$ of $T_m(x)$ and $T_n(y)$,

© 2003 by CRC Press LLC

respectively, on $y = \pm 1$ and $x = \pm 1$, respectively. This gives the two pairs of equations

$$\begin{aligned}
\sum_{i=0}^{m}{}' \sum_{j=0}^{n}{}' a_{ij} T_i(X_k) T_j(\pm 1) &= g(X_k, \pm 1), \\
\sum_{i=0}^{m}{}' \sum_{j=0}^{n}{}' a_{ij} T_i(\pm 1) T_j(Y_\ell) &= g(\pm 1, Y_\ell).
\end{aligned} \tag{11.92}$$

If we add/subtract these pairs of equations, noting that $T_j(1) = 1$ and that $T_j(-1) = (-1)^j$, we deduce that

$$\begin{aligned}
\sum_{i=0}^{m}{}' \sum_{\substack{j=0 \\ j \text{ even}}}^{n}{}' a_{ij} T_i(X_k) &= G_{k0} \equiv \tfrac{1}{2}(g(X_k, 1) + g(X_k, -1)), (k = 1, \ldots, m) \\
\sum_{i=0}^{m}{}' \sum_{\substack{j=1 \\ j \text{ odd}}}^{n} a_{ij} T_i(X_k) &= G_{k1} \equiv \tfrac{1}{2}(g(X_k, 1) - g(X_k, -1)), (k = 1, \ldots, m) \\
\sum_{\substack{i=0 \\ i \text{ even}}}^{m}{}' \sum_{j=0}^{n}{}' a_{ij} T_j(Y_\ell) &= H_{\ell 0} \equiv \tfrac{1}{2}(g(1, Y_\ell) + g(-1, Y_\ell)), (\ell = 1, \ldots, n) \\
\sum_{\substack{i=1 \\ i \text{ odd}}}^{m} \sum_{j=0}^{n}{}' a_{ij} T_j(Y_\ell) &= H_{\ell 1} \equiv \tfrac{1}{2}(g(1, Y_\ell) - g(-1, Y_\ell)), (\ell = 1, \ldots, n)
\end{aligned} \tag{11.93}$$

where the arrays G_{kp}, $H_{\ell q}$ are defined above for $p = 0, 1$; $q = 0, 1$.

Now, defining w_i to be $2/(i+1)$ for all i, multiplying the first pair of equations in (11.93) by $w_m T_r(X_k)$ and summing over k, multiplying the second pair of equations by $w_n T_s(Y_\ell)$ and summing over ℓ, and exploiting discrete orthogonality, it follows that

$$\begin{aligned}
R_{r0} \equiv \sum_{\substack{j=0 \\ j \text{ even}}}^{n}{}' a_{rj} &= J_{r0} \equiv w_m \sum_{k=1}^{m} T_r(X_k) G_{k0}, (r = 0, \ldots, m-1) \\
R_{r1} \equiv \sum_{\substack{j=1 \\ j \text{ odd}}}^{n} a_{rj} &= J_{r1} \equiv w_m \sum_{k=1}^{m} T_r(X_k) G_{k1}, (r = 0, \ldots, m-1) \\
C_{s0} \equiv \sum_{\substack{i=0 \\ i \text{ even}}}^{m}{}' a_{is} &= K_{s0} \equiv w_n \sum_{\ell=1}^{n} T_s(Y_\ell) H_{\ell 0}, (s = 0, \ldots, n-1)
\end{aligned}$$

© 2003 by CRC Press LLC

$$C_{s1} \equiv \sum_{\substack{i=1 \\ i \text{ odd}}}^{m} a_{is} = K_{s1} \equiv w_n \sum_{\ell=1}^{n} T_s(Y_\ell) H_{\ell 1}, (s = 0, \ldots, n-1) \quad (11.94)$$

where R, C, J, K are defined to form left-hand sides (R, C) or right-hand sides (J, K) of the relevant equations. In addition each R or C is a linear sum of alternate elements in a row or column, respectively, of the matrix $\mathbf{A} = [a_{ij}]$.

Now we claim that this set of $2m+2n$ linear equations (11.94) in $a_{00}, \ldots, a_{mn}$ is not of rank $2m + 2n$ but rather of rank $2m + 2n - 1$. Indeed, it is easy to verify that a sum of alternate rows of $\mathbf{A}$ equals a sum of alternate columns; specifically

$$\sum_{\substack{i=0 \\ n-i \text{ odd}}}^{n-1}{}' R_{ip} = \sum_{\substack{j=0 \\ m-j \text{ odd}}}^{m-1}{}' C_{jq} = \sum_{\substack{i=0 \\ m-i \text{ odd}}}^{m-1}{}' \sum_{\substack{j=0 \\ n-j \text{ odd}}}^{n-1}{}' a_{ij}, \quad (11.95)$$

where $p = 0, 1$ for $m - 1$ even, odd, respectively, and $q = 0, 1$ for $n - 1$ even, odd, respectively. For example, for $m = n = 4$,

$$R_{11} + R_{31} = C_{11} + C_{31} = a_{11} + a_{13} + a_{31} + a_{33}, \quad (11.96)$$

and, for $m = n = 3$,

$$\tfrac{1}{2}R_{00} + R_{20} = \tfrac{1}{2}C_{00} + C_{20} = \tfrac{1}{2}(\tfrac{1}{2}a_{00} + a_{02} + a_{20}) + a_{22}. \quad (11.97)$$

Clearly we must delete one equation from the set (11.94) and add an additional independent equation in order to restore full rank. For simplicity we shall only discuss the cases where m, n are both even or both odd, leaving the even/odd and odd/even cases to the reader.

For m, n both even, we delete $C_{11} = K_{11}$ from (11.94) and add an averaged "even/even" boundary collocation equation

$$\begin{aligned} &\tfrac{1}{4}[u(1,1) + u(-1,1) + u(-1,-1) + u(1,-1)] = \\ &\lambda_{00} := \tfrac{1}{4}[g(1,1) + g(-1,1) + g(-1,-1) + g(1,-1)]. \end{aligned} \quad (11.98)$$

This simplifies to

$$\sum_{\substack{i=0 \\ i \text{ even}}}^{m}{}' R_{i0} = \lambda_{00} \quad (11.99)$$

where R_{m0} is defined by extending the definition (11.94) of R_{r0} to $r = m$ and where λ_{00} is as defined in (11.98). We may eliminate every R except R_{m0} from this equation, by using (11.94), to give a simplified form for the extra equation

$$R_{m0} = J_{m0} = \lambda_{00} - \sum_{\substack{i=0 \\ i \text{ even}}}^{m-2}{}' J_{i0} \quad (11.100)$$

© 2003 by CRC Press LLC

where the right-hand side J_{m0} is defined as shown.

For m, n both odd, we delete $C_{00} = K_{00}$ from (11.94) and add an averaged "odd/odd" boundary collocation equation

$$\begin{aligned} &\tfrac{1}{4}[u(1,1) - u(-1,1) + u(-1,-1) - u(1,-1)] = \\ &\lambda_{11} := \tfrac{1}{4}[g(1,1) - g(-1,1) + g(-1,-1) - g(1,-1)]. \end{aligned} \tag{11.101}$$

This simplifies to

$$\sum_{\substack{j=1 \\ j \text{ odd}}}^{n} C_{j1} = \lambda_{11} \tag{11.102}$$

where C_{n1} is defined by extending the definition (11.94) of C_{s1} to $s = n$ and where λ_{11} is as defined in (11.101).

We may eliminate every C except C_{n1} from this equation, by using (11.94), to give a simplified form for the extra equation

$$C_{n1} = K_{n1} \equiv \lambda_{11} - \sum_{\substack{j=1 \\ j \text{ odd}}}^{n-2} K_{j1} \tag{11.103}$$

where the right-hand side K_{n1} is defined as shown.

We now have $2m + 2n$ non-singular equations for the coefficients a_{ij}, and it remains to handle the elliptic equation by collocation at $(m-1)(n-1)$ suitable Chebyshev points in D.

For a general quasilinear equation we should set $Lu = f$ at a tensor product of $(m-1)\times(n-1)$ Chebyshev zeros, giving the same number of linear algebraic equations for $\{a_{ij}\}$, and these equations together with the $2m+2n$ boundary collocation equations would then be solved as a full system.

For simplicity, and so that we can give fuller illustrative details, we concentrate on the Poisson equation, as a special example of (11.89a), corresponding to the form

$$Lu \equiv \Delta u \equiv u_{xx} + u_{yy} = f(x,y) \text{ in } D. \tag{11.104}$$

Now second derivatives of Chebyshev sums are readily seen (see Chapter 2) to take the form

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} T_k(x) = {\sum_{\substack{r=0 \\ k-r \text{ even}}}^{k-2}}' (k-r)k(k+r)T_r(x) \quad (k \geq 2), \tag{11.105a}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2} T_\ell(y) = {\sum_{\substack{s=0 \\ \ell-s \text{ even}}}^{\ell-2}}' (\ell-s)\ell(\ell+s)T_s(y) \quad (\ell \geq 2). \tag{11.105b}$$

© 2003 by CRC Press LLC

Hence, on substituting (11.91) into (11.104), we obtain

$$\begin{aligned}\Delta u_{mn} &= \sum_{k=2}^{m} {\sum_{\ell=0}^{n}}' a_{k\ell} T_\ell(y) \sum_{\substack{r=0 \\ k-r \text{ even}}}^{k-2}{}' (k-r)k(k+r)T_r(x) \\ &\quad + \sum_{\ell=2}^{n} {\sum_{k=0}^{m}}' a_{k\ell} T_k(x) \sum_{\substack{s=0 \\ \ell-s \text{ even}}}^{\ell-2}{}' (\ell-s)\ell(\ell+s)T_s(y) \\ &= A_{mn}, \text{ say.}\end{aligned} \tag{11.106}$$

Setting Δu_{mn} equal to $f(x,y)$ at the abscissae x_i, y_j , where x_i are zeros of $T_{m-1}(x)$ and y_j are zeros of $T_{n-1}(y)$ (for $i = 1, \ldots, m-1$; $j = 1, \ldots, n-1$), multiplying by $T_p(x_i)T_q(y_j)$, and summing over i, j, we deduce that, for every $p = 0, \ldots, m-2$; $q = 0, \ldots, n-2$:

$$\begin{aligned}E_{pq} &\equiv \sum_{i=1}^{m-1}\sum_{j=1}^{n-1} A_{mn}(x_i, y_j)T_p(x_i)T_q(yj) \\ &= \sum_{i=1}^{m-1}\sum_{j=1}^{n-1} f(x_i, y_j)T_p(x_i)T_q(y_j) \\ &\equiv f_{pq},\end{aligned} \tag{11.107}$$

where f_{pq} represents the discrete Chebyshev transform of f with respect to $T_p(x)T_q(y)$. Substituting for A_{mn},

$$\begin{aligned}E_{pq} &= \sum_{i=1}^{m-1}\sum_{j=1}^{n-1}\sum_{k=2}^{m}{\sum_{\ell=0}^{n}}' a_{k\ell}T_\ell(y_j)T_q(y_j) \sum_{r=0}^{k-2}{}' (k-r)k(k+r)T_r(x_i)T_p(x_i) \\ &\quad + \sum_{i=1}^{m-1}\sum_{j=1}^{n-1}\sum_{\ell=2}^{n}{\sum_{k=0}^{n}}' a_{k\ell}T_k(x_i)T_p(x_i) \sum_{s=0}^{\ell-2}{}' (\ell-s)\ell(\ell+s)T_s(y_j)T_q(j_j) \\ &= \sum_{k=2}^{m}{\sum_{\ell=0}^{n}}'\sum_{j=1}^{n-1} T_\ell(y_j)T_q(y_j)a_{k\ell} \sum_{\substack{r=0 \\ k-r \text{ even}}}^{k-2}{}' (k-r)k(k+r) \sum_{i=1}^{m-1} T_r(x_i)T_p(x_i) \\ &\quad + \sum_{\ell=2}^{n}{\sum_{k=0}^{m}}'\sum_{i=1}^{m-1} T_k(x_i)T_p(x_i)a_{k\ell} \sum_{\substack{s=0 \\ \ell-s \text{ even}}}^{\ell-2}{}' (\ell-s)\ell(\ell+s) \sum_{j=1}^{n-1} T_s(y_j)T_q(y_j).\end{aligned} \tag{11.108}$$

© 2003 by CRC Press LLC

Using the discrete orthogonality property that, for example,

$$\sum_{j=1}^{n-1} T_\ell(y_j)T_q(y_j) = \begin{cases} 0, & \ell \neq q \\ (n-1)/2, & \ell = q \neq 0 \\ n-1, & \ell = q = 0 \end{cases},$$

we deduce that

$$\begin{aligned} E_{pq} \;=\; & \left[\sum_{\substack{k=p+2 \\ k-p \text{ even}}}^{m} \tfrac{1}{2}(n-1)a_{kq} + \sum_{j=1}^{n-1} T_{n-1}(y_j)T_q(y_j)a_{k,n-1} \right. \\ & \left. + \sum_{j=1}^{n-1} T_n(y_j)T_q(y_j)a_{kn} \right] \tfrac{1}{2}(m-1)(k-p)k(k+p) \\ & + \left[\sum_{\substack{\ell=q+2 \\ \ell-q \text{ even}}}^{n} \tfrac{1}{2}(m-1)a_{p\ell} + \sum_{i=1}^{m-1} T_{m-1}(x_i)T_p(x_i)a_{m-1,\ell} \right. \\ & \left. + \sum_{i=1}^{m-1} T_m(x_i)T_p(x_i)a_{m\ell} \right] \tfrac{1}{2}(n-1)(\ell-q)\ell(\ell+q). \end{aligned} \tag{11.109}$$

Now, by the definition of x_i, y_j, we know that $T_{m-1}(x_i)$ and $T_{n-1}(y_j)$ are zero. Also, using the three-term recurrence at x_i,

$$T_m(x_i) = 2xT_{m-1}(x) - T_{m-2}(x_i) = -T_{m-2}(x_i), \quad T_n(y_j) = -T_{n-2}(y_j). \tag{11.110}$$

Substituting these values into (11.109), using discrete orthogonality, and using the Kronecker delta notation

$$\delta_{rs} = 1 \ (r = s), \quad \delta_{rs} = 0 \ (r \neq s), \tag{11.111}$$

we deduce that

$$\begin{aligned} E_{pq} \;\equiv\; & \tfrac{1}{4}(m-1)(n-1) \left[\sum_{\substack{k=p+2 \\ k-p \text{ even}}}^{m} \left(a_{kq} - \delta_{q,n-2}a_{kn}\right)(k-p)k(k+p) + \right. \\ & \left. + \sum_{\substack{l=q+2 \\ l-q \text{ even}}}^{n} \left(a_{pl} - \delta_{p,m-2}a_{ml}\right)(l-q)l(l+q) \right] \\ = \; & f_{pq} \quad (p = 0, \ldots, m-2; \; q = 0, \ldots, n-2). \end{aligned} \tag{11.112}$$

© 2003 by CRC Press LLC

For example, for $m = n = 3$ we have this set of four collocation equations:

$$\begin{aligned}
E_{11} &\equiv \frac{4}{4}[(a_{31} - a_{33}) + (a_{13} - a_{33})]2.3.4 \\
&= 24(a_{13} + a_{31} - 2a_{33}) = f_{11} = 24F_{11}, \\
E_{10} &\equiv a_{30}2.3.4 + (a_{12} - a_{32})2.2.2 \\
&= 8(a_{12} + 3a_{30} - a_{32}) = f_{10} = 8F_{10}, \\
E_{01} &\equiv (a_{21} - a_{23})2.2.2 + a_{03}2.3.4 \\
&= 8(3a_{03} + a_{21} - a_{23}) = f_{01} = 8F_{01}, \\
E_{00} &\equiv a_{20}2.2.2 + a_{02}2.2.2 \\
&= 8(a_{02} + a_{20}) = f_{00} = 8F_{00},
\end{aligned} \tag{11.113}$$

where F_{ij} are defined as shown by scaling f_{ij}. For $m = n = 4$, the system (11.112) gives the following nine equations, where we leave the reader to confirm the details:

$$\begin{aligned}
E_{22} &\equiv 108(a_{24} + a_{42} - 2a_{44}) \\
&= f_{22} = 108F_{22}, \\
E_{21} &\equiv 54(a_{23} + 2a_{41} - a_{43}) \\
&= f_{21} = 54F_{21}, \\
E_{20} &\equiv 18(a_{22} + 8a_{24} + 6a_{40} - a_{42} - 8a_{44}) \\
&= f_{20} = 18F_{20}, \\
E_{12} &\equiv 54(2a_{14} + a_{32} - a_{34}) \\
&= f_{12} = 54F_{12}, \\
E_{11} &\equiv 54(a_{13} + a_{31}) \\
&= f_{11} = 54F_{11}, \\
E_{10} &\equiv 18(a_{12} + 8a_{12} + 3a_{30}) \\
&= f_{10} = 18F_{10}, \\
E_{02} &\equiv 18(6a_{04} + a_{22} - a_{24} + 8a_{42} - 8a_{44}) \\
&= f_{02} = 18F_{02}, \\
E_{01} &\equiv 18(3a_{03} + a_{21} + 8a_{41}) \\
&= f_{01} = 18F_{01}, \\
E_{00} &\equiv 18(a_{02} + 8a_{04} + a_{20} + 8a_{40}) \\
&= f_{00} = 18F_{00}.
\end{aligned} \tag{11.114}$$

For $m = n = 4$, the complete set of 25 collocation equations, 16 boundary equations and 9 interior PDE equations, namely (11.94) for $m = n = 4$ and

© 2003 by CRC Press LLC

(11.114), may be written in the matrix form

$$\mathbf{Ma} = \mathbf{b}, \tag{11.115}$$

where $\mathbf{M}$ is the matrix of equation entries and $\mathbf{a}$ is the column vector of approximation coefficients

$$\mathbf{a} = (a_{00}, a_{01}, \ldots, a_{04}, a_{10}, a_{11}, \ldots, a_{14}, a_{20}, \ldots, a_{30}, \ldots, a_{40}, \ldots, a_{44})' \tag{11.116}$$

and $\mathbf{b}$ is the set of right-hand sides, either boundary or PDE terms, in appropriate order. In Table 11.2 we display the matrices $\mathbf{M}$, $\mathbf{a}$, $\mathbf{b}$ for $m = n = 4$, blank entries denoting zeros. The column of symbols to the left of $\mathbf{M}$ indicates which equation has been used to construct each row. Note that we have ordered the equations to give a structure in $\mathbf{M}$ as close to lower triangular as possible. The order used is based on:

$$R_{4*}, R_{3*}, E_{2*}, R_{2*}, E_{1*}, R_{1*}, E_{0*}, R_{0*}, C_{3*}, C_{2*}, C_{1*}, C_{0*} \tag{11.117}$$

where $*$ is a wild subscript, E indicates a PDE term, and R, C indicate boundary conditions.

On studying Table 11.2, some important facts emerge. The coefficients a_{ij} appearing in any equation are exclusively in one of the four symmetry classes: i, j both even, i, j both odd, i odd and j even, and i even and j odd. Thus the set of 25 equations can be separated into four wholly independent subsystems, respectively involving 4 subsets of a_{ij}. These four subsystems are shown in Table 11.3 for $m = n = 4$, and they consist of 8,5,6,6 equations in 9,4,6,6 coefficients a_{ij}, respectively.

This immediately confirms that we have a surplus equation in the odd/odd subsystem (for $a_{11}, a_{13}, a_{31}, a_{33}$) and one too few equations in the even/even subsystem. As proposed in earlier discussions, we therefore delete equation C_{11} and replace it by equation R_{40}, as indicated in Table 11.2.

The extremely sparse nature of the matrix $\mathbf{M}$ is clear from Table 11.2, and moreover the submatrices formed from even and/or odd subsystems remain relatively sparse, as is illustrated in Tables 11.3 to 11.6.

The odd/odd subsystem (for $m = n = 4$) (in Table 11.6) is remarkably easy to solve in the case $g \equiv 0$ of zero boundary conditions, when $J_{**} = K_{**} = 0$. The solution is then

$$-a_{11} = a_{13} = a_{31} = -a_{33} = \tfrac{1}{2} F_{11}. \tag{11.118}$$

In Table 11.7, we also show the full algebraic system for the odd degrees $m = n = 3$, and in Tables 11.8 to 11.11 the equations are separated into their four even/odd subsystems. The equation C_{00} is noted and is to be deleted, while the equation C_{31} has been added. Equations are again ordered so as to optimise sparsity above the diagonal. The $m = n = 3$ subsystems are easily

© 2003 by CRC Press LLC

Table 11.2: Full collocation matrix—Poisson problem: $m = n = 4$

R_{40}																					$\frac{1}{2}$	0	1	0	1	J_{40}
R_{31}																0	1	0	1	0						J_{31}
R_{30}																$\frac{1}{2}$	0	1	0	1						J_{30}
E_{22}											0	0	0	0	1						0	0	1	0	-2	F_{22}
E_{21}											0	0	0	1	0						0	2	0	-1	0	F_{21}
E_{20}											0	0	1	0	8						6	0	-1	0	-8	F_{20}
R_{21}											0	1	0	1	0											J_{21}
R_{20}											$\frac{1}{2}$	0	1	0	1											J_{20}
E_{12}						0	0	0	0	2						0	0	1	0	-1						F_{12}
E_{11}						0	0	0	1	0						0	1	0	0	0						F_{11}
E_{10}						0	0	1	0	8						3	0	0	0	0						F_{10}
R_{11}						0	1	0	1	0																J_{11}
R_{10}						$\frac{1}{2}$	0	1	0	1																J_{10}
E_{02}	0	0	0	0	6						0	0	1	0	-1						0	0	8	0	-8	F_{02}
E_{01}	0	0	0	3	0						0	1	0	0	0						0	8	0	0	0	F_{01}
E_{00}	0	0	1	0	8						1	0	0	0	0						8	0	0	0	0	F_{00}
R_{01}	0	1	0	1	0																					J_{01}
R_{00}	$\frac{1}{2}$	0	1	0	1																					J_{00}
C_{30}	0	0	0	$\frac{1}{2}$	0						0	0	0	1	0						0	0	0	1	0	K_{30}
C_{31}						0	0	0	1	0						0	0	0	1	0						K_{31}
C_{21}						0	0	1	0	0						0	0	1	0	0						K_{21}
C_{20}	0	0	$\frac{1}{2}$	0	0						0	0	1	0	0						0	0	1	0	0	K_{20}
C_{10}	0	$\frac{1}{2}$	0	0	0						0	1	0	0	0						0	1	0	0	0	K_{10}
C_{11}						0	1	0	0	0						0	1	0	0	0						K_{11}
C_{01}						1	0	0	0	0						1	0	0	0	0						K_{01}
C_{02}	$\frac{1}{2}$	0	0	0	0						1	0	0	0	0						1	0	0	0	0	K_{02}

Unknown vector (multiplying the matrix, $=$ right-hand side):

$$\begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{20} & a_{21} & a_{22} & a_{23} & a_{24} & a_{30} & a_{31} & a_{32} & a_{33} & a_{34} & a_{40} & a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}^T$$

© 2003 by CRC Press LLC

Table 11.3: $m = n = 4$, partial system odd/even in x/y

$$\begin{array}{c} R_{30} \\ E_{12} \\ E_{10} \\ R_{10} \\ C_{21} \\ C_{01} \end{array} \left[\begin{array}{cccccc} & & & \frac{1}{2} & 1 & 1 \\ 0 & 0 & 2 & 0 & 1 & -1 \\ 0 & 1 & 8 & 3 & 0 & 0 \\ \frac{1}{2} & 1 & 1 & & & \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{array}\right] \left[\begin{array}{c} a_{10} \\ a_{12} \\ a_{14} \\ a_{30} \\ a_{32} \\ a_{34} \end{array}\right] = \left[\begin{array}{c} J_{30} \\ F_{12} \\ F_{10} \\ J_{10} \\ K_{21} \\ K_{01} \end{array}\right]$$

Table 11.4: $m = n = 4$, partial system even/odd in x/y

$$\begin{array}{c} E_{21} \\ R_{21} \\ E_{01} \\ R_{01} \\ C_{30} \\ C_{10} \end{array} \left[\begin{array}{cccccc} & & 0 & 1 & 2 & -1 \\ & & 1 & 1 & & \\ 0 & 3 & 1 & 0 & 8 & 0 \\ 1 & 1 & & & & \\ 0 & \frac{1}{2} & 0 & 1 & 0 & 1 \\ \frac{1}{2} & 0 & 1 & 0 & 1 & 0 \end{array}\right] \left[\begin{array}{c} a_{01} \\ a_{03} \\ a_{21} \\ a_{23} \\ a_{41} \\ a_{43} \end{array}\right] = \left[\begin{array}{c} F_{21} \\ J_{21} \\ F_{01} \\ J_{01} \\ K_{30} \\ K_{10} \end{array}\right]$$

Table 11.5: $m = n = 4$, partial system even/even in x/y

$$\begin{array}{c} R_{40} \\ \hline E_{22} \\ E_{20} \\ R_{20} \\ E_{02} \\ E_{00} \\ R_{00} \\ C_{20} \\ C_{00} \end{array} \left[\begin{array}{cccccccccc} & & & & & & \frac{1}{2} & 1 & 1 \\ \hline & & & 0 & 0 & 1 & 0 & 1 & -2 \\ & & & 0 & 1 & 8 & 6 & -1 & -8 \\ & & & \frac{1}{2} & 1 & 1 & & & \\ 0 & 0 & 6 & 0 & 1 & 1 & 0 & 8 & -8 \\ 0 & 1 & 8 & 1 & 0 & 0 & 8 & 0 & 0 \\ \frac{1}{2} & 1 & 1 & & & & & & \\ 0 & \frac{1}{2} & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{array}\right] \left[\begin{array}{c} a_{00} \\ a_{02} \\ a_{04} \\ a_{20} \\ a_{22} \\ a_{24} \\ a_{40} \\ a_{42} \\ a_{44} \end{array}\right] = \left[\begin{array}{c} J_{40} \\ \hline F_{22} \\ F_{20} \\ J_{20} \\ F_{02} \\ F_{00} \\ J_{00} \\ K_{20} \\ K_{00} \end{array}\right]$$

Table 11.6: $m = n = 4$, partial system odd/odd in x/y

$$\begin{array}{c} R_{31} \\ E_{11} \\ R_{11} \\ \hline C_{31} \\ C_{11} \end{array} \left[\begin{array}{cccc} & & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & & \\ \hline 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{array}\right] \left[\begin{array}{c} a_{11} \\ a_{13} \\ a_{31} \\ a_{33} \end{array}\right] = \left[\begin{array}{c} J_{31} \\ F_{11} \\ J_{11} \\ \hline K_{31} \\ K_{11} \end{array}\right]$$

© 2003 by CRC Press LLC

Table 11.7: Full collocation matrix—Poisson problem: $m = n = 3$ (blank spaces contain zero entries)

$$
\begin{array}{c|cccccccccccccccc|c|c}
C_{31} & & & & & 0 & 0 & 0 & 1 & & & & & 0 & 0 & 0 & 1 & a_{00} & K_{31} \\
\hline
R_{21} & & & & & & & & & 0 & 1 & 0 & 1 & & & & & a_{01} & J_{21} \\
R_{20} & & & & & & & & & \tfrac{1}{2} & 0 & 1 & 0 & & & & & a_{02} & J_{20} \\
E_{11} & & & & & 0 & 0 & 0 & 1 & & & & & 0 & 1 & 0 & -2 & a_{03} & F_{11} \\
E_{10} & & & & & 0 & 0 & 1 & 0 & & & & & 3 & 0 & -1 & 0 & a_{10} & F_{10} \\
R_{11} & & & & & 0 & 1 & 0 & 1 & & & & & & & & & a_{11} & J_{11} \\
R_{10} & & & & & \tfrac{1}{2} & 0 & 1 & 0 & & & & & & & & & a_{12} & J_{10} \\
E_{01} & 0 & 0 & 0 & 3 & & & & & 0 & 1 & 0 & -1 & & & & & a_{13} & F_{01} \\
E_{00} & 0 & 0 & 1 & 0 & & & & & 1 & 0 & 0 & 0 & & & & & a_{20} & F_{00} \\
R_{01} & 0 & 1 & 0 & 1 & & & & & & & & & & & & & a_{21} & J_{01} \\
R_{00} & \tfrac{1}{2} & 0 & 1 & 0 & & & & & & & & & & & & & a_{22} & J_{00} \\
C_{20} & 0 & 0 & \tfrac{1}{2} & 0 & & & & & 0 & 0 & 1 & 0 & & & & & a_{23} & K_{20} \\
C_{21} & & & & & 0 & 0 & 1 & 0 & & & & & 0 & 0 & 1 & 0 & a_{30} & K_{21} \\
C_{11} & & & & & 0 & 1 & 0 & 0 & & & & & 0 & 1 & 0 & 0 & a_{31} & K_{11} \\
C_{10} & 0 & \tfrac{1}{2} & 0 & 0 & & & & & 0 & 1 & 0 & 0 & & & & & a_{32} & K_{10} \\
\hline
C_{00} & \tfrac{1}{2} & 0 & 0 & 0 & & & & & 1 & 0 & 0 & 0 & & & & & a_{33} & K_{00} \\
\hline
C_{01} & & & & & 1 & 0 & 0 & 0 & & & & & 1 & 0 & 0 & 0 & & K_{01}
\end{array}
$$

solved to give explicit formulae in the case $g \equiv 0$, as we now show. We leave it as an exercise to the reader to generate a set of tables for the case $m = n = 5$.

We may readily determine formulae for all coefficients a_{ij} for $m = n = 3$ by eliminating variables in Tables 11.8 to 11.11, and we leave this as an exercise to the reader (Problem 7).

We deduce from Table 11.8 that, for $g \equiv 0$, and hence $J_{**} = K_{**} = 0$, the even/even coefficients are

$$-a_{00} = 2a_{02} = 2a_{20} = -4a_{22} = F_{00}. \tag{11.119}$$

From Table 11.9, for $g \equiv 0$, the odd/odd coefficients are

$$-a_{11} = a_{13} = a_{31} = -a_{33} = \tfrac{1}{4}F_{11}. \tag{11.120}$$

From Table 11.10, for $g \equiv 0$, the even/odd coefficients are

$$-a_{01} = a_{03} = 2a_{21} = -2a_{23} = \tfrac{1}{4}F_{01}. \tag{11.121}$$

From Table 11.11, for $g \equiv 0$, the odd/even coefficients are

$$-a_{10} = 2a_{12} = a_{30} = -2a_{32} = \tfrac{1}{4}F_{10}. \tag{11.122}$$

© 2003 by CRC Press LLC

Table 11.8: $m = n = 3$, partial system even/even in x/y

$$\begin{array}{c} R_{20} \\ E_{00} \\ R_{00} \\ C_{20} \\ \hline C_{00} \end{array} \left[\begin{array}{cccc} & & \frac{1}{2} & 1 \\ 0 & 1 & 1 & 0 \\ \frac{1}{2} & 1 & & \\ 0 & \frac{1}{2} & 0 & 1 \\ \hline \frac{1}{2} & 0 & 1 & 0 \end{array}\right] \left[\begin{array}{c} a_{00} \\ a_{02} \\ a_{20} \\ a_{22} \end{array}\right] = \left[\begin{array}{c} J_{20} \\ F_{00} \\ J_{00} \\ K_{20} \\ \hline K_{00} \end{array}\right]$$

Table 11.9: $m = n = 3$, partial system odd/odd in x/y

$$\begin{array}{c} C_{31} \\ \hline E_{11} \\ R_{11} \\ C_{11} \end{array} \left[\begin{array}{cccc} 0 & 1 & 0 & 1 \\ \hline 0 & 1 & 1 & -2 \\ 1 & 1 & & \\ 1 & 0 & 1 & 0 \end{array}\right] \left[\begin{array}{c} a_{11} \\ a_{13} \\ a_{31} \\ a_{33} \end{array}\right] = \left[\begin{array}{c} K_{31} \\ \hline F_{11} \\ J_{11} \\ K_{11} \end{array}\right]$$

Table 11.10: $m = n = 3$, partial system even/odd in x/y

$$\begin{array}{c} R_{21} \\ E_{01} \\ R_{01} \\ C_{10} \end{array} \left[\begin{array}{cccc} & & 1 & 1 \\ 0 & 3 & 1 & -1 \\ 1 & 1 & & \\ \frac{1}{2} & 0 & 1 & 0 \end{array}\right] \left[\begin{array}{c} a_{01} \\ a_{03} \\ a_{21} \\ a_{23} \end{array}\right] = \left[\begin{array}{c} J_{21} \\ F_{01} \\ J_{01} \\ K_{10} \end{array}\right]$$

Table 11.11: $m = n = 3$, partial system odd/even in x/y

$$\begin{array}{c} E_{10} \\ R_{10} \\ C_{21} \\ C_{01} \end{array} \left[\begin{array}{cccc} 0 & 1 & 3 & -1 \\ \frac{1}{2} & 1 & & \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{array}\right] \left[\begin{array}{c} a_{10} \\ a_{12} \\ a_{30} \\ a_{32} \end{array}\right] = \left[\begin{array}{c} F_{10} \\ J_{10} \\ K_{21} \\ K_{01} \end{array}\right]$$

© 2003 by CRC Press LLC

Thus for zero boundary conditions, the approximate solution u is given very simply for $m = n = 3$. Indeed we see, not surprisingly, (Problem 8) that u_{mn} can be written exactly in the form

$$u_{mn} = (x^2 - 1)(y^2 - 1)(a + bx + cy + dxy). \tag{11.123}$$

If J_{**} and K_{**} are not both zero, then no simplification such as (11.123) occurs, but we still obtain four separate sparse subsystems to solve for the coefficients a_{ij} for all m, n.

An alternative but closely related approach to the special case of the Poisson problem is given by Haidvogel & Zang (1979).

11.9 Computational fluid dynamics

One of the most important PDE problems in computational fluid dynamics is the *Navier–Stokes equation*

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = -\nabla p + \nu \Delta \mathbf{v}, \quad \nabla \cdot \mathbf{v} = 0 \tag{11.124}$$

where $\mathbf{v}$ is the velocity vector, p is the pressure divided by the (constant) density, ν is the kinematic viscosity and Δ denotes the Laplacian. This problem is studied in detail in the lecture notes by Deville (1984), as well as in Canuto et al. (1988). Deville considers, as preparatory problems, the Helmholtz equation, the Burgers equation and the Stokes problem. We shall here briefly discuss the Burgers equation.

The Burgers equation is the nonlinear equation

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = \nu \frac{\partial^2 u}{\partial x^2}, \tag{11.125}$$

which we shall take to have the boundary and initial conditions

$$u(-1, t) = u(1, t) = 0, \quad u(x, 0) = u_0(x). \tag{11.126}$$

The general procedure for solution is to discretise (11.125) into a first-order system of ordinary differential equations in t, which is solved by a scheme that is explicit as regards the nonlinear part and implicit for the linear part. Using Chebyshev collocation at the $n + 1$ points $\{y_j\}$, the discretisation can be written (Canuto et al. 1988) as

$$\mathbf{Z}_n \left(\frac{\partial \mathbf{u}_n}{\partial t} + \mathbf{U}_n \mathbf{D}_n \mathbf{u}_n - \nu \mathbf{D}^2 \mathbf{u}_n \right) = \mathbf{0}, \tag{11.127}$$

where $\mathbf{D}_n$ is the appropriate Chebyshev collocation differentiation matrix, $\mathbf{u}_n$ is a vector with elements $u_n(y_j, t)$, $\mathbf{U}_n$ is a diagonal matrix with the

© 2003 by CRC Press LLC

elements of $\mathbf{u}_n$ on its diagonal and $\mathbf{Z}_n$ is a unit matrix with its first and last elements replaced by zeros. The boundary conditions are imposed by requiring $u_n(y_0, t) = u_n(y_n, t) = 0$. The method as it stands involves $O(n^2)$ operations at each time step for the implicit term, but this can be reduced to $O(n \log n)$ operations by using FFT techniques.

A Chebyshev tau method may instead be applied, defining

$$\langle f\,,\, T_k \rangle = \frac{2}{\pi c_k} \int_{-1}^{1} \frac{f(x) T_k(x)}{\sqrt{1-x^2}}\, \mathrm{d}x. \tag{11.128}$$

Then, defining

$$u_n(x,t) = {\sum_{k=0}^{n}}' a_k(t) T_k(x),$$

we have

$$\left\langle \frac{\partial u_n}{\partial t} + u_n \frac{\partial u_n}{\partial x} - \nu \frac{\partial^2 u_n}{\partial x^2}\,,\, T_k \right\rangle = 0, \tag{11.129}$$

which reduces to

$$\frac{\mathrm{d}a_k}{\mathrm{d}t} + \left\langle u_n \frac{\partial u_n}{\partial x}\,,\, T_k \right\rangle - \nu a_k^{(2)} = 0. \tag{11.130}$$

Again a mixed explicit/implicit method may be adopted for each time step, the inner product being evaluated explicitly.

For discussion of the Stokes and Navier–Stokes equations, the reader is referred to Deville (1984), Canuto et al. (1988), Fornberg (1996), and Gerritsma & Phillips (1998, 1999).

11.10 Particular issues in spectral methods

It is important to remember that the key advantages of spectral and pseudospectral methods lie in

1. the rapid (e.g., exponential) convergence of the methods for very smooth data and PDEs, which makes Chebyshev methods so powerful;
2. the use of discrete orthogonality, which greatly simplifies collocation equations;
3. the use of the FFT, which speeds up computations typically from $O(n^2)$ to $O(n \log n)$ operations;
4. the possibility of a matrix representation of derivatives, which simplifies the representation of the solution and boundary conditions in certain problems.

© 2003 by CRC Press LLC

For the reasons above, the method will always be restricted to somewhat special classes of problems if it is to compete with more general methods like the finite element method. However, the spectral method shares with the finite element method a number of common features, including the pointwise and continuous representation of its solution (as in the differentiation matrix method) and the possibility of determining good preconditioners (Fornberg & Sloan 1994).

We now raise some further important issues that arise in spectral/pseudospectral methods. We do not have the scope to illustrate these issues in detail but can at least make the reader aware of their significance.

Aliasing (see Section 6.3.1) is an interesting feature of trigonometric and Chebyshev polynomials on discrete meshes. There is a potential for ambiguity of definition when a Chebyshev or Fourier series attempts to match a PDE on too coarse a grid. Fortunately, aliasing is not generally to be regarded as threatening, especially not in linear problems, but nonlinear problems do give cause for some concern on account of the possible occurrence of high-frequency modes which may be misinterpreted as low-frequency ones. Canuto et al. (1988, p.85) note that aliases may be removed by phase shifts, which can eliminate special relationships between low and high frequency modes.

Preconditioners are frequently used in finite-element methods to improve the conditioning of linear equations. Their use with finite differences for Chebyshev methods is discussed for example by Fornberg (1996), Fornberg & Sloan (1994) and Phillips et al. (1986). The idea is, for example, to take a system of linear equations whose matrix is neither diagonally dominant nor symmetric, and to find a multiplying matrix that yields a result that is strictly diagonally dominant, and therefore amenable to Gauss–Seidel iteration. More broadly, it improves the conditioning of the system matrix.

Basis functions in spectral methods may be not only Chebyshev polynomials, but also Legendre polynomials or trigonometric polynomials (Canuto et al. 1988). Legendre polynomials are sometimes preferred for Galerkin methods and Chebyshev polynomials for collocation methods (because of discrete orthogonality). Trigonometric polynomials are versatile but normally suitable for periodic functions only, because of the Gibbs phenomenon (see page 118, footnote). Clearly we are primarily interested in Chebyshev polynomials here, and so shall leave discussion of Legendre polynomials and other possible bases to others.

11.11 More advanced problems

The subject of partial differential equations is a huge one, and we cannot in this broadly-based book do full credit to spectral and pseudospectral methods. We have chosen to illustrate some key aspects of the methods, mainly for linear

© 2003 by CRC Press LLC

and quasilinear problems, and to emphasise some of the technical ideas that need to be exploited.

For discussion of other problems and, in particular, more advanced PDE problems including nonlinear problems, the reader is referred to such books as:

- Canuto et al. (1988) — for many fluid problems of varying complexity and solution structures, as well as an abundance of background theory;
- Trefethen (2000) — for a very useful collection of software and an easy-to-read discussion of the spectral collocation approach;
- Boyd (2000) — for a modern treatment including many valuable results;
- Guo Ben-yu (1998) — for an up-to-date and very rigorous treatment;
- Fornberg (1996) — as it says, for a practical guide to pseudospectral methods;
- Deville (1984) — for a straightforward introduction mainly to fluid problems;
- Gottlieb & Orszag (1977) — for an early and expository introduction to the spectral approach.

11.12 Problems for Chapter 11

1. Apply the method of separation of variables in (r, θ) coordinates to

$$\Delta u = r^2 \frac{\partial^2 u}{\partial r^2} + r \frac{\partial u}{\partial r} + \frac{\partial^2 u}{\partial \theta^2} = 0$$

(see (11.11a) above) in the disc $S : r \leq 1$, where $u(1, \theta) = g(\theta)$ on $\partial S : r = 1$, and $g(\theta)$ is a known 2π-periodic function of the orientation of a point P of the boundary. Determine the solution as a series in the cases in which

(a) $g(\theta) = \pi^2 - \theta^2$;

(b) $g(\theta) = \begin{cases} -1, & -\pi \leq \theta \leq 0 \\ +1, & 0 \leq \theta \leq \pi. \end{cases}$

2. In addition to satisfying $(m-1)(n-1)$ specified linear conditions in the interior of the square domain $D : \{|x| < 1, |y| < 1\}$, the form

$$\sum_{i=0}^{m+1}{}' \sum_{j=0}^{n+1}{}' a_{ij} T_i(x) T_j(y)$$

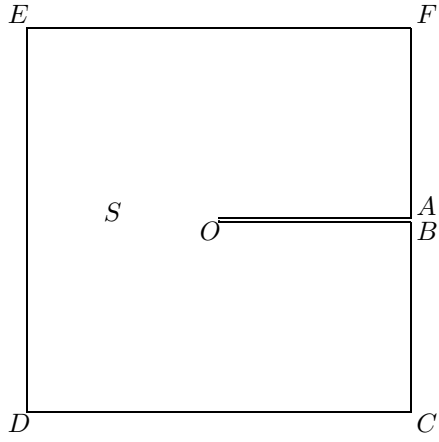
© 2003 by CRC Press LLC

is collocated to a function $g(x, y)$ at $2(m + n)$ points on its boundary ∂D. The latter points are chosen at the zeros of $(1 - x^2)U_{m-2}(x)$ on $y = \pm 1$ and at the zeros of $(1-y^2)U_{n-2}(y)$ on $x = \pm 1$, where each of the four corners of the boundary (which occur in both sets of zeros) is only counted once. Investigate whether or not the resulting linear system is singular and determine its maximum rank.

(This question is an analogue of a result in Section 11.8.5, where the zeros of $T_m(x)$ and $T_n(y)$ were adopted.)

3. The diagram shows a square membrane with a slit from the midpoint A of one side to the centre O. We wish to determine solutions of the eigenvalue problem

$$\Delta u + \lambda u = 0 \text{ in } S,$$
$$u = 0 \text{ on } \partial S.$$



Follow the style of Section 11.8.2 to transform the domain and problem into one which may be approximated by Chebyshev collocation. Use the mapping $z' = z^{\frac{1}{2}}$ about O to straighten the cut AOB, determine equations for the mapped (curved) sides of the domain, determine the mapped PDE and deduce the form of approximation u_{mn} to u. Describe the method of solution for λ and u.

[Note: The boundary equation is again $y'(x^2y^2 - r^2 + 1) = 0$ before mapping.]

4. Investigate whether or not there is any gain of efficiency or accuracy in practice in using the Chebyshev form $\sum\sum c_{ij}T_{2i+t}(x/a)T_j(y/b)$ rather than $\sum\sum c_{ij}x^{2i+t}y^j$ in the L-membrane approximation of Section 11.8.2 and, similarly, for the relevant forms in the method for Problem 3 above. Is it possible, for example, to exploit discrete orthogonality in the collocation equations?

© 2003 by CRC Press LLC

5. As a variant on the separation of variables method, consider the solution of

$$\Delta u = f(x,y) \text{ in the ellipse } D : \frac{x^2}{a^2} + \frac{y^2}{b^2} \le 1, \qquad (A)$$

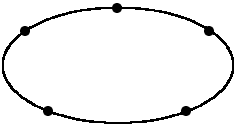
$$u = g(x,y) \text{ on } \partial D : \phi(x,y) \equiv \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \qquad (B)$$

where $f \equiv 0$ and g is given explicitly on ∂D.

Then the form

$$u_n = a_0 + \sum_{k=1}^{n} (a_k \cos k\theta + b_k \sin k\theta) r^k,$$

where $x = r\cos\theta$ and $y = r\sin\theta$, satisfies (A) for all coefficients a_0, a_k, b_k. Compute $a_0, \ldots, a_n$, $b_1, \ldots, b_n$ so that (B) is collocated at $2n+1$ suitably chosen points of ∂D. It is suggested that equal angles of θ should be used on $[0, 2\pi]$; discuss some of the possible choices. What set of points would remain distinct as $b \to 0$, if the ellipse has a small eccentricity?



[Hint: Start at $\theta = \frac{1}{2}\pi/(2n+1)$; the nodes for $n = 2$ are then chosen as in the figure and occur at $\pi/10$, $5\pi/10$, $9\pi/10$, $13\pi/10$, $17\pi/10$. Choose simple non-polynomial functions for g; e.g., $g(x,y) = \cosh(x+y)$.]

6. Repeat Problem 5, but with $g \equiv 0$ and f given explicitly on D, using the Chebyshev polynomial approximation

$$u_{mn} = \phi(x,y). \sum_{i=0}^{m}{}' \sum_{j=0}^{n}{}' a_{ij} T_i(x) T_j(y)$$

and collocating the PDE at a tensor product of the zeros of $T_{m+1}(x/a)$ and $T_{n+1}(y/b)$ on the rectangle

$$R : \{-a \le x \le a; -b \le y \le b\}.$$

Compute results for small choices of m, n.

[Note: This is a method which might be extended to more general boundary $\phi(x,y)$, and ϕ does not need to be a polynomial in x, y.]

7. Generate a set of tables similar to Tables 11.7–11.11 for the odd/odd case $m = n = 5$, showing the 36×36 linear algebraic system for $\{a_{ij}\}$ and the four subsystems derived from this.

© 2003 by CRC Press LLC

8. For $m = n = 3$ (see Section 11.8 above) show that the approximate solution u_{mn} of (11.75) with $g \equiv 0$, given by (11.91) with coefficients (11.119)–(11.122), may be simplified exactly into the form

$$u_{mn} = (1 - x^2)(1 - y^2)[a + bx + cy + dxy].$$

What are the values of a, b, c, d?

Derive u_{mn} directly in this form by collocating the PDE at the Chebyshev zeros. (Note that this method cannot be applied unless $g(x, y) \equiv 0$.)

9. For $m = n = 3$ in (11.75), in the case where g is *not* identically zero, obtain formulae for the coefficients a_{ij} in u_{mn} from Tables 11.8–11.11, namely the four linear subsystems that define them.

© 2003 by CRC Press LLC