HYBRID CHEBYSHEV COLLOCATION-SERIES METHODS FOR ELLIPTIC PROBLEMS

by

J.C. Mason
Royal Military College of Science, Shrivenham, England
and Centre for Mathematical Analysis, A.N.U., Canberra
and
I.D. Coope
Department of Mathematics, University of Canterbury, Christchurch, New Zealand

No.60 December, 1990

HYBRID CHEBYSHEV COLLOCATION-SERIES METHODS FOR ELLIPTIC PROBLEMS

J.C. Mason, I.D. Coope

1. INTRODUCTION

Bivariate Chebyshev polynomial approximations of the form

$$u^* = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} \, T_{i-1}(x) \, T_{j-1}(y) \tag{1.1}$$

have been applied successfully for many years to the numerical solution of a wide range of partial differential equations in two variables, and an excellent discussion, mainly of more recent work, is given in the book of Canuto, Hussaini, Quarteroni and Zang [1]. These "spectral methods" can be very accurate, because of the rapid convergence of Chebyshev series expansions to functions which have a number of continuous derivatives. The methods can also be very efficient, provided that they are designed to exploit special structures in the given problem or in the algebraic problem which is derived from it. Assuming that m is comparable with n or proportional to it, such spectral methods can require only $O(n^2 \log n)$ or $O(n^3)$ arithmetic operations for problems as simple as the Poisson equation, using a series method such as that of Haidvogel and Zang [2]. However, at the other extreme, in the case of a general linear partial differential equation with variable coefficients, it may be necessary to apply a bivariate collocation method, such as that of an early paper of Mason [3], which involves $O(n^6)$ operations if no special structure is exploited. Even in this extreme case, spectral methods may still provide competition for other methods, such as finite elements or finite difference methods. For although these competing methods are almost invariably able to exploit sparsity in the resulting linear algebraic system, they typically involve larger numbers of parameters than spectral methods. However, spectral methods are clearly going to be at their most efficient and competitive when the number of their arithmetic operations can be reduced to the level of $O(n^4)$ or less. The overall aim of the current paper is to extend the range of problems for which such efficiency is achievable.

In a recent paper, Mason [4] and Mason and Olaofe [5] noted that there was a class of problems, which might be crudely described as "simple in one variable, but more complicated in the other variable", where structure could be exploited by using vector/matrix techniques. They gave an algorithm which adopted a Lanczos tau method (see [6]) in the simpler variable and a collocation method in the other variable, and which required only $O(n^4)$ arithmetic operations. In the present paper we show that, by exploiting a matrix eigenvalue decomposition, the number of operations can be further reduced to $O(n^3)$. However, this tau-collocation method has the disadvantage that the approximation u^* to the solution u is expressed in the power form

$$u^* = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \, x^{i-1} \, y^{j-1} \, . \tag{1.2}$$

It is well known that the coefficients a_{ij} in this form can grow rapidly and oscillate in sign with i and j, and that this can lead to accumulator overflow and/or large significance errors for larger values of m and n. This has been observed to occur in practice, and so the method can only at present be used with confidence for modest values of m and n. Nevertheless, useful solutions (of, for example, up to four figure accuracy) may be obtained very rapidly for a variety of problems by the method.

The main aim of the present paper is to derive a numerically stable method in the spirit of the method of Mason and Olaofe [5]. We achieve this by adopting instead of (1.2) the Chebyshev form (1.1), substituting this in the differential equation and equating Chebyshev coefficients in the simpler variable (x), while simultaneous collocating at Chebyshev zeros in y. Vector/matrix techniques are again exploited to achieve a sparse linear algebraic system for c_{ij}, which may be solved by a numerically stable method in $O(n^4)$ operations. Very accurate numerical solutions are achieved in this way for a variety of model problems, with values of m and n up to 16 or more. The new method is thus both efficient and numerically sound.

2. CHEBYSHEV POLYNOMIALS AND FORMULAE

The Chebyshev polynomial of the first kind of degree n is defined trigonometrically by the formula

(2.1) $$T_n(x) = \cos n\theta \quad \text{where} \quad x = \cos\theta$$

for x in $[-1, 1]$ corresponding to θ in $[0, \pi]$. In the methods of this paper we need formulae for the following polynomials, whose degrees are indicated by their subscripts:

(2.2) $$\psi_{n+2}(x) = (1 - x^2)T_n(x)$$

(2.3) $$\theta_{n+1}(x) = \psi'_{n+2}(x)$$

(2.4) $$\phi_n(x) = \psi''_{n+2}(x) .$$

From (2.1) and (2.2), we see that

$$\begin{aligned}\psi_{n+2}(\cos\theta) &= \sin^2\theta\ \cos n\theta = \frac{1}{2}(1 - \cos 2\theta)\cos n\theta \\ &= \frac{1}{2}\cos n\theta - \frac{1}{4}[\cos(n-2)\theta + \cos(n+2)\theta]\end{aligned}$$

and hence that, for all $n \geq 0$,

(2.5) $$\psi_{n+2}(x) = -\frac{1}{4}T_{n+2}(x) + \frac{1}{2}T_n(x) - \frac{1}{4}T_{|n-2|}(x) .$$

From (2.1) and (2.3), it follows that

$$\begin{aligned}\theta_{n+1}(x) &= (-\sin\theta)^{-1}\frac{d}{d\theta}(\sin^2\theta\ \cos n\theta) \\ &= -2\cos\theta\ \cos n\theta + n\ \sin\theta\ \sin n\theta\end{aligned}$$

and hence that, for all $n \geq 0$,

$$\theta_{n+1}(x) = -\frac{1}{2}(n+2)\, T_{n+1}(x) + \frac{1}{2}(n-2)\, T_{|n-1|}(x)\,. \tag{2.6}$$

Finally, we leave it as an exercise to the reader to confirm that

$$\phi_n(x) = -(n+2)(n+1)\, T_n(x) - 6n[T_{n-2}(x) + T_{n-4}(x) + \cdots] \tag{2.7}$$

where the last term of the series in brackets is

$$\begin{array}{ll} T_1(x) & \text{if } n \text{ is odd,} \\ \frac{1}{2}T_0(x) & \text{if } n \text{ is even.} \end{array}$$

3. MODEL PROBLEMS

We consider in this paper the problem

$$L(u) \equiv u_{xx} + \gamma(y)u_{yy} + \delta(y)u_y + f(x,y) = 0 \tag{3.1}$$

subject to

$$u = 0 \quad \text{on} \quad x = \pm 1\,,\ y = \pm 1\,. \tag{3.2}$$

Apart from considering the equation (3.1) in its general form, we also consider a number of important examples that correspond to special choices of γ, δ and f.

In particular we consider the case

$$\delta(y) = \gamma'(y)\,, \tag{3.3}$$

for which (3.1) takes the form

$$u_{xx} + \frac{\partial}{\partial y}\left(\gamma(y)\frac{\partial u}{\partial y}\right) + f(x,y) = 0 \tag{3.4}$$

and corresponds to the equation for steady state flow in a confined aquifer.

Other classical problems are obvious special cases of (3.1), such as the Poisson equation ($\gamma = 1, \delta = 0$), which gives the torsion equation when f is a positive constant. A useful intermediate problem corresponds to $\delta = 0$, in the case in which $\gamma(y)$ is still dependent on y.

4. THE COLLOCATION – TAU METHOD

Before describing our new collocation-series Chebyshev method (in §5 below), it is appropriate to recall the method of Mason and Olaofe [5], to show how it may be applied to the general problem (3.1), and to describe how it may in principle be speeded up from $O(n^4)$ to $O(n^3)$ arithmetic operations. This sets the scene for the more stable and accurate method of §5.

Consider the general problem (3.1), (3.2) and approximate u in the power form

$$u^* = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}(1-x^2)x^{2i-2}(1-y^2)y^{2j-2}\,, \tag{4.1}$$

where we assume for convenience that the functions γ and f are even in x and y and that δ is odd in y. Note that (4.1) automatically satisfies (3.2) exactly. Suppose also that $f(x,y)$ is approximated in the form

$$f^* = \sum_{i=1}^{m} f_i(y)x^{2i-2} \tag{4.2}$$

where f_i is a polynomial of degree $2n-2$ in y, defined by requiring that f^* should interpolate f in the tensor product of mn points $\{(x_k, y_\ell)\}$ $(k=1,\ldots,m;\ \ell=1,\ldots,n)$, where x_k and y_ℓ are, respectively, positive zeros of $T_{2m}(x)$ and $T_{2n}(y)$. Using the Chebyshev zeros $\{y_\ell\}$ again, as collocation points for (3.1), we now define a collocation-tau method by replacing (3.1) by the following perturbed equation for u^* (compare Lanczos' tau method [6]):

$$\begin{aligned} L(u^*) = u^*_{xx} + \gamma(y)u^*_{yy} + \delta(y)u^*_y + f^*(x,y) = \tau(y)\, T_{2m}(x) \\ \text{at} \quad y = y_\ell, \quad (\ell = 1,\ldots,m)\,. \end{aligned} \tag{4.3}$$

The substitution of (4.1) into (4.3) gives

$$\begin{aligned} \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\left[p_j(y_\ell)x^{2i} + q_{ij}(y_\ell)x^{2i-2} + r_{ij}(y_\ell)x^{2i-4}\right] + \sum_{i=1}^{m} f_i(y_\ell)x^{2i-2} \\ = \tau(y_\ell)\sum_{i=1}^{m+1} t_i x^{2i-2} \end{aligned} \tag{4.4}$$

where t_i is the coefficient of x^{2i-2} in $T_{2m}(x)$, and where

$$p_j(y_\ell) = \alpha_{\ell j} \tag{4.5}$$

$$q_{ij}(y_\ell) = -2i(2i-1)\beta_{\ell j} - \alpha_{\ell j} \tag{4.6}$$

$$r_{ij}(y_\ell) = (2i-2)(2i-3)\beta_{\ell j} \tag{4.7}$$

$$\begin{aligned} \alpha_{\ell j} = \gamma(y_\ell)\left[-2j(2j-1)y_\ell^{2j-2} + (2j-2)(2j-3)y_\ell^{2j-4}\right] \\ +\delta(y_\ell)\left[-2j\, y_\ell^{2j-1} + (2j-2)y_\ell^{2j-3}\right] \end{aligned} \tag{4.8}$$

$$\beta_{\ell j} = y_\ell^{2j-2}\,. \tag{4.9}$$

To simplify the algebra, define $n \times n$ matrices $\mathbf{A}, \mathbf{B}, \mathbf{P}, \mathbf{Q}_i$, and $\mathbf{R}_i$ and $n \times 1$ vectors $\mathbf{f}_i, \boldsymbol{\tau}$, $\mathbf{c}_i$ as follows:

$$(\mathbf{A})_{\ell j} = \alpha_{\ell j}, \quad (\mathbf{B})_{\ell j} = \beta_{\ell j} \tag{4.10}$$

(4.11) $$(\mathbf{P})_{\ell j} = p_j(y_\ell),\ \ (\mathbf{Q}_i)_{\ell j} = q_{ij}(y_\ell),\ \ (\mathbf{R}_i)_{\ell j} = r_{ij}(y_\ell)$$

(4.12) $$(\mathbf{f}_i)_\ell = f_i(y_\ell),\ \ (\boldsymbol{\tau})_\ell = \tau(y_\ell),\ \ (\mathbf{c}_i)_j = c_{ij}\,.$$

Then, on equating coefficients of x^{2i} $(i = 0, \ldots, m)$ in (4.4) and using (4.11), (4.12), we obtain the linear algebraic system

(4.13) $$\mathbf{P}\ \mathbf{c}_i + \mathbf{Q}_{i+1}\ \mathbf{c}_{i+1} + \mathbf{R}_{i+2}\ \mathbf{c}_{i+2} = t_{i+1}\ \boldsymbol{\tau} - \mathbf{f}_{i+1},\quad (i{=}0,\ldots,m),$$

provided we set undefined vectors and matrices to zero as follows:

(4.14) $$\begin{cases} \mathbf{c}_0 = \mathbf{c}_{m+1} = \mathbf{c}_{m+2} = \mathbf{0}\,,\ \ \mathbf{f}_{m+1} = \mathbf{0}\,, \\ \mathbf{Q}_0 = \mathbf{R}_0 = \mathbf{R}_1 = \mathbf{Q}_{m+1} = \mathbf{R}_{m+1} = \mathbf{R}_{m+2} = \mathbf{0} \end{cases}$$

On substituting in (4.13) for $\mathbf{c}_i$ in the form

(4.15) $$\mathbf{c}_i = \mathbf{V}_i \boldsymbol{\tau} + \mathbf{w}_i\,,$$

where $\mathbf{V}_i$ and $\mathbf{w}_i$ are an undetermined matrix and vector, we deduce recurrence relations for the latter:

(4.16) $$\mathbf{V}_i = \mathbf{P}^{-1}(t_{i+1}\,\mathbf{I} - \mathbf{Q}_{i+1}\mathbf{V}_{i+1} - \mathbf{R}_{i+2}\mathbf{V}_{i+2})$$

(4.17) $$\mathbf{w}_i = \mathbf{P}^{-1}(\mathbf{f}_{i+1} - \mathbf{Q}_{i+1}\mathbf{w}_{i+1} - \mathbf{R}_{i+2}\mathbf{w}_{i+2})$$

where

(4.18) $$\mathbf{V}_m = t_{m+1}\mathbf{P}^{-1},\ \ \mathbf{w}_m = \mathbf{0}\,.$$

A formula for $\boldsymbol{\tau}$ follows from the first equation of (4.13):

(4.19) $$\boldsymbol{\tau} = -(t_1\mathbf{I} - \mathbf{Q}_1\mathbf{V}_1 - \mathbf{R}_2\mathbf{V}_2)^{-1}(-\mathbf{f}_1 - \mathbf{Q}_1\mathbf{w}_1 - \mathbf{R}_2\mathbf{w}_2)\,.$$

Hence $\mathbf{c}_i$ is defined explicitly from (4.15), and the approximate solution u^* is determined in the form (4.1).

The above is the algorithm of Mason and Olaofe [5], appropriate to the general problem (3.1), (3.2). The operations count is $O(mn^3)$, or, assuming m is related linearly to n, $O(n^4)$.

4.1 Limitations on This Method

The algorithm above has been tested on a number of problems and gives good results for modest values of m and n (e.g. $m{=}n{=}4$ for the test problems of §3). However, the coefficients c_{ij} in (4.1) can grow rapidly with i and j and this can lead to accumulator overflow and/or significance errors in the recurrences (4.16), (4.17) for large values of n.

The situation improves somewhat if an alternative approach is adopted to the solution of the system (4.13), namely the use of Gauss elimination with interchanges (compare §5 below) together with suitable scaling. However, this does not alter the basic fact that the use of the power form (4.1) is sometimes unsatisfactory. We therefore suggest that the use of the algorithm above should be reduced to the determination of solutions of modest accuracy, and in that context it is efficient.

4.2 An $O(n^3)$ Version of the Method

The operations count in this method can be reduced to $O(n^3)$ by exploiting the (unused) fact that the matrices $\mathbf{P}, \mathbf{Q}_i$ and $\mathbf{R}_i$ are all linear combinations of just 2 matrices $\mathbf{A}$ and $\mathbf{B}$ given by (4.10), namely

(4.20) $$\mathbf{P} = \mathbf{A}, \quad \mathbf{Q}_i = -2i(2i-1)\mathbf{B} - \mathbf{A}, \quad \mathbf{R}_i = (2i-2)(2i-3)\mathbf{B}\,.$$

Defining

(4.21) $$\mathbf{T} = \mathbf{A}^{-1}\,, \quad \mathbf{S} = \mathbf{TB}\,, \quad \mathbf{g}_i = \mathbf{T}\,\mathbf{f}_i$$

and using (4.20), the recurrences (4.16), (4.17) become

(4.22) $$(\mathbf{V}_i - \mathbf{V}_{i+1}) = t_{i+1}\,\mathbf{T} + (2i+2)(2i+1)\mathbf{S}(\mathbf{V}_{i+1} - \mathbf{V}_{i+2})$$

(4.23) $$(\mathbf{w}_i - \mathbf{w}_{i+1}) = \mathbf{g}_{i+1} + (2i+2)(2i+1)\mathbf{S}(\mathbf{w}_{i+1} - \mathbf{w}_{i+2})\,.$$

The recurrence (4.22) for $\{\mathbf{V}_i - \mathbf{V}_{i+1}\}$ represents the evaluation of a polynomial in the matrix $\mathbf{S}$ by nested multiplication, and this can in principle be done much more efficiently by determining, once and for all, the eigenvalue decomposition of $\mathbf{S}$:

(4.24) $$\mathbf{S} = \mathbf{K}\,\mathbf{D}\,\mathbf{K}^{-1}$$

where $\mathbf{D}$ is a diagonal matrix of eigenvalues, and $\mathbf{K}$ is a matrix of eigenvectors. Assuming this decomposition is regular, then (4.22) simplifies to

(4.25) $$\mathbf{G}_i = t_{i+1}\,\mathbf{K}^{-1}\,\mathbf{T} + (2i+2)(2i+1)\mathbf{D}\,\mathbf{G}_{i+1}$$

where

(4.26) $$\mathbf{G}_i = \mathbf{K}^{-1}(\mathbf{V}_i - \mathbf{V}_{i+1}).$$

Clearly the computation (4.25) is now of $O(n^3)$ complexity.

5. A NEW COLLOCATION-SERIES METHOD

Consider again the elliptic problem (3.1), (3.2), namely

(5.1) $$L(u) = u_{xx} + \gamma(y)u_{yy} + \delta(y)u_y + f(x,y) = 0\,,$$

subject to

(5.2) $$u = 0 \quad \text{on} \quad x = \pm 1,\ y = \pm 1\,.$$

Again assume for convenience that γ and f are even in x, y and δ is odd in y, but now approximate u in the more appropriate Chebyshev form

(5.3) $$u^* = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\,(1-x^2)T_{2i-2}(x)\,(1-y^2)T_{2j-2}(y)\,.$$

Using the definitions (2.2), (2.3), (2.4) of §2, we have

$$\psi_{2i}(x) = (1-x^2)T_{2i-2}(x)\,, \quad \theta_{2i-1}(x) = \psi'_{2i}(x)\,, \quad \phi_{2i-2}(x) = \psi''_{2i}(x)\,, \tag{5.4}$$

and the formulae (2.5), (2.6), (2.7) for these polynomials in terms of Chebyshev polynomials become, for $i = 1,\ldots,m$:

$$\psi_{2i}(x) = -\frac{1}{4}\,T_{2i}(x) + \frac{1}{2}\,T_{2i-2}(x) - \frac{1}{4}\,T_{|2i-4|}(x) \tag{5.5}$$

$$\theta_{2i-1}(x) = -i\,T_{2i-1}(x) + (i-2)T_{|2i-3|}(x) \tag{5.6}$$

$$\phi_{2i-2}(x) = -2i(2i-1)T_{2i-2}(x) - 6(2i-2)\sum_{k=0}^{i-2}{}' \,T_{2k}(x) \tag{5.7}$$

where the dash indicates that the first term of the sum is halved. Note that the form (5.3) already satisfies the boundary conditions (3.2) of the problem.

In our new method we substitute u^*, given by (5.3), into the equation (5.1) and equate coefficients of $\frac{1}{2}T_0(x), T_2(x), T_4(x), \ldots, T_{2m-2}(x)$, while collocating at the n positive zeros y_ℓ of $T_{2n}(y)$. (This is similar to the method of §4, but we adopt a Clenshaw-type approach in the x variable (see for example [7]) in place of the Lanczos tau method). Substituting (5.3) in (5.1), setting $y = y_\ell$, and using (5.5), (5.6), (5.7), we obtain:

$$\begin{aligned} L(u^*(x,y_\ell)) = &\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\,\psi_{2j}(y_\ell)\Big[-2i(2i-1)T_{2i-2}(x) - 6(2i-2)\sum_{k=0}^{i-2}{}'\,T_{2k}(x)\Big] \\ &+\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\,[\gamma(y_\ell)\,\phi_{2j-2}(y_\ell) + \delta(y_\ell)\,\theta_{2j-1}(y_\ell)] \\ &\qquad\times\left[-\tfrac{1}{4}T_{2i}(x) + \tfrac{1}{2}T_{2i-2}(x) - \tfrac{1}{4}T_{|2i-4|}(x)\right] \\ &+ f(x,y_\ell)\,. \end{aligned} \tag{5.8}$$

The function $f(x,y_\ell)$ is now replaced by the polynomial $f^*(x,y_\ell)$, which interpolates it in the positive zeros x_k of $T_{2m}(x)$, namely

$$f(x,y_\ell) \simeq f^*(x,y_\ell) = \sum_{i=1}^{m}{}' \,f_{i\ell}\,T_{2i-2}(x)\,, \tag{5.9}$$

where

$$f_{i\ell} = \frac{2}{m}\sum_{k=1}^{m} f(x_k,y_\ell)T_{2i-2}(x_k)\,. \tag{5.10}$$

Then (5.8) gives

$$L(u^*(x,y_\ell)) = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\Big[p_{ij}(y_\ell)T_{2i}(x) + q_{ij}(y_\ell)T_{2i-2}(x) + r_{ij}(y_\ell)T_{2i-4}(x) + s_{ij}(y_\ell)\sum_{k=0}^{i-3}{}' T_{2k}(x)\Big] + \sum_{i=1}^{m}{}' f_{i\ell}T_{2i-2}(x) \tag{5.11}$$

where

$$\begin{aligned} p_{ij}(y_\ell) &= (1+\lambda)\alpha_{\ell j} \\ q_{ij}(y_\ell) &= -2i(2i-1)\beta_{\ell j} - 2\alpha_{\ell j} \\ r_{ij}(y_\ell) &= -3\mu(2i-2)\beta_{\ell j} + (1-\lambda)\alpha_{\ell j} \\ s_{ij}(y_\ell) &= -6(2i-2)\beta_{\ell j} \end{aligned}$$

and

$$\begin{aligned} \alpha_{\ell j} &= -\tfrac{1}{4}[\gamma(y_\ell)\phi_{2j-2}(y_\ell) + \delta(y_\ell)\theta_{2j-1}(y_\ell)] \\ \beta_{\ell j} &= \psi_{2j}(y_\ell) \end{aligned}$$

with

$$\lambda = 1 \quad \text{for} \quad i=1, \quad \lambda = 0 \ \text{otherwise}$$
$$\mu = 1 \quad \text{for} \quad i=2, \quad \mu = 2 \ \text{otherwise}.$$

To simplify this, we again use matrices and vectors (as in §4), and in particular define $\mathbf{A}$ and $\mathbf{B}$ as

$$(\mathbf{A})_{\ell j} = \alpha_{\ell j}\,, \quad (\mathbf{B})_{\ell j} = \beta_{\ell j}\,. \tag{5.12}$$

Equating to zero the coefficients of $\frac{1}{2}T_0(x),\ T_2(x), \ldots T_{2m-2}(x)$ in (5.11), we may derive the following system of equations for the vectors $\mathbf{c}_i$

$$\begin{bmatrix} \mathbf{Q}_1 & \mathbf{R}_2 & \mathbf{S}_3 & \mathbf{S}_4 & \mathbf{S}_5 & \cdots & \mathbf{S}_m \\ 2\mathbf{A} & \mathbf{Q}_2 & \mathbf{R}_3 & \mathbf{S}_4 & \mathbf{S}_5 & \cdots & \mathbf{S}_m \\ & \mathbf{A} & \mathbf{Q}_3 & \mathbf{R}_4 & \mathbf{S}_5 & \cdots & \mathbf{S}_m \\ & & & & & & \vdots \\ & & & \mathbf{A} & \mathbf{Q}_{m-2} & \mathbf{R}_{m-1} & \mathbf{S}_m \\ & & & & \mathbf{A} & \mathbf{Q}_{m-1} & \mathbf{R}_m \\ & & & & & \mathbf{A} & \mathbf{Q}_m \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \\ \vdots \\ \mathbf{c}_{m-2} \\ \mathbf{c}_{m-1} \\ \mathbf{c}_m \end{bmatrix} = - \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \\ \vdots \\ \mathbf{f}_{m-2} \\ \mathbf{f}_{m-1} \\ \mathbf{f}_m \end{bmatrix} \tag{5.13}$$

where

$$\left.\begin{aligned} \mathbf{Q}_i &= \begin{cases} 2[-2i(2i-1)\mathbf{B} - 2\mathbf{A}] & \text{for } i=1 \\ [-2i(2i-1)\mathbf{B} - 2\mathbf{A}] & \text{otherwise} \end{cases} \\ \mathbf{R}_i &= \begin{cases} -6(2i-2)\mathbf{B} + 2\mathbf{A} & \text{for } i=2 \\ -6(2i-2)\mathbf{B} + \mathbf{A} & \text{otherwise} \end{cases} \\ \mathbf{S}_i &= -6(2i-2)\mathbf{B} \end{aligned}\right\} \tag{5.14}$$

and where $(\mathbf{c}_i)_j = c_{ij}$, $(\mathbf{f}_i)_\ell = f_{i\ell}$.

Now the equations (5.13) have a block Hessenberg form, (when viewed as a system of equations in c_{ij}), and they are readily simplified by subtracting each equation in (5.13) from its successor to give

$$\text{(5.15)}\quad \begin{bmatrix} \mathbf{Q}_1^* & \mathbf{R}_2^* & -\mathbf{A} & & & & & \\ 2\mathbf{A} & \mathbf{Q}_2^* & \mathbf{R}_3^* & -\mathbf{A} & & & & \\ & \mathbf{A} & \mathbf{Q}_3^* & \mathbf{R}_4^* & -\mathbf{A} & & & \\ & & & \vdots & & & & \\ & & & \mathbf{A} & \mathbf{Q}_{m-2}^* & \mathbf{R}_{m-1}^* & -\mathbf{A} \\ & & & & \mathbf{A} & \mathbf{Q}_{m-1}^* & \mathbf{R}_m^* \\ & & & & & \mathbf{A} & \mathbf{Q}_m^* \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \\ \vdots \\ \mathbf{c}_{m-2} \\ \mathbf{c}_{m-1} \\ \mathbf{c}_m \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \\ \vdots \\ \mathbf{h}_{m-2} \\ \mathbf{h}_{m-1} \\ \mathbf{h}_m \end{bmatrix}$$

where

$$\text{(5.16)}\quad \mathbf{Q}_i^* = \begin{cases} -2[2i(2i-1)\mathbf{B}+3\mathbf{A}] & i=1 \\ -\ [2i(2i-1)\mathbf{B}+2\mathbf{A}] & i=m \\ -\ [2i(2i-1)\mathbf{B}+3\mathbf{A}] & \text{otherwise} \end{cases}$$

$$\mathbf{R}_i^* = \begin{cases} [-6(2i-2)+2i(2i-1)]\mathbf{B}+4\mathbf{A} & i=2 \\ [-6(2i-2)+2i(2i-1)]\mathbf{B}+3\mathbf{A} & \text{otherwise} \end{cases}$$

The linear system (5.15) for c_{ij} has a sparse matrix, with at most four $n \times n$ blocks in each row as shown in Figure 1. The use of Gauss elimination with interchanges (applied to the scalar equations for c_{ij}) leads to an upper triangular matrix of the form shown in Figure 2, with at most 4 non-zero $n \times n$ blocks (the first being upper triangular) in each row.

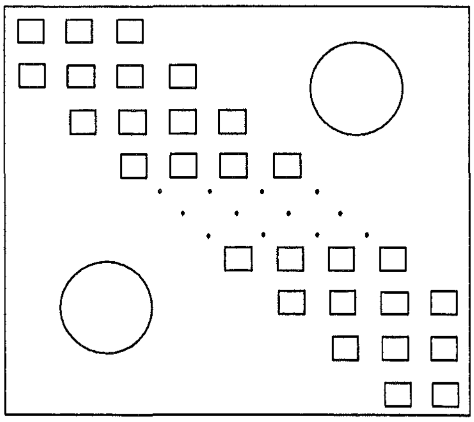

Figure 1. Original block structure

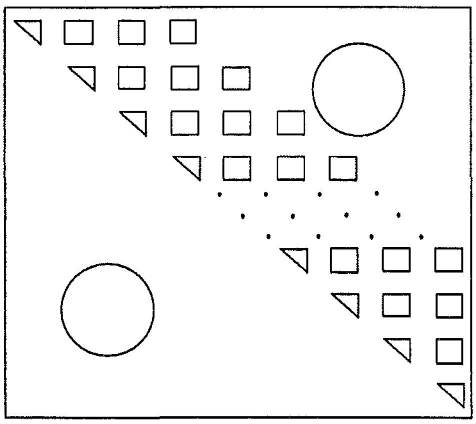

Figure 2. After Gauss elimination

The number of multiplications involved in forming the upper triangular matrix is at most

$$\text{(5.17)}\quad m \sum_{k=1}^{n} (2n-k)(4n-k) \simeq 5\tfrac{1}{3}\, mn^3 .$$

Thus, if m is linear in n, the operations count is $O(n^4)$ and the algorithm, in its basic form, is comparable in efficiency with the collocation-tau method of §4.

This Gauss elimination procedure was found to work extremely well in practice, and systems of 256 or more equations (e.g. m=n=16 or more) were readily solved accurately and speedily for $\{c_{ij}\}$.

5.1 More Efficient Versions of the Method

The algorithm, as described above, does not exploit the fact that the blocks $\mathbf{Q}_i^*$ and $\mathbf{R}_i^*$, given by (5.16), are linear combinations of the fixed matrices A and B. Nor does it exploit the fact that the block-banded matrix has a lower diagonal of matrices all but one of which are A, and an upper band of matrices all of which are $-\mathbf{A}$. It is not difficult to use this information to speed up the Gauss elimination procedure considerably, by carrying out matrix decompositions of A and B. Our efforts so far have succeeded in reducing the operations count (5.17) by a factor of about 4, while maintaining the numerical stability of the procedure.

The operations count can in principle be reduced to $O(n^3)$ in a number of ways, but none of the techniques that we have tested has yet led to a numerically effective algorithm. For example, recurrence relations can be obtained for $\mathbf{V}_i$ and $\mathbf{w}_i$ if we assume the form

$$\mathbf{c}_i = \mathbf{V}_i \, \mathbf{c}_n + \mathbf{w}_i$$

and substitute into (5.15). A fast $O(n^3)$ procedure analogous to that of §4.2 can then be derived by exploiting the relations (5.16). However, in practice the entries in the resulting matrices $\mathbf{V}_i$ seem to be very large, and large numerical errors are consequently generated. Other approaches that we have adopted also seem to have analogous difficulties. Our present approach is therefore to keep to the very satisfactory Gauss elimination procedure and to optimise the operations count in this method at the $O(n^4)$ level, as discussed above.

6. NUMERICAL RESULTS

All problems now considered belong to the general category (3.1) (3.2) (i.e. (5.1) (5.2)) above. Three specific examples, of progressively more complicated form are adopted and solved by the collocation-series method of §5.

Example 1. Torsion Problem

Consider first a standard torsion problem, with no δ term and constant γ and f terms, namely

$$\gamma = .5625, \quad \delta = 0, \quad f = 3600 \quad \text{in (3.1)}.$$

This corresponds, after transformation to the square $-1 \leq x, y \leq 1$, to the Poisson problem

$$u_{xx} + u_{yy} + 14400 = 0 \quad \text{on} \quad |x| \leq \tfrac{1}{2}, \; |y| \leq \tfrac{2}{3}.$$

The solution has a second derivative discontinuity at the corners of the region, and so some loss of accuracy may well occur in numerical solutions around these points (although this should not prevent convergence). In Table 1 we give numerical results for u^* obtained at 3 sample points $(0,0)$, $(.9,0)$ and $(.9,.9)$ (central to the region and close to the boundary) for a variety of values of m and n, and it is clear that very high accuracy has been achieved. For m=n=16, u^* is clearly correct to at least 11, 5, and 5 figures at the respective points. Relative to its maximum value, u^* has about 7 figures correct at the point $(.9,.9)$.

Table 1	Torsion Problem - Sample Numerical Solutions		
m(=n)	$u^*(0,0)$	$u^*(.9,0)$	$u^*(.9,.9)$
4	1349.56667070	269.61744	83.37992
6	1349.56166361	271.22896	82.34699
8	1349.56174907	271.41149	82.44040
10	1349.56174675	271.38574	82.41425
12	1349.56174557	271.37507	82.42026
14	1349.56174519	271.37858	82.41956
16	1349.56174506	271.37941	82.41930

In Table 2 we give the computed Chebyshev series coefficients c_{ij} for m=n=4, which define the approximate solution u^* in the form (5.3).

Table 2	Torsion Problem - Chebyshev Series Coefficients : m=n=4		
c_{11}	$.18027938 \times 10^4$	c_{31}	$.62316317 \times 10^2$
c_{12}	$.48073771 \times 10^3$	c_{32}	$.10627902 \times 10^3$
c_{13}	$.10009029 \times 10^3$	c_{33}	$.63997927 \times 10^2$
c_{14}	$.29627816 \times 10^2$	c_{34}	$.26766070 \times 10^2$
c_{21}	$.31705951 \times 10^3$	c_{41}	$.11473890 \times 10^2$
c_{22}	$.37866829 \times 10^3$	c_{42}	$.20675183 \times 10^2$
c_{23}	$.14930920 \times 10^3$	c_{43}	$.14517153 \times 10^2$
c_{24}	$.49836904 \times 10^2$	c_{44}	$.69583688 \times 10^1$

Example 2. Torsion with Variable Coefficients

For this problem, δ is again taken to be zero, but $\gamma(y)$ is allowed to vary with y; specifically we consider the case

$$\gamma = \left(1 - \tfrac{1}{2}y^2\right)^2; \quad \delta = 0, \quad f = 2 \ \text{in (3.1)} .$$

The results u^* obtained for this problem at the sample points are given in Table 3 and it is clear that at $(0,0)$, $(.9,0)$ and $(.9,.9)$ the solution is correct to 7–9 decimal places for m=n=16. High accuracy has again been achieved.

Table 3	Torsion With Variable Coefficients - Sample Solutions		
m(=n)	$u^*(0,0)$	$u^*(.9,0)$	$u^*(.9,.9)$
4	.66330955411	.13454191413	.051612213
6	.66386936129	.13688050186	.049780573
8	.66387856009	.13723959239	.049970897
10	.66387878560	.13725129342	.049955295
12	.66387879786	.13724333440	.049962553
14	.66387879977	.13724505733	.049962609
16	.66387880026	.13724534863	.049962065

Example 3. Confined Aquifer Problem

Consider the steady state flow in a confined aquifer, given by the equation (3.4), corresponding to the specific case

$$\gamma = \left(1 - \tfrac{1}{2}y^2\right)^2, \quad \delta = \gamma'(y), \quad f = \exp\left(x^2 - y^2\right) \text{ in } (3.1),$$

namely

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial}{\partial y}\left[\left(1 - \tfrac{1}{2}y^2\right)^2 \frac{\partial u}{\partial y}\right] + \exp\left(x^2 - y^2\right) = 0 .$$

The results u^* at $(0,0)$, $(.9,0)$ and $(.9,.9)$ are given in Table 4, and, in this case, fairly uniform results of about 5–6 decimal places of accuracy are achieved for m=n=10.

Table 4 Steady State Flow in a Confined Aquifer - Sample Solutions

m(=n)	$u^*(0,0)$	$u^*(.9,0)$	$u^*(.9,.9)$
4	.389807095	.09631361	.03300478
6	.391141736	.09811100	.03190094
8	.391173149	.09835505	.03202999
10	.391173991	.09836919	.03202387

In Table 5 are given the computed Chebyshev series coefficients c_{ij} for m=n=4.

Table 5 Steady State Flow in a Confined Aquifer - Chebyshev Coefficients m=n=4

c_{11}	$.67575820 \times 10^0$	c_{31}	$.27246550 \times 10^{-1}$
c_{12}	$.27653375 \times 10^0$	c_{32}	$.41748915 \times 10^{-1}$
c_{13}	$.93330349 \times 10^{-1}$	c_{33}	$.29633952 \times 10^{-1}$
c_{14}	$.31186375 \times 10^{-1}$	c_{34}	$.14365799 \times 10^{-1}$
c_{21}	$.17025069 \times 10^0$	c_{41}	$.46038077 \times 10^{-2}$
c_{22}	$.14565509 \times 10^0$	c_{42}	$.81336266 \times 10^{-2}$
c_{23}	$.83935323 \times 10^{-1}$	c_{43}	$.60895092 \times 10^{-2}$
c_{24}	$.35627641 \times 10^{-1}$	c_{44}	$.31358589 \times 10^{-2}$

7. CONCLUSIONS

(i) The collocation-series method, based on approximation in the Chebyshev polynomial form (5.3), has yielded numerical solutions of high accuracy.

(ii) The method is applicable and has good numerical properties for relatively large degrees m and n in the form of approximation.

(iii) The method requires only $O(n^4)$ operations (where m is proportional to n) and is therefore relatively efficient.

(iv) Although it is in principle possible to reduce the operation count in the method to $O(n^3)$, this has not yet been achieved without sacrificing numerical stability. The challenge therefore remains to adapt the algorithm into (a stable) one with only $O(n^3)$ operations.

(v) The earlier collocation-tau method of Mason and Olaofe may be carried out in $O(n^3)$ operations, but it is only numerically workable for modest degrees m and n in x and y.

ACKNOWLEDGEMENTS

We sincerely thank Dr R.S.Anderssen (CSIRO, Canberra) and Dr A.R. Davies (University of Aberystwyth), the organisers of this Mini-Symposium, for their encouragement and patience, and we are grateful also to Dr A. McNabb (Massey University, N.Z.) for his advice on relevant applications. We acknowledge the academic support of the Mathematics Departments at RMCS (Cranfield) Shrivenham, Australian Defence Force Academy (University of New South Wales), and University of Canterbury, and of the Centre for Mathematical Analysis at Australian National University.

REFERENCES

[1] C. CANUTO, M.Y. HUSSAINI, A. QUARTERONI, and T. ZANG, *Spectral Methods in Fluid Dynamics*, Springer Verlag, Berlin, 1988.
[2] D.B. HAIDVOGEL and T. ZANG, *The accurate solution of Poisson's equation by expansion in Chebyshev polynomials*, J. Comp. Phys. **30** (1979), 167–180.
[3] J.C. MASON, *Chebyshev polynomial approximations for the L–membrane eigenvalue problem*, SIAM J. Appl. Math. **15** (1967), 172–186.
[4] J.C. MASON, *Some properties and applications of Chebyshev polynomial and rational approximation.* In "Rational Approximation and Interpolation", P.R. Graves-Morris, E.B. Saff and R.S. Varga (Eds.), Springer Verlag, Berlin, (1984), 27–48.
[5] J.C. MASON and G.O. OLAOFE, *Vector collocation – tau method for linear partial differential equations*, Math. Comput. Modelling **11** (1988), 656–660.
[6] C. LANCZOS, *Applied Analysis*, Prentice Hall (1956).
[7] C.W. CLENSHAW, *The numerical solution of linear differential equations in Chebyshev series*, Proc. Camb. Phil. Soc. **53** (1957), 134–149.

Applied and Computational Mathematics Group
Royal Military College of Science
Shrivenham
Swindon
Wiltshire
SN6 8LA
ENGLAND

Department of Mathematics
University of Canterbury
Christchurch
NEW ZEALAND