IMA Journal of Numerical Analysis (1987) **7,** 15–37

A Note on C^0 Chebyshev Methods for Parabolic P.D.E.s

M. Berzins and P. M. Dew

Department of Computer Studies, The University, Leeds LS2 9JT

[Received 19 May 1983 and in final revised form 10 July 1986]

A family of spatial discretization formulas, based on piecewise Chebyshev expansions with C^0 continuity, is given for the solution of a general class of parabolic equations. These formulas are obtained by first expressing the generalized Chebyshev method of Berzins & Dew (1981) in a Galerkin framework, and then using this framework to simplify the method. An analysis of the new and old discretization formulas is given and a comparison made with a finite-difference method. A method is described for obtaining an indication of the error in the numerical solution that takes account of both the spatial and temporal approximations.

1. Introduction

In recent years there has been considerable interest in the development of general-purpose codes for time-dependent partial differential equations (see the survey by Machura & Sweet (1980)). These codes are generally based on the method of lines using Gear's method for the temporal integration. Two examples are the C^1 collocation code PDECOL written by Madsen & Sincovec (1978) and the finite-difference code of Dew & Walsh (1981). PDECOL is the first widely available general-purpose code to provide the user with the option of selecting the order of the approximation to be used in spatial discretization. The formulas provided by PDECOL have been derived using C^1 polynomial approximations.

In this paper we are concerned with deriving a family of spatial discretization formulas that are based upon piecewise polynomials, with C^0 continuity. The formulas are easy to implement and apply to a wide range of parabolic equations. The advantage of using C^0 continuity, compared with C^1 continuity, is its much wider applicability (e.g. problems with material interfaces and with discontinuous initial and boundary conditions) coupled with the fact that it is possible to derive a complete class of formulas, including first- and second-order ones. The formulas have been derived using an improved form of the generalized Chebyshev method of Berzins & Dew (1981). A distinctive feature of the method is the use of a Chebyshev weighted inner product that allows the well-known advantages of Chebyshev polynomial approximations to be exploited.

The family of discretization formulas derived in this paper have been implemented in a general-purpose p.d.e. solver SGENCO, which is part of the SPRINT package of Berzins, Dew, & Furzeland (1985). The main drawback of providing such a general-purpose routine is that the user now has to make an additional choice in selecting the order of the formula to be used. In order to help

with this task we have been investigating various ways of providing an error indicator which combines the errors in both spatial discretization and temporal integration using Gear's method. One such technique is described in Section 4. Although such indicators are in their infancy, this is clearly a necessary first step towards selecting automatically the order of the polynomial to be used in the spatial discretization.

2. Preliminaries

For the sake of clarity we shall consider a problem class that is sufficiently general to illustrate the main features of the method and remark that the method extends naturally to systems of partial differential equations and to boundary conditions of much greater generality. The class of parabolic equations considered is given by

$$\frac{\partial}{\partial x} r\left(x, t, u, \frac{\partial u}{\partial x}\right) = q\left(x, t, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial t}\right) \quad ((x, t) \in \Omega), \tag{2.1}$$

where $\Omega = [a, b] \times (0, t_e]$ and $u = u(x, t)$, with

$$q\left(x, t, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial t}\right) = c(x, t, u)\frac{\partial u}{\partial t} - f\left(x, t, u, \frac{\partial u}{\partial x}\right), \tag{2.2}$$

c being bounded by constants c_1 and c_2:

$$0 < c_1 < c(x, t, u) < c_2 \quad \forall (x, t) \in \Omega.$$

The boundary conditions are taken to be of the form

$$u(a, t) = 0, \qquad r\left(x, t, u, \frac{\partial u}{\partial x}\right) = g(t, u) \quad \text{at } x = b \quad (t \in (0, t_e]) \tag{2.3}$$

and the initial condition has the form

$$u(x, 0) = k(x) \quad (x \in [a, b]). \tag{2.4}$$

We assume that the p.d.e. defined by the above equation is well posed and has a unique continuous solution $u(x, t)$, for all $(x, t) \in \Omega$. For each $t \in (0, t_e]$, we shall approximate the solution by a C_0-continuous piecewise polynomial $U(\bullet, t)$, with C^0 continuity in t, and introduce the approximate functions

$$\bar{q}(x, t) = c(x, t, U)\frac{\partial U}{\partial t} - f\left(x, t, U, \frac{\partial U}{\partial x}\right), \qquad \bar{r}(x, t) = r\left(x, t, U, \frac{\partial U}{\partial x}\right). \tag{2.5}$$

It is assumed throughout that:

1. the functions $\bar{q}$ and $\bar{r}$ tend to q and r respectively as U tends to u;
2. r is continuous on Ω;
3. $\bar{q}$ is piecewise continuous in the x variable with known points of discontinuity that are independent of t.

The spatial mesh is defined by $\delta = \{X_0, \ldots, X_J\}$, where

$$a = X_0 < X_1 < \cdots < X_J = b; \tag{2.6}$$

the X_j are referred to as break points. This mesh partitions the interval $[a, b]$ into J subintervals,

$$I_j = [X_{j-1}, X_j] \quad (j = 1, \ldots, J) \tag{2.7}$$

of length h_j, where

$$h_j = X_j - X_{j-1} \quad (j = 1, \ldots, J). \tag{2.8}$$

We shall assume that the break points are chosen to include any points in the interval $[a, b]$ at which the function $q(\bullet)$ is discontinuous with respect to the spatial variable x.

It is helpful at this stage to define the piecewise-polynomial spaces that are used in this paper. Let

$$M^k(r, \delta) := \{v \in C^k[a, b] : v \in P_r(I_j) \ (j = 1, \ldots, J)\}, \tag{2.9}$$

where $P_r(I_j)$ denotes the set of polynomials of degree $\leqslant r$ defined on the interval I_j; further, let

$$M^{*k}(r, \delta) := M^k(r, \delta) \cap \{v : v(X_j) = 0 \quad (j = 1, \ldots, J)\} \tag{2.10}$$

and

$$\bar{M}(r, \delta) := M^0(r, \delta) \cap \{v : v(a) = 0\}. \tag{2.11}$$

These spaces have been used by a number of authors to analyse spatial discretization methods; see, for example, Dupont (1976).

The following two t-parametrized families of inner products are also used: the L_2 *inner product*

$$(u, v)^t = \int_a^b u(x, t)v(x, t)\,dx \quad (t \in (0, t_e])$$

and the *piecewise-Chebyshev inner product*

$$(u, v)^t_\rho = \frac{4}{\pi} \sum_{j=1}^{J} \int_{X_{j-1}}^{X_j} \rho_i(x)u(x, t)v(x, t)\,dx \quad (t \in (0, t_e]),$$

where $u(\bullet, t)$ and $v(\bullet, t)$ are in $M^k(N, \delta)$ and

$$\rho_j(x) = [1 - W_j^2(x)]^{-\frac{1}{2}} \quad (x \in I_j);$$

here, W_j is the linear mapping of I_j onto $[-1, 1]$. The norm associated with the Chebyshev inner product is denoted by

$$\|u(\bullet, t)\|_\rho = \sqrt{(u, u)^t_\rho}$$

3. Chebyshev C^0 collocation method

The generalized Chebyshev method of Berzins & Dew (1981) was derived by using a series-expansion method that exploited the properties of Chebyshev polynomials. This section shows that the same method can also be derived by applying a Galerkin-type method to the original p.d.e. This allows the relationship between the generalized method and the C^0-collocation method discussed by

Diaz (1977) and Wheeler (1977) to be established, and a new discretization method to be devised. This new method, referred to as the *Chebyshev* C^0 *collocation method,* retains the advantages of the generalized method but is simpler and more efficient to implement in a computer program.

The general idea of the generalized method is first to approximate the solution $u(\bullet, t)$ of equations (2.1)–(2.3) by a piecewise polynomial approximation $U(\bullet, t) \in \bar{M}(N, \delta)$, where N is the degree of the polynomial used in each subinterval. The approximate solution may then be written as

$$U_j(x, t) = \sum_{i=0}^{N} a_{j,i}(t) T_i(W_j(x)) \quad (x \in I_j; j = 1, \ldots, J), \tag{3.1}$$

for $t \in (0, t_e]$, where $U_j(\bullet, t)$ is the restriction of $U(\bullet, t)$ to I_j, with T_i the Chebyshev polynomial of the first kind of degree i and w_j as defined above.

To derive the generalized method we introduce the piecewise polynomial approximations $Q(\bullet, t)$ and $R(\bullet, t)$ which interpolate the functions $\bar{q}(\bullet, t)$ and $\bar{r}(\bullet, t)$ respectively (see equations (2.5)) at the transformed Chebyshev points $x_{j,i}$ defined by

$$W_j(x_{j,i}) = \cos\frac{(N-i)\pi}{N} \quad (j = 1, \ldots, J; i = 0, \ldots, N). \tag{3.2}$$

The reader should note that the transformed Chebyshev points $x_{j,N}$ and $x_{j+1,0}$ are both equal to the break point X_j for $j = 1, \ldots, J-1$. The functions $Q(\bullet, t)$ and $R(\bullet, t)$ can both be written as Chebyshev polynomial expansions of the form given by equation (3.1); the polynomial coefficients are respectively denoted by $q_{j,i}(t)$ and $r_{j,i}(t)$ $(j = 1, \ldots, J;\ i = 0, \ldots, N)$. The function $\bar{q}(\bullet, t)$ is assumed to be piecewise continuous (in the x variable) with any discontinuities at the break points X_j and hence the approximating piecewise polynomial $Q(\bullet, t)$ may also be discontinuous at these points. The discretization methods described here require the values of $Q(x, t)$ as x tends to the break point from above and below. To handle this situation we use the following convention:

$$Q(x_{j+1,0}, t) = \lim_{\varepsilon \to 0+} Q(X_j + \varepsilon, t), \qquad Q(x_{j,N}, T) = \lim_{\varepsilon \to 0+} Q(X_j - \varepsilon, t). \tag{3.3}$$

The same convention is also used for the piecewise polynomials $\partial U/\partial x$, R, and $\partial R/\partial x$, which may also be discontinuous at the break points.

Having defined the polynomials Q and R, the discretization method can be derived from the approximate identity

$$\frac{\partial R}{\partial x} \approx Q.$$

The details of the discretization formulation are fully given in Berzins & Dew (1981). It is sufficient for our purposes to remark that the generalized method can

also be derived from the following pair of orthogonality conditions:

$$\left(\frac{\partial R}{\partial x} - Q, w\right)_\rho^t = 0 \quad \forall\ w(\cdot, t) \in \mathrm{M}^{*k}(N, \delta) \quad (t \geqslant 0), \tag{3.4}$$

$$\left(\frac{\partial R}{\partial x} - Q, v\right)^t = 0 \quad \forall\ v(\cdot, t) \in \bar{\mathrm{M}}(1, \delta), \quad (t \geqslant 0). \tag{3.5}$$

We shall assume that the inner products are continuous in the parameter t. By using the basis

$$w_{i,j}(x) = \frac{\mathrm{T}_i(W_j(x)) - \mathrm{T}_{i-2}(W_j(x))}{2i-2} \quad (i = 2, \ldots, N; j = 1, \ldots, J)$$

for the space $\mathrm{M}^{*k}(N, \delta)$, and integrating (3.4) by parts, it is possible to derive the simple relation given by equation (2.8) in Berzins & Dew (1981). Full details of the derivation of these conditions can be found in Berzins 1982. Both the conditions (3.4) and (3.5) are evaluated in their weak forms; for example, condition (3.5) becomes

$$\left(R, \frac{\mathrm{d}v}{\mathrm{d}x}\right)^t + (Q, v)^t - g(t, U(b, t))v(b) = 0 \quad (t \geqslant 0). \tag{3.6}$$

The generalized method was derived by using the properties of Chebyshev polynomials to evaluate the inner products analytically. The formulation above shows that the terminology of Wheeler (1977) may be used to refer more accurately to the generalized method as a C^0 Chebyshev–Galerkin Method.

In the next part of the paper, sub-optimal quadrature rules (lumping) are used to approximate the inner products, since this leads to a simpler and more efficient method. This is the *Chebyshev* C^0 *collocation method.* We shall consider each of the orthogonality conditions (3.4) and (3.5) separately by applying different quadrature rules to take account of the different weight functions in each of the inner products.

3.1 *Orthogonality Condition* (3.4)

Following a similar approach to that of Diaz (1977) we can apply the $(N+1)$-point Gauss–Lobatto quadrature rule with the Chebyshev weight function to the inner product in equation (3.4) to obtain the equations

$$\frac{\partial R}{\partial x}(x_{j,i}, t) - Q(x_{j,i}, t) = 0 \quad (j = 1, \ldots, J; i = 1, \ldots, N-1), \tag{3.7}$$

where the points $x_{j,i}$ are defined by equation (3.2). In the case when

$$\frac{\partial R}{\partial x}(x_{j,i}, t) = \frac{\partial \bar{r}}{\partial x}(x_{j,i}, t),$$

equations (3.7) are identical to those obtained by collocating at the transformed Chebyshev points $x_{j,i}$. This condition holds when $\bar{r}(x, t)$ is a polynomial of degree

N or less on each subinterval, e.g. when

$$\bar{r}(x, t) = (ax + b)\frac{\partial U}{\partial x} + cU$$

for real constants a, b, and c. For this reason, we refer to (3.7) as collocation-like equations and to the method as a Chebyshev C^0 collocation method to distinguish it from classical Galerkin methods.

Equation (3.7) can be written as an explicit equation for $\partial U/\partial t$ at the point $x_{j,i}$:

$$c(x_{j,t}, t, U_{j,i})\frac{\mathrm{d}U_{j,i}}{\mathrm{d}t} = \frac{\partial R}{\partial x}(x_{j,i}, t) + f\left(x_{j,i}, t, U_{j,i}, \frac{\mathrm{d}U_{j,i}}{\mathrm{d}x}\right)$$

where $U_{j,i} = U(x_{j,i}, t)$ $(j = 1, \ldots, J; i = 1, \ldots, N-1)$. The \right-hand side of this equation is much simpler than the expression we obtained using the method of Berzins & Dew (1981). Any loss of accuracy due to the use of the quadrature rule is unlikely to be significant, since we can show that the approximate solution $U \in \bar{\mathrm{M}}(N, \delta)$ that satisfies equations (3.7) is also the exact solution of the perturbed differential equation

$$\frac{\partial R}{\partial x} - \frac{\partial P_j}{\partial x} = Q(x, t) \quad ((x, t) \in (X_{j-1}, X_j) \times (0, T_e];\ j = 1, \ldots, J), \tag{3.8}$$

where

$$P_j(x, t) = -\frac{h_j}{4}\left[\frac{q_{j,N-1}(t)}{N}T_N(y) + q_{j,N}(t)\left(\frac{T_{N+1}(y)}{N+1} + \varphi\right)\right] + r_{j,N}(t)T_N(y),$$

in which

$$\varphi = T_{N-1}(y)/(N-1), \qquad y = W_j(x) \quad (x \in I_j), \qquad y \in [-1, 1].$$

In the case of the generalized method, the parameter φ is zero. Both of these results depend on the elementary properties of Chebyshev polynomials; see Berzins (1982).

3.2. *Orthogonality Condition* (3.5)

By the application of a quadrature rule to equation (3.4) we were able to approximate this orthogonality condition by a set of equations, each one of which explicitly defines the time derivative at a single collocation point. In this section we similarly use the Clenshaw–Curtis quadrature rule to approximate the orthogonality condition (3.5) by a set of equations, each of which explicitly defines the time derivative at a break point. Let $\{v_j\}_{j=1}^{J}$ denote the set of linear basis (hat) functions that span the space $\bar{\mathrm{M}}(1, \delta)$ where

$$v_j(X_i) = 1 \quad (i = j), \qquad v_j(X_i) = 0 \quad \text{otherwise.} \tag{3.9}$$

Using these basis functions, equation (3.5) can be written as

$$\left(R, \frac{dv_j}{dx}\right)^t + (Q, v_j)^t = 0 \quad (j = 1, \ldots, J-1), \tag{3.10}$$

$$\left(R, \frac{dv_J}{dx}\right)^t + (Q, v_J)^t = g(t, U(b, t)). \tag{3.11}$$

For $j = 1, \ldots, J-1$, the integrand Qv_j in the inner product $(Q, v_j)^t$ is a polynomial of degree $N+1$ defined on the intervals I_j and I_{j+1} and is zero elsewhere. The function v_J, and hence the integrand, is only nonzero on the final interval I_J. We approximate the integrand by an interpolating polynomial (of degree N) at the transformed Chebyshev points $x_{j,i}$ and integrate the resulting inner product *exactly*. Equivalently, we can use an $(N+1)$-point Clenshaw–Curtis quadrature rule. For simplicity, we shall consider the interior intervals (i.e. $j \neq J$). In this case,

$$(Q, v_j)^t \approx \sum_{i=0}^{N} \frac{\lambda_i}{2} [h_j Q(x_{j,i}, t) v_j(x_{j,i}) + h_{j+1} Q(x_{j+1,i}, t) v_j(x_{j+1,i})], \tag{3.12}$$

where $\{\lambda_i\}_{i=0}^N$ denote the weights of the $(N+1)$-point Clenshaw–Curtis quadrature rule. The values of the weights may be found, for example, in Imhof (1963). Using equation (3.7) we can rewrite equation (3.12) as

$$(Q, v_j)^t \approx \tfrac{1}{2} h_j \lambda_N Q(x_{j,N}, t) + \tfrac{1}{2} h_{j+1} \lambda_0 Q(x_{j+1,0}, t) + \sum_{i=1}^{N-1} \frac{\lambda_i}{2} \left(h_j \frac{\partial R}{\partial x}(x_{j,i}, t) v_j(x_{j,i}) + h_{j+1} \frac{\partial R}{\partial x}(x_{j+1,i}, t) v_j(x_{j+1,i}) \right)$$

where the evaluation of the functions at $x_{j,N}$ and $x_{j+1,0}$ is defined by equation (3.3). Similarly we can apply the $(N+1)$-point Clenshaw–Curtis quadrature rule to evaluate exactly the inner product $\left(R, \frac{dv_j}{dx}\right)^t$. Thus, on noting that $\lambda_0 = \lambda_N$, we obtain from equation (3.10), for $j = 1, \ldots, J-1$, the expression

$$\begin{aligned} &h_j Q(x_{j,N}, t) + h_{j+1} Q(x_{j+1,0}, t) \\ &\quad = R(x_{j+1,0}, t) + R(x_{j+1,N}, t) - R(x_{j,0}, t) - R(x_{j,N}, t) + \lambda_N^{-1} \sum_{i=1}^{N-1} \lambda_i \Big(R(x_{j+1,i}, t) \\ &\qquad - R(x_{j,i}, t) - h_{j+1} \frac{\partial R}{\partial x}(x_{j+1,i}, t) v_j(x_{j+1,i}) - h_j \frac{\partial R}{\partial x}(x_{j,i}, t) v_j(x_{j,i}) \Big). \end{aligned} \tag{3.13}$$

A similar expression for the case when $j = J$ can be easily found by using equation (3.11).

In the case when $N > 1$, equation (3.13) can be simplified further at the interior break points, by using integration by parts on $\left(\frac{\partial R}{\partial x}, v_j\right)^t$ and noting that $v_j(x)$ is zero at $x = a$ and $x = b$, to give:

$$\left(\frac{\partial R}{\partial x}, v_j\right)^t + \left(R, \frac{dv_j}{dx}\right)^t = 0 \quad (j = 1, \ldots, J-1).$$

Applying the Clenshaw–Curtis rule to evaluate exactly these inner products gives

$$\sum_{i=0}^{N} \lambda_i \Big(R(x_{j+1,i}, t) - R(x_{j,i}, t) - h_{j+1} \frac{\partial R}{\partial x}(x_{j+1,i}, t) v_j(x_{j+1,i}) - h_j \frac{\partial R}{\partial x}(x_{j,i}, t) v_j(x_{j,i}) \Big) = 0 \quad (3.14)$$

$(j = 1, \ldots, J-1)$; and on substitution into equation (3.13) we obtain the simpler expression

$$h_j Q(x_{j,N}, t) + h_{j+1} Q(x_{j+1,0}, t) = h_{j+1} \frac{\partial R}{\partial x}(x_{j+1,0}, t) + h_j \frac{\partial R}{\partial x}(x_{j,N}, t). \quad (3.15)$$

$(j = 1, \ldots, J-1)$. In the case when $N = 1$, the function R may depend only on $\partial U/\partial x$ and so be piecewise constant. The identity (3.14) then only holds trivially, and so cannot be used to simplify equation (3.13).

Assuming that $\partial U/\partial t$ is continuous in x for all $t \in (0, t_e]$, we can use the definition of $Q(x, t)$ to write the left-hand side of equation (3.13) as

$$h_j Q(x_{j,N}, t) + h_{j+1} Q(x_{j+1,0}, t) = \quad (3.16)$$

$$[h_j c(x_{j,N}, t, U_j) + h_{j+1} c(x_{j+1,0}, t, U_j)] \frac{\partial U}{\partial t}(X_j, t) - h_j f\Big(x_{j,n}, t, U_j, \frac{\partial U}{\partial x}(x_{j,N}, t)\Big) - h_{j+1} f\Big(x_{j+1,0}, t, U_j, \frac{\partial U}{\partial x}(x_{j+1,0}, t)\Big)$$

where $U_j = U(X_j, t)$. This gives an explicit formula for $\partial U/\partial t$ at the break point X_j. This break-point condition is also simpler than the one that is used in the generalized Chebyshev method. The expression derived from equation (3.11) can be treated in the same way and the Dirichlet boundary condition is handled directly by enforcing the condition $U(a, t) = 0$. Finally on combining equations (3.7), the interface conditions (3.13), and the boundary conditions, we can define the *Chebyshev* C^0 *collocation method.* An algorithmic description of the method is given in Appendix I.

THEOREM 1. *The function* $U(\bullet, t) \in \bar{M}(N, \delta)$, *that satisfies equations* (3.7) *and* (3.13) *and the boundary conditions, also satisfies the orthogonality conditions*

$$\Big(\frac{\partial R}{\partial x} - Q, w\Big)_\rho^t = \Big(\frac{\partial G}{\partial x}, w\Big)_\rho^t \quad \forall\ w \in M^{*k}(N, \delta) \quad (N > 1), \quad (3.17)$$

$$\Big(\frac{\partial R}{\partial x} - Q, v_j\Big)^t = Z_j(t) \quad (j = 1, \ldots, J-1), \quad (3.18)$$

where

(i) $G(x, t) = -\dfrac{h_j}{4} q_{j,N}(t) \dfrac{T_{N-1}(y)}{N-1}, \quad y = W_j(x) \quad (x \in I_j, N > 1),$

(ii) $v_j(x)$ *is defined by equation* (3.9),

(iii) $$Z_j(t) = \begin{cases} \dfrac{2}{N(N^2-4)}\left[h_j q_{j,N}(t) - \varepsilon_{j+1} h_{j+1} q_{j+1,N}(t)\right] & (N \text{ odd}), \\ 0 \quad (N \text{ even}), \end{cases}$$

with $\varepsilon_i = 1$ $(i \neq J)$ *and* $\varepsilon_i = 0$ $(i = J)$.

Remark. Equation (3.18) may be written as an equation for the continuity of the flux R at the break point X_j by integrating it by parts. This is the approach applied by Berzins & Dew (1981) to equation (3.5) to arrive at the flux continuity condition in Theorem 1 of that paper.

Proof. Equation (3.17) follows directly from equation (3.8) by using the perturbed equations satisfied by the Chebyshev C^0 collocation method and by using the orthogonality of the Chebyshev polynomials with respect to the inner product $(\bullet, \bullet)_\rho^t$. In applying quadrature to equation (3.5) we have interpolated the function $-Qv_j$ by a polynomial of degree N on the intervals I_j and I_{j+1} and then integrated this polynomial exactly. In other words, the function

$$\psi_{j,N}(x, t) = \begin{cases} \dfrac{q_{k,N}(t)(-1)^{k-j}}{4h_k}\left[T_{N+1}(W_k(x)) - T_{N-1}(W_k(x))\right] & (x \in I_k;\ k = j, j+1), \\ 0 \quad \text{elsewhere}, \end{cases}$$

has been added to $-Qv_j$ and the resulting polynomial has been exactly integrated. Note that, from the properties of Chebyshev polynomials,

$$\psi_{j,N}(x_{k,i}, t) = 0 \quad (k = j, j+1;\ i = 0, \ldots, N). \quad \square$$

This theorem shows that the effect of applying the quadrature rules (lumping) to the generalized Chebyshev method is to perturb the inner product in equations (3.5) and (3.6) by a function that depends on the least significant polynomial coefficient of $Q(x, t)$, namely $q_{j,N}(t)$. Indeed, from page 61 of Fox & Parker (1968), we can show that $q_{j,N}(t)$ is $O(h_j^N/N!)$, and we would therefore expect that this perturbation has relatively little effect on the accuracy of the solution.

Numerical Example. From Theorem 1 we would expect no significant difference between the solutions obtained by using the two methods. To show that this is indeed the case, consider the following example in spherical polar coordinates·

$$u\frac{\partial u}{\partial t} = \frac{1}{x^2}\frac{\partial}{\partial x}\left(x^2 u \frac{\partial u}{\partial x}\right) + 5u^2 + 4xu\frac{\partial u}{\partial x}$$

for $(x, t) \in [0, 1] \times (0, 1]$; the left-hand boundary condition is the symmetry condition

$$\frac{\partial u}{\partial x}(0, t) = 0$$

and the right-hand Dirichlet condition and the initial condition are consistent with the analytic solution of $u(x, t) = e^{1-x^2-t}$.

Three codes were applied to this problem: the generalized method of Berzins &

TABLE 1
Estimates L^2 *error norm*

N	CODE	Time 0·01	0·25	0·50	0·75	1·00
5	GENERL	7·89 E -5	3·43 E -4	4·78 E -5	5·46 E -4	5·61 E -4
	SGENCO	1·07 E -5	3·53 E -4	4·84 E -4	5·57 E -4	5·73 E -4
	PDECOL	7·57 E -5	6·31 E -4	1·09 E -3	1·41 E -3	1·65 E -3
7	GENERL	2·31 E -6	3·55 E -6	4·27 E -6	4·66 E -6	4·71 E -6
	SGENCO	2·47 E -6	3·60 E -6	4·25 E -6	4·60 E -6	4·64 E -6
	PDECOL	2·51 E -6	1·47 E -5	2·51 E -5	3·55 E -5	4·00 E -5
9	GENERL	5·23 E -8	5·49 E -8	6·13 E -8	6·44 E- 8	6·39 E -8
	SGENCO	5·90 E -8	5·96 E -8	6·43 E -8	6·64 E -8	6·55 E -8
	PDECOL	3·91 E -8	2·63 E -7	4·51 E -7	6·00 E- 7	7·17 E -7

N is the degree of the polynomial used to spatially discretize the p.d.e.

Dew (1981)—GENERL; the Chebshev C^0 collocation method—GENCOL and the PDECOL code of Madsen and Sincovec (1978). In each code, a single polynomial expansion of degree 5, 7, and 9 was used to represent the solution. The o.d.e. integration was sufficiently accurate in each case to ensure that the spatial discretization error dominated. Estimates of the polar—weighted L^2 error norm, formed by using the trapezoidal rule with 100 equally spaced spatial mesh points are given in Table 1.

Similar results were obtained by Berzins (1982) on a range of other parabolic p.d.e.s; coupled with the relative simplicity of the new method, they lead us to recommend it over the generalized method. SGENCO can also be used to solve problems for which the C^1 continuity of PDECOL is unsuitable.

3.3 *Linear Basis Functions*

For many of the parabolic equations that arise in practice, low-order finite-difference methods are often acceptable. The Chebyshev C^0 collocation method described above defines a family of formulas, starting with linear basis functions. It is therefore of interest to compare the Chebyshev C^0 collocation method with one of the more-commonly used finite-difference formulas: the box scheme of Keller (1970). The Chebyshev C^0 collocation method with linear basis function can be written as

$$(h_j c_j^- + h_{j+1} c_j^+) \frac{\partial U}{\partial t}(X_j, t) = R_{j+1}^- + R_j^+ - R_j^- - R_{j-1}^+ + h_j f_j^- + h_{j+1} f_j^+, \quad (3.19)$$

where R_j^+ and R_j^- denote the function R evaluated at the break point X_j above and below with

$$\frac{\partial U^+}{\partial x_j} = \frac{U_{j+1} - U_j}{h_{j+1}}, \qquad \frac{\partial U^-}{\partial x_j} = \frac{U_j - U_{j-1}}{h_j},$$

and $U_j = U(X_j, t)$. The quantities c_j^+, c_j^-, f_j^+, and f_j^- are defined similarly and $U_j = U(X_j, t)$. Equation (3.19) was obtained in Section (3.2) by applying the trapezoidal rule to the weak form of the inner product in equation (3.5). Theorem 1 shows that the effect of using the trapezoidal rule is that a perturbed form of the inner product is exactly satisfied, namely

$$\left(\frac{\partial R}{\partial x} - Q, v_j\right)^t = Z_j(t) \quad (j = 1, \ldots, J-1).$$

If, on the other hand, we apply the mid-point rule to the inner product in equation (3.5) and note that the functions R and Q are piecewise linear, we obtain the set of equations

$$\tfrac{1}{2}h_j\left(c_{j-1}^+ \frac{\partial U_j}{\partial t} + c_j^- \frac{\partial U_j}{\partial t}\right) + \tfrac{1}{2}h_{j+1}\left(c_j^+ \frac{\partial U_j}{\partial t} + c_{j+1}^- \frac{\partial U_{j+1}}{\partial t}\right)$$
$$= (R_{j+1}^- + R_j^+ - R_j^- - R_{j-1}^+) + \tfrac{1}{2}h_j(f_j^- + f_{j-1}^+) + \tfrac{1}{2}h_{j+1}(f_j^+ + f_{j+1}^-) \quad (3.20)$$

which satisfies exactly the perturbed inner product

$$\left(\frac{\partial R}{\partial x} - Q, v_j\right)^t = -\tfrac{1}{2}Z_t(t) \quad (j = 1, \ldots, J-1).$$

This result follows immediately from Theorem 1 and the property of the mid-point rule that it is twice as accurate as the trapezoidal rule when both are applied to the same piecewise-quadratic integrand.

Equation (3.20) is an intermediate form which can easily be used to derive Keller's box scheme. It is easily shown from Taylor's series expansions that

$$R_{j+1}^- + R_j^+ = 2R_{j+\frac{1}{2}} + O(h_{j+1}^2)$$

where

$$R_{j+\frac{1}{2}} = r\left(\frac{X_j + X_{j+1}}{2}, t, \frac{U_j + U_{j+1}}{2}, \frac{U_{j+1} - U_j}{h_{j+1}}\right).$$

The functions $R_{j-\frac{1}{2}}, c_{j+\frac{1}{2}}, c_{j-\frac{1}{2}}, f_{j-\frac{1}{2}}, f_{j+\frac{1}{2}}$ are defined analogously. We can also show that

$$c_{j-1}^+ \frac{\partial U_j}{\partial t} + c_j^- \frac{\partial U_j}{\partial t} = c_{j-\frac{1}{2}}\left(\frac{\partial U_{j-1}}{\partial t} + \frac{\partial U_j}{\partial t}\right) + O(h_j^2).$$

Hence, ignoring terms of $O(h^2)$, we can apply Taylor's series to each of the terms in equation (3.20) to obtain the box scheme (Keller, 1970):

$$\frac{h_j}{2}c_{j-\frac{1}{2}}\left(\frac{\partial U_{j-1}}{\partial t} + \frac{\partial U_j}{\partial t}\right) + \frac{h_{j+1}}{2}c_{j+\frac{1}{2}}\left(\frac{\partial U_j}{\partial t} + \frac{\partial U_{j+1}}{\partial t}\right)$$
$$= 2(R_{j+\frac{1}{2}} - R_{j-\frac{1}{2}}) + h_{j+1}f_{j+\frac{1}{2}} + h_j f_{j-\frac{1}{2}} \quad (j = 2, \ldots, J-1).$$

The box scheme is second-order (Keller, 1970), and from the analysis we see that it satisfies the perturbed orthogonality condition

$$\left(\frac{\partial R}{\partial x} - Q, v_j\right)^t = \tfrac{1}{2}Z_j(t) + O(h_j^2 + h_{j+1}^2) \quad (j = 1, \ldots, J-1)$$

Noting that $Q(x, t)$ is piecewise linear in x and consequently

$$q_{j,1}(t) = Q(x_{j,N}, t) - Q(x_{j,0}, t) = h_j \frac{\partial Q}{\partial x}(\xi, t) \quad (\xi \in I_j; j = 1, \ldots, J),$$

we see that $Z_j(t)$ is $O(h_j^2) + O(h_{j+1}^2)$. Since both the Galerkin method with linear basis functions and the perturbation $Z_j(t)$ introduced by quadrature are second-order, it follows that the Chebyshev C^0 collocation method with linear basis functions is also second-order.

A possible difficulty with the Chebyshev C^0 collocation method is that the user has to supply both the left and right limit values of any function that is discontinuous at the break points. In practice this is often not a problem; the user interface described by Berzins, Dew, & Furzeland (1983) provides a solution to this problem.

3.4. *Open Quadrature Rules*

The slightly improved accuracy when using the mid-point rule and the possible problem of evaluating left and right limit values of functions at break points suggest that it may be worth considering open quadrature rules for lumping, instead of the closed quadrature rules used here. The difficulty is caused by the inner product in equation (3.5); applying open quadrature rules, and using the approach described above, results in implicit ordinary differential equations in time such as equation (3.20).

Recently, Skeel (1981) derived a modified form of the box scheme that overcomes this problem. (In the case of nonpolar parabolic equations the Skeel scheme is identical to the lumped finite-element scheme of Bakker (1977), providing that the function c in equation (2.2) is constant.) At present it is not clear if this approach can be extended to the discretization formulas described here that are based on polynomials of degree ≥ 2 and give rise to a system in normal form.

4. Error indicators for the Chebyshev C^0 collocation method

In the first part of this paper we showed how the use of quadrature rules (lumping) simplified the generalized Chebyshev method to give the *Chebyshev* C^0 *collocation method.* We now consider the problem of providing some indication of the error in the numerical solution. Two approaches are considered. The first—introduced by Delves (1976)—only estimates the spatial discretization error, while in the second approach we devise a new technique for estimating the combined error that is introduced by the spatial and temporal discretizations. This is achieved by combining the general approach of Delves with a global error indicator for Gear's method.

For notational convenience, now denoting $u = u(\bullet, \bullet)$, equations (2.1) and (2.2) will be written in the form

$$\frac{\partial u}{\partial t} = Su \quad ((x, t) \in \Omega) \tag{4.1}$$

where S is a nonlinear differential operator defined by

$$Su(x, t) = \frac{\partial}{\partial x} r\left(x, t, u(x, t), \frac{\partial u}{\partial x}(x, t)\right) + f\left(x, t, u(x, t), \frac{\partial u}{\partial x}(x, t)\right) \tag{4.2}$$

For simplicity of derivation, and without loss of generality, we have restricted equation (2.1) to the case when

$$c(x, t, u(x, t)) = 1 \quad ((x, t) \in \Omega)$$

and we assume that: (i) the function

$$(x, t) \mapsto r\left(x, t, u(x, t), \frac{\partial u}{\partial x}(x, t)\right)$$

is differentiable at the break points, (ii) the degree N of the approximating polynomial is greater than 1, and (iii) the boundary conditions are of Dirichlet type. The first and second assumptions allow the simplified form of the break-point condition, equation (3.15), to be used.

4.1 *Delves' Error Indicator for the Spatial Discretization Error*

The *spatial discretization error*, $e_s(x, t)$, is defined by

$$e_s(x, t) = u(x, t) - U(x, t) \quad ((x, t) \in \Omega).$$

Suppose that the exact solution of the p.d.e. defined by equations (2.1), (2.3), and (2.4) on the interval I_j is given by the uniformly convergent Chebyshev series

$$u_j(x, t) = \sum_{i=0}^{\infty} b_{j,i}(t) \mathrm{T}_i(W_j(x)) \quad (x \in I_j) \tag{4.3}$$

and that the truncated form is given by

$$u_j^*(x, t) = \sum_{i=0}^{N} b_{j,i}(t) \mathrm{T}_i(W_j(x)) \quad (x \in I_j). \tag{4.4}$$

Delves (1976) made two assumptions. The first assumption is that the polynomial coefficients $b_{j,i}$ converge at some power rate:

$$|b_{j,i}(t)| \approx B_j(t) i^{-r}, \qquad r > \tfrac{1}{2},$$

so that (Delves, 1976)

$$\|u_j - u_j^*\|_\rho \approx (N h_j b_{j,N}^2)^{\frac{1}{2}};$$

and the second assumption is that

$$U_j \approx u_j^* \quad (j = 0, \ldots, J).$$

It follows from these two assumptions that

$$\|u_j - U_j\|_\rho \approx (N h_j a_{j,N}^2)^{\frac{1}{2}}, \tag{4.5}$$

$$\|e_s(x, t)\|_\rho \approx \left(N \sum_{j=1}^{J} h_j a_{j,N}^2(t)\right)^{\frac{1}{2}}. \tag{4.6}$$

Although there are no proofs that we are aware of to justify these assumptions for the Chebyshev C^0 collocation method, the Delves indicator appears to work well in practice.

In the estimation of the combined error due to the spatial and temporal approximations, it is necessary to estimate the error at the individual mesh points. A simple pointwise error estimate that satisfied equation (4.6) may be derived by assuming that the error in any interval may be approximated by a spatially constant function, i.e. for the jth interval:

$$e_s(x, t) \approx k_j(t) \quad (x \in I_j; j = 1, \ldots, J). \tag{4.7}$$

On substituting the right-hand side of equation (4.7) into equation (4.5) we see that a suitable value for $k_j(t)$ is given by

$$|k_j(t)| = N^{\frac{1}{2}} |a_{j,N}(t)|$$

and hence an estimate of the spatial discretization error at the collocation points is given by

$$e_s(x_{j,i}, t) \approx N^{\frac{1}{2}} a_{j,n}(t) \quad (i = 1, \ldots, N-1; j = 1, \ldots, J). \tag{4.8}$$

The estimate used at the break point is the weighted average of the values in the adjoining two intervals:

$$e_s(X_j, t) = \frac{N^{\frac{1}{2}}}{h_j + h_{j+1}} [h_{j+1} a_{j+1,N}(t) + h_j a_{j,N}(t)] \quad (j = 1, \ldots, J-1). \tag{4.9}$$

4.2 *Integration in Time*

The system of ordinary differential equations in time defined by the Chebyshev C^0 collocation method (e.g. equations (3.7), (3.13), and (3.16)) can equivalently be written as

$$\dot{\boldsymbol{u}}(t) = \boldsymbol{f}_N(\boldsymbol{u}(t)). \tag{4.10}$$

where $(JN + 1)$-dimensional vectors are defined by

$$\boldsymbol{u}(t) = \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_J \end{bmatrix}, \qquad \boldsymbol{u}_j = \begin{bmatrix} U_{j,0} \\ U_{j,1} \\ \vdots \\ U_{j,N-1} \end{bmatrix} \quad (j = 1, \ldots, J-1), \qquad \boldsymbol{u}_J = \begin{bmatrix} U_{J,0} \\ U_{J,1} \\ \vdots \\ U_{J,N} \end{bmatrix},$$

and $U_{j,i} = U(x_{j,i}, t)$. The initial condition for $\boldsymbol{u}(t)$ is found by evaluating $k(x)$ at the transformed Chebyshev points in each element.

We now define a *restriction* operator $\boldsymbol{r}_h$ that maps from a C^0 continuous function defined on the interval $[a, b]$ to a $(NJ + 1)$-dimensional vector of the function values at the transformed Chebyshev points in each interval, including the break points. For $f \in C^0(\Omega)$, we similarly define $\boldsymbol{r}_h f$ by:

$$(\boldsymbol{r}_h f)(t) = \boldsymbol{r}_h f(\bullet, t) \quad (t \in (0, t_e].$$

Associated with the restriction operator is a *prolongation* operator

$$p_h : \mathbb{R}^{N-J+1} \to \mathrm{M}^0(N, \delta),$$

which defines a piecewise Chebyshev polynomial of degree N on each interval I_j, by using polynomial interpolation between the function values at the transformed Chebyshev points in I_j. Thus, in the case of the function U, we have

$$U(\bullet, t) = p_h \boldsymbol{u}(t), \qquad \boldsymbol{u}(t) = r_h U(\bullet, t). \tag{4.11}$$

In practice, equation (4.10) is a stiff or mildly stiff system of ordinary differential equations in time that is usually solved by Gear's method. The o.d.e. global error is defined by

$$\boldsymbol{e}_{\mathrm{g}}(t) = \boldsymbol{u}(t) - \boldsymbol{v}(t) \quad (t \in (0, t_e])$$

where $\boldsymbol{v}(t)$ is the approximation to $\boldsymbol{u}(t)$ that is computed by Gear's method. It is well known that (Shampine, 1979) $\boldsymbol{e}_{\mathrm{g}}(t)$ satisfies:

$$\boldsymbol{f}_N(\boldsymbol{v}(t_{n+1}) + \boldsymbol{e}_{\mathrm{g}}(t_{n+1})) - \boldsymbol{f}_N(\boldsymbol{v}(t_{n+1})) + \boldsymbol{e}_{\mathrm{t}}(t_{n+1}) \frac{1}{h\beta} + \left[\sum_{i=1}^{p} \alpha_i \boldsymbol{e}_{\mathrm{g}}(t_{n+1-i}) - \boldsymbol{e}_{\mathrm{g}}(t_{n+1}) \right] \frac{1}{h\beta} = 0, \qquad \boldsymbol{e}_{\mathrm{g}}(0) = 0, \tag{4.12}$$

where $h = t_{n+1} - t_n$ and $\boldsymbol{e}_{\mathrm{t}}(t_{n+1})$ is the truncation error associated with the backward-difference formula of order p, defined by the coefficients α_i $(i = 1, \ldots, p)$ and β (Gear, 1971). One of the simplest methods of estimating the global error using (4.12) is that of Dew & West (1979). This method calculates $\boldsymbol{e}_{\mathrm{g}}(t_{n+1})$ from the equations

$$(\boldsymbol{I} - h\beta \boldsymbol{G}) \boldsymbol{e}_{\mathrm{g}}(t_{n+1}) \approx \boldsymbol{e}_{\ell}(t_{n+1}) + \boldsymbol{e}_{\mathrm{g}}(t_n)$$

where $\boldsymbol{e}_{\ell}(t_{n+1})$ is the local error estimate that is used to approximate $\boldsymbol{e}_{\mathrm{t}}(t_{n+1})$ at time t_{n+1} and the matrix $\boldsymbol{G}$ is the approximation to the Jacobian matrix $\partial \boldsymbol{f}_N / \partial \boldsymbol{u}$, of the right-hand side of equations (4.10) that is used by Gear's method.

4.3 *A Combined O.D.E.–P.D.E. Error Indicator*

We shall now consider an error indicator that also takes the propagation of the error in time into consideration. Our aim is to estimate the overall error in the numerical solution due to spatial discretization and to temporal integration. The combined p.d.e. spatial discretization error and o.d.e. global error is defined by

$$\boldsymbol{e}(t) = r_h u(\bullet, t) - \boldsymbol{v}(t) = \boldsymbol{e}_{\mathrm{S}}(t) + \boldsymbol{e}_{\mathrm{g}}(t)$$

where $\boldsymbol{e}_{\mathrm{S}}(t) = r_h e_{\mathrm{s}}(\bullet, t)$. The function $\boldsymbol{e}_{\mathrm{S}}(t)$ represents the accumulation of the spatial discretization error at the mesh points, and is termed by Cullen & Morton (1980) the evolutionary error. It is convenient to introduce the polynomial function $\bar{u}(\bullet, t)$ which is defined by

$$\bar{u}(\bullet, t) = p_h r_h u(\bullet, t),$$

since this allows us to obtain a more suitable expression for the error at the mesh

points. From equation (4.1) and from the initial condition (2.4) it follows that, on applying the restriction operator r_h,

$$r_h \frac{\partial e_s}{\partial t} = f_N(r_h \bar{u}) - f_N(r_h U) + r_h S u - f_N(r_h \bar{u}), \qquad r_h e_s(\bullet, 0) = 0.$$

This can be written in vector notation, by using equation (4.11), as

$$\dot{e}_S(t) = f_N(\bar{u}(t)) - f_N(u(t)) + e_T(t), \qquad e_S(0) = 0 \tag{4.13}$$

The vector $e_T(t)$ is defined by applying the restriction operator r_h to the p.d.e. truncation error

$$e_T(x, t) = (Su)(\bullet, t) - p_h f_N(r_h \bar{u}(\bullet, t)).$$

The vectors $\dot{e}_S(t)$ and $\bar{u}(t)$ are similarly defined by

$$\dot{e}_S(t) = r_h \frac{\partial e_s}{\partial t}(\bullet, t), \qquad \bar{u}(t) = r_h \bar{u}(\bullet, t).$$

We now apply the same backward difference formula to equation (4.13) as in equation (4.12) to give

$$f_N(\bar{u}(t_{n+1})) - f_N(u(t_{n+1})) + e_T(t_{n+1}) + \left(\sum_{i=1}^{p} \alpha_i e_S(t_{n+1-i}) - e_S(t_{n+1}) + e_t^*(t_{n+1})\right) \frac{1}{h\beta} = 0$$

where $e_t^*(t_{n+1})$ is the truncation error from using the backward-difference formula in this calculation. Using equation (4.12) to eliminate the term $f_N(u(t_{n+1}))$ gives

$$f_N(v(t_{n+1}) + e(t_{n+1})) - f_N(v(t_{n+1})) + e_T(t_{n+1}) + [e_T(t_{n+1}) + e_t^*(t_{n+1})] \frac{1}{h\beta}$$
$$+ \left(\sum_{i=1}^{p} \alpha_i e(t_{n+1-i}) - e(t_{n+1})\right) \frac{1}{h\beta} = 0, \qquad e(0) = 0. \tag{4.14}$$

The only differences between equations (4.14) and (4.12) are that in (4.14) we have $e(t)$, not $e_g(t)$, and we have added two extra terms, $e_T(t_{n+1})$ and $e_t^*(t_{n+1})$, to the o.d.e. truncation-error term. Consequently the same techniques can be applied to both equations. For simplicity, we shall use the Dew–West method, although any of the methods described by Shampine (1979) could be used; it is not clear at the moment which is the most suitable. Applying the Dew–West method gives

$$(I - h\beta G) e(t_{n+1}) \approx e_e(t_{n+1}) + e_t^*(t_{n+1}) + h\beta e_T(t_{n+1}) + e(t_n). \tag{4.15}$$

The weakness of this approach is that it is a zero-order approximation to the linearized form of equation (4.13). Better approximations are possible using methods described by Shampine (1979) and these will be part of our future investigations.

4.4 *Implementation of the Error Indicator*

Equation (4.15) can be used to estimate the components of the error vector $e(t_{n+1})$, providing that we can estimate the vectors $e_t^*(t)$ and $e_T(t)$. The vector

$e_t^*(t)$ is the o.d.e. truncation error that arises integrating numerically the spatial discretization-error equation (4.13). At present, it is not clear how the term $e_t^*(t)$ can be estimated and whether or not it needs to be included to provide a reliable error indicator. In the numerical experiments described below, we have not attempted to estimate this term. The p.d.e. truncation error $e_T(t)$ can be estimated at the collocation point $x_{j,i}$ by using the pointwise error indicator derived in Section (4.1). The component of the vector $e_T(t)$ corresponding to the collocation point $x_{j,i}$ is then given, for $i = 1, \ldots, N-1$ and $j = 1, \ldots, J$, by

$$[e_T(t)]_l = E_T(x_{j,i}, t) \approx N^{\frac{1}{2}}(s_{j,N}(t)) \tag{4.16}$$

where $l = N \times (j-1) + i$ and the coefficient $s_{j,N}(t)$ is the Nth Chebyshev coefficient of $S_j(u(\bullet, t))$ at time, t, with $S_j(u(\bullet, t))$ denoting the restriction of $S(u(\bullet, t))$ to the interval I_j. The coefficient $s_{j,N}$ is used in this estimate because, at any given time t, the p.d.e. truncation error defined by equation (4.13) depends on the exact solution to the p.d.e. $(u(x, t))$ and its polynomial interpolant $(\tilde{u}(x, t))$, and not directly on the computed solution at that time. The truncation error at the break-point X_j, for $j = 1, \ldots, J-1$, is estimated by (see equation (4.9))

$$[e_T(t)]_k = E_T(X_j, t) \approx \frac{N^{\frac{1}{2}}}{h_j + h_{j+1}} [h_{j+1} s_{j+1,N}(t) + h_j s_{j,N}(t)], \tag{4.17}$$

where $k = jN$. This estimate is consistent with the form of the collocation equations in the break-point equation (4.4).

The coefficients $s_{j,i}(t)$ may be estimated by noting that, from (4.1),

$$r_h[Su)(\bullet, t)] = \dot{v}(t) + \dot{e}(t), \tag{4.18}$$

where $\dot{v}(t)$ is the time derivative produced by Gear's method but $\dot{e}(t)$ is the time derivative of the error at time t. The coefficient $s_{j,N}$ is approximated by the Nth Chebyshev coefficient of the function

$$p_h r_h[(Su)(\bullet, t)]$$

which is a polynomial of degree N on each interval. The implementation algorithm for this error indicator is described in Appendix II.

5. Numerical examples

The following two examples illustrate some of the properties of the error indicator derived above. The first example is the p.d.e. used in example (3.2.1) and the second example is defined by:

$$\frac{\partial u}{\partial t} = \begin{cases} \dfrac{1}{C_1}\dfrac{\partial^2 u}{\partial x^2} + C_1 e^{-2u} + e^{-u} & \text{if } x \in [-1, 0), \\ \dfrac{1}{C_2}\dfrac{\partial^2 u}{\partial x^2} + C_2 e^{-2u} + e^{-u} & \text{if } x \in (0, 1], \end{cases}$$

subject to the boundary conditions

$$u(-1, t) = \log_e(-C_1 + t + P),$$

$$u(1, t) + (C_2 + t + P)\frac{\partial u}{\partial x}(1, t) = \log_e(C_2 + t + P) + 1{\cdot}0,$$

for $t \geq 0$. The initial condition is consistent with the analytic solution

$$u(x, t) = \begin{cases} \log_e(C_1 x + t + P) & \text{if } x \leq 0, \\ \log_e(C_2 x + t + P) & \text{if } x > 0, \end{cases}$$

and $P = 1{\cdot}0$, $C_1 = 0{\cdot}1$, $C_2 = 1{\cdot}0$. The problem was integrated from $t = 0$ to $t = 1{\cdot}0$.

The Chebyshev C^0 collocation method was applied with two equally spaced elements, using an approximating polynomial of degree 7 in each element for Problem 1 and a polynomial of degree 9 in each element for Problem 2. The interior break point for each problem was situated at 0·5 and 0·0 respectively. Each problem was integrated in time with three different absolute local error tolerances: 10^{-3}, 10^{-7}, and 10^{-9}. The L_2 error norm was measured at ten time levels during the integration (the weighted norm was not used for Problem 1, since the Delves indicator defined by equation (4.6) does not estimate this norm directly).

In the case when the local error tolerance is 10^{-3}, the time integration error dominates and the error is estimated with the same accuracy as in the examples of Dew & West (1979). The Delves indicator is not appropriate in this case. At a local error tolerance of 10^{-7}, each indicator provides acceptable estimates of the error norm; the Delves indicator is preferred for Problem 1 and the new indicator for Problem 2, see Figs 1 and 3. At a tolerance of 10^{-10} the new indicator proves to be slightly superior to the Delves indicator.

An additional advantage of the new error indicator is that it is also possible to estimate the maximum error at the solution mesh points.

6. Summary

The Chebyshev C^0 collocation method allows a wide range of Chebyshev polynomial approximations to be applied to many p.d.e. problems in one space dimension. Two advantages of the method over that of Berzins & Dew (1981) are that an explicit o.d.e. system is solved and that, for linear basis functions, the method reduces to one based upon a second-order difference approximation.

From our practical experience, the error indicator derived above seems to be a promising means of estimating the total error in the numerical solution. Over a limited range of simple parabolic equations, we have found it to be more reliable than either of the two error indicators of Delves (1976) and Dew (1978), with the additional advantage that the o.d.e. local error tolerance no longer has to be restricted so that the time integration error is of a smaller magnitude than the spatial discretization error. This is particularly important when high-order Chebyshev polynomials are used in the spatial discretization of problems with smooth solutions.

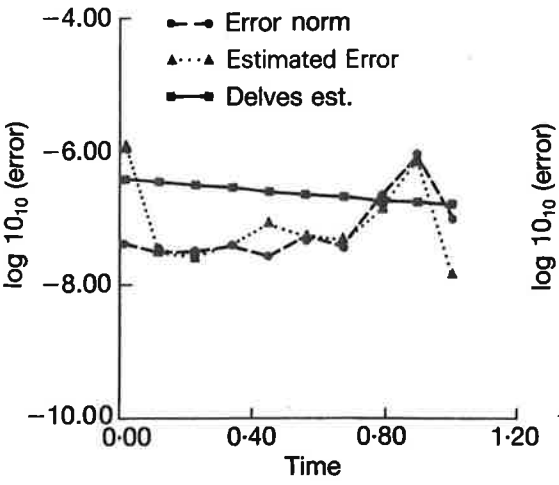

FIG. 1. Problem 1, $\varepsilon = 10^{-7}$

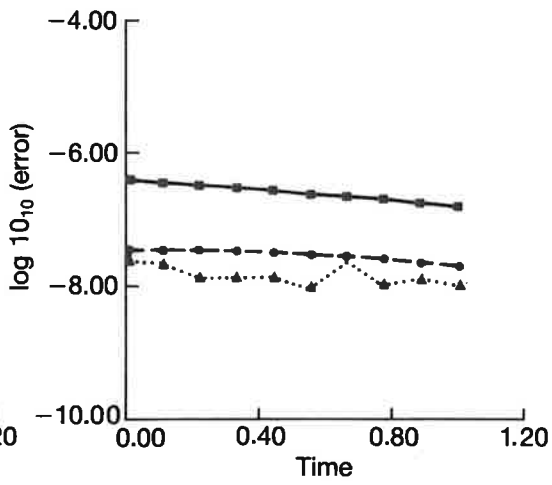

FIG. 2. Problem 1, $\varepsilon = 10^{-10}$

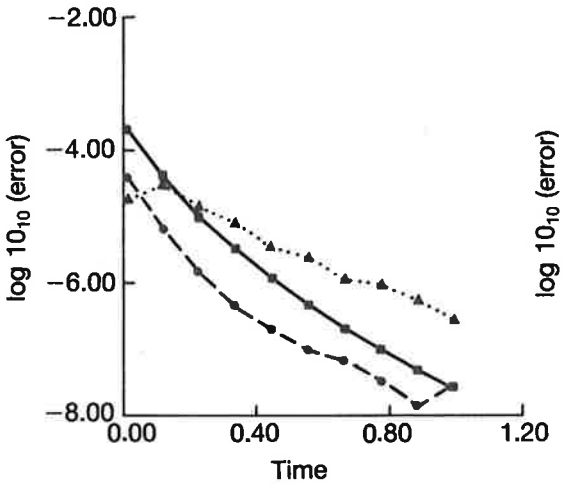

FIG. 3. Problem 2, $\varepsilon = 10^{-7}$

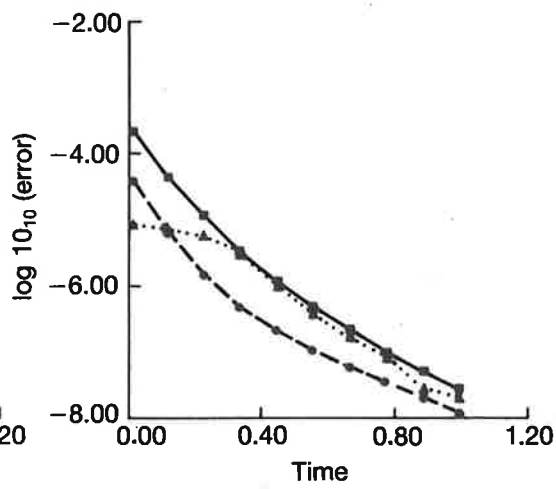

FIG. 4. Problem 2, $\varepsilon = 10^{-10}$

There are two outstanding problems with this indicator: (i) it is necessary to determine the best o.d.e. global error formula for equation (4.13) and (ii) a suitable method of estimating the p.d.e. truncation error for a wide range of problems and for different choices of spatial basis functions is required.

Acknowledgements

Thanks are due to Shell Research Limited for funding one of us (M.B.) on a two-year research project on software for parabolic equations, during which time this research was completed. The constructive comments of the referees are also acknowledged.

REFERENCES

BAKKER, M. 1977 Galerkin methods in circular and spherical regions *Report* NW 50/77, Mathematisch Centrum, Amsterdam.

BERZINS, M. 1982 Chebyshev Polynomial Methods for Parabolic Equations (Ph.D. Thesis). Department of Computer Studies, The University, Leeds LS2 9JT.

BERZINS, M., & DEW, P. M. 1981 A generalized Chebyshev method for parabolic p.d.e.s in one space variable. *I.M.A. J. numer. Anal.* **1,** 469–487.

BERZINS, M., DEW, P. M., & FURZELAND, R. M. 1985 A user's manual for SPRINT: Part two solving partial differential equations, Department of Computer Studies, The University, Leeds LS2 9JT, Report no. 202.

CULLEN, M. J. P., & MORTON, K. W. 1980 Analysis of evolutionary error in finite element and other methods. *J. comput. Physics* **34,** 245–269.

DELVES, L. M. 1976. Expansion methods. In: *Modern Numerical Methods for O.D.E.s* (G. Hall & J. M. Watt, Eds). Clarendon Press, Oxford. Pp. 269–281.

DEW, P. M. 1978 A note on the numerical solution of quasi-linear parabolic p.d.e.s with error estimates. *J.I.M.A.* **25,** 401–417.

DEW, P. M., & WALSH, J. E. 1981 A set of library routines for the numerical solution of parabolic equations in one space variable. *A.C.M. Trans. Math. Soft.* **7,** 295–314.

DEW, P. M., & WEST, M. 1979 Estimating and controlling the global error in Gear's method. *BIT* **19,** 135–137.

DIAZ, J. C. 1977 A collocation—Galerkin method for the two point boundary value problem using continuous piecewise polynomial spaces. *SIAM J. numer. Anal.* **14,** 844–858.

DUPONT, T. 1976 A unified theory of superconvergence for Galerkin methods for two-point boundary value problems. *SIAM J. numer. Anal.* **13,** 362–368.

FOX, L., & PARKER, I. B. 1968 *Chebyshev Polynomials in Numerical Analysis.* Oxford University Press.

IMHOF, J. P. 1963 On the method for numerical integration of Clenshaw and Curtis. *Numer. Math.* **5,** 138.

KELLER, H. B. 1971 A new difference scheme for parabolic problems. In: *Numerical Solution of P.D.E.s* – II, Synspade (1970) (B. Hubbard, Ed.). Academic Press.

MACHURA, M., & SWEET, R. A. 1980 A survey of software for partial differential equations. *A.C.M. Trans. math. Soft.* **6,** 461–488.

MADSEN, N., & SINCOVEC, R. F. 1978 PDECOL: general collocation software for partial differential equations. *A.C.M. Trans. math. Soft.* **5,** 326–351.

PETZOLD, L. R. 1982 A description of DASSL. *Report* SAND 82-8367, Applied Math. Division, Sandia National Laboratories, Livermore, California.

SHAMPINE, L. F. 1979 Global error estimation for stiff o.d.e.s. *Report* SAND 79-1587. Sandia Laboratories, Numerical Math. Division 5642, Albuquerque, New Mexico 87185. U.S.A.

SKEEL, R. D. 1981 Improving routines for solving parabolic equations in one space variable. *Numer. Anal. Report* 63, Department of Mathematics, University of Manchester.

WHEELER, M. F. 1977 A C^0 collocation finite element method for two-point boundary value problems and one space dimensional parabolic problems. *SIAM J. numer. Anal.* **14,** 71–90.

Appendix I: Implementation details

This appendix provides a concise description of the spatial discretization method which is derived in the first part of the paper. The method requires that a set of break points $X_0, \ldots, X_j$ are provided by the user and that these break points include any discontinuities in the p.d.e. functions c and f in equation (2.2). The spatial discretization method has been designed to reduce the p.d.e. to an

implicit system of ordinary differential equations of the form

$$\boldsymbol{F}(t, \boldsymbol{u}(t), \dot{\boldsymbol{u}}(t)) = \boldsymbol{A}\dot{\boldsymbol{u}} - \boldsymbol{G}(t, \boldsymbol{u}(t)) = 0, \qquad \boldsymbol{u}(0) = \boldsymbol{k}, \tag{A1}$$

where the matrix $\boldsymbol{A}$ may be singular. The vectors $\boldsymbol{f}$, $\boldsymbol{u}(t)$, and $\dot{\boldsymbol{u}}$ are ordered as in equation (4.9). Several o.d.e. solvers for problems of this type have recently been developed, e.g. Petzold (1982). These solvers automatically calculate an estimate of $\dot{\boldsymbol{u}}(0)$ thus removing the need for a special starting procedure when the matrix A is singular; see Berzins & Dew (1981).

AI.1 *Preliminaries*

Before defining precisely the vector $\boldsymbol{f}$, it is helpful to introduce two square matrices and three vectors of dimension $N+1$.

The matrix $\boldsymbol{\Omega}$ The square matrix $\boldsymbol{\Omega}$, of dimension $N+1$, is defined by (Berzins & Dew, 1981):

$$[\boldsymbol{\Omega}]_{i,k} = \frac{N_{ik}}{N} \mathrm{T}_i(y_k) \tag{A2}$$

where

$$y_k = \cos\frac{(N-k)\pi}{N}, \qquad N_{i,k} = \begin{cases} 2 & \text{if } (i,k) \in \{1, \ldots, N-1\}^2, \\ 1 & \text{otherwise} \end{cases}$$

$(i = 0, \ldots, N;\ k = 0, \ldots, N)$.

The matrix $\boldsymbol{D}$ The second matrix is used to estimate the values of $\partial U/\partial x$ at the break-points and collocation points in each element. We first define a matrix $\bar{\boldsymbol{D}}$ by

$$[\bar{\boldsymbol{D}}]_{i,k} = \mathrm{T}_i'(y_k),$$

the derivative of T_i at y_k. The elements of this matrix are then given by (Fox & Parker, 1968),

$$[\bar{\boldsymbol{D}}]_{i,0} = [\bar{\boldsymbol{D}}]_{i,N} = 0 \quad (i = 1, \ldots, N-1),$$

$$[\bar{\boldsymbol{D}}]_{N,i} = i^2, \quad [\bar{\boldsymbol{D}}]_{0,i} = i^2(-1)^{(i-1)} \qquad (i = 0, \ldots, N),$$

$$[\bar{\boldsymbol{D}}]_{i,k} = \left(\sin\frac{\pi k(i-N)}{N}\right) \Big/ \sin\frac{\pi(i-N)}{N} \quad (i = 1, \ldots, N-1;\ k = 1, \ldots, N-1).$$

The matrix $\boldsymbol{D}$ is then defined by $\boldsymbol{D} = \bar{\boldsymbol{D}}\boldsymbol{\Omega}$.

Temporary vectors Consider the jth element $(j = 1, \ldots, J)$. The derivative $\partial U/\partial x$ at the collocation points $x_{j,i}$ $(i = 0, \ldots, N)$ in the jth element is then given by

$$\boldsymbol{u}_j^{(x)} = \frac{2}{X_{j+1} - X_j} \boldsymbol{D}\boldsymbol{u}_j^*(t)$$

where

$$\boldsymbol{u}_j^* = \begin{cases} [U_{j,0}, \ldots, U_{j,N-1}, U_{J,N}]^T & (j = J), \\ [U_{j,0}, \ldots, U_{j,N-1}, U_{j+1,0}]^T & (j = 1, \ldots, J-1). \end{cases}$$

From the definition of the p.d.e. functions r, c, and f we can then define the vectors $\boldsymbol{q}_j$ and $\boldsymbol{r}_j$ whose components are given by

$$Q_{j,i} = c_{j,i}\dot{U}_{j,i} - f_{j,i}, \qquad R_{j,i} = r(t, x_{j,i}, U^*_{j,i}, U^{(x)}_{j,i})$$

$(i = 0, \ldots, N)$, where $c_{j,i}$ and $f_{j,i}$ are defined in the same way as $R_{j,i}$. The space derivative $\dot{\boldsymbol{r}}_j$ of the flux R at the collocation points is then given by

$$\dot{\boldsymbol{r}}_j = \frac{2}{X_{j+1} - X_j} \boldsymbol{D}\boldsymbol{r}_j.$$

The components of $\dot{\boldsymbol{r}}_j$ are denoted by $\dot{R}_{j,i}$.

AI.2 *Definition of the Vector* $\boldsymbol{f}$

Given the values of t, $\boldsymbol{u}(t)$, and $\dot{\boldsymbol{u}}$, we shall construct the vector $\boldsymbol{f}$ in equation (A1) by using the vectors and matrices defined above.

Collocation points The component of the vector $\boldsymbol{f}$ corresponding to the collocation equation at the point $x_{j,i}$ is given by (see equation (3.7))

$$F_{j,i} = \dot{R}_{j,i} - Q_{j,i} \quad (i = 1, \ldots, N-1; j = 1, \ldots, J).$$

Break points The contribution of the $(j+1)$th element to the break-point condition at X_{j+1} is given by

$$F_{j+1,L} = h_j Q_{j,N} + R_{j,0} + R_{j,N} + \sum_{i=1}^{N-1} \frac{\lambda_i}{\lambda_N} R_{j,i} + \tfrac{1}{2}(1 + y_i)\frac{\lambda_i}{\lambda_N}\dot{R}_{j,i}h_j. \tag{A3}$$

The constants y_i are defined by equation (A2) and λ_i by equation (3.12). The contribution of the $(j+2)$th element to the same break-point condition is given by

$$F_{j+1,R} = h_{j+1}Q_{j+1,0} - R_{j+1,0} - R_{j+1,N} + \sum_{i=1}^{N-1} -\frac{\lambda_i}{\lambda_N} R_{j+1,i} + \tfrac{1}{2}(1 - y_i)\frac{\lambda_i}{\lambda_N}\dot{R}_{j+1,i}h_{j+1}; \tag{A4}$$

so that, for internal break points,

$$F_{j+1,0} = (F_{j+1,L}h_j + F_{j+1,R}h_{j+1})\frac{1}{h_j + h_{j+1}} \quad (j = 2, \ldots, J-1). \tag{A5}$$

This equation (A5) may be seen to correspond to the interface condition (3.13).

Dirichlet boundary condition at $x = a$ The value of $F_{1,0}$ is given by (see equation (2.3)) $F_{1,0} = U_{1,0}$. Other Dirichlet conditions at $x = a$ are treated by defining $F_{j,0}$ as the residual of an equation involving $U_{1,0}$. The same approach can be applied to Dirichlet conditions at $x = b$.

Neumann boundary condition at $x = b$ In this case $F_{J,N}$ has the value of $F_{J+1,L}$ in equation (A3) and the boundary condition (see equation (2.3)) is treated by replacing $R_{J,N}$ in the equation

$$F_{J,N} = h_J Q_{J,N} + R_{J,0} + R_{J,N} + \sum_{i=1}^{N-1} \frac{\lambda_i}{\lambda_N} R_{J,i} + \tfrac{1}{2}(1 + y_i) \frac{\lambda_i}{\lambda_N} \dot{R}_{J,i} h_J.$$

by using $R_{J,N} = g(t, U_{J,N})$. A Neumann boundary condition at $x = a$ could be treated in exactly the same way except that, in equation (A4) with $j = 0$, the flux $R_{1,0}$ would be replaced by its value according to the boundary condition, and $F_{1,0}$ would be given the value of $F_{1,R}$ from equation (A4).

Appendix II: Implementation algorithm for the error indicator

The estimation of $e_T(t_{n+1})$ involves the vector $e(t_{n+1})$ (see equation (4.18)) and so equation (4.15) cannot be used directly. Assume that the combined error $e(t_n)$ at time t_n is known, and that the Gear's method has calculated $v(t_{n+1})$ and $\dot{v}(t_{n+1})$. The following algorithm can then be used to calculate $e(t_{n+1})$.

Algorithm—Error Indicator

(i) Estimate $\dot{e}(t_{n+1})$ by

$$\dot{e}(t_{n+1}) \approx [e(t_n) - e(t_{n-1})]/(t_n - t_{n-1}).$$

(ii) Calculate the Chebyshev coefficients $s_{j,N}(t_{n+1})$ from the polynomial

$$p_h r_h[(Su)(\cdot, t_{n+1})] = p_h[\dot{v}(t_{n+1}) + \dot{e}(t_{n+1})].$$

(iii) Use the coefficients $s_{j,N}(t)$ to provide an estimate of the p.d.e. truncation error $e_T(t_{n+1})$ by using the estimates of equations (4.16) and (4.17).

(iv) Use equation (4.15) to calculate $e(t_{n+1})$ by a back-substitution using the factored form of the matrix $I - h\beta G$ that is stored by Gear's method.

$$(I - h\beta G)e(t_{n+1}) \approx e_e(t_{n+1}) + h\beta e_T(t_{n+1}) + e(t_n).$$

(v) Estimate the time derivative of the error at time t_{n+1} by

$$\dot{e}(t_{n+1}) = [e(t_{n+1}) - e(t_n)]/(t_{n+1} - t_n).$$

(vi) Repeat steps (ii) to (v) until the iteration for $e(t_{n+1})$ has converged. Two iterations have proved to be sufficient in the experiments that we have conducted.

End of Algorithm

Correction for A Note on C^0 - Chebyshev Methods for Parabolic P.D.E.s .

There is an error in this paper concerning the simplified form of equation (3.13). Although the actual flux $r(x,t)$ is continuous at the break-points the numerically computed flux $R(x\ ,\ t)$ need not be so . The paragraph below equation (3.13) must be changed as follows to allow for the possible jump in the flux at the break-point.

In the case when $N>1$ equation (3.13) can be further simplified at the interior breakpoints by using integration by parts on $\left(\frac{\partial R}{\partial x}\ ,\ v_j \right)$ and noting that $v_j(x)$ is zero at $x = a$ and $x = b$ to give :

$$\left(\frac{\partial R}{\partial x}\ ,\ v_j \right) + \left(R\ ,\ \frac{dv_j}{dx} \right) = R(x_{j,N},t) - R(x_{j+1,0}\ ,\ t) \quad j = 1, \ldots, N-1.$$

Applying the Clenshaw-Curtis rule to exactly evaluate these inner products gives

$$\sum_{i=0}^{N} \frac{\lambda_i}{2} [R(x_{j+1,i},t) - R(x_{j,i},t) - h_{j+1} \frac{\partial R}{\partial x}(x_{j+1,i},t)\, v_j(x_{j+1,i}) - h_j \frac{\partial R}{\partial x}(x_{j,i},t)\, v_j(x_{j,i})] = R(x_{j+1,0},t) - R(x_{j,N}\ ,\ t)$$
$$j = 1,2, \ldots, J-1 \qquad (3.14)$$

and on substitution into equation (3.13) we obtain the simpler expression

$$h_j\ Q(x_{j,N}\ ,\ t) + h_{j+1}\ Q(x_{j+1,0}\ ,\ t) = h_{j+1} \frac{\partial R}{\partial x}(x_{j+1,0},t) + h_j \frac{\partial R}{\partial x}(x_{j,N},t) + 2 [R(x_{j+1,0},t) - R(x_{j,N}\ ,\ t)] (\lambda_N)^{-1} \quad j = 1,2, \ldots, J-1\ . \qquad (3.15)$$

In the case when $N = 1$ the function R may depend only on $\frac{\partial U}{\partial x}$ and so be piecewise constant . The identity (3.14) then only holds trivially and so cannot be used to simplify equation (3.13). The method is then defined by the interface equations above and the collocation equations.

The boundary condition in Berzins and Dew can be treated in much the same way to get

$$h_J\ \beta(b\ ,\ t) [\ Q(b\ ,\ t) - \frac{\partial R}{\partial x}(b,t)\ ] = -2 [\ g(b,t) - \beta(b,t)\ R(b\ ,\ t)\ ] (\lambda_N)^{-1} \qquad (b)$$

A flux boundary condition at $x = a$ would similarly give;

$$h_1 \beta(a\ ,\ t)\ [\ Q(a\ ,\ t)\ -\ \frac{\partial R}{\partial x}(a,t)\ ]\ =\ 2\ [\ g(a,t) - \beta(a,t)\ R(a\ ,\ t)\ ]\ (\lambda_N)^{-1} \qquad \text{(a)}$$

The method is then defined by the collocation equations

$$\frac{\partial R}{\partial x}\ (\ x_{j,i}\ ,\ t)\ -\ Q(\ x_{j,i}\ ,\ t)\ =\ 0$$

$$j\ =\ 1,2,\ .\ .\ .\ ,J\ \ ;\ i\ =\ 1,2,\ .\ .\ .\ ,N-1 \qquad (3.7)$$

where the points $x_{j,i}$ are defined by equation (3.2) of Berzins and Dew, the interface conditions (3.15) and the boundary conditions (a) and (b).

The following theorem is a simple consequence of the fact that both $Q(x,t)$ and $R(x\ ,\ t)$ are polynomials of degree N on each sub-interval which satisfy the boundary , interface and collocation equations given above.

Theorem

The function $U(x,t)\ \varepsilon\ \bar{M}(N\ ,\ \delta)$ which satisfies the collocation boundary and interface conditions above also satisfies the peturbed differential equation

$$\frac{\partial R}{\partial x}\ +\ P_j\ =\ Q(x\ ,\ t)\ ,$$

$$(x\ ,\ t)\ \varepsilon\ (\ X_{j-1}\ ,\ X_j)\ \times\ (\ 0\ ,\ T_e]\ ,\quad i\ =\ 1,2,\ .\ .\ .\ ,J$$

where from Berzins (1982)

$$P_j(x\ ,\ t) =\ \frac{dT_N(y)}{dy}\ (\ \frac{q_{j,N-1}(t)}{2N}\ +\ \frac{q_{j,N}(t)}{N}\ y\ -\ r_{j,N}(t)\ \frac{2}{h_j}\ )$$

$$y\ =\ W_j(x)\ ,\quad x\ \varepsilon\ I_j\ ,\quad y\ \varepsilon\ [-1\ ,\ 1]\ ,$$

and from the simplified form of the interface conditions we see that

$$h_j\ P_j(x_{j,N}\ ,\ t)\ +\ h_{j+1}\ P_{j+1}(x_{j+1,0}\ ,\ t)\ \ =\ \ 2\ [\ R(x_{j+1,0},t)\ -\ R(x_{j,N}\ ,\ t)\ ]\ (\lambda_N)^{-1}$$

$$j\ =\ 1,2,\ .\ .\ .\ ,J-1\ .$$

while from the boundary conditions (a) and (b) we see that

$$h_J\ \beta(b\ ,\ t)\ P_J(b\ ,\ t)\ \ =\ -2\ [\ g(b,t) - \beta(b,t)\ R(b\ ,\ t)\ ]\ (\lambda_N)^{-1}$$

$$h_1\ \beta(a\ ,\ t)\ P_1(a\ ,\ t)\ \ =\ 2\ [\ g(a,t) - \beta(a,t)\ R(a\ ,\ t)\ ]\ (\lambda_N)^{-1}$$

Similar Numerical Methods

The simplified description given above shows that the Chebyshev C^0 collocation method of Berzins and Dew (1987) , as outlined as the generalized collocation method by Berzins (1982), is related to the collocation-like method introduced by Leyk(1986). The main difference is that Leyk uses the transformed Legendre points given by

$$W_j(x_{j,i}) = i \text{ th zero of } \frac{d\ L_N(y)}{dy} \quad , \quad y \varepsilon (-1\ ,\ 1)$$

where $W_j(x)$ is the linear map defined by Berzins and Dew (1987) and $L_N(y)$ is the Legendre polynomial of degree N on $(-1\ ,\ 1)$.

The quadrature rule used by Leyk instead of the Curtis Clenshaw rule is the Gauss Lobatto formula for the Legendre weight $w(x) = 1$. Leyk proves optimal rates of convergence for his method and superconvergence at the break-points .

References.

Berzins M. (1982) Chebyshev Polynomial Methods for Parabolic Equations. (Ph.D. Thesis) , Department of Computer Studies , The University, Leeds LS2 9JT.

Berzins M. and Dew P.M. (1987) . A Note on C^0 Chebyshev Methods for Parabolic P.D.E.s , Vol 7 , I.M.A. Journal of Numerical Analysis pp 15-37.

Leyk Z. (1986) A C^0 Collocation-like Method for Two Point Boundary Value Problems . Numer Math , Vol 49 , pages 39 - 53.