Chemical Engineering Science, 1975. Vol. 30, pp. 587–596. Pergamon Press. Printed in Great Britain

ORTHOGONAL COLLOCATION ON FINITE ELEMENTS

G. F. CAREY† and BRUCE A. FINLAYSON
Department of Chemical Engineering, University of Washington, Seattle, WA 98195, U.S.A.

(*Received* 3 *July* 1974; *accepted* 1 *November* 1974)

Abstract—The effectiveness factor problem for heat and mass transfer with chemical reaction in a catalyst pellet is solved with a new technique especially suited to situations corresponding to high Thiele modulus when the solution is confined to a thin boundary region near the catalyst surface. The method of orthogonal collocation on finite elements combines the rapid convergence of the orthogonal collocation method with the convenience associated with finite difference methods of locating grid points or elements where the solution is important or has large gradients. The efficiency of the method results from block $\overline{LU}$ decompositions employed in the iterative schemes devised. The method is applied to two problems to illustrate the rate of convergence, the efficiency (as expressed by error versus computation time curves), and the use of the residual for optimum location of the finite elements. Comparisons are also made to usual orthogonal collocation and finite difference methods.

INTRODUCTION

Consider the diffusion of heat and mass into a catalyst pellet where chemical reaction takes place. The mathematical problem describing this situation is a two-point nonlinear boundary value problem which has interested chemical engineers for two reasons: (1) models of chemical reactors (especially transient models) may require the solution of the problem hundreds or thousands of times, and (2) the problem provides a testing ground for new methods, since the character of the solution can be changed completely by changes in the Thiele modulus and the nonlinearity can be affected dramatically by the choice of parameters in the reaction rate expression. In an effectiveness factor problem in which diffusion is very fast (small Thiele modulus), the concentration of each species is everywhere close to its boundary value. As the Thiele modulus increases, the concentration profiles develop a boundary layer near the pellet surface, and the concentration in the interior of the pellet is at chemical equilibrium.

The orthogonal collocation method is useful for solving effectiveness factor problems [1–3], even for quite large Thiele modulii, far into the range of validity of the asymptotic solution. For example, for isothermal pellets with an effectiveness factor greater than 0·2, a two-term orthogonal collocation solution ($N = 2$) gave 1% accuracy in predicting the effectiveness factor [3]. Indeed, the high accuracy is one of the prime advantages of the orthogonal collocation method. Ferguson and Finlayson [4] showed in one example that the error in the solution decreased in proportion to $(1/N)^{1.72N}$, where N was the number of interior collocation (or grid) points. As N changes from 5 to 6 the error decreases by a factor of over 100.

For large Thiele moduli, however, when the solution has a steep gradient near the pellet surface, the orthogonal collocation method becomes unwieldy because a large number of collocation points is needed in order to have any at all in the boundary layer. For problems of this type finite difference methods are possible because a large number of grid points can be used, and calculations can be done efficiently due to the tri-diagonal matrix which results. It is for problems of this type that we develop the method of orthogonal collocation on finite elements in an effort to combine the small truncation error associated with the orthogonal collocation method with the ability of the finite difference method to locate grid (or collocation) points where needed.

In orthogonal collocation on finite elements we divide the domain into small subdomains, which we call finite elements. For the problems discussed below this is just dividing the line $0 \leq x \leq 1$ into elements of width Δx. We apply orthogonal collocation within elements and require that the function and its first derivative be continuous at the boundaries between elements. A first step in this direction was made by Paterson and Cresswell [5]. They postulated an effective reaction zone, the boundary layer, in which all the reaction took place and solved an approximated problem in the layer by low-order orthogonal collocation. The layer width of this "outer element" entered the problem description as a function of the expansion coefficients. The remainder of the domain, the "inner element" does not enter their formulation and the concentration was identically zero there. For effectiveness factor problems with multiple solutions they achieved agreement within a few per cent of the exact solution for the full range of Thiele moduli. Orthogonal collocation on finite elements extends this concept to include the "inner element" and generalizes to include several subdomains or elements and an efficient method of solution of the resulting algebraic equations. Furthermore the residual is used to guide the choice of the location of elements, which can be clustered in regions where the solution has steep gradients.

The method may also be viewed as an extension to two-point boundary value problems of the application of orthogonal collocation to integrate ordinary differential equations as initial-value problems [6]. In that application the time domain is divided into steps Δt, and orthogonal

†Department of Aeronautics and Astronautics.

collocation is applied to integrate from t_k to $t_{k-1} = t_k + \Delta t$, and continued stepwise to t_f. In an initial value problem the solution is obtained for t from 0 to t_f, but the nonlinear algebraic equations involve terms only in one of the elements Δt, for $(k-1)\,\Delta t \le t \le k\,\Delta t$, and not for larger t. For a two-point boundary value problem, such as the effectiveness factor problem, x, the spatial position, takes the place of time, t, but the solution at $x = 1$ does influence the solution at smaller x. Thus the algebraic equations for an element Δx are coupled with those of all the other elements.

The idea of piecewise polynomial trial functions has been used before[7], and especially by Varga and associates in the context of two-point boundary-value problems[8]. Usually these authors use a variational or Galerkin method in place of the collocation method. If the problem is nonlinear, this may lead to the necessity to use quadrature formulas to evaluate integrals, and if the quadrature must be evaluated at each iteration, or time step, as is often the case, lengthy computations result. This difficulty is eliminated by a collocation method since no quadratures are involved. We remark that it is much simpler to develop the computer codes if the equations are written in terms of the solution at the collocation points rather than the coefficients in the trial function. Thus all three methods discussed below—orthogonal collocation, orthogonal collocation on finite elements, and finite difference—are programmed in terms of the solution at the collocation or grid points.

Below we describe the method of orthogonal collocation on finite elements and compare its central features to other methods. An efficient algorithm is developed to solve effectiveness factor problems and the same method can be used for transient or multi-dimensional problems. Results are given for numerical experimentation on two problems involving combined heat and mass transfer with reaction in a catalyst. For one problem, for which we can derive error bounds on the solution, we examine the truncation error and curves of error versus computation time for three methods—orthogonal collocation, orthogonal collocation on finite elements and finite difference. For the other problem, which has a steep gradient of concentration near the boundary of the catalyst, we apply orthogonal collocation on finite elements and use the residual to place new elements where they are needed to achieve the best solution.

DESCRIPTION OF METHOD

Consider the diffusion and reaction of a species in an isothermal catalyst pellet, which is governed by the equations

$$\frac{d^2c}{dx^2} + \frac{a-1}{x}\frac{dc}{dx} = f(c), \quad 0 < x < 1 \tag{1}$$

$$\frac{dc}{dx}(0) = 0, \quad -\frac{dc}{dx}(1) = Bi_m[c(1) - 1] \tag{2}$$

where $a = 1, 2, 3$ for planar, cylindrical and spherical geometry. We divide the domain $0 \le x \le 1$ into NE elements by placing the dividing points at x_l, $l = 1, \ldots, NE+1$, with $x_1 = 0{\cdot}0$ and $x_{NE-1} = 1{\cdot}0$ as shown in Fig. 1

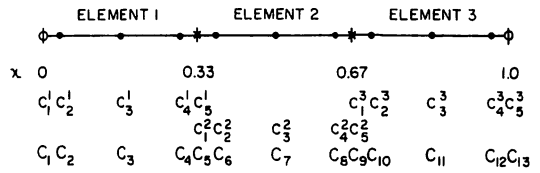

Fig. 1. Location of collocation points: ○, bounding points; x, continuity between elements; ●, interior collocation points within each element.

for NE = 3. Within each element we define a new variable $u^l = (x - x_l)/\Delta x_l$, $\Delta x_l = x_{l+1} - x_l$ and place interior collocation points at the roots to $P_n(u) = 0$, where P_N is a shifted Legendre polynomial defined on $0 \le u \le 1$. For $N = 3$ the collocation points are shown in Fig. 1. Within the lth element the variable u^l goes from zero to one. Applying the usual procedures of orthogonal collocation, Finlayson[2], we write the differential equation at the interior collocation points in terms of the value of the solution at the collocation points in the same element. For the jth interior collocation point on the lth element we have

$$\frac{1}{\Delta x_l^2}\sum_{i=1}^{N+2} B_{ji}\, c_i^l + \frac{a-1}{x_l + u_j^l \Delta x_l}\frac{1}{\Delta x_l}\sum_{i=1}^{N+2} A_{ji}\, c_i^l = f(c_j^l) \tag{3}$$

$$l = 1, \ldots, NE;\ j = 2, \ldots, N+1$$

where we have used $c_i^l = c(u_i^l) = c(x_l + u_i^l \Delta x_l)$. The matrices $\bar{B}$ and $\bar{A}$ approximate the second and first derivatives and are given in detail in Table 5.5 of Finlayson[2]. To satisfy the boundary conditions we have

$$\frac{1}{\Delta x_1}\sum_{i=1}^{N+2} A_{1i}\, c_i^1 = 0 \tag{4}$$

$$\frac{1}{\Delta x_{NE}}\sum_{i=1}^{N+2} A_{N+2,i}\, c_i^{NE} + Bi_m\, c_{N+2}^{NE} = Bi_m. \tag{5}$$

At the division between elements we do not collocate but require continuity of the function and its first derivative.

$$c_{N+2}^l \equiv c_1^{l+1} \tag{6}$$

$$\left.\frac{1}{\Delta x_l}\sum_{i=1}^{N+2} A_{N+2,i}\, c_i^l - \frac{1}{\Delta x_{l+1}}\sum_{i=1}^{N+2} A_{1i}\, c_i^{l+1} = 0\right\}$$

$$l = 1, \ldots, NE-1 \tag{7}$$

Detailed equations

Collecting Eqs. (3)–(7) determines the nonlinear system of equations which may be written in the form

$$\bar{M}\bar{C} = \bar{F}(\bar{C}) \tag{8}$$

where the components of $\bar{C}$ are the unknown solution values at the interior collocation points and the element end points,

$$C((N+1)(l-1)+i) = c_i^l$$

$$l = 1, \ldots, NE,\ i = 1, \ldots, N+1;\ i = N+2 \text{ for } l = NE.$$

This expresses the matrix c_i^l as a vector $\bar{C}$ as indicated in

Fig. 1 and insures that Eq. (6) is satisfied. The matrix $\bar{\bar{M}}$ has the block diagonal structure shown in Fig. 2, with overlap only in one entry between adjacent blocks, and the row vector $\bar{F}$ has the structure shown.

For computation, the diagonal blocks of $\bar{\bar{M}}$ are stored in a three-dimensional array $\bar{S}(j, i, l)$, $j, i = 1, \ldots, N+2$; $l = 1, \ldots, NE$. The elements of $\bar{S}$ and $\bar{F}$ are given by the following relations:

$$l = 1, j = 1 \text{ (boundary condition)}$$

$$S(1, i, 1) = A_{1i} \quad i = 1, \ldots, N+2$$

$$F(1) = 0$$

$$l = 1 \rightarrow NE - 1, j = N+2 \text{ (continuity of derivative)}$$

$$S(N+2, i, l) = \begin{cases} A_{N+2,i} & i = 1, \ldots, N+1 \\ A_{N+2,i} - \dfrac{\Delta x_l}{\Delta x_{l+1}} A_{1i} & i = N+2 \end{cases}$$

$$l = 2 \rightarrow NE, j = 1 \text{ (continuity of derivative)}$$

$$S(1, i, l) = \begin{cases} S(N+2, N+2, l-1) & i = 1 \\ -\dfrac{\Delta x_{l-1}}{\Delta x_l} A_{1i} & i = 2, \ldots, N+2 \end{cases}$$

$$F((N+1)(l-1)+1) = 0$$

$$l = NE, j = N+2 \text{ (boundary condition)}$$

$$S(N+2, i, NE) = A_{N+2,i} + Bi_m \Delta x_{NE} \delta_{N+2,i}$$

$$F((N+1)NE+1) = Bi_m \Delta x_{NE} \qquad i = 1, \ldots, N+2$$

$$l = 1 \rightarrow NE, j = 2, \ldots, N+1 \text{ (residual)}$$

$$S(j, i, l) = B_{ji} + \frac{a-1}{x_l + u_j \Delta x_l} \Delta x_l A_{ji} \quad i = 1, \ldots, N+2$$

$$F((N+1)(l-1)+j) = \Delta x_l^2 f(C((N+1)(l-1)+j))$$

δ_{ji} is the Kronecker delta. Once these equations are set up we can deal entirely with the system[8].

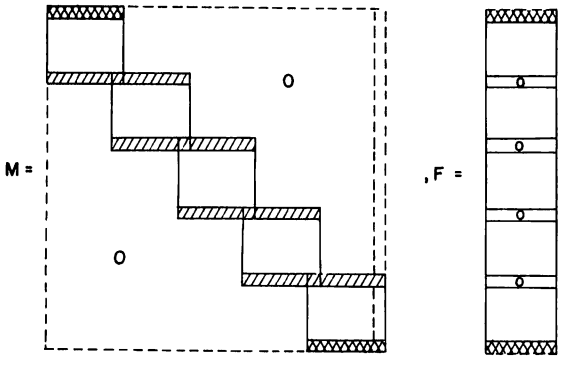

Fig. 2. Matrix structure for orthogonal collocation on finite elements. Cross-hatched areas, equations arising from boundary conditions; hatched areas, equations arising from continuity of first derivatives at boundaries of elements; clear areas, equations arising from residuals at interior collocation points of each element. The single column on the right hand side of $\bar{\bar{M}}$ arises when the Newton iteration is used for boundary-layer reactions (then $\bar{\bar{M}} \equiv \bar{\bar{J}}$, Jacobian) since the residual at each point depends on $c(x = 1)$.

Iterative solution of equations

The algebraic system of Eq. (8) is solved by iteration using a Picard or successive substitution iterative method.

$$\bar{\bar{M}} \bar{C}^{k+1} = \bar{F}(\bar{C}^k). \tag{9}$$

The matrix $\bar{\bar{M}}$ is inverted using an $\bar{\bar{L}}\bar{\bar{U}}$ decomposition[9] applied in turn to each block shown in Fig. 1. Since the matrix does not depend on the solution, the inversion can be done once and for all. Then each iteration requires only inexpensive forward and backward substitution sweeps. There is no fill outside the blocks during decomposition and the triangular matrices $\bar{\bar{L}}$ and $\bar{\bar{U}}$ are stored directly over $\bar{\bar{M}}$ during their calculation. Thus the storage requirements for the array are $NE(N+2)^2$ words, whereas a full system of this size would require $(NE(N+1)+1)^2$ words. Then the important features of this implementation are its low storage requirements and its computational speed, even for calculations involving many elements.

Standard iteration theory[10] can be used to show that the method converges provided

$$P = (\Delta x_{\max})^2 \left\| \bar{\bar{M}}^{-1} \right\| \left| \frac{\partial f}{\partial c} \right|_{\max} < 1. \tag{11}$$

Since the norm of the matrix inverse $\|\bar{\bar{M}}^{-1}\|$ is bounded for any Δx, if $\partial f/\partial c$ has a maximum (as is the case for the examples treated below) we can make $P < 1$ by choosing Δx small enough. For the second problem treated below, however, $\partial f/\partial c = 1{\cdot}1 \times 10^5$. Such a value would have required a prohibitively small Δx, so for that case the Newton Raphson method was used, in the form

$$\bar{\bar{J}}(\bar{C}^{k+1} - \bar{C}^k) = -\bar{G}(\bar{C}^k). \tag{12}$$

The block diagonal algebraic system produces a block diagonal Jacobian $\bar{\bar{J}}$ which requires decomposition within each iteration. A further complication in Problem II below is that the reaction rate at x depends on $c(x)$ and $c(1)$. The Jacobian structure is augmented by an entire vector in the last column of Fig. 2, corresponding to terms involving $\partial f/\partial c(1)$. During the decomposition of $\bar{\bar{M}}$, dual operations are performed on the final column in forming $\bar{\bar{U}}$, and this column vector is used in the back substitution. Except for the additional Jacobian calculation and $\bar{\bar{L}}\bar{\bar{U}}$ decomposition at each iteration, a similar efficiency in storage and computation is achieved when using the Newton-Raphson method.

Use of residual

The collocation method, using any trial function, is one of the methods of weighted residuals[2]. The residual is Eq. (1) with the approximate solution substituted into it. Of course the residual is zero at the collocation points, but it is generally nonzero at other positions, $0 \le x \le 1$. The residual can be evaluated after an approximate solution has been found, and this information will give valuable insight into the optimum location of elements for improving the solution.

Within each element, suppose we have found c_i^l,

$i = 1, \ldots, N+2$. We can write the trial function within the l th element as

$$c^l(u) = \sum_{i=1}^{N-2} d_i^l u^{i-1} \tag{12}$$

and evaluate the expression at the collocation points.

$$c_j^l = c^l(u_j) = \sum_{i=1}^{N-2} Q_{ji} d_i^l, \; Q_{ji} = u_j^{i-1}, \; j = 1, \ldots, N+2$$

Since $\bar{d}^l = \bar{Q}^{-1}\,\bar{c}^l$, knowing the solution at the collocation points, c_j^l, gives the vector d_i^l, and hence the solution $c^l(u)$ throughout the element. This method of interpolation suffices for low N(8–10 on a machine using 15 digits in single precision) and more refined techniques[11] are suitable for large N (up to 80 have been used). Once the function $c^l(u)$ is known within the element, derivatives of c^l can be found, and thus the residual can be defined throughout the element, and thereby throughout the domain. If we call $R_l(u)$ the residual in the l th element, then one indication of the accuracy of the solution is the mean-squared residual.

$$RS = \left[\sum_{l=1}^{NE} \Delta x_l \int_0^1 R_l^2(u)\,(u\Delta x_l + x_l)^{a-1}\,du\right]^{1/2}.$$

Ferguson and Finlayson[4] have shown for some problems that the point-wise error in the solution is bounded by the mean-squared residual. They applied the theorem to orthogonal collocation solutions, and the theorem holds provided $R(x)$ is piecewise continuous, as it is in the method of orthogonal collocation on finite elements.

For second-order finite difference calculations the same idea can be used. By taking any three adjacent grid points a quadratic polynomial can be defined to pass through the solution at the grid points. The quadratic polynomial then permits definition of the residual between the grid points in a manner which is consistent with the difference formula at the center grid point. Thus the mean-squared residual can be calculated for a finite difference solution, and an error bound thereby derived.

Location of elements

Since we want the residual to be small everywhere, after a given calculation we can examine the residual everywhere to determine where it is large and insert additional elements there. In this way as the number of elements is increased the solution should converge faster than for a uniform spacing of the elements because the elements are placed where needed. This is done in Problem II below. There the criterion was based on the residual integrated over each element. However, the largest residual usually occurred at the end point of the element (see Fig. 10 below) since a continuity condition is imposed there rather than setting a residual to zero. Thus in other calculations it would probably suffice to calculate the residual at the end points of the element and use that residual to determine the location of additional elements. This is a more convenient method, since no interpolation using Eq. (12) and no integrations are involved. The residual is evaluated as in Eq. (3) for $j = 1$ or $N+2$.

An alternative to this means for locating elements is a modification of the procedure developed by Pearson[12] for finite difference calculations. After a calculation with a given grid placement, consider the set of concentrations at the end points of each element,

$$\{C(m)\} = \{C(i) \mid i = (N+1)m + 1, m = 1, \ldots, NE - 1\}.$$

The value of $|C(m) - C(m+1)|$ is compared to

$$\delta^* = 0{\cdot}01\,[\max_m C(m) - \min_m C(m)].$$

If $|C(m) - C(m+1)| > \delta^*$, additional elements are inserted between x_m and $x_m + \Delta x_m$. The number of elements inserted is $|C(m) - C(m+1)|NP/\delta^*$, rounded to an integer value. Now let $\{x_m\}$ denote the new set of positions between the elements. The location of x_m are smoothed (to avoid abrupt changes in Δx_m in successive elements) by using the algorithm

$$x'_m = \frac{1}{2}(x'_{m-1} + x_{m+1})$$

in turn, beginning with $m = 1, 2, \ldots$, where $\{x'_m\}$ are the final locations of positions between elements.

Truncation error

Douglas and DuPont[13] have studied a collocation method on finite elements for parabolic differential equations. They showed that the spatial discretization error was proportional to h^4, where $h = 1/NE$, when cubic functions were used in each element ($N = 2$) and the collocation points were the Gaussian quadrature points, $u = 0{\cdot}21132\ldots, 0{\cdot}78867\ldots$. These are, of course, the collocation points for orthogonal collocation using $N = 2$ (cubic polynomial). They also mention that the error is proportional to h^2 if the collocation points are distributed uniformly on the element. Thus the seemingly trivial change of using certain collocation points dramatically reduces the truncation error. These results apply to the global error in the solution, i.e. at any point in the domain, not just the collocation points.

DeBoor and Swartz[14] have proven a far more reaching result. They consider higher degree polynomials, and use as collocation points in the element the zeroes of Legendre polynomials (this is what is done in orthogonal collocation). We specialize DeBoor and Swartz's results here for the case of a second order equation and let N be the number of interior collocation points in each element so that the degree of the polynomial on an element is $N+1$. DeBoor and Swartz prove that the global truncation error is proportional to h^{N+2}, where the global error is the error at any point in the domain. Their next result is more important: the error in the function and its first derivative at the end of each element is proportional to h^{2N}. For $N = 2$ the global error agrees with Douglas and DuPont and the global error converges with the same rate as the error at an element end-point. For larger N we

obtain an improved truncation error in the function and its derivative at the element end-point. For N as low as 3, 4 and 5 we obtain $0(h^6, h^8, h^{10}$, etc.). The work of DeBoor and Swartz also proves the method converges since the error can be made as small as desired by adding more elements. The proofs apply to equations of the form $D^2y = f(y)$, so they are not strictly applicable to the problems solved below (for spherical geometry), but the numerical results confirm the predictions nevertheless. Douglas[15] proves similar results for linear, time dependent problems.

NUMERICAL EXPERIENCE

We apply the method of orthogonal collocation on finite elements to two problems, one for a small Thiele modulus which gives a smoothly varying, almost parabolic, solution and the other for a highly non-isothermal problem with a large Thiele modulus which leads to a boundary layer solution, with the solution varying from nearly zero at $x = 0{\cdot}997$ to its boundary value at $x = 1{\cdot}0$.

Problem I

We solve Eqs. (1–2) with $a = 3$ (spherical domain), $Bi_m \to \infty$ so that the boundary condition is $c(1) = 1$, and with

$$f(c) = \phi^2 c \exp[\gamma(1 - 1/T)]$$

$$T = 1 + \beta - \beta c$$

corresponding to an irreversible, first order, non-isothermal reaction in a spherical catalyst pellet. Parameter values are $\phi = 0{\cdot}5$, $\gamma = 18$, $\beta = 0{\cdot}3$, which lead to a unique solution with effectiveness factor $\eta = 1{\cdot}086$. The concentration profile is shown in Fig. 3, although the problem was actually solved in terms of the temperature variable. The initial guess for all iterations was $T(x) = 1$. This problem was solved using orthogonal collocation by Ferguson and Finlayson[4] because error bounds could be derived for it. The theorems proved by Ferguson and Finlayson give

$$\text{error in } T(x) \le 0{\cdot}69\, RS$$

$$E_F \equiv \text{error in } \frac{dT}{dx}(1) \le 0{\cdot}69\, RS.$$

The fact that the numerical values are the same is only coincidence. Thus we can solve the problem using any method, compute the mean-squared residual RS, and then calculate the maximum error in the solution, even though the exact solution is now known.

The problem was solved using orthogonal collocation and double precision arithmetic on a CDC 6400 (thus retaining 30 digits in the calculations). Double precision was required to guarantee the numerical results truly represent the high accuracy of the method. The most important feature of the solution is the effectiveness factor, or flux at the catalyst boundary which is determined by dT/dx at $x = 1$. The error bounds for this quantity are shown in Fig. 4. The most accurate solution, for $N = 8$, is proved to have an error less than 10^{-14} and gives at $x = 1$ the solution $dT/dx = -0{\cdot}02716089570333$. Actually the solution for $N = 6$ and $N = 8$ were the same to at least $0(10^{-17})$, the maximum number of digits printed out. If we use the solution for $N = 8$, proved accurate to 10^{-14}, as the exact solution, we can determine the actual error for $N = 1 \to 5$, which is plotted in Fig. 4. The error bound is very conservative, being up to 10^6 times larger than the actual error for very accurate solutions. The actual error approaches a straight line as N increases, showing that the error is proportional to $(1/N)^{3{\cdot}75N}$. Ferguson and Finlayson[4] report the error as propor-

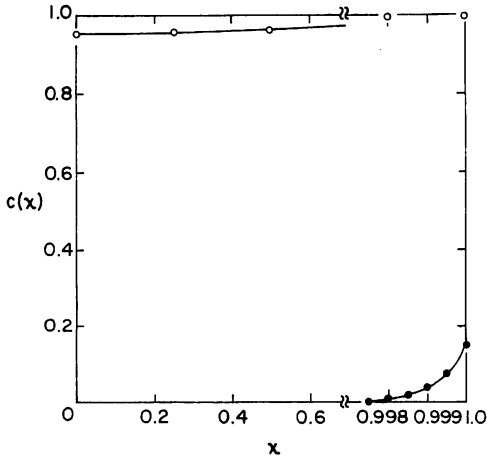

Fig. 3. Concentration profiles in catalyst pellets. ○, Problem 1; ●, Problem II.

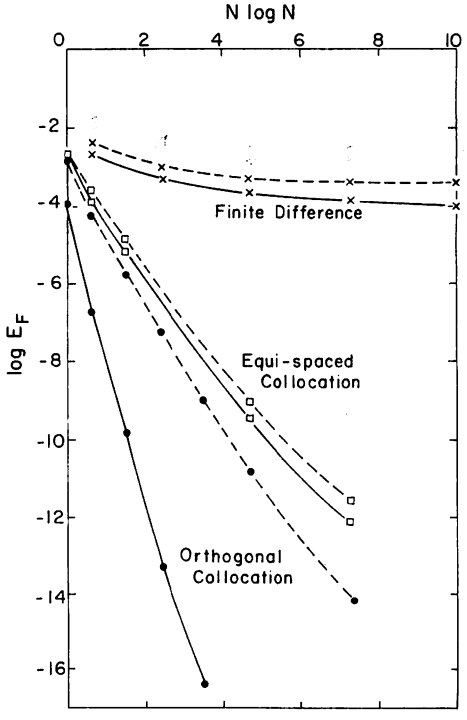

Fig. 4. Error in boundary flux for Problem I. ●, Orthogonal collocation; □, equi-spaced collocation; x, finite difference; ——, actual error; ---, error bound.

tional to $(1/N)^{1.72N}$ due to an incorrect choice of collocation points for a cylindrical instead of spherical domain. Figure 4 also illustrates the importance of the location of collocation points. If equi-distant collocation points are used the error is considerably larger, up to 10^7 times as big for the most accurate solutions. All the collocation results show very rapid convergence in N, which is characteristic of the method.

Also shown in Fig. 4 are finite difference results. The solution was found using central difference expressions for first and second derivatives, and incorporating false boundaries to achieve $0(h^2)$ truncation error at the boundaries. The iteration was similar to that illustrated in Eq. (9), except the matrix $\bar{M}$ was tri-diagonal. Only single precision arithmetic was used on the CDC 6400 for these calculations. The error in the finite difference results goes as h^2, but with extrapolation techniques [16] can be made $0(h^4)$ as shown in other figures. The rate of convergence of finite difference results is much slower than orthogonal collocation, so that a much larger N is needed in the finite difference methods.

Results from orthogonal collocation on finite elements are shown in Fig. 5. Plotted here are the actual error in the temperature flux at the catalyst boundary versus the number of elements, NE. For Problem I each element has the same size. These calculations were done using single precision arithmetic, and round off errors begin to affect the results with $N = 7$ and 9. The residual as well as actual error begins to rise as N is increased above 5. Even for $N = 5$ round off error prevented getting solutions with large enough NE to obtain the slope of the curve. The truncation error results of DeBoor and Swartz show that the error E_F should be proportional to h^{2N}. For $N = 2$ the slope of 4·05 agrees well with the theoretical slope of 4 while for $N = 3$ the slope of 6·4 is close to 6. The finite difference results show the error going as $0(h^2)$, or with extrapolation as $0(h^4)$, but the curve is still above the corresponding result for orthogonal collocation on finite elements, $N = 2$, which also gives $0(h^4)$ error. This demonstrates the reduction in error by going from finite difference to orthogonal collocation on finite elements, while keeping the same grid spacing. In comparing the two methods the reader must keep in mind that the tolerance on the solution (to stop the iterations) was 10^{-7} for the finite difference results, requiring 7 iterations, and 10^{-12} for orthogonal collocation on finite elements, requiring 13 iterations. For collocation solutions no more accurate than 10^{-7}, the computation time is needlessly long to reach a tolerance of 10^{-12}.

Figure 6 plots the error in center temperature as it depends on the number of elements. Since the center ($x = 0$) is the end of one element, the error should be $0(h^{2N})$, as is found. Using these values we can extrapolate the finite element results by applying

$$E_1 = E + b\left(\frac{1}{NE_1}\right)^{2N}$$

$$E_2 = E + b\left(\frac{1}{NE_2}\right)^{2N}$$

where E_i is the result when using NE_i elements. Solution of these two equations gives b and E, the best estimate of the flux. Such an extrapolation can decrease the error by factors ranging from 30 to 100.

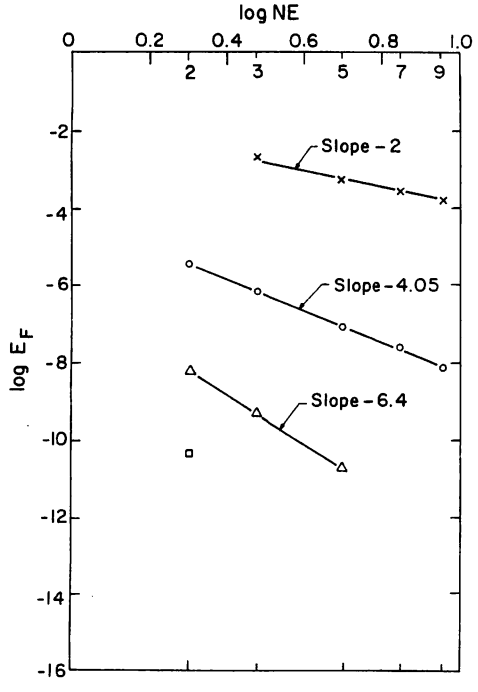

Fig. 5. Error in boundary heat flux for Problem I as a function of the number of elements. Orthogonal collocation on finite elements: ○, $N = 2$; △, $N = 3$; □, $N = 5$; finite difference x.

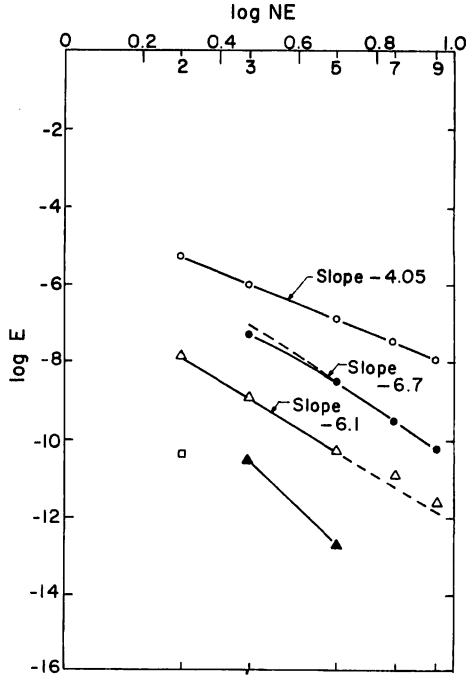

Fig. 6. Error in center temperature, $T(x = 0)$, for Problem I as a function of the number of elements in orthogonal collocation on finite elements: ○, $N = 2$; △, $N = 3$, □, $N = 5$. Filled symbols obtained by extrapolation and NE refers to the largest NE used in the extrapolation.

Finally the error in flux versus the computation time (CPU) on the CDC 6400 is shown in Fig. 7. Provided the number of grid points is the same for all methods, the finite difference method is fastest, since computation involves solution of a tri-diagonal system. Orthogonal collocation on finite elements is more time consuming, because a block diagonal system must be solved. Finally, orthogonal collocation is the slowest (for the same N) since a dense matrix results, with nearly every element non-zero. Of course the accuracy as a function of N has the reverse behavior: orthogonal collocation is most accurate for the same N, etc. Consequently the real comparison of the methods depends on efficiency: accuracy achieved per computation time. We recall that the orthogonal collocation calculations used double precision arithmetic and the Newton-Raphson method (requiring 4 iterations to obtain the solution to full precision). The finite difference and orthogonal collocation on finite element methods used an iterative scheme, Eq. (8), that had to decompose the matrix $\bar{M}$ only once. For this class of problems and calculations to 7 significant digits, computation times are approximately 10 times larger if the Newton-Raphson method is used. Since the orthogonal collocation recognizes the symmetry in $x = 0$, only an even expansion is necessary. On the other hand, orthogonal collocation on finite elements uses a general polynomial approximant on each element.

The most efficient solution, in terms of the smallest error for a given amount of computation, is the one found with orthogonal collocation, and the advantage over other methods increases as the allowed error decreases. For very high accuracy the orthogonal collocation method is certainly preferred. The next best method is orthogonal collocation on finite elements with the highest N being preferred. The finite difference method is the least efficient, giving the largest error for a given computation time. The extrapolated finite difference results are slightly better than the unextrapolated results for orthogonal collocation on finite elements with $N = 2$ (cubic trial functions). If these latter results are extrapolated, too, then they regain their advantage. The improvement through extrapolation is not as dramatic for the higher order methods, and for $N = 3$ appears to give no advantage in efficiency at all. Compared to the extrapolated finite difference results, the orthogonal collocation method is about 2 times as fast at an accuracy of 0·3%, 2·9 times at 0·1%, and 3·5 times at 0·01%. Compared to the unextrapolated finite difference results, the orthogonal collocation method is 3·5, 5·5 and 11 times as fast for errors of 0·3, 0·1 and 0·01%, respectively. Solutions with larger errors cannot be compared since the first approximation by orthogonal collocation gives an accuracy of 0·3%.

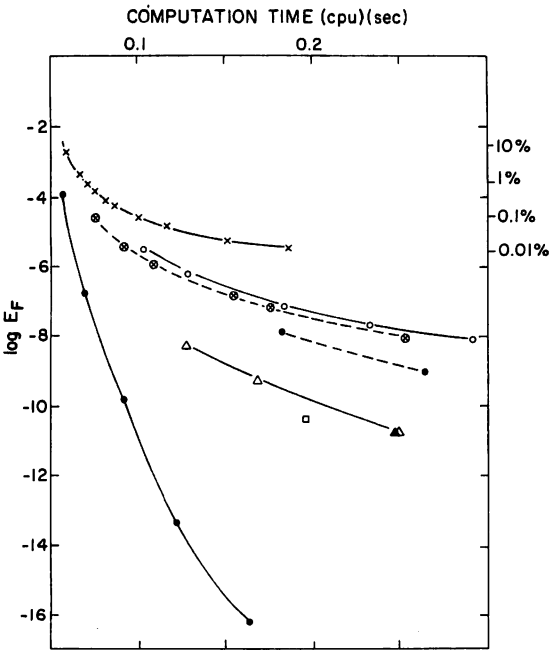

Fig. 7. Error in boundary flux for Problem I as a function of computation time. ●, Orthogonal collocation; Orthogonal Collocation on finite elements: ○, $N = 2$; △, $N = 3$; □, $N = 5$; filled symbols obtained by extrapolation, with CPU time being the total time; x, Finite difference; ⊗, extrapolated to obtain Δx^4 truncation error.

In terms of the total number of grid points, for an accuracy of 0·3% the unextrapolated finite difference method needed 10 times as many interior grid points as the collocation method needed interior collocation points; at 0·1% the factor is 14 while at 0·01% it is 35. The corresponding numbers for the extrapolated finite difference results are 3·5, 4 and 6. The method of orthogonal collocation on finite elements for $N = 2$ requires 4·5 times as many elements and 3 times as many total collocation points as for $N = 3$, for an accuracy of 10^{-8} in the error in flux. Clearly then the number of elements needed is reduced by going to higher degree polynomials, and the total number of collocation points (for a given accuracy) is reduced as well. Even fewer collocation points are needed in orthogonal collocation. Such a reduction in total number of grid points is particularly important in multi-dimensional calculations. These considerations may be valid for all problems with smooth solutions, but they are not valid when the solution exhibits a boundary layer structure.

Problem II

Consider the same type of problem—first-order irreversible, non-isothermal reaction in a spherical catalyst ($a = 3$)—but with more realistic boundary conditions and parameters. The Biot number for mass transfer (modified Sherwood number) is taken as 250 in Eq. (2) while the Biot number for heat transfer is taken as 5. A smaller, more realistic $\beta = 0{\cdot}02$ is taken; with $\gamma = 20$, and $\phi = 14{\cdot}44$, which give a steep concentration gradient as shown in Fig. 3. The complete η vs ϕ curve for these parameters is in Fig. 4 of [3]. Here the reaction rate expression is

$$f(c) = \phi^2 c \exp[\gamma(1 - 1/T)]$$

$$T(x) = 1 + \beta\delta + \beta(1 - \delta)c(1) - \beta c(x)$$

with $\delta = Bi_m/Bi = 50$. For this problem the iteration scheme, Eq. (8), did not converge due to the large Lipschitz constant so the Newton-Raphson method was

used. We used as the initial guess in all solutions the one-term orthogonal collocation solution derived using Paterson and Cresswell's method [5] described in the introduction. The Newton-Raphson scheme also converged for the initial guess $c(x) = 0$.

Error bounds cannot be derived for this solution due to the large reaction rate and the very large Lipschitz constant. We use as the exact solution the value of η corresponding to $N = 5$ and $NE = 15$, $\eta = 3{\cdot}0651065174126$. The results for lower-degree polynomials converge to this answer to within 10^{-9} and the rate of convergence for $N = 5$ as NE is increased suggests that the $NE = 15$ answer should be the most precise.

We did calculations only for orthogonal collocation on finite elements, to evaluate the capability of this technique for this type of boundary layer problem. The calculations proceeded as follows. The initial guess was found using Paterson and Cresswell's method [5]. This gives the layer location at 0·997. For each degree of polynomial $(N + 1)$, the finite element calculations began with $NE = 5$, with the boundaries between elements at 0·5, 0·997, 0·998, 0·999. After this solution was obtained the residual was examined. Five new elements were added within those previous elements having the largest integrated residual. If the residual for one element was more than ten times as large as the next highest residual, two elements were added there, but only five new elements were added each time. This procedure resulted in a non-uniform distribution of collocation points, but as NE increased the residual forced a more gradual transition in element sizes. This is shown in Fig. 8 for $N = 3$. The location of grid points was done manually by interaction with the computer for this pilot investigation, and prior to developing a strategy for automated mesh refinement. Each calculation started from the same initial guess, i.e. the solution for $NE = 15$ was not used to begin the calculations for $NE = 20$. If one is interested in solving the problem, rather than in displaying the features of the method, one would use the last solution as the first guess for the next value of NE to decrease the number of iterations, and hence the computation time.

In Fig. 9 are plotted the error in η vs NE for polynomials with $N = 2$, 3 and 5, respectively, in each element. Shown there is the combined effect of the truncation error, which depends on N, and the effect of the optimum location of the elements. Consider first the solid curves, which correspond to element location by the residual. For $N = 2$, the slope of the line is $-4{\cdot}4$, just

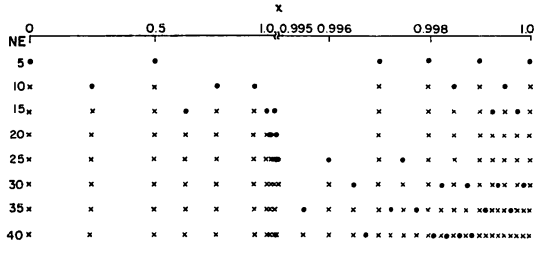

Fig. 8. Location of elements for Problem II, $N = 3$. x, end points of existing elements; ., new elements; calculation proceeds from top to bottom.

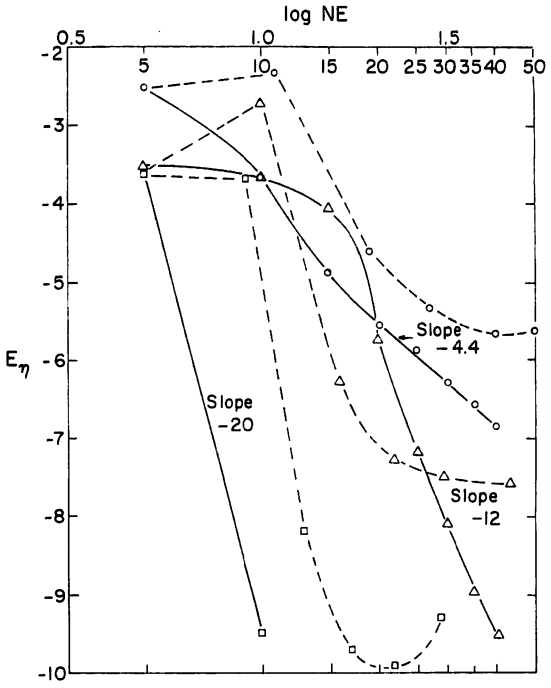

Fig. 9. Error in effectiveness factor as a function of the number of elements for Problem II. (○, $N = 2$; △, $N = 3$; □, $N = 5$; ——, residual used to locate elements; ---, concentration solution used to locate elements).

slightly larger than the $-4{\cdot}0$ for uniform distribution. Of course a uniform distribution of elements would require a much larger number of elements for the same accuracy. Indeed, 50 uniformly spaced elements gave errors larger than 0·1 for all N. For $N = 3$ the improvement due to the optimum location of elements is more dramatic, giving a slope of -12 instead of the theoretical value of -6 for uniform spacing. For $N = 5$ the slope is -20 (admittedly based on only two points since the next point is used as the exact solution), compared to a theoretical value of -10 for a uniform grid. Clearly the use of the residual to locate the element positions greatly improves the rate of convergence. The hump in the curve for $N = 3$ is caused by a placement of elements far from the boundary, but as NE increases the additional elements are placed near the boundary and the error decreases rapidly.

When the solution is used to locate the elements, giving the dotted curves in Fig. 9, the error initially increases, because the element nearest $x = 1{\cdot}0$ is enlarged, but thereafter the error decreases. The error eventually stops decreasing, showing that adding more elements is not advantageous. Compared to the residual location of the elements, both schemes are feasible, but the residual location gives slightly better results, at the expense of increased computation time.

The mean square residual also decreases as N increases for a fixed location of elements or as the number of elements increases for fixed N. Even though the mean square residual cannot be proved to be related to the pointwise error for this problem, it can serve as an indicator. For $N = 2$ the value of RS went from 1500 for $NE = 5$ to 2·7 for $NE = 40$. For $N = 3$ the corresponding

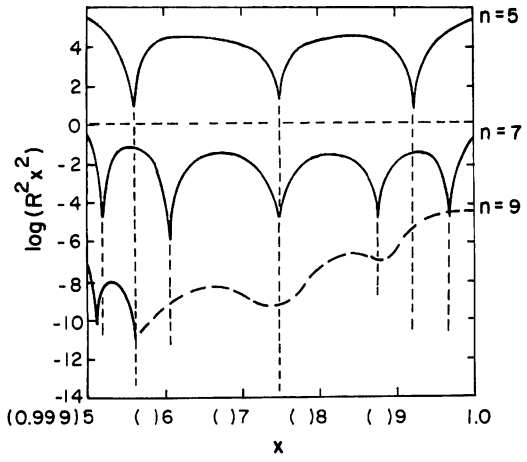

Fig. 10. Pointwise residual in last element.

values were 1200 to 0·12, while for $N = 5$ they were 930 for $NE = 5$, 0·25 for $NE = 10$ and 0·010 for $NE = 15$. The value of the residual also suggests that the solution for $NE = 15$, $N = 5$ is the most accurate one. The residuals do not decrease as fast as the error, in agreement with Fig. 4. The pointwise residual is plotted in Fig. 10. It goes to zero at the collocation points and decreases as NE increases. The irregularity for $N = 7$ is probably due to the influence of cancellation error in single precision computations.

Figure 11 gives a comparison of the error in effectiveness factor as it depends on computation time. At large

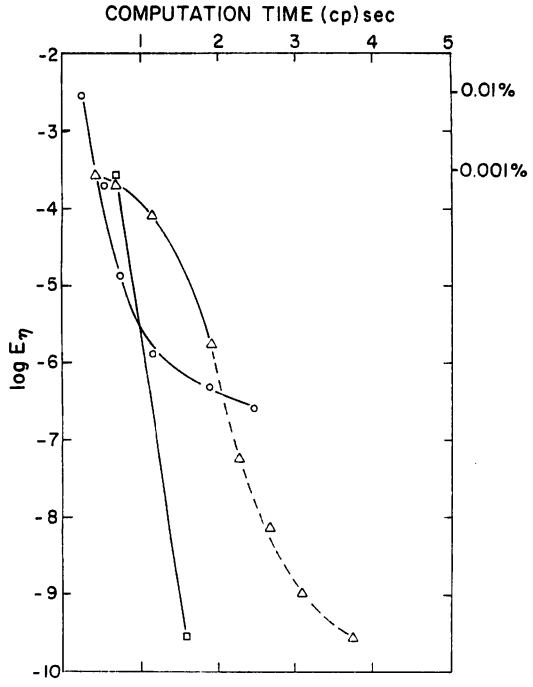

Fig. 11. Error in effectiveness factor as a function of computation time for Problem II (Caption as in Fig. 9). The dotted portion of the $N = 3$ curve is based on an interpolation of computation for 12 iterations (the number used by nearly all the other calculations). More iterations were actually used and the computations stopped at the iteration limit due to inappropriate choices of convergence criteria.

errors any of the methods can be used with equal efficiency, whereas for small errors the highest N feasible should be used.

Extension to partial differential equations

Although the method of orthogonal collocation on finite elements is applied above to the one-dimensional effectiveness factor problem, it can also be applied to multi-dimensional and transient problems. The important feature is the efficient LU-decomposition of the block diagonal matrices. For transient effectiveness factor problems, or plug flow reactors, a Crank-Nicolson-type method leads to the same block diagonal structure provided one iterates on one variable at a time. Such applications are discussed elsewhere[17].

CONCLUSION

The method of orthogonal collocation on finite elements is outlined. An $\bar{L}\bar{U}$-decomposition is a convenient method to solve the problem which generates block diagonal matrices. The residual from a previous calculation can be used to locate additional elements to reduce the error. The truncation error of the function and first derivative at the element end points is $0(h^{2N})$ where $h = 1/NE$, NE is the number of elements and $N + 1$ is the degree of polynomial in each element.

Numerical studies on one typical problem indicate the method is efficient in generating an accurate solution in small amounts of computation time, and the efficiency increases as N increases. For one effectiveness factor problem, corresponding to a small Thiele modulus, the method of orthogonal collocation on finite elements is more efficient than a finite difference method and less efficient than orthogonal collocation, using one polynomial over the whole domain. For another effectiveness factor problem, with a boundary layer type of solution, the global orthogonal collocation method is unsatisfactory. Orthogonal collocation on finite elements is very suitable, yielding accurate answers, and the location of elements by examination of the residual improves the method by reducing the error in the solution.

Acknowledgement—This research was supported by the National Science Foundation, Grant GK 12517.

NOTATION

- a 1, 2, 3 for planar, cylindrical, and spherical geometry
- A_{ji} Orthogonal collocation matrix representing first derivative
- b Constant in extrapolation equation
- B_{ji} Orthogonal collocation matrix representing second derivative
- Bi Biot number for heat transfer
- Bi_m Biot number for mass transfer (modified Sherwood number)
- c Concentration
- $\bar{C}$ Concentration vector at collocation points
- d Coefficient in polynomial expansion
- E_d Iteration tolerance on solution
- E_F Error in flux at $x = 1$
- E_i Error in i th calculation

f reaction rate
$\bar{F}$ Vector in Fig. 2
$\bar{G}$ Vector in Eq. (12)
h Element size, Δx
$\bar{\bar{J}}$ Jacobian matrix in Newton–Raphson method
K Lipschitz constant
$\bar{\bar{M}}$ Block diagonal matrix in Fig. 2
N Number of interior collocation points in each element
NE Number of elements
NP $N+2$, the number of collocation points in one element
P $\Delta x^2 K\|\bar{\bar{M}}\|$
Q_{ji} Interpolation matrix
R Residual
RS Mean squared residual
S_{jil} Stacked version of matrix M
t Time
t_f Final time
t_k $k\Delta t$
T Temperature
u Position in element
x Position in catalyst pellet
Δx Element size

Greek symbols

β Dimensionless heat of reaction
γ Dimensionless activation energy
δ Bi_m/Bi
δ^* Parameter used in locating elements
δ_{ji} Kronecker delta
η Effectiveness factor
ϕ Thiele modulus

Subscripts and Superscripts

l Element index
k Iteration index
j j th collocation point
m Element index

REFERENCES

[1] Villadsen J. V. and Stewart W. E., *Chem. Engng Sci.* 1967 **22** 1483–1501.
[2] Finlayson B. A., *The Method of Weighted Residuals and Variational Principles.* Academic Press, New York (1972).
[3] Finlayson B. A., *Cat. Rev.-Sci. Engng* 1974 **10** 69–138.
[4] Ferguson N. B. and Finlayson B. A., *A.I.Ch.E.J.* 1972 **18** 1053–1059.
[5] Paterson W. R. and Cresswell, D. L., *Chem. Engng Sci.* 1971 **26** 605–616.
[6] Villadsen J. and Sorensen J. P., *Chem. Engng Sci.* 1969 **24** 1337–1349.
[7] Martin H. C. and Carey G. F., *Introduction to Finite Element Analysis.* McGraw-Hill, New York, 1973.
[8] Ciarlet P. G., Schultz M. H. and Varga R. S., *Num. Math.* 1967 **9** 394–430.
[9] Forsythe G. and Moler C. B., *Computer Solution of Linear Algebraic Systems.* Prentice-Hall, New Jersey, 1967.
[10] Traub J. F., *Iterative Methods for the Solution of Equations.* Prentice Hall, New York, 1964.
[11] Michelsen M. L. and Villadsen J., *Chem. Engng J.*, 1972 **4** 64–68.
[12] Pearson C., *J. Math. Phys.*, 1968 **47** 351–358.
[13] Douglas J., Jr. and DuPont T., *Math Comp.* 1973 **27** 17–28.
[14] DeBoor C. and Swartz B., *SIAM J. Numer. Anal.* 1973 **10** 582–606.
[15] Douglas J., Jr. *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations* (Edited by Aziz A. K.), pp. 475–490. Academic Press, New York, 1972.
[16] Keller H. B., *Numerical Methods for Two-Point Boundary-Value Problems.* Blaisdell, New York, 1968.
[17] Finlayson B. A., *Finite Elements in Fluids: Mathematical Foundations, Aerodynamics and Lubrication* (Edited by R. H. Gallagher, J. T. Oden, C. Taylor and O. C. Zienkiewicz), Chap. 1. Wiley, New York, 1974.