INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN ENGINEERING, VOL. 15, 1681–1689 (1980)

OSCILLATION LIMITS FOR WEIGHTED RESIDUAL METHODS APPLIED TO CONVECTIVE DIFFUSION EQUATIONS

OLE KROGH JENSEN AND BRUCE A. FINLAYSON

Department of Chemical Engineering, University of Washington, Seattle, Washington, U.S.A.

SUMMARY

Convection–diffusion equations are difficult to solve when the convection term dominates because most solution methods give solutions which oscillate in space. Previous criteria based on the one-dimensional convection–diffusion equation have shown that finite difference and Galerkin (linear or quadratic basis functions) will not give oscillatory solutions provided the Peclet number times the mesh size (Pe Δx) is below a critical value. These criteria are based on the solution at the nodes, and ensure that the nodal values are monotone. Similar criteria are developed here for other methods: quadratic Galerkin with upwind weighting, cubic Galerkin, orthogonal collocation on finite elements with quadratic, cubic or quartic polynomials using Lagrangian interpolation, cubic or quartic polynomials using Hermite interpolation, and the method of moments. The nodal values do not oscillate for collocation or moments methods with Hermite cubic polynomials regardless of the value of Pe Δx.

A new criterion is developed for all methods based on the monotonicity of the solutions throughout the domain. This criterion is more restrictive than one based only on the nodal values. All methods that are second order (Δx^2) or better in truncation error give oscillatory solutions (based on the entire domain) unless Pe Δx is below a critical value. This value ranges from 2 for finite difference methods to 4·6 for Hermite, quartic, collocation methods.

INTRODUCTION

Convection–diffusion equations are difficult to solve when the convection terms dominate. Similarly the Navier–Stokes equation suffers from the same difficulty for large Reynolds numbers. The difficulty is manifest by solutions that oscillate in space, in contrast to the real behaviour. These oscillations are not present if the mesh size is sufficiently small. It is not always possible to use small enough mesh sizes, however, for economical reasons, and then the analyst must turn to various upwind weighting or differencing schemes. These schemes eliminate the oscillations but degrade the accuracy of the solution. Because of these problems it is important to know *a priori* when oscillations are expected.

Price *et al.*[1] first showed for the one-dimensional transient equation that the finite difference method would give solutions which did not oscillate provided the mesh spacing satisfied

$$\mathrm{Pe}\Delta x \leqslant 2 \tag{1}$$

Christie *et al.*[2] provided a criterion for the one-dimensional steady-state equation with the Galerkin finite element method and linear or quadratic trial functions. Linear functions give the criterion (1), whereas quadratic functions give the criterion

$$\mathrm{Pe}\Delta x \leqslant 4 \tag{2}$$

0029–5981/80/1115–1681$01.00

Received 17 October 1979
Revised 4 February 1980

Both criteria are based on examining the conditions needed to make the solution at successive nodes monotone. For the finite difference method or Galerkin method with linear shape functions this also makes the solution between the nodes monotone, since the interpolation is linear between nodes. For the quadratic Galerkin method, however, it is possible for the solution at the nodes to be monotone but still violate the monotonicity between the nodes. This is illustrated in Figure 1 for a case in which the solution is 1 at the first two nodes of an element and zero at the last node. Clearly the quadratic interpolant achieves a maximum and is not monotone.

If only the steady-state solution is desired, and the solution is sampled only at the nodes, monotonicity throughout the domain may not be important. For the transient problem, however, the steep front moves in time, and a maximum which occurs interior to an element at one time may be a maximum at a node at a later time. Thus monotonicity throughout the domain in the steady problem is necessary to insure monotonicity at the nodes in transient problems. New criteria are derived to ensure monotonicity throughout the domain, based on the steady-state equations. Galerkin and collocation finite element methods are considered for a variety of basis functions. For most methods the monotonicity criterion is more restrictive when based on the whole domain instead of just the nodes.

MODEL PROBLEM

The model problem and boundary conditions are

$$\frac{d^2c}{dx^2} - \mathrm{Pe}\frac{dc}{dx} = 0 \tag{3}$$

$$c = 1 \quad \text{at } x = 0$$

$$c = 0 \quad \text{at } x = 1$$

The theoretical solution of (3) is

$$c = A + B\,e^{\mathrm{Pe}x} = (e^{\mathrm{Pe}} - e^{\mathrm{Pe}x})/(e^{\mathrm{Pe}} - 1) \tag{4}$$

The solution satisfies

$$c_i > c_j \quad \text{if } x_i < x_j \tag{5}$$

The function c is thus a monotone function over the whole interval, zero to one.

CRITERION FOR OSCILLATIONS

The finite element method approximates a variable within an element k by a polynomial of degree $NP-1$.

$$c_k(x) = \sum_{i=0}^{NP-1} a_i' x^i$$

Normalizing x for one element we have

$$c(u) = \sum_{i=0}^{NP-1} a_i u^i, \qquad u \in [0, 1] \tag{6}$$

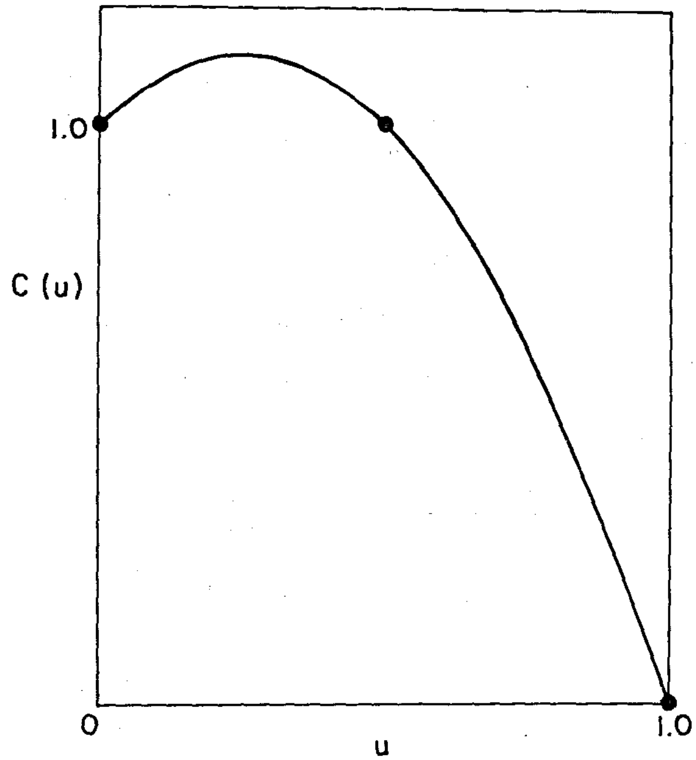

Figure 1. Galerkin quadratic interpolation on one element

The solution (4) is monotone decreasing, so the most restrictive criterion to ensure no oscillations is that the numerical solution satisfy

$$\frac{\mathrm{d}c}{\mathrm{d}x}<0 \qquad \text{for all } x\in[0, 1]$$

The constants a_i in (6) can be found once the analytical expression for the value of c at all nodes in an element are known. Then

$$\frac{\mathrm{d}}{\mathrm{d}u}(c(u))=\sum_{i=1}^{NP-1} ia_iu^{i-1}$$

and we require

$$\frac{\mathrm{d}c(u)}{\mathrm{d}u}<0 \qquad \text{for } u\in[0, 1] \tag{7}$$

for any element. The Galerkin method with quadratic shape functions and orthogonal collocation on finite elements (OCFE) with cubic Hermite shape functions are examined using this criterion. For all the finite element methods we divide domain $0\leqslant x\leqslant 1$ into NE equal elements of length $\Delta x=1/NE$. We denote by c_i the concentration at the ith node, and reserve integer i for the nodes at the ends of the element. Fractional i correspond to points interior to the element.

Galerkin method with quadratic shape functions

Christie *et al.*[2] derived the difference equations for this mode:

$$\begin{aligned}(1+\mathrm{Pe}\Delta x/2)c_{i-1}+(-8-2\mathrm{Pe}\Delta x)c_{i-1/2}+14c_i+(-8+2\mathrm{Pe}\Delta x)c_{i+1/2}\\ +(1-\mathrm{Pe}\Delta x/2)c_{i+1}=0\end{aligned} \tag{8}$$

for nodes at the end of an element and

$$(4+\mathrm{Pe}\Delta x)c_i - 8c_{i+1/2} + (4-\mathrm{Pe}\Delta x)c_{i+1} = 0 \tag{9}$$

for nodes at the mid-side of an element. Combining (8) and (9) gives

$$\begin{aligned}(1+\mathrm{Pe}\Delta x/2-\mathrm{Pe}^2\Delta x^2/12)c_{i-1}+(2+\mathrm{Pe}^2\Delta x^2/6)c_i\\ +(1-\mathrm{Pe}\Delta x/2+\mathrm{Pe}^2\Delta x^2/12)c_{i+1}=0\end{aligned} \tag{10}$$

which only involves integer nodes. The exact solution to the difference equation (10) is

$$c_i = \mathrm{A} + \mathrm{B}\phi^i \tag{11}$$

where

$$\phi = \frac{1+\mathrm{Pe}\Delta x/2+\mathrm{Pe}^2\Delta x^2/12}{1-\mathrm{Pe}\Delta x/2+\mathrm{Pe}^2\Delta x^2/12}, \qquad \phi > 1 \tag{12}$$

and A and B are determined by the boundary conditions. Making $c_0 = 1$ and $c_{NE} = 0$ ($N = NE =$ the number of elements):

$$\mathrm{A} = \frac{\phi^N}{\phi^N - 1}, \qquad \mathrm{B} = \frac{-1}{\phi^N - 1}, \qquad c_i = \frac{\phi^N - \phi^i}{\phi^N - 1} \tag{13}$$

Substituting equation (11) into equation (9) gives the solution for the mid-side nodes:

$$c_{i+1/2} = \mathrm{A} + \mathrm{B}\phi^i \frac{1-\mathrm{Pe}^2\Delta x^2/24}{1-\mathrm{Pe}\Delta x/2+\mathrm{Pe}^2\Delta x^2/12} \tag{14}$$

Equations (13) and (14) thus give the exact solution to the Galerkin equations. We can now evaluate the constants a_i in any element. We have

$$c(u) = a_0 + a_1 u + a_2 u^2$$

where

$$c(0) = c_i, \qquad c(\tfrac{1}{2}) = c_{i+1/2}, \qquad c(1) = c_{i+1}$$

which gives

$$\begin{aligned} a_0 &= c_i \\ a_1 &= -3c_i + 4c_{i+1/2} - c_{i+1} \\ a_2 &= 2c_i - 4c_{i+1/2} + 2c_{i+1} \end{aligned} \tag{15}$$

Thus criterion (7) gives

$$\begin{aligned}\frac{\mathrm{d}c}{\mathrm{d}u} &= a_1 + 2a_2 u < 0 \\ &= 2a_2\left(u + \frac{a_1}{2a_2}\right) < 0\end{aligned} \tag{16}$$

For a maximum or minimum in c to occur in the element $-a_1/2a_2$ must be between -1 and zero. Substituting equations (11), (12), (14) into equation (15) gives

$$-a_1/2a_2 = \tfrac{1}{2} - 1/\mathrm{Pe}\Delta x \tag{17}$$

For $\mathrm{Pe}\Delta x < 2$, c has a maximum outside the element, which is of no concern since the element interpolation is only inside the element. Furthermore, c will be monotone decreasing inside the element. For $\mathrm{Pe}\Delta x > 2$, one maximum occurs inside the element. By looking at

$$c_i > c_{i+1} \quad \text{and} \quad c_i > c_{i+1/2}$$

the apparent criterion would be $\mathrm{Pe}\Delta x < 4$ The restrictive criterion to ensure no oscillations (7) is then $\mathrm{Pe}\Delta x < 2$ for quadratic shape functions and the Galerkin method.

Collocation with cubic Hermite interpolation

The formulation of OCFE is given elsewhere.[3,4] When applied to equation (2) we get for an arbitrary element

$$\left(-\frac{\sqrt{12}}{\Delta x^2}+\frac{\mathrm{Pe}}{\Delta x}\right)c_{i-1}+\left(\frac{-(1+\sqrt{3})}{\Delta x^2}-\frac{1}{\sqrt{12}}\frac{\mathrm{Pe}}{\Delta x}\right)c'_{i-1}+\left(\frac{\sqrt{12}}{\Delta x^2}-\frac{\mathrm{Pe}}{\Delta x}\right)c_i+\left(\frac{1-\sqrt{3}}{\Delta x^2}+\frac{1}{\sqrt{12}}\frac{\mathrm{Pe}}{\Delta x}\right)c'_i=0$$

and

$$\left(\frac{\sqrt{12}}{\Delta x^2}+\frac{\mathrm{Pe}}{\Delta x}\right)c_{i-1}+\left(\frac{\sqrt{(3)}-1}{\Delta x^2}+\frac{1}{\sqrt{12}}\frac{\mathrm{Pe}}{\Delta x}\right)c'_{i-1}+\left(-\frac{\sqrt{12}}{\Delta x^2}-\frac{\mathrm{Pe}}{\Delta x}\right)c_i+\left(\frac{1+\sqrt{3}}{\Delta x^2}-\frac{1}{\sqrt{12}}\frac{\mathrm{Pe}}{\Delta x}\right)c'_i=0$$

Here c_i and c'_i are the function and first derivative at the element boundary x_i. With two additional equations for the neighbouring element (using nodes c_i, c'_i, c_{i+1}, c'_{i+1}) we can reduce these four equations to

$$(1+\mathrm{Pe}\Delta x/2+\mathrm{Pe}^2\Delta x^2/12)c_{i+1}+(-2-\mathrm{Pe}^2\Delta x^2/6)c_i+(1-\mathrm{Pe}\Delta x/2+\mathrm{Pe}^2\Delta x^2/12)c_{i+1}=0, \tag{18}$$

which only involves node values, and

$$(1+\mathrm{Pe}\Delta x/2+\mathrm{Pe}^2\Delta x^2/12)c_{i-1}-(1+\mathrm{Pe}\Delta x/2+\mathrm{Pe}^2\Delta x^2/12)c_i+c'_i=0, \tag{19}$$

involving derivatives too. The solution to (18) is also equation (11) with ϕ given by equation (12). The nodal values c_i are given by equation (13). From (19) we obtain an expression for the derivative

$$c'_i=\mathrm{Pe}\Delta x\,.\,\mathrm{B}\,.\,\phi^i \tag{20}$$

We thus have solved the collocation Hermite equations exactly. The nodal values are monotone and the first derivative at the nodes is always negative. Thus oscillations do not occur in these values for *any* $\mathrm{Pe}\Delta x$.

To evaluate the limit using the entire element, we look at the cubic expansion in an arbitrary element

$$c(u)=a_0+a_1u+a_2u^2+a_3u^3 \tag{21}$$

with the conditions

$$c(0)=c_i \qquad \frac{\mathrm{d}c}{\mathrm{d}u}(0)=c'_1$$

$$c(1)=c_{i+1} \qquad \frac{\mathrm{d}c}{\mathrm{d}u}(1)=c'_{i-1} \tag{22}$$

The coefficients are

$$a_0 = c_i, \qquad a_1 = c_i'$$

$$a_2 = -3c_i - 2c_i' + 3c_{i+1} - c_{i+1}'$$

$$a_3 = 2c_i + c_i' - 2c_{i+1} + c_{i+1}'$$

Using equations (12), (13) and (20), the constants a_i can now be calculated.

The criterion (7) is

$$\frac{dc}{du} = a_1 + 2a_2 u + 3a_3 u^2 < 0$$

which can be written as

$$\frac{dc}{du} = 3a_3(u - u_1)(u - u_2) < 0$$

with

$$u_{1,2} = \frac{1}{2} - \frac{1}{\text{Pe}\Delta x} \pm \sqrt{\left(\frac{1}{12} - \frac{1}{\text{Pe}^2 \Delta x^2}\right)} \tag{23}$$

Both u_1 and u_2 must be outside [0, 1] to avoid oscillations. This gives

$\text{Pe}\Delta x < \sqrt{12}$, no oscillations

$\text{Pe}\Delta x = \sqrt{12}$, one maximum in c in $u \in [0, 1]$

$Pe\Delta x > \sqrt{12}$, one maximum and one minimum in c

The most restrictive criterion is then $\text{Pe}\Delta x < \sqrt{12}$. If we only examine the nodal values and the derivatives at the nodes, there is no limit; oscillations do not occur for any $\text{Pe}\Delta x$. Furthermore, if

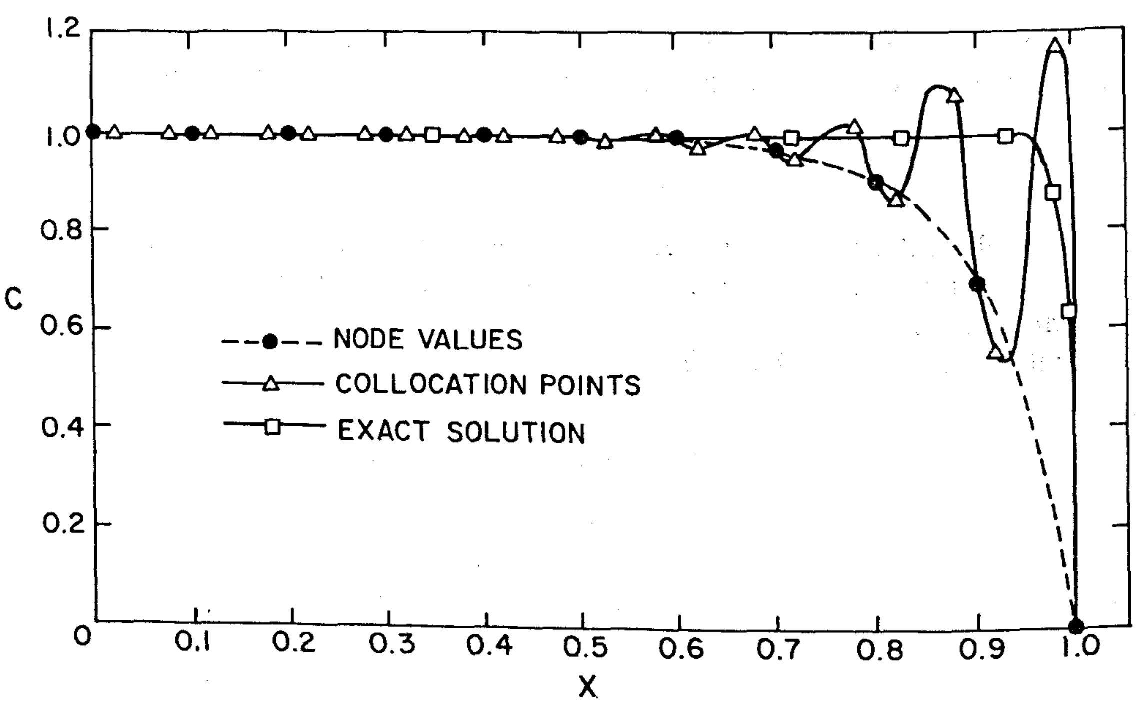

Figure 2. Solution of convection–diffusion equation with cubic Hermite polynomials: $\text{Pe} = 100$, $\Delta x = 0{\cdot}1$, $\text{Pe}\Delta x = 10$

Table I. Oscillation limits for convective diffusion, $\mathrm{Pe}\Delta x \leq \mathrm{B}$

Method	ϕ in equation (11)	Criterion based on monotonicity, equation (7), B	Criterion based on monotonicity at nodes, B	Equation order of method	Work†	Comparative work‡
Finite difference:						
second-order	$(1+\mathrm{Pe}\Delta x/2)/(1-\mathrm{Pe}\Delta x/2)$	2	2	Δx^2	$5NE$	2·5 Pe
upstream weighting	$(1+\mathrm{Pe}\Delta x)$	∞	∞	Δx	$5NE$	0 Pe
Galerkin C^0:						
linear	$(1+\mathrm{Pe}\Delta x/2)/(1-\mathrm{Pe}\Delta x/2)$	2	2	Δx^2	$5NE$	2·5 Pe
linear with upstream weighting	$(1+(\alpha+1)\mathrm{Pe}\Delta x/2)/(1+(\alpha-1)\mathrm{Pe}\Delta x/2)$	$2/(1-\alpha)$	$2/(1-\alpha)$	Δx	$5NE$	$2{\cdot}5(1-\alpha)$ Pe
quadratic	$(1+\mathrm{Pe}\Delta x/2+\mathrm{Pe}^2\Delta x^2/12)/(1-\mathrm{Pe}\Delta x/2+\mathrm{Pe}^2\Delta x^2/12)$	2	4	Δx^3	$16NE$	8·0 Pe
quadratic with upstream weighting and optional α, β	$(6+3)(1+\beta/2)\mathrm{Pe}\Delta x+(2+\alpha+\beta)\mathrm{Pe}^2\Delta x^2/4)/(6+3)(-1+\beta/2)\mathrm{Pe}\Delta x+(2-\alpha-\beta)\mathrm{Pe}^2\Delta x^2/4)$	2	∞	Δx	$16NE$	8·0 Pe
cubic, C^0	same as quartic, $NP=5$, OCFE	4·644	4·644	Δx^4	$35NE$	7·5 Pe
OCFE–Lagrange, C^1:						
quadratic, $NP=3$	same as Galerkin linear	2	2	Δx^3	$16NE$	8·0 Pe
cubic, $NP=4$	same as Galerkin quadratic	$\sqrt{12}=3{\cdot}4641$	4·3914	Δx^4	$35NE$	10·1 Pe
quartic, $NP=5$	$(1+\mathrm{Pe}\Delta x/2+\mathrm{Pe}^2\Delta x^2/10+\mathrm{Pe}^3\Delta x^3/120)/(1-\mathrm{Pe}\Delta x/2+\mathrm{Pe}^2\Delta x^2/10-\mathrm{Pe}^2\Delta x^3/120)$	4·644	4·644	Δx^5	$64NE$	13·8 Pe
OCFE–Hermite, C^1:						
cubic, $NP=4$	same as Lagrangian, $NP=4$	$\sqrt{12}=3{\cdot}4641$	∞	Δx^4	$12NE$	3·5 Pe
quartic, $NP=5$	same as Lagrangian, $NP=5$	4·644	∞	Δx^2	$17NE$	5·0 Pe
Moments, C^1:						
cubic	same as Galerkin quadratic	$\sqrt{12}=3{\cdot}4641$	∞	Δx^4	$12NE$	3·5 Pe

† Work is defined as the number of multiplications to perform the LU decomposition of the matrix, optimized for each method, for a problem with NE elements.
‡ Comparative work is the number of multiplications to perform an LU decomposition and one fore and aft sweep when the number of elements is chosen by column 1. This work is appropriate for a time integration scheme with variable step size, so that the LU decomposition must be done frequently.

we only look at the nodes and the collocation points, the criterion would be $\text{Pe}\Delta x < 4{\cdot}391$, as reported in Reference 5.

The numerical solution agrees with these deductions. Shown in Figure 2 is a solution of (3) using OCFE with first-order Hermitian shape functions and $\text{Pe} = 100$, $\Delta x = 1/10$, so $\text{Pe}\Delta x = 10$. The solution does not oscillate if we only look at the nodal values, or derivatives at the nodes, but inside the element both a maximum and minimum occur. This analysis suggests that when Hermite polynomials are used it is always necessary to look at values inside the elements.

Other finite element methods

The oscillation limits for other finite element methods are summarized in Table I. Details of the analysis can be found in Reference 6. Criteria based on any point and only the node points are included. The function ϕ appearing in equation (11) is given and is a Padé approximation to exp ($\text{Pe}\Delta x$).

We notice that the criterion based on the nodal values alone says that many methods will not oscillate. If the criteria are based on monotonicity over the entire domain then all methods, but one, oscillate for large enough Δx. The only exception is linear basis functions with upstream weighting (finite difference), which is low order (Δx) and inaccurate. The quadratic elements provide no improvement over linear elements since they have the same criterion based on the mesh size but have twice as many unknowns per element. The higher order methods with cubic or quartic basis functions do provide an improvement Furthermore, the mesh size limitation increases faster than the number of nodes per element, so that net improvement is obtained by using higher order elements. However, the work requirements increase more rapidly than the allowable mesh size. If we consider a computation with the mesh size chosen by criteria (7), and compare work effort, the smallest work is provided by the finite difference or Galerkin linear element. The next best method is the Hermite cubic collocation moments. Comparisons based on equivalent accuracy, rather than $\text{Pe}\Delta x = \text{B}$, may lead to other conclusions. For extremely large Pe the key problem is to use as large a Δx as possible without inducing oscillations, and in that case $\Delta x = \text{B}/\text{Pe}$ would be used, giving the work requirements in the last column of Table I.

CONCLUSION

Criteria are presented to eliminate oscillations in solutions to the convection-diffusion equation when solved by finite difference, Galerkin or collocation finite element methods. Criteria based on the nodal values alone are least restrictive and smaller Δx are needed when the criteria are based on monotonicity over the entire domain. All methods have a limitation on the allowable mesh size if oscillations are to be avoided, unless artificial dispersion is included.

APPENDIX: NOMENCLATURE

a_i = Constants to define element polynomial.
A,B = Constants defined by boundary conditions.
c = Dimensionless concentration
c_i = Concentration at ith node.
k = Element number.
NE = Number of uniform elements on $0 \leq x \leq 1$.
NP = Degree of polynomial is $NP - 1$.
Pe = Peclet number.

u = Spatial co-ordinate within element.
x = Spatial co-ordinate.
Δx = Grid spacing (uniform elements).
ϕ = Function.

REFERENCES

1. H. S. Price, R. S. Varga and J. E. Warren, 'Application of oscillation matrices to diffusion-convection equations', *J. Math. Phys.* **45**, 301–311 (1966).
2. I. Christie, D. F. Griffiths, A. R. Mitchell and O. C. Zienkiewicz, 'Finite element methods for second order differential equations with significant first derivatives', *Int. J. num. Meth. Engng* **10**, 1389–1396 (1976).
3. G. F. Carey and B. A. Finlayson, 'Orthogonal collocation on finite elements', *Chem. Eng. Sci.* **30**, 587–596 (1975).
4. B. A. Finlayson, *Nonlinear Analysis in Chemical Engineering*, McGraw-Hill, London and New York, 1980.
5. O. K. Jensen and B. A. Finlayson, 'Solution of the convection–diffusion equation using a moving coordinate system', in *Finite Elements in Water Resources* (*FE2*) (Eds C. A. Brebbia, W. G. Gray and G. F. Pinder), Pentech Press, London, 1978, pp. 4.21–4.32.
6. O. K. Jensen, 'Numerical modeling with a moving coordinate system: application to flow through porous media', *Ph.D. thesis*, University of Washington (1979).