INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN ENGINEERING, VOL. 28, 1077–1101 (1989)

CONSISTENT HIGHER DEGREE PETROV–GALERKIN METHODS FOR THE SOLUTION OF THE TRANSIENT CONVECTION–DIFFUSION EQUATION

J. J. WESTERINK

Ocean Engineering Program, Civil Engineering Department, Texas A & M University, College Station, TX 77843, U.S.A.

D. SHEA

R. M. Parsons Laboratory, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

SUMMARY

The solution of the convection–diffusion equation for convection dominated problems is examined using both $N+1$ and $N+2$ degree Petrov–Galerkin finite element methods in space and a Crank–Nicolson finite difference scheme in time. While traditional $N+1$ degree Petrov–Galerkin methods, which use test functions one polynomial degree higher than the trial functions, work well for steady-state problems, they fail to adequately improve the solution for the transient problem. However, using novel $N+2$ degree Petrov–Galerkin methods, which use test functions two polynomial degrees higher than the trial functions, yields dramatically improved solutions which in fact get better as the Courant number increases to 1·0. Specifically, cubic test functions with linear trial functions and quartic test functions in conjunction with quadratic trial functions are examined.

Analysis and examples indicate that $N+2$ degree Petrov–Galerkin methods very effectively eliminate space and especially time truncation errors. This results in substantially improved phase behaviour while not adversely affecting the ratio of numerical to analytical damping.

INTRODUCTION

For more than two decades the numerical solution of convective–diffusive transport phenomena has been a very active area of research for finite difference and finite element modellers. As is well documented, the central type difference expressions which are obtained from a central finite difference or standard Bubnov–Galerkin finite element approach give rise to spurious oscillations when transport is convection dominated. These oscillations result from difficulties associated with both the spatial and temporal discretizations.

The standard technique for dealing with oscillations stemming from the spatial discretization arose from extensive study of the steady-state case and involves adding artificial diffusion in an optimal way so that it balances the apparent negative diffusion inherent in central differencing of the convective term. The addition of artificial diffusion has been accomplished through a variety of schemes. In the finite difference regime, artificial diffusion can be implemented via a backward or upwind difference expression for the convective first derivative term, which essentially introduces a truncation error which has the same form as the physical diffusion term.

Artificial diffusion techniques for Galerkin finite element methods are accomplished by using modified forms of the standard test functions as weighting functions. We can highlight the

0029–5981/89/051077–25$12.50

Received 15 December 1987
Revised 14 September 1988

development of these upwinded or Petrov–Galerkin formulations as follows. Christie *et al.*[1] introduced upwinding or linear basis elements by adding a quadratic modification to the standard linear test functions. Heinrich *et al.*[2] extended upwinding to two dimensional linear elements. Hughes[3] then showed how upwinding could be accomplished more efficiently by shifting the quadrature points within an element. The concept of upwinding then proceeded to higher-order elements. Heinrich and Zienkiewicz[4] developed upwinding for 9-node Lagrangian elements and 8-node serendipity elements, and Christie and Mitchell[5] briefly looked at upwinding of one-dimensional cubic elements. Most of these early schemes were similar in that the effect of adding artificial diffusion was accomplished by adding a modification function to the test functions which was one degree higher than the basis functions. In addition, through difference equation analysis for the one-dimensional steady-state problem, formulae were developed which specified the optimal amount of upwinding needed to obtain the exact solution at the nodes.

Through the study of two-dimensional problems, it was independently discovered by Hughes and Brooks[6] and Kelly *et al.*[7] that artificial diffusion was only desirable, and could be implemented, in the direction of flow, thereby eliminating the problem of crosswind diffusion. The streamline upwind approach, as it came to be known, allowed oscillations to be suppressed simply by adding an optimal amount of artificial diffusion directly to the physical diffusion tensor while using a Bubnov–Galerkin formulation. Hughes and Brooks[8] and Brooks and Hughes[9] then proposed a formulation which modifies the usual test functions by a perturbation dependent upon the velocity field and the derivative of the basis functions. This scheme is called the streamline upwind/Petrov–Galerkin (SU/PG) method. It differs from the original technique of upwinding of finite elements in that the test functions are no longer modified by a higher degree polynomial, but are actually perturbed by a lower degree function than the element shape functions. SU/PG offered the benefits of eliminating crosswind diffusion in a consistent framework and as such eliminated some of the problems of streamline upwinding. Donea *et al.*[10] then developed an alternative approach to streamline upwinding by perturbing the steady-state transport equation, by subtracting from the original equation the scalar product of its gradient and a vector of upwinding parameters, and applying the standard Bubnov–Galerkin method to discretize this perturbed equation. They apply their scheme in conjunction with bi-quadratic elements.

With a good grasp of the steady-state problem, attention then began to shift to the time-dependent case. As the transient problem came under closer scrutiny, it was observed that the time dependency introduces new numerical difficulties, such as additional numerical dispersion and the inaccurate representation of peak phase speed. It was quickly recognized that merely adding artificial diffusion using the techniques and optimal formulae derived for the steady-state problem generally produced overly diffusive solutions. Especially for difficult temporal discretizations (i.e. high Courant number), remedial methods originally derived for steady-state problems are incapable of addressing characteristics associated with the time dependency. To this end, a number of modellers have recently developed approaches which successfully bring the time dependency into more active consideration. Tezduyar and Hughes[11] and Tezduyar and Ganjoo[12] have modified the perturbation terms for SU/PG so that they are also dependent upon Courant number. Yu and Heinrich[13] have considered a Petrov–Galerkin approach which utilizes space–time finite elements and also includes the time dependency in their upwind modifying functions.

In this paper, we shall explore the use of upwind modifying functions which are two polynomial degrees higher than the basis functions used to approximate the variable. This approach is specifically geared towards improving the accuracy of time-dependent convection dominated problems and results in modifications to the transient term. This class of methods was proposed by Dick,[14] who examined a cubic modification to linear test functions in combination with the original standard quadratic modification. In this paper we extend this approach by developing a quartic modification for quadratic test functions. Through analysis and examples we shall examine the effects and roles of cubic and quadratic upwinding on linear elements and of quartic and cubic upwinding on quadratic elements.

GOVERNING EQUATIONS AND DISCRETIZATION DEVELOPMENT

We consider the one-dimensional form of the convective–diffusive transport equation:

$$\frac{\partial\phi}{\partial t}+u\frac{\partial\phi}{\partial x}=D\frac{\partial^2\phi}{\partial x^2} \tag{1}$$

on the interval Γ. We apply the finite element method to resolve this equation in space. Thus, we first develop a weighted residual form of (1) by weighting it with some test function w, integrating over the interval Γ and adding to this the weighted error incurred at at most one diffusive flux specified boundary point. Finally, integrating the diffusion term by parts and accounting for boundary relationships leads to the desired weak weighted residual form of (1):

$$\int_\Gamma\left\{\left(\frac{\partial\phi}{\partial t}+u\frac{\partial\phi}{\partial x}\right)w+D\frac{\partial\phi}{\partial x}\frac{\partial w}{\partial x}\right\}\mathrm{d}x=D\frac{\partial\phi}{\partial x}w\bigg|_{x_{\text{flux boundary}}} \tag{2}$$

The finite element approach is now applied by assuming that the variable ϕ and the weighting function w may be represented by C^0 piecewise continuous functions over a sequence of finite elements. This leads to a global system of differentially time-dependent equations:

$$\mathbf{M}\frac{\mathrm{d}\boldsymbol{\phi}}{\mathrm{d}t}+(\mathbf{A}+\mathbf{B})\boldsymbol{\phi}=\mathbf{P} \tag{3}$$

where $\boldsymbol{\phi}$ = vector of nodal unknowns, $\mathbf{M}$ = mass matrix, $(\mathbf{A}+\mathbf{B})$ = stiffness matrix, where $\mathbf{A}$ = convection matrix and $\mathbf{B}$ = diffusion matrix, and $\mathbf{P}$ = diffusive boundary flux loading vector.

The structure and characteristics of the matrices $\mathbf{M}$ and $(\mathbf{A}+\mathbf{B})$ are determined both by the degree of the C^0 trial functions used to represent the variable ϕ and by the degree of the C^0 test or weighting functions used to represent the weighting function w. For the Bubnov–Galerkin method, the same interpolating functions are used both as trial and test functions. Furthermore, the convective contribution to the stiffness matrix is non-symmetric. When convection is the dominant transport mechanism, the highly non-symmetric stiffness matrix results in the well-known oscillations. For the Petrov–Galerkin method, different orders of interpolation are used to represent the trial and test functions. Traditionally the test or weighting functions have been modified by a polynomial one degree higher or lower than the trial functions. For steady-state problems, these traditional Petrov–Galerkin methods have typically aimed to increase the degree of symmetry of the convection matrix. However, for time-dependent problems the mass matrix is significant and its role in obtaining accurate solutions should receive as much consideration as the convection matrix in any upwinding scheme.

The time discretization of (3) is implemented through the use of a Crank–Nicolson finite difference scheme to obtain

$$\left[\mathbf{M}+\frac{\Delta}{2}(\mathbf{A}^{n+1}+\mathbf{B}^{n+1})\right]\boldsymbol{\phi}^{n+1}=\left[\mathbf{M}-\frac{\Delta}{2}(\mathbf{A}^{n}+\mathbf{B}^{n})\right]\boldsymbol{\phi}^{n}+\frac{\Delta}{2}\mathbf{P}^{n+1}+\frac{\Delta}{2}\mathbf{P}^{n} \tag{4}$$

where $n+1$ and n represent the future and current time levels, and $\Delta =$ time step. This scheme is second order accurate in time.

DEVELOPMENT OF $N+1$ AND $N+2$ UPWIND WEIGHTING FUNCTIONS

(i) Linear elements with quadratic and cubic upwind weighting functions

Upwinding for linear finite elements was first developed by Christie *et al.*[1] for the steady-state form of the convection–diffusion equation. These investigators introduced a quadratic modifying function of the form

$$F_{\mathrm{QD}}(\xi)=\tfrac{3}{4}(1+\xi)(1-\xi) \tag{5}$$

This quadratic function is added to the standard linear trial functions ψ_1 and ψ_2 to produce the following upwinded weighting functions:

$$w_1=\psi_1-\alpha F_{\mathrm{QD}}(\xi) \tag{6a}$$

$$w_2=\psi_2+\alpha F_{\mathrm{QD}}(\xi) \tag{6b}$$

where α equals the quadratic upwinding coefficient. The standard trial functions, the quadratic modification function and the resulting upwinded test functions are all plotted in Figure 1(a). Since we designate the degree of the basis function interpolation as order N and we add an upwind bias which is one polynomial degree higher than the basis functions, we will classify this as an $N+1$ degree upwind method.

The coefficient α determines the amount of quadratic upwind bias. For the steady-state case it is possible to choose α such that the exact solution is obtained at the nodes. Christie *et al.*[1] showed that this optimal α is given by

$$\alpha_{\mathrm{opt}}=\coth\frac{\gamma}{2}-\frac{2}{\gamma} \tag{7}$$

where $\gamma=uh/D$ is the Peclet number. However, for the transient case, the truncation errors can not be eliminated with the specification of any amount of quadratic bias α and thus there is no precise definition of an optimal α. If we try to use (7) for transient problems, we are in effect contributing a low-order truncation error term which successfully suppresses oscillations but does so at the expense of degrading general solution accuracy.

In order to improve the solution to the time-dependent problem, Dick[14] proposed that, in addition to the quadratic bias, a cubic modifying function be added to the weighting functions. This cubic function has the form

$$F_{\mathrm{CU}}(\xi)=\tfrac{5}{8}\xi(\xi+1)(\xi-1) \tag{8}$$

and results in weighting functions of the form

$$w_1=\psi_1-\beta F_{\mathrm{CU}}(\xi) \tag{9a}$$

$$w_2=\psi_2+\beta F_{\mathrm{CU}}(\xi) \tag{9b}$$

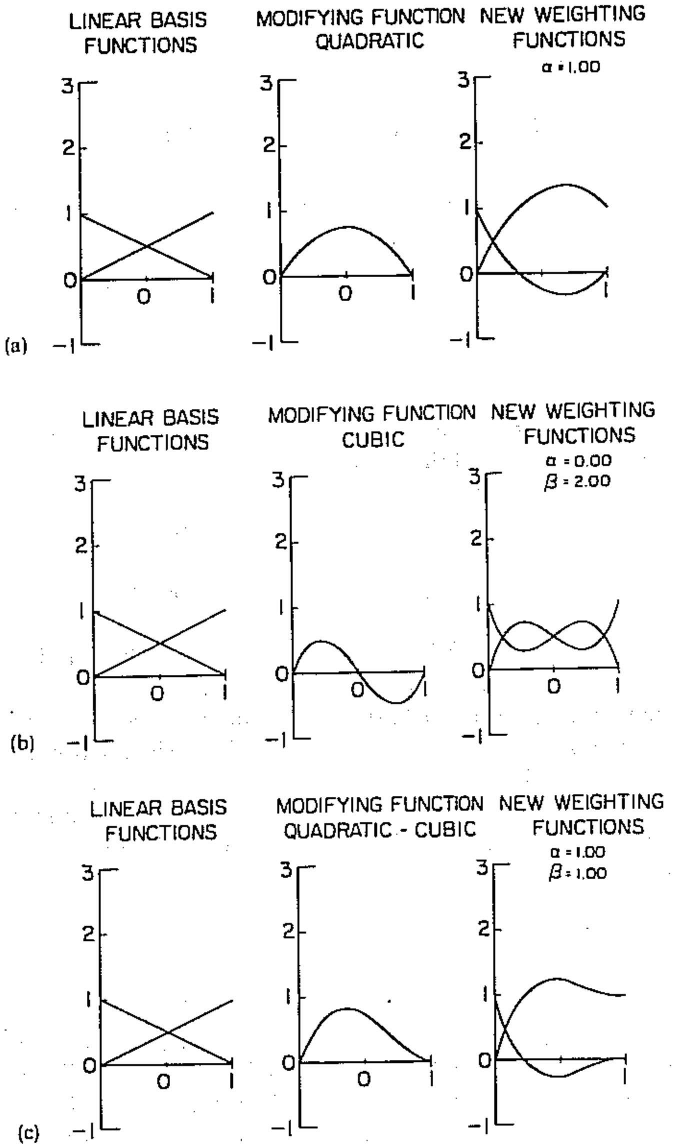

Figure 1. Trial functions, upwinded modifying function and resulting upwinded test functions for linear elements: (a) quadratic ($N+1$ degree) modification; (b) cubic ($N+2$ degree) modification; (c) combined quadratic and cubic modification

where β equals the amount of cubic upwind bias. This cubic modification together with the cubic upwinded weighting functions are shown in Figure 1(b). It is noted that these cubic upwinded weighting functions are symmetrical with respect to each other over the element, unlike the traditional quadratic upwinded weighting functions. Symmetrical weighting functions lead to flow direction invariant upwinding coefficients for the one-dimensional problem. Since the resulting upwind weighting functions are the trial functions modified by a polynomial which is two degrees greater, we will refer to this as an $N+2$ degree upwind method. We note that Dick actually considered only the combined quadratic and cubic biased weighting functions of the form

$$w_1=\psi_1-\alpha F_{\mathrm{QD}}(\xi)-\beta F_{\mathrm{CU}}(\xi) \tag{10a}$$

$$w_2=\psi_2+\alpha F_{\mathrm{QD}}(\xi)+\beta F_{\mathrm{CU}}(\xi) \tag{10b}$$

The combined modifying function and resulting weighting functions are shown in Figure 1(c). These combined weighting functions have lost their symmetry by including the quadratic bias. We shall consider both analytically and experimentally the effects of each modifying function $F_{\mathrm{QD}}(\xi)$ and $F_{\mathrm{CU}}(\xi)$, separately and in combination.

Using the weighting functions given in (10), we can readily show that the elemental matrices which combine to form the global matrices in (4) equal

$$\mathbf{M}^{(e)}=\frac{h}{6}\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}+\alpha\frac{h}{4}\begin{bmatrix}-1 & -1\\ 1 & 1\end{bmatrix}+\beta\frac{h}{24}\begin{bmatrix}-1 & 1\\ 1 & -1\end{bmatrix} \tag{11a}$$

$$\mathbf{A}^{(e)}=\frac{u}{2}\begin{bmatrix}-1 & 1\\ -1 & 1\end{bmatrix}+\alpha\frac{u}{2}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix} \tag{11b}$$

$$\mathbf{B}^{(e)}=\frac{D}{h}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix} \tag{11c}$$

where h = node to node distance. The leading term in each expression represents the elemental matrix derived from the standard Galerkin method, whereas the non-zero second and third terms respectively represent the additions due to the quadratic and cubic modifying functions. It is noted that the quadratic bias influences both the mass and convection matrices, while the cubic bias influences the mass matrix only. Hence the cubic modification affects only a time-dependent problem. Furthermore, the cubic upwind contribution to these elemental matrices would vanish in the case of a lumped mass formulation.

By assembling the element matrices in (11) and substituting into (4), global equations for any non-boundary node, j, are obtained:

$$\begin{aligned}
&\phi_{j-1}^{n+1}\left(\frac{1}{6\Delta}+\frac{\alpha}{4\Delta}+\frac{\beta}{24\Delta}-\frac{u}{4h}-\frac{\alpha u}{4h}-\frac{D}{2h^2}\right)\\
&\quad+\phi_{j-1}^{n}\left(-\frac{1}{6\Delta}-\frac{\alpha}{4\Delta}-\frac{\beta}{24\Delta}-\frac{u}{4h}-\frac{\alpha u}{4h}-\frac{D}{2h^2}\right)\\
&\quad+\phi_{j}^{n+1}\left(\frac{2}{3\Delta}-\frac{\beta}{12\Delta}+\frac{\alpha u}{2h}+\frac{D}{h^2}\right)\\
&\quad+\phi_{j}^{n}\left(-\frac{2}{3\Delta}+\frac{\beta}{12\Delta}+\frac{\alpha u}{2h}+\frac{D}{h^2}\right)\\
&\quad+\phi_{j+1}^{n+1}\left(\frac{1}{6\Delta}-\frac{\alpha}{4\Delta}+\frac{\beta}{24\Delta}+\frac{u}{4h}-\frac{\alpha u}{4h}-\frac{D}{2h^2}\right)\\
&\quad+\phi_{j+1}^{n}\left(-\frac{1}{6\Delta}+\frac{\alpha}{4\Delta}-\frac{\beta}{24\Delta}+\frac{u}{4h}-\frac{\alpha u}{4h}-\frac{D}{2h^2}\right)=0
\end{aligned} \tag{12}$$

We note that (12) can be related to the generalized hyperbolic difference equation given by Tezduyar and Hughes[11] by selecting $\alpha=2\tau u/h$, where τ equals a generalized SU/PG upwinding parameter, and selecting $\beta=24(r-1/6)$, where r equals a mass matrix element quadrature coefficient ($r=1/4$ corresponds to a 1-point Gauss rule and $r=1/6$ corresponds to the exact 2-point Gauss rule). Thus, while $N+1$ upwinding on linear one-dimensional elements corresponds to standard SU/PG, $N+2$ upwinding corresponds to selecting a different integration rule for the mass matrix. Furthermore, (12) is also in part similar to the discretized equation obtained for Yu and Heinrich's[13] quadratic-in-time–linear-in-space weights. However, while our cubic bias affects the transient term, their time weighting component modifies the convective term.

(ii) *Quadratic elements with cubic and quartic upwind weighting functions*

Heinrich and Zienkiewicz[4] first introduced an upwind formulation for use with quadratic interpolation. The cubic modification function these investigators developed is also given by equation (8). This cubic modification is added to the standard quadratic trial functions ψ_1, ψ_2 and ψ_3 as follows:

$$w_1=\psi_1-\alpha_c F_{\mathrm{CU}}(\xi) \tag{13a}$$

$$w_2=\psi_2+4\alpha_m F_{\mathrm{CU}}(\xi) \tag{13b}$$

$$w_3=\psi_3-\alpha_c F_{\mathrm{CU}}(\xi) \tag{13c}$$

where α_c equals the cubic bias for the corner nodes and α_m equals the cubic bias for the mid-element nodes. The standard quadratic trial functions, the cubic modification function $F_{\mathrm{CU}}(\xi)$ and the resulting upwinded weighting functions are all shown in Figure 2(a). The cubic upwinded weighting functions are non-symmetric over the element, as was the case for the quadratic upwind biased linear weighting functions. Furthermore, we note that, since the cubic weighting functions are only one polynomial degree greater than the quadratic trial functions, we can also classify this as an $N+1$ degree upwind method. Owing to the distinctly different nature of the corner and mid-element nodes the best values of α_m and α_c will not be the same. Optimal values for these coefficients have been determined for the steady-state case,[15] but again these do not carry over to the transient case.

We now introduce a new upwinding function which is two polynomial degrees higher than the quadratic trial functions being considered. The resulting quartic modifying function is given by

$$F_{\mathrm{QR}}(\xi)=\tfrac{21}{16}(-\xi^4+\xi^2) \tag{14}$$

and the resulting weighting functions are given by

$$w_1=\psi_1-\beta_c F_{\mathrm{QR}}(\xi) \tag{15a}$$

$$w_2=\psi_2+4\beta_m F_{\mathrm{QR}}(\xi) \tag{15b}$$

$$w_3=\psi_3-\beta_c F_{\mathrm{QR}}(\xi) \tag{15c}$$

where β_c and β_m respectively represent the quartic corner node bias and mid-element node bias. The pertinent functions are shown in Figure 2(b). These quartic biased quadratic upwind ($N+2$ degree) weighting functions are symmetrical with respect to each other over the element (w_1 with

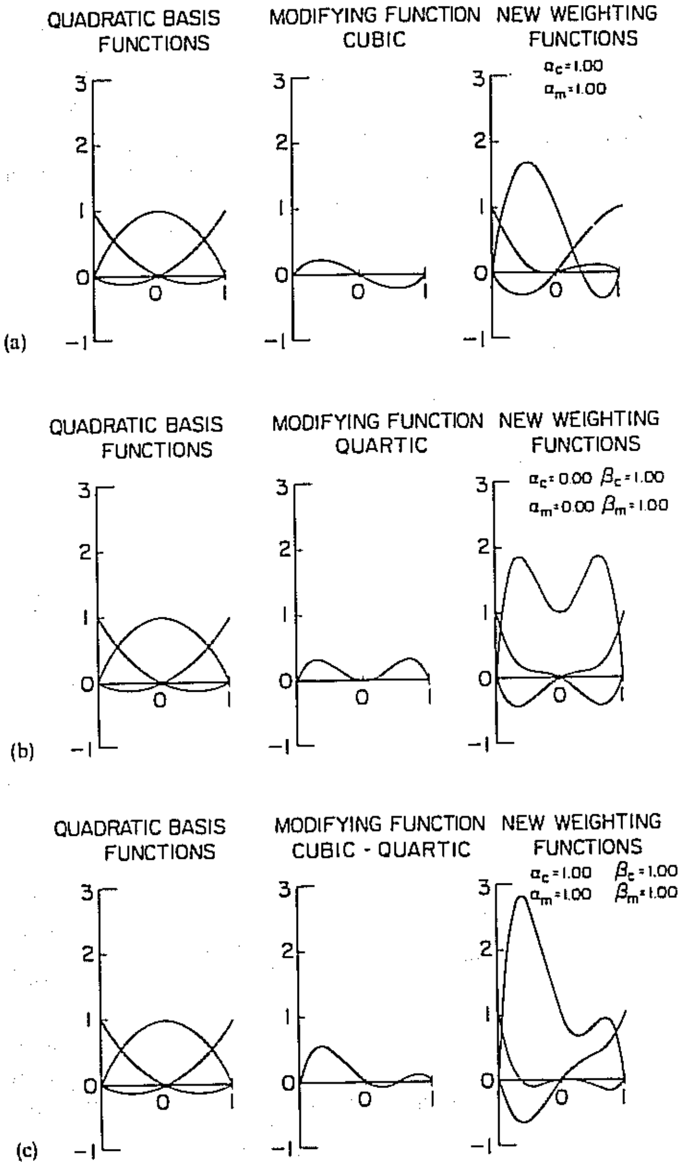

Figure 2. Trial functions, upwind modifying function and resulting upwinded test functions for quadratic elements: (a) cubic ($N+1$ degree) modification; (b) quartic ($N+2$ degree) modification; (c) combined cubic and quartic modification

respect to w_3 and w_2 with respect to itself), as was the case for the cubic biased linear ($N+2$ degree) weighting functions. We note that, for linear elements, quadratic and cubic upwinding are respectively classified as $N+1$ and $N+2$ degree upwinding, while for quadratic elements, cubic and quartic upwinding respectively represent the $N+1$ and $N+2$ degree bias. The combined cubic and quartic biased weighting functions may be expressed as

$$w_1=\psi_1-\alpha_c F_{\mathrm{CU}}(\xi)-\beta_c F_{\mathrm{QR}}(\xi) \tag{16a}$$

$$w_2=\psi_2+4\alpha_m F_{\mathrm{CU}}(\xi)+4\beta_m F_{\mathrm{QR}}(\xi) \tag{16b}$$

$$w_3=\psi_3-\alpha_c F_{\mathrm{CU}}(\xi)-\beta_c F_{\mathrm{QR}}(\xi) \tag{16c}$$

The resulting combined modifying and weighting functions are shown in Figure 2(c). We shall consider both cubic and quartic upwinding, separately and in combination.

The use of quadratic trial functions together with upwinded cubic/quartic test functions leads to the following elemental matrices:

$$\mathbf{M}^{(e)}=\frac{h}{15}\begin{bmatrix}4 & 2 & -1\\ 2 & 16 & 2\\ -1 & 2 & 4\end{bmatrix}+\frac{h}{120}\begin{bmatrix}-10\alpha_c & 0 & 10\alpha_c\\ 40\alpha_m & 0 & -40\alpha_m\\ -10\alpha_c & 0 & 10\alpha_c\end{bmatrix}$$

$$+\frac{h}{120}\begin{bmatrix}-9\beta_c & -24\beta_c & -9\beta_c\\ 36\beta_m & 96\beta_m & 36\beta_m\\ -9\beta_c & -24\beta_c & -9\beta_c\end{bmatrix} \tag{17a}$$

$$\mathbf{A}^{(e)}=\frac{u}{6}\begin{bmatrix}-3 & 4 & -1\\ -4 & 0 & 4\\ 1 & -4 & 3\end{bmatrix}+\frac{u}{120}\begin{bmatrix}20\alpha_c & -40\alpha_c & 20\alpha_c\\ -80\alpha_m & 160\alpha_m & -80\alpha_m\\ 20\alpha_c & -40\alpha_c & 20\alpha_c\end{bmatrix}$$

$$+\frac{u}{120}\begin{bmatrix}21\beta_c & 0 & -21\beta_c\\ -84\beta_m & 0 & 84\beta_m\\ 21\beta_c & 0 & -21\beta_c\end{bmatrix} \tag{17b}$$

$$\mathbf{B}^{(e)}=\frac{D}{6h}\begin{bmatrix}7 & -8 & 1\\ -8 & 16 & -8\\ 1 & -8 & 7\end{bmatrix}+\frac{D}{20h}\begin{bmatrix}7\beta_c & -14\beta_c & 7\beta_c\\ -28\beta_m & 56\beta_m & -28\beta_m\\ 7\beta_c & -14\beta_c & 7\beta_c\end{bmatrix} \tag{17c}$$

The cubic modifying function on quadratic elements affects the transient and convective terms, as did the quadratic upwinding on linear elements. Quartic upwinding on quadratic elements, unlike cubic upwinding on linear elements, affects all terms in our transport equation. Assembling these elemental equations leads to corner node equations which involve 5-point approximations, and mid-element node equations which involve 3-point approximations.

ANALYSIS OF $N+1$ AND $N+2$ DEGREE UPWINDING FOR LINEAR ELEMENTS

In order to better understand the effects that $N+1$ and $N+2$ degree upwinding have on the solution, we perform a truncation error and Fourier analysis for the linear element case.

(*i*) *Truncation error analysis*

The truncation error is readily computed by first Taylor series expanding the nodal difference equation (12) about ϕ_j^n. We then consider the form of the original partial differential equation (1) and perform sequential substitutions and/or take spatial derivatives such that equations of the following form are generated:

$$\frac{\partial\phi}{\partial t}=-u\frac{\partial\phi}{\partial x}+D\frac{\partial^2\phi}{\partial x^2} \tag{18a}$$

$$\frac{\partial^2\phi}{\partial t^2}=u^2\frac{\partial^2\phi}{\partial x^2}-2uD\frac{\partial^3\phi}{\partial x^3}+D^2\frac{\partial^4\phi}{\partial x^4} \tag{18b}$$

and

$$\frac{\partial^2\phi}{\partial x\,\partial t}=-u\frac{\partial^2\phi}{\partial x^2}+D\frac{\partial^3\phi}{\partial x^3} \tag{18c}$$

$$\frac{\partial^3\phi}{\partial x\,\partial t^2}=u^2\frac{\partial^3\phi}{\partial x^3}-2uD\frac{\partial^4\phi}{\partial x^4}+D^2\frac{\partial^5\phi}{\partial x^5} \tag{18d}$$

and so forth.

Substituting these expressions into the Taylor series expanded form of (12) will allow the grouping of terms with equal order spatial derivative of ϕ. Thus, collecting and re-arranging terms leads to the truncation error

$$\begin{aligned}\mathscr{T}={}&h^2\left[(2C^2-\beta)-\frac{1}{\gamma}(12\alpha)\right]\frac{u}{24}\frac{\partial^3\phi_j^n}{\partial x^3}\\&+h^3\left[-\frac{C}{2}(2C^2-\beta)+\alpha(1-C^2)+\frac{1}{\gamma}(2-6C^2+\beta+6C\alpha)\right]\frac{u}{24}\frac{\partial^4\phi_j^n}{\partial x^4}\\&+h^4\left[\left(-\frac{2}{15}+\frac{1}{3}C^2+\frac{3}{10}C^4-\frac{1}{12}\beta-\frac{1}{6}C^2\beta\right)-\alpha\frac{C}{2}(1-C^2)\right.\\&\left.+\frac{1}{\gamma}(-C+3C^2-C\beta-2\alpha)+\frac{1}{\gamma^2}(6C^2-6C\alpha)\right]\frac{u}{24}\frac{\partial^5\phi_j^n}{\partial x^5}\\&+\text{H.O.T.}\end{aligned} \tag{19}$$

where $C=u\Delta/h$ is the Courant number.

Let us first examine the case of pure convection with $D=0$ and $\gamma=\infty$. For this case, the standard Bubnov–Galerkin solution is such that $O(h)^2$ and $O(h)^3$ truncation terms are associated only with the time discretization whereas the $O(h)^4$ term reflects the combined effects of time and space discretization errors. Since for $\gamma=\infty$, α no longer appears in the $O(h)^2$ term in equation (19), there is no mechanism by which $N+1$ degree upwinding can eliminate this leading-order truncation term. However, it is readily apparent that we can eliminate the $O(h)^2$ truncation term through $N+2$ degree upwinding by selecting

$$\beta=2C^2 \tag{20}$$

This choice of β also forces the non-α-dependent portion of the $O(h)^3$ term to vanish. Thus the truncation error now appears as

$$\mathscr{T}=h^3[\alpha(1-C^2)]\frac{u}{24}\frac{\partial^4\phi_j^n}{\partial x^4}+h^4\left[\left(-\frac{2}{15}+\frac{1}{6}C^2-\frac{1}{30}C^4\right)-\alpha\frac{C}{2}(1-C^2)\right]\frac{u}{24}\frac{\partial^5\phi_j^n}{\partial x^5}+\text{H.O.T.} \quad (21)$$

For Courant numbers other than unity, we can only eliminate the $O(h)^3$ term in (21) by selecting $\alpha=0$. This yields a solution which is in general $O(h)^4$ accurate. Any combination of non-zero α and β still requires that $\beta=2C^2$ in order to eliminate the $O(h)^2$ term. However, the solution will in general be only $O(h)^3$ accurate. Finally, we note that at $C=1$ and with our selection of β, both the $O(h)^3$ and $O(h)^4$ truncation terms vanish regardless of the choice of α, yielding a solution which is at least $O(h)^5$ accurate.

For problems with combined convection and diffusion, $N+1$ upwinding does allow for the elimination of the $O(h)^2$ truncation term in equation (19) by setting

$$\alpha=\frac{C^2\gamma}{6} \quad (22)$$

However, this selection of α tends to increase the magnitude of the coefficient of the $O(h)^3$ term by up to several orders of magnitude. This increase in the $O(h)^3$ term becomes more substantial with higher γ values. In order to eliminate the $O(h)^2$ truncation term for $\gamma<\infty$ with $N+2$ degree upwinding, we keep the same value of β as for the pure convection case. In general, this choice of β also tends to markedly decrease the magnitude of the coefficient of the $O(h)^3$ term. This decrease becomes increasingly more significant with larger γ and C values.

Thus, our truncation error analysis indicates that $N+1$ degree upwinding does not offer an effective mechanism for eliminating truncation terms in time-dependent problems. It can not delete the leading-order (time) truncation term for purely convective cases, while for cases with diffusion it eliminates this $O(h)^2$ term (with a Peclet- and Courant-number-dependent α coefficient) at the cost of increasing the importance of the $O(h)^3$ term. $N+2$ upwinding, on the other hand, offers a very attractive mechanism for eliminating truncation terms. The choice $\beta=2C^2$ eliminates the leading-order (time) truncation term for both purely convecting and convective–diffusive cases. Furthermore, this non-Peclet-dependent selection of β eliminates the $O(h)^3$ term entirely for $\gamma=\infty$ while it reduces it substantially for $\gamma<\infty$. Finally, it is indicated that combining $N+1$ and $N+2$ degree upwinding still requires that $\beta=2C^2$ to eliminate the leading $O(h)^2$ truncation term but will re-introduce an $O(h)^3$ truncation term.

(ii) *Fourier analysis*

We now study the effects of $N+1$ and $N+2$ upwinding methods by applying standard Fourier analysis to the nodal difference equation (12). We substitute into (12)

$$\phi_j^n=\xi_\lambda^n\,e^{\sqrt{-1}\,2\pi jh/\lambda} \quad (23)$$

where $\xi_\lambda=$numerical amplification factor for a component with spatial wavelength λ. After some manipulation we obtain

$$\xi_\lambda=\frac{\text{Z1}}{\text{Z2}} \quad (24)$$

where

$$Z1=\left(8+2\beta+12\alpha C+24\frac{C}{\gamma}\right)\cos\left(\frac{2\pi h}{\lambda}\right)+\left(16-2\beta-12\alpha C-24\frac{C}{\gamma}\right)$$

$$+\sqrt{-1}\,(-12\alpha-12C)\sin\left(\frac{2\pi h}{\lambda}\right) \tag{25a}$$

$$Z2=\left(8+2\beta-12\alpha C-24\frac{C}{\gamma}\right)\cos\left(\frac{2\pi h}{\lambda}\right)+\left(16-2\beta+12\alpha C+24\frac{C}{\gamma}\right)$$

$$+\sqrt{-1}\,(-12\alpha+12C)\sin\left(\frac{2\pi h}{\lambda}\right) \tag{25b}$$

Taking into consideration the analytical amplification factor for a component with wavelength λ,

$$\xi_{a-\lambda}=\left[\cos\left(2\pi C\frac{h}{\lambda}\right)-\sqrt{-1}\sin\left(2\pi C\frac{h}{\lambda}\right)\right]\exp\left[-\left(\frac{C}{\gamma}\right)\left(2\pi\frac{h}{\lambda}\right)^2\right] \tag{26}$$

in addition to the number of time steps necessary to propagate a component through its entire wavelength, allows the definition of the following error criteria. First, the error in the amplification of a component of wavelength λ which occurs over one wavelength or the ratio of numerical to analytical damping may be expressed as

$$R=\left[\frac{|\xi_\lambda|}{|\xi_{a-\lambda}|}\right]^{(\lambda/h)(1/C)} \tag{27}$$

Second, the phase error of a component of wavelength λ after one complete wavelength equals:

$$\theta=\frac{\lambda}{h}\frac{1}{C}\tan^{-1}\left(\frac{\operatorname{Im}\xi_\lambda}{\operatorname{Re}\xi_\lambda}\right)-2\pi \tag{28}$$

The origin of the wiggle problem in linear equations is the poor phase propagation behaviour of the numerical scheme. Thus, it is natural that our goal in using any upwinding scheme should be to substantially improve the phase error while not adversely affecting the numerical to analytical damping ratio. To this end, we examine the required $N+1$ degree upwinding coefficient α to produce perfect phase behaviour (i.e. $\theta=0$) for a given λ/h ratio at $\gamma=\infty$. Figure 3 shows that the required α values vary widely, depending on the λ/h value. However, an optimal range of α values is indicated by the shaded region. This range of α values tends to significantly improve the phase behaviour at the short wavelengths while not deteriorating the phase properties at the intermediate and/or longer wavelengths. Table I compares the damping ratio R and phase lag θ of the standard solution and the $N+1$ degree upwinded solution obtained by using an α value for each C value from the middle of the optimal range indicated in Figure 3. We note that $N+1$ degree upwinding can be made very effective in improving phase behaviour, especially at low Courant numbers. However, along with the dramatic reduction in θ comes a substantial decrease in R over a relatively broad range of λ/h values such that the $N+1$ degree upwinded solution is markedly overdamped at short and even intermediate wavelengths. It is not generally perceived as desirable to introduce excessive damping to either eliminate or control wiggles.[16] Reductions in α large

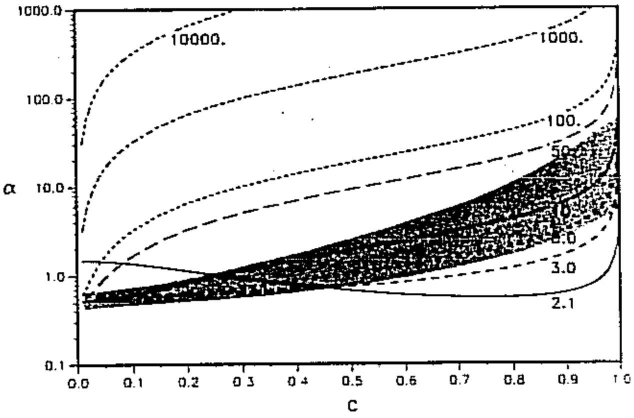

Figure 3. α values required to produce perfect phase behaviour ($\theta=0$) for various λ/h ratios at $\gamma=\infty$

Table I. Comparison of damping ratio and phase lag behaviour for various schemes at $\gamma=\infty$

		Standard Bubnov–Galerkin		$N+1$ Petrov–Galerkin		$N+2$ Petrov–Galerkin		Combined $N+1/N+2$ Petrov–Galerkin	
C	λ/h	R	θ	R	θ	R	θ	R	θ
0·1	2·6667	1·0000	−1·9176	0·0425	−0·4661	1·0000	−1·7240	0·4960	−1·6631
	4·0	1·0000	−0·2944	0·4956	−0·0035	1·0000	−0·1964	0·8665	−0·1855
	8·0	1·0000	−0·0175	0·9288	−0·0015	1·0000	0·0045	0·9869	0·0050
	50·0	1·0000	−0·0001	0·9997	−0·0001	1·0000	0·0004	1·0000	0·0004
0·5	2·6667	1·0000	−2·1311	0·0704	1·0675	1·0000	−1·2786	0·5527	−1·2014
	4·0	1·0000	−0·5429	0·4391	0·1257	1·0000	−0·1371	0·8996	−0·1282
	8·0	1·0000	−0·0927	0·8866	−0·0437	1·0000	0·0032	0·9902	0·0036
	50·0	1·0000	−0·0021	0·9995	−0·0020	1·0000	0·0003	1·0000	0·0003
0·9	2·6667	1·0000	−2·5143	0·7596	0·4458	1·0000	−0·2744	0·8527	−0·2387
	4·0	1·0000	−1·0054	0·8041	0·1913	1·0000	−0·0279	0·9755	−0·0255
	8·0	1·0000	−0·2563	0·8988	0·0192	1·0000	0·0007	0·9976	0·0008
	50·0	1·0000	−0·0067	0·9971	−0·0055	1·0000	0·0001	1·0000	0·0001

enough to significantly limit this adverse numerical damping lead to phase behaviour which is again similar to the standard solution. Thus $N+1$ degree upwinding is not a viable approach to improve the overall quality of the solution.

Figure 4 shows the $N+2$ degree upwinding coefficients β required to produce perfect phase behaviour at various λ/h values at $\gamma=\infty$. The β values which ensure that $\theta=0$ for the range of λ/h values converge to $\beta=2{\cdot}0$ as $C\rightarrow 1{\cdot}0$. Furthermore, the curves for higher λ/h values lie increasingly closer and, in the limit as $\lambda/h\rightarrow\infty$, merge into the curve defined by $\beta=2C^2$. We recall that this is the value of β which eliminates the $O(h)^2$ and $O(h)^3$ truncation terms for the case $\gamma=\infty$. The fact that the curves for lower λ/h values do not coincide with the upper limit curve can be explained by considering the magnitude of the spatial derivatives and their associated coefficients in the

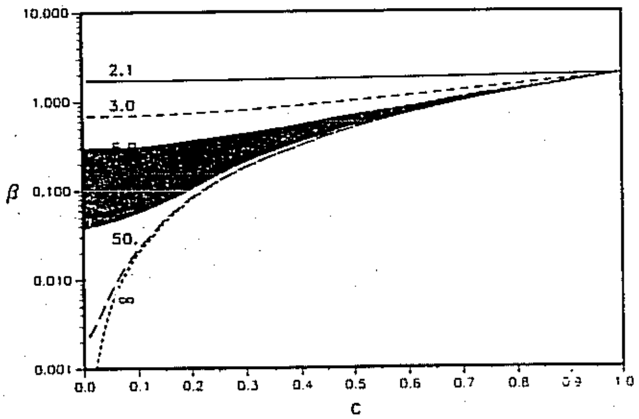

Figure 4. β values required to produce perfect phase behaviour ($\theta = 0$) for various λ/h ratios at $\gamma = \infty$

truncation series (19). The shaded region in Figure 4 defines the overall optimal range of β values for a given C which substantially reduces phase lag at the shorter wavelengths while not developing a significant phase lead at intermediate and/or long wavelengths. Table I compares the damping ratio and phase lag of the $N+2$ upwinded solution which uses β values defined in the middle of the shaded region in Figure 4 to the $N+1$ upwinded and standard solutions. $N+2$ degree upwinding increasingly improves the overall phase behaviour as C increases. In fact, at $C = 1{\cdot}0$ the phase behaviour is perfect. However, this improvement in phase still leads to perfect damping behaviour, as is also the case for the standard Bubnov–Galerkin method with a Crank–Nicolson time scheme. Finally, we note that for $C > 1$, $N+2$ degree upwinding continues to be able to eliminate truncation terms in a stable manner. However, the required values of β exceed 2, which will adversely affect the properties of the system matrix which must be solved.

Combining $N+1$ and $N+2$ degree upwinding again introduces artificial damping into the solution. The amount of α specified affects the extent of damping over the λ/h range and large values lead to damping behaviour similar to that of $N+1$ degree upwinding alone. However, applying small α values in conjunction with the optimal β values used for $N+2$ degree upwinding alone does allow for the introduction of controlled damping which only affects the very short wavelengths and not the intermediate or longer wavelength range. This is illustrated in Table I where α equals $0{\cdot}1$ and the same β values are used as for the pure $N+2$ degree upwinding case. We also note that this limited introduction of α improves phase behaviour slightly.

For cases with $\gamma < \infty$, $N+1$ degree upwinding is still able to substantially reduce phase errors. However, the α values necessary are highly γ and C number dependent. Furthermore, the technique.remains overly diffusive over almost the entire range of convection dominated γ values. Only at very low γ values, such as 2, is $N+1$ degree upwinding able to improve phase behaviour (mostly in the intermediate λ/h range) without introducing excessive numerical diffusion. On the other hand, $N+2$ degree upwinding continues to be able to effectively improve phase behaviour while not adversely affecting the damping ratio over most of the convection dominated γ range. The optimal β values appear only weakly dependent on γ, although the relative improvement achieved with $N+2$ upwinding decreases at very low γ. Furthermore, the effect of β decreases with lower C values as in the pure convection case.

Thus, the results of Fourier analysis indicate that, while both $N+1$ and $N+2$ degree upwinding can improve phase behaviour for convection dominated problems, only $N+2$ degree upwinding can do so without introducing excessive artificial damping. $N+2$ degree upwinding is especially effective in improving the quality of the solution for difficult temporal discretization cases. Furthermore, combined $N+1$ and $N+2$ degree upwinding also degrades the damping behaviour. However, very small amounts of α in conjunction with the optimal β values do allow for controlled damping of only the shortest wavelengths.

NUMERICAL EXAMPLES

In order to illustrate the performance characteristics of the previously described upwinding techniques, we now examine a one-dimensional example problem with a Gaussian plume of standard deviation $\sigma=264$ travelling in pure convection ($D=0$, $u=0{\cdot}5$). The node to node distance remains constant at $h=200$ for all examples shown. All numerical solutions are compared to the analytical solution at $t=9600$, although a number of different time steps are used such that a range of Courant numbers can be examined. This allows us to discern the difference between time and spatial discretization difficulties. The example cases are examined using the standard or Bubnov–Galerkin method in addition to $N+1$ and $N+2$ degree upwind or Petrov–Galerkin methods, both separately and in combination. The Crank–Nicolson scheme is the only time discretization scheme considered. A variety of error criteria are listed in the Appendix for each case.

(i) *Linear elements with quadratic and cubic upwind weighting functions*

In this sub-section we shall examine the use of Lagrange linear interpolation for the trial functions in conjunction with linear test functions with added quadratic ($N+1$ degree) and/or cubic ($N+2$ degree) modifying functions.

Figure 5(a) shows the standard Galerkin solution after 100 time steps of $\Delta=96$. The Courant number, $C=0{\cdot}24$, is relatively low. This solution is plagued by trailing oscillations and a drop in peak concentration. The first attempt to improve this result is shown in Figure 5(b), for which the classical quadratic upwind method has been used. The quadratic bias of $\alpha=0{\cdot}7$ manages to improve phase behaviour and for the most part eliminates trailing oscillations, but does so at the expense of the introduction of artificial diffusion which causes a further drop in the peak. No optimal value of α exists for this problem and the value of α chosen will always represent a trade-off between overall accuracy, peak accuracy and oscillation amplitude.

Figure 5(c) illustrates that the use of cubic upwinding results in a much better solution compared to either the standard Galerkin or the quadratically upwinded solution. Both the amplitude and symmetry of the distribution have substantially improved owing to the improved phase properties of $N+2$ upwinding. However, oscillations still follow the plume although their amplitudes have been reduced and their character changed as compared to the standard Galerkin results (shown in Figure 5(a)). We note that, when experimenting with choices of β, an optimal value becomes apparent when the majority of the error criteria listed in the Appendix are minimized. For this case the optimal β was determined to be $0{\cdot}30$. Increasing β beyond $0{\cdot}30$ eventually results in a leading dip, an overall phase lead of the plume itself and a deterioration of plume symmetry owing to the phase lead at intermediate wavelengths becoming important.

Finally, Figure 5(d) illustrates the combined effects of quadratic and cubic upwinding. The addition of quadratic bias to the cubic bias eliminates most of the remaining oscillations but at the cost of damping the peak and a decrease in overall accuracy. The symmetry of the plume achieved with the cubic bias has been retained. Again the optimal value of β is readily established and is unaffected by the addition of the quadratic bias. However, no optimal value of α can be established

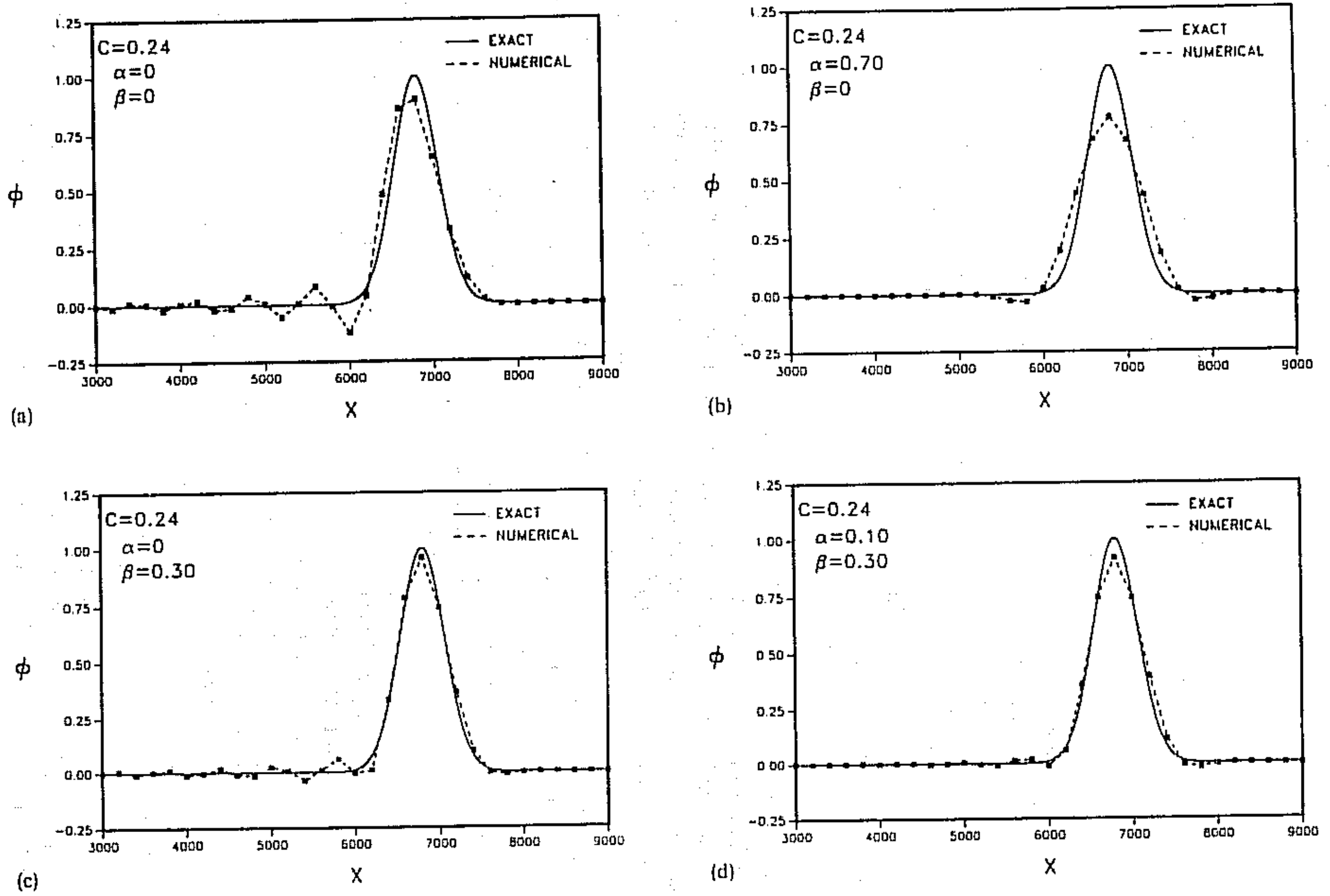

Figure 5. Low Courant number ($C=0{\cdot}24$) example of the pure convection of a 1-D Gaussian plume using linear elements ($D=0$, $u=0{\cdot}5$, $h=200$, $\Delta=96$): (a) standard Bubnov–Galerkin solution; (b) quadratic upwinded ($N+1$ degree) Petrov–Galerkin solution; (c) cubic upwinded ($N+2$ degree) Petrov–Galerkin solution; (d) combined quadratic-cubic upwinded Petrov–Galerkin solution

since the use of any amount of quadratic bias again represents a trade-off between the existence of oscillations and the addition of artificial damping. The value of α used for the case shown in Figure 5(d) is equal to 0·10, substantially reduced from that used in the quadratic upwinding only case. However, we deem that the best overall plume results when α is set to zero and only cubic upwinding is applied.

We now consider the same problem but decrease the number of time steps to 30 and increase the time step to $\Delta = 320$ such that the plume is convected to the same position, but at a higher Courant number, $C = 0{\cdot}80$. Thus we have increased the difficulty of the time discretization. This is reflected in the results for the standard Galerkin solution shown in Figure 6(a). The overall quality of the solution has substantially deteriorated from the previously considered lower Courant number case which was solved with the standard Galerkin method (Figure 5(a)). Specifically, the amplitude of the trailing oscillations has increased, the peak is further depressed and the plume itself now exhibits a substantial phase lag. Figure 6(b) again indicates that $N+1$ degree upwinding can improve phase properties reasonably well. However, this is again at the cost of substantially deteriorating peak accuracy and gradients. While lowering α does allow for slightly less damping of the peak (although the peak will not be any better than the already excessively lowered standard peak in Figure 6(a)), the solution again becomes overall lagged, loses its symmetry and is left with a single large trailing dip.

Cubic upwinding, on the other hand, does a superb job at improving the accuracy of the solution, as is shown in Figure 6(c). Peak amplitude and overall plume phase have been dramatically improved and, furthermore, oscillations have been essentially eliminated. The optimal value of β at this Courant number is readily determined to be 1·37. It is noted that the overall quality of this high Courant number solution is much better than the optimal cubic biased solution at low Courant number (case in Figure 5(c)). The effect of combined quadratic and cubic bias is shown in Figure 6(d). Although the addition of a small amount of α does slightly improve the solution at the foot of the plume and reduce the few remaining oscillations which were seen in Figure 6(c), it is at the cost of adding artificial damping, as is reflected by the lowered peak. Again the optimal value of β remains the same as when cubic upwinding is applied by itself and no optimal value of α can be defined.

When the Courant number increases to unity, cubic upwinding with $\beta = 2{\cdot}00$ yields the exact solution at the nodes. Furthermore, the solution for this case is entirely insensitive to α as long as the optimal value of $\beta = 2{\cdot}00$ is retained.

Thus, all the essential features of $N+1$ and $N+2$ degree upwinding predicted in the analysis section have been well verified. Although quadratic upwinding on linear elements can be made effective in improving phase behaviour, it does so at the cost of introducing excessive artificial damping. The optimal cubic upwinded solution gives excellent phase behaviour without introducing any damping. This solution is always dramatically better than either the standard Galerkin or any quadratic upwinded solution. Extensive numerical experimentation over a Courant number range of $0{\cdot}1 \leqslant C \leqslant 1{\cdot}0$ and using a variety of plume widths indicates that the optimal cubic bias β increases with C in the manner shown in Figure 7. These optimal experimental β values essentially correspond to the curve shown in Figure 4 which defines β such that $\theta = 0$ for $\lambda/h = 5{\cdot}0$. The cubic biased solution improves as the Courant number increases with overall accuracy, peak amplitude and overall phase getting better and oscillation amplitudes decreasing. Thus both temporal and, when C is high, spatial accuracy are enhanced. The solution at $C = 1{\cdot}0$ is very special in that a perfect numerical solution can be attained by using the correct amount of cubic bias ($\beta = 2{\cdot}0$) and, furthermore, in that the addition of any amount of quadratic bias has absolutely no effect on the solution. In general, combined $N+1$ and $N+2$ degree upwinding introduces artificial damping. However, when very small values of α are applied with

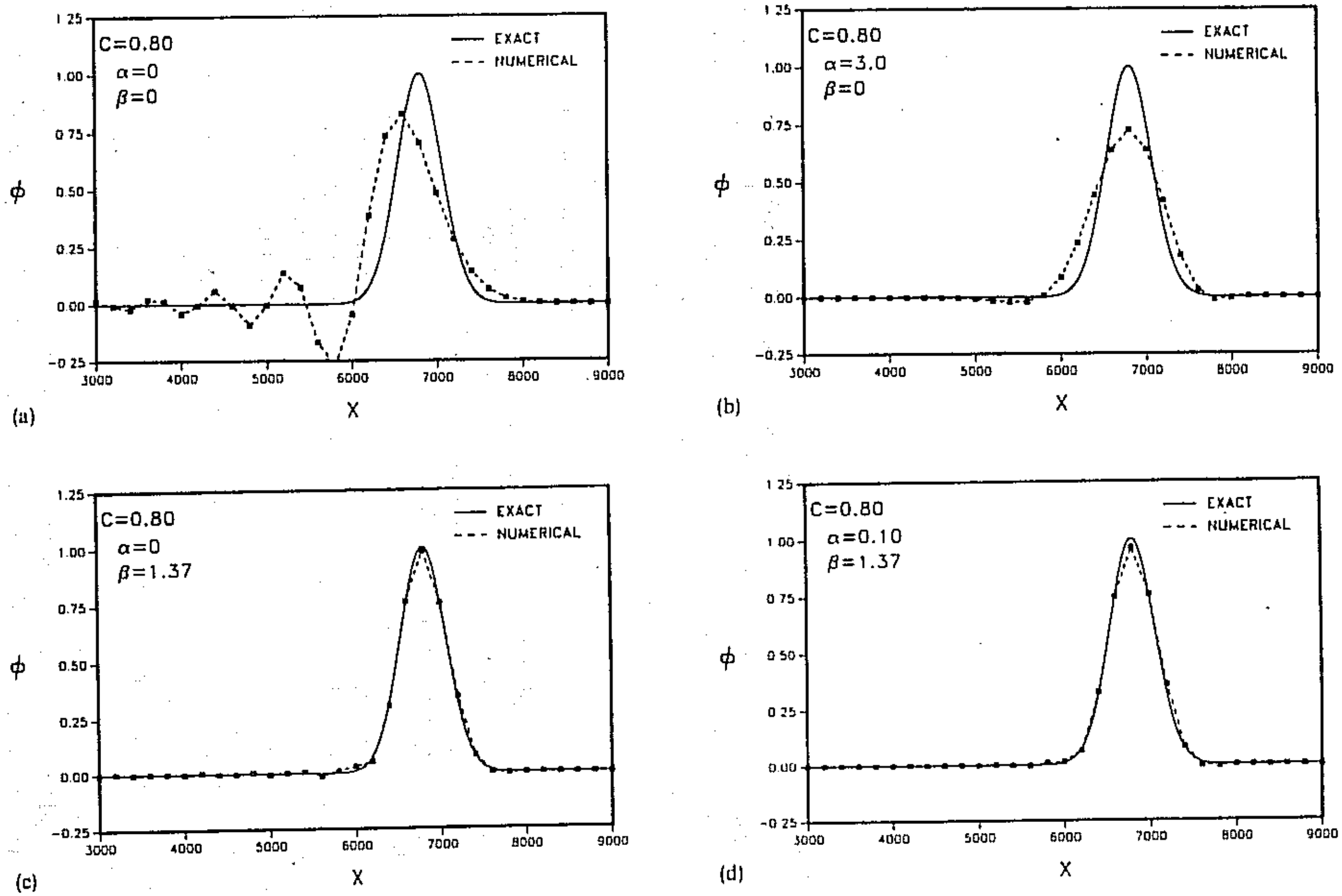

Figure 6. High Courant number ($C = 0{\cdot}8$) example of the pure convection of a 1-D Gaussian plume using linear elements ($D = 0$, $u = 0{\cdot}5$, $h = 200$, $\Delta = 320$): (a) standard Bubnov–Galerkin solution; (b) quadratic upwinded ($N + 1$ degree) Petrov–Galerkin solution; (c) cubic upwinded ($N + 2$ degree) Petrov–Galerkin solution; (d) combined quadratic–cubic upwinded Petrov–Galerkin solution

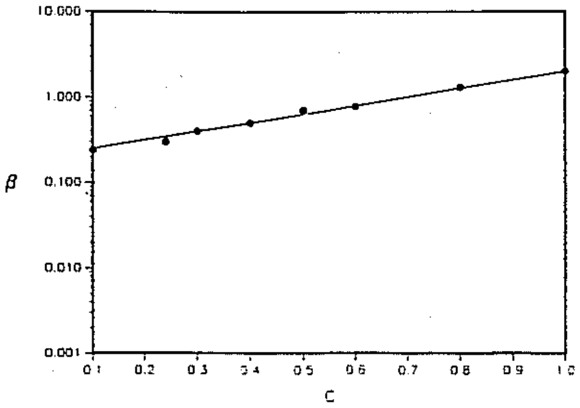

Figure 7. Optimal cubic upwinding on linear elements factor β, found through numerical experimentation, as a function of Courant number

the previously defined optimal β values, this damping can be effectively limited. Nonetheless, if the problem is such that significant energy exists at these short wavelengths, overall solution and peak accuracy will deteriorate. Finally, the introduction of physical diffusion into our test problems indicates that the optimal β values found for our pure convection cases are not very sensitive to Peclet number γ and that these values continue to improve the solution over almost the entire convection dominated γ range.

(*ii*) *Quadratic elements with cubic and quartic upwind weighting functions*

We now examine the use of Lagrange quadratic interpolation for the trial functions in conjunction with quadratic test functions which have an added cubic ($N+1$ degree) bias in addition to our newly developed quartic ($N+2$ degree) bias.

First, we re-examine the low Courant number problem, with $C=0{\cdot}24$, previously considered with linear elements. The standard Galerkin solution with quadratic elements is excellent, as is illustrated in Figure 8(a). The plume is very well represented and only a few very small oscillations lag the plume. Adding a small amount of cubic bias does not really improve the overall solution. In fact, Figure 8(b) shows that the peak is pushed up slightly beyond the analytical value. We deem that the optimal values of cubic upwinding for both the mid-element node α_m and for the corner node α_c are zero.

A small amount of quartic upwinding, Figure 8(c), is able to slightly enhance the quality of the already very good standard solution, mainly by slightly improving the symmetry of the solution. For this case the optimal quartic corner node bias is $\beta_c=0{\cdot}15$ and the optimal mid-element node bias is $\beta_m=0{\cdot}075$. Finally, the combined use of cubic and quartic bias does not improve the solution. In fact, the tendency of the cubic bias to push up the peak too far is again seen in Figure 8(d) and thus the optimal cubic bias values are zero. It is noted that the optimal quartic bias values remained the same regardless of the amount of cubic bias introduced.

The same problem is now computed using a high Courant number, $C=0{\cdot}8$. The more difficult time discretization has substantially degraded the quality of the solution, as is indicated in Figure 9(a). The peak is down and substantially lags the exact solution. In addition, a number of rather large oscillations follow behind the plume. Figure 9(b) illustrates that, while cubic upwinding on

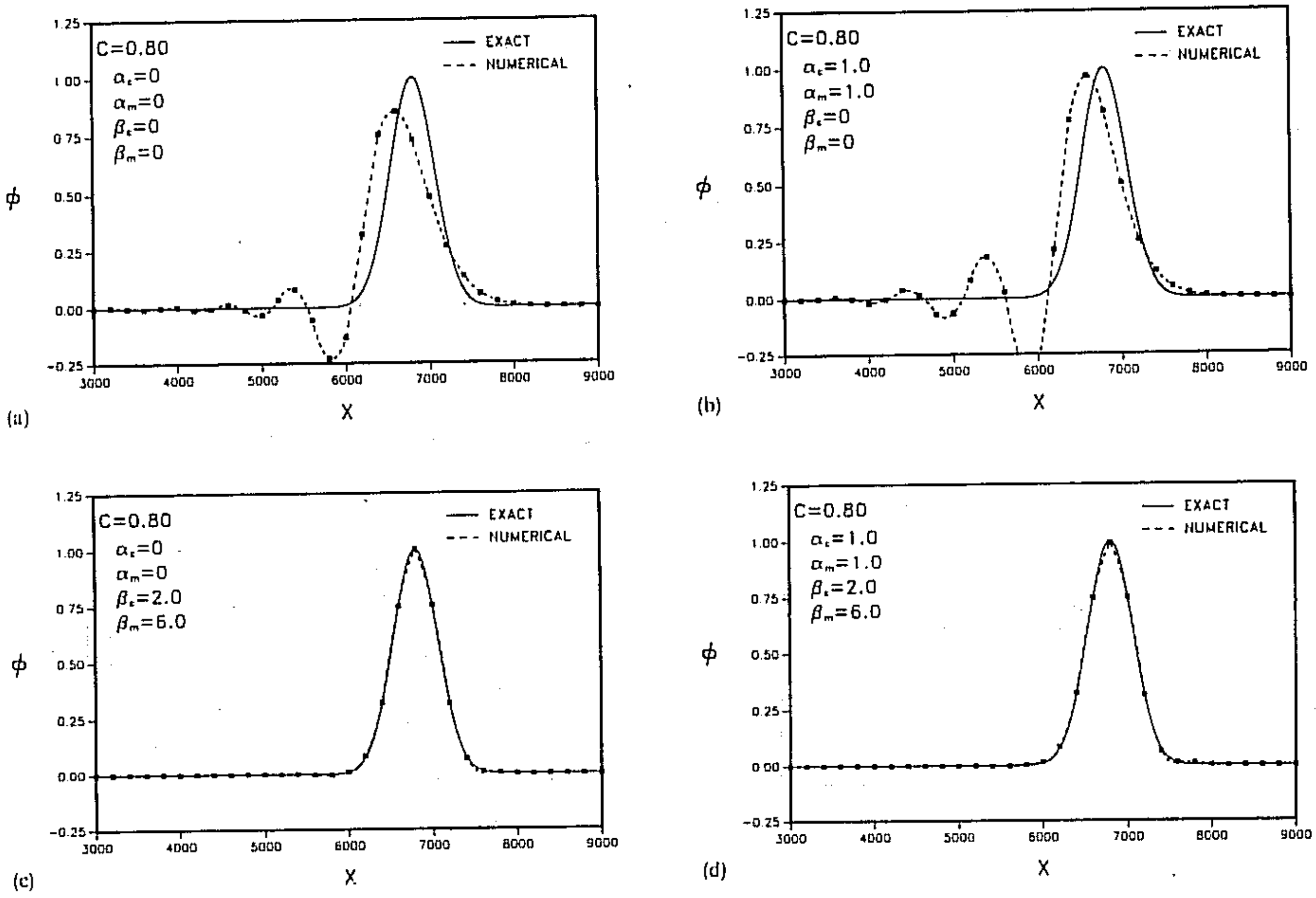

Figure 9. High Courant number ($C = 0{\cdot}8$) example of the pure convection of a 1-D Gaussian plume using quadratic elements ($D = 0$, $u = 0{\cdot}5$, $h = 200$, $\Delta = 320$): (a) standard Bubnov–Galerkin solution; (b) cubic upwinded ($N + 1$ degree) Petrov–Galerkin solution; (c) quartic upwinded ($N + 2$ degree) Petrov–Galerkin solution; (d) combined cubic–quartic upwinded Petrov–Galerkin solution

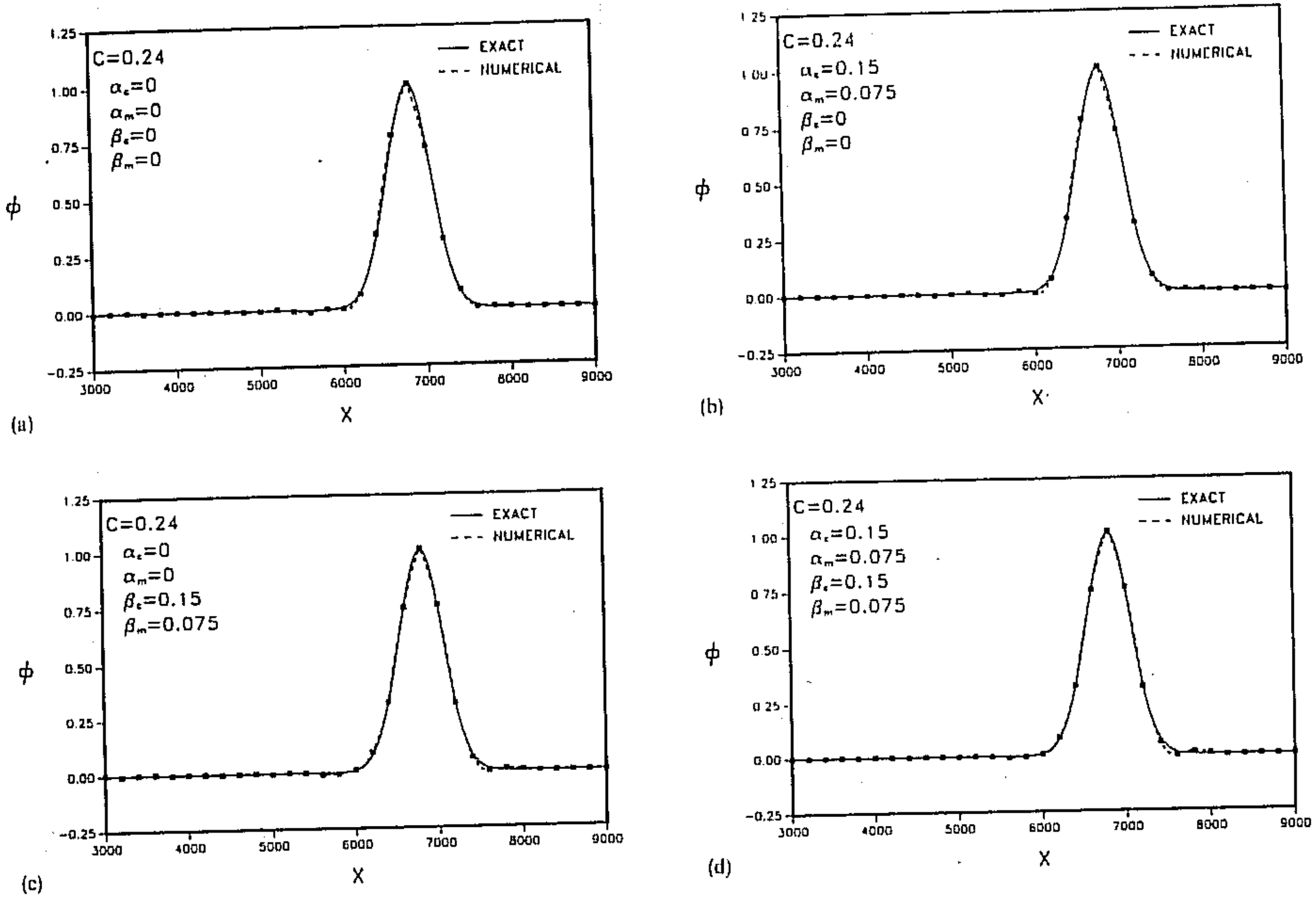

Figure 8. Low Courant number ($C=0{\cdot}24$) example of the pure convection of a 1-D Gaussian plume using quadratic elements ($D=0$, $u=0{\cdot}5$, $h=200$, $\Delta=96$): (a) standard Bubnov–Galerkin solution; (b) cubic upwinded ($N+1$ degree) Petrov–Galerkin solution; (c) quartic upwinded ($N+2$ degree) Petrov–Galerkin solution; (d) combined cubic–quartic upwinded Petrov–Galerkin solution

quadratic elements tends to push the peak up, it does not improve the overall solution at all. In fact, increasing α_m and α_c excessively can lead to unstable solutions.

Quartic upwinding, Figure 9(c), restores a truly excellent solution. The overall quality of this solution at a high Courant number is much better than the best solution at a low Courant number, which was already a very good solution. The optimal quartic upwinding coefficients are again readily established, and for this Courant number, $\beta_c=2{\cdot}00$ and $\beta_m=6{\cdot}00$. Finally, as is seen in Figure 9(d), the combined use of cubic and quartic upwinding has little effect although it again degrades the general solution quality as compared to the use of quartic upwinding alone.

Through numerical experimentation we have established optimal values of quartic mid-element node and corner node bias. Optimal values depend only on C and are shown in Figure 10. For time-dependent problems, the optimal amount of cubic upwinding on quadratic elements is always equal to zero, both when used in conjunction with quartic upwinding and by itself. In general, standard quadratic element solutions at low C are already very good and only a very small amount of quartic bias is necessary to achieve the optimal upwinded solution. Thus, quadratic interpolation is by itself able to effectively handle the spatial discretization for problems with relatively easy time discretizations without really necessitating the use of quartic upwinding. At higher C values, the time discretization difficulties cause the standard solution to deteriorate substantially. Quartic upwinding effectively returns an excellent solution and in fact the optimal quartic upwinded solution improves as the Courant number increases. The optimal quartic bias values increase along with increasing Courant number. We note that quartic upwinding in conjunction with quadratic elements in general produces better results than cubic upwinding on linear elements. Above $C=0{\cdot}8$, quartic upwinding is no longer effective at improving the quality of the solution. Finally, we note that problems which include diffusion indicate that optimal quartic bias values do not depend on the Peclet number γ.

CONCLUSIONS

We have examined the solution of the transient convection dominated transport problem using $N+1$ and $N+2$ degree Petrov–Galerkin methods to resolve the spatial dependency and a Crank–Nicolson scheme to resolve the time dependency. Although traditional $N+1$ degree upwinding methods, which employ test functions one degree higher than the trial functions, have

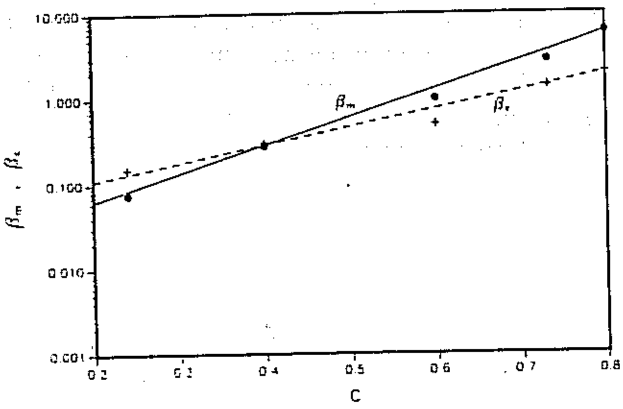

Figure 10. Optimal quartic upwinding on quadratic elements, found through numerical experimentation, as a function of Courant number. Values for both mid-element node bias β_m and corner node bias β_c are shown

been quite successful in accurately solving steady-state problems, they are incapable of improving the solution for difficult time problems. Quadratic upwinding on linear elements eliminates the spurious oscillations in the solution at the cost of adding excessive artificial damping. Cubic upwinding on quadratic elements does not generally improve the quality of any solution and, in fact, can even lead to unstable solutions. In general, $N+1$ degree upwinding is unable to produce a well defined optimal solution and at best yields some type of compromise solution.

$N+2$ degree Petrov–Galerkin methods, which use test functions two degrees higher than the trial functions, show a remarkable ability to improve both spatial and especially temporal accuracy. Cubic upwinding on linear elements leads to dramatically improved results as compared to both the standard Bubnov–Galerkin solution and the $N+1$ degree Petrov–Galerkin solution. Although optimal cubic upwinded solutions at low C values still include some small oscillations, the peak amplitude and phase have clearly improved. Furthermore, these solutions improve as the Courant number approaches unity and a perfect numerical solution is obtained at $C=1$. It is stressed that these cubic upwinded solutions, unlike quadratic upwinding, do not add any artificial damping to the solution. Quartic upwinding on quadratic elements yields truly excellent results over the entire Courant number range. In fact, these quartic upwinded solutions are almost perfect in all aspects. Only a very small amount of quartic upwinding is required at low C values since the standard quadratic solution is already very good. The quality of the optimal quartic upwinded solution also improves as the Courant number increases up to $C=0{\cdot}8$. Furthermore, quartic upwinding also yields solutions which exhibit perfect analytical damping. The optimal $N+2$ degree upwinded solution for both linear and quadratic elements is always well defined and the associated $N+2$ degree upwinding coefficients are readily established. These optimal upwinding coefficients do not vary with the amount of diffusion for convection dominated problems.

In general, we feel that for time-dependent linear problems, $N+2$ degree upwinding on linear elements should be applied in and of itself and not in conjunction with the traditional $N+1$ degree upwinding which always adds some amount of artificial damping. This is because the small remaining wiggles in the solution are manifestations of the gradients of the distribution being such that significant energy exists at λ/h ratios within a range where these components of the solution are still not sufficiently well propagated. An appropriate amount of mesh refinement will allow for the improvement of phase behaviour and thus will eliminate wiggles without resorting to damping the peak and degrading overall solution quality. However, for non-linear problems where the governing equations may permit the physical transfer of energy to the smallest resolvable scales, controlled numerical damping may be useful when physical damping is not sufficient at these lowest scales. It is noted that a reasonable amount of mesh refinement is unable to solve this problem. Thus, for non-linear problems, a small amount of $N+1$ degree upwinding together with $N+2$ degree upwinding may be appropriate for linear elements. Finally, for quadratic elements, $N+2$ degree upwinding should be applied by itself for both linear and non-linear problems since for this case $N+1$ degree upwinding does not offer the possibility of controlled damping at low λ/h ratios.

ACKNOWLEDGEMENT

This research was sponsored in part by NSF under grant EET–8718436.

APPENDIX

To quantify how well the numerical solutions represent the exact solution, a table of errors is presented for the example problems based on the following definitions.[17]

Error E1: Integral measure of the overall error of the numerical solution.

$$\mathrm{E1}=\frac{1}{m(t)}\left[\int_0^L(\phi^{\mathrm{num}}(x,t)-\phi^{\mathrm{ex}}(x,t))^2\,\mathrm{d}x\right]^{1/2}$$

Value for exact solution = 0·0

Error E2: Discrete measure of the overall error of the numerical solution.

$$\mathrm{E2}=\frac{1}{m(t)}\left[\sum_i(\phi_i^{\mathrm{num}}(t)-\phi^{\mathrm{ex}}(x_i,t))^2\right]^{1/2}$$

Value for exact solution = 0·0

Error E3: Point measure of the artificial damping of the numerical solution (peak depression).

$$\mathrm{E3}=\left|\frac{\phi_{\max}^{\mathrm{ex}}(t)-\phi_{\max}^{\mathrm{num}}(t)}{\phi_{\max}^{\mathrm{ex}}(t)}\right|$$

Value for exact solution = 0·0

Error E4: Point measure of the maximum spurious oscillation in the numerical solution.

$$\mathrm{E4}=\left|\frac{\phi_{\max,\mathrm{neg}}^{\mathrm{num}}(t)}{\phi_{\max}^{\mathrm{ex}}(t)}\right|$$

Value for exact solution = 0·0

Error E5: Point measure of the phase shift introduced in the numerical solution.

$$\mathrm{E5}=\frac{x_{\max}^{\mathrm{ex}}(t)-x_{\max}^{\mathrm{num}}(t)}{x_{\max}^{\mathrm{ex}}(t)}$$

Table II. Table of error criteria for example problems

Fig.	E1	E2	E3	E4	E5	E6
5(a)	0·006182	0·000463	0·102794	0·121774	0·000000	−0·000048
5(b)	0·007775	0·000517	0·223424	0·032215	0·000000	0·000000
5(c)	0·002910	0·000201	0·041088	0·037532	0·000000	0·000212
5(d)	0·003532	0·000201	0·084018	0·019359	0·000000	0·000007
6(a)	0·015666	0·001161	0·173435	0·274874	0·041667	0·000119
6(b)	0·009227	0·000621	0·274061	0·022135	0·000000	−0·000062
6(c)	0·001740	0·000079	0·012801	0·011506	0·000000	0·000052
6(d)	0·002254	0·000092	0·038199	0·005854	0·000000	0·000002
8(a)	0·001471	0·000068	0·000541	0·009944	0·000000	0·000194
8(b)	0·001418	0·000069	0·007314	0·005994	0·000000	0·000406
8(c)	0·001214	0·000045	0·004231	0·007821	0·000000	0·000226
8(d)	0·001200	0·000048	0·008116	0·006171	0·000000	0·000353
9(a)	0·014804	0·001065	0·147398	0·230520	0·044167	0·000327
9(b)	0·016274	0·001184	0·033778	0·338399	0·039167	0·000074
9(c)	0·001094	0·000020	0·000005	0·004461	0·000000	0·000073
9(d)	0·001259	0·000036	0·010708	0·002652	0·000000	0·000092

Value for exact solution = 0·0

Error E6: Integral measure of mass preservation.

$$E6 = 1{\cdot}0 - \frac{1}{m(t)}\int_0^L \phi^{\text{num}}(x, t)\,dx$$

Value for exact solution: 0·0

REFERENCES

1. 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. numer. methods eng.*, **10**, 1389–1396 (1976).
2. J. C. Heinrich, P. S. Huyakorn, O. C. Zienkiewicz and A. R. Mitchell, 'An upwind finite element scheme for two-dimensional convective transport equation', *Int. j. numer. methods eng.*, **11**, 131–143 (1977).
3. T. J. R. Hughes, 'A simple scheme for developing upwind finite elements', *Int. j. numer. methods eng.*, **12**, 1359–1365 (1978).
4. J. C. Heinrich and O. C. Zienkiewicz, 'Quadratic finite element schemes for two-dimensional convective transport problems', *Int. j. numer. methods eng.*, **11**, 1831–1844 (1977).
5. I. Christie and A. R. Mitchell, 'Upwinding of high order Galerkin methods in conduction–convection problems', *Int. j. numer. methods eng.*, **12**, 1764–1771 (1978).
6. T. J. R. Hughes and A. N. Brooks, 'A multidimensional upwind scheme with no crosswind diffusion', in T. J. R. Hughes (ed.), *Finite Element Methods for Convection Dominated Flows*, AMD. Vol. 34, ASME, 1979.
7. D. W. Kelly, S. Nakazawa, O. C. Zienkiewicz and J. C. Heinrich, 'A note on upwinding and anisotropic balancing dissipation in finite element approximations to convective diffusion problems', *Int. j. numer. methods eng.*, **15**, 1705–1711 (1980).
8. T. J. R. Hughes and A. N. Brooks, 'A theoretical framework for Petrov–Galerkin methods with discontinuous weighting functions: Applications to the streamline upwind procedure', in R. H. Gallagher (ed.), *Finite Elements in Fluids, Vol. 4*, Wiley, London, 1982.
9. A. N. Brooks and T. J. R. Hughes, 'Streamline upwind/Petrov–Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier–Stokes equations', *Comp. Methods Appl. Mech. Eng.*, **32**, 199–259 (1982).
10. J. Donea, T. Belytschko and P. Smolinski, 'A generalized Galerkin method for steady convection–diffusion problems with application to quadratic shape function elements', *Comp. Methods Appl. Mech. Eng.*, **48**, 25–43 (1985).
11. T. E. Tezduyar and T. J. R. Hughes, 'Development of time-accurate finite element techniques for first-order hyperbolic systems with particular emphasis on the compressible Euler equations', *Final Report*, NASA–Ames University Consortium, Interchange No. NCA2-OR745-104, 1982.
12. T. E. Tezduyar and D. K. Ganjoo, 'Petrov–Galerkin formulations with weighting functions dependent upon spatial and temporal discretization: Applications to transient convection–diffusion problems', *Comp. Methods Appl. Mech. Eng.*, **59**, 49–71 (1986).
13. C.-C. Yu and J. C. Heinrich, 'Petrov–Galerkin methods for the time-dependent convective transport equation', *Int. j. numer. methods eng.*, **23**, 883–901 (1986).
14. E. Dick, 'Accurate Petrov–Galerkin methods for transient convective diffusion problems', *Int. j. numer. methods eng.*, **19**, 1425–1433 (1983).
15. R. S. Marshall, 'On upwinded quadratic elements for the one-dimensional convective transport equations', N.J. Institute of Technology, Department of Mechanical Engineering, preprint, 1978.
16. P. M. Gresho and R. L. Lee, 'Don't suppress the wiggles—They're telling you something', *Comp. Fluids*, **9**, 223–253 (1981).
17. A. Baptista, E. Adams and P. Gresho, 'Reference problems for the convection–diffusion forum', unpublished report, 1988.