INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2003; **41**:151–183 (DOI: 10.1002/fld.433)

Positivity-preserving, flux-limited finite-difference and finite-element methods for reactive transport

Robert J. MacKinnon[1,*,†] and Graham F. Carey[2]

[1] *Sandia National Laboratories, Albuquerque, New Mexico, 87185, U.S.A.*
[2] *College of Engineering, University of Texas at Austin, Austin, Texas, 78712, U.S.A.*

SUMMARY

A new class of positivity-preserving, flux-limited finite-difference and Petrov–Galerkin (PG) finite-element methods are devised for reactive transport problems. The methods are similar to classical TVD flux-limited schemes with the main difference being that the flux-limiter constraint is designed to preserve positivity for problems involving diffusion and reaction. In the finite-element formulation, we also consider the effect of numerical quadrature in the lumped and consistent mass matrix forms on the positivity-preserving property. Analysis of the latter scheme shows that positivity-preserving solutions of the resulting difference equations can only be guaranteed if the flux-limited scheme is both implicit and satisfies an additional lower-bound condition on time-step size. We show that this condition also applies to standard Galerkin linear finite-element approximations to the linear diffusion equation. Numerical experiments are provided to demonstrate the behavior of the methods and confirm the theoretical conditions on time-step size, mesh spacing, and flux limiting for transport problems with and without nonlinear reaction. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: convection–diffusion–reaction equation; positivity preserving; total variation diminishing; upwinding; Petrov–Galerkin method; finite-difference method

1. INTRODUCTION

During the last decade, there has been a significant increase in the use of adaptive stencil methods for convection-dominated flow and transport problems. In particular, total variation diminishing (TVD) high-resolution schemes [1–3] have proven to be very effective for a diverse range of applications. These applications include, for example, inviscid and viscous flows [4–9], flow and transport in porous media [10–12], and shallow water equations [13, 14]. This body of work has focused predominately on problems such as oscillations near sharp fronts and excessive numerical diffusion, with little attention given to problems with nonlinear reaction. Positivity of the approximate solution is sometimes recognized to be a desirable attribute, but in many applications may not be a strict requirement. Furthermore, standard

[*] Correspondence to: R. J. MacKinnon, Sandia National Laboratories, P.O. Box 5800, MS 0778, Albuquerque, NM 87123-0778, U.S.A.
[†] E-mail: rjmacki@sandia.gov

Received 16 August 2001
Revised 24 June 2002

approaches for maintaining, say, concentration $C \geqslant 0$ may be inadequate or inaccurate (e.g. one common strategy is to simply set negative nodal solution values to zero).

In TVD flux-limited methods, a feedback mechanism extracts information from the approximate solution and uses this information to decide where in the solution domain it is permissible to increase the accuracy of the difference stencil. A shortcoming of this reliance on a feedback mechanism is that flux-limited methods are non-linear, even for the case of linear convection. However, it is increasingly the case that contemporary applications are predominantly non-linear because of nonlinear reaction terms or constitutive relations. Since these problems are inherently non-linear to begin with, the added computational effort introduced by a nonlinear flux-limiting scheme is less significant and worth the potential increase in solution accuracy. However, the difficulty of preserving positivity of the approximate solution for the reactive transport problem with these types of formulations still remains.

TVD methods do not necessarily provide positivity-preserving solutions to convection-dominated transport problems with non-linear reaction terms. Moreover, a lack of positivity is possible even for simple convection problems in multidimensions, particularly on unstructured grids. This lack of positivity is in part due to the fact that TVD theory only guarantees positivity for pure hyperbolic conservation laws in one dimension. The theory does not include problems with diffusion and nonlinear reaction. In recent years, research has been directed towards developing positivity-preserving flux-limited schemes for hyperbolic problems on unstructured grids. These schemes usually satisfy a maximum principle or are based on positivity of coefficients. Both of these properties have the advantage of being relatively easy to satisfy in multidimensions (see References [14, 15] and references therein). In related work, the maximum principle was used to ensure positivity of finite-element approximations to convection-dominated elliptic problems [see Reference [16], references therein]. Recently, Berzins [17] proposed a modification to the standard Galerkin finite-element mass matrix with the aim of preserving positivity of discrete solutions for hyperbolic and parabolic PDEs. In the present work, a new class of flux-limited finite-difference and Petrov–Galerkin (PG) finite-element schemes are devised for reactive transport problems. These methods employ a new flux-limiter constraint that is designed to account for diffusion and reaction. In the case of pure convection, the new constraint reduces to the well-known TVD condition. Hence, we refer to the schemes as TVD-like methods. The key point is that, in addition to being high resolution, the resulting approximate solutions are guaranteed to be positivity preserving. An outline of the paper is as follows.

The class of reaction–diffusion problems to be analysed is concisely stated in Section 2. Then, in Section 3, we give a flux-limited finite-difference formulation for a representative parabolic convection–diffusion–reaction problem. Conditions for mesh spacing, time-step size, and flux limiting are given for this formulation that ensure positivity and accuracy for solutions of the difference equations. These conditions are obtained by requiring that the matrix of the final algebraic system maintain three properties during its evolution in time: (i) positive diagonal entries, (ii) negative off-diagonal entries, and (iii) diagonal dominance. A matrix with these properties is sometimes referred to as a matrix of non-negative type [16]. In addition, the right hand side vector is required to always have all positive entries. We show that if the final algebraic system meets these requirements, the approximate solution is guaranteed to be positivity preserving as it evolves in time. In Section 4, we extend the ideas to construct a Petrov–Galerkin finite-element method for this problem class. The method employs quadratic test functions weighted in the upstream direction similar to the classical approach [18]. However, the amount of upwind weighting introduced is now a function of the approximate solution and is limited using a standard TVD flux limiter in conjunction with a new limiter constraint that guarantees positivity. As is typical in standard finite-element approximations, the final form of the algebraic system of difference equations is dependent on the numerical quadrature employed to evaluate the element integral contributions. When the element integral contributions are approximated by Newton Cotes under-integration, we obtain a system matrix identical to that of the finite-difference method presented previously. It therefore follows that the conditions for mesh spacing, time-step size, and flux limiting are the same. A related case is then considered wherein all terms are integrated as before with the exception of the element mass matrices, which are integrated using a two-point Gauss quadrature (i.e. consistent mass matrix). The key finding here is that positivity preserving solutions of the resulting difference equations can only be guaranteed if the flux-limited scheme is both implicit and satisfies an additional lower-bound condition on time-step size. We show that this condition also applies for standard Galerkin linear finite-element approximations to the diffusion equation. In Section 5, a non-linear solution algorithm for the flux-limiting methods is described along with representative numerical examples that demonstrate the performance of the new methods. Extension of the 1D scheme to two dimensions is briefly commented on in Section 6. A detailed treatment of the 2D extension with numerical experiments is described in a subsequent study. Concluding remarks are provided in Section 7.

2. CONVECTION–DIFFUSION–REACTION PROBLEM

We consider flux-limited finite-difference and finite-element schemes for the prototype non-linear convection–diffusion–reaction (CDR) equation:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(D\,\frac{\partial C}{\partial x}\right) - \frac{\partial vC}{\partial x} + R(C) \quad \text{on } \Omega\times I \tag{1}$$

We further assume that coefficient functions $v(x)$, $D(x)$, and reaction term $R(C)$ are sufficiently well behaved that solutions exist and the subsequent finite-difference and finite-element models can be constructed to stated accuracy. For convenience, we will restrict the derivation to essential boundary conditions

$$C(0,t) = g_0(t) \geqslant 0, \quad C(L,t) = g_L(t) \geqslant 0, \quad t\in I \tag{2}$$

and initial conditions

$$C(x,0) = C_0(x) \geqslant 0, \quad x\in\Omega \tag{3}$$

where Ω is a one-dimensional domain $\Omega = [0,L]$, I is the time interval $I = (0,T]$, $C(x,t)$ is the scalar solution field, and the convective velocity $v(x)$, the diffusion coefficient $D(x)$, and reaction term $R(C)$ are given real-valued functions of x and C, respectively.

3. FINITE-DIFFERENCE APPROXIMATION

In this section, we develop a new positivity-preserving, flux-limited, finite-difference approximation to the CDR problem (1). To begin, let Ω be subdivided into E segments (cells) with mesh size $h_i = x_{i+1} - x_i$, $i \in \{1, 2, \dots, E\}$. Let $\Delta t > 0$ denote the time-step size and $t^n = n\Delta t$, $n = 0, 1, 2 \dots$. Let the difference approximation to C at (x_i, t^n) be denoted by c_i^n and let the time difference approximation be given by

$$\left.\frac{\partial C_i}{\partial t}\right|_{t^{n+\theta}} \approx \frac{c_i^{n+1} - c_i^n}{\Delta t} \tag{4}$$

where $t^{n+\theta} \equiv (1-\theta)t_n + \theta t_{n+1}$, $0 \leqslant \theta \leqslant 1$.

We denote the approximation to $C(x_i, t^{n+\theta})$ as

$$C(x_i, t^{n+\theta}) \approx c_i^{n+\theta} = (1-\theta)c_i^n + \theta c_i^{n+1} \tag{5}$$

Next, let the spatial derivatives at (x_i, t^n) be approximated by the following cell-centered differences

$$\frac{\partial}{\partial x}\left(D\frac{\partial C}{\partial x}\right)_i^n \approx \frac{1}{h_{mi}}\left[\frac{1}{h_i}D_{mi}(c_{i+1}^n - c_i^n) - \frac{1}{h_{i-1}}D_{mi-1}(c_i^n - c_{i-1}^n)\right] \tag{6}$$

and

$$\left(\frac{\partial vC}{\partial x}\right)_i^n \approx \frac{1}{h_{mi}}[f_{i+1/2}^n - f_{i-1/2}^n] \tag{7}$$

where $h_{mi} = (h_i + h_{i-1})/2$, $D_{mi} = (D_{i+1} + D_i)/2$, and $f_{i+1/2}$ denotes the convective flux function $f = vc$ at the center of cell i.

Of particular interest here are problems where the reaction term $R(c)$ in (1) has the form

$$R(C) \approx R(c) = [R_1(c) - R_2(c)]c \tag{8}$$

where the individual reaction functions have the property $R_k \geqslant 0$, $k = 1, 2$ for $c \geqslant 0$. A number of important reaction problems involve rate expressions that can be recast in this form. These problems involve rate expressions that are polynomial functions of c, rate expressions that are sufficiently smooth functions of c that can be cast into the appropriate form directly, or sufficiently smooth functions of c that can be approximated by polynomials. Examples include zeroth-order kinetics such as radioactive decay ($R = -\mu c$, $R_1 = 0$ and $R_2 = \mu$, where μ is a decay constant), a combination of first-order and second-order kinetics to represent population growth ($R = \mu_1 c - \mu_2 c^2$, $R_1 = \mu_1$ and $R_2 = \mu_2 c$, where μ_1 and μ_2 are growth and decline constants) [19], reactions that are fractional order ($R = -\mu c^{3/2}$, $R_1 = 0$ and $R_2 = \mu c^{1/2}$, where μ is a reaction rate constant) [20], and Monod kinetics used to simulate biodegradation ($R = \mu c/(K + c)$, $R_1 = \mu/(K + c)$, $R_2 = 0$, where μ is the maximum growth rate and K is the half-saturation constant [21]; note that in this form R_1 would have to be linearized by, for example, using the previous iterate value for c in the denominator).

In practice it has been observed that more robust results are obtained if the reaction term at time $t^{n+\theta}$ is represented by introducing averaging weights α_k, $k = 1, 2$. Moreover, we use this weighting later in our positivity analysis. Accordingly, let the reaction term be approximated by

$$R_i^{n+\theta} = (1-\alpha_1)R_{1i}^n c_i^n + \alpha_1 R_{1i}^{n+1} c_i^{n+1} - (1-\alpha_2)R_{2i}^n c_i^n - \alpha_2 R_{2i}^{n+1} c_i^{n+1} \tag{9}$$

where $0 \leqslant \alpha_k \leqslant 1$, $k=1,2$. For reasons that will become clear later, α_1 and α_2 will in general not be chosen equal to each other or to θ. Although this flexibility introduces an $O(\Delta t)$ time truncation error when $\theta \neq \alpha_1 \neq \alpha_2$, it is perfectly permissible since it does not violate consistency requirements and allows construction of schemes that retain favorable solution properties such as reduced phase errors and solution positivity. These favorable properties are due in part to the fact that the sign of reaction term contributions to the final matrix system of approximate equations can always be made to be positive if so desired. For example when $\alpha_1 = 0$ and $\alpha_2 = 1$ in (9), the reaction contributions to the RHS (i.e. at t^n) and to the LHS (i.e. at t^{n+1}) of the final matrix system of approximate equations will both be positive. Such positive contributions help guarantee positivity (see the comment after (27)) and alleviate restrictions on time-step size and mesh spacing.

Using (4)–(9), the cell-centered finite difference approximation of (1) at each interior grid point x_i, $i=2,3,\ldots,E$, can be written as

$$\begin{aligned}\left(\frac{c_i^{n+1}-c_i^n}{\Delta t}\right)h_{mi} = {} & \frac{\theta D_{mi}}{h_i}[c_{i+1}-c_i]^{n+1} - \frac{\theta D_{mi-1}}{h_{i-1}}[c_i - c_{i-1}]^{n+1} + \frac{(1-\theta)D_{mi}}{h_i}[c_{i+1}-c_i]^n \\ & - \frac{(1-\theta)D_{mi-1}}{h_{i-1}}[c_i - c_{i-1}]^n - \theta[f_{i+1/2} - f_{i-1/2}]^{n+1} \\ & - (1-\theta)[f_{i+1/2} - f_{i-1/2}]^n + (1-\alpha_1)R_{1i}^n c_i^n h_{mi} + \alpha_1 R_{1i}^{n+1} c_i^{n+1} h_{mi} \\ & - (1-\alpha_2)R_{2i}^n c_i^n h_{mi} - \alpha_2 R_{2i}^{n+1} c_i^{n+1} h_{mi}\end{aligned} \tag{10}$$

Following conventional flux-limiter terminology [4, 6], the flux function $f = vc$ at cell centers is given by

$$f_{i+1/2}^n = \tfrac{1}{2}[(f_i^n + f_{i+1}^n) - \mathrm{sign}(v_i)(1-\lambda_i^n)(f_{i+1}^n - f_i^n)], \quad i = 1,2,\ldots,E \tag{11}$$

where the flux function at node i is $f_i = v_i c_i$ and the flux limiter $\lambda_i^n = \lambda_i(\beta_i^n)$ is a function of the difference ratio β_i at time t^n given by

$$\beta_i^n = \frac{f_{s+1}^n - f_s^n}{f_{i+1}^n - f_i^n} \tag{12}$$

where we assume that $f_{i+1} - f_i > 0$ and $s = i - \mathrm{sign}(v_i)$.

In this study, we consider the following common flux limiter [22]

$$\lambda_i^n = \frac{2\beta_i^n}{1+\beta_i^n} \tag{13}$$

For the case of uniform grid size h, if flux limiter (13) is used in (11), the flux derivative in (7) will be approximated to second-order accuracy ($O(h^2)$) [2].

Flux approximation (11) takes on familiar forms depending on the form or values of λ_i^n. For example, when $\lambda_i^n = \beta_i^n$, 0, 1 or 2, flux approximation (11) becomes a two-point upstream-weighted approximation, a single-point upstream-weighted approximation, a centered-difference approximation, or a single-point downstream-weighted approximation, respectively. Note that when $\beta_i^n \neq 0$, 1 or 2, the calculation of β_i^n at cell centers next to the boundaries (i.e. at $x=0$ for $v>0$ and $x=L$ for $v<0$) may require upwind flux information that is outside the model domain. In this study we always use a first-order single-point upwind approximation at the cell center next to the upwind boundary, that is, $\lambda_{1+1/2}^n = 0$, $v>0$ and $\lambda_{E+1/2}^n = 0$, for $v<0$. This approach introduces a local first-order truncation error near the upstream boundary and has a negligible effect on solution accuracy since the significant solution gradients in test problems considered here are typically located in the interior of the problem domain away from the boundary. In this case a standard second-order scheme may also be used near the boundary. For problems where significant gradients occur near the upstream boundary one can use a first-order approximation in combination with local grid refinement.

Equations (10)–(12), along with (13), define the flux-limited finite-difference approximation at grid point x_i. Observe that because limiter λ_i^n is a function of the approximate solution c^n, the algorithm is nonlinear if any of the weighting parameters θ, α_1, or α_2 are nonzero, even if the original PDE problem is linear. Conditions on λ_i^n and Δt necessary to ensure a positive and accurate approximate solution are formulated next.

To simplify the following formulation, we restrict our derivation to the case where sign(v) >0. Incorporating (11) into (10) leads to the following flux-limited finite-difference approximation to model Equation (1) at each interior grid point x_i.

$$\begin{aligned}\left(\frac{c_i^{n+1} - c_i^n}{\Delta t}\right) h_{mi} = {} & \frac{\theta D_{mi}}{h_i}(c_{i+1} - c_i)^{n+1} - \frac{\theta D_{mi-1}}{h_{i-1}}(c_i - c_{i-1})^{n+1} + \frac{(1-\theta)D_{mi}}{h_i}(c_{i+1} - c_i)^n \\ & - \frac{(1-\theta)D_{mi-1}}{h_{i-1}}(c_i - c_{i-1})^n - \theta(f_i - f_{i-1})^{n+1} - \frac{\theta\lambda_i^{n+1}}{2}(f_{i+1} - f_i)^{n+1} \\ & + \frac{\theta\lambda_{i-1}^{n+1}}{2}(f_i - f_{i-1})^{n+1} - (1-\theta)(f_i - f_{i-1})^n - \frac{(1-\theta)\lambda_i^n}{2}(f_{i+1} - f_i)^n \\ & + \frac{(1-\theta)\lambda_{i-1}^n}{2}(f_i - f_{i-1})^n + (1-\alpha_1)R_{1i}^n c_i^n h_{mi} + \alpha_1 R_{1i}^{n+1} c_i^{n+1} h_{mi} \\ & - (1-\alpha_2)R_{2i}^n c_i^n h_{mi} - \alpha_2 R_{2i}^{n+1} c_i^{n+1} h_{mi} \qquad (14)\end{aligned}$$

We next formulate the upwind algebraic equivalent of approximation (14). This upwind form of (14) allows us to use simple linear algebra concepts to derive functional relationships between the flux-limiter value, time-step size, node-point spacing, diffusivity, and reaction terms. These relationships ensure positive approximations to c_i^n. In addition, they ensure that difference approximation (14) is locally $O(h^2)$ accurate in a Taylor series sense in those regions where the approximate solution is sufficiently smooth. To recast (14) into an algebraically equivalent upwind form, we assume that $0 < |\beta_i| \leqslant B < \infty$ for some finite constant B and replace the downwind convection terms (i.e, $f_{i+1} - f_i$) in equation (14) with upwind difference terms using Equation (12). Collecting terms we obtain

$$\begin{aligned}
&-c_{i-1}^{n+1}\left[\theta\left(\frac{D_{mi-1}}{h_{i-1}}+v_{i-1}\left(1+\frac{\lambda_i^{n+1}}{2\beta_i^{n+1}}-\frac{\lambda_{i-1}^{n+1}}{2}\right)\right)\right]\\
&\quad+c_i^{n+1}\left[\frac{h_{mi}}{\Delta t}+\theta\left(\left(\frac{D_{mi-1}}{h_{i-1}}+\frac{D_{mi}}{h_i}\right)+v_i\left(1+\frac{\lambda_i^{n+1}}{2\beta_i^{n+1}}-\frac{\lambda_{i-1}^{n+1}}{2}\right)\right)-\alpha_1 R_{1i}^{n+1}h_{mi}+\alpha_2 R_{2i}^{n+1}h_{mi}\right]\\
&\quad-c_{i+1}^{n+1}\left[\theta\,\frac{D_{mi}}{h_i}\right]=c_{i-1}^{n}\left[(1-\theta)\left(\frac{D_{mi-1}}{h_{i-1}}+v_{i-1}\left(1+\frac{\lambda_i^{n}}{2\beta_i^{n}}-\frac{\lambda_{i-1}^{n}}{2}\right)\right)\right]\\
&\quad+c_i^{n}\left[\frac{h_{mi}}{\Delta t}-(1-\theta)\left(\left(\frac{D_{mi-1}}{h_{i-1}}+\frac{D_{mi}}{h_i}\right)+v_i\left(1+\frac{\lambda_i^{n}}{2\beta_i^{n}}-\frac{\lambda_{i-1}^{n}}{2}\right)\right)\right.\\
&\quad\left.+(1-\alpha_1)R_{1i}^{n}h_{mi}-(1-\alpha_2)R_{2i}^{n}h_{mi}\right]\\
&\quad+c_{i+1}^{n}\left[(1-\theta)\frac{D_{mi}}{h_i}\right]
\end{aligned}\tag{15}$$

This equation can be written compactly as

$$a_{ii}^{n+1}c_i^{n+1}-\sum_{j\neq i}a_{ij}^{n+1}c_j^{n+1}=\sum_j b_{ij}^n c_j^n=F_i^n \tag{16a}$$

where the coefficients in the expression on the left depend on the unknown solution approximation. We write the corresponding system in matrix form as

$$\boldsymbol{A}^{n+1}\boldsymbol{c}^{n+1}=\boldsymbol{F}^n \tag{16b}$$

To ensure positivity we require coefficients in (16a) and (16b) to satisfy the following two conditions:

(i) non-negativeness

$$b_{ij}\geqslant 0 \text{ and } a_{ij}>0,\quad j=i-1,i,i+1 \;\; i=2,3,\ldots,E \tag{17}$$

(ii) strict diagonal dominance

$$\sum_{i\neq j}a_{ij}<a_{ii}\quad i=2,3,\ldots,E \tag{18}$$

We remark that the above conditions are sufficient but not strictly necessary for the numerical scheme to be positivity preserving.

Theorem
Under conditions of nonnegative initial and boundary data, the approximate solution at each time step will be nonnegative everywhere provided conditions (17) and (18) are satisfied.

Proof
The initial data by definition satisfies

$$c_i^0 \geqslant 0, \quad i = 1, 2, \ldots, E+1 \tag{19}$$

It follows on substitution of (19) in the right side of (15) that F_i^0 in (16) satisfies

$$F_i^0 \geqslant 0, \quad i = 2, 3, \ldots, E \tag{20}$$

and hence, at each interior grid point $i = 2, 3, \ldots, E$

$$a_{ii}^1 c_i^1 - \sum_{j \neq i} a_{ij}^1 c_j^1 \geqslant 0 \tag{21}$$

Using properties (17) in (21) we have,

$$a_{ii}^1 c_i^1 \geqslant \min_j c_j^1 \sum_{j \neq i} a_{ij}^1 \tag{22}$$

and using (18) it follows that

$$\left[a_{ii}^1 \middle/ \sum_{j \neq i} a_{ij}^1 \right] c_i^1 \geqslant \min_j c_j^1 \tag{23}$$

Inequality (23) applies at each interior grid point i, until a grid point with a neighbor on a boundary is reached. Since each boundary value is nonnegative, each adjacent interior grid point value is also nonnegative. This condition applies to each subsequent interior grid point and positivity of the approximation everywhere at t^1 is guaranteed. Hence we have

$$c_i^1 \geqslant 0, \quad i = 1, 2, \ldots, E+1$$

Further, because the approximation is nonnegative everywhere, condition (21) is satisfied for the subsequent time step and the same line of reasoning as for the first time step again applies. Repeating this process for each time step we get

$$c_i^n \geqslant 0, \quad i = 1, 2, \ldots, E+1 \tag{24}$$

and by induction the desired result is proved. □

Now, let us apply this result to our CDR scheme. Analyzing (15), condition (17) is satisfied if

$$\frac{D_{mi-1}}{h_{i-1}} + v_{i-1}\left(1 + \frac{\lambda_i^n}{2\beta_i^{n+1}} - \frac{\lambda_{i-1}^n}{2}\right) \geqslant 0 \tag{25}$$

$$\frac{h_{mi}}{\Delta t} + \theta\left(\left(\frac{D_{mi-1}}{h_{i-1}} + \frac{D_{mi}}{h_i}\right) + v_i\left(1 + \frac{\lambda_i^n}{2\beta_i^n} - \frac{\lambda_{i-1}^n}{2}\right)\right) - \alpha_1 R_{1i}^n h_{mi} + \alpha_2 R_{2i}^n h_{mi} > 0 \tag{26}$$

and

$$(1-\theta)\left(\left(\frac{D_{mi-1}}{h_{i-1}}+\frac{D_{mi}}{h_i}\right)+v_i\left(1+\frac{\lambda_i^n}{2\beta_i^n}-\frac{\lambda_{i-1}^n}{2}\right)\right)$$
$$-(1-\alpha_1)R_{1i}^n h_{mi}+(1-\alpha_2)R_{2i}^n h_{mi})\neq\frac{h_{mi}}{\Delta t} \tag{27}$$

Next recall the comments immediately following equation (9). Note here that conditions are most favorable for (26) and (27) when $\alpha_1=0$ and $\alpha_2=1$. This observation provides the main motivation for introducing parameters α_1 and α_2 and suggesting the choice $\theta\neq\alpha_1\neq\alpha_2$.

To simplify the derivation of conditions that must be placed on λ_i^n to ensure satisfaction of (25)–(27), we make the following standard assumption [2, 6] for the flux limiter

$$\lambda_i^n=0,\quad \text{if }\ \beta_i^n\neq 0 \tag{28}$$

Condition (28) gives rise to two consequences: (i) (14) is a standard $O(h)$ upwinded approximation to (1) when the slopes of the approximate solution have opposite signs at grid point x_i and (ii) $\lambda_i^n\geqslant 0$ when calculated from (13).

Constraint (25) will always be satisfied for each λ_i^n if

$$0\leqslant\lambda_i^n\leqslant 2\left(1+\frac{D_{mi}}{h_i v_i}\right) \tag{29a}$$

since $\lambda_i^n/\beta_i^n\geqslant 0$. In convection-dominated problems, the term $D_{mi}/(h_i v_i)$ is small. Hence, we choose to replace (29a) by the stronger inequality

$$0\leqslant\lambda_i^n\leqslant 2 \tag{29b}$$

which is identical to the standard TVD constraint that imposes a maximum bound on the flux limiter for hyperbolic problems [2, 6].

Condition (29b) also guarantees that the coefficients of the velocity terms in (25)–(27), and hence in (15), are always non-negative.

From condition (26) we obtain the time-step size restriction

$$\frac{1}{\Delta t}\neq\alpha_1 R_1^n-\alpha_2 R_2^n \tag{30}$$

where we again choose a stronger inequality by neglecting the diffusion term for a reason that will be come apparent shortly.

Condition (27) is always satisfied for each x_i if

$$0\leqslant\frac{\lambda_i^n}{\beta_i^n}\leqslant\frac{2h_{mi}}{v_i\Delta t(1-\theta)}-\frac{2}{v_i}\left(\frac{D_{mi-1}}{h_{i-1}}+\frac{D_{mi}}{h_i}\right)+\frac{2(1-\alpha_1)}{v_i(1-\theta)}R_{1i}^n h_{mi}$$
$$-\frac{2(1-\alpha_2)}{v_i(1-\theta)}R_{2i}^n h_{mi}-2\equiv G^n \tag{31}$$

since $\lambda_i^n\geqslant 0$. We note that if diffusion and reaction terms are zero, condition (31) reduces to the standard limiter constraint [6] that is a companion to (29b).

The diagonal dominance condition (18) is always satisfied if

$$\frac{h_{mi}}{\Delta t}+\theta\Delta v_i\left(1+\frac{\lambda_i^n}{2\beta_i^n}-\frac{\lambda_{i-1}^n}{2}\right)-\alpha_1 R_{1i}^n h_{mi}+\alpha_2 R_{2i}^n h_{mi}>0 \tag{32}$$

where $\Delta v_i = v_i - v_{i-1}$. Note that because of conditions (29b) and (30), (32) is always satisfied when $\Delta v_i = v_i - v_{i-1} \geqslant 0$. However, when $\Delta v_i = v_i - v_{i-1} < 0$, the following upper bound for λ_i^n/β_i^n is required to ensure the diagonal dominance requirement is satisfied

$$0\leqslant\frac{\lambda_i^n}{\beta_i^n}\leqslant\frac{4h_{mi}}{\theta(|\Delta v_i|-\Delta v_i)\Delta t}-\frac{4\alpha_1 R_{1i}^n}{\theta(|\Delta v_i|-\Delta v_i)}h_{mi}+\frac{4\alpha_2 R_{2i}^n}{\theta(|\Delta v_i|-\Delta v_i)}h_{mi}-2\equiv H^n \tag{33}$$

Note that this expression implies that $H^n\to\infty$ when $\Delta v_i \geqslant 0$ and therefore guarantees that the flux limiter is not constrained by (33) when $\Delta v_i \geqslant 0$.

In practice, λ_i is bounded above by relationships (29b), (31), and (33); that is

$$\lambda_i^n \leqslant \min(2, \beta_i^n G^n, \beta_i^n H^n) \tag{34}$$

In most practical applications, $G^n < H^n$ since $|\Delta v_i|$ will typically be small relative to v_i.

Relationships between Δt and h_{mi} can be determined from expressions (31) and (33) by satisfying their lower bounds. That is

$$\frac{2h_{mi}}{v_i\Delta t(1-\theta)}-\frac{2}{v_i}\left(\frac{D_{mi-1}}{h_{i-1}}+\frac{D_{mi}}{h_i}\right)+\frac{2(1-\alpha_1)}{(1-\theta)}\frac{R_{1i}^n h_{mi}}{v_i}-\frac{2(1-\alpha_2)}{(1-\theta)}\frac{R_{2i}^n h_{mi}}{v_i}-2\geqslant 0 \tag{35}$$

and

$$\frac{4h_{mi}}{\theta(|\Delta v_i|-\Delta v_i)\Delta t}-\frac{4\alpha_1 R_{1i}^n}{\theta(|\Delta v_i|-\Delta v_i)}h_{mi}+\frac{4\alpha_2 R_{2i}^n}{\theta(|\Delta v_i|-\Delta v_i)}h_{mi}-2\geqslant 0 \tag{36}$$

For convenience of interpretation, we introduce the cell Courant, Peclet, and Damkholer numbers, respectively,

$$Co_{hi}=\frac{v_i\Delta t}{h_{mi}} \tag{37}$$

$$P_{evi}=\frac{v_i}{\frac{1}{2}(D_{mi-1}/h_{i-1}+D_{mi}/h_i)} \tag{38}$$

and

$$D_{aki}^n=\frac{R_{ki}^n h_{mi}}{v_i}, \quad k=1,2 \tag{39}$$

Then (35) and (36) imply that the cell Courant number satisfy

$$\frac{1}{Co_{hi}}\geqslant\max\left(\frac{\theta}{2v_i}(|\Delta v_i|-\Delta v_i)+\alpha_1 D_{a1i}^n-\alpha_2 D_{a2i}^n,\right.$$
$$\left.\frac{(1-\theta)(2+P_{evi})+(1-\alpha_2)D_{a2i}^n P_{evi}-(1-\alpha_1)D_{a1i}^n P_{evi})}{P_{evi}},0\right) \tag{40}$$

Table I. Conditions for cell Courant number and time-step size.*

			Convection dominated $P_{evi} \to \infty$	Diffusion dominated $v \to 0$
θ	α_2	$Co_{hi} \leqslant$	$Co_{hi} \leqslant$	$\Delta t \leqslant$
0	0	$1/((2/P_{evi}+1)+D_{a2i})$	$1/(1+D_{a2i})$	$h^2/(2D+R_2h^2)$
0	1/2	$2/(2(2/P_{evi}+1)+D_{a2i})$	$2/(2+D_{a2i})$	$2h^2/(4D+R_2h^2)$
0	1	$1/(2/P_{evi}+1)$	1	$h^2/2D$
1/2	0	$2/((2/P_{evi}+1)+2D_{a2i})$	$2/(1+2D_{a2i})$	$h^2/(D+R_2h^2)$
1/2	1/2	$2/((2/P_{evi}+1)+D_{a2i})$	$2/(1+D_{a2i})$	$2h^2/(2D+R_2h^2)$
1/2	1	$2/(2/P_{evi}+1)$	2	h^2/D
1	0	$1/D_{a2i}$	$1/D_{a2i}$	$1/R_2$
1	1/2	$2/D_{a2i}$	$2/D_{a2i}$	$2/R_2$
1	1	∞	∞	∞

*$\alpha_1 = 0$ for all cases.

Remarks

In practice, expression (40) is used to determine the upper bound on Δt. If we ignore the contributions on the RHS of (40) from D_{a2i} in the first term and D_{a1i} in the second term, expression (40) can be simplified to yield the stronger inequality

$$Co_{hi} \leqslant \min\left(\frac{1}{\dfrac{\theta}{2v_i}(|\Delta v_i| - \Delta v_i) + \alpha_1 D^n_{a1i}}, \frac{P_{evi}}{(1-\theta)(2+P_{evi}) + (1-\alpha_2)D^n_{a2i}P_{evi})}\right) \tag{41}$$

From (41), we find that in the convective limit $P_{evi} \to \infty$

$$Co_{hi} \leqslant \min\left(\frac{1}{\dfrac{\theta}{2v_i}(|\Delta v_i| - \Delta v_i) + \alpha_1 D^n_{a1i}}, \frac{1}{(1-\theta) + (1-\alpha_2)D^n_{a2i})}\right) \tag{42a}$$

and in the diffusive limit $v \to 0$, this implies

$$\Delta t \leqslant \frac{h_{mi}}{(1-\theta)(D_{mi-1}/h_{i-1} + D_{mi}/h_i) + (1-\alpha_2)R^n_{2i}h_{mi})} \tag{42b}$$

Table I summarizes limits for Co_h and Δt given by (3.39) for cases where $\Delta v_i = 0$, grid spacing is a uniform h, the diffusion coefficient is a constant D, and θ and α_2 take on values of 0, $\frac{1}{2}$, and 1. To simplify the limit expressions we set $\alpha_1 = 0$.

From Table I we see that the limits for Co_h and Δt are independent of the reaction terms whenever $\alpha_1 = 0$ and $\alpha_2 = 1$. These cases have favourable solution properties (see comments following equations (9) and (27)) and permit the largest Courant number and time-step size for $\theta = 0$, $\frac{1}{2}$, and 1, respectively. Note that the limits for Co_h and Δt are determined by the reaction terms only when $\theta = 1$, except in the case $\alpha_2 = 1$.

4. FINITE-ELEMENT APPROXIMATION

Next we develop a finite-element Petrov–Galerkin formulation for the prototype convection–diffusion–reaction problem. This formulation employs quadratic test functions weighted in the upstream direction similar to the classical approach [18]. However, the amount of upwind weighting introduced now depends on limiter (13) and constraint (34). As is typical in standard finite-element approximations, the final form of the algebraic system of difference equations is dependent on the numerical quadrature formulae employed to evaluate the element integral contributions. When the mass and reaction matrices are lumped by Newton-Cotes under-integration and the remaining contributions are integrated with the usual Gauss rule we obtain a system matrix identical to that of the finite-difference method presented previously. It therefore follows that conditions for mesh spacing, time-step size, and flux limiting are the same. We then consider the case wherein all terms are integrated as just described with the exception of the element mass matrices, which are integrated using a two-point Gauss quadrature (i.e. consistent mass matrix). We find that positivity preserving solutions of the resulting algebraic equations can only be guaranteed if the flux-limited scheme is both implicit and satisfies an additional lower-bound condition on time-step size. We show that this condition also applies for standard Galerkin linear finite-element approximations to the diffusion equation.

Returning to the governing equation (1) and introducing a test function w, we proceed in the usual manner from weighted-residual concepts to construct a weak semi-discrete formulation as follows. Projecting with the test function and applying the Gauss-divergence theorem to both the diffusion and convection terms, the semi-discrete weak integral statement is: find $C \in H$ satisfying the essential boundary conditions (2) and initial conditions (3) such that

$$\int_0^L \frac{\partial C}{\partial t} w \, \mathrm{d}x = \int_0^L \left[-\frac{\partial w}{\partial x}\left(D \frac{\partial C}{\partial x} \right) + vC \frac{\partial w}{\partial x} + R(C)w \right] \mathrm{d}x \tag{43}$$

holds for all admissible $w \in W$ with $w=0$ at $x=0$ and $x=L$. Here H and W denote trial and test spaces.

Introducing the finite-element discretization and trial and test subspaces H_h and W_h of H and W respectively, we obtain from (43) a corresponding Petrov–Galerkin finite-element problem: For $t \in I$, find $c \in H_h$ satisfying the essential boundary conditions and initial conditions and such that

$$\int_0^L \frac{\partial c}{\partial t} w_h \, \mathrm{d}x + \int_0^L \left[\frac{\partial w_h}{\partial x}\left(D \frac{\partial c}{\partial x} \right) - f \frac{\partial w_h}{\partial x} - R(c,t) w_h \right] \mathrm{d}x = 0 \tag{44}$$

for all $w_h \in W_h$ with $w_h = 0$ at $x=0$ and $x=L$. In (44), f again denotes the convective flux function $f = vc$.

In the Petrov–Galerkin method developed here, the trial and test spaces consist of standard piecewise linear functions and quadratic upwind biased functions, respectively. It is important to note that the amount of upwind biasing specified in the test functions will be an explicit function of the approximate solution c. This explicit function is constructed based on flux-limiting ideas with the positivity preserving algebraic analysis presented in the previous section now adapted to the PG framework presented here.

As in Section 2, let $\Omega = [0,L]$ be subdivided into E segments (elements) with mesh size $h_i = x_{i+1} - x_i$, $i \in \{1,2,\ldots,E\}$. The approximate solution can be expressed in the standard way as the product series expansion

$$c(x,t) = \sum_{j=1}^{E+1} c_j(t)\phi_j(x) \tag{45}$$

where $c_j(t)$ are the unknown nodal approximate values and $\{\phi_j\}$ denote the familiar Lagrange piecewise-linear 'hat' functions

$$\begin{aligned} \phi_j(\xi) &= \frac{1}{2}(1+\xi), \quad \xi(x) = \frac{x-x_j}{x_j-x_{j-1}} + \frac{x-x_{j-1}}{x_j-x_{j-1}}, \quad x \in [x_{j-1}, x_j] \\ \phi_j(\xi) &= \frac{1}{2}(1-\xi), \quad \xi(x) = \frac{x-x_{j+1}}{x_{j+1}-x_j} + \frac{x-x_j}{x_{j+1}-x_j}, \quad x \in [x_j, x_{j-1}] \end{aligned} \tag{46}$$

on the patch for node j and $\phi_j(\xi)$ zero otherwise.

We use the group variable approximation for the convective flux function $f = vc$. That is

$$f(x,t) = vc = \sum_{j=1}^{E+1} f_j(t)\phi_j(x) = \sum_{j=1}^{E+1} v_j c_j(t)\phi_j(x) \tag{47}$$

Upwind test functions $\{w_{hi}\}$ are solution-dependent and based on the standard form (for example, see References [18, 23])

$$\begin{aligned} w_{hi}^n &= \phi_i + \text{sign}(v)\frac{\omega_i^n}{4}(1-\xi^2), \quad x \in [x_{i-1}, x_i] \\ w_{hi}^n &= \phi_i - \text{sign}(v)\frac{\omega_i^n}{4}(1-\xi^2), \quad x \in [x_i, x_{i+1}] \end{aligned} \tag{48}$$

on the patch centered at node i with the maps $\xi(x)$ defined as before and where upwinding parameter ω_i^n specifies the amount of upwind bias desired at t^n. In this study, ω_i^n is formulated so that the PG scheme is spatially higher-order accurate (based on local Taylor series truncation error) in regions where the solution is sufficiently smooth and first-order accurate otherwise. Specifically we set

$$\begin{aligned} \omega_i^n &= \sigma(1-\lambda_-^n), \quad x \in [x_{i-1}, x_i] \\ \omega_i^n &= \sigma(1-\lambda_+^n), \quad x \in [x_i, x_{i+1}] \end{aligned} \tag{49}$$

where σ is a constant, the magnitude of which depends on the numerical quadrature formula used to evaluate the element integral contributions involving ω_i^n. For example, in (49) σ takes on values of 3 or 1 if two-point Gauss quadrature or two-point Newton–Cotes quadrature is used, respectively. The 'limiter values' λ_-^n and λ_+^n in (49) are defined locally by

$$\begin{aligned} \lambda_-^n &= \lambda_{i-1}(\beta_{i-1}^n), \quad \lambda_+^n = \lambda_i(\beta_i^n), \quad v > 0 \\ \lambda_+^n &= \lambda_{i+1}(\beta_{i+1}^n), \quad \lambda_-^n = \lambda_i(\beta_i^n), \quad v < 0 \end{aligned} \tag{50}$$

where β_i^n and λ_i^n are defined previously in (12) and (13), respectively. Note that choices $\lambda_i^n = 1$ and $\lambda_i^n = 0$, respectively yield the standard Galerkin and fully upwinded PG approximations to the convective term. Lastly, we see from (49) and (50) that ω_i^n is a function of the gradient ratio β_i^n, with its specific form dependent on the flow direction.

Introducing expressions (45) through (50) into (44) and using the variably weighted time-stepping scheme (15) and reaction term approximation (9), the non-linear system again has the form

$$\mathbf{A}^{n+1}\boldsymbol{c}^{n+1} = \mathbf{B}^n\boldsymbol{c}^n = \mathbf{F}^n \tag{51}$$

where

$$\begin{aligned} A_{ij}^{n+1} = {} & \frac{1}{\Delta t}\int_0^L w_{hi}^{n+1}\phi_j \,\mathrm{d}x + \theta\int_0^L \left[\frac{\partial w_{hi}^{n+1}}{\partial x} D\frac{\partial \phi_j}{\partial x} + \frac{\partial w_{hi}^{n+1}}{\partial x} v_j\phi_j\right]\mathrm{d}x \\ & - \int_0^L w_{hi}^{n+1}[\alpha_1 R_{1j}^{n+1} - \alpha_2 R_{2j}^{n+1}]\phi_j\,\mathrm{d}x \end{aligned} \tag{52a}$$

$$\begin{aligned} B_{ij}^{n} = {} & \frac{1}{\Delta t}\int_0^L w_{hi}^{n}\phi_j \,\mathrm{d}x - (1-\theta)\int_0^L \left[\frac{\partial w_{hi}^{n}}{\partial x} D\frac{\partial \phi_j}{\partial x} + \frac{\partial w_{hi}^{n}}{\partial x} v_j\phi_j\right]\mathrm{d}x \\ & + \int_0^L w_{hi}^{n}[(1-\alpha_1) R_{1j}^{n} - (1-\alpha_2) R_{2j}^{n}]\phi_j\,\mathrm{d}x \end{aligned} \tag{52b}$$

As in the previous finite-difference scheme (16b), conditions on λ_i^n and Δt must be imposed to ensure that the PG approximation given by (51) is positive and higher-order accurate (based on local Taylor series truncation error) in regions where the approximate solution is sufficiently smooth. In this study, we consider the case where the integrals in (52a) and (52b) are 'lumped' by integrating approximately using a two-point Newton–Cotes formula, with $\sigma = 1$ in (49). Performing these integrations and assembling the element matrices, the resulting discrete finite-element approximation at each interior grid point x_i is

$$\begin{aligned} \left(\frac{c_i^{n+1} - c_i^n}{\Delta t}\right) h_{mi} = {} & \frac{\theta D_{mi}}{h_i}(c_{i+1} - c_i)^{n+1} - \frac{\theta D_{mi-1}}{h_{i-1}}(c_i - c_{i-1})^{n+1} + \frac{(1-\theta)D_{mi}}{h_i}(c_{i+1} - c_i)^n \\ & - \frac{(1-\theta)D_{mi-1}}{h_{i-1}}(c_i - c_{i-1})^n - \theta(f_i - f_{i-1})^{n+1} \\ & - \frac{\theta\lambda_i^{n+1}}{2}(f_{i+1} - f_i)^{n+1} + \frac{\theta\lambda_{i-1}^{n+1}}{2}(f_i - f_{i-1})^{n+1} \\ & - (1-\theta)(f_i - f_{i-1})^n - \frac{(1-\theta)\lambda_i^n}{2}(f_{i+1} - f_i)^n + \frac{(1-\theta)\lambda_{i-1}^n}{2}(f_i - f_{i-1})^n \\ & + (1-\alpha_1)R_{1i}^n c_i^n h_{mi} + \alpha_1 R_{1i}^{n+1} c_i^{n+1} h_{mi} \\ & - (1-\alpha_2)R_{2i}^n c_i^n h_{mi} - \alpha_2 R_{2i}^{n+1} c_i^{n+1} h_{mi} \end{aligned} \tag{53}$$

where we have reintroduced the notation for the flux function $f_i = v_i c_i$ at node i (see Equation (11)). This finite-element approximation is identical to the finite-difference approximation (14). Therefore, the analysis following (14) applies here as well, so constraints (34) and (40) again define the conditions on λ_i^n, h and Δt necessary to ensure a stable, positive, and accurate solution. We remark that (53) would also be obtained if $\sigma = 3$ in (49) and the middle integrals in (52a) and (52b) were evaluated using a two-point Gauss formula instead of the Newton–Cotes formula.

We next consider a consistent mass-matrix case where the first integrals in (52a) and (52b) are evaluated approximately using a two-point Gauss formula while other integrals in (52a) and (52b) are treated as before. We begin by going directly to the resulting representative finite-element approximation corresponding to (53)

$$\begin{aligned}
&\frac{1}{6}\left(\frac{c_{i-1}^{n+1}-c_{i-1}^{n}}{\Delta t}\right)h_{i-1}+\frac{2}{3}\left(\frac{c_{i}^{n+1}-c_{i}^{n}}{\Delta t}\right)h_{mi}+\frac{1}{6}\left(\frac{c_{i+1}^{n+1}-c_{i+1}^{n}}{\Delta t}\right)h_{i}\\
&\quad=\frac{\theta D_{mi}}{h_i}(c_{i+1}-c_i)^{n+1}-\frac{\theta D_{mi-1}}{h_{i-1}}(c_i-c_{i-1})^{n+1}+\frac{(1-\theta)D_{mi}}{h_i}(c_{i+1}-c_i)^{n}\\
&\qquad-\frac{(1-\theta)D_{mi-1}}{h_{i-1}}(c_i-c_{i-1})^{n}-\theta(f_i-f_{i-1})^{n+1}\\
&\qquad-\frac{\theta\lambda_i^{n+1}}{2}(f_{i+1}-f_i)^{n+1}+\frac{\theta\lambda_{i-1}^{n+1}}{2}(f_i-f_{i-1})^{n+1}-(1-\theta)(f_i-f_{i-1})^{n}\\
&\qquad-\frac{(1-\theta)\lambda_i^{n}}{2}(f_{i+1}-f_i)^{n}+\frac{(1-\theta)\lambda_{i-1}^{n}}{2}(f_i-f_{i-1})^{n}+(1-\alpha_1)R_{1i}^{n}c_i^{n}h_{mi}\\
&\qquad+\alpha_1R_{1i}^{n+1}c_i^{n+1}h_{mi}-(1-\alpha_2)R_{2i}^{n}c_i^{n}h_{mi}-\alpha_2R_{2i}^{n+1}c_i^{n+1}h_{mi}
\end{aligned}\tag{54}$$

Following the development presented in Section 3, Equation (54) can be expressed in a form corresponding to Equation (15):

$$\begin{aligned}
&-c_{i-1}^{n+1}\left[\theta\left(\frac{D_{mi-1}}{h_{i-1}}+v_{i-1}\left(1+\frac{\lambda_i^{n+1}}{2\beta_i^{n+1}}-\frac{\lambda_{i-1}^{n+1}}{2}\right)\right)-\frac{h_{i-1}}{6\Delta t}\right]\\
&+c_i^{n+1}\left[\frac{2h_{mi}}{3\Delta t}+\theta\left(\left(\frac{D_{mi-1}}{h_{i-1}}+\frac{D_{mi}}{h_i}\right)+v_i\left(1+\frac{\lambda_i^{n+1}}{2\beta_i^{n+1}}-\frac{\lambda_{i-1}^{n+1}}{2}\right)\right)-\alpha_1R_{1i}^{n+1}h_{mi}+\alpha_2R_{2i}^{n+1}h_{mi}\right]\\
&-c_{i+1}^{n+1}\left[\theta\frac{D_{mi}}{h_i}-\frac{h_i}{6\Delta t}\right]=c_{i-1}^{n}\left[\frac{h_{i-1}}{6\Delta t}+(1-\theta)\left(\frac{D_{mi-1}}{h_{i-1}}+v_{i-1}\left(1+\frac{\lambda_i^{n}}{2\beta_i^{n}}-\frac{\lambda_{i-1}^{n}}{2}\right)\right)\right]\\
&+c_i^{n}\left[\frac{2h_{mi}}{3\Delta t}-(1-\theta)\left(\left(\frac{D_{mi-1}}{h_{i-1}}+\frac{D_{mi}}{h_i}\right)+v_i\left(1+\frac{\lambda_i^{n}}{2\beta_i^{n}}-\frac{\lambda_{i-1}^{n}}{2}\right)\right)\right.\\
&\left.-(1-\alpha_1)R_{1i}^{n}h_{mi}+(1-\alpha_2)R_{2i}^{n}h_{mi}\right]+c_{i+1}^{n}\left[\frac{h_i}{6\Delta t}+(1-\theta)\frac{D_{mi}}{h_i}\right]
\end{aligned}\tag{55}$$

Analysing (55), the positive coefficient condition (17) is satisfied if

$$\frac{h_i^2}{\Delta t}\leqslant 6\theta D_{mi}\tag{56}$$

We remark that (56) implies positivity preserving solutions for the consistent mass matrix case are only guaranteed for $\theta > 0$ and $D_{mi} > 0$.

$$v_{i-1}\left(1+\frac{\lambda_i^{n+1}}{2\beta_i^{n+1}}-\frac{\lambda_{i-1}^{n+1}}{2}\right) \geqslant \frac{h_{i-1}}{\theta 6\Delta t}-\frac{D_{mi-1}}{h_i} \tag{57}$$

and

$$\begin{aligned}(1-\theta)\Bigg(&\left(\frac{D_{mi-1}}{h_{i-1}}+\frac{D_{mi}}{h_i}\right)\\ &+v_i\left(1+\frac{\lambda_i^n}{2\beta_i^n}-\frac{\lambda_{i-1}^n}{2}\right)-\frac{(1-\alpha_1)}{(1-\theta)}R_{1i}^n h_{mi}+\frac{(1-\alpha_2)}{(1-\theta)}R_{2i}^n h_{mi}\Bigg)\Bigg) \neq \frac{2h_{mi}}{3\Delta t}\end{aligned} \tag{58}$$

Since the lower bound of expression (57) is less than or equal to zero in accordance with condition (56), it is sufficient for each λ_i^n to again satisfy (recall (29)),

$$0 \leqslant \lambda_i^n \leqslant 2 \tag{59}$$

Also, (58) is always satisfied for each x_i if

$$\begin{aligned}0 \leqslant \frac{\lambda_i^n}{\beta_i^n} \leqslant \frac{4h_{mi}}{3v_i\Delta t(1-\theta)}-\frac{2}{v_i}\left(\frac{D_{mi-1}}{h_{i-1}}+\frac{D_{mi}}{h_i}\right)+\frac{2(1-\alpha_1)}{v_i(1-\theta)}R_{1i}^n h_{mi}\\ -\frac{2(1-\alpha_2)}{v_i(1-\theta)}R_{2i}^n h_{mi}-2 \equiv G'^n\end{aligned} \tag{60}$$

since $\lambda_i^n \geqslant 0$.

The diagonal dominance condition (18) is always satisfied if

$$\frac{h_{mi}}{3\Delta t}+\theta(v_i-v_{i-1})\left(1+\frac{\lambda_i^n}{2\beta_i^{n+1}}-\frac{\lambda_{i-1}^n}{2}\right)-\alpha_1 R_{1i}^n h_{mi}+\alpha_2 R_{2i}^n h_{mi} \neq 0 \tag{61}$$

which leads to the condition

$$0 \leqslant \frac{\lambda_i^n}{\beta_i^n} \leqslant \frac{4h_{mi}}{3\theta(|\Delta v_i|-\Delta v_i)\Delta t}+\frac{4\alpha_1 R_{1i}^n}{\theta(|\Delta v_i|-\Delta v_i)}h_{mi}-\frac{4\alpha_2 R_{2i}^n}{\theta(|\Delta v_i|-\Delta v_i)}h_{mi}-2 \equiv H'^n \tag{62}$$

From (59), (60), and (62)

$$\lambda_i^n \leqslant \min(2, \beta_i^n G'^n, \beta_i^n H'^n) \tag{63}$$

Finally, again introducing cell Courant, Peclet, and Damkholer numbers (37), (38), and (39), respectively, we obtain the condition

$$\begin{aligned}\frac{1}{Co_{hi}} \geqslant \max\Bigg(&\frac{3\theta}{2v_i}(|\Delta v_i|-\Delta v_i)+\alpha_1 D_{a1i}^n-\alpha_2 D_{a2i}^n,\\ &\frac{3((1-\theta)(2+P_{evi})+(1-\alpha_2)D_{a2i}^n P_{evi}-(1-\alpha_1)D_{a1i}^n P_{evi})}{2P_{evi}}, 0\Bigg)\end{aligned} \tag{64}$$

and, in addition to (64), Δt must satisfy the minimum time-step size requirement given by condition (56).

Table II. Conditions for cell Courant number and time-step size.*

			Convection dominated	Diffusion dominated
			$P_{evi} \to \infty$	$v \to 0$
θ	α_2	$Co_{hi} \leqslant$	$Co_{hi} \leqslant$	$\Delta t \leqslant$
0	0	$2/(3(2/P_{evi}+1)+3D_{a2i})$	$2/3(1+D_{a2i})$	$2h^2/3(2D+R_2h^2)$
0	1/2	$4/(6(2/P_{evi}+1)+3D_{a2i})$	$4/3(2+D_{a2i})$	$4h^2/3(4D+R_2h^2)$
0	1	$2/(3(2/P_{evi}+1))$	$2/3$	$h^2/3D$
1/2	0	$4/(3(2/P_{evi}+1)+6D_{a2i})$	$4/3(1+2D_{a2i})$	$2h^2/3(D+R_2h^2)$
1/2	1/2	$4/(3(2/P_{evi}+1)+3D_{a2i})$	$4/3(1+D_{a2i})$	$4h^2/3(2D+R_2h^2)$
1/2	1	$4/(3(2/P_{evi}+1))$	$4/3$	$2h^2/3D$
1	0	$2/3_{Da2i}$	$2/3D_{a2i}$	$2/3R_2$
1	1/2	$4/3_{Da2i}$	$4/3D_{a2i}$	$4/3R_2$
1	1	∞	∞	∞

*$\alpha_1 = 0$ for all cases.

Remark

Following the line of reasoning presented in Section 3 for Equations (41) and (42), we obtain from (64) the simplified Courant condition

$$Co_{hi} \leqslant \min\left(\frac{1}{\dfrac{3\theta}{2v_i}(|\Delta v_i| - \Delta v_i) + \alpha_1 D^n_{a1i}}, \frac{2P_{evi}}{3((1-\theta)(2+P_{evi}) + (1-\alpha_2)D^n_{a2i}P_{evi})}\right) \tag{65}$$

In the convective limit $P_{evi} \to \infty$ (65) simplifies to

$$Co_{hi} \leqslant \min\left(\frac{1}{\dfrac{3\theta}{2v_i}(|\Delta v_i| - \Delta v_i) + \alpha_1 D^n_{a1i}}, \frac{2}{3((1-\theta) + (1-\alpha_2)D^n_{a2i})}\right) \tag{66a}$$

and in the diffusive limit $v \to 0$ we now get

$$\Delta t \leqslant \frac{2h_{mi}}{3((1-\theta)(D_{mi-1}/h_{i-1} + D_{mi}/h_i) + (1-\alpha_2)R^n_{2i}h_{mi})} \tag{66b}$$

Table II presents limits for Co_h and Δt given by (66) for cases where $\Delta v_i = 0$, grid spacing is a uniform h, the diffusion coefficient is a constant D, and θ and α_2 take on values of 0, $\frac{1}{2}$, and 1 as indicated. As before we set $\alpha_1 = 0$.

The results in Table II can be compared with those in Table I for the finite-difference scheme (and, equivalently, the lumped finite-element scheme). We immediately see that for the consistent finite-element scheme, all limits are reduced by a factor of 2/3 as compared to the finite difference and lumped finite-element schemes. Again we see that the limits for Co_h and Δt are independent of the reaction terms whenever $\alpha_1 = 0$ and $\alpha_2 = 1$ and permit the largest Courant number and time-step size for $\theta = 0$, $\frac{1}{2}$, and 1, respectively. As before, limits for Co_h and Δt are determined by the reaction terms only when $\theta = 1$, except in the case $\alpha_2 = 1$.

5. NUMERICAL STUDIES

We present results of several numerical experiments that were designed to assess the performance of the new flux-limited reactive transport methods. The test cases include both hyperbolic and parabolic transport problems with (and without) reaction.

5.1. Solution algorithm

The following solution algorithm is applied in the numerical experiments. Recall that the flux-limited scheme presented here is a nonlinear scheme if any of the weighting parameters θ, α_1, or α_2 are nonzero, even if the original PDE problem is linear. Therefore, the solution c^{n+1} at the end of the time step must be obtained by iterative solution of a system of nonlinear equations using, for instance, successive approximation, Newton iteration or a similar scheme. For example, in our time stepping scheme the solution at the end of the previous time step provides a good starting iterate and successive approximation is a natural choice in view of the structure of the non-linear coefficients in (16b) or (51). Let k denote the iteration level in a successive approximation scheme within each time step. The system of equations (16b) or (51) is then linearized at each iteration by evaluating the coefficient matrix at the current iterate. That is at each time step n, solve the linearized system

$$A^{n+1,k}c^{n+1,k+1}=F^{n,k} \tag{67}$$

for each iterate $k=0,1,2\dots$.

A simple solution algorithm for the new positivity-preserving difference scheme follows as:

```
For time step $n$ until $N$ steps
    Post-process the solution at $t^n$ to obtain $\beta_i^n$ and $\lambda_i^n$ (from Equations (12) and (13)) at each interior node point $i$. (During this step, constraint (34) on $\lambda_i$ must be checked. If the computed $\lambda_i$ exceeds the acceptable limit defined by the minimum value of (34), it is set equal to the acceptable limit. In the present implementation, the acceptable $\lambda_i^n$ is calculated and stored in an array.)

    For each iterate $k$ until $k\max$ do
    If $k=0$, then
    Set $\lambda_i^{n+1,k}=\lambda_i^n$ and $R_{pi}^{n+1,k}=R_{pi}^n$ for $p=1,2$ and each $i=2,3,\dots,E$.
    Else
    Post process the solution from iterate $k$ to obtain $\beta_i^{n+1,k+1}$ and $\lambda_i^{n+1,k+1}$ at each interior node point $i$.
    End If
    Solve (67) using $\lambda_i^n$, $R_{pi}^n$ (in the R.H.S.) and $\lambda_i^{n+1,k+1}$, $R_{pi}^{n+1,k+1}$ (in the L.H.S.) to obtain $c^{n+1,k+2}$
    Replace $c^{n+1,k+1}$ by current estimate $c^{n+1,k+2}$
    If
    $|c^{n+1,k+2}-c^{n+1,k+1}|<\varepsilon$ (test at end of loop or if $k=k$ max) then exit
    else continue
    Replace $c^n$ with $c^{n+1}$
    End For.
End For.
```

5.2. Test problems

In the test problems presented here, we assume that the convective velocity $v \geqslant 0$ and diffusion coefficient $D \geqslant 0$ are constants in space and time. With these assumptions, and using (8) in (1), the governing transport equation simplifies in dimensionless form to

$$\frac{\partial C_D}{\partial T} = D^* \frac{\partial^2 C_D}{\partial X^2} - \frac{\partial C_D}{\partial X} + D_{a1} C_D - D_{a2} C_D, \quad 0 \leqslant X \leqslant 1, \;\; T \neq 0 \tag{68}$$

where

$$X = x/L, \quad T = vt/L, \quad D^* = D/vL, \quad C_D = C/C_0, \quad \text{and} \quad D_{ak} = RL/v, \quad k = 1,2 \tag{69}$$

Here, C_0 is an arbitrary non-dimensionalizing concentration and all other variables have been defined previously. To facilitate the discussion of the numerical results, flux-limiter constraints given by (29b) and (31) are provided here in dimensionless form as

$$0 \leqslant \lambda_i^n \leqslant 2 \tag{70a}$$

$$0 \leqslant \frac{\lambda_i^n}{\beta_i^n} \leqslant \frac{2}{Co_{hi}(1-\theta)} - \frac{2}{P_{evi}} + \frac{2(1-\alpha_1)}{(1-\theta)} D_{a1i}^n - \frac{2(1-\alpha_2)}{(1-\theta)} D_{a2i}^n - 2 \tag{70b}$$

where Co_{hi}, P_{evi}, and D_{aki}, $k = 1,2$ are given by (37)–(39). The latter two parameters are related to the dimensionless variables in (69) in the following way

$$P_{evi} = D^* L/h \quad \text{and} \quad D_{aki} = D_{ak} h/L, \quad k = 1,2 \tag{71}$$

where h is the grid spacing for the uniform grids used in the numerical examples presented here.

5.2.1. Convection–diffusion–reaction case studies. The scheme is tested for the following CDR problems:

Case 1: Convective transport of an initial square wave with a decreasing solution.
Case 2: Convective transport of an initial square wave with an increasing solution.
Case 3: Convective-diffusive transport of a solute injected for a short period into the left end of the problem domain; non-reactive and reactive solutes are considered.

Uniform grids are used with mesh sizes $h = 1/20$ and $1/40$. Results for grid Courant numbers of 3/8 and 3/4 are computed for both grids using non-dimensional time-step sizes of $\Delta T = 0.01875$ and 0.0375 with the coarse grid, and $\Delta T = 0.009375$ and 0.01875 with the fine grid. The non-dimensionalizing concentration is $C_0 = 1$. Limiter (13) is used for all calculations. Numerical results computed using new limiter constraint (70b) and the standard TVD limiter constraint (i.e. (70b) with $P_{evi} = D_{a1i}^n = D_{a2i}^n = 0$) are compared with analytical solutions for each case. When the numerical scheme is implicit with either $\theta > 0$ or $\alpha_1 > 0$ or $\alpha_2 > 0$, nonlinear solutions are computed using 1 iteration only (2 solves per time step).

Case 1: This simple test case can cause numerical difficulties (e.g. excessive numerical diffusion and spurious oscillations) for many convection and convection-diffusion schemes. The initial concentration profile is

$$C_{D0} = 1 \quad \text{if} \;\; 0.1 < x_i < 0.3 \quad \text{and} \;\; C_{D0} = 0 \quad \text{otherwise}$$

For this case the system diffusivity is $D^* = 1 \times 10^{-9}$ (a non-zero value for D^* is specified to avoid zero pivot values in the Gaussian elimination solver routine used to solve system (16b) at each interaction step). The following simple non-linear reaction is considered:

$$R = -\mu C^{3/2} \quad \text{therefore } R_1 = 0 \text{ and } R_2 = \mu C^{1/2} \tag{72}$$

This non-linear rate expression represents a practical class of reactions sometimes involving chain mechanisms [20]. The corresponding system Damkohler number is given by

$$D_a = D_{a2} = \left(-\mu C_0^{1/2} \frac{L}{v}\right) C_D^{1/2} = -\mu^* C_D^{1/2} \tag{73}$$

In this case we consider a problem that has a relatively high rate of reaction to convection and set $\mu^* = 4$.

Since D^* is negligibly small, the analytic solution to Equation (68) is very close to the solution to the pure hyperbolic case with $D^* = 0$. This limit case can be written with respect to a reference frame that moves with the initial solution profile. That is,

$$\frac{\mathrm{d}C_D}{\mathrm{d}T} = -\mu^* C_D^{3/2} \tag{74}$$

where material derivative $\mathrm{d}C_D/\mathrm{d}T$ is the time derivative in the moving reference frame (characteristic frame). This expression can be integrated to yield ([20, p. 29])

$$C_D(T - T_0) = \frac{C_{D0}}{[1 + C_{D0}^{1/2}(\mu^*/2)(T - T_0)]^2} \tag{75}$$

where C_{D0} is the initial concentration profile at time $T = T_0$.

We first demonstrate how the lack of positivity can occur if the standard TVD constraints on the limiter are used. Figure 1 shows the exact (for the case $D^* = 0$) and computed solution profiles at $T = 0$ and 0.3 for the new scheme and the standard scheme for grid spacing $h = 1/20$ and 1/40 and a Courant number $Co_{hi} = \frac{3}{4}$. These results were obtained from fully explicit calculations with time-weighting parameters $\theta = 0$ and $\alpha_2 = 0$. For both grid sizes, the *standard TVD scheme yields negative solution values* at the trailing edge of the solution profile whereas the solution obtained using the new scheme is positive everywhere. Both numerical solutions are similar and slightly out of phase with the exact solution. As expected, both the accuracy and propagation speed of the approximate solutions improve as the grid spacing is decreased. It should be noted that negative solution values exhibited by the standard scheme are nonphysical and unacceptable in (68). This is particularly true when the reaction term has a nonlinear form such as in (72), which is not defined for negative concentration values and thus will cause the numerical solution to abort. To avoid this problem with the standard scheme, we set negative solution values to zero when evaluating the reaction term in (68). We remark that this simple ‘fix’ is only used here to facilitate a comparison between schemes and to show that the standard scheme will produce unacceptable negative solution values.

The dependence of the solution profiles on h, ΔT, and parameters θ and α_2 is further depicted in Figures 2 and 3. Numerical solutions, at $T = 0.3$, are presented for Courant numbers $Co_{hi} = 3/8$ and 3/4. All results are shown to be positive, including those calculated with the standard TVD scheme. Moreover, they tend to be slightly more diffusive with increasing

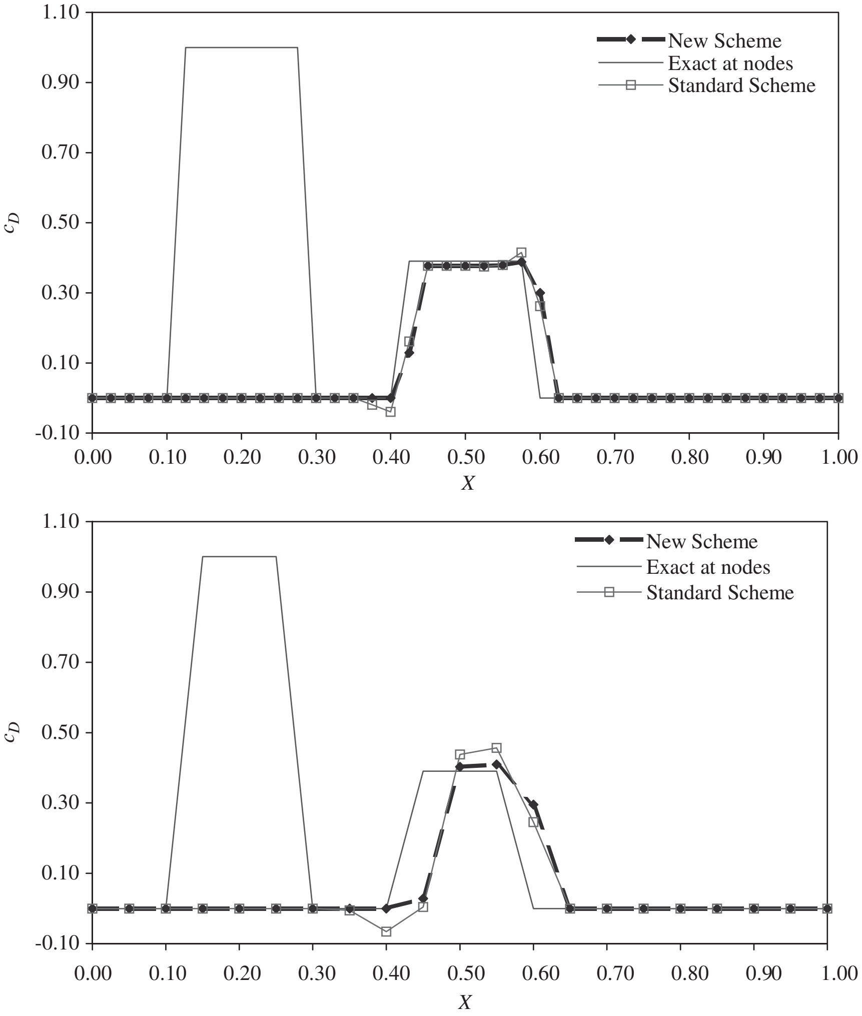


Figure 1. Square wave with decreasing amplitude at $T = 0.0$ and 0.3 with $Co_{hi} = 3/4$, $\theta = 0$, $\alpha_2 = 0$, and $h = 1/20$ (bottom) and $h = 1/40$ (top).

h, increasing θ, and decreasing Co_{hi}. These results are consistent with those observed in standard one-dimensional first-order and TVD upwind approximations to wave propagation without reaction [24]. Results also indicate that increasing the implicitness of the reaction term improves the propagation speed of the numerical approximation (Figure 2). Interestingly, the new scheme and the standard scheme give identical results for all cases except the fully explicit case presented in Figure 1. This behaviour can be easily explained for the implicit reaction case shown in Figure 2 by the fact that according to condition (70b) when $\alpha_2 = 1$ the new condition reduces to the standard condition (recall that $R_1 = 0$ and $D^* \sim 0$). The other case in Figure 3 gives identical results because, as the solution becomes more diffuse, the raw limiter value calculated from (13) always falls below the standard and new constrained

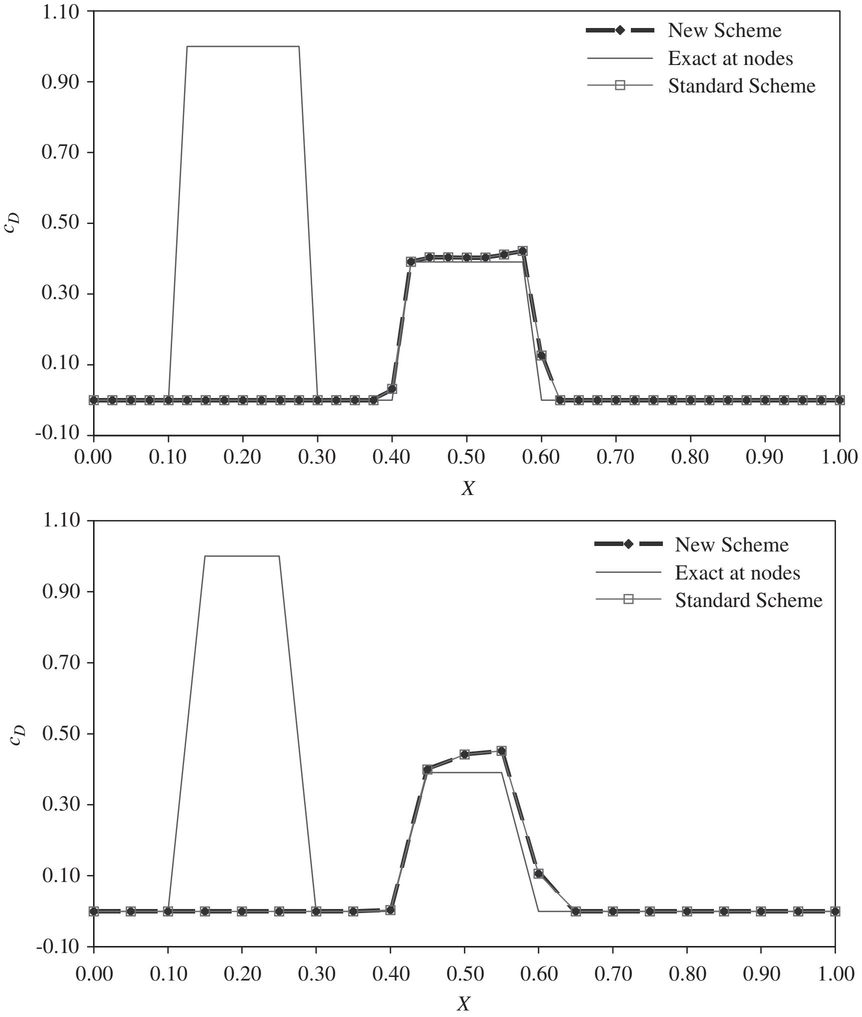


Figure 2. Square wave with decreasing amplitude at $T = 0.0$ and 0.3 with $Co_{hi} = 3/4$, $\theta = 0$, $\alpha_2 = 1$, and $h = 1/20$ (bottom) and $h = 1/40$ (top).

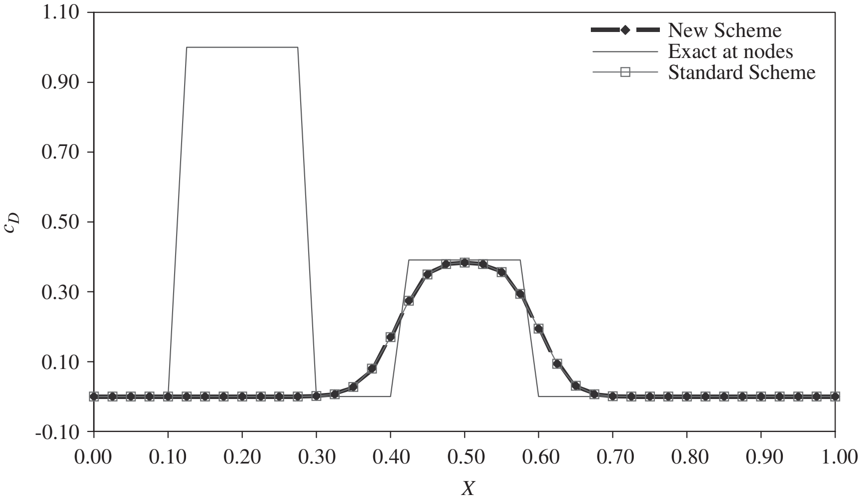
values at each grid point as illustrated in Figure 4. In this figure, results are depicted after 4 time steps ($\Delta T = 0.0375$) for the case $Co = \frac{3}{4}$, $h = 1/20$, $\theta = \frac{1}{2}$ and $\alpha_2 = \frac{1}{2}$.

Case 2: In this case, the square wave increases in amplitude as it propagates in the positive direction. The initial concentration profile is

$$C_{D0} = 0.5 \quad \text{if } 0.1 < x_I < 0.3 \quad \text{and } C_{D0} = 0 \quad \text{otherwise}$$

and $D^* = 1 \times 10^{-9}$. The reaction rate expression is given by

$$R = \mu(C - C^2) \quad \text{therefore } R_1 = \mu \text{ and } R_2 = \mu C \tag{76}$$



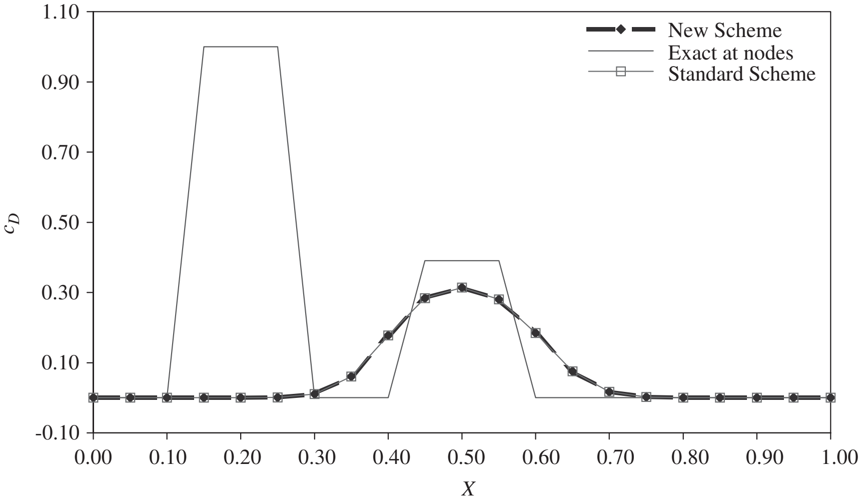


Figure 3. Square wave with decreasing amplitude at $T = 0.0$ and 0.3 with $Co_{hi} = 3/4$, $\theta = 1/2$, $\alpha_2 = 1/2$, and $h = 1/20$ (bottom) and $h = 1/40$ (top).

This expression represents population growth [19]. The system Damkohler numbers corresponding to R_1 and R_2 are given by

$$D_{a1} = \mu C_0 \frac{L}{v} = \mu^*, \quad D_{a2} = \left(\mu C_0 \frac{L}{v}\right) C_D = \mu^* C_D \tag{77}$$

Following the same line of development presented for Case 1, the characteristic solution for the hyperbolic limit is given by ([19, p. 404])

$$C_D(T - T_0) = \frac{C_{D0} e^{\mu^*(T-T_0)}}{1 - C_{D0}(1 - e^{\mu^*(T-T_0)})} \tag{78}$$

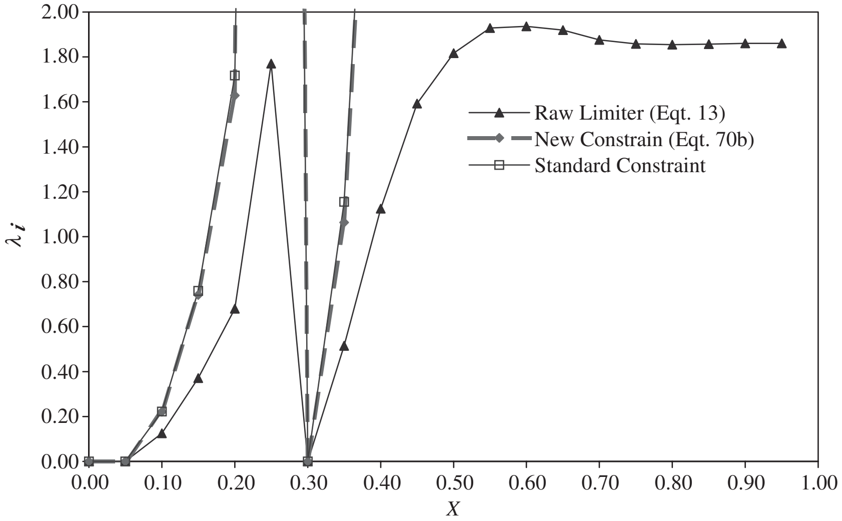


Figure 4. Comparison of raw limiter values calculated from Equation (13) with the standard TVD limiter constraint and the new limiter constraint (calculated from (44b) with $R_1 = 0$, $\alpha_2 = 1/2$, $\theta = 1/2$, $D = 0$).

The dependence of the solution profiles on h, ΔT, and parameters θ, α_1, and α_2 is depicted in Figures 5 and 6. Numerical solutions, at $T = 0.3$, are presented for Courant number $Co_{hi} = 3/4$. Results obtained using a smaller Co_{hi} exhibited the same trends presented in Case 1 and are not presented here.

In this case, both the standard and new schemes produce positive results as shown in Figure 5. The reason for this outcome in the fully explicit case can be explained by examination of equation (70b). Since $D_{a1i} \geqslant D_{a2i}$ the correction to the standard constraint is always positive or zero if $\alpha_1 = \alpha_2 = 1$. Hence, if this correction is neglected as in the standard scheme, the condition for positivity will still be met. However, because the limiter values in general differ slightly between the two cases, the results will also differ slightly as illustrated in Figure 5. As shown, the standard scheme is slightly more diffusive at the trailing edge of the square wave, but it more accurately represents the peak concentration. Differences between the two schemes reduce with decreasing grid spacing. Although not presented here, results for $\theta = 0$ and $\alpha_1 = \alpha_2 = 1$ were again found to be identical for the new and standard schemes since the new and standard limiter constraints are equal for this case. Finally, the behavior of the solutions for different values of α_1 and α_2 are shown in Figure 6. The standard and new schemes yield comparable results with the new scheme producing a concentration peak at the leading edge of the wave. Both schemes yield solutions that travel too fast ($\alpha_1 = 1$, $\alpha_2 = 0$) or too slow ($\alpha_1 = 0$, $\alpha_2 = 1$) depending on the combination of reaction weighting. However, for the case $\alpha_1 = 1$ and $\alpha_2 = 0$ the standard scheme produces negative results at the trailing edge of the wave since (70b) becomes more restrictive for this case.

Case 3: In this test case we consider transport with stronger diffusion both with and without reaction. For the reactive-transport case we consider first-order decay:

$$R = R_2 = -\mu C \tag{79}$$

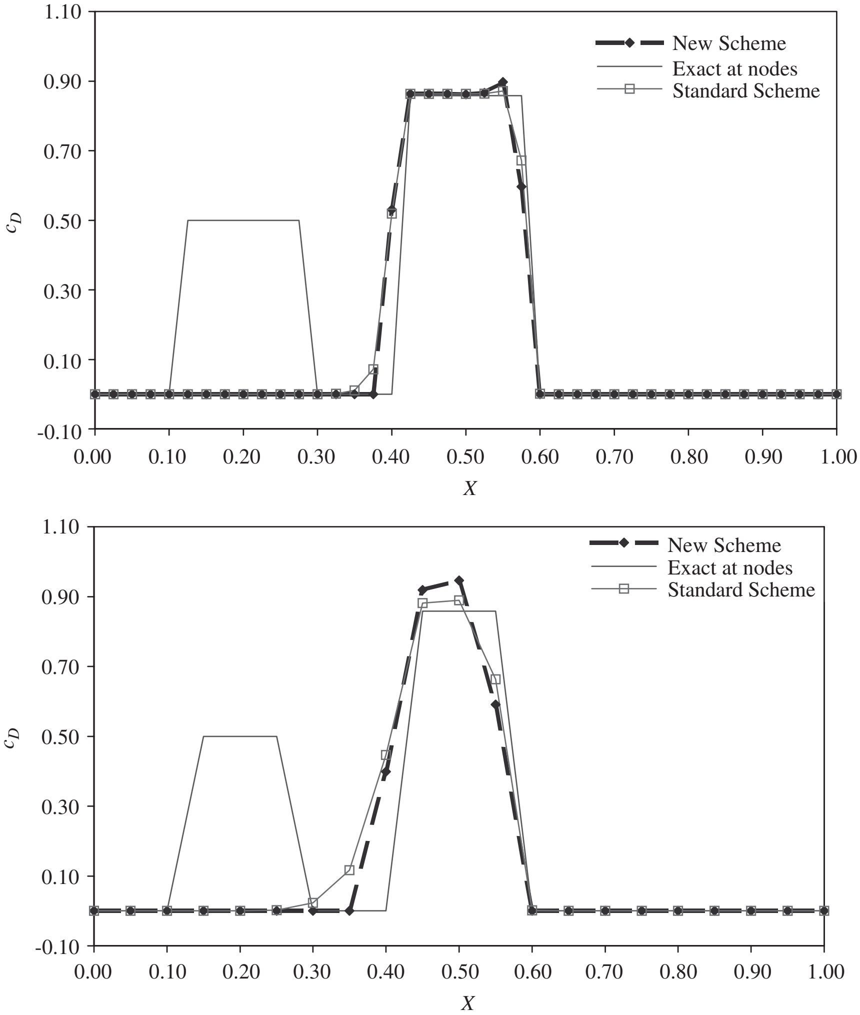


Figure 5. Square wave with increasing amplitude at $T=0.0$ and 0.3 with $Co_{hi}=3/4$, $\theta=0$, $\alpha_1=0$, $\alpha_2=0$ and $h=1/20$ (bottom) and $h=1/40$ (top).

The corresponding system Damkohler number is

$$D_{a1}=\mu C_0\,\frac{L}{v}=\mu^* \tag{80}$$

At the left end of the problem domain a solute is injected for a short time $T\leqslant 0.15$ into a clean solvent, which is flowing left to right. The corresponding essential boundary and initial conditions are:

$$C_D(0,T)=1\ \ 0\leqslant T\leqslant 0.15,\quad C_D(0,t)=0\ \ T\geqslant 0.15,\quad C_D(1,T)=0\ \ T\geqslant 0$$

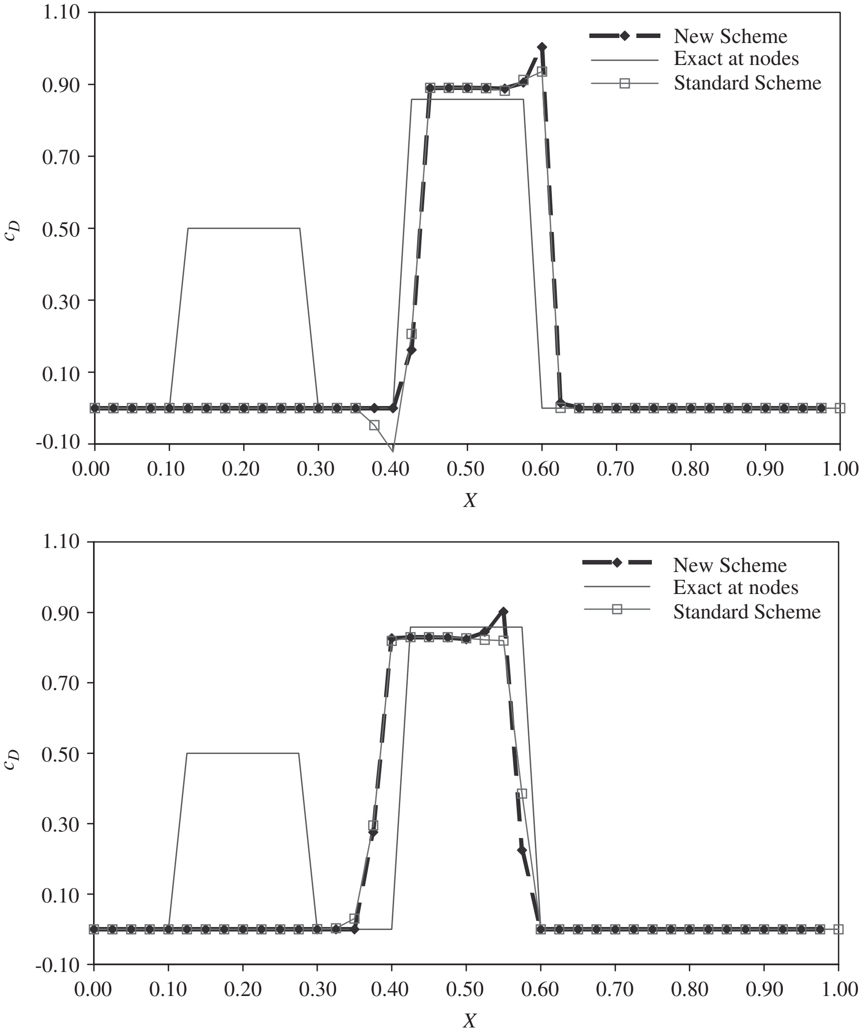


Figure 6. Square wave with increasing amplitude at $T = 0.0$ and 0.3 with $Co_{hi} = \frac{3}{4}$, $\theta = 0$, $h = 1/40$ $\alpha_1 = 0$, $\alpha_2 = 1$ (bottom) and $\alpha_1 = 1$, $\alpha_2 = 0$ (top).

and

$$C_D(X, 0) = 0$$

At sufficiently early times, the analytical solution to this test case, with and without reaction, can be represented by well known analytical solutions for an infinite domain [25] and these results are used here for comparative purposes.

For this test case $D^* = 0.0025$ so diffusion is not negligible. We first consider the case without reaction. Numerical solutions, at $T = 0.2625$ and 0.525, are presented for Courant number $Co_{hi} = \frac{3}{4}$ on grids $h = 1/20$ and $1/40$ and for Courant number $Co_{hi} = 3/8$ on grid $h = 1/40$ in

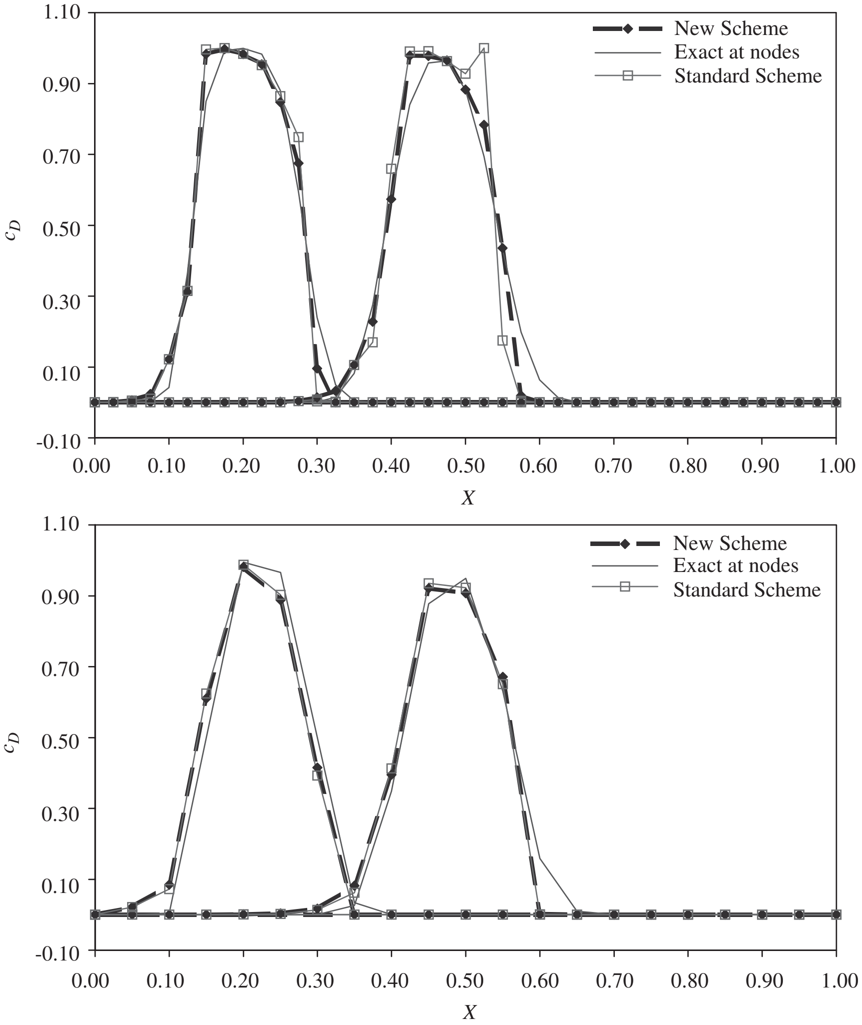


Figure 7. Solution profiles for convection–diffusion at $T = 0.0$ and 0.3 with $Co_{hi} = 3/4$, $\theta = 0$, and $h = 1/20$ (bottom) and $h = 1/40$ (top).

Figure 7. All solutions are positive. Both the new and standard schemes yield comparable and good explicit approximations to the exact solution, except the standard scheme develops a local extremum in the solution at $T = 0.525$. Note that the diffusion term in Equation (31) becomes more important as the grid is refined, leading to increased potential for negative or oscillatory solution behavior. Again, the new and standard schemes again yield identical solution profiles for implicit solutions since the raw limiter values are always less than the constrained values. Finally, reducing the Courant number while keeping the grid spacing fixed leads to an improvement in the accuracy of the solution profile and the new and standard schemes become indistinguishable (Figure 8). Also note that the local extremum in the

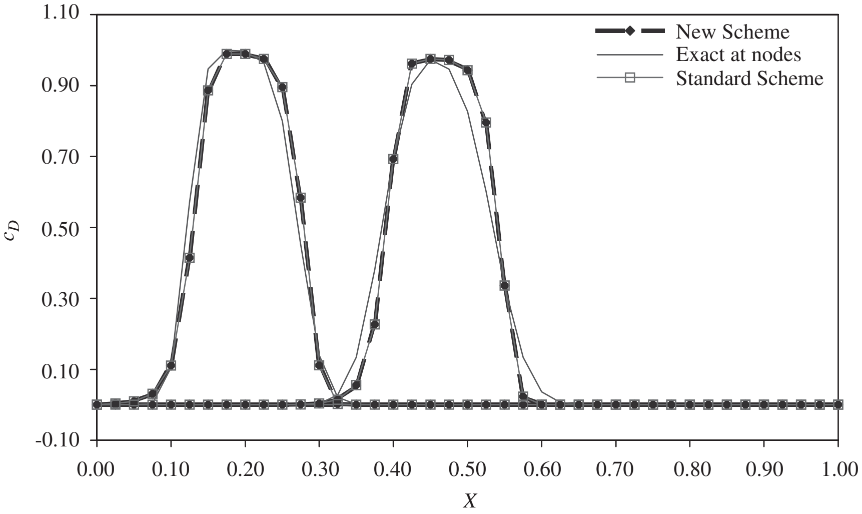


Figure 8. Solution profiles for convection–diffusion at $T = 0.2625$ and 0.525, $Co_{hi} = 3/8$, $h = 1/40$, and $\theta = 0$.

solution no longer occurs, in part because the diffusion term in Equation (31) becomes less significant as the leading term increases in magnitude because of the reduction in Co_{hi}.

The preceding calculations were repeated for the case with reaction. The Damkohler number is set to $D_{a2i} = 4$. Figure 9 shows the solution profiles at times $T = 0.2625$ and 0.525 for $Co_{hi} = \frac{3}{4}$ and different values of h. Results show trends similar to those presented earlier. The new scheme produces solutions that are positive with better approximations of the peak concentration. Further, the standard scheme produces small negative concentrations on both grids (e.g. on the coarse grid, $c(0.3, 0.525) = -6.17 \times 10^{-4}$). For the implicit reaction case (see Figure 10), the new and standard schemes produced identical solution profiles for $h = 1/20$ and slightly different solution profiles for $h = 1/40$. In the latter case the new and standard limiter values are slightly different because the diffusion term in constraint (31) becomes more significant as the grid is refined.

5.2.2. Consistent mass finite-element test case. This final example illustrates in the present context how the consistent mass-matrix Galerkin formulation for a model diffusion problem can produce negative solution values if the lower bound time-step size constraint given by (56) is violated. The issue of preserving positivity for the diffusion equation has also been considered in recent work [17, 26]. For this problem $v = 0$, $D^* = 1.0$, $\theta = 1.0$, and $h = 1/40$. The boundary and initial conditions for this standard model problem, are, respectively,

$$C_D(0, T) = 1, \quad C_D(1, T) = 0, \quad T \geqslant 0$$

and

$$C_D(X, 0) = 0$$

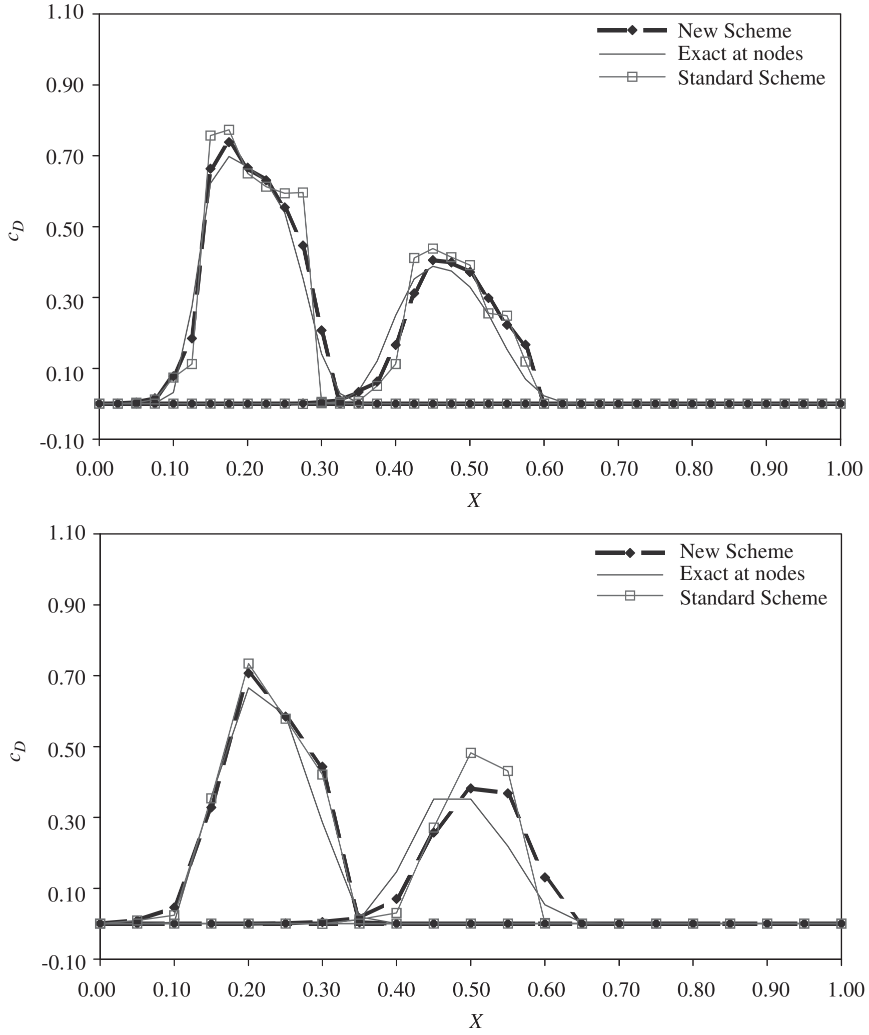


Figure 9. Solution profiles for convection–diffusion–reaction at $T = 0.2625$ and 0.525 with $Co_{hi} = 3/4$, $\theta = 0$, $\alpha_2 = 0$, and $h = 1/20$ (bottom) and $h = 1/40$ (top).

At sufficiently early times, the analytical solution to this test problem can be represented by the well-known analytical conduction solution for an infinite domain ([27, p. 60]).

From condition (56), we find that the consistent mass-matrix Galerkin finite-element approximation to this simple problem must satisfy a minimum time-step size of $\Delta T \geqslant 1.04 \times 10^{-4}$ in order to guarantee a positivity preserving solution. Note that the upper bound on time-step size is given by (66b). Table III presents illustrative results for the first ten grid points after four time steps for $\Delta T = 1.0 \times 10^{-4}$ and $\Delta T = 1.2 \times 10^{-4}$. Positive results were obtained at every time step and grid point for the case where the time-step size condition was satisfied (column 5). Negative solution values were obtained when the time-step size restriction was violated

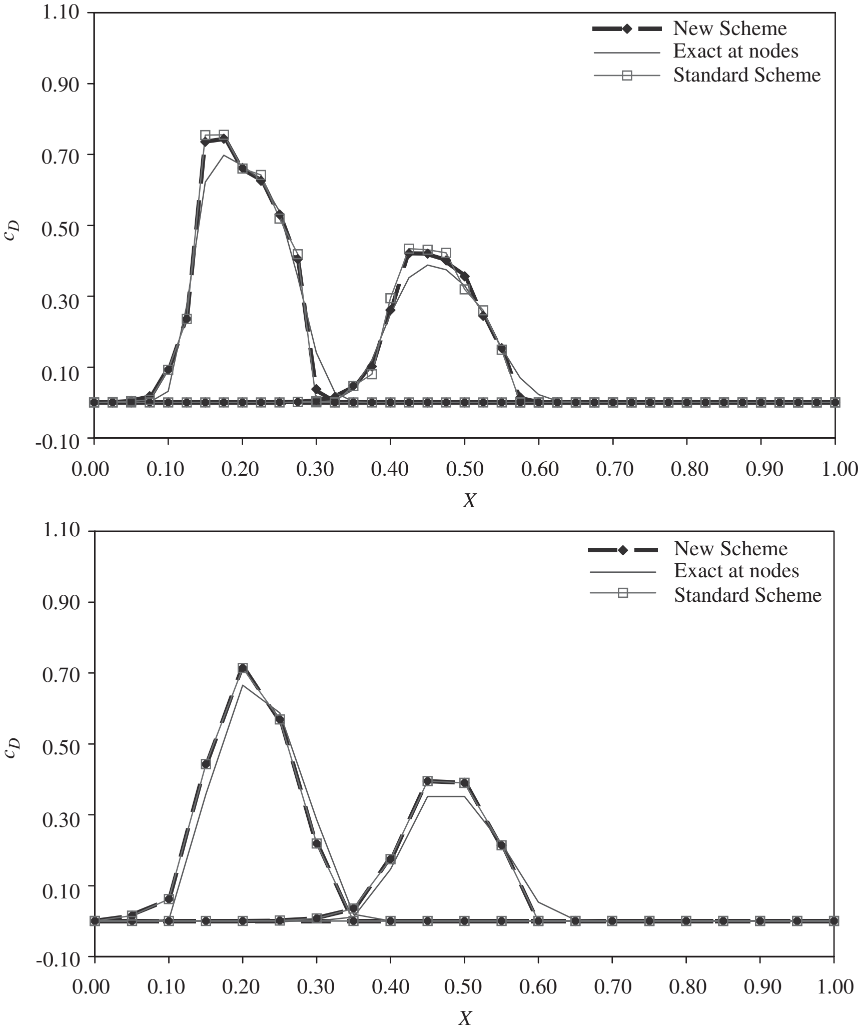


Figure 10. Solution profiles for convection–diffusion–reaction at $T = 0.2625$ and 0.525 with $Co_{hi} = 3/4$, $\theta = 0$, $\alpha_2 = 1$, and $h = 1/40$ (top) and $h = 1/20$ (bottom).

(column 3). Eventually, the negative solution values damped out as the solution evolved in time.

6. EXTENSION TO MULTIDIMENSIONS

Here we briefly outline a simple approach for implementing the new scheme in higher dimensions. In this approach the algorithm is applied separately in each coordinate direction.

Consider the two-dimensional version of model equation (1) discretized in each direction as in the 1D case where Ω is a spatial domain in $\Re^2$. For simplicity, we assume Ω is a rectangular

Table III. Nodal solution results for consistent mass finite-element case.

Node number	x	C_D $\Delta T = 1.0 \times 10^{-4}$	C_D Exact	C_D $\Delta T = 1.2 \times 10^{-4}$	C_D Exact
1	0.00E + 00	1.00E + 00	1.00E + 00	1.00E + 00	1.00E + 00
2	2.50E − 02	2.71E − 01	3.07E − 01	3.00E − 01	3.51E − 01
3	5.00E − 02	2.48E − 02	4.12E − 02	3.91E − 02	6.24E − 02
4	7.50E − 02	−3.48E − 04	2.20E − 03	1.71E − 03	5.19E − 03
5	1.00E − 01	3.57E − 06	4.46E − 05	5.98E − 05	1.94E − 04
6	1.25E − 01	−3.23E − 08	3.34E − 07	1.89E − 06	3.19E − 06
7	1.50E − 01	2.74E − 10	9.14E − 10	5.64E − 08	2.27E − 08
8	1.75E − 01	−2.23E − 12	9.04E − 13	1.62E − 09	6.94E − 11
9	2.00E − 01	1.76E − 14	3.22E − 16	4.54E − 11	9.09E − 14
10	2.25E − 01	−1.36E − 16	4.09E − 20	1.25E − 12	5.06E − 17

domain with sides aligned with the x and y directions and subdivided into $E = E_x \times E_y$ square cells with $h = x_{i+1} - x_i$, $i \in \{1, 2, \ldots, E_x\}$, $h = y_{j+1} - y_j$, $j \in \{1, 2, \ldots, E_y\}$ and the convective fluxes in the x and y directions, vc and uc, are denoted by f_x and f_y, respectively. Following an approach similar to that for the 1D case, we obtain the following sufficient conditions in the x and y directions, respectively,

$$0 \leqslant \frac{\lambda^n_{xi,j}}{\beta^n_{xi,j}} \leqslant \frac{2h\tau^n_{i,j}}{v_{xi,j}\Delta t(1-\theta)} - \frac{4D}{v_{xi,j}h} + \frac{2(1-\alpha_1)}{v_{xi,j}(1-\theta)}\,\tau^n_{i,j}R^n_{1i,j}h - \frac{2(1-\alpha_2)}{v_{xi,j}(1-\theta)}\,\tau^n_{i,j}R^n_{2i,j}h - 2 \quad (81)$$

$$\begin{aligned} 0 \leqslant \frac{\lambda^n_{yi,j}}{\beta^n_{yi,j}} \leqslant & \frac{2h(1-\tau^n_{i,j})}{u_{yi,j}\Delta t(1-\theta)} - \frac{4D}{u_{yi,j}h} + \frac{2(1-\alpha_1)}{u_{yi,j}(1-\theta)}(1-\tau^n_{i,j})R^n_{1i,j}h - \frac{2(1-\alpha_2)}{u_{yi,j}(1-\theta)} \\ & (1-\tau^n_{i,j})R^n_{2i,j}h - 2 \end{aligned} \quad (82)$$

where λ_x and λ_y denote the directional limiters and τ weights terms according to the magnitude of the flux gradient with $0 \leqslant \tau \leqslant 1$. Note that even for the standard hyperbolic case without reaction an extension of the 1D scheme to 2D obtained by simply applying the 1D scheme in each coordinate direction will not guarantee positivity. This is because the leading terms on the R.H.S. of (81) and (82) are in general a fraction of their 1D counterparts. A more detailed treatment is provided in a subsequent study.

7. CONCLUDING REMARKS

We have presented TVD-like, flux-limited, finite-difference and finite-element schemes for solving the scalar convection-diffusion-reaction equation. Sufficient conditions to ensure that these schemes are positivity preserving have been derived based on standard linear algebra concepts. The key contribution is analysis and construction of a flux-limiter constraint that is designed to explicitly account for diffusion and reaction. Numerical examples show that the new schemes are capable of producing both accurate and positive solutions. In particular, numerical results demonstrate that the accuracy of the new and standard TVD-like schemes are comparable but the new schemes have the advantage of guaranteed positivity. Both the new and standard schemes produce solution profiles that become increasingly smeared with increasing θ. For the standard TVD scheme, the implicit cases tend to produce more diffuse solution profiles that lead to raw limiter values that automatically satisfy the upper bound provided by the new limiter constraint condition. Hence the standard TVD scheme produced negative values for the explicit ($\theta = \alpha_1 = \alpha_2 = 0$) cases only. Moreover, we emphasize that the conditions on time-step size, mesh spacing, and flux limiting that are derived here are sufficient but not necessary for positivity. How well this finding holds for problems other than the ones tested and for problems in multidimensions is not known. Finally, we have shown that the consistent mass matrix form of the Petrov–Galerkin method will produce positivity-preserving solutions only if the flux-limited scheme is both implicit and satisfies an additional lower-bound condition on time-step size. We show that this result also applies to standard Galerkin linear finite-element approximation to the diffusion equation.

REFERENCES

1. Harten A. High resolution schemes for hyperbolic conservation laws. *Journal of Computational Physics* 1983; **49**:357.
2. Sweby PK. High resolution schemes using flux limiters for hyberbolic conservation laws. *SIAM Journal on Numerical Analysis* 1984; **21**:995–1011.
3. LeVeque RJ. *Numerical Methods for Conservation Laws*. Birkhauser Verlag: Basal, 1992.
4. Yee HC. Construction of explicit and implicit symmetric TVD schemes and their applications. *Journal of Computational Physics* 1987; **68**:151–179.
5. Yee HC, Klopfer GH, Montagne JL. High-resolution shock-capturing schemes for inviscid and viscous hypersonic flows. *Journal of Computational Physics* 1990; **88**:31–61.
6. Wang Z, Richards BE. High resolution schemes for steady flow computation. *Journal of Computational Physics* 1991; **97**:53–72.
7. Arora M, Roe PL. A well-behaved TVD limiter for high-resolution calculations of unsteady flow. *Journal of Computational Physics* 1997; **132**:3–11.
8. Daru V, Tenaud C. Evaluation of TVD high resolution schemes for unsteady viscous shocked flows. *Computers in Fluids* 2001; **30**:89–113.
9. Andrews MJ. Accurate computation of convective transport in transient two-phase flow. *International Journal for Numerical Methods in Fluids* 1995; **21**:205–222.
10. Gupta AD, Lake LW, Pope GA, Sepehrnoori K. High-resolution monotonic schemes for reservoir fluid flow simulation. *In-situ* 1991; **15**(3):289–317.
11. Blunt M, Rubin B. Implicit flux limiting for petroleum reservoir simulation. *Journal of Computational Physics* 1992; **102**:194–210.
12. Liu J, Pope GA, Sepehrnoori K. A high-resolution finite-difference scheme for nonuniform grids. *Applied Mathematical Modelling* 1995; **19**:162–172.
13. Shi J, Toro EF. Fully discrete high-order shock-capturing numerical schemes. *International Journal for Numerical Methods in Fluids* 1996; **23**:241–269.
14. Hubbard ME. Multidimensional slope limiters for MUSCL-type finite volume schemes on unstructured grids. *Journal of Computational Physics* 1999; **155**:54–74.
15. Suresh A. Positivity-preserving schemes in multidimensions. *SIAM Journal on Scientific Computing* 2000; **22**:1184–1198.
16. Codina R. Comparison of some finite element methods for solving the diffusion-convection-reaction equation. *Computer Methods in Applied Mechanics and Engineering* 1998; **156**:185–210.
17. Berzins M. Modified mass matrices and positivity preservation for hyperbolic and parabolic PDEs. *Communication in Numerical Methods and Engineering* 2001; **17**:658–666.
18. Christie I, Griffiths DF, Mitchell AR, Zienkiewicz OC. Finite element methods for second order differential equations with significant first derivatives. *International Journal for Numerical Methods in Engineering* 1976; **10**:1389–1396.
19. Bailey JE, Ollis DF. *Biochemical Engineering Fundamentals* (2nd edn). McGraw-Hill: New York, 1986.
20. Espenson JH. *Chemical Kinetics and Reaction Mechanisms* (2nd edn). McGraw-Hill: New York, 1995.
21. Kindred JS, Celia MA. Contaminant transport and biodegradation, 2, conceptual model and test simulations. *Water Resources Research* 1989; **25**(6):1149–1160.
22. van Leer B. Towards the ultimate conservative difference scheme. II. *Journal of Computational Physics* 1974; **14**:361–370.

23. Carey GF, Oden JT. *Finite Elements Fluid Mechanics*, vol. VI. Prentice-Hall: Englewood Cliffs, NJ, 1986.
24. Wilcoxson M, Manousiouthakis V. On an implicit ENO scheme. *Journal of Computational Physics* 1994; **115**:376–389.
25. Javandel I, Doughty C, Tsang C. *Groundwater Transport*: *Handbook of Mathematical Models*. Water Resources Monograph Series 10, American Geophysical Union: Washington, DC, 1984.
26. Farago I, Horvath R. On the nonnegativity conservation of finite element solution of parabolic problems. In *3D Finite Element Conference Proceedings*, Krizek M. *et al.* (eds). Gakkotosho Co.: Tokyo, 2001.
27. Carslaw HS, Jaeger JC. *Conduction of Heat in Solids* (2nd edn). Oxford University Press: Oxford, UK, 1959.