SIAM J. SCI. COMPUT.
Vol. 27, No. 4, pp. 1202–1223

ANALYSIS OF BOX SCHEMES FOR REACTIVE FLOW PROBLEMS*

S. L. MITCHELL†, K. W. MORTON‡, AND A. SPENCE‡

Abstract. Key properties of the box scheme are shown to be advantageous for reactive flow problems. Unconditional stability and compact conservation are shown by a detailed modified equation analysis to enable the scheme to reflect exactly the "reduced speed," enhanced diffusion, and dispersion which are typical of such "hyperbolic conservation laws with relaxation." A novel modified equation analysis is also used to show how the spurious checkerboard mode behaves and can be controlled. Numerical experiments for some nonlinear one-dimensional problems and a two-dimensional problem demonstrate that the behavior of the scheme deduced from a simple model problem has general validity.

Key words. box scheme, reactive flow problems, modified equation analysis

AMS subject classifications. 65M06, 65M12, 35L65

DOI. 10.1137/030601910

1. Introduction.

Our objectives in this paper are twofold. First, we wish to develop effective numerical methods for the differential equations which arise in the transport of reacting substances by groundwater flow of the general form

$$\mathbf{u}_t + \operatorname{div} \mathcal{F}(\mathbf{u}) = \mathbf{R}(\mathbf{u}), \quad \mathbf{u}(\mathbf{x}, t) \,:\, \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^p, \tag{1.1}$$

where the fluxes $\mathcal{F}$ represent the groundwater transport and the terms $\mathbf{R}$ on the right represent the chemical reactions. In typical problems the reaction rates vary widely, and some may be on a time scale much shorter than the advection velocities occurring in the transport terms. The result is the phenomenon of "reduced speed" whereby the transport occurs at a speed very much lower than the advection velocity; this precludes the use of simple explicit numerical schemes for modeling the transport.

Such situations are common in other applications, such as river-flow modeling. And there the standard numerical remedy is to use the box scheme, often called the Preissmann box scheme; see [6] and [15]. The name of the scheme derives from its construction by integrating the differential equation over a box in $(\mathbf{x}, t)$-space, applying the divergence theorem to the differential terms, and approximating all the integrals by quadrature formulae based on values at the corners of the box. This gives a very compact difference scheme on a uniform mesh (or a finite volume scheme on a general mesh) which has natural conservation properties. It is also implicit and unconditionally stable. For it to be fully effective, two steps are necessary. First, an efficient marching scheme is needed to solve the implicit system of equations by sweeping out from the given initial and boundary conditions. Second, the time step needs to be chosen so that the dominant flow of "information" corresponds roughly to a main diagonal of the discretization box. Then the remaining disadvantage of the scheme is the presence of spurious oscillations in the numerical solution, which have to be controlled in some way.

*Received by the editors November 19, 2003; accepted for publication (in revised form) January 10, 2005; published electronically January 6, 2006.
http://www.siam.org/journals/sisc/27-4/60191.html
†Institute of Applied Mathematics, University of British Columbia, Vancouver, BC, V6T 1Z2, Canada (sarah@iam.ubc.ca).
‡Department of Mathematical Sciences, University of Bath, Claverton Down, BA2 7AY, UK (bill.morton@comlab.ox.ac.uk, as@maths.bath.ac.uk).

Thus our second objective is to explore the properties of the box scheme and show how it may best be used for the problems under consideration. In the next section we begin with the study of a linear model problem in one space dimension in order to introduce most of the key ideas. We show by an asymptotic analysis how the reaction terms lead to the reduced speed, and we also introduce diffusion and dispersion into the simple model. In section 3 we consider the box scheme for this problem—its stability by an energy method and its modified equation analysis. For smooth approximations it reproduces the reduced speed, diffusion, and dispersion of the differential equation, with the last achieved by choosing the time step so that the CFL number based on the reduced speed is set to unity; a novel modified equation analysis of the spurious oscillatory modes shows how they move at a completely different speed and how they can be damped by applying a general θ-weighting to the advective terms with $\theta = \frac{1}{2} + \mathcal{O}(\Delta t)$. To demonstrate that these key properties extend to more general problems, in section 4 the method is applied to one-dimensional problems with nonlinear reaction terms and in section 5 to a two-dimensional problem with an incompressible flow field and a nonuniform mesh.

To conclude this introduction, we should point out that the standard schemes in widespread use for modeling reactive transport in groundwater flow make use of operator-splitting techniques—see, for example, [8], [19], [7], [24], [2], [3], and [1]. That is, one solves alternately the equilibrium equation system $\mathbf{R}(\mathbf{u}) = 0$ and the transport equation. There are several disadvantages of such a procedure, as has been pointed out in [3] and [7]. These include the fact that the transport can occur at the incorrect speed unless the reaction rates are sufficiently fast, an $\mathcal{O}(\Delta t)$ error arises due to the splitting, and stability constraints determined by the velocity of the transport equation are possible. Nonetheless, splitting schemes are computationally attractive and form the basis for several software packages, notably PARSIM [1] and MINTRAN [19].

2. A linear model in one space dimension. The simplest model that displays the key features we wish to study is for a single chemical component being transported through rock by steady groundwater flow and reacting with the rock. Let c be the concentration of the chemical in the water, s the concentration in the rock, λ the rate of adsorption into the rock, μ the rate of desorption from the rock into the water, and V the groundwater flow velocity. Then we have the following pair of equations, with λ, μ, and V all positive,

$$\frac{\partial c}{\partial t} + V\frac{\partial c}{\partial x} = \mu s - \lambda c, \tag{2.1}$$

$$\frac{\partial s}{\partial t} = \lambda c - \mu s, \tag{2.2}$$

to hold on $(x,t) \in [0,\infty) \times [0,\infty)$.

In a typical problem the concentrations are zero initially, but then there is a release of the chemical at one point over a short period of time, followed by its dispersal through the rock. Thus we shall take as initial and boundary conditions

$$c(x,0) = s(x,0) = 0 \text{ for } x \geq 0, \quad c(0,t) = h(t), \text{ given, for } t > 0; \tag{2.3}$$

then $s(0,t)$ is found from integrating (2.2) with respect to t.

The general behavior of this system is easy to deduce. At each point in the rock the concentration s is obtained by integrating (2.2), while the concentration c

is advected to the right with velocity V but decays through the term $-\lambda c$ and is augmented by μs. For this linear model the result can be expressed in the form of two integral equations which can be solved by the use of Laplace transforms—see [16] and [14] for details. Examples of the resulting (exact) solution are shown in Figure 1, which can be used to judge the accuracy of our numerical methods; note that, in common with the usual practice in this field, we plot the solution against time at a given point in space. Key features of the solutions obtained for different values of λ and μ are the speed of propagation of the initial pulse and its change of form. Some of these features can be deduced by simple, generally applicable, asymptotic analysis, the validity of which can again be checked for this model problem against the Laplace transform solution.

2.1. Asymptotic analysis. For large values of λ and μ, one may apply the analysis of [13] (see also [4]) for *hyperbolic conservation laws with relaxation*. These are defined as systems of the form (1.1) in which there exists a constant matrix Q with rank $r < p$ such that $Q\mathbf{R}(\mathbf{u}) = 0$ and $\mathbf{R} = \epsilon^{-1}\tilde{\mathbf{R}}$, where ϵ is a small parameter. Then this defines an *equilibrium manifold* given by the mapping $\mathcal{M} : \mathbf{v} \in \mathbb{R}^r \to \mathbf{u}^e \in \mathbb{R}^p$, on which $\tilde{\mathbf{R}}(\mathbf{u}^e) = 0$. In one space dimension, where we write (1.1) as $\mathbf{u}_t + \mathbf{f}_x = \mathbf{R}$ with $\mathbf{f} : \mathbb{R}^p \to \mathbb{R}^p$, we can define a new flux $\mathbf{g} : \mathbb{R}^r \to \mathbb{R}^r$ by $\mathbf{g}(\mathbf{v}) = Q\mathbf{f}(\mathcal{M}\mathbf{v})$; thence, by operating on the equation with the constant matrix Q, we obtain an *equilibrium model*

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{g}(\mathbf{v})}{\partial x} = 0, \tag{2.4}$$

whose solution provides a lowest order approximation to $\mathbf{u}$ as $\epsilon \to 0$.

For the linear model (2.1), (2.2), the equilibrium manifold is given by $\lambda c = \mu s$ with Q the row vector $(1, 1)$. Hence we add the two equations to give

$$s_t + c_t + Vc_x = 0, \tag{2.5}$$

which is an important equation in its own right. Then, substituting for s in this equation gives the equilibrium model

$$c_t + \frac{\mu V}{\lambda + \mu} c_x = 0, \tag{2.6}$$

which shows how the reduced speed $V' = \mu V/(\lambda + \mu)$ comes into play in the limit of large λ and μ. Strictly speaking, in the sense of [13], the equilibrium model should be written in terms of $c + s$, but for this linear problem c, s, and $c + s$ all satisfy (2.6) so we use the simpler form.

Further levels of approximation are obtained by expanding the mapping $\mathcal{M}$ in powers of the small parameter, the reciprocals of λ and μ in the present case. From differentiating (2.2) we obtain

$$s_t = (\lambda/\mu)c_t - (1/\mu)s_{tt}, \tag{2.7}$$

and then to the lowest order we can derive the approximation

$$s_{tt} = (\lambda/\mu)c_{tt} - (1/\mu)s_{ttt} \sim (\lambda/\mu)V'^2 c_{xx}. \tag{2.8}$$

Substitution back into the conservation equation (2.5) then gives

$$c_t + \frac{\mu V}{\lambda + \mu} c_x = \frac{\lambda \mu V^2}{(\lambda + \mu)^3} c_{xx}, \tag{2.9}$$

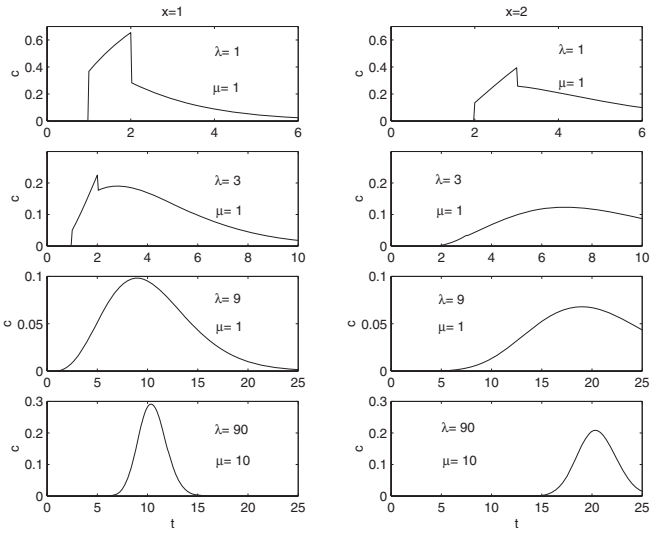

FIG. 1. *The Laplace transform solution* c *at* $x = 1$ *and* $x = 2$ *for* $V = 1$ *for various values of* (λ, μ). *The boundary condition is a square pulse of height and width* 1.

commonly called the *improved equilibrium model*. Continuing to one order higher introduces a dispersion term to give

$$c_t + \frac{\mu V}{\lambda + \mu} c_x = \frac{\lambda \mu V^2}{(\lambda + \mu)^3} c_{xx} - \frac{\lambda \mu (\lambda - \mu) V^3}{(\lambda + \mu)^5} c_{xxx}. \tag{2.10}$$

From the plots in Figure 1 we can clearly see the effect not only of the reduced speed but also of the diffusion term in the improved model (2.9) and the dispersion term in (2.10). In the four cases shown we have $V = 1$ and the values of λ and μ are $(1, 1)$, $(3, 1)$, $(9, 1)$, and $(90, 10)$. These give values of the reduced velocity $V' = \mu V/(\lambda + \mu)$ equal to $1/2$, $1/4$, $1/10$, and $1/10$, respectively. In Table 2.1 these are compared with estimates of the actual speed of the initial pulse deduced from the solutions in the figures. Note too how the solutions are smoothed and damped, particularly in the last two cases, and in the table the values of the diffusion and dispersion coefficients in the improved equilibrium models are also given.

TABLE 2.1

Table showing observed and reduced speed, and comparing the estimate for $c_{\max}$ *from* (2.12) *with the observed value (at* $x = 2$ *with the speed* $V = 1$*). Also shown are the diffusion and dispersion coefficients in* (2.10).

	$(\lambda, \mu) = (1, 1)$	$(\lambda, \mu) = (3, 1)$	$(\lambda, \mu) = (9, 1)$	$(\lambda, \mu) = (90, 10)$
Observed speed	1.00	0.31	0.11	0.10
Reduced speed	0.50	0.25	0.10	0.10
Max observed c	0.394	0.123	0.068	0.208
$c_{\max}$ (from (2.12))	0.200	0.115	0.067	0.210
Diffusion coefficient	0.125	0.047	0.009	9×10^{-4}
Dispersion coefficient	0	0.006	7.2×10^{-4}	7.2×10^{-6}

We have already referred to the similarity of the present problems to those encountered in river modeling. In a classic paper [11], [12], Lighthill and Whitham carried out an asymptotic analysis of flood waves in which they derived an expression

for the maximum flood height as well as the speed of the wave. If we differentiate (2.1) with respect to t and add to it μ times the conservation equation (2.5), we can eliminate s and obtain the following second order equation for c:

$$c_{tt} + Vc_{tx} + (\lambda + \mu)c_t + \mu V c_x = 0. \tag{2.11}$$

Following the procedure used in [11], [12], and [22] we can apply a Laplace transform to this equation, with the initial and boundary conditions given in (2.3), and obtain an expression for the solution in terms of a contour integral. Then the solution near the dynamic wave-front can be deduced from the behavior of the integrand for large values of the transform variable. In particular, we find that the maximum occurs where $x = V't$ and for large values of $\lambda + \mu$ is given by

$$c_{\max} \sim \frac{1}{2}\left(\frac{(\lambda+\mu)^3}{\pi\lambda\mu t}\right)^{\frac{1}{2}}. \tag{2.12}$$

Values of this expression are also compared with the observed maxima for the four exact solutions in Table 2.1.

3. The box scheme for the linear model. The box scheme became a popular method for approximating the St. Venant equations of river flow through the work of Preissmann and his colleagues [6]. The scheme is also associated with the name of Keller, who applied it to parabolic problems [10] and to two-point boundary value problems [9]. It was also analyzed in the early paper [18] and very recently in [15]. It can be identified as the valuable cell-vertex finite volume method for approximating the Euler equations (and also the Navier–Stokes equations) of steady transonic aerodynamics (see, for example, [5]), and its multisymplectic structure has been recognized and has led to its application to the KdV equation [23]. When applied to these cases, a valuable feature of the scheme, which follows from the derivation outlined in the introduction, is the accurate and compact discretization that it yields on a nonuniform quadrilateral mesh. However, we shall concentrate here on its finite difference formulation on a uniform rectangular mesh, because we can demonstrate that its asymptotic behavior closely mirrors that of the differential system as outlined in the previous section. We shall also begin by using the simple trapezoidal rule weighting in time, although we shall show later, as is well known, that the more general θ-weighting is important in controlling the spurious checkerboard mode.

Suppose we have a uniform mesh over the rectangle $0 \le t \le T$, $0 \le x \le L$, with time step $\Delta t = T/N$ and spatial step $\Delta x = L/J$ for a given N and J. Let U_j^n denote the numerical approximation of u at $(x_j, t^n) := (j\Delta x, n\Delta t)$ for $j = 0, 1, \ldots, J$ and $n = 0, 1, \ldots, N$. On the mesh we define the finite difference operators

$$\delta_x U_{j+\frac{1}{2}}^{n+\frac{1}{2}} = U_{j+1}^{n+\frac{1}{2}} - U_j^{n+\frac{1}{2}}, \quad \mu_x U_{j+\frac{1}{2}}^{n+\frac{1}{2}} = \tfrac{1}{2}\left(U_{j+1}^{n+\frac{1}{2}} + U_j^{n+\frac{1}{2}}\right) \tag{3.1}$$

and, similarly, δ_t and μ_t. Then the box scheme consistently uses the operator $\mu_x\delta_t$ to approximate ∂_t and, similarly, $\mu_t\delta_x$ to approximate ∂_x in a partial differential equation (PDE). If we apply this scheme to the conservation law (2.6), we obtain

$$\mu_x\delta_t C_{j+\frac{1}{2}}^{n+\frac{1}{2}} + \mu_x\delta_t S_{j+\frac{1}{2}}^{n+\frac{1}{2}} + p\mu_t\delta_x C_{j+\frac{1}{2}}^{n+\frac{1}{2}} = 0, \text{ where } p = V\Delta t/\Delta x \tag{3.2}$$

and p denotes the CFL number. This is a very compact scheme which uses four neighboring values of the unknowns, and it is actually *implicit* as it involves two points

at the new time level. However, for appropriate initial and boundary conditions such as those given in (2.3), the discrete equations can be solved *explicitly* by marching away from a boundary. To complete the approximation we have to discretize the reaction equation (2.2), which is an ordinary differential equation in time, to which we apply the one-dimensional box scheme (i.e., the trapezoidal rule) to obtain

$$\delta_t S_j^{n+\frac{1}{2}} = \mu_t\big(\lambda' C_j^{n+\frac{1}{2}} - \mu' S_j^{n+\frac{1}{2}}\big), \tag{3.3}$$

where $\lambda' := \lambda\Delta t$ and $\mu' := \mu\Delta t$. Note that this equation needs no boundary condition as its application for $j = 0$ integrates the boundary data for c up the t-axis.

Equations (3.2) and (3.3) can be written out as

$$\begin{aligned}(1+p)C_{j+1}^{n+1} + (1-p)C_j^{n+1} - (1-p)C_{j+1}^n - (1+p)C_j^n \\ + \big(S_{j+1}^{n+1} - S_{j+1}^n\big) + \big(S_j^{n+1} - S_j^n\big) = 0,\end{aligned} \tag{3.4}$$

$$S_{j+1}^{n+1} = \frac{\frac{1}{2}\lambda'}{1+\frac{1}{2}\mu'}\big(C_{j+1}^{n+1} + C_{j+1}^n\big) + \left(\frac{1-\frac{1}{2}\mu'}{1+\frac{1}{2}\mu'}\right) S_{j+1}^n, \tag{3.5}$$

and they are applied for $j = 0, \ldots, (J-1)$ and $n = 0, \ldots, (N-1)$. This is an explicit scheme when the conditions of (2.3) are used: from (3.5), S_{j+1}^{n+1} can be substituted into (3.4) to obtain an explicit formula to solve for C_{j+1}^{n+1}, by marching away from the boundary.

3.1. Stability analysis. The box scheme is centered in x and t and the trapezoidal scheme is centered in t, so all the odd order terms in a local truncation error analysis cancel and the scheme is second order accurate. Moreover, a Fourier analysis of the method shows that it is Lax–Richtmyer stable for all Δx and Δt. (See [14] for details.) Unfortunately, the scheme suffers from the notorious checkerboard mode $(-1)^{j+n}$. This is clearly a solution of (3.2), and when the reaction equation approximation (3.3) is space-averaged (a form we shall sometimes use in the analysis below), it is a solution of that too. Thus it is clearly a spurious solution of the complete system. However, applying the unaveraged (3.3) along the t-axis provides special boundary data which minimizes its effect, so that it is then mainly initiated by nonsmooth boundary data for c.

A more illuminating analysis of stability is provided by an energy analysis. We consider first the differential equations (2.1) and (2.2): multiplying (2.1) by c and (2.2) by s and integrating in x over $(0, \infty)$ gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^\infty \frac{1}{2}c^2\,\mathrm{d}x + V\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{2}c^2\right)\mathrm{d}x = -\int_0^\infty c(\lambda c - \mu s)\,\mathrm{d}x, \tag{3.6}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^\infty \frac{1}{2}s^2\,\mathrm{d}x = \int_0^\infty s(\lambda c - \mu s)\,\mathrm{d}x. \tag{3.7}$$

We assume the boundary data is zero for $t > t^*$ and $c(x,t) \to 0$ as $x \to \infty$ for all finite t, so the second integral in (3.6) is then zero; hence if we multiply (3.6) and (3.7) by λ and μ, respectively, and add the results, we deduce that for $t > t^*$

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^\infty \frac{1}{2}\left(\lambda c^2 + \mu s^2\right)\,\mathrm{d}x = -\int_0^\infty (\lambda c - \mu s)^2\,\mathrm{d}x \le 0. \tag{3.8}$$

To obtain a similar result for the box scheme we first introduce $R := \lambda C - \mu S$. Then we rewrite the box scheme of (3.2) and (3.3) by substituting for S in the first

and applying the space average to the second to obtain

$$\mu_x\big[C^{n+1}_{j+\frac12}-C^n_{j+\frac12}\big]+\tfrac12 p\delta_x\big[C^{n+1}_{j+\frac12}+C^n_{j+\frac12}\big]=-\tfrac12\Delta t\mu_x\big[R^{n+1}_{j+\frac12}+R^n_{j+\frac12}\big], \tag{3.9}$$

$$\mu_x\big[S^{n+1}_{j+\frac12}-S^n_{j+\frac12}\big]=\tfrac12\Delta t\mu_x\big[R^{n+1}_{j+\frac12}+R^n_{j+\frac12}\big]. \tag{3.10}$$

For convenience, we introduce the notation $\bar C^n := \mu_x C^n_{j+1/2}$, etc. and $\langle \bar C^n, \bar S^n\rangle_2$ for a typical l^2 inner product. To emulate the procedure carried out in the continuous case, we multiply (3.9) and (3.10) by $\bar C^{n+1}+\bar C^n$ and $\bar S^{n+1}+\bar S^n$, respectively, and sum over the mesh. Then from the identity $\langle \mu_x W, \delta_x W\rangle_2 \equiv \frac12(W_J^2-W_0^2)$, and assuming that $t^n > t^*$ so that $C^n_0=0$, we obtain

$$\begin{aligned}\|\bar C^{n+1}\|_2^2-\|\bar C^n\|_2^2+\tfrac12 V\Delta t\big[C^{n+1}_J+C^n_J\big]^2&=-\tfrac12\Delta t\langle \bar C^{n+1}+\bar C^n,\bar R^{n+1}+\bar R^n\rangle_2\\ \|\bar S^{n+1}\|_2^2-\|\bar S^n\|_2^2&=\tfrac12\Delta t\langle \bar S^{n+1}+\bar S^n,\bar R^{n+1}+\bar R^n\rangle_2.\end{aligned}$$

Multiplying the first expression by λ and the second by μ, and adding gives

$$\begin{aligned}&\lambda\|\bar C^{n+1}\|_2^2+\mu\|\bar S^{n+1}\|_2^2-\big[\lambda\|\bar C^n\|_2^2+\mu\|\bar S^n\|_2^2\big]\\ &\qquad=-\tfrac12\Delta t\|\bar R^{n+1}+\bar R^n\|_2^2-\tfrac12 V\Delta t\big[C^{n+1}_J+C^n_J\big]^2\le 0.\end{aligned} \tag{3.11}$$

Hence we have established the strong stability of the *cell averages* $\mu_x C$ and $\mu_x S$.

However, this does not cover the spurious mode $(-1)^{j+n}$ oscillations. We could treat these by mapping between the cell averages and the nodal values, which we can do since we have prescribed the boundary value, setting it to zero for $t^n > t^*$: every nodal value can then be obtained from the averages by a recurrence ($C^n_1 = 2\bar C^n_{1/2}$, $C^n_2 = 2(\bar C^n_{3/2}-\bar C^n_{1/2})$, etc.), which defines an oscillation matrix, from the l^2 norm of which we deduce that $\|C^n\|_2 \le \sqrt{2J(J+1)}\|\bar C^n\|_2$. Hence there is a potential linear growth which is consistent with the well-known phenomenon that imposing inappropriate boundary conditions may cause a linear growth in the oscillatory mode [17]. It is more useful, though, to consider these modes by means of a modified equation analysis; by this means we can also show how they can be damped by introducing a θ-weighting in the time averaging for the spatial derivative. This is done next.

3.2. Modified equation analysis. The use of a modified equation analysis is a well-known technique for understanding the properties of a finite difference approximation of a given PDE. Ignoring the highest frequencies, the analysis yields an asymptotic series of PDEs that can provide a more accurate representation of the behavior of the solution of the difference scheme. Such an analysis can be especially helpful in giving a qualitative understanding of features like diffusion and dispersion in a numerical scheme (see [21] and [15]).

The finite difference operators defined by (3.1) can be expanded in terms of differential operators using Taylor series expansions, i.e.,

$$\delta_x=\Delta x\big[1+\tfrac{1}{24}\Delta x^2\partial_x^2+\mathcal{O}(\Delta x^4)\big]\partial_x,\quad \mu_x=1+\tfrac18\Delta x^2\partial_x^2+\mathcal{O}(\Delta x^4), \tag{3.12}$$

with similar expansions for δ_t and μ_t. Subsequently, we shall use the notation "..." to mean either $\mathcal{O}(\Delta x^4)$ or $\mathcal{O}(\Delta t^4)$. Strictly speaking, we need to replace the discrete variables by suitable prolongations (for example, using Fourier expansions or high order polynomials) before we can apply the differential operators to them, but we shall not make that distinction in our notation here.

If we define the operators

$$\mathcal{D}_x := \frac{\delta_x \mu_x^{-1}}{\Delta x}, \quad \mathcal{D}_t := \frac{\delta_t \mu_t^{-1}}{\Delta t}, \tag{3.13}$$

then we readily obtain

$$\mathcal{D}_x = \left(1 - \tfrac{1}{12}\Delta x^2 \partial_x^2 + \ldots\right) \partial_x, \tag{3.14}$$

while for the corresponding expansion for $\mathcal{D}_t$ it is convenient to invert the difference operator bracket to obtain

$$\partial_t = \left(1 + \tfrac{1}{12}\Delta t^2 \partial_t^2 + \ldots\right) \mathcal{D}_t. \tag{3.15}$$

Moreover, we can rewrite (3.2) and (3.3) as

$$\mathcal{D}_t C + \mathcal{D}_t S + V \mathcal{D}_x C = 0, \quad (\mathcal{D}_t + \mu) S = \lambda C. \tag{3.16}$$

Since all the operators commute, we can eliminate S to obtain a second order equation for C,

$$[\mathcal{D}_t^2 + (\lambda + \mu + V\mathcal{D}_x)\mathcal{D}_t + \mu V \mathcal{D}_x] C = 0 \tag{3.17}$$

(cf. (2.11)). Because the model is linear this equation also holds for S and for the total concentration $C + S$. It can also be interpreted as a quadratic for $\mathcal{D}_t$, with the two roots corresponding to two wave modes.

The solutions to the quadratic equation are functions of the operator $\mathcal{D}_x$: if we expand these as power series, it is easily seen that we get

$$\mathcal{D}_t = -\frac{\mu V}{\lambda + \mu}\mathcal{D}_x + \frac{\mu\lambda V^2}{(\lambda+\mu)^3}\mathcal{D}_x^2 - \frac{\mu\lambda(\lambda-\mu)V^3}{(\lambda+\mu)^5}\mathcal{D}_x^3 + \ldots \tag{3.18}$$

for the positive root, which corresponds to the main advected wave, and

$$\mathcal{D}_t = -(\lambda + \mu) - \frac{\lambda V}{\lambda + \mu}\mathcal{D}_x - \frac{\mu\lambda V^2}{(\lambda+\mu)^3}\mathcal{D}_x^2 + \frac{\mu\lambda(\lambda-\mu)V^3}{(\lambda+\mu)^5}\mathcal{D}_x^3 - \ldots \tag{3.19}$$

for the negative root, which corresponds to a damped wave which is also advected but with a different speed.

The correspondence of the leading terms in (3.18) to those in the improved equilibrium models (2.9) and (2.10) is very clear: it follows, of course, from the consistent way that the box scheme discretizes the differential equation, which is the basis of its multisymplectic properties already referred to. But now we can use the expansion in (3.14) for all the terms in $\mathcal{D}_x$. Moreover, we can apply the similar expansion in (3.15) to $\mathcal{D}_t$ on the left and obtain, after a straightforward calculation and applying the operators to C, the following expansion (to second order in Δx) for C_t:

$$\begin{aligned} C_t = &-\frac{\mu V}{\lambda+\mu}C_x + \frac{\mu\lambda V^2}{(\lambda+\mu)^3}C_{xx} \\ &- \frac{\mu V}{\lambda+\mu}\left(\frac{\lambda(\lambda-\mu)V^2}{(\lambda+\mu)^4}C_{xxx} + \frac{1}{12}\Delta t^2 C_{xtt} - \frac{1}{12}\Delta x^2 C_{xxx}\right) + \ldots. \end{aligned} \tag{3.20}$$

This is in the usual form found in the first stage of a conventional modified equation analysis, namely, with some higher order t-derivatives on the right-hand side;

these can then be eliminated in favor of x-derivatives by successive differentiation and substitution from this equation. Alternatively, we can use finite difference operator calculus to write $\mathcal{D}_t = (\frac{1}{2}\Delta t)^{-1}\tanh(\frac{1}{2}\Delta t\partial_t)$, take the inverse of this relation, and, from the Taylor expansion of $\tanh^{-1}$, obtain the relation

$$\partial_t = \left(1 + \tfrac{1}{12}\Delta t^2\mathcal{D}_t^2 + \ldots\right)\mathcal{D}_t, \tag{3.21}$$

into which we can substitute (3.18). Either procedure produces

$$\begin{aligned} C_t = &-V'C_x + \frac{\lambda\mu V^2}{(\lambda+\mu)^3}C_{xx} \\ &- V'\left(\frac{\lambda(\lambda-\mu)V^2}{(\lambda+\mu)^4} + \frac{1}{12}\left[V'^2\Delta t^2 - \Delta x^2\right]\right)C_{xxx} + \ldots, \end{aligned} \tag{3.22}$$

where $V' = V\mu/(\lambda+\mu)$ is the reduced speed. Similarly and to the same order, from (3.19) we obtain for the damped wave

$$\begin{aligned} C_t = &-(\lambda+\mu)\left(1 + \frac{1}{12}(\lambda+\mu)^2\Delta t^2\right)C - \bar{V}\left(1 + \frac{1}{4}(\lambda+\mu)^2\Delta t^2\right)C_x \\ &- \bar{V}\left(\frac{V'}{\lambda+\mu} + \frac{1}{4}(\lambda+\mu)V\Delta t^2\right)C_{xx} \\ &+ \bar{V}\left(\frac{V'(\bar{V}-V')}{(\lambda+\mu)^2} - \frac{1}{12}(V^2+VV'+V'^2)\Delta t^2 + \frac{1}{12}\Delta x^2\right)C_{xxx}\ldots, \end{aligned} \tag{3.23}$$

where $\bar{V} = V\lambda/(\lambda+\mu)$ is the basic speed of this wave.

For the main advective wave, the lowest order discretization terms appear in the dispersive term—consistent with the second order accuracy. But the key point to note is that these become zero when the mesh ratio is based on the reduced speed, i.e., when $V'\Delta t = \Delta x$; the dispersive coefficient then reduces to that obtained in section 2.1 for the differential equation. The improved accuracy obtained with this choice is demonstrated in the numerical results given in section 3.5. Note, however, that for the damped wave the dominant effect of the discretization is to increase the damping, although it similarly affects the advective speed and the diffusion term.

The above modified equation analysis, with its emphasis on Taylor expansions of smooth solutions, is unsuitable for studying high frequency modes and in particular the checkerboard mode. The importance of this mode and some of the ways in which it can affect the result of using the box scheme are demonstrated in Figures 2, 3, and 4; there we see that it is greatly enhanced by discontinuous boundary data and that its behavior is markedly different for different values of λ and μ. In the next section we therefore present an extension of the above analysis to describe this behavior, pointing to some of the ways in which the mode may be controlled.

3.3. Separating the smooth and oscillatory numerical solution. Suppose we consider a Fourier mode $C_j = \mathrm{e}^{\mathrm{i}kj\Delta x}$ substituted into the box scheme. The Fourier symbols of the spatial difference operators μ_x and δ_x appearing in (3.2) are given by the relations

$$\mu_x C_{j+\frac{1}{2}} = (\cos\tfrac{1}{2}k\Delta x)C_{j+\frac{1}{2}}, \quad \delta_x C_{j+\frac{1}{2}} = (2\mathrm{i}\sin\tfrac{1}{2}k\Delta x)C_{j+\frac{1}{2}}; \tag{3.24}$$

for small values of $k\Delta x$, Taylor expansions of these symbols correspond to the expansions in (3.12) when the Fourier symbol $\mathrm{i}k$ of ∂_x is substituted in the latter. However,

the checkerboard mode has $k\Delta x = \pi$ and for modes near this frequency the symbol has a very different behavior. Indeed, if we write $k\Delta x = \pi + k'\Delta x$ where $k'\Delta x$ is small, then we get instead

$$\mu_x C_{j+\frac{1}{2}} = (-\sin\tfrac{1}{2}k'\Delta x)C_{j+\frac{1}{2}}, \quad \delta_x C_{j+\frac{1}{2}} = (2\mathrm{i}\cos\tfrac{1}{2}k'\Delta x)C_{j+\frac{1}{2}}. \tag{3.25}$$

These expressions can then be expanded in powers of $k'\Delta x$. Moreover, if we write $C_j = (-1)^j C_j^o$, we get from (3.25) and (3.24)

$$\mu_x C_{j+\frac{1}{2}} = -\sin\tfrac{1}{2}k'\Delta x(-1)^{j+\frac{1}{2}}C^o_{j+\frac{1}{2}} = (-1)^{j+1}\tfrac{1}{2}\delta_x C^o_{j+\frac{1}{2}}, \tag{3.26}$$

and similarly

$$\delta_x C_{j+\frac{1}{2}} = (-1)^{j+1}2\mu_x C^o_{j+\frac{1}{2}}. \tag{3.27}$$

These considerations prompt the following approach.

An arbitrary mesh function can be represented by an expansion in Fourier modes for which $|k\Delta x| \le \pi$. Suppose we split this range at $\pi/2$ for an expansion of both C and S before substitution in the box scheme. Suppose, moreover, that there is little interaction between the two sets of modes. Then the behavior of the smooth set of modes is described by the modified equation analysis of the previous subsection. Now we use the relations (3.26) and (3.27) and an assumption that C^o and S^o are smooth to develop a modified equation analysis of the oscillatory modes and, in particular, the checkerboard mode. Thus we write C_j^n as a sum of smooth and oscillatory components,

$$C_j^n \equiv (C^s)_j^n + (-1)^{j+n}(C^o)_j^n, \tag{3.28}$$

and treat S similarly. Then, with similar relations to (3.26) and (3.27) holding for the time difference operators, (3.2) and (3.3) become, for the oscillatory part, after cancellation of the factor $(-1)^{j+n}$,

$$\delta_x\mu_t(C^o + S^o) + p\delta_t\mu_x C^o \approx 0, \quad -2\mu_t S^o \approx -\tfrac{1}{2}\delta_t\left(\lambda' C^o - \mu' S^o\right). \tag{3.29}$$

We can write these in terms of the operators (3.13) to obtain

$$\mathcal{D}_x(C^o + S^o) + p_2\mathcal{D}_t C^o \approx 0, \quad 4S^o \approx \mathcal{D}_t\left(\lambda'' C^o - \mu'' S^o\right), \tag{3.30}$$

where $p_2 := V(\Delta t/\Delta x)^2$, $\lambda'' := \lambda\Delta t^2$, and $\mu'' := \mu\Delta t^2$.

As in the treatment of the smooth part in the previous subsection, we can eliminate S^o to obtain a second order equation for C^o,

$$\left[\mu'' p_2\mathcal{D}_t^2 + (4p_2 + (\lambda'' + \mu'')\mathcal{D}_x)\,\mathcal{D}_t + 4\mathcal{D}_x\right] C^o = 0. \tag{3.31}$$

Then we rewrite this so that it is precisely of the form of (3.17), namely,

$$\left[\mathcal{D}_t^2 + [(4/\mu'') + ((\lambda+\mu)/\mu p_2)\mathcal{D}_x]\mathcal{D}_t + (4/\mu'' p_2)\mathcal{D}_x\right] C^o = 0. \tag{3.32}$$

To simplify application to the present case of the result proved in section 3.2 for that equation, and also to allow for generalization of the analysis to the general weighted scheme in the next section, we restate the main derivation as a lemma; its proof consists of a straightforward manipulation, which we omit.

LEMMA 3.1. *Suppose that in terms of the difference operators $\mathcal{D}_t$ and $\mathcal{D}_x$ of (3.13), a quantity U satisfies the operator equation (with $\alpha > \beta$)*

$$\left[(\mathcal{D}_t+\alpha)(\mathcal{D}_t+\beta)+(a\mathcal{D}_t^2+b\mathcal{D}_t+c)\mathcal{D}_x\right]U=0. \tag{3.33}$$

Then we can deduce modified equations for the main advected wave mode of U from the asymptotic relation

$$D_t=\sum_{j=0}^{3}(-a\mathcal{D}_x)^j\left\{-\beta-\tfrac{1}{2}b\mathcal{D}_x+\tfrac{1}{2}(\alpha-\beta)[d\mathcal{D}_x+\tfrac{1}{2}(e-d^2)(1-d\mathcal{D}_x)\mathcal{D}_x^2]\right\}+\ldots, \tag{3.34}$$

where

$$d=[(\alpha+\beta)b-2(\alpha\beta a+c)]/(\alpha-\beta)^2,\quad e=(b^2-4ac)/(\alpha-\beta)^2;$$

the corresponding relation for the other mode is obtained by interchanging α and β.

Hence, by comparing the coefficients in (3.32) with those in the lemma, we deduce that the main advected wave for the oscillatory mode C^o is described by the following expansion:

$$\begin{aligned} C_t^o &= -\frac{V}{p^2}C_x^o+\left(\frac{\lambda\Delta t^2V^2}{4p^4}\right)C_{xx}^o \\ &\quad -\frac{V}{p^2}\left(\frac{\lambda(\lambda-\mu)\Delta t^4V^2}{16p^4}+\frac{\Delta x^2}{12p^2}\left[1-p^2\right]\right)C_{xxx}^o+\ldots. \end{aligned} \tag{3.35}$$

The first coefficient here gives the speed at which the envelope of a set of checkerboard oscillations move. It is easy to check that this speed V/p^2 is in fact just the group velocity of the box scheme applied to the linear advection equation when one sets $k\Delta x=\pi$. Note that it is independent of λ and μ. Also, it is quite small for a reasonably large choice of CFL number. The second, damping coefficient is proportional to λ, and it depends on the mesh through a factor $\Delta x^4/\Delta t^2$. Note that for $\lambda=\mu$ the dispersion coefficient is positive for $p>1$ and negative for $p<1$. We shall see the effect of these terms in the numerical experiments presented in section 3.5.

3.4. The weighted box scheme. The traditional method of controlling the checkerboard mode is to introduce θ-weighting in the time integration (see [6]). In this section we describe this scheme and modify the analysis given above to cover this change. We confine the change to the transport equation and so obtain, instead of (3.2),

$$\mu_x\delta_tC_{j+\frac{1}{2}}^{n+\frac{1}{2}}+\mu_x\delta_tS_{j+\frac{1}{2}}^{n+\frac{1}{2}}+p\theta_t\delta_xC_{j+\frac{1}{2}}^{n+\frac{1}{2}}=0, \tag{3.36}$$

where we define the θ-averaging operator as

$$\theta_tC_j^{n+\frac{1}{2}}=\theta C_j^{n+1}+(1-\theta)C_j^n. \tag{3.37}$$

The scheme is second order accurate provided $\theta=\frac{1}{2}+O(\Delta t)$, and a Fourier analysis shows it is Lax–Richtymer stable for all Δx and Δt provided $\theta\geq\frac{1}{2}$; we shall therefore make this assumption in all that follows.

We first reconsider the energy analysis, making use of the identities and notation used earlier. As implied above, we continue to use the trapezoidal weighting for the source term; thus the equations become

$$\mu_x \delta_t C^{n+\frac{1}{2}} + p\theta_t \delta_x C^{n+\frac{1}{2}} = -\Delta t \mu_x \mu_t R^{n+\frac{1}{2}}, \tag{3.38}$$

$$\mu_x \delta_t S^{n+\frac{1}{2}} = \Delta t \mu_x \mu_t R^{n+\frac{1}{2}}, \tag{3.39}$$

where the discretization of the reaction equation has again been space-averaged.

Following the same procedure as before, we multiply (3.38) and (3.39) by $\theta_t \bar{C}^{n+1/2}$ and $\theta_t \bar{S}^{n+1/2}$, respectively, and then sum the resulting equations over j. The advection inner product collapses and, with the assumption that $t^n > t^*$, the boundary term at $j = 0$ vanishes and we obtain

$$\begin{aligned} &\theta \|\bar{C}^{n+1}\|_2^2 + (1-2\theta)\langle \bar{C}^{n+1}, \bar{C}^n\rangle_2 - (1-\theta)\|\bar{C}^n\|_2^2 \\ &\qquad + \tfrac{1}{2} V \Delta t \big[\theta_t C_J^{n+\frac{1}{2}}\big]^2 = -\Delta t \langle \theta_t \bar{C}^{n+\frac{1}{2}}, \mu_t \bar{R}^{n+\frac{1}{2}}\rangle_2, \end{aligned} \tag{3.40}$$

$$\begin{aligned} &\theta \|\bar{S}^{n+1}\|_2^2 + (1-2\theta)\langle \bar{S}^{n+1}, \bar{S}^n\rangle_2 - (1-\theta)\|\bar{S}^n\|_2^2 \\ &\qquad = \Delta t \langle \theta_t \bar{S}^{n+\frac{1}{2}}, \mu_t \bar{R}^{n+\frac{1}{2}}\rangle_2. \end{aligned} \tag{3.41}$$

Now from the Cauchy–Schwarz inequality we have, for $\theta \geq \frac{1}{2}$,

$$(2\theta - 1)\langle \bar{A}^{n+1}, \bar{A}^n\rangle_2 \leq \big(\theta - \tfrac{1}{2}\big)\big[\|\bar{A}^{n+1}\|_2^2 + \|\bar{A}^n\|_2^2\big].$$

Hence the terms on the left of (3.40) can be greatly simplified to give

$$\tfrac{1}{2}\|\bar{C}^{n+1}\|_2^2 - \tfrac{1}{2}\|\bar{C}^n\|_2^2 \leq -\Delta t \langle \theta_t \bar{C}^{n+\frac{1}{2}}, \mu_t \bar{R}^{n+\frac{1}{2}}\rangle_2. \tag{3.42}$$

Similarly, from (3.41), we get

$$\tfrac{1}{2}\|\bar{S}^{n+1}\|_2^2 - \tfrac{1}{2}\|\bar{S}^n\|_2^2 \leq \Delta t \langle \theta_t \bar{S}^{n+\frac{1}{2}}, \mu_t \bar{R}^{n+\frac{1}{2}}\rangle_2. \tag{3.43}$$

Combining these results as in (3.11) we obtain

$$\begin{aligned} &\lambda \|\bar{C}^{n+1}\|_2^2 + \mu \|\bar{S}^{n+1}\|_2^2 - \lambda \|\bar{C}^n\|_2^2 - \mu \|\bar{S}^n\|_2^2 \\ &\qquad \leq -\Delta t \big[\theta \|\bar{R}^{n+1}\|_2^2 + (1-\theta)\|\bar{R}^n\|_2^2 + \langle \bar{R}^{n+1}, \bar{R}^n\rangle_2\big]. \end{aligned} \tag{3.44}$$

Now we can rewrite the terms on the right by setting $\theta = \frac{1}{2}(1+\xi)$, and apply the Cauchy–Schwarz inequality again, to get

$$\begin{aligned} &\tfrac{1}{2}\big[(1+\xi)\|\bar{R}^{n+1}\|_2^2 + (1-\xi)\|\bar{R}^n\|_2^2 + 2\langle \bar{R}^{n+1}, \bar{R}^n\rangle_2\big] \\ &\qquad \geq \tfrac{1}{2}\xi\big[\|\bar{R}^{n+1}\|_2^2 - \|\bar{R}^n\|_2^2\big]. \end{aligned} \tag{3.45}$$

Substituting this result in (3.44) then yields the required stability of the cell averages, since we have $\xi \geq 0$, in terms of the energy $\lambda\|\bar{C}^n\|_2^2 + \mu\|\bar{S}^n\|_2^2 + \frac{1}{2}\xi\Delta t\|\bar{R}^n\|_2^2$. We should also note at this point how the weighting affects the propagation of the oscillatory mode into the interior from the boundary data—and hence the mapping from the cell averages to the nodal values. In applying a Godunov–Ryabenkii stability analysis, we can substitute a mode of the form $\alpha^n(-1)^j$ into (3.36); only the advection term contributes and yields the result $\alpha = 1 - 1/\theta$. So this mode is damped if $\theta > \frac{1}{2}$.

However, it is the modified equation analysis which gives the most information on the behavior of the weighted box scheme. The averaging operator θ_t can be written as

$$\theta_t = \mu_t + \big(\theta - \tfrac{1}{2}\big)\,\delta_t, \tag{3.46}$$

so we can introduce the operator

$$\mathcal{M}_t := \mu_t^{-1}\theta_t = 1 + \left(\theta - \tfrac{1}{2}\right)\Delta t\mathcal{D}_t. \tag{3.47}$$

The modified equation analysis can now proceed in exactly the same way as in sections 3.2 and 3.3, with the only changes being the replacement of $\mathcal{D}_x$ by $\mathcal{M}_t\mathcal{D}_x$ and the operator equations for both the smooth waves and the oscillatory waves being in the form covered by Lemma 3.1. For the main advected wave which previously led to the expansion in (3.22), it is a straightforward calculation to find that we get instead

$$\begin{aligned} C_t = &-V'C_x + V'^2\left\{\frac{\lambda}{\mu(\lambda+\mu)} + \left(\theta - \frac{1}{2}\right)\Delta t\right\}C_{xx} \\ &- V'\left\{\frac{\lambda(\lambda-\mu)V^2}{(\lambda+\mu)^4} + \frac{1}{12}\left[V'^2\Delta t^2 - \Delta x^2\right] + 3\frac{\lambda V'^2}{\mu(\lambda+\mu)}\left(\theta - \frac{1}{2}\right)\Delta t\right. \\ &\left. + V'^2\left(\theta - \frac{1}{2}\right)^2\Delta t^2\right\}C_{xxx} + \ldots. \end{aligned} \tag{3.48}$$

On comparing this with the unweighted case, we see that diffusion is enhanced for $\theta > \frac{1}{2}$ as expected, but the advection speed is unaffected; also there are extra terms in the dispersion, which will affect the choice of mesh ratio to minimize the numerical dispersion—note, however, that the term $\left(\theta - \frac{1}{2}\right)^2\Delta t^2$ is negligible if $\theta = \frac{1}{2} + O(\Delta t)$.

The most significant change is in the expansion for the main oscillatory wave, where the leading terms of (3.35) are replaced by

$$\begin{aligned} C_t^o = &-4\frac{\theta - \frac{1}{2}}{\Delta t}(1+\gamma)C^o \\ &- \frac{V}{p^2}\left(\frac{1 - (\lambda+\mu)(\theta - \frac{1}{2})\Delta t}{1 - \mu(\theta - \frac{1}{2})\Delta t}\right)(1+3\gamma)C_x^o + \ldots, \end{aligned} \tag{3.49}$$

where $\gamma = (4/3)(\theta - \frac{1}{2})^2$ and the factors involving this quantity arise from the transformation of an expansion for $\mathcal{D}_t$ into one for ∂_t. Thus even for $\theta = \frac{1}{2} + O(\Delta t)$, exponential damping of the oscillatory modes occurs, while the advective speed is little changed.

3.5. Numerical experiments. In all of the numerical experiments with the linear model problem that we present here, we have taken $V = 1.0$ and, in the early cases, we use the simple scheme with $\theta = \frac{1}{2}$. The plots are generally of the concentration C at a fixed point $x = 1$, as a function of t. Figure 2, in which we take $\Delta x = 0.025$, shows three results for the case of $(\lambda, \mu) = (1, 1)$: the bottom graph shows the smooth result obtained with Gaussian boundary data, while the top two show the corresponding results obtained with a square pulse boundary data (of unit height on the interval $(0, 1)$, which is our standard test data). The exact solution is shown as an unbroken line. These two plots show how devastating the checkerboard mode can be. The top plot is for the CFL number, $p = V\Delta t/\Delta x$, equal to 0.8 and shows the oscillations moving faster than the main pulse; the middle plot is with $p = 1.25$ and shows the oscillations moving more slowly, as predicted by the V^2/p velocity component in (3.35). In Figure 3 corresponding results are shown for the case $\lambda = 90$, $\mu = 10$, the top plot for square pulse data and the bottom plot for Gaussian data, both with $p = 0.8$. The middle plot has $p = 1.25$. The results are quite accurate and show the mollifying effect of increasing λ and μ. In Figure 4 this case is repeated with square pulse initial data, with $\Delta x = 0.04$, and varying values of the CFL number.

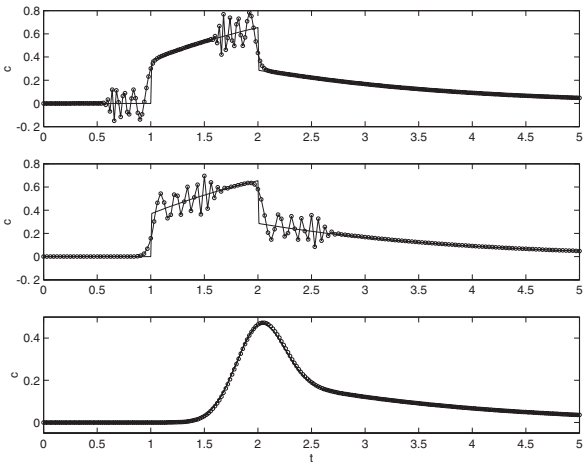

FIG. 2. *Solution c for $V = 1$ and $x = 1$: the thin unbroken line indicates the exact solution, the dots joined by an unbroken line indicate the box scheme. In all cases $(\lambda, \mu) = (1, 1)$, $\Delta x = 0.025$. The top plot has a square pulse initial data with CFL number $p = 0.8$, the middle plot has the same initial data with $p = 1.25$, and the bottom plot has Gaussian initial data with $p = 0.8$.*

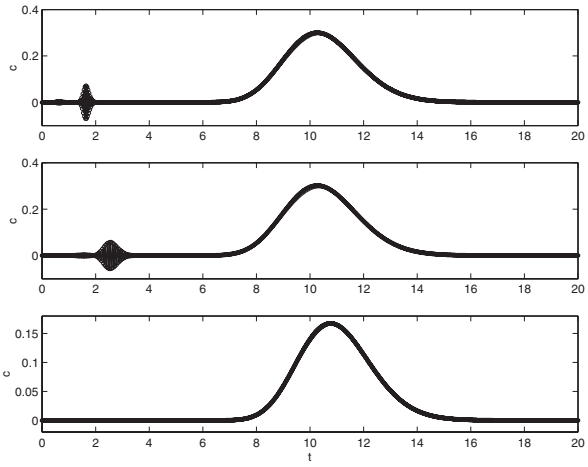

FIG. 3. *As for Figure 2 but with $(\lambda, \mu) = (90, 10)$.*

Choosing p to match the reduced speed (that is, $p = (\lambda + \mu)/\mu = 10$) eliminates the effects of the discretization in the dispersion term in (3.22). However, the best results are obtained by taking $p = 6.4103$, which is calculated to set the coefficient of the dispersion term in (3.22) to zero (which is possible in this model problem but unlikely to hold in applications). These results confirm the modified equation analysis of section 3.2 showing that the accuracy is improved by tuning p to match the reduced speed rather than the speed in the transport equation. They also show how the effect of the oscillations is reduced by such a choice. These experiments are repeated with $\theta = .51$ in Figure 4, showing how such a choice completely eliminates the oscillations, as predicted by the analysis in section 3.4. Finally, we show that the spurious oscillations move with speed V/p^2 independent of λ and μ as predicted by

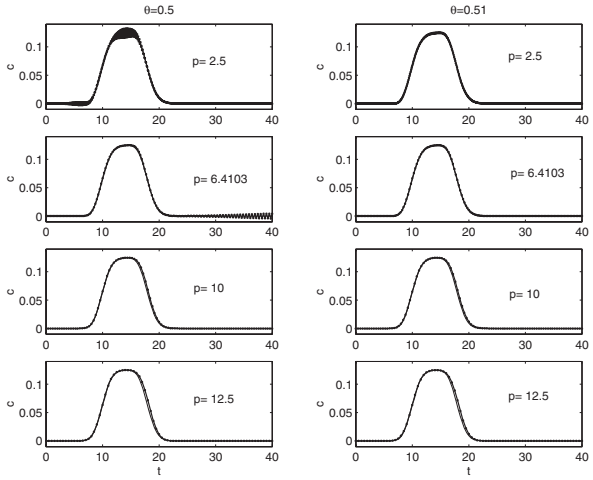

FIG. 4. *Solution c for $V = 1$ at $x = 1$: the thin unbroken line indicates the exact solution and the dotted line the box scheme solution. In all cases $\Delta x = 0.04$ $(\lambda, \mu) = (90, 10)$. The values of p are 2.5, 6.4103, 10, and 12.5. The left-hand column has $\theta = 0.5$ and the right-hand column has $\theta = 0.51$.*

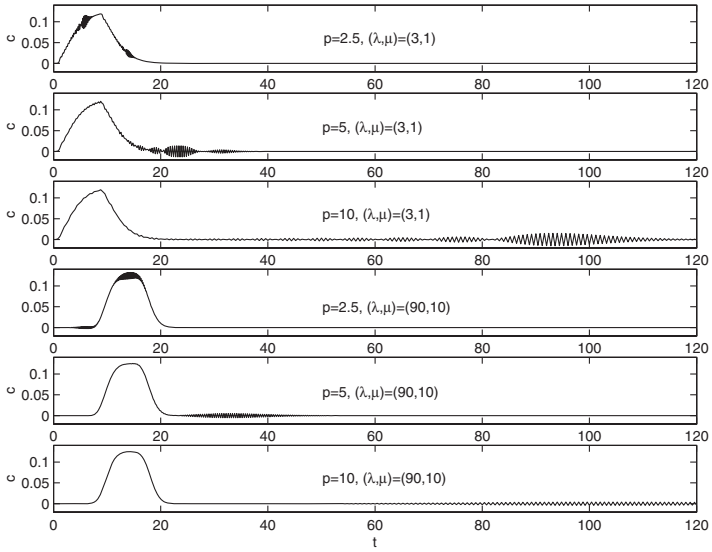

FIG. 5. *Solution c for $V = 1$ at $x = 1$ for $(\lambda, \mu) = (3, 1)$ and $(\lambda, \mu) = (90, 10)$, for three values of $p = \Delta t/\Delta x$: the thin line indicates the box scheme solution. It is easily seen that the spurious oscillations move with speed V/p^2.*

(3.35). This is easily observed in Figure 5 for $(\lambda, \mu) = (90, 10)$ and $(\lambda, \mu) = (3, 1)$, where the spurious oscillations exhibit the same qualitative behavior for the two sets of parameter values.

4. A nonlinear model in one space dimension. We now turn our attention to the following problem with a nonlinear reaction term:

$$c_t + Vc_x = r(c,s), \tag{4.1}$$

$$s_t = -r(c,s). \tag{4.2}$$

If $r(c,s) = -\lambda c + \mu s$, then we recover the linear problem discussed in section 2. Common nonlinear expressions are the Langmuir reaction (see [16], p. 166)

$$r(c,s) = -\lambda c(\mathcal{B} - s) + \mu s, \tag{4.3}$$

where $\mathcal{B}$ represents the maximum capacity for absorption, and thus the rate of absorption of c into the rock reduces as s increases, and the Freundlich reaction

$$r(c,s) = -\lambda c^{1+\beta} + \mu s, \tag{4.4}$$

where the first term on the right-hand side represents a $(1+\beta)$th-order reaction for the adsorption of c into the rock. Often $-1 < \beta < 0$, but we do not impose this restriction here.

Adding (4.1) and (4.2) gives the conservation equation

$$(c+s)_t + Vc_x = 0, \tag{4.5}$$

and, following [13], let us assume that the reaction term $r(c,s)$ in (4.1) and (4.2) can be written as $r(c,s) = (s - S(c))/\epsilon$ for some smooth $S(c)$ and some small parameter ϵ. Thus (4.2) is written as

$$s_t = \frac{1}{\epsilon}(S(c) - s), \tag{4.6}$$

and the equilibrium manifold (see the first paragraph in section 2.1) is $s = S(c)$.

The equilibrium model (see (2.4)) arises by replacing s in (4.5) by $S(c)$ to give the nonlinear conservation law

$$c_x + \frac{1}{V}(c + S(c))_t = 0, \tag{4.7}$$

where the roles of x and t are reversed in the application of standard theory of hyperbolic conservation laws. Consider this equation subject to Riemann boundary data

$$c(0,t) = \begin{cases} c_L, & t \le \tau, \\ c_R, & t > \tau, \end{cases} \tag{4.8}$$

for some $\tau \ge 0$. If $S''(c) > 0$, as is the case for the Freundlich nonlinearity for $\beta > 0$, and $c_L > c_R$, (4.7) develops a shock with speed

$$U_s := \frac{V(c_L - c_R)}{S(c_L) - S(c_R) + c_L - c_R}, \tag{4.9}$$

but if $c_L < c_R$, (4.7) describes a rarefaction wave. If $S''(c) < 0$, as is the case for the Langmuir nonlinearity, the two cases are interchanged. The improved equilibrium model is (cf. (2.9))

$$c_x + \frac{1}{V}(c + S(c))_t = \frac{\epsilon}{V}(S'(c)c_t)_t, \tag{4.10}$$

where we have replaced s_t in (4.5) using $s_t = (S(c))_t - \epsilon s_{tt}$ and have approximated s_{tt} using $s_{tt} \approx (S'(c)c_t)_t$. Here we see that if $S'(c) > 0$, as is the case in both our examples, then the right-hand side of (4.10) provides a "viscous damping" term that prevents shocks from forming. The numerical results presented in section 4.1 show that discontinuities are indeed gradually smoothed and damped.

As in the linear model, we use the predicted pulse speed to choose the time step; this can be done dynamically as the pulse develops. The shock speed U_s, given by (4.9), is our main guide in the case of a data pulse of height $|c_L - c_R|$; but note that as c_L tends to c_R this becomes the *reduced speed* $V' = V/(1 + S'(c))$, the characteristic speed of the equilibrium model (4.7) which in general differs from the shock speed. We note also that a traveling wave solution of (4.1), (4.2) propagates at the shock speed (cf. [22, pp. 101–102]).

4.1. Numerical results. Here we present numerical results for two examples with nonlinear reactions. We shall see that the weighted box scheme with appropriately chosen mesh ratio produces good results for the problems under consideration. A likely strategy in a practical situation would be to fix Δx and tune Δt based either on the shock speed, so that $U_s \Delta t/\Delta x \approx 1$, or on the reduced speed, so that $\frac{V}{1+S'(c)}\Delta t/\Delta x \approx 1$, for some appropriate choice of c. Here we carry out experiments with fixed Δt and Δx to fully understand the performance of the box scheme, and to compare the nonlinear case with the linear case discussed in section 3.5.

First, let us consider system (4.1), (4.2) with the Freundlich nonlinearity (4.4) where $\beta = 1$, i.e., $r(c, s) = -\lambda c^2 + \mu s$, and hence $S(c) = \lambda c^2/\mu$, the reduced speed is $\mu V/(\mu + 2\lambda c)$, but the shock speed is $\mu V/(\mu + \lambda|c_L - c_R|)$. We take the boundary condition

$$c(0, t) = \begin{cases} 1/3, & 0 \le t \le 1, \\ 0, & \text{otherwise}, \end{cases} \tag{4.11}$$

and since $S'' > 0$ we expect the "switch-off" discontinuity at $t = 1$ to lead to a steep front. Let $V = 1$, $(\lambda, \mu) = (90, 10)$, so the shock speed is $U_s = 0.25$, which suggests the choice $\Delta t = 4\Delta x$. In Figure 6 we present results for several values of $\Delta t/\Delta x$ at $x = 1$, and in Figure 7 we repeat the experiment using the weighted box scheme with $\theta = 0.52$. First, we see in Figure 6 that the spurious oscillations are strongly evident for all values of p shown here, and they move with speed V/p^2 as predicted for the linear theory by (3.35). However, the θ-weighting eliminates the spurious oscillations in all cases. As x increases the pulse diffuses, as indicated by (4.10), so for optimal results we should tune Δt using the reduced speed so that Δt would gradually reduce from the value $4\Delta x$ derived using the shock speed.

Next, let us consider system (4.1), (4.2) with the Langmuir nonlinearity (4.3) with $\mathcal{B} = 1$, i.e., $r(c, s) = -\lambda c(1 - s) + \mu s$, subject to the same boundary condition as the previous example. Now we have $S(c) = \lambda c/(\mu + \lambda c)$, the reduced speed is $V[1 + \lambda\mu(\mu + \lambda c)^{-2}]$, and the shock speed is $V[1 + \lambda\mu\{(\mu + \lambda c_R)(\mu + \lambda c_L)\}^{-1}] = 4/13$ for the data given in (4.11). Since $S''(c) < 0$ we expect the "switch-on" discontinuity at $t = 0$ to give the steep front and we choose Δt based on the shock speed so that $\Delta t = 3.25\Delta x$. Numerical results for the box scheme using $\Delta x = 1/30$ with both $\theta = 0.5$ (i.e., no weighting) and $\theta = 0.51$ are given in Figure 8. Again we see the significant effect of the spurious oscillations in the unweighted scheme, though for $p = \Delta t/\Delta x = 10/3$ rather good results are obtained. As expected, the introduction of a small amount of weighting produces much improved results for all values of p. As in the first example, we see that the spurious oscillations introduced by the discontinuity

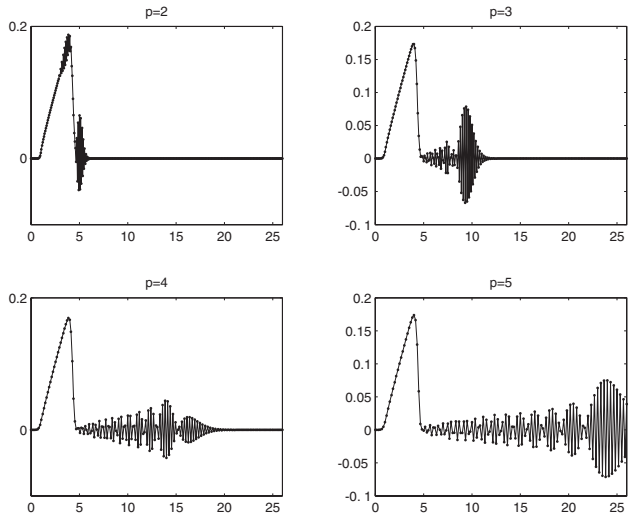

FIG. 6. *The box scheme with no θ-weighting is applied to* (4.1), (4.2) *with Freundlich nonlinearity* (4.4), *and with a square pulse boundary condition* (4.11). *Here* $\beta = 2$, $\lambda = 90$, $\mu = 10$, $V = 1$, $\Delta x = 1/30$, *and* $p := \Delta t/\Delta x$.

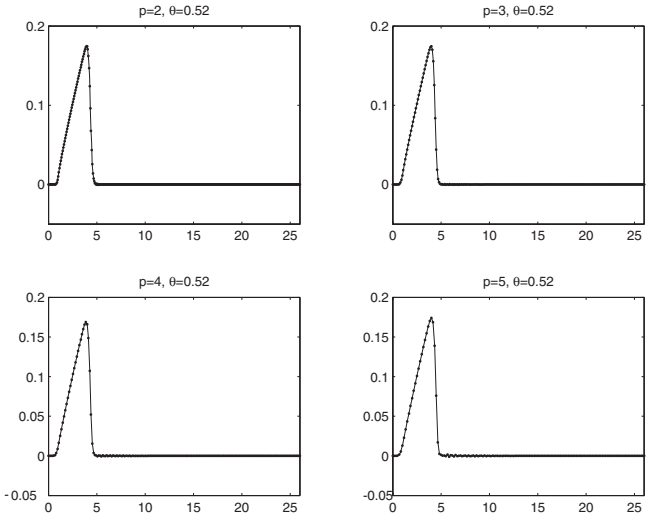

FIG. 7. *As for Figure* 6 *but with* $\theta = 0.52$.

at $x = 0$, $t = 1$ for $\theta = 0.5$ propagate with speed V/p^2, as indicated by (3.35) for the linear problem.

5. The box scheme for two-dimensional problems. Most practical problems are posed in two or three space dimensions with variable velocity profiles. We shall see that the theory of the box scheme in one dimension predicts well its behavior in two dimensions and that the weighted scheme successfully eliminates any spurious oscillations.

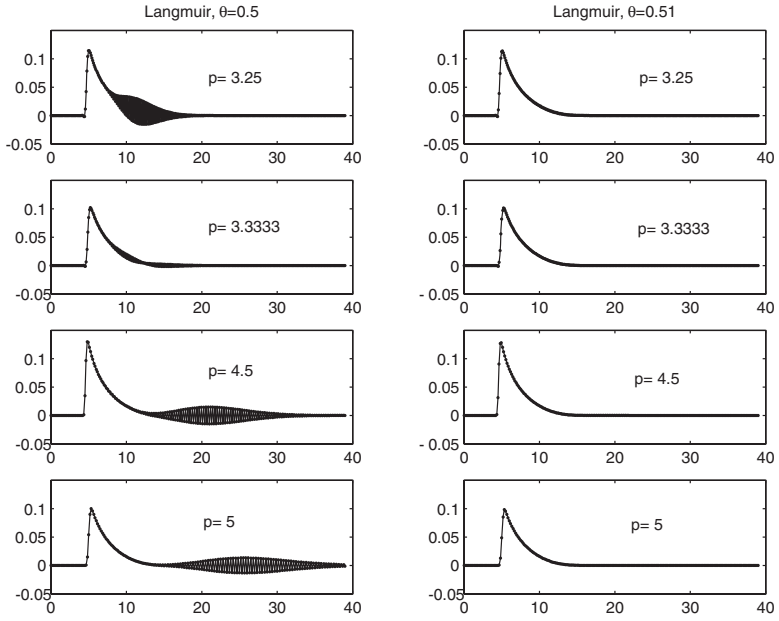

FIG. 8. *The box scheme with* $\theta = 0.5$ *and* 0.51 *is applied to* (4.1), (4.2) *with Langmuir nonlinearity* (4.3), *and with a square pulse boundary condition* (4.11). *Here* $\mathcal{B} = 1$, $(\lambda, \mu) = (90, 10)$, $V = 1$, $x = 1$, *and* $p := \Delta t/\Delta x$.

In [19] and [20] the authors consider the mobility of potentially toxic dissolved metals discharged from a mine tailings source into an aquifer in which the flow is incompressible (see Figure 2 in [20]). We simplify this situation by considering a single chemical pollutant with a reacting term as in the linear model problem. The system to be solved, in conservation form, is

$$c_t + s_t + (U(x,z)c)_x + (V(x,z)c)_z = 0, \tag{5.1}$$

$$s_t = \lambda c - \mu s, \tag{5.2}$$

in the domain $\mathcal{D} := 0.5 \le x \le 2.0$, $0.8 \le z \le 1.0$, for $t \ge 0$, with the incompressible flow field $(U(x,z), V(x,z))$ being given by

$$U(x,z) = \frac{2}{\pi \sinh\left(\frac{\pi Z}{2X}\right)} \cosh\left(\frac{\pi z}{2X}\right) \sin\left(\frac{\pi x}{2X}\right), \tag{5.3}$$

$$V(x,z) = \frac{-2}{\pi \sinh\left(\frac{\pi Z}{2X}\right)} \sinh\left(\frac{\pi z}{2X}\right) \cos\left(\frac{\pi x}{2X}\right). \tag{5.4}$$

(These expressions arise by solving $\nabla^2 \phi(x,z) = 0$ in $0 \le x \le X$, $0 \le z \le Z$, subject to the boundary conditions $\phi_z(x,0) = 0$, $\phi_z(x,Z) = -0.5$, $\phi_x(0,z) = 0$, $\phi(X,z) = 1$, using separation of variables and then taking the first terms in the expressions for $\phi_x(x,z)$ and $\phi_z(x,z)$ as the respective velocity components.)

The initial condition is

$$c(x,z,0) = s(x,z,0) = 0 \quad \text{for } (x,z) \in \mathcal{D}, \tag{5.5}$$

and the boundary condition is

$$c(x,Z,t)=\begin{cases} g(x), & 0\le t\le t^*,\\ 0, & t>t^*. \end{cases} \tag{5.6}$$

In our examples, $t^*=2$ and

$$g(x)=\begin{cases} 0, & x\le\tau,\\ \sin^2\left(\frac{\alpha\pi}{2}(x-\tau)\right), & \tau<x<\frac{1}{\alpha}+\tau,\\ 1, & \frac{1}{\alpha}+\tau\le x\le\beta-\frac{1}{\alpha}+\tau,\\ \sin^2\left(\frac{\alpha\pi}{2}(\beta+\tau-x)\right), & \beta-\frac{1}{\alpha}+\tau\le x\le\beta+\tau,\\ 0, & x>\beta+\tau. \end{cases} \tag{5.7}$$

Here α can be varied, with a small α giving a smooth pulse and a large α making the pulse more "squarelike." In our numerical experiments $\alpha=6$, $\beta=0.8$, and $\tau=0.6$.

In terms of difference operators, the box scheme applied to (5.1), (5.2) produces

$$\begin{aligned} &\frac{\mu_x\mu_z\delta_t}{\Delta t}C^{n+\frac12}_{i+\frac12,j+\frac12}+\frac{\mu_x\mu_z\delta_t}{\Delta t}S^{n+\frac12}_{i+\frac12,j+\frac12}\\ &\quad+\frac{\mu_z\mu_t\delta_x}{\Delta x}(U\ C)^{n+\frac12}_{i+\frac12,j+\frac12}+\frac{\mu_x\mu_t\delta_z}{\Delta z}(V\ C)^{n+\frac12}_{i+\frac12,j+\frac12}=0,\\ &\frac{\delta_t}{\Delta t}S^{n+\frac12}_{i+1,j+1}=\lambda\,\mu_t C^{n+\frac12}_{i+1,j+1}-\mu\,\mu_t S^{n+\frac12}_{i+1,j+1}. \end{aligned}$$

The box scheme is an implicit scheme, but in this linear problem, with the fluxes flowing in the same general direction in all the cells, an efficient explicit scheme is obtained by sweeping from the top left-hand corner. We use an averaged mesh in the x and z direction calculated as follows. Set

$$\hat U(x)=\frac{1}{0.2}\int_{0.8}^{1}U(x,z)\,\mathrm{d}z,\qquad \hat V(z)=\frac{1}{1.5}\int_{0.5}^{2}V(x,z)\,\mathrm{d}x$$

and choose variable step sizes Δx_{i+1} and Δz_{j+1} such that

$$\frac{\hat U(x_{i+1})+\hat U(x_i)}{2\Delta x_{i+1}}=\frac{\hat V(z_{j+1})+\hat V(z_j)}{2\Delta z_{j+1}}.$$

In our numerical experiments we take $\lambda=90$, $\mu=10$. We expect the solution to be close to that for the equilibrium model, and hence the solution speed is reduced by a factor of $\frac{\mu}{\lambda+\mu}$. Thus we expect the optimal choice for Δt to be given by

$$\Delta t=\frac{\lambda+\mu}{\mu}\frac{2\Delta x_{i+1}}{\hat U(x_{i+1})+\hat U(x_i)}=\frac{\lambda+\mu}{\mu}\frac{2\Delta z_{i+1}}{\hat V(x_{i+1})+\hat V(x_i)}.$$

If we introduce the quantity p given by

$$p=\frac{\hat U(x_{i+1})+\hat U(x_i)}{2}\frac{\Delta t}{\Delta x_{i+1}} \tag{5.8}$$

(cf. (3.2) for the one-dimensional problem), then to tune the mesh for optimal results we take p to be $\frac{\lambda+\mu}{\mu}$. Numerical results obtained using the box scheme with $p=\frac{\lambda+\mu}{\mu}=10$ are shown in Figure 9, and for the θ-weighted scheme in Figure 10. The unweighted scheme clearly shows the spurious mode moving at a different speed from the pulse, and the weighted scheme has again successfully removed the oscillations. To illustrate the behavior of the spurious oscillations, we show in Figures 11 and 12 the

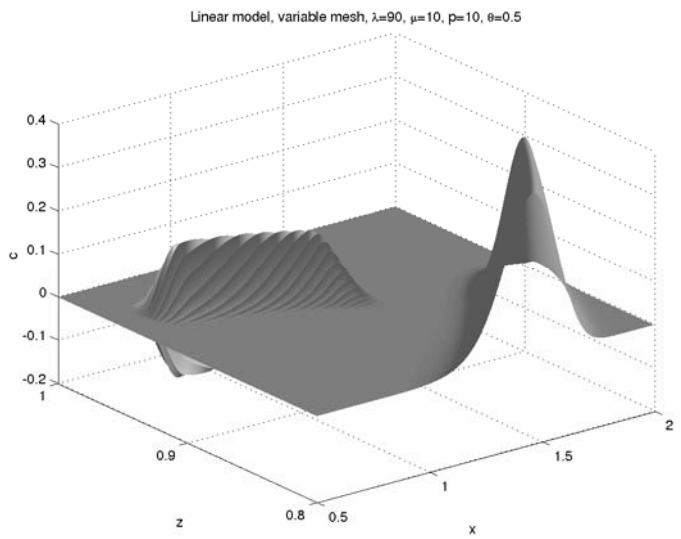

FIG. 9. *Box scheme applied to* (5.1), (5.2) *with* $\theta = 0.5$. *Here the solution pulse moves independently of the spurious oscillations.*

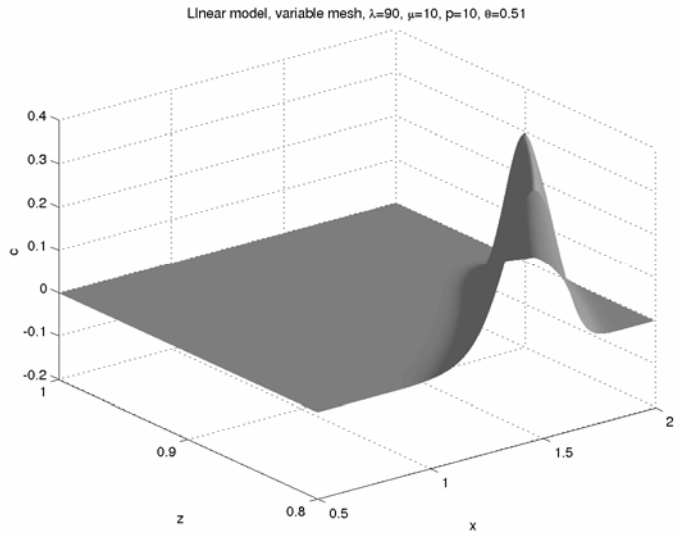

FIG. 10. *Box scheme applied to* (5.1), (5.2) *with* $\theta = 0.51$ *has removed the spurious oscillations.*

computed values of $c(1, 0.95, t)$ for $0 \le t \le 25$ for four values of p, namely, $p = 1.25$, 2.5, 5, and 10. Figure 11 shows the results from the unweighted box scheme. As p doubles we see the spurious oscillations move at a speed determined by the mesh and not the parameters in the differential equation. In fact, they move at a speed roughly like $1/p^2$, as predicted by (3.35) for one-dimensional problems (cf. also Figure 5). Figure 12 shows improved results obtained by the weighted box scheme with $\theta = 0.51$, and in particular the spurious oscillations have disappeared for the optimal value $p = 10$, as expected.

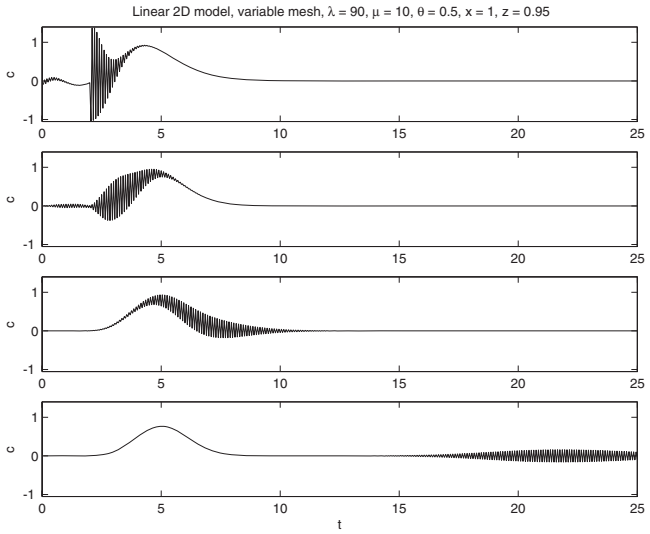

FIG. 11. *Plot of the computed value of* $c(1, 0.95, t)$ *for* $0 \leq t \leq 25$ *for* $\Delta t = 0.05$ *and* $p = 1.25, 2.5, 5, 10$*, respectively. The pulse of spurious oscillations moves like* p^{-2} *just as predicted in the one-dimensional theory.*

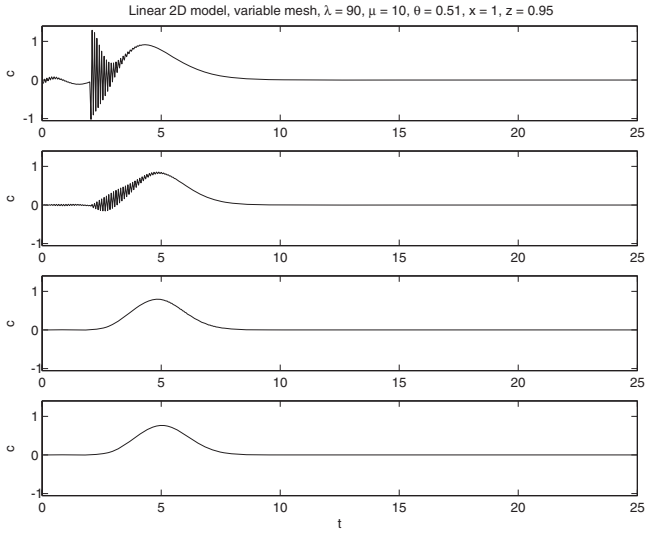

FIG. 12. *As for Figure* 11 *but with* $\theta = 0.51$*. The spurious oscillations have been greatly reduced and for* $p = 10$ *have disappeared.*

6. Conclusions. In this paper we discuss the application of the box scheme to hyperbolic conservation laws with reactive source terms. Our discussion centers initially on a model problem with linear reaction terms. First, we provide a straightforward asymptotic analysis that explains the phenomena of reduced speed, enhanced diffusion, and dispersion. An energy analysis for the box scheme shows the stability of cell averages but also indicates how the notorious checkerboard mode may grow linearly. A detailed modified equation analysis shows how to tune the mesh parameters

in the box scheme to capture the main advected wave with minimal dispersion. A novel modified equation analysis, based on separating smooth and oscillatory solutions, shows how the checkerboard mode moves at a nonphysical speed determined by the mesh parameters. The θ-weighted box scheme is also discussed and, in particular, the modified equation analysis clearly shows how the θ-weighting exponentially damps the checkerboard mode. We show how the deductions from the analysis for the linear problem apply to problems with nonlinear reactions; this has applications in other physically important problems, such as the St. Venant equations for open channel flow. Numerical results for linear and nonlinear reactions in one dimension and linear reactions in two dimensions illustrate the applicability of the theory, including the nonphysical speed of the checkerboard mode and the effectiveness of θ-weighting, across a variety of problems.

REFERENCES

[1] T. Arbogast, S. Bryant, C. Dawson, F. Saaf, C. Wang, and M. Wheeler, *Computational methods for multiphase flow and reactive transport problems arising in subsurface contaminant remediation*, J. Comput. Appl. Math., 74 (1996), pp. 19–32.

[2] D. A. Barry, K. Bajracharya, and C. T. Miller, *Alternative split-operator approach for solving chemical reaction/groundwater transport models*, Advances in Water Resources, 19 (1996), pp. 261–275.

[3] D. A. Barry, C. T. Miller, P. J. Culligan, and K. Bajracharya, *Analysis of split operator methods for nonlinear and multispecies groundwater chemical transport models*, Math. Comput. Simulation, 43 (1997), pp. 331–341.

[4] G.-Q. Chen, D. Levermore, and T.-P. Liu, *Hyperbolic conservation laws with stiff relaxation terms and entropy*, Commun. Pure Appl. Math., 47 (1994), pp. 787–830.

[5] P. I. Crumpton, J. A. MacKenzie, and K. W. Morton, *Cell vertex algorithms for the compressible Navier-Stokes equations*, J. Comput. Phys., 109 (1993), pp. 1–15.

[6] J. A. Cunge, F. M. Holly, Jr., and A. Verwey, *Practical Aspects of Computational River Hydraulics*, Pitman, Boston, 1980.

[7] J. Herzer and W. Kinzelbach, *Coupling of transport and chemical processes in numerical transport models*, Geoderma, 44 (1989), pp. 115–127.

[8] A. A. Jennings, D. J. Kirkner, and T. L. Theis, *Multicomponent equilibrium chemistry in groundwater quality models*, Water Resources Research, 18 (1982), pp. 1089–1096.

[9] H. B. Keller, *Numerical Methods for Two-Point Boundary-Value Problems*, Blaisdell, London, 1968.

[10] H. B. Keller, *A new finite difference scheme for parabolic problems*, in Numerical Solution of Partial Differential Equations II, SYNSPADE 1970, B. Hubbard, ed., Academic Press, New York, 1971, pp. 327–350.

[11] M. J. Lighthill and G. B. Whitham, *On kinematic waves:* I. *Flood movement in long rivers*, Proc. Roy. Soc. A, 229 (1955), pp. 281–316.

[12] M. J. Lighthill and G. B. Whitham, *On kinematic waves:* II. *A theory of traffic flow on long crowded roads*, Proc. Roy. Soc. A, 229 (1955), pp. 317–345.

[13] T.-P. Liu, *Hyperbolic conservation laws with relaxation*, Commun. Math. Phys., 108 (1987), pp. 153–175.

[14] S. L. Mitchell, *Coupling Transport and Chemistry: Numerics, Analysis and Applications*, Ph.D. thesis, University of Bath, UK, 2003.

[15] K. W. Morton and D. F. Mayers, *Numerical Solution of Partial Differential Equations*, 2nd ed., Cambridge University Press, Cambridge, UK, 2005.

[16] H.-K. Rhee, R. Aris, and N. R. Amundson, *First-Order Partial Differential Equations, Volume* I*, Theory and Application of Single Equations*, Prentice Hall, Englewood Cliffs, NJ, 1986.

[17] R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial Value Problems*, Interscience, New York, 1967.

[18] V. Thomée, *A stable difference scheme for the mixed boundary value problem for a hyperbolic first order system in two dimensions*, J. Soc. Indust. Appl. Math., 10 (1962), pp. 229–245.

[19] A. L. Walter, E. O. Frind, D. W. Blowes, C. J. Ptacek, and J. W. Molson, *Modeling of multicomponent reactive transport in groundwater* 1. *Model development and evaluation*,

Water Resources Research, 30 (1994), pp. 3117–3148.
[20] A. L. Walter, E. O. Frind, D. W. Blowes, C. J. Ptacek, and J. W. Molson, *Modeling of multicomponent reactive transport in groundwater* 2. *Metal mobility in aquifers impacted by acidic mine tailings discharge*, Water Resources Research, 30 (1994), pp. 3149–3158.
[21] R. F. Warming and B. J. Hyett, *The modified equation approach to the stability and accuracy analysis of finite-difference methods*, J. Comput. Phys., 14 (1974), pp. 159–179.
[22] G. B. Whitham, *Linear and Nonlinear Waves*, John Wiley, New York, 1974.
[23] P. F. Zhao and M. Z. Qin, *Multisymplectic geometry and multisymplectic Preissmann scheme for the KdV equation*, J. Phys. A, 33 (2000), pp. 3613–3626.
[24] A. Zysset, F. Stauffer, and T. Dracos, *Modeling of chemically reactive groundwater transport*, Water Resources Research, 30 (1994), pp. 2217–2228.