Godunov-Mixed Methods for Advective
Flow Problems in One Space Dimension

Clint N. Dawson

March, 1991

TR91-05

GODUNOV-MIXED METHODS FOR ADVECTIVE FLOW PROBLEMS IN ONE SPACE DIMENSION

CLINT N. DAWSON*

Abstract. A time-splitting method for solving advection-dominated, parabolic, partial differential equations is presented. In this method, a higher-order Godunov procedure approximates advection and a mixed finite element procedure approximates diffusion. Several variations on the basic scheme are formulated for solving one-dimensional, quasilinear, parabolic problems with Dirichlet boundary conditions. A maximum principle for one variant of the scheme is demonstrated, and $L^\infty(L^2)$ and $L^2(L^2)$ error estimates for the approximate solution and the diffusive flux, respectively, are derived. These estimates indicate that one variant of the scheme is L^∞ -stable in certain situations, but possibly sub-optimal in error, while another variant is optimal and L^2 -stable.

Keywords: Higher-order Godunov method, mixed finite element method, error estimates

AMS(MOS) subject classification: 65N10, 65N30

1. Introduction. Many problems in science and engineering are described by advection-dominated parabolic partial differential equations. Besides advection, these equations model diffusion and dispersion, and, in some cases, chemical reactions. Problems of this type occur, for example, in flow in porous media, aerodynamics, and the modelling of semiconductors.

There are several numerical difficulties which must be addressed when solving these types of equations, including the approximation of sharp fronts, grid orientation effects, and, for reactive flow in particular, time-scale considerations. It is well-known that standard finite difference and finite element methods exhibit nonphysical oscillations for problems with sharp fronts. In response to this phenomenon, so-called upwinding schemes were introduced which suppress oscillations by incorporating artificial diffusion. In its basic form, upwinding suffers not only from excessive smearing of fronts but also from extreme dependence on grid orientation [11, 13]. In recent years however, numerous schemes have been introduced which combine numerical stability with minimal artificial diffusion and dependence on grid orientation. We refer to the recent work on the modified method of characteristics for the miscible displacement problem (see [12] and the references therein) as a case in point. Moreover, by incorporating time-splitting, the time-scale difficulties which occur when instantaneous chemical reactions are included can be handled [18].

In recent years, there has been much interest in the application of higher-order Godunov techniques to advective flow problems. In this paper, we investigate combining these techniques with mixed finite element methods for advection-dominated advection-diffusion equations. Hereafter, we will refer to this general approach as the Godunov-mixed method, or GMM. The combining of these two schemes is natural because both are based on a weak form of the differential equation and utilize discontinuous approximating spaces. Though we restrict our attention in this paper to one space dimension, the methodology has been extended to two dimensions, using an unsplit, higher-order Godunov scheme described in [1]. Theoretical results for this algorithm will be given in a later paper. Numerical results for the scheme presented here applied to the Buckley-Leverett equation can be found in [19]. These results indicate that the scheme accurately resolves sharp fronts. Two-dimensional results given in [1] also indicate the scheme suffers minimal sensitivity to grid orientation.

The higher-order Godunov schemes we consider are the second-order method developed

* Dept. of Mathematical Sciences, Rice University, Houston, TX 77251. This research was supported by NSF Grant DMS-8807257.

by van Leer [14], and the generalizations of this scheme developed by Harten, *et. al.*, [8]. The former scheme, called MUSCL, was based on a scheme given in [7] and was later refined and extended by Colella [3] and Colella and Woodward [5]. The latter method is the second-order version of the Essentially Non-Oscillatory (ENO) schemes. These schemes have been used for solving conservation laws, such as those arising in gas dynamical simulations. Bell and Shubin [2] were the first to apply higher-order Godunov techniques to flow in porous media. This methodology and work of Colella [4] and van Leer [15] were the impetus for the two-dimensional scheme derived in [1].

The mixed finite element method was first developed for second-order elliptic equations by Raviart and Thomas [10]. Here we concentrate on the mixed method with the lowest-order Raviart-Thomas approximating spaces.

The major results of this paper are derivation of a maximum principle and error estimates for Godunov-mixed methods applied to one-dimensional partial differential equations of the form

$$(1.1) \quad s_t + f(s)_x - ((a(x, t)s_x)_x = 0, \quad (x, t) \in I \times (0, T] \equiv Q_T,$$

with initial condition

$$(1.2) \quad s(x, 0) = s^0(x), \quad x \in I,$$

and Dirichlet boundary conditions,

$$(1.3) \quad \begin{aligned} s(0, t) &= g_0(t), \\ s(1, t) &= g_1(t), \quad 0 < t \leq T. \end{aligned}$$

Similar results for boundary conditions of Neumann and mixed type, and extension of the arguments to $a = a(x, t, s)$ and nonhomogeneous right hand side, can be found in [6]. In our analysis, we assume for simplicity that $f'(s) \geq 0$ and we assume $0 < a_* \leq a(x, t) \leq a^*$ for some positive constants a_* , a^* . We remark on the extension of the arguments to general $f \in C^2$ in Section 7. Throughout we assume s and the coefficients are sufficiently smooth so that optimal order estimates can be obtained. We also assume compatibility between boundary and initial data; i.e., $g_0(0) = s^0(0)$, etc.

The remainder of this paper is organized as follows. Notation is established in the next section. The Godunov-mixed method is then derived in Section 3. Theoretical results for the basic scheme, namely a maximum principle and error estimates, are stated in Section 4. The maximum principle is then proved in Section 5, and the error estimates in Section 6. Extensions of these results to a more general flux function, and to more general time-stepping procedures, as well as experimental rates of convergence, are presented in Section 7. In Section 7, we also introduce a second order in time scheme.

2. Notation and definitions. Let J be a positive integer, and

$$\delta_x : 0 = x_{\frac{1}{2}} < x_{\frac{3}{2}} < \dots < x_{J+\frac{1}{2}} = 1$$

be a partition of $[0, 1]$ into grid blocks $B_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$. Let x_j be the midpoint of B_j , and $\Delta x_j = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$, $\Delta x_{j+\frac{1}{2}} = x_{j+1} - x_j$, for $j = 1, \dots, J-1$. Let $g_j = g(x_j)$ and $g_{j+\frac{1}{2}} = g(x_{j+\frac{1}{2}})$. We assume the mesh defined by δ_x is quasi-uniform; i.e., there exist positive constants C_0 , C^0 , independent of the mesh, such that

$$(2.4) \quad C_0 \leq \frac{\Delta x_i}{\Delta x_j} \leq C^0, \quad i, j = 1, \dots, J.$$

Define

$$\mathcal{M}_{-1}^k(\delta_x) = \{v \mid v \text{ is a polynomial of degree } \leq k \text{ on } B_j\},$$

and set $\mathcal{M}_0^k(\delta_x) = \mathcal{M}_{-1}^k(\delta_x) \cap C^0(I)$. We note that a standard basis for $\mathcal{M}_0^1(\delta_x)$ is the set of “hat” functions $\{v_{j+\frac{1}{2}}(x)\}_{j=0}^J$, given by

$$(2.5) \quad v_{j+\frac{1}{2}}(x) = \begin{cases} (x - x_{j-\frac{1}{2}})/\Delta x_j, & x_{j-\frac{1}{2}} \leq x \leq x_{j+\frac{1}{2}}, \\ (x_{j+\frac{3}{2}} - x)/\Delta x_{j+1}, & x_{j+\frac{1}{2}} \leq x \leq x_{j+\frac{3}{2}}, \\ 0, & \text{otherwise,} \end{cases}$$

for $j = 1, \dots, J-1$;

$$(2.6) \quad v_{1/2}(x) = \begin{cases} x/\Delta x_1, & 0 \leq x \leq x_{3/2}, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$(2.7) \quad v_{J+1/2}(x) = \begin{cases} (1-x)/\Delta x_J, & x_{J-1/2} \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, for $g \in \mathcal{M}_0^1(\delta_x)$,

$$(2.8) \quad g(x) = \sum_{j=0}^J g(x_{j+\frac{1}{2}}) v_{j+\frac{1}{2}}(x);$$

i.e., g is determined by its values at $x_{j+\frac{1}{2}}$.

Given a sequence of time steps $\Delta t^0, \Delta t^1, \dots, \Delta t^{n-1}$, let

$$t^n = \sum_{k=0}^{n-1} \Delta t^k,$$

and let N^* be such that $t^{N^*} = T$. Throughout this paper, Δt^n will be chosen in accordance with a CFL constraint as described below. Finally, let $\lambda_j^n = \Delta t^n / \Delta x_j$, and let $g^n(x) = g(x, t^n)$.

We will employ the following discrete difference operators. Given b_j , $j = 1, \dots, J$, let

$$\Delta_+ b_j = \frac{b_{j+1} - b_j}{\Delta x_{j+\frac{1}{2}}}, \quad j = 1, \dots, J-1.$$

For $j = 2, \dots, J$, define

$$\Delta_- b_j = \frac{b_j - b_{j-1}}{\Delta x_{j-\frac{1}{2}}}.$$

For $b = b^n$ time-dependent, let

$$\partial_t b_j^{n+1} = \frac{b_j^{n+1} - b_j^n}{\Delta t^n}.$$

For Γ a measurable spatial domain, let $L^p(\Gamma)$ denote the standard Banach space, $1 \leq p \leq \infty$, with $\|\cdot\|_{L^p}$ the corresponding norm, and let $\|\cdot\| \equiv \|\cdot\|_{L^2}$. Let $W_p^q(\Gamma)$ denote the standard Sobolev space of functions whose derivatives of order $\leq q$ are in $L^p(\Gamma)$, with $1 \leq p \leq \infty$ and q a nonnegative integer, and let $\|\cdot\|_{W_p^q(\Gamma)}$ denote the corresponding norm. When $p = 2$, we denote $\|\cdot\|_{W_2^q}$ by $\|\cdot\|_q$. When $\Gamma = I$, we write L^p for $L^p(I)$, etc.

Let $(\cdot, \cdot)$ denote the L^2 inner product on I , and for $g(x, t), h(x, t) \in L^2(I)$ for each t , let

$$(g, h)_a \equiv \left(\frac{g(\cdot, t)}{a(\cdot, t)}, h(\cdot, t) \right).$$

Let $\langle \cdot, \cdot \rangle$ denote the trapezoidal rule approximation to $(\cdot, \cdot)$ on the mesh δ_x ; i.e.,

$$\langle g, h \rangle = \sum_{j=0}^{J-1} g_{j+\frac{1}{2}} h_{j+\frac{1}{2}} \Delta x_{j+\frac{1}{2}},$$

and let

$$\langle g, h \rangle_a = \left\langle \frac{g}{a}, h \right\rangle.$$

Let $\langle g, g \rangle \equiv |||g|||^2$ and note that, for $g \in \mathcal{M}_0^1(\delta_x)$, and quasi-uniform mesh, there exists a constant C , independent of Δx , such that

$$(2.9) \quad |||g|| \leq C |||g|||.$$

In the uniform mesh case, $C = 1$. Moreover, let $|||g|||_a^2 = \langle g, g \rangle_a$.

Let $[t_1, t_2] \subset [0, T]$, and let $\mathcal{X} = \mathcal{X}(\Gamma)$ denote a Banach or Sobolev space defined on Γ . To incorporate time dependence, we employ the notation $L^p(t_1, t_2; \mathcal{X})$ and $\|\cdot\|_{L^p(t_1, t_2; \mathcal{X})}$ to denote the space and norm, respectively, of $\mathcal{X}$ -valued functions g with the map $t \rightarrow \|f(\cdot, t)\|_{\mathcal{X}}$ belonging to $L^p(t_1, t_2)$. If $[t_1, t_2] = [0, T]$, we simplify our notation and write $L^p(\mathcal{X})$ for $L^p(0, T; \mathcal{X})$. Similar definitions hold with L^p replaced by a Sobolev space W_p^q .

In our analysis, unless indicated otherwise, we define $\Delta x = \max_j \Delta x_j$ and $\Delta t = \max_n \Delta t^n$. Moreover, C represents a generic constant, independent of Δx and Δt . The notation $C = C(\phi)$ means C is a constant which depends on ϕ . We will also use the standard inequality,

$$bc \leq \epsilon b^2 + \frac{1}{4\epsilon} c^2, \quad b, c, \epsilon \in \mathbf{R}, \quad \epsilon > 0.$$

3. The Godunov-mixed method. The approach we use to solve (1.1)-(1.3) is based on time-splitting. Assume we have an approximation $S^n(x)$ to $s(x, t^n)$ satisfying (1.1), and we want an approximation to $s(x, t^{n+1})$. We first apply the MUSCL (or ENO) scheme to the equation

$$(3.1) \quad \bar{s}_t + f(\bar{s})_x = 0, \quad (x, t) \in I \times (t^n, t^{n+1}],$$

with initial condition

$$(3.2) \quad \bar{s}(x, t^n) = S^n(x), \quad x \in I.$$

We then apply the mixed method to the equation

$$(3.3) \quad s_t^* - (a(x, t)s_x^*)_x = 0, \quad (x, t) \in I \times (t^n, t^{n+1}],$$

with initial condition

$$(3.4) \quad s^*(x, t^n) = \bar{S}^{n+1}(x), \quad x \in I,$$

where $\bar{S}^{n+1}(x)$ is the MUSCL (ENO) approximation to $\bar{s}(x, t^{n+1})$. The result of these two steps is an approximation, $S^{n+1}(x)$, to $s(x, t^{n+1})$; $S^{n+1}(x)$ is then the initial condition for the next advection step.

We now describe these schemes.

The MUSCL scheme. In the MUSCL algorithm, the initial condition $S^n \in \mathcal{M}_{-1}^0(\delta_x)$. Given S^n , we construct a function $R(x; S^n) \in \mathcal{M}_{-1}^1(\delta_x)$ which is of the form

$$(3.5) \quad R(x; S^n) = S_j^n + (x - x_j)\delta S_j^n, \quad x \in B_j,$$

where δS_j^n is determined below.

The MUSCL algorithm can be derived by integrating (3.1) over the space-time domain $B_j \times [t^n, t^{n+1}]$ and applying the midpoint rule in time. Doing so one obtains, for f and s sufficiently smooth,

$$(3.6) \quad \int_{B_j} (\bar{s}^{n+1}(x) - S^n(x)) dx + \Delta t^n [f(\bar{s}(x_{j+\frac{1}{2}}, t^{n+\frac{1}{2}})) - f(\bar{s}(x_{j-\frac{1}{2}}, t^{n+\frac{1}{2}}))] + \mathcal{O}(\Delta t^2) = 0,$$

where $t^{n+\frac{1}{2}} = (t^n + t^{n+1})/2$.

For the case $f' \geq 0$, the term $f(\bar{s}(x_{j+\frac{1}{2}}, t^{n+\frac{1}{2}}))$ is approximated by $f(S_{j,L}^n)$, where $S_{j,L}^n$ is determined by characteristic tracing from the point $(x_{j+\frac{1}{2}}, t^{n+\frac{1}{2}})$ back to time t^n . We construct the line $\bar{x}(t)$ satisfying

$$(3.7) \quad \bar{x}'(t) = f'(S_j^n), \quad \bar{x}(t^{n+\frac{1}{2}}) = x_{j+\frac{1}{2}}.$$

Assuming the CFL constraint

$$(3.8) \quad \sup_j f'(S_j^n) \lambda_j^n \leq 1,$$

we find that $\bar{x}(t)$ crosses the $t = t^n$ axis at the point $x_{j,L}$, where

$$x_{j,L} = x_{j+\frac{1}{2}} - \frac{\Delta t^n}{2} f'(S_j^n) \in B_j;$$

By (3.5), $S_{j,L}^n$ is then given by

$$(3.9) \quad S_{j,L}^n = R(x_{j,L}; S^n) = S_j^n + \frac{\Delta x_j}{2} (1 - f'(S_j^n) \lambda_j^n) \delta S_j^n.$$

This is valid for $j = 1, \dots, J$; $S_{0,L}^n$ is determined by the boundary conditions and is given below.

Given S_j^n for $j = 1, \dots, J$, and $S_{j,L}^n$ for $j = 0, \dots, J$, define $F(x; S_L^n) \in \mathcal{M}_0^1(\delta_x)$ such that

$$(3.10) \quad F(x_{j+\frac{1}{2}}; S_L^n) = f(S_{j,L}^n), \quad j = 0, \dots, J.$$

Then, approximate $\bar{s}(x, t^{n+1})$ by $\bar{S}^{n+1} \in \mathcal{M}_{-1}^0(\delta_x)$, where $\bar{S}_j^{n+1}$ is determined by

$$(3.11) \quad \bar{S}_j^{n+1} = S_j^n - \frac{\Delta t^n}{\Delta x_j} (F(x_{j+\frac{1}{2}}; S_L^n) - F(x_{j-\frac{1}{2}}; S_L^n)),$$

which follows from (3.6). Multiplying by a constant $w_j \Delta x_j$, and summing on j , (3.11) can be written as

$$(3.12) \quad (\bar{S}^{n+1}, w) = (S^n, w) - \Delta t^n (F(S_L^n)_x, w), \quad w \in \mathcal{M}_{-1}^0(\delta_x).$$

For general, smooth $f(s)$, the approximation to $f(\bar{s}(x_{j+\frac{1}{2}}, t^{n+\frac{1}{2}}))$ is determined by calculating left and right states, $S_{j,L}^n$ and $S_{j,R}^n$, and solving a Riemann problem. Let $\eta_j^+ = \max(0, f'(S_j^n))$, and $\eta_j^- = \min(0, f'(S_j^n))$. Set

$$(3.13) \quad S_{j,L}^n = S_j^n + \frac{\Delta x_j}{2} (1 - \eta_j^+ \lambda_j^n) \delta S_j^n,$$

and

$$(3.14) \quad S_{j,R}^n = S_{j+1}^n - \frac{\Delta x_{j+1}}{2} (1 - \eta_{j+1}^- \lambda_{j+1}^n) \delta S_{j+1}^n.$$

The numerical flux $h^G(S_{j,L}^n, S_{j,R}^n)$ which approximates $f(\bar{s}_{j+\frac{1}{2}}^{n+\frac{1}{2}})$ is given by the formula

$$(3.15) \quad h^G(v_L, v_R) = \begin{cases} \min_{[v_L, v_R]} f(v), & \text{if } v_L \leq v_R, \\ \max_{[v_R, v_L]} f(v), & \text{otherwise;} \end{cases}$$

h^G is referred to as the Godunov flux [7]. In this case, defining $F(x; S_L^n, S_R^n) \in \mathcal{M}_0^1(\delta_x)$ by

$$(3.16) \quad F(x_{j+\frac{1}{2}}; S_L^n, S_R^n) = h^G(S_{j,L}^n, S_{j,R}^n), \quad j = 0, \dots, J,$$

we obtain

$$(3.17) \quad (\bar{S}^{n+1}, w) = (S^n, w) - \Delta t^n (F(S_L^n, S_R^n)_x, w), \quad w \in \mathcal{M}_{-1}^0(\delta_x).$$

The ENO scheme. The ENO extension of the MUSCL scheme varies from the MUSCL scheme only in the definitions of the slope δS_j^n and the left and right states $S_{j,L}^n$ and $S_{j,R}^n$. The calculation of δS_j^n is given below. The left and right states are given by (3.13) and (3.14) with $\eta_j^+ = \eta_j^- = f'(S_j^n)$.

The lowest-order mixed method. Given $\bar{S}^{n+1}$ as the initial condition, we derive a mixed method for (3.3) and (3.4) using the lowest-order Raviart-Thomas approximating spaces.

Let s^* satisfy (3.3). Let

$$(3.18) \quad u(x, t) = -a(x, t) s_x(x, t), \quad x \in I,$$

and let $U^{n+1} \approx u^{n+1}$, $U^{n+1} \in \mathcal{M}_0^1(\delta_x)$. Let $S^{n+1} \approx s^*(x, t^{n+1})$, where $S^{n+1} \in \mathcal{M}_{-1}^0(\delta_x)$. The approximations U^{n+1} and S^{n+1} are determined by

$$(3.19) \quad \left(\frac{1}{a^{n+1}} U^{n+1}, v \right) - (S^{n+1}, v_x) = -s^{n+1} v|_0^1, \quad v \in \mathcal{M}_0^1(\delta_x),$$

and

$$(3.20) \quad \left(\frac{S^{n+1} - \bar{S}^{n+1}}{\Delta t^n}, w \right) + (U_x^{n+1}, w) = 0, \quad w \in \mathcal{M}_{-1}^0(\delta_x).$$

Combining (3.19), (3.20), and (3.17), the GMM is given by

$$(3.21) \quad \left(\frac{1}{a^{n+1}} U^{n+1}, v \right) - (S^{n+1}, v_x) = -s^{n+1} v|_0^1, \quad v \in \mathcal{M}_0^1(\delta_x),$$

$$(3.22) \quad (\partial_t S^{n+1}, w) + (F(S_L^n, S_R^n)_x, w) + (U_x^{n+1}, w) = 0, \quad w \in \mathcal{M}_{-1}^0(\delta_x).$$

When the advective component of (3.22) is approximated by MUSCL, we will refer to (3.21)-(3.22) as the MUSCL-mixed method, or MMM, and when it is approximation by the ENO scheme, we will refer to it as the ENO-mixed method, or EMM.

Approximating the first integral in (3.19) using the trapezoidal rule, one obtains

$$(3.23) \quad U^{n+1}(x_{j+\frac{1}{2}}) = -a(x_{j+\frac{1}{2}}, t^{n+1}) \Delta_+ S_j^{n+1}, \quad j = 1, \dots, J-1;$$

the terms $U_{1/2}^{n+1}$ and $U_{J+\frac{1}{2}}^{n+1}$ are determined by the boundary conditions. Moreover, choosing $w = 1$ on B_j and 0 elsewhere in (3.20), and substituting (3.23), we see that

$$(3.24) \quad \frac{S_j^{n+1} - \bar{S}_j^{n+1}}{\Delta t^n} - \Delta_-(a_{j+\frac{1}{2}}^{n+1} \Delta_+ S_j^{n+1}) = 0.$$

Thus, the mixed method with lowest-order approximating spaces and the trapezoidal rule of integration is equivalent to block-centered finite differences applied to (3.3) [11, 16]. We note that, for uniform mesh, the truncation error associated with (3.24) is $\mathcal{O}(\Delta x^2 + \Delta t)$, away from boundaries, but for nonuniform mesh, the truncation error is $\mathcal{O}(1)$.

Combining (3.17) with (3.24) we obtain

$$(3.25) \quad (\partial_t S^{n+1}, w) + (F(S_L^n, S_R^n)_x, w) + (U_x^{n+1}, w) = 0, \quad w \in \mathcal{M}_{-1}^0(\delta_x),$$

and

$$(3.26) \quad \langle U^{n+1}, v \rangle_a - (S^{n+1}, v_x) = -s^{n+1} v|_0^1, \quad v \in \mathcal{M}_0^1(\delta_x).$$

Hence, to determine S^{n+1} from (3.25) and (3.26), we solve a positive definite, symmetric, tridiagonal system of equations at each time t^{n+1} . In this case, when the advective term is determined by the MUSCL algorithm, we will refer to the scheme (3.25)-(3.26) as the MMM-TR (TR - trapezoidal rule), with an analogous definition for EMM-TR.

Calculation of slopes. Given S^{n+1} , the final step in the algorithm is the construction of $R(x; S^{n+1})$. By (3.5), this involves calculating the slopes δS_j^{n+1} .

In the MUSCL scheme, δS_j^n is calculated for each n using the following “slope-limiting” procedure, which allows for piecewise linear approximations in some grid cells, but reduces the approximation to piecewise constants at local extrema. Let

$$(3.27) \quad \mu_j = \frac{\Delta x_{j+\frac{1}{2}}}{\Delta x_{j-\frac{1}{2}}},$$

and set

$$(3.28) \quad \Delta_c S_j^n = \frac{S_{j+1}^n - (1 - \mu_j^2)S_j^n - \mu_j^2 S_{j-1}^n}{\mu_j(\Delta x_{j+\frac{1}{2}} + \Delta x_{j-\frac{1}{2}})},$$

for $j = 2, \dots, J-1$. The limited slope δS_j^n for $j = 2, \dots, J-1$ is given by

$$(3.29) \quad \delta S_j^n = \min(\Delta_{lim} S_j^n, |\Delta_c S_j^n|) \cdot \text{sgn}(\Delta_c S_j^n),$$

where

$$\Delta_{lim} S_j^n = \begin{cases} \alpha_{l,j} \min(|\Delta_+ S_j^n|, |\Delta_- S_j^n|), & \text{if } (\Delta_+ S_j^n) \cdot (\Delta_- S_j^n) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Here $\alpha_{l,j}$ is a parameter satisfying

$$(3.30) \quad 0 \leq \alpha_{l,j} \leq 2 \frac{\Delta x_{j+\frac{1}{2}}}{\Delta x_j}, \quad j = 2, \dots, J-1.$$

Limited slopes δS_1^n and δS_J^n incorporating the boundary conditions are given below. In practice one normally chooses the parameter $\alpha_{l,j}$ to be close to its upper bound. This allows for steeper representation of sharp fronts. It also has the effect of allowing the slope-limiter to choose the centered slope $\Delta_c S_j^n$ more often. We note that when $\alpha_{l,j} \equiv 0$, the MUSCL scheme reduces to the first-order Godunov scheme [7].

In the ENO scheme, the slope is given by

$$(3.31) \quad \delta S_j^n = \begin{cases} \Delta_+ S_j^n, & \text{if } |\Delta_+ S_j^n| \leq |\Delta_- S_j^n|, \\ \Delta_- S_j^n, & \text{otherwise.} \end{cases}$$

Enforcing boundary and initial conditions. In the MUSCL scheme, a limited slope can be calculated for the first interval by setting

$$(3.32) \quad \delta S_1^n = \min(\Delta_{lim} S_1^n, |\Delta_c S_1^n|) \cdot \text{sgn}(\Delta_c S_1^n),$$

where

$$(3.33) \quad \Delta_c S_1^n = \frac{S_2^n - (1 - 4\mu_1^2)S_1^n - 4\mu_1^2 g_0^n}{\mu_1(\Delta x_1 + 2\Delta x_{\frac{3}{2}})},$$

and

$$\Delta_{lim} S_1^n = \begin{cases} \alpha_{l,1} \min(|\Delta_+ S_1^n|, |\Delta_- S_1^n|), & \text{if } (\Delta_+ S_1^n) \cdot (\Delta_- S_1^n) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Here $\alpha_{l,1}$ satisfies

$$(3.34) \quad 0 \leq \alpha_{l,1} \leq \frac{\Delta x_{3/2}}{\Delta x_1},$$

$$(3.35) \quad \Delta_- S_1^n = \frac{2(S_1^n - g_0^n)}{\Delta x_1},$$

and

$$(3.36) \quad \mu_1 = \frac{\Delta x_{\frac{3}{2}}}{\Delta x_1}.$$

A similar formula for the term δS_J^n can be derived with $\alpha_{l,J}$ satisfying

$$(3.37) \quad 0 \leq \alpha_{l,J} \leq 1;$$

i.e.,

$$(3.38) \quad \delta S_J^n = \min(\Delta_{lim} S_J^n, |\Delta_c S_J^n|) \cdot \text{sgn}(\Delta_c S_J^n),$$

where

$$(3.39) \quad \Delta_c S_J^n = \frac{4\mu_J^2 g_1^n - (4\mu_J^2 - 1)S_J^n - S_{J-1}^n}{\mu_J(\Delta x_J + 2\Delta x_{J-\frac{3}{2}})},$$

and

$$\Delta_{lim} S_J^n = \begin{cases} \alpha_{l,J} \min(|\Delta_+ S_J^n|, |\Delta_- S_J^n|), & \text{if } (\Delta_+ S_J^n) \cdot (\Delta_- S_J^n) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Here,

$$(3.40) \quad \Delta_+ S_J^n = \frac{2(g_1^n - S_J^n)}{\Delta x_J},$$

and

$$(3.41) \quad \mu_J = \frac{\Delta x_{J-\frac{3}{2}}}{\Delta x_J}.$$

The ENO slope in the first interval is given by (3.31) with $\Delta_- S_1^n$ given by (3.35). A similar definition holds in the last interval.

Dirichlet boundary conditions (1.3) are further enforced by setting

$$(3.42) \quad S_{0,L}^n = g_0(t^n), \quad S_{J,R}^n = g_1(t^n).$$

When applying the trapezoidal rule in (3.19), we obtain

$$(3.43) \quad U_{1/2}^n = -a_{1/2}^n \frac{2(S_1^n - g_0^n)}{\Delta x_1},$$

$$(3.44) \quad U_{J+1/2}^n = -a_{J+1/2}^n \frac{2(g_1^n - S_J^n)}{\Delta x_J}.$$

Calculating δS_j^0 by (3.29), $j = 2, \dots, J-1$, and calculating δS_1^0 and δS_J^0 by the proper formulas as given above, we approximate the initial condition (1.2) by setting

$$(3.45) \quad S^0(x) = S_j^0 + (x - x_j)\delta S_j^0, \quad x \in B_j,$$

where

$$(3.46) \quad S_j^0 = (\Delta x_j)^{-1} \int_{B_j} s^0(x) dx.$$

Remarks on general time-splitting techniques. The procedure outlined above can be generalized so that one performs M advection steps per diffusion step, where M is a positive integer. This approach may be useful in cases where the CFL time-step constraint associated with the higher-order Godunov scheme limits the overall efficiency of the method. Note that there is no stability time-step constraint associated with the mixed method, since we employed a fully implicit time-stepping procedure.

Let $\mathcal{E}_A^{\Delta x}(t)$, $\mathcal{E}_D^{\Delta x}(t)$ denote the approximate advection and diffusion solution operators associated with the MUSCL (or ENO) and mixed methods, respectively. Let $R(x; S^n)$ be given by (3.5). The GMM can then be written in the abstract form

$$S^{n+1} = \mathcal{E}_D^{\Delta x}(\Delta t^n) \circ \mathcal{E}_A^{\Delta x}(\Delta t^n) \circ R(x; S^n).$$

Let Δt_A and Δt_D denote time-steps for advection and diffusion, respectively, where

$$(3.47) \quad \Delta t_D = M \Delta t_A,$$

and Δt_A is determined by a CFL constraint. Taking M advection steps per diffusion step then gives the following scheme:

$$(3.48) \quad S^{n+1} = \mathcal{E}_D^{\Delta x}(\Delta t_D) \circ \left(\prod_{m=1}^M \mathcal{E}_A^{\Delta x}(\Delta t_A) \circ R(x; S^{n,m}) \right),$$

where $S_j^{n,1} = S_j^n$, $j = 1, \dots, J$.

For more details on this idea, see [6].

4. Summary of Theoretical Results. In this section we summarize theoretical results which are proved in the next two sections. We will concentrate on the basic scheme presented in Section 3. Error estimates for the scheme given by (3.48) can be found in [6]. We will remark on this estimate in Section 7. We will assume $f' \geq 0$; the extension of the error estimates to more general f is also discussed in Section 7.

The first result we will state concerns the stability of the scheme just derived. We show that with certain time-step restrictions and assumptions on the slope-limiting parameter $\alpha_{l,j}$, the MMM-TR solution S is L^∞ -stable.

LEMMA 4.1. *Let S be the MMM-TR approximation to the solution of (1.1)-(1.3) with $f' \geq 0$; Let the slope-limiting parameters $\alpha_{l,j}$ satisfy (3.30)-(3.34). Moreover, set*

$$(4.1) \quad \Delta x = \min_j \Delta x_j,$$

$$(4.2) \quad \alpha_l = \max_j \frac{\alpha_{l,j} \Delta x_j}{\Delta x_{j-\frac{1}{2}}}.$$

Let Δt^n satisfy

$$(4.3) \quad \frac{\Delta t^n}{\Delta x} \sup_{\hat{S}^n} f'(s) \leq \frac{1}{1 + \frac{\alpha_l}{2}}.$$

where

$$\hat{S}^n = [\min(g_0^n, g_1^n, \min_j S_j^n), \max(g_0^n, g_1^n, \max_j S_j^n)].$$

Then, S satisfies

$$(4.4) \quad m^n \leq S_j^n \leq M^n$$

for $j = 1, \dots, J$ and $n = 0, \dots, N^*$; where

$$m^n = \min(\min_j S_j^0, \inf_{[0, t^n]} g_0(t), \inf_{[0, t^n]} g_1(t))$$

and

$$M^n = \max(\max_j S_j^0, \sup_{[0, t^n]} g_0(t), \sup_{[0, t^n]} g_1(t)).$$

We now state $L^\infty(L^2)$ -type estimates for the error in the approximation solution, and $L^2(L^2)$ -type estimates for the error in the diffusive flux u , for the GMM applied to (1.1)-(1.3).

In our analysis, we make the following smoothness assumptions on the coefficients and the solution, namely,

- (i) $f \in W_\infty^2(\mathbf{R})$, Lipschitz continuous with Lipschitz constant L_f , and $f' \geq 0$,
- (ii) $a \in W_\infty^1(W_\infty^2)$, $0 < a_* < a(x, t) < a^* < \infty$,
- (iii) $s^0 \in W_\infty^2(I)$,
- (iv) $g_0, g_1 \in W_\infty^2(0, T)$,
- (v) $u \in L^\infty(W_2^2)$,
- (vi) $s \in W_\infty^1(W_\infty^1) \cap L^\infty(W_\infty^2) \cap W_\infty^2(L^\infty)$.

Let C^* be a constant which satisfies

$$\|f\|_{W_\infty^2(\mathbf{R})}, \|a\|_{W_\infty^1(W_\infty^2)}, \|s^0\|_{W_\infty^2(I)}, \|g_0\|_{W_\infty^2(0, T)}, \|g_1\|_{W_\infty^2(0, T)} \leq C^*,$$

and

$$\|s\|_{W_\infty^1(W_\infty^1)}, \|s\|_{L^\infty(W_\infty^2)}, \|s\|_{W_\infty^2(L^\infty)}, \|u\|_{L^\infty(W_2^2)} \leq C^*.$$

Let $\tilde{s}$ be the L^2 -projection of s into $\mathcal{M}_{-1}^0(\delta_x)$; i.e.,

$$(4.5) \quad (s - \tilde{s}, w) = 0, \quad w \in \mathcal{M}_{-1}^0(\delta_x),$$

and let u_I be the interpolant of u in the space $\mathcal{M}_0^1(\delta_x)$. Then

$$(4.6) \quad \tilde{s}_j = (\Delta x_j)^{-1} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} s(x) dx = s(x_j) + \mathcal{O}(\Delta x_j^2), \quad j = 1, \dots, J,$$

and by approximation theory

$$(4.7) \quad \|u - u_I\| \leq C\Delta x^2 \|u\|_2.$$

THEOREM 4.2. *Assume Δt^n satisfies the constraints given in Lemma 4.1. Let S be the MMM approximation to s , and U the approximation to the diffusive flux u . Then,*

$$(4.8) \quad \max_{0 \leq n \leq N^*} \|S^n - \tilde{s}^n\| + \left(\sum_{n=1}^{N^*} \|U^n - u_I^n\|_a^2 \Delta t^n \right)^{1/2} \leq C(\Delta x + \Delta t),$$

where C is independent of Δx and Δt ,

$$(4.9) \quad C = C(a_*, C^*, L_f).$$

For (S, U) the EMM approximations to (s, u) ,

$$(4.10) \quad \max_{0 \leq n \leq N^*} \|S^n - \tilde{s}^n\| + \left(\sum_{n=1}^{N^*} \|U^n - u_I^n\|_a^2 \Delta t^n \right)^{1/2} \leq C(\Delta x^2 + \Delta t).$$

The above theorem shows the MMM to be first-order in $L^\infty(L^2)$, which is not optimal when approximating by piecewise linears. This represents a worst-case scenario based on the accuracy of the slope-limiting procedure. First, note that we have not precluded the case $\alpha_{l,j} = 0$, which is the Godunov scheme and which is clearly first-order accurate. For $\alpha_{l,j} \neq 0$, the MUSCL scheme uses linear functions in cells where the approximate solution does not have a local extrema; in neighborhoods of local extrema, piecewise constants are used. Thus, the truncation error for the MUSCL scheme can vary between $\mathcal{O}(1)$ and $\mathcal{O}(\Delta x^2)$ depending on the slope-limiting procedure. If the slope-limiter forces $\delta S_j^n = 0$ for “most” j ’s and n ’s, then one would expect the estimate (4.8) to be optimal, since a piecewise constant approximation is used over most of the space-time domain. Heuristically, however, one expects that in most cases the truncation error for the MUSCL scheme will approach $\mathcal{O}(\Delta x^2)$ as the mesh tends to zero; i.e., that piecewise linear approximations will be used over “most” of the domain. We will examine this effect numerically in Section 7.

We now state results for the MMM-TR and EMM-TR schemes.

THEOREM 4.3. *Let the assumptions of Theorem 4.2 hold. Let $\bar{u}^n \in \mathcal{M}_0^1(\delta_x)$ satisfy*

$$(4.11) \quad \langle \bar{u}^n, v \rangle_a = (s^n, v_x) - s^n v|_0^1, \quad v \in \mathcal{M}_0^1(\delta_x).$$

Let (S, U) be the MMM-TR or EMM-TR approximations to (s, u) . Then,

$$(4.12) \quad \max_{0 \leq n \leq N^*} \|S^n - \tilde{s}^n\| + \left(\sum_{n=1}^{N^*} \|U^n - \bar{u}^n\|_a^2 \Delta t^n \right)^{1/2} \leq C(\Delta x^p + \Delta x^q + \Delta t),$$

where $p = 1$ for MMM-TR, $p = 2$ for EMM-TR, and $q = 3/2$ for uniform mesh, $q = 1$ for nonuniform mesh.

The $\Delta x^{3/2}$ term in (4.12) in the uniform mesh case is a result of the way the mixed method approximates Dirichlet boundary conditions. Based on experimental evidence given in Section 7, this effect appears to be real and not just an artifact of the proof.

Finally, we prove a result for the L^2 error of the EMM-TR scheme which is valid for nonuniform mesh.

THEOREM 4.4. *Let the assumptions of Theorem 4.2 hold, with the added smoothness assumptions*

$$(4.13) \quad u_t \in L^\infty(W_2^2).$$

Let S be the EMM-TR approximation to s , then

$$(4.14) \quad \max_{0 \leq n \leq N^*} \|S^n - \bar{s}^n\| \leq C(\Delta x^2 + \Delta t).$$

5. Proof of Lemma 4.1. In this section, we extend earlier results of Wheeler [17] to prove Lemma 4.1.

Let $\Delta_-^u S_j^n = S_j^n - S_{j-1}^n$, $j = 2, \dots, J$, $\Delta_-^u S_1^n = 2(S_1^n - g_0^n)$, and $h_j(s) = f'(s) \frac{\Delta t^n}{\Delta x_j}$.

Applying the Mean Value Theorem to f in (3.11) we find

$$(5.1) \quad \begin{aligned} \bar{S}_j^{n+1} &= S_j^n - h_j(\beta_j)(S_{j,L}^n - S_{j-1,L}^n) \\ &= S_j^n - D_{j-\frac{1}{2}} \Delta_-^u S_j^n, \end{aligned}$$

where

$$D_{j-\frac{1}{2}} = \begin{cases} h_j(\beta_j) [\Delta_-^u S_j^n + \frac{\Delta x_j}{2} (1 - h_j(S_j^n)) \delta S_j^n \\ \quad - \frac{\Delta x_{j-1}}{2} (1 - h_{j-1}(S_{j-1}^n)) \delta S_{j-1}^n] / \Delta_-^u S_j^n, & \text{if } \Delta_-^u S_j^n \neq 0 \text{ and } j \neq 1, \\ h_1(\beta_1) [\Delta_-^u S_1^n + \frac{\Delta x_1}{2} (1 - h_1(S_1^n)) \delta S_1^n] / \Delta_-^u S_1^n, & \text{if } \Delta_-^u S_1^n \neq 0 \text{ and } j = 1, \\ 0, & \text{otherwise,} \end{cases}$$

and β_j is some point between $S_{j,L}^n$ and $S_{j-1,L}^n$. We now show $0 \leq D_{j-\frac{1}{2}} \leq 1$.

First, note that by the definitions of δS_j^n and $\alpha_{l,j}$ (see (3.29), (3.30), (3.37), and (3.38)), and the definition of $S_{j,L}^n$ (see (3.9)), that

$$(5.2) \quad \min(S_j^n, S_{j+1}^n) \leq S_{j,L}^n \leq \max(S_j^n, S_{j+1}^n),$$

for $1 \leq j \leq J-1$, and

$$(5.3) \quad \min(S_j^n, g_1^n) \leq S_{j,L}^n \leq \max(S_j^n, g_1^n).$$

Since β_j is between $S_{j,L}^n$ and $S_{j-1,L}^n$, then by (5.2), (5.3), and (3.42), $\beta_j \in \hat{S}^n$ for $1 \leq j \leq J$, thus by (4.3),

$$(5.4) \quad h_j(\beta_j) \leq \frac{1}{1 + \frac{\alpha_l}{2}}.$$

Note that by the slope-limiting procedure, δS_j^n and δS_{j-1}^n are zero or have the same sign as $\Delta_-^u S_j^n$. Hence by (3.30), for $j \neq 1$,

$$(5.5) \quad D_{j-\frac{1}{2}} \geq h_j(\beta_j) (1 - \frac{\Delta x_{j-1} \alpha_{l,j-1}}{2 \Delta x_{j-\frac{1}{2}}}) \geq 0,$$

and, for $j = 1$,

$$(5.6) \quad D_{\frac{1}{2}} \geq h_1(\beta_1) \geq 0.$$

Moreover, by (5.4), (4.2), and (4.3),

$$(5.7) \quad D_{j-\frac{1}{2}} \leq h_j(\beta_j)(1 + \frac{\alpha_{l,j}\Delta x_j}{2\Delta x_{j-\frac{1}{2}}}) \leq 1.$$

By (5.1), and (5.5)-(5.7),

$$(5.8) \quad \min(g_0^n, \min_{1 \leq j \leq J} S_j^n) \leq \bar{S}_j^{n+1} \leq \max(g_0^n, \max_{1 \leq j \leq J} S_j^n).$$

Recall

$$(5.9) \quad S_j^{n+1} - \frac{\Delta t^n}{\Delta x_j} \left[a_{j+\frac{1}{2}}^{n+1} \frac{S_{j+1}^{n+1} - S_j^{n+1}}{\Delta x_{j+\frac{1}{2}}} - a_{j-\frac{1}{2}}^{n+1} \frac{S_j^{n+1} - S_{j-1}^{n+1}}{\Delta x_{j-\frac{1}{2}}} \right] = \bar{S}_j^{n+1}.$$

Suppose

$$S_{\hat{j}}^{n+1} = \max_j S_j^{n+1}$$

for some $\hat{j}$ between 2 and $J-1$. Since $a > 0$, it follows easily that

$$(5.10) \quad S_{\hat{j}}^{n+1} \leq \bar{S}_{\hat{j}}^{n+1}.$$

Suppose

$$S_1^{n+1} = \max_j S_j^{n+1},$$

and recall $S_0^{n+1} = S_1^{n+1} - 2g_0^{n+1}$. We have two cases. Either $S_1^{n+1} \leq g_0^{n+1}$ or $S_1^{n+1} \geq g_0^{n+1}$. In the latter case, we obtain

$$S_1^{n+1} \leq \bar{S}_1^{n+1}.$$

Thus, in either case,

$$(5.11) \quad S_1^{n+1} \leq \max(\bar{S}_1^{n+1}, g_0^{n+1}).$$

Similarly,

$$(5.12) \quad S_J^{n+1} \leq \max(\bar{S}_J^{n+1}, g_1^{n+1}).$$

Combining (5.10), (5.11), and (5.12) with (5.8) we find

$$\max_j S_j^{n+1} \leq \max(\max_j S_j^n, \sup_{[t^n, t^{n+1}]} g_0(t), \sup_{[t^n, t^{n+1}]} g_1(t)).$$

By similar arguments,

$$\min_j S_j^{n+1} \geq \min(\min_j S_j^n, \inf_{[t^n, t^{n+1}]} g_0(t), \inf_{[t^n, t^{n+1}]} g_1(t)).$$

The lemma now follows by induction on n . //

6. Proofs of Theorems 4.2-4.4. *Proof of Theorem 4.2.* By (1.1), the solution pair $s, u = -as_x$, satisfies

$$(6.1) \quad (s_t^{n+1}, w) + (u_x^{n+1}, w) = -(f(s^{n+1})_x, w), \quad w \in \mathcal{M}_{-1}^0(\delta_x),$$

$$(6.2) \quad (u^{n+1}, v)_a - (s^{n+1}, v_x) = -s^{n+1}v|_0^1, \quad v \in \mathcal{M}_0^1(\delta_x).$$

Let

$$(6.3) \quad \tilde{s}_{j,L} = \tilde{s}_j^n + \frac{\Delta x_j}{2}(1 - f'(\tilde{s}_j^n)\lambda_j^n)\delta\tilde{s}_j^n, \quad j = 1, \dots, J,$$

where, for the MMM,

$$(6.4) \quad \delta\tilde{s}_j^n = \begin{cases} \alpha_{l,j}\Delta_+\tilde{s}_j^n, & \text{if } \delta S_j^n = \alpha_{l,j}\Delta_+S_j^n, \\ \alpha_{l,j}\Delta_-\tilde{s}_j^n, & \text{if } \delta S_j^n = \alpha_{l,j}\Delta_-S_j^n, \\ \Delta_c\tilde{s}_j^n, & \text{if } \delta S_j^n = \Delta_cS_j^n, \\ 0, & \text{if } \delta S_j^n = 0. \end{cases}$$

For the EMM,

$$(6.5) \quad \delta\tilde{s}_j^n = \begin{cases} \Delta_+\tilde{s}_j^n, & \text{if } \delta S_j^n = \Delta_+S_j^n, \\ \Delta_-\tilde{s}_j^n, & \text{if } \delta S_j^n = \Delta_-S_j^n, \end{cases}$$

Moreover, we define

$$(6.6) \quad \tilde{s}_{0,L}^n = g_0^n.$$

Let $\xi = S - \tilde{s}$, $\eta = U - u_I$, then by (3.22), (4.5), and (6.1),

$$(6.7) \quad \begin{aligned} (\partial_t \xi^{n+1}, w) + (\eta_x^{n+1}, w) &= -(F(S_L^n)_x, w) - (\partial_t s^{n+1}, w) - ((u_I)_x^{n+1}, w) \\ &= (f(s^{n+1})_x, w) - (F(S_L^n)_x, w) + (s_t^{n+1}, w) - (\partial_t s^{n+1}, w) \\ &\quad + ((u^{n+1} - u_I^{n+1})_x, w), \quad w \in \mathcal{M}_{-1}^0(\delta_x). \end{aligned}$$

Note that the last term on the right side of (6.7) vanishes since u_I interpolates u and w is piecewise constant. By (6.2), (4.5), and (3.21), and since $v \in \mathcal{M}_0^1(\delta_x)$ implies $v_x \in \mathcal{M}_{-1}^0(\delta_x)$,

$$(6.8) \quad \begin{aligned} (\eta^{n+1}, v)_a - (\xi^{n+1}, v_x) &= -(u_I^{n+1}, v)_a + (\tilde{s}^{n+1}, v_x) - s^{n+1}v|_0^1 \\ &= (u^{n+1} - u_I^{n+1}, v)_a, \quad v \in \mathcal{M}_0^1(\delta_x). \end{aligned}$$

Thus, setting $w = \xi^{n+1}$, $v = \eta^{n+1}$, and adding (6.7) and (6.8), we find

$$(6.9) \quad \begin{aligned} &\frac{1}{2\Delta t^n} [||\xi^{n+1}||^2 - ||\xi^n||^2 + ||\xi^{n+1} - \xi^n||^2] + ||\eta^{n+1}||_a^2 \\ &= (f(s^{n+1})_x - F(S_L^n)_x, \xi^{n+1}) + (s_t^{n+1} - \partial_t s^{n+1}, \xi^{n+1}) + (u^{n+1} - u_I^{n+1}, \eta^{n+1})_a \\ &= I_1 + I_2 + I_3. \end{aligned}$$

Let $F(x; s^{n+1}) \in \mathcal{M}_0^1(\delta_x)$ satisfy

$$(6.10) \quad F(x_{j+\frac{1}{2}}; s^{n+1}) = f(s_{j+\frac{1}{2}}^{n+1}), \quad j = 0, \dots, J.$$

Then, by (6.8),

$$\begin{aligned}
I_1 &= ((F(s^{n+1}) - F(S_L^n))_x, \xi^{n+1}) \\
&= (\eta^{n+1}, F(s^{n+1}) - F(S_L^n))_a - (u^{n+1} - u_I^{n+1}, F(s^{n+1}) - F(S_L^n))_a \\
(6.11) \quad &\leq \frac{1}{4} \|\eta^{n+1}\|_a^2 + C(a_*^{-1}) \|u_I^{n+1} - u^{n+1}\|^2 + C(a_*^{-1}) \|F(s^{n+1}) - F(S_L^n)\|^2.
\end{aligned}$$

By (2.9), and the Lipschitz continuity of f ,

$$\begin{aligned}
\|F(s^{n+1}) - F(S_L^n)\|^2 &\leq C \|F(s^{n+1}) - F(S_L^n)\|^2 \\
&\leq C \sum_{j=0}^J |f(s_{j+\frac{1}{2}}^{n+1}) - f(\tilde{s}_{j,L}^n)|^2 \Delta x_{j+\frac{1}{2}} \\
&\quad + C \sum_{j=0}^J |f(\tilde{s}_{j,L}^n) - f(S_{j,L}^n)|^2 \Delta x_{j+\frac{1}{2}} \\
(6.12) \quad &\leq C(L_f) \sum_{j=0}^J |s_{j+\frac{1}{2}}^{n+1} - \tilde{s}_{j,L}^n|^2 \Delta x_{j+\frac{1}{2}} + C(L_f) \|\tilde{s}_L^n - S_L^n\|^2.
\end{aligned}$$

By Taylor series expansion and the differential equation, for $j > 0$,

$$\begin{aligned}
(6.13) \quad s(x_{j+\frac{1}{2}}, t^{n+\frac{1}{2}}) &= s_j^n + \frac{\Delta x_j}{2} (1 - f'(s_j^n) \frac{\Delta t^n}{\Delta x_j}) s_x^n(x_j) - \frac{\Delta t^n}{2} u_x^n(x_j) \\
&\quad + \mathcal{O}(\Delta t^2 + \Delta x^2).
\end{aligned}$$

Thus, by (6.13) and (6.3),

$$\begin{aligned}
s_{j+\frac{1}{2}}^{n+1} - \tilde{s}_{j,L}^n &= s_{j+\frac{1}{2}}^{n+1} - s_{j+\frac{1}{2}}^{n+\frac{1}{2}} + s_{j+\frac{1}{2}}^{n+\frac{1}{2}} - \tilde{s}_{j,L}^n \\
(6.14) \quad &= \int_{t^{n+\frac{1}{2}}}^{t^{n+1}} s_t(x_j, t) dt + [s_j^n - \tilde{s}_j^n] \\
&\quad + \frac{\Delta x_j}{2} (1 - f'(\tilde{s}_j^n) \frac{\Delta t^n}{\Delta x_j}) (s_x^n(x_j) - \delta \tilde{s}_j^n) + \frac{\Delta t^n}{2} (f'(\tilde{s}_j^n) - f'(s_j^n)) s_x^n(x_j) \\
&\quad + \frac{\Delta t^n}{2} u_x^n(x_j) + \mathcal{O}(\Delta t^2 + \Delta x^2).
\end{aligned}$$

Note that, for the MMM, the third term is $\mathcal{O}(\Delta x)$ under certain choices of $\delta \tilde{s}_j^n$. Heuristically, one would expect $\delta \tilde{s}_j^n$ to converge to $s_x^n(x_j)$ as $\Delta x \rightarrow 0$, however, proving this theoretically remains an open question. Also note that, for the EMM, this term is always $\mathcal{O}(\Delta x^2)$, regardless of the choice of $\delta \tilde{s}_j^n$. For $j = 0$,

$$(6.15) \quad s_{1/2}^{n+1} - \tilde{s}_{0,L}^n = g_0^{n+1} - g_0^n \leq C(C^*) \Delta t^n.$$

Thus, by (6.14), (4.6), and (6.15),

$$(6.16) \quad |s_{j+\frac{1}{2}}^{n+1} - \tilde{s}_{j,L}^n| \leq C(C^*) (\Delta t + \Delta x^p),$$

where $p = 1$ for the MMM, $p = 2$ for the EMM. Moreover, by definition, for $j > 0$,

$$\begin{aligned}
\tilde{s}_{j,L}^n - S_{j,L}^n &= \xi_j^n + \frac{\Delta x_j}{2} (1 - f'(S_j^n) \frac{\Delta t^n}{\Delta x_j}) (\delta \tilde{s}_j^n - \delta S_j^n) \\
&\quad + \frac{\Delta t^n}{2} (f'(\tilde{s}_j^n) - f'(S_j^n)) \delta \tilde{s}_j^n.
\end{aligned}$$

For $j = 0$, $\tilde{s}_{0,L}^n - S_{0,L}^n = 0$. Thus, by the definitions of $\delta \tilde{s}_j^n$ and δS_j^n , and the smoothness of f , we find

$$(6.17) \quad |||\tilde{s}_L^n - S_L^n|||^2 \leq C(C^*) \|\xi^n\|^2.$$

Combining (6.16), (6.17), and (4.7) with (6.11),

$$(6.18) \quad I_1 \leq \frac{1}{4} \|\eta^{n+1}\|_a^2 + C(C^*, a_*^{-1}, L_f)(\Delta x^{2p} + \Delta t^2 + \|\xi^n\|^2).$$

By time truncation error analysis,

$$(6.19) \quad I_2 \leq C \Delta t^n \|s_{tt}\|_{L^2(t^n, t^{n+1}; L^2)}^2 + C \|\xi^{n+1}\|^2,$$

and by (4.7),

$$(6.20) \quad I_3 \leq C(a_*^{-1}) \Delta x^4 \|u^{n+1}\|_2^2 + \frac{1}{4} \|\eta^{n+1}\|_a^2.$$

Combining (6.18)-(6.20) with (6.9), multiplying by Δt^n and summing on n , $n = 0, \dots, N-1$, where $N \leq N^*$ is arbitrary, we obtain

$$(6.21) \quad \|\xi^N\|^2 + \sum_{n=1}^N \|\eta^n\|_a^2 \Delta t^n \leq C(\Delta x^4 + \Delta x^{2p} + \Delta t^2) + C \sum_{n=0}^{N-1} \|\xi^{n+1}\|^2 \Delta t^n.$$

Note that by (3.46), $\xi^0 \equiv 0$. Applying a discrete version of Gronwall's Lemma [9], we complete the proof of Theorem 4.2. //

Proof of Theorem 4.3. Let $\bar{u}^n \in \mathcal{M}_0^1(\delta_x)$ be defined by

$$(6.22) \quad \langle \bar{u}^n, v \rangle_a = (s^n, v_x) - s^n v|_0^1, \quad v \in \mathcal{M}_0^1(\delta_x).$$

Hence

$$(6.23) \quad \bar{u}_{j+\frac{1}{2}}^n = -a_{j+\frac{1}{2}}^n \frac{\tilde{s}_{j+1}^n - \tilde{s}_j^n}{\Delta x_{j+\frac{1}{2}}}, \quad j = 1, \dots, J-1,$$

and

$$(6.24) \quad \bar{u}_{1/2}^n = -a_{1/2}^n \frac{2(\tilde{s}_1^n - g_0^n)}{\Delta x_1},$$

$$(6.25) \quad \bar{u}_{J+1/2}^n = -a_{J+1/2}^n \frac{2(g_1^n - \tilde{s}_J^n)}{\Delta x_J}.$$

Let $\tau = U - \bar{u}$, $\zeta = u - \bar{u}$, and $\xi = S - \tilde{s}$, where $\tilde{s}$ is defined by (4.5), and S, U are the MMM-TR or EMM-TR approximations to s, u . Then by (6.1) and (3.25),

$$(6.26) \quad (\partial_t \xi^{n+1}, w) + (\tau_x^{n+1}, w) = (\zeta_x^{n+1}, w) + (s_t^{n+1} - \partial_t s^{n+1}, w) \\ + (f(s^{n+1})_x - F(S_L^n)_x, w), \quad w \in \mathcal{M}_{-1}^0(\delta_x),$$

and

$$(6.27) \quad \langle \tau^{n+1}, v \rangle_a - (\xi^{n+1}, v_x) = -\langle \bar{u}^{n+1}, v \rangle_a + (\tilde{s}^{n+1}, v_x) - s^{n+1} v|_0^1 \\ = 0, \quad v \in \mathcal{M}_0^1(\delta_x),$$

by (3.26), (6.22), and (4.5).

Thus, setting $w = \xi^{n+1}$, $v = \tau^{n+1}$, and adding (6.26) and (6.27), we obtain

$$\begin{aligned}
 & \frac{1}{2\Delta t^n} \left[\|\xi^{n+1}\|^2 - \|\xi^n\|^2 + \|\xi^{n+1} - \xi^n\|^2 \right] + \|\tau^{n+1}\|_a^2 \\
 & = (f(s^{n+1})_x - F(S_L^n)_x, \xi^{n+1}) + (s_t^{n+1} - \partial_t s^{n+1}, \xi^{n+1}) + (\zeta_x^{n+1}, \xi^{n+1}) \\
 (6.28) \quad & = I_1 + I_2 + I_3.
 \end{aligned}$$

Defining $F(x; s^{n+1}) \in \mathcal{M}_0^1(\delta_x)$ by (6.10), then by (6.27), (6.16), and (6.17),

$$\begin{aligned}
 I_1 & = (F(s^n)_x - F(S_L^n)_x, \xi^{n+1}) \\
 & = \langle \tau^{n+1}, F(s^{n+1}) - F(S_L^n) \rangle_a \\
 (6.29) \quad & \leq \frac{1}{4} \|\tau^{n+1}\|_a^2 + C(L_f, a_*, C^*)(\Delta x^{2p} + \Delta t^2 + \|\xi^n\|^2).
 \end{aligned}$$

An estimate for I_2 is given by (6.19).

For I_3 , note that by (6.27),

$$(6.30) \quad (\zeta_x^{n+1}, \xi^{n+1}) = \langle \tau^{n+1}, \zeta^{n+1} \rangle_a.$$

Thus

$$(6.31) \quad I_3 \leq \frac{1}{4} \|\tau^{n+1}\|_a^2 + C \|\zeta^{n+1}\|_a^2.$$

By (6.23) and Taylor expansion, for $j \neq 0, J$,

$$\begin{aligned}
 (a_{j+\frac{1}{2}}^n)^{-1} |\zeta_{j+\frac{1}{2}}^n| & = \left| \frac{\tilde{s}_{j+1}^n - \tilde{s}_j^n}{\Delta x_{j+\frac{1}{2}}} - s_x(x_{j+\frac{1}{2}}, t^n) \right| \\
 (6.32) \quad & \leq C \begin{cases} \Delta x^2 \Delta x_{j+\frac{1}{2}}^{-1/2} \|s_{xx}^n\|_{L^2(x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})}, & \text{uniform mesh,} \\ \Delta x \Delta x_{j+\frac{1}{2}}^{-1/2} \|s_x^n\|_{L^2(x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})}, & \text{nonuniform mesh.} \end{cases}
 \end{aligned}$$

For $j = 0$,

$$\begin{aligned}
 (a_{1/2}^n)^{-1} |\zeta_{1/2}^n| & = \left| \frac{2(\tilde{s}_1^n - g_0^n)}{\Delta x_1} - s_x(0, t^n) \right| \\
 (6.33) \quad & \leq C \Delta x \|s_x^n\|_{L^\infty(0, x_1)}.
 \end{aligned}$$

A similar estimate holds for $j = J$. Thus, by (6.32), (6.33),

$$(6.34) \quad \|\zeta^n\|_a^2 \leq C(C^*)(\Delta x^{2q} + \Delta x^3),$$

and consequently,

$$(6.35) \quad I_3 \leq \frac{1}{4} \|\tau^{n+1}\|_a^2 + C(C^*)(\Delta x^{2q} + \Delta x^3),$$

where $q = 1$ for nonuniform mesh, $q = 2$ otherwise.

Combining (6.35), (6.29), (6.19), and (6.28), multiplying by Δt^n , summing on n , and applying Gronwall's Lemma we obtain the estimate (4.12). //

Proof of Theorem 4.4. In this proof, we compare the EMM-TR approximation S to $\hat{S}$ defined as follows. For each $t \in [0, T]$, let $\hat{U}(t) \in \mathcal{M}_0^1(\delta_x)$, $\hat{S}(t) \in \mathcal{M}_{-1}^0(\delta_x)$ be given by

$$(6.36) \quad \langle \hat{U}, v \rangle_a - (\hat{S}, v_x) = -sv|_0^1, \quad v \in \mathcal{M}_0^1(\delta_x),$$

$$(6.37) \quad (\hat{U}_x, w) = (u_x, w) = -(s_t + f(s)_x, w), \quad w \in \mathcal{M}_{-1}^0(\delta_x).$$

In the following, for a given, smooth function $g = g(x)$, let $\tilde{g}$ denote the L^2 -projection of g into $\mathcal{M}_{-1}^0(\delta_x)$. It is easily shown that

$$(6.38) \quad \|g - \tilde{g}\| \leq C\|g\|_1 \Delta x.$$

We first estimate $\hat{S} - \tilde{s}$.

LEMMA 6.1. *Let $\hat{S}$ be given by (6.36)-(6.37). Then*

$$(6.39) \quad \|\hat{S} - \tilde{s}\| \leq C_1 \Delta x^2,$$

where $C_1 = C(\|u(\cdot, t)\|_2)$. Moreover

$$(6.40) \quad \|\partial_t(\hat{S}^{n+1} - \tilde{s}^{n+1})\| \leq C_2 \Delta x^2,$$

where C_2 depends on $\|u_{xxt}\|_{L^\infty(L^2)}$ and C_1 .

Proof. Set $\nu = s - \hat{S}$, $\tilde{\nu} = \tilde{s} - \hat{S}$, and $\gamma = u - \hat{U}$. Then subtracting (6.36) from (6.2), we obtain

$$(6.41) \quad \langle \gamma, v \rangle_a - (\nu, v_x) = T_1(u, v), \quad v \in \mathcal{M}_0^1(\delta_x),$$

where

$$(6.42) \quad T_1(u, v) = \langle u, v \rangle_a - (a^{-1}u, v).$$

Moreover, from (6.37),

$$(6.43) \quad (\gamma_x, w) = 0, \quad w \in \mathcal{M}_{-1}^0(\delta_x),$$

By (6.43) and (6.38), setting $w = \hat{U}_x - \tilde{u}_x$, we find

$$(6.44) \quad \begin{aligned} \|\gamma_x\|^2 &= (\gamma_x, \gamma_x - w) \\ &= (\gamma_x, u_x - \tilde{u}_x) \\ &\leq C \Delta x \|\gamma_x\| \|u(\cdot, t)\|_2. \end{aligned}$$

Thus

$$(6.45) \quad \|\gamma_x\| \leq C \|u(\cdot, t)\|_2 \Delta x.$$

Let

$$(6.46) \quad -(a\phi_x)_x = \tilde{\nu}, \quad x \in I,$$

$$(6.47) \quad \phi(0) = \phi(1) = 0.$$

Note that by elliptic regularity,

$$(6.48) \quad \|\phi\|_2 \leq C\|\tilde{\nu}\|.$$

Setting $\bar{v} = -a\phi_x$, we have by (6.46) that $\bar{v} \in \mathcal{M}_0^1(\delta_x)$, and by (6.41) and (6.43),

$$(6.49) \quad \begin{aligned} \|\tilde{\nu}\|^2 &= (\nu, \bar{v}_x) \\ &= \langle \gamma, \bar{v} \rangle_a - T_1(u, \bar{v}) \\ &= -(\gamma, \phi_x) + T_1(\gamma, \bar{v}) - T_1(u, \bar{v}) \\ &\equiv R_1 + R_2 + R_3. \end{aligned}$$

Integrating by parts and applying (6.43), (6.45), (6.38), and (6.48), we find

$$(6.50) \quad \begin{aligned} R_1 &= (\gamma_x, \phi - \tilde{\phi}) \\ &\leq C\Delta x^2 \|u\|_2 \|\phi\|_1 \\ &\leq C\Delta x^4 + \frac{1}{8} \|\tilde{\nu}\|^2. \end{aligned}$$

Consider R_3 . By error analysis for the trapezoidal rule of integration, using the fact that $\bar{v}_{xx} = 0$ on B_j by (6.46), and applying (6.48),

$$(6.51) \quad R_3 \leq C\Delta x^4 \|u\|_2^2 + \frac{1}{8} \|\tilde{\nu}\|^2.$$

Similarly, since $\gamma_{xx} = u_{xx}$ on B_j , and $\|\gamma_x\|$ is bounded,

$$(6.52) \quad R_2 \leq C\Delta x^4 \|u\|_2^2 + \frac{1}{8} \|\tilde{\nu}\|^2.$$

Combining (6.50)-(6.52) with (6.49) completes the proof of (6.39).

The proof of (6.40) involves differencing (6.41) and (6.43) in time and following an argument similar to the one given above. We leave the details of this proof to the reader.

Let $\beta = S - \hat{S}$ and $\kappa = U - \hat{U}$, where S, U are the EMM-TR approximations to s, u . Then, by (6.36), (6.37), and (3.25), we obtain

$$(6.53) \quad \begin{aligned} (\partial_t \beta^{n+1}, w) + (\kappa_x^{n+1}, w) &= (s_t^{n+1} - \partial_t \hat{S}^{n+1}, w) + (f(s^{n+1})_x - F(S_L^n)_x, w) \\ &= (s_t^{n+1} - \partial_t s^{n+1}, w) + (\partial_t \nu^{n+1}, w) \\ &\quad + (f(s^{n+1})_x - F(S_L^n)_x, w), \quad w \in \mathcal{M}_{-1}^0(\delta_x), \end{aligned}$$

and

$$(6.54) \quad \langle \kappa^{n+1}, v \rangle_a - (\beta^{n+1}, v_x) = 0, \quad v \in \mathcal{M}_0^1(\delta_x).$$

Setting $w = \beta^{n+1}$ and $v = \kappa^{n+1}$, we find

$$(6.55) \quad \begin{aligned} &\frac{1}{2\Delta t^n} \left[\|\beta^{n+1}\|^2 - \|\beta^n\|^2 + \|\beta^{n+1} - \beta^n\|^2 \right] + \|\kappa^{n+1}\|_a^2 \\ &= -(\partial_t \hat{S}^{n+1}, w) - (\hat{U}_x^{n+1}, w) - (F(S_L^n)_x, w) \\ &= (f(s^{n+1})_x - F(S_L^n)_x, \beta^{n+1}) + (s_t^{n+1} - \partial_t s^{n+1}, \beta^{n+1}) \\ &\quad + (\partial_t \tilde{\nu}^{n+1}, \beta^{n+1}) \\ &\equiv I'_1 + I'_2 + I'_3. \end{aligned}$$

Estimates for I'_1 and I'_2 can be obtained by arguments similar to those used to derive (6.29) and (6.19). Defining $\hat{S}_{j,L}^n$ analogously to $\bar{s}_{j,L}^n$ (see (6.3)-(6.6)), we find

$$\begin{aligned}
 I'_1 &\equiv (f(s^{n+1})_x - F(S_L^n)_x, \beta^{n+1}) \\
 &= \langle \kappa^{n+1}, F(s^{n+1}) - F(S_L^n) \rangle_a \\
 &\leq \frac{1}{4} \| \kappa^{n+1} \|_a^2 + C(a_*^{-1}) \| F(s^{n+1}) - F(\bar{s}_L^n) \|^2 \\
 &\quad + C(a_*^{-1}) [\| F(\bar{s}_L^n) - F(\hat{S}_L^n) \|^2 + \| F(\hat{S}_L^n) - F(S_L^n) \|^2] \\
 (6.56) \quad &\leq \frac{1}{4} \| \kappa^{n+1} \|_a^2 + C(L_f, C_1, a_*^{-1}, C^*)(\Delta x^4 + \Delta t^2 + \|\beta^n\|^2).
 \end{aligned}$$

Here we have also applied (6.39).

Moreover,

$$\begin{aligned}
 I'_2 &\equiv (s_t^{n+1} - \partial_t s^{n+1}, \beta^{n+1}) \\
 (6.57) \quad &\leq C\Delta t^n \| s_{tt} \|_{L^2(t^n, t^{n+1}; L^2)}^2 + C\|\beta^{n+1}\|^2.
 \end{aligned}$$

By (6.40), we obtain

$$\begin{aligned}
 I'_3 &\equiv (\partial_t \tilde{v}^{n+1}, \beta^{n+1}) \\
 (6.58) \quad &\leq C(C_2)\Delta x^4 + C\|\beta^{n+1}\|^2.
 \end{aligned}$$

Thus, combining (6.55)-(6.58), multiplying by Δt^n and summing on n , and applying Gronwall's Lemma, we obtain

$$(6.59) \quad \|\beta^N\| \leq C(\Delta x^2 + \Delta t) + \|\beta^0\|^2.$$

Apply the triangle inequality and (6.37) to obtain

$$(6.60) \quad \|\xi^N\| \leq C(\Delta x^2 + \Delta t).$$

7. Extensions and experimental rates of convergence. *Extension of the estimates to the scheme (3.48).* The scheme outlined in (3.48) is given by

$$\begin{aligned}
 (\partial_t S^{n+1}, w) + \frac{\Delta t_A}{\Delta t_D} \sum_{m=1}^M (F(S_L^{n,m})_x, w) + (U_x^{n+1}, w) &= 0, \quad w \in \mathcal{M}_{-1}^0(\delta_x), \\
 (U^{n+1}, v)_a - (S^{n+1}, v_x) &= -s^{n+1} v|_0^1, \quad v \in \mathcal{M}_0^1(\delta_x),
 \end{aligned}$$

where

$$(7.1) \quad S^{n,1} = S^n,$$

$$(7.2) \quad S_{j,L}^{n,m} = S_j^{n,m} + \frac{\Delta x_j}{2} (1 - f'(S_j^{n,m}) \frac{\Delta t_A}{\Delta x_j}) \delta S_j^{n,m}, \quad j \neq 1,$$

$$(7.3) \quad S_{0,L}^{n,m} = g_0^{n,m},$$

and

$$(7.4) \quad (S^{n,m+1}, w) = (S^{n,m}, w) - \Delta t_A (F(S_L^{n,m})_x, w), \quad w \in \mathcal{M}_{-1}^0(\delta_x).$$

Let $G(x; S^n) = \frac{\Delta t \Delta}{\Delta t_D} \sum_{m=1}^M F(x; S_L^{n,m})$.

In this case, emulating the proof of Theorem 4.2:

$$(7.5) \quad (\partial_t \xi^{n+1}, w) + (\eta_x^{n+1}, w) = (f(s^{n+1})_x - G(S^n)_x, w) + (s_t^{n+1} - \partial_t s^{n+1}, w),$$

$$(7.6) \quad (\eta^{n+1}, v)_a - (\xi^{n+1}, v_x) = (u^{n+1} - u_I^{n+1}, v)_a.$$

Setting $v = \eta^{n+1}$, $w = \xi^{n+1}$, we obtain an equation of the form (6.9) with

$$(7.7) \quad \begin{aligned} I_1 &= (f(s^{n+1}) - G(S^n), \eta^{n+1}) + (u^{n+1} - u_I^{n+1}, f(s^{n+1}) - G(S^n)) \\ &\leq \frac{1}{4} \|\eta^{n+1}\|_a^2 + C \Delta x^4 + C(a_*^{-1}) \|F(s^{n+1}) - G(S^n)\|^2. \end{aligned}$$

An estimate for the last term on the right side of (7.7) is straightforward but tedious. The details are given in [6], where it is shown that

$$(7.8) \quad \|F(s^{n+1}) - G(S^n)\| \leq C(M, C^*)(\Delta x^2 + \Delta t^2) + C(M) \|\xi^n\|^2.$$

Finally, we find

$$(7.9) \quad \|\xi^N\|^2 + \sum_{n=1}^N \|\eta^n\|_a^2 \Delta t^n \leq C(\Delta x^2 + \Delta t^2).$$

In this case, the error is provably only first order in space for both the EMM and the MMM.

Extension to general $f \in C^2$. Under the assumptions of Theorem 4.2, we define

$$\tilde{s}_{j,L}^n = \tilde{s}_j^n + \frac{\Delta x_j}{2} (1 - h_j^+(\tilde{s}_j^n)) \delta \tilde{s}_j^n,$$

where, for the analysis of the MMM, we set

$$h_j^+(\tilde{s}_j^n) = \lambda_j^n \cdot \begin{cases} f'(\tilde{s}_j^n), & \text{if } f'(S_j^n) \geq 0, \\ 0, & \text{otherwise;} \end{cases}$$

and

$$\tilde{s}_{j,R}^n = \tilde{s}_{j+1}^n - \frac{\Delta x_{j+1}}{2} (1 + h_{j+1}^-(\tilde{s}_{j+1}^n)) \delta \tilde{s}_{j+1}^n,$$

where

$$h_j^-(\tilde{s}_j^n) = \lambda_j^n \cdot \begin{cases} f'(\tilde{s}_j^n), & \text{if } f'(S_j^n) \leq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Also, $\tilde{s}_{0,L}^n = g_0^n$, and $\tilde{s}_{j,R}^n = g_1^n$. For the analysis of the EMM, we define $\tilde{s}_{j,L}^n$ and $\tilde{s}_{j,R}^n$ by the above formulas with $h_j^+(\tilde{s}_j^n) = h_j^-(\tilde{s}_j^n) = f'(\tilde{s}_j^n)$.

Note that the Godunov flux h^G given by (3.15) is consistent; i.e., $h^G(v, v) = f(v)$. It is also easily shown that h^G is Lipschitz in its arguments, in fact,

$$(7.10) \quad |h^G(v_L^1, v_R^1) - h^G(v_L^2, v_R^2)| \leq L_f \left[|v_L^1 - v_L^2| + |v_R^1 - v_R^2| \right],$$

for arbitrary left and right states, v_L^1 , v_L^2 , v_R^1 , and v_R^2 .

Emulating the proof of Theorem 4.2, the analogous bound for I_1 given in (6.11) is

$$(7.11) \quad I_1 \leq \frac{1}{4} \|\eta^{n+1}\|_a^2 + C\Delta x^4 + C\|F(s^{n+1}) - F(S_L^n, S_R^n)\|^2.$$

In this case, we have, by the consistency of h^G and (7.10),

$$(7.12) \quad \begin{aligned} \left| f(s_{j+\frac{1}{2}}^{n+1}) - F(S_{j,L}^n, S_{j,R}^n) \right| &\leq \left| f(s_{j+\frac{1}{2}}^{n+1}) - f(s_{j+\frac{1}{2}}^{n+\frac{1}{2}}) \right| + \left| h^G(s_{j+\frac{1}{2}}^{n+\frac{1}{2}}, s_{j+\frac{1}{2}}^{n+\frac{1}{2}}) - h^G(\tilde{s}_{j,L}^n, \tilde{s}_{j,R}^n) \right| \\ &\quad + \left| h^G(\tilde{s}_{j,L}^n, \tilde{s}_{j,R}^n) - h^G(S_{j,L}^n, S_{j,R}^n) \right| \\ &\leq C(L_f) \left[\left| s_{j+\frac{1}{2}}^{n+1} - s_{j+\frac{1}{2}}^{n+\frac{1}{2}} \right| + \left| s_{j+\frac{1}{2}}^{n+\frac{1}{2}} - \tilde{s}_{j,L}^n \right| + \left| s_{j+\frac{1}{2}}^{n+\frac{1}{2}} - \tilde{s}_{j,R}^n \right| \right. \\ &\quad \left. + \left| \tilde{s}_{j,L}^n - S_{j,L}^n \right| + \left| \tilde{s}_{j,R}^n - S_{j,R}^n \right| \right]. \end{aligned}$$

By the same arguments which led to (6.18), we find that

$$(7.13) \quad \|F(s^{n+1}) - F(S_L^n, S_R^n)\|^2 \leq C(\Delta x^{2p} + \Delta t^2) + C\|\xi^n\|^2,$$

where, as before, $p = 1$ for the MMM, $p = 2$ for the EMM. Thus, the estimate given in Theorem 4.2 holds. Extensions to the estimates given in Theorems 4.3 and 4.4 are straightforward.

A second-order-in-time scheme. The EMM and EMM-TR can be modified so that they are fully second order in time as well as in space. This is accomplished using Crank-Nicholson time-stepping in the mixed method, and modifying the definitions of $S_{j,L}^n$ and $S_{j,R}^n$ so that they are second-order accurate in the sense of truncation error.

In particular, for $f' \geq 0$ (with the obvious extension to more general f), the scheme can be written as

$$(7.14) \quad (\partial_t S^{n+1}, w) + (F(S_L^{n+\frac{1}{2}})_x, w) + (U_x^{n+\frac{1}{2}}, w) = 0, \quad w \in \mathcal{M}_{-1}^0(\delta_x),$$

$$(7.15) \quad (U^{n+\frac{1}{2}}, v)_a - (S^{n+\frac{1}{2}}, v_x) = -s^{n+\frac{1}{2}} v|_0^1, \quad v \in \mathcal{M}_0^1(\delta_x),$$

where $S^{n+\frac{1}{2}} = (S^n + S^{n+1})/2$,

$$(7.16) \quad S_{j,L}^{n+\frac{1}{2}} = S_{j,L}^n - \begin{cases} \frac{\Delta t}{2} u_x^0(x_j), & n = 0, \\ \frac{\Delta t}{2} U_x^{n-\frac{1}{2}}(x_j), & n > 0, \end{cases}$$

for $j = 1, \dots, J$, and for $j = 0$,

$$(7.17) \quad S_{0,L}^{n+\frac{1}{2}} = g_0^{n+\frac{1}{2}}.$$

Assume the trapezoidal rule is used to evaluate $(U^{n+\frac{1}{2}}, v)_a$. Define $\hat{S}$ and $\hat{U}$ by (6.36) and (6.37). Then $\beta = S - \hat{S}$ and $\kappa = U - \hat{U}$ satisfy

$$(7.18) \quad \begin{aligned} (\partial_t \beta^{n+1}, w) + (\kappa_x^{n+\frac{1}{2}}, w) &= (s_t^{n+\frac{1}{2}} - \partial_t s^{n+1}, w) + (\partial_t \nu^{n+1}, w) \\ &\quad + (f(s^{n+\frac{1}{2}})_x - F(S_L^{n+\frac{1}{2}})_x, w), \quad w \in \mathcal{M}_{-1}^0(\delta_x), \\ (\kappa^{n+\frac{1}{2}}, v)_a - (\beta^{n+\frac{1}{2}}, v_x) &= 0, \quad v \in \mathcal{M}_0^1(\delta_x). \end{aligned}$$

Assume the hypothesis of Theorem 4.4 holds, with an additional assumption on Δt , given below, and assume $s_{ttt} \in L^2(0, T; L^2)$. Setting $w = \beta^{n+\frac{1}{2}}$, $v = \kappa^{n+\frac{1}{2}}$, multiplying by $2\Delta t$, summing on n , applying time truncation error estimates, and Lemma 2, we find

$$\begin{aligned}
 (7.19) \quad & \|\beta^N\|^2 + 2\Delta t \sum_{n=0}^{N-1} \|\kappa^{n+\frac{1}{2}}\|_a^2 \\
 & \leq 2\Delta t \sum_{n=0}^{N-1} (f(s^{n+\frac{1}{2}})_x - F(S_L^{n+\frac{1}{2}})_x, \beta^{n+\frac{1}{2}}) \\
 & \quad + C(\Delta x^4 + \Delta t^4) + C \sum_{n=0}^{N-1} \|\beta^{n+\frac{1}{2}}\|^2 \Delta t.
 \end{aligned}$$

The first term on the right side of (7.19) is equal to

$$(7.20) \quad 2\Delta t \sum_{n=0}^{N-1} \langle F(s^{n+\frac{1}{2}}) - F(S_L^{n+\frac{1}{2}}), \kappa^{n+\frac{1}{2}} \rangle_a$$

by (7.18).

Defining $\tilde{s}_{j,L}^{n+\frac{1}{2}}$ analogously to (7.16), (7.17), with $U_x^{n-\frac{1}{2}}$ replaced by $\hat{U}_x^{n-\frac{1}{2}}$, we find by Taylor expansion (see (6.13)) that,

$$(7.21) \quad s_{j+\frac{1}{2}}^{n+\frac{1}{2}} - \tilde{s}_{j,L}^{n+\frac{1}{2}} = \mathcal{O}(\Delta t^2 + \Delta x^2) + \mathcal{O}(\Delta t)[u_x^{n-\frac{1}{2}}(x_j) - \hat{U}_x^{n-\frac{1}{2}}(x_j)].$$

Noting by (6.43) that

$$\hat{U}_x(x_j) = \frac{u_{j+\frac{1}{2}} - u_{j-\frac{1}{2}}}{\Delta x_j},$$

we find that

$$(7.22) \quad \sum_{j=0}^J |s_{j+\frac{1}{2}}^{n+\frac{1}{2}} - \tilde{s}_{j,L}^{n+\frac{1}{2}}|^2 \Delta x_{j+\frac{1}{2}} \leq C(\Delta x^4 + \Delta t^4).$$

Moreover, for $n > 0$,

$$\begin{aligned}
 (7.23) \quad & \langle F(\tilde{s}_L^{n+\frac{1}{2}}) - F(S_L^{n+\frac{1}{2}}), \kappa^{n+\frac{1}{2}} \rangle_a = \sum_{j=1}^J \left(\frac{\kappa}{a} \right)_{j+\frac{1}{2}}^{n+\frac{1}{2}} \\
 & \quad \times f'(\theta_j^n) \left[(\tilde{s}_{j,L}^n - S_{j,L}^n) + \frac{\Delta t}{2} \kappa_x^{n-\frac{1}{2}}(x_j) \right],
 \end{aligned}$$

where θ_j^n is between $\tilde{s}_{j,L}^{n+\frac{1}{2}}$ and $S_{j,L}^{n+\frac{1}{2}}$. The term involving $\tilde{s}_{j,L}^n - S_{j,L}^n$ has been dealt with before, and is bounded by

$$(7.24) \quad \frac{\epsilon}{4} \|\kappa^{n+\frac{1}{2}}\|_a^2 + C(\epsilon) \|\beta^n\|^2,$$

where ϵ is a small, positive constant. The remaining term is equal to

$$\begin{aligned}
 & \sum_{j=1}^J \left(\frac{\kappa}{a} \right)_{j+\frac{1}{2}}^{n+\frac{1}{2}} f'(\theta_j^n) \frac{\Delta t}{2\Delta x_j} \left[\kappa_{j+\frac{1}{2}}^{n-\frac{1}{2}} - \kappa_{j-\frac{1}{2}}^{n-\frac{1}{2}} \right] \Delta x_{j+\frac{1}{2}} \\
 (7.25) \quad & \leq \frac{\Delta t}{2} \left[\sup_{j,n} \left[(a_{j+\frac{1}{2}}^{n-\frac{1}{2}}/a_{j+\frac{1}{2}}^{n+\frac{1}{2}})^{1/2} \frac{f'(\theta_j^n)}{\Delta x_j} \right] + \sup_{j,n} \left[(a_{j-\frac{1}{2}}^{n-\frac{1}{2}}/a_{j+\frac{1}{2}}^{n+\frac{1}{2}})^{1/2} \frac{f'(\theta_j^n)}{\Delta x_j} \right] \right] \\
 & \quad \times (|||\kappa^{n-\frac{1}{2}}|||_a^2 + |||\kappa^{n+\frac{1}{2}}|||_a^2).
 \end{aligned}$$

For Δt sufficiently small such that

$$(7.26) \quad \Delta t \left[\sup_{j,n} \left[(a_{j+\frac{1}{2}}^{n-\frac{1}{2}}/a_{j+\frac{1}{2}}^{n+\frac{1}{2}})^{1/2} \frac{f'(\theta_j^n)}{\Delta x_j} \right] + \sup_{j,n} \left[(a_{j-\frac{1}{2}}^{n-\frac{1}{2}}/a_{j+\frac{1}{2}}^{n+\frac{1}{2}})^{1/2} \frac{f'(\theta_j^n)}{\Delta x_j} \right] \right] \leq 1 - \epsilon,$$

the right side of (7.23) is bounded by

$$(7.27) \quad \left(\frac{1}{2} - \frac{\epsilon}{4} \right) [|||\kappa^{n+\frac{1}{2}}|||_a^2 + |||\kappa^{n-\frac{1}{2}}|||_a^2] + C(\epsilon) \|\beta^n\|^2.$$

For $n = 0$, the first term on the right of (7.23) is bounded by (7.24) (with $n = 0$), while the second term is zero by the definition of $\tilde{s}_{j,L}^{\frac{1}{2}}$.

Thus

$$\begin{aligned}
 2\Delta t \sum_{n=0}^{N-1} \langle F(s^{n+\frac{1}{2}}) - F(S_L^{n+\frac{1}{2}}), \kappa^{n+\frac{1}{2}} \rangle_a &= 2\Delta t \sum_{n=0}^{N-1} \langle F(s^{n+\frac{1}{2}}) - F(\tilde{s}_L^{n+\frac{1}{2}}), \kappa^{n+\frac{1}{2}} \rangle_a \\
 &\quad + 2\Delta t \sum_{n=0}^{N-1} \langle F(\tilde{s}_L^{n+\frac{1}{2}}) - F(S_L^{n+\frac{1}{2}}), \kappa^{n+\frac{1}{2}} \rangle_a \\
 &\leq C(\epsilon)(\Delta x^4 + \Delta t^4) + C(\epsilon) \sum_{n=0}^{N-1} \|\beta^n\|^2 \Delta t \\
 &\quad + 2\Delta t \sum_{n=0}^{N-1} |||\kappa^{n+\frac{1}{2}}|||^2.
 \end{aligned}$$

Substituting into (7.19) and applying Gronwall's Lemma, we find that

$$(7.28) \quad \|\beta^N\|^2 \leq C(\Delta t^4 + \Delta x^4).$$

Thus, by Lemma 2,

$$(7.29) \quad \|\xi^N\| \leq C(\Delta x^2 + \Delta t^2).$$

Experimental rates of convergence. We conclude by presenting some experimental rates of convergence for smooth flow problems. We consider

$$\begin{aligned}
 s_t + s_x - a s_{xx} &= r(x, t), \quad 0 < x < 1, \quad 0 < t \leq .125, \\
 s(x, 0) &= \sin(\pi x),
 \end{aligned}$$

where a is a constant.

TABLE 1
 $a = 1$, error in approximate solution

MMM-TR			EMM-TR		Δt^2 Scheme	
J	$\ \xi\ * 10^4$	Rate	$\ \xi\ * 10^4$	Rate	$\ \xi\ * 10^4$	Rate
20	6.35	–	6.41	–	12.5	–
40	1.64	1.95	1.63	1.97	3.14	1.99
80	.416	1.98	.416	1.97	.788	1.99
160	.105	1.99	.105	1.99	.197	2.00

TABLE 2
 $a = 1$, error in diffusive flux

MMM-TR			Δt^2 Scheme	
J	$\ \tau\ _{L^2(L^2)} * 10^3$	Rate	$\ \tau\ _{L^2(L^2)} * 10^3$	Rate
20	3.09	–	4.29	–
40	1.09	1.50	1.44	1.58
80	.386	1.50	.495	1.54
160	.136	1.50	.173	1.51

We first set $a = 1$ and determine r such that $s(x, t) = \sin(\pi(x - t))$. The experimental rates of convergence for $\|\xi\|$ and $\|\tau\|_{L^2(L^2)}$ as defined in the proof of Theorem 4.3 for the MMM-TR and EMM-TR schemes are given in Tables 1 and 2, respectively, indicating that ξ is second order accurate, and τ is of order $\Delta x^{3/2}$, as predicted. The errors are computed at time $t = .125$. These results show that our heuristic intuition concerning the second order convergence of the MMM-TR scheme holds in this case. In these runs, $\Delta t = \Delta x^2$. We also give results for the Δt^2 scheme derived above; here $\Delta t = \Delta x/2$.

We next set $a = 10^{-2}$, and again construct r such that $s(x, t) = \sin(\pi(x - t))$. The results in this case are presented in Table 3.

Results for nonuniform mesh for the latter case are given in Table 4, these results agree with the estimates of Theorem 4.4. The coarsest mesh in this case satisfied

$$(7.30) \quad \Delta x_j = \begin{cases} .02, & j \text{ odd,} \\ .04, & j \text{ even.} \end{cases}$$

The latter meshes were determined by dividing each block in the previous mesh in half.

The final results we present are for $a = 10^{-4}$. Here we consider only the Δt^2 scheme, since the differences in the schemes for this value of a are negligible. In this case, several grid refinements are necessary before the L^2 error approaches second order accuracy. However, it was observed that $\|\xi\|_{L^1}$ approaches second order accuracy even on the coarser meshes.

TABLE 3
 $a = 10^{-2}$

MMM-TR			EMM-TR		Δt^2 Scheme	
J	$\ \xi\ * 10^4$	Rate	$\ \xi\ * 10^4$	Rate	$\ \xi\ * 10^4$	Rate
20	38.64	—	39.51	—	22.07	—
40	11.02	1.81	11.14	1.83	5.88	1.91
80	2.89	1.93	2.90	1.94	1.48	1.99
160	.729	1.97	.730	1.99	.370	2.00

TABLE 4
 $a = 10^{-2}$, *nonuniform mesh*

EMM-TR			Δt^2 Scheme	
Δx	$\ \xi\ * 10^4$	Rate	$\ \xi\ * 10^4$	Rate
.04	17.93	—	12.69	—
.02	5.09	1.81	3.64	1.80
.01	1.39	1.87	1.00	1.86
.005	.364	1.93	.262	1.93

TABLE 5
 $a = 10^{-4}$ Δt^2 Scheme

J	$\ \xi\ * 10^4$	Rate
40	8.34	—
80	2.68	1.64
160	.845	1.67
320	.255	1.73
640	.071	1.85

Acknowledgments. The author expresses his deepest gratitude to Professors Mary F. Wheeler and Todd F. Dupont for their helpful suggestions and discussions related to this work.

REFERENCES

- [1] J. B. Bell, C. N. Dawson, and G. R. Shubin, *An unsplit higher-order Godunov scheme for scalar conservation laws in two dimensions*, J. Comput. Phys., **74**, pp. 1-24 (1988).
- [2] J. B. Bell and G. R. Shubin, *Higher-order Godunov methods for reducing numerical dispersion in reservoir simulation*, SPE 13514, Proceedings of SPE Reservoir Simulation Symposium, Dallas, February 1985.
- [3] P. Colella, *A direct Eulerian MUSCL scheme for gas dynamics*, SIAM J. Sci. Stat. Comput., **6**, pp. 104-117 (1985).
- [4] P. Colella, *A multidimensional second-order Godunov scheme for conservation laws*, LBL-17023, Lawrence Berkeley Laboratory (1984).
- [5] P. Colella and P. R. Woodward, *The piecewise-parabolic method (PPM) for gas-dynamical simulations*, J. Comput. Phys., **54**, pp. 174-201 (1984).
- [6] C. N. Dawson, *Error Estimates for Godunov Mixed Methods for Nonlinear Parabolic Equations*, Ph. D. Thesis, Rice University (1988).
- [7] S. K. Godunov, *A difference scheme for numerical computation of discontinuous solutions of equations in fluid dynamics*, Math. Sbornik, **47**, pp. 271-306 (1959).
- [8] A. Harten, B. Engquist, S. Osher, and S. R. Chakravarthy, *Uniformly high order accurate essentially non-oscillatory schemes III*, J. Comput. Phys., **71**, pp. 231-303, 1987.
- [9] M. Lees, *A priori estimates for the solution of a difference approximation to parabolic differential equations*, Duke Math. Journal, **27**, pp. 297-312 (1960).
- [10] P. A. Raviart and J. M. Thomas, *A mixed finite element method for 2nd order elliptic problems*, in Mathematical Aspects of the Finite Element Method, Rome 1975, Lecture Notes in Mathematics, Springer-Verlag, Berlin (1977).
- [11] T. F. Russell and M. F. Wheeler, *Finite element and finite difference methods for continuous flow problems*, in The Mathematics of Reservoir Simulation (R. E. Ewing, ed.), Frontiers in Science, Society for Industrial and Applied Mathematics, Philadelphia (1983).
- [12] T. R. Russell, M. F. Wheeler, and C. Chiang, *Large-scale simulation of miscible displacement by mixed and characteristic finite element methods*, in Proceedings of the SEG/SIAM/SPE Conference on Mathematical and Computational Methods in Seismic Exploration and Reservoir Modeling (W. E. Fitzgibbon, ed.), SIAM (1985).
- [13] G. R. Shubin and J. B. Bell, *An analysis of the grid orientation effect in numerical simulation of miscible displacement*, Comp. Meth. in Appl. Mech. and Eng., **47**, pp. 47-71 (1983).
- [14] B. van Leer, *Towards the ultimate conservative difference scheme v.a. second-order sequel to Godunov's method*, J. Comput. Phys., **32**, pp. 101-136 (1979).
- [15] B. van Leer, *Multidimensional explicit difference schemes for hyperbolic conservation laws*, Comp. Meth. in Appl. Sci. and Eng., VI, Elsevier Science Publishers, pp. 493-499 (1984).
- [16] A. Weiser and M. F. Wheeler, *On convergence of block-centered finite differences for elliptic problems*, this Journal, **25**, pp. 351-375 (1988).
- [17] M. F. Wheeler, *Higher-order Godunov for scalar conservation laws with monotone flux functions*, EPR Special Report, Exxon Production Research (1985).
- [18] M. F. Wheeler and C. N. Dawson, *An operator-splitting method for advection-diffusion-reaction problems*, MAFELAP Proceedings VI (J. A. Whiteman, ed.), Academic Press, London, pp. 463-482 (1988).
- [19] M. F. Wheeler, W. A. Kinton, and C. N. Dawson, *Time-splitting for advection-dominated parabolic problems in one space variable*, Communications in Applied Numerical Methods, **4**, pp. 413-423 (1988).