JOURNAL OF COMPUTATIONAL PHYSICS 84, 90–113 (1989)

TVB Runge–Kutta Local Projection Discontinuous Galerkin Finite Element Method for Conservation Laws III: One-Dimensional Systems

BERNARDO COCKBURN AND SAN-YIH LIN

School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

AND

CHI-WANG SHU

Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912

Received April 28, 1988; revised October 19, 1988

This is the third paper in a series in which we construct and analyze a class of TVB (total variation bounded) discontinuous Galerkin finite element methods for solving conservation laws $\mathbf{u}_t + \sum_{i=1}^{d} (\mathbf{f}_i(\mathbf{u}))_{x_i} = 0$. In this paper we present the method in a system of equations, stressing the point of how to use the weak form in the component spaces, but to use the local projection limiting in the characteristic fields, and how to implement boundary conditions. A 1-dimensional system is thus chosen as a model. Different implementation techniques are discussed, theories analogous to scalar cases are proven for linear systems, and numerical results are given illustrating the method on nonlinear systems. Discussions of handling complicated geometries via adaptive triangle elements will appear in future papers. © 1989 Academic Press, Inc.

1. INTRODUCTION

In [3, 4] we constructed and analyzed a new class of finite element methods—we call them RK$\Lambda\Pi$P^k, or $(k+1)$th order TVB Runge–Kutta local projection discontinuous Galerkin finite element method—for solving the hyperbolic conservation law

$$\mathbf{u}_t + \sum_{i=1}^{d} (\mathbf{f}_i(\mathbf{u}))_{x_i} = 0, \tag{1.1}$$

with suitable initial or initial-boundary conditions. In (1.1) $\mathbf{u} = (u_1, ..., u_m)^T$, $\mathbf{x} = (x_1, ..., x_d)$, and $\sum_{i=1}^{d} \xi_i (\partial \mathbf{f}_i / \partial \mathbf{u})$ always has m real eigenvalues and a complete set of eigenvectors, with real ξ_i. In [4] we presented the general framework in the case $d = m = 1$, keeping in mind about the possibility of natural extensions to $d > 1$ and/or $m > 1$. For details, history and related work, see [1–4, 8, 10, 11, 13–15]. As

0021-9991/89 $3.00

indicated in [4], the main advantages of these methods over most other finite element methods are their time explicitness, hence they can be equipped with high order TVD Runge–Kutta type time discretizations in [15], and their TVB provableness in 1D, scalar nonlinear case; the main advantage over finite difference methods is the flexibility in handling complicated geometries and boundary conditions. In this paper we carry out the generalization to one dimensional system $m>1$, $d=1$. The multidimensional case $d>1$ needs adaptive triangle elements to handle complicated geometries, and will be discussed in future papers.

Theories about (1.1) as well as about numerical methods solving (1.1) are far less advanced for the system case $m>1$ than for the scalar case $m=1$. A numerical method will be considered acceptable for practical purposes if it verifies a convergence theory for the linear system case $f(u)=Au$ (the scheme in this case is usually still nonlinear) as well as for the scalar nonlinear case, if it is easily implementable to nonlinear systems and if it gives good results for solving nonlinear systems with discontinuities. We will follow this conventional approach in this paper. In Section 2 we will consider initial value problems. We present the weak form and the general framework of the scheme and consider different possible avenues to apply monotone fluxes at the interfaces and the local projection limiters. A total variation bounded estimate, similar to the scalar case $m=1$, is proven for linear systems. In Section 3 boundary condition implementation is considered. A similar total variation bounded estimate is proven, again for linear systems. Section 4 includes some numerical results, mainly for the (nonlinear) Euler's equation in gas dynamics, to illustrate the behavior of our schemes for nonlinear systems. The test problems chosen are standard. For comparison with non-oscillatory finite difference schemes and other finite element methods, we refer the readers to [5, 7, 14]. Concluding remarks are contained in Section 5.

2. Initial Value Problems

We consider in this section Eq. (1.1) with $m>1$, $d=1$ and with a pure initial condition (periodic or compactly supported):

$$\mathbf{u}(x,0)=\mathbf{u}^0(x). \tag{2.1}$$

As in the scalar case $m=1$, we shall first discretize (1.1)–(2.1) in the spatial variable x. Let $I_j=(x_{j-1/2}, x_{j+1/2})$, $I=\bigcup_j I_j$ be a partition of the real line. Denote $\Delta x_j=x_{j+1/2}-x_{j-1/2}$ and $h=\sup_j \Delta x_j$. The finite element method we are going to use is a Galerkin method for which the finite dimensional space $\mathbf{V}_h$ to which the approximate solution $\mathbf{u}^h(t)$ belongs for $t\in[0,T]$ is taken as

$$\mathbf{V}_h=\mathbf{V}_h^k=\{\mathbf{p}\text{: each of its components } p_i\in BV\cap L^1\text{:}\ p_i|_{I_j} \text{ is a polynomial of degree } \leqslant k\}, \tag{2.2}$$

All the general framework developed in [4] for the scalar case, except for the choice of monotone flux and the local projection limiting, can be applied here, just componentwisely. We have

$$\mathbf{u}^h(x,t)=\sum_{l=0}^{k} a_l \mathbf{u}_j^{(l)}(t)\, v_l^{(j)}(x) \qquad \text{for} \quad x\in I_j, \tag{2.3}$$

where $v_l^{(j)}(x)$ form a local orthogonal basis over I_j:

$$v_0^{(j)}(x)=1, \qquad v_1^{(j)}(x)=x-x_j, \qquad v_2^{(j)}(x)=(x-x_j)^2-\tfrac{1}{12}\Delta x_j^2, \ldots, \tag{2.4}$$

the coefficients a_l are

$$a_l=\frac{\Delta x_j^{l+1}}{\int_{I_j}(v_l^{(j)}(x))^2\,dx}; \qquad \text{i.e.,} \quad a_0=1, \quad a_1=\frac{12}{\Delta x_j}, \quad a_2=\frac{180}{\Delta x_j^2}, \ldots, \tag{2.5}$$

and the degrees of freedom, $\mathbf{u}_j^{(l)}(t)$, are defined by

$$\mathbf{u}_j^{(l)}=\mathbf{u}_j^{(l)}(t)=\frac{1}{\Delta x_j^{l+1}}\int_{I_j}\mathbf{u}(x,t)\, v_l^{(j)}(x)\,dx, \qquad l=0,1,\ldots,k. \tag{2.6}$$

In order to determine the degrees of freedom of $\mathbf{u}^h$ we proceed as in [4]. We multiply (1.1) by $\mathbf{v}^h\in\mathbf{V}_h^k$, integrate over I_j, and integrate by parts formally to obtain

$$\begin{aligned}\frac{d}{dt}\int_{I_j}\mathbf{u}(x,t)\,\mathbf{v}^h(x)\,dx+[\Delta_+(\mathbf{v}^h(x_{j-1/2})\,\mathbf{f}(\mathbf{u}(x_{j-1/2},t))]\\ -\int_{I_j}\mathbf{f}(\mathbf{u}(x,t))\frac{d}{dx}\mathbf{v}^h(x)\,dx=0, \qquad \forall\mathbf{v}^h\in\mathbf{V}_h^k,\end{aligned} \tag{2.7}$$

where $\Delta_\pm$ are the usual difference operators $\Delta_\pm a_j=\pm(a_{j\pm1}-a_j)$. As indicated in [4], we choose the basis functions (2.4) and the corresponding degrees of freedom (2.6) only for easy presentation. The essential ingredients are the weak formulation (2.7) and the choice of space (2.2). Other basis functions can of course be used as well.

Next, we replace the exact solution $\mathbf{u}$ by its approximation $\mathbf{u}^h$, and $\mathbf{f}(\mathbf{u}(x_{j-1/2},t))$ by some monotone flux (whose choice in the current system case will be determined later) $\mathbf{h}_{j-1/2}=\mathbf{h}(\mathbf{u}_{j-1/2}^-,\mathbf{u}_{j-1/2}^+)$, where $\mathbf{u}_{j-1/2}^\pm=\mathbf{u}^h(x_{j-1/2}^\pm,t)$ are defined by (2.3), subject to some local projection limiting to be discussed later. We obtain, after some simple algebraic manipulations,

$$\begin{aligned}\frac{d}{dt}\mathbf{u}_j^{(l)}+\frac{1}{\Delta x_j^{l+1}}[\Delta_+(v_l^{(j)}(x_{j-1/2})\,\mathbf{h}_{j-1/2}]\\ -\frac{1}{\Delta x_j^{l+1}}\int_{I_j}\mathbf{f}(\mathbf{u}^h(x,t))\frac{d}{dx}v_l^{(j)}(x)\,dx=0, \qquad l=0,1,\ldots,k.\end{aligned} \tag{2.8}$$

The integration in (2.8) can be approximated by a suitable quadrature whose

error is at most $O(h^{k+l+2})$, and (2.8) is solved in time by a TVD Runge–Kutta type method [4, 15]:

$$(\mathbf{u}^h)^{(i)} = \sum_{l=0}^{i-1} [\alpha_{il}(\mathbf{u}^h)^{(l)} + \beta_{il}\,\Delta t\,\mathbf{L}_h((\mathbf{u}^h)^{(l)}, t^n + d_l\,\Delta t)], \qquad i = 1, ..., r, \tag{2.9a}$$

$$(\mathbf{u}^h)^{(0)} = (\mathbf{u}^h)^n, \qquad (\mathbf{u}^h)^{(r)} = (\mathbf{u}^h)^{n+1}. \tag{2.9b}$$

For example,

$$\text{second order } (r=2): \quad \alpha_{10} = \beta_{10} = 1,\ \alpha_{20} = \alpha_{21} = \beta_{21} = \tfrac{1}{2},\ \beta_{20} = 0;\ d_0 = 0,\ d_1 = 1;\ \text{CFL} = 1, \tag{2.10a}$$

$$\text{third order } (r=3): \quad \alpha_{10} = \beta_{10} = 1,\ \alpha_{20} = \tfrac{3}{4},\ \beta_{20} = 0,\ \alpha_{21} = \beta_{21} = \tfrac{1}{4},\ \alpha_{30} = \tfrac{1}{3},\ \beta_{30} = \alpha_{31} = \beta_{31} = 0,\ \alpha_{32} = \beta_{32} = \tfrac{2}{3},\ d_0 = 0,\ d_1 = 1,\ d_2 = \tfrac{1}{2};\ \text{CFL} = 1, \tag{2.10b}$$

etc. The starting point of (2.9)–(2.10) is (2.8) written in a concise ODE form:

$$\frac{d}{dt}\mathbf{u}^h = \mathbf{L}_h(\mathbf{u}^h, t), \tag{2.11}$$

where we include the time variable t in case there are time dependent forcing terms or boundary conditions.

We now turn to the problem of choosing $\mathbf{h}_{j+1/2}$ in (2.8) and applying the local projection limiting. For this purpose we write (see (2.3))

$$\mathbf{u}^-_{j+1/2} = \mathbf{u}^{(0)}_j + \tilde{\mathbf{u}}_j, \qquad \mathbf{u}^+_{j-1/2} = \mathbf{u}^{(0)}_j - \tilde{\tilde{\mathbf{u}}}_j, \tag{2.12}$$

and apply the local projection limiting on $\tilde{\mathbf{u}}_j$, $\tilde{\tilde{\mathbf{u}}}_j$, as in the scalar case [4].

One simple way is to do everything just componentwisely. We define

$$\tilde{\mathbf{u}}^{(\text{mod})}_j = \mathbf{m}(\tilde{\mathbf{u}}_j, \Delta_+\mathbf{u}^{(0)}_j, \Delta_-\mathbf{u}^{(0)}_j), \qquad \tilde{\tilde{\mathbf{u}}}^{(\text{mod})}_j = \mathbf{m}(\tilde{\tilde{\mathbf{u}}}_j, \Delta_+\mathbf{u}^{(0)}_j, \Delta_-\mathbf{u}^{(0)}_j), \tag{2.13}$$

where $\mathbf{m}$ is a vector minmod function with TVB correction:

$$\mathbf{m}(\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n) = \begin{pmatrix} m((a_1)_1, (a_2)_1, ..., (a_n)_1) \\ \vdots \\ m((a_1)_m, (a_2)_m, ..., (a_n)_m) \end{pmatrix}, \tag{2.14}$$

where $\mathbf{a}_i = ((a_i)_1, (a_i)_2, ..., (a_i)_m)^{\mathrm{T}}$ and

$$m(b_1, ..., b_n) = \begin{cases} b_1, & \text{if } |b_1| \leqslant Mh^2, \\ s \cdot \min_{1 \leqslant i \leqslant n} |b_i|, & \text{if } |b_1| > Mh^2, \text{ and } \operatorname{sign}(b_1) = \cdots = \operatorname{sign}(b_n) = s, \\ 0, & \text{otherwise,} \end{cases} \tag{2.15a}$$

with

$$M = \tfrac{2}{3}M_2, \tag{2.15b}$$

or

$$M = M_{j,l} = \tfrac{2}{9}(3 + 10M_2)\, M_2 \frac{h^2}{h^2 + |(\Delta_+ \mathbf{u}_j^{(0)})_l| + |(\Delta_-(\mathbf{u}_j^{(0)}))_l|}, \tag{2.15c}$$

and M_2 is some estimate of the absolute value of the second derivatives of the solution near smooth critical points.

We then use the (local) Lax–Friedrichs flux

$$\mathbf{h}_{j+1/2} = \mathbf{h}(\mathbf{u}_{j+1/2}^-, \mathbf{u}_{j+1/2}^+) = \tfrac{1}{2}[\mathbf{f}(\mathbf{u}_{j+1/2}^-) + \mathbf{f}(\mathbf{u}_{j+1/2}^+) - \alpha_{j+1/2}(\mathbf{u}_{j+1/2}^+ - \mathbf{u}_{j+1/2}^-)], \tag{2.16a}$$

with

$$\alpha_{j+1/2} = \max_{1 \leqslant p \leqslant m} (|\lambda_{j+1/2}^{(p)+}|, |\lambda_{j+1/2}^{(p)-}|) \qquad \text{(local Lax–Friedrichs)}, \tag{2.16b}$$

or

$$\alpha_{j+1/2} \equiv \alpha = \max_{j,p} |\lambda_{j+1/2}^{(p)\pm}| \qquad \text{(Lax–Friedrichs)}, \tag{2.16c}$$

where $\lambda_{j+1/2}^{(p)\pm}$, $p = 1, ..., m$, are the m real eigenvalues of the Jacobian $(\partial \mathbf{f}/\partial \mathbf{u})|_{\mathbf{u} = \mathbf{u}_{j+1/2}^{\pm}}$.

Notice that we do not need to evaluate the Jacobian or its eigenvectors, just its eigenvalues. Thus computationally it is very simple. Unfortunately this simple version does not have a TVB theory, even for linear systems. Computationally we do observe wriggles (see Section 4), although these wriggles are usually rather small for second-order schemes.

To achieve better qualities at the price of more complicated computations, we use characteristic field decompositions. In terms of formal accuracy any average can be used. For first order schemes Roe average gives better shock transitions. Our experience is that for higher order methods the difference between the Roe average and the simple arithmetic mean diminishes. We denote by $A_{j+1/2} = (\partial \mathbf{f}/\partial \mathbf{u})_{\mathbf{u} = \mathbf{u}_{j+1/2}}$ some "average" Jacobian, e.g., $\mathbf{u}_{j+1/2} = (\mathbf{u}_j^{(0)} + \mathbf{u}_{j+1}^{(0)})/2$ (simple arithmetic mean) or (for Euler equations of gas dynamics) $\mathbf{u}_{j+1/2} = \mathbf{R}(\mathbf{u}_j^{(0)}, \mathbf{u}_{j+1}^{(0)})$, where $\mathbf{R}$ is the Roe average [12]. We denote the eigenvalues and left and right eigenvectors of $A_{j+1/2}$ by $\lambda_{j+1/2}^{(p)}$, $\mathbf{l}_{j+1/2}^{(p)}$, $\mathbf{r}_{j+1/2}^{(p)}$, $p = 1, ..., m$, normalized so that $\mathbf{l}_{j+1/2}^{(p)} \cdot \mathbf{r}_{j+1/2}^{(q)} = \delta_{pq}$. Then, in computing $\mathbf{h}_{j+1/2}$, we project everything to the eigenspace of $A_{j+1/2}$,

$$a^{(p)} = \mathbf{l}_{j+1/2}^{(p)} \cdot \mathbf{a}, \tag{2.17}$$

where we take $\mathbf{a} = \tilde{\mathbf{u}}_j$, $\tilde{\tilde{\mathbf{u}}}_{j+1}$, $\mathbf{u}_j^{(0)}$, $\mathbf{u}_{j+1}^{(0)}$, $\Delta_- \mathbf{u}_j^{(0)}$, $\Delta_+ \mathbf{u}_j^{(0)}$, $\Delta_+ \mathbf{u}_{j+1}^{(0)}$. We then apply the local projection limiting in each characteristic field:

$$(\tilde{u}_j)^{(p)(\mathrm{mod})} = m((\tilde{u}_j)^{(p)}, (\Delta_+ u_j^{(0)})^{(p)}, (\Delta_- u_j^{(0)})^{(p)}), \tag{2.18a}$$

$$(\tilde{\tilde{u}}_{j+1})^{(p)(\mathrm{mod})} = m((\tilde{\tilde{u}}_{j+1})^{(p)}, (\Delta_+ u_j^{(0)})^{(p)}, (\Delta_+ u_{j+1}^{(0)})^{(p)}). \tag{2.18b}$$

(We will drop the superscript "mod" in the following.) We form

$$(u_{j+1/2}^-)^{(p)} = (u_j^{(0)})^{(p)} + (\tilde{u}_j)^{(p)}, \qquad (u_{j+1/2}^+)^{(p)} = (u_{j+1}^{(0)})^{(p)} - (\tilde{\tilde{u}}_{j+1})^{(p)}, \tag{2.19}$$

return to the component space

$$\mathbf{a} = \sum_{p=1}^{m} a^{(p)} \mathbf{r}_{j+1/2}^{(p)}, \tag{2.20}$$

by taking $\mathbf{a} = \mathbf{u}_{j+1/2}^{\pm}$, compute $\mathbf{f}_{j+1/2}^{\pm} = \mathbf{f}(\mathbf{u}_{j+1/2}^{\pm})$, compute $(f_{j+1/2}^{\pm})^{(p)}$ by (2.17), then use any scalar monotone flux or E-flux [11] in the pth characteristic field, $p = 1, ..., m$. For example, we may use the local Lax–Friedrichs flux

$$h_{j+1/2}^{(p)} = \tfrac{1}{2}[(f_{j+1/2}^+)^{(p)} + (f_{j+1/2}^-)^{(p)} - \alpha_{j+1/2}^{(p)}((u_{j+1/2}^+)^{(p)} - (u_{j+1/2}^-)^{(p)})], \tag{2.21a}$$

with

$$\alpha_{j+1/2}^{(p)} = \max(|\lambda_j^{(p)}|, |\lambda_{j+1}^{(p)}|), \tag{2.21b}$$

(for convex case only, otherwise the maximum should be taken in the whole interval) or the Roe flux with entropy correction

$$h_{j+1/2}^{(p)} = \begin{cases} (f_{j+1/2}^+)^{(p)}, & \text{no sonic point and } \lambda_{j+1/2}^{(p)} < 0, \\ (f_{j+1/2}^-)^{(p)}, & \text{no sonic point and } \lambda_{j+1/2}^{(p)} \geqslant 0, \\ \text{same as in (2.21)}, & \text{otherwise.} \end{cases} \tag{2.22}$$

We finally get $\mathbf{h}_{j+1/2}$ by (2.20) with $\mathbf{a} = \mathbf{h}_{j+1/2}$:

$$\mathbf{h}_{j+1/2} = \sum_{p=1}^{m} h_{j+1/2}^{(p)} \mathbf{r}_{j+1/2}^{(p)}. \tag{2.23}$$

Computationally this approach needs much more work. However, it works better both theoretically and numerically. Theoretically we have the following proposition, similar to the results for scalar case [4], for linear systems.

PROPOSITION 2.1. *Scheme* (2.8)–(2.9)–(2.23) *is TVBM* (*total variation bounded in the means* $\mathbf{u}^{(0)}$ *and TVB, under the total variation definition* (*see* (2.17) *for notations*)

$$TV(\mathbf{u}^{(0)}) = \sum_j \sum_{p=1}^{m} |(u_{j+1}^{(0)})^{(p)} - (u_j^{(0)})^{(p)}|, \qquad TV(\mathbf{u}) = \sum_j \sum_{p=1}^{m} |(u_{j+1})^{(p)} - (u_j)^{(p)}|, \tag{2.24}$$

for linear system $\mathbf{f}(\mathbf{u}) = A\mathbf{u}$, *where* A *is a constant matrix, hence has a convergent subsequence in this case.*

Proof. Since $\lambda_{j+1/2}^{(p)}$, $\mathbf{l}_{j+1/2}^{(p)}$, $\mathbf{r}_{j+1/2}^{(p)}$ do not vary with j, in each characteristic field, $(u^{(0)})^{(p)}$ will satisfy a scalar TVB scheme in [4]. The same argument as in the scalar case [4] now leads to TVBM and TVB. ∎

Notice that the scheme in this case is still nonlinear. The result can be generalized to $A = A(x)$. For numerical results see Section 4.

3. Initial Boundary Value Problems

We now turn our attention to the initial boundary value problems. For simplicity we take the interval $(0, +\infty)$ and consider one boundary at $x = 0$ only. The general case of two boundaries can be handled similarly. If at the boundary $x = 0$

$$\lambda^{(1)} \leqslant \cdots \leqslant \lambda^{(s)} < 0 < \lambda^{(s+1)} \leqslant \cdots \leqslant \lambda^{(m)}, \tag{3.1}$$

where $\lambda^{(p)}$ are the eigenvalues of $\partial \mathbf{f}/\partial \mathbf{u}|_{x=0}$, then a well-posed boundary condition takes the form (see (2.17) for notations):

$$\begin{pmatrix} (u)^{(s+1)}(0,t) \\ \vdots \\ (u)^{(m)}(0,t) \end{pmatrix} = B(t) \begin{pmatrix} (u)^{(1)}(0,t) \\ \vdots \\ (u)^{(s)}(0,t) \end{pmatrix} + g(t), \tag{3.2}$$

where $B(t)$ is a $(m-s)\times s$ matrix with Lipschitz continuous components, $g(t)$ is a $(m-s)\times 1$ vector with bounded variation.

We put the boundary at $x_{-1/2} = 0$ and implement the boundary condition (3.2) as

$$\begin{gathered} (u^{-}_{-1/2})^{(p)} = (u^{+}_{-1/2})^{(p)}, \qquad (\tilde{u}_0)^{(p)(\mathrm{mod})} = m((\tilde{u}_0)^{(p)}, (\Delta_+ u_0^{(0)})^{(p)}), \\ (\tilde{\tilde{u}}_0)^{(p)(\mathrm{mod})} = m((\tilde{\tilde{u}}_0)^{(p)}, (\Delta_+ u_0^{(0)})^{(p)}), \end{gathered} \tag{3.3a}$$

for $p = 1, \ldots, s$;

$$\begin{pmatrix} (u^{-}_{-1/2})^{(s+1)} \\ \vdots \\ (u^{-}_{-1/2})^{(m)} \end{pmatrix} = B(t) \begin{pmatrix} (u^{-}_{-1/2})^{(1)} \\ \vdots \\ (u^{-}_{-1/2})^{(s)} \end{pmatrix} + g(t),$$

$$\begin{aligned} (\tilde{u}_0)^{(p)(\mathrm{mod})} = m\Bigg(& (\tilde{u}_0)^{(p)}, (\Delta_+ u_0^{(0)})^{(p)}, 2\Bigg((u_0^{(0)}) \\ & - \Bigg(\sum_{l=1}^{s} B_{p-s,l}(t)(u^{-}_{-1/2})^{(l)} + (g(t))_{p-s} \Bigg) \Bigg) \Bigg), \end{aligned} \tag{3.3b}$$

$$(\tilde{\tilde{u}}_0)^{(p)(\mathrm{mod})} = m((\tilde{u}_0)^{(p)}, (\Delta_+ u_0^{(0)})^{(p)}),$$

for $p = s+1, \ldots, m$.

Notice that, as in the scalar case [4], accuracy is not affected by the limiters. We then have the following proposition.

PROPOSITION 3.1. *Scheme* (2.8)–(2.9)–(2.23)–(3.3) *is TVBM, under the total variation definition for the mean* $\mathbf{u}^{(0)}$,

$$TV(\mathbf{u}^{(0)}) = \sum_{j \geqslant -1} \left(Q \sum_{p=1}^{s} + \sum_{p=s+1}^{m} \right) |(u_{j+1}^{(0)})^{(p)} - (u_j^{(0)})^{(p)}|, \tag{3.4}$$

with $Q = \max(1, 2\,\|B\|_1) = \max(1, 2 \max_{0 \leqslant t \leqslant T, 1 \leqslant p \leqslant s} \sum_{l=1}^{m-s} |B_{lp}(t)|)$, *and*

$$(u_{-1}^{(0)})^{(p)} \equiv \begin{cases} (u_0^{(0)})^{(p)}, & p = 1, \ldots, s \\ \displaystyle\sum_{l=1}^{s} B_{p-s,l}(t)(u_0^{(0)})^{(l)} + (g(t))_{p-s}, & p = s+1, \ldots, m, \end{cases} \tag{3.5}$$

and TVB under the total variation definition (2.24), *for linear constant coefficiented system* $\mathbf{f}(\mathbf{u}) = A\mathbf{u}$.

Proof. We only need to prove the result for the Euler forward version of (2.8). Following the lines of proofs in [4, Proposition 3.1] we have

$$((u_j^{(0)})^{(p)})^{n+1} = ((u_j^{(0)})^{(p)})^n + C_{j+1/2}^{(p)} \Delta_+((u_j^{(0)})^{(p)})^n - D_{j-1/2}^{(p)} \Delta_-((u_j^{(0)})^{(p)})^n,$$

for $j \geqslant 1$, $p = 1, \ldots, m$ and $j = 0$, $p = s+1, \ldots, m$, and

$$((u_0^{(0)})^{(p)})^{n+1} = ((u_j^{(0)})^{(p)})^n + C_{1/2}^{(p)} \Delta_+((u_0^{(0)})^{(p)})^n,$$

for $p = 1, \ldots, s$, where all the C's and D's are non-negative, module $O(h^2)$.

Following the lines of proof in [14, Theorem 3.1], we then have, for $p = s+1, \ldots, m$,

$$\begin{aligned} \Delta_+((u_{-1}^{(0)})^{(p)})^{n+1} &= (1 - D_{-1/2}^{(p)})\, \Delta_+((u_{-1}^{(0)})^{(p)})^n \\ &\quad + C_{1/2}^{(p)} \Delta_+((u_0^{(0)})^{(p)})^n - (g(t^{n+1}) - g(t^n))_{p-s} \\ &\quad - \sum_{l=1}^{s} (B_{p-s,l}(t^{n+1})((u_0^{(0)})^{(l)})^{n+1} - B_{p-s,l}(t^n)((u_0^{(0)})^{(l)})^n). \end{aligned}$$

Hence, after some technical manipulations similar to [14, Theorem 3.1], we arrive at

$$\begin{aligned} TV((\mathbf{u}^{(0)})^{n+1}) &= \sum_{j \geqslant -1} \left(Q \sum_{p=1}^{s} + \sum_{p=s+1}^{m} \right) |\Delta_+((u_j^{(0)})^{(p)})^{n+1}| \\ &\leqslant TV((\mathbf{u}^{(0)})^n) - \sum_{p=1}^{s} \left(Q - 2 \sum_{l=1}^{m-s} |B_{lp}^{n+1}| \right) C_{1/2}^{(p)} |\Delta_+(u_0^{(0)})^{(p)}| \\ &\quad + L\,\Delta t\, \mathrm{TV}((\mathbf{u}^{(0)})^n) + \sum_{p=1}^{m-s} |(g(t^{n+1}) - g(t^n))_p| + H\,\Delta t, \end{aligned}$$

where the $L\,\Delta t\,\mathrm{TV}(\mathbf{u}^{(0)})$ term is related to the Lipschitz condition of $B(t)$, it does not appear if $B(t)$ is a constant matrix. The $H\,\Delta t$ term is related to the TVB correction constant M in (2.15); it does not appear if $M \equiv 0$.

The remainder of the proof is straightforward. ∎

Remark 3.2. As in [14], the total variation definition in (3.4) has different weights for incoming and outgoing components. This is motivated by the differential equation theory and guarantees total variation diminishing of the boundary treatment (3.3) under this definition of total variation; i.e., if $B(t)$ is a constant matrix, $g \equiv 0$, $M = 0$ in (2.15), then $\mathrm{TV}((\mathbf{u}^{(0)})^{n+1}) \leqslant \mathrm{TV}((\mathbf{u}^{(0)})^n)$.

4. Numerical Results

For simplicity we use equally spaced cells, although the methods can be easily implemented for non-uniform cells. For time-dependent problems with the location of discontinuities moving in time, the cells should also move accordingly. We shall not discuss this further here.

Example 1. We consider the Riemann problems of the Euler equation of gas dynamics for a polytropic gas,

$$\mathbf{u}_t + \mathbf{f}(\mathbf{u})_x = 0, \qquad \mathbf{u}(x, 0) = \mathbf{u}^0(x) = \begin{cases} \mathbf{u}_L, & x < 0, \\ \mathbf{u}_R, & x > 0, \end{cases} \tag{4.1a}$$

$$\mathbf{u} = (\rho, m, E)^{\mathrm{T}}, \qquad \mathbf{f}(\mathbf{u}) = q\mathbf{u} + (0, p, qp)^{\mathrm{T}}, \tag{4.1b}$$

with

$$p = (\gamma - 1)(E - \tfrac{1}{2}\rho q^2), \qquad m = \rho q; \tag{4.1c}$$

$\gamma = 1.4$ is used in the following computation. For details of the Jacobian, its eigenvalues, eigenvectors, etc., see [5, 12].

Two sets of initial conditions are considered. One is proposed by Sod [17]:

$$(\rho_L, q_L, p_L) = (1, 0, 1); \qquad (\rho_R, q_R, p_R) = (0.125, 0, 0.10). \tag{4.2a}$$

The other is used by Lax [9]:

$$(\rho_L, q_L, p_L) = (0.445, 0.698, 3.528); \qquad (\rho_R, q_R, P_R) = (0.5, 0, 0.571). \tag{4.2b}$$

We test our second-order and third-order schemes, i.e., $k = 2$ and 3 in (2.8), $r = 2$ and 3 in (2.10). Both componentwise limiters (2.13)–(2.16a), (2.16b) and characteristicwise limiters (2.18)–(2.22) are tested. Local Lax–Friedrichs flux (2.16a), (2.16b), and (2.21) are used. The results of componentwise limiters for (4.2a) are in Figs. 1–6.[1] We have not included the figures for (4.2b) because they are

[1] In all figures, solid lines are for the exact solutions or converged solutions, and "+" or "0" are for the numerical solutions (just one point per cell is printed). In Figs. 1–20, we solve (4.1)–(4.2a) to $t = 2.0$, and (4.1)–(4.2b) to $t = 1.3$, using 100 cells. Density, velocity, and pressure are pictured for each case.

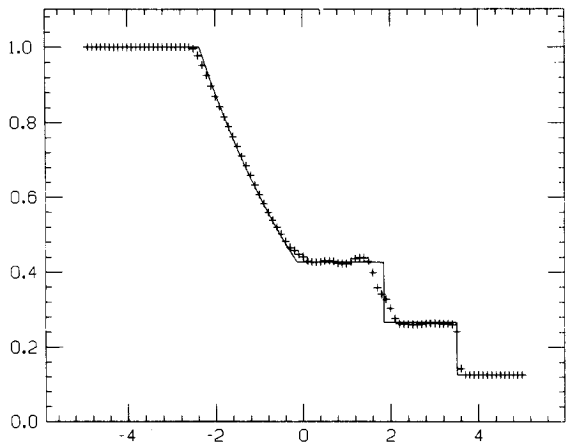

FIG. 1. Second-order, componentwise limiter (4.2a), density.

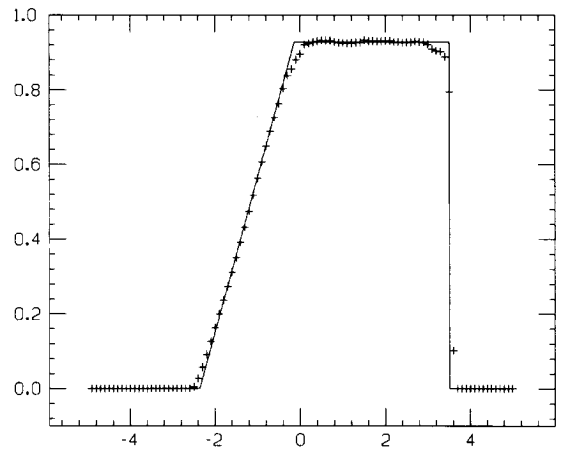

FIG. 2. Second-order, componentwise limiter (4.2a), velocity.

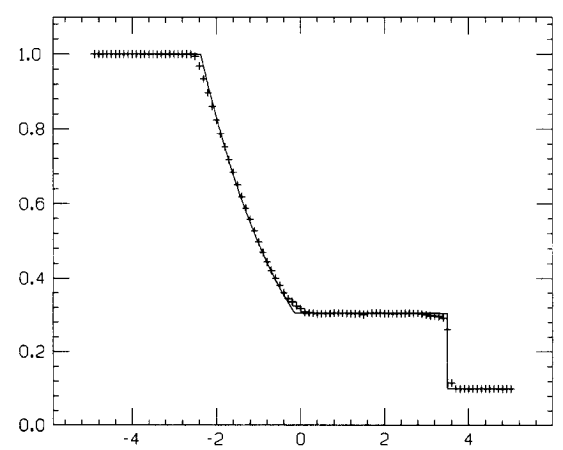

FIG. 3. Second-order, componentwise limiter (4.2a), pressure.

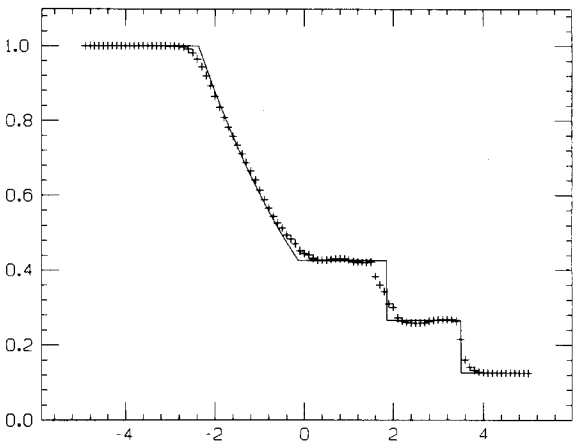

FIG. 4. Third-order, componentwise limiter (4.2a), density.

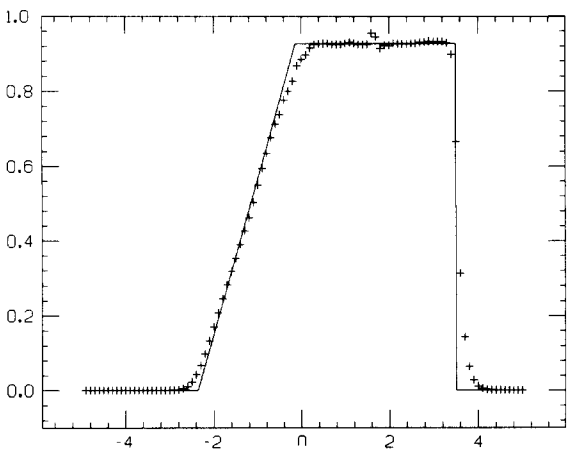

FIG. 5. Third-order, componentwise limiter (4.2a), velocity.

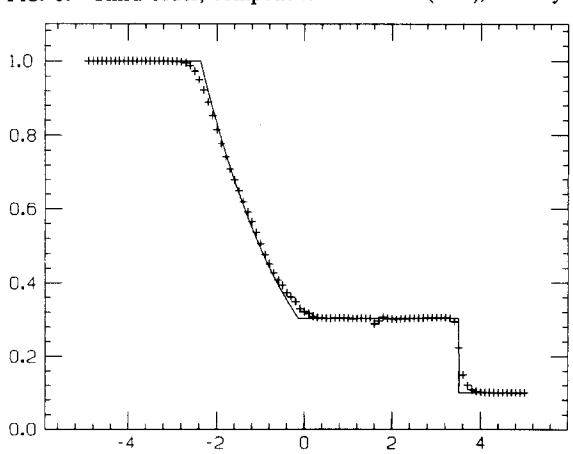

FIG. 6. Third-order, componentwise limiter (4.2a), pressure.

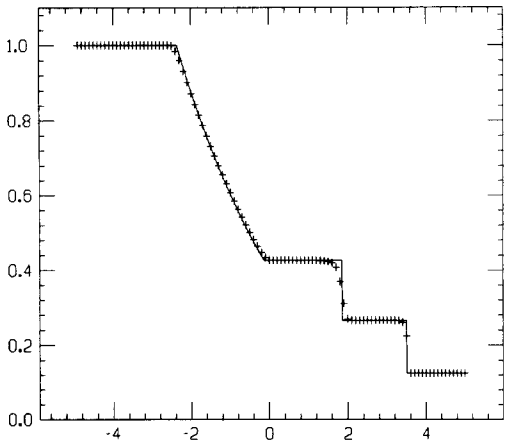

FIG. 7. Second-order, characteristicwise limiter (4.2a), density.

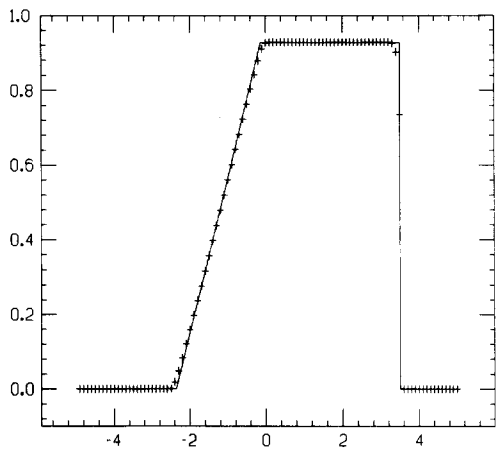

FIG. 8. Second-order, characteristicwise limiter (4.2a), velocity.

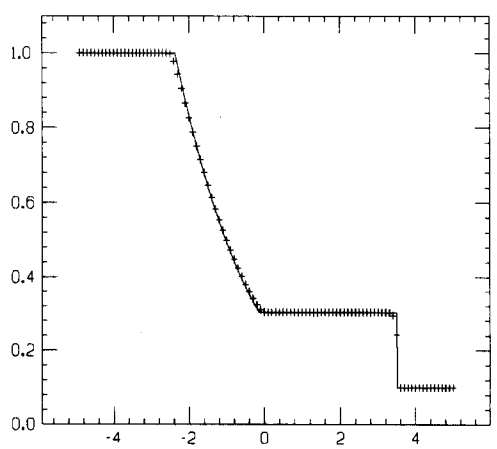

FIG. 9. Second-order, characteristicwise limiter (4.2a), pressure.

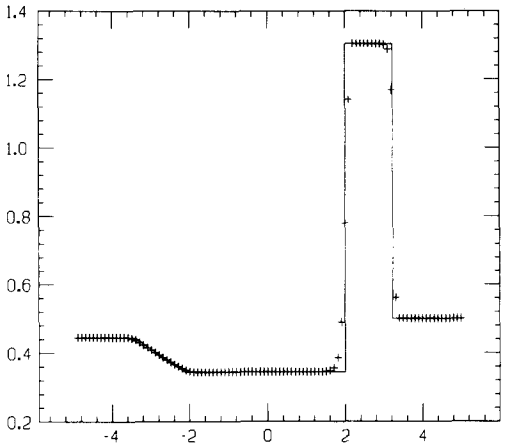

FIG. 10. Second-order, characteristicwise limiter (4.2b), density.

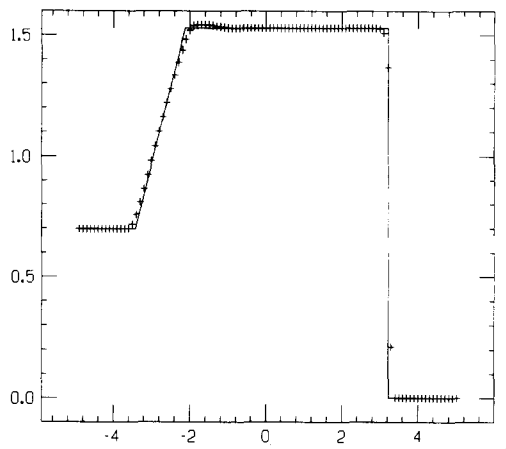

FIG. 11. Second-order, characteristicwise limiter (4.2b), velocity.

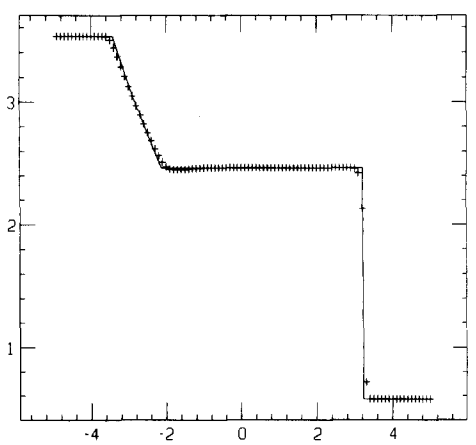

FIG. 12. Second-order, characteristicwise limiter (4.2b), pressure.

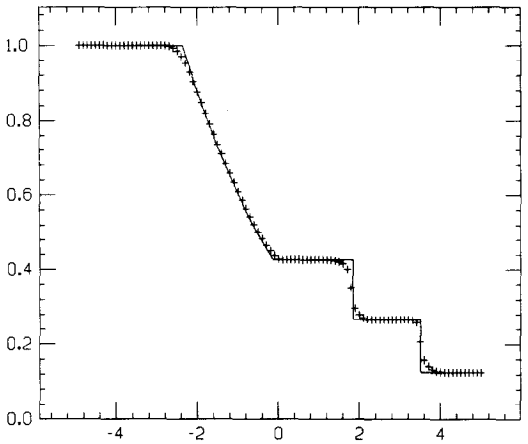

FIG. 13. Third-order, characteristicwise limiter (4.2a), density.

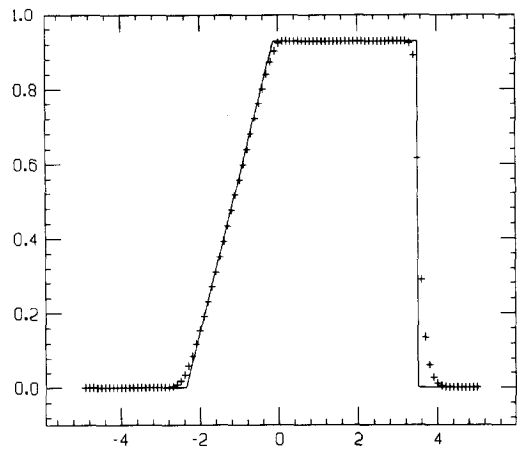

FIG. 14. Third-order, characteristicwise limiter (4.2a), velocity.

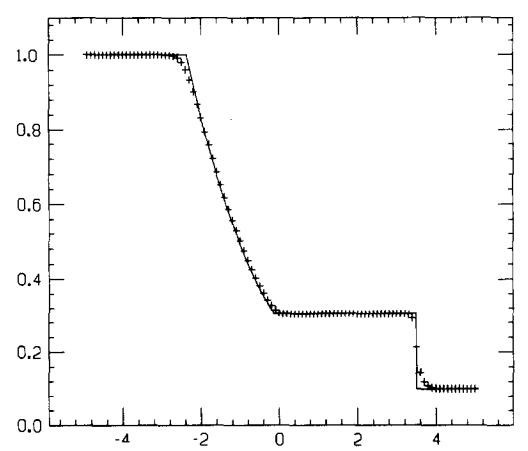

FIG. 15. Third-order, characteristicwise limiter (4.2a), pressure.

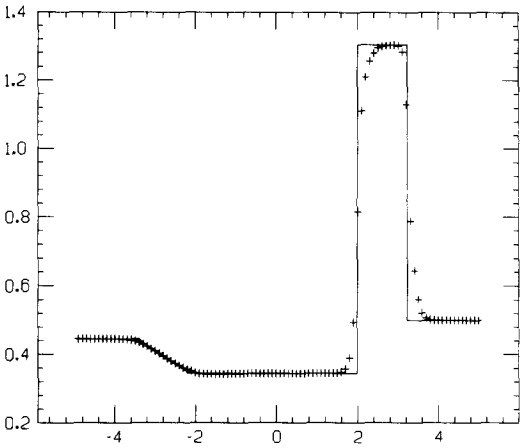

FIG. 16. Third-order, characteristicwise limiter (4.2b), density.

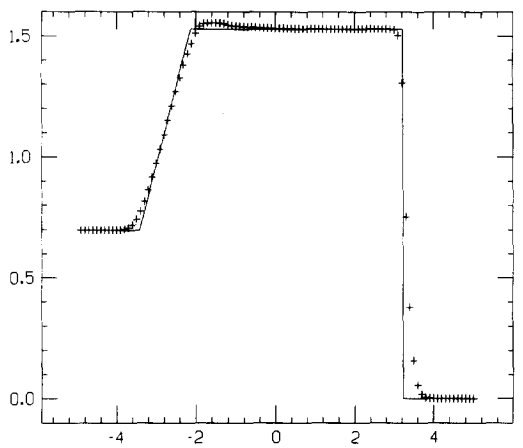

FIG. 17. Third-order, characteristicwise limiter (4.2b), velocity.

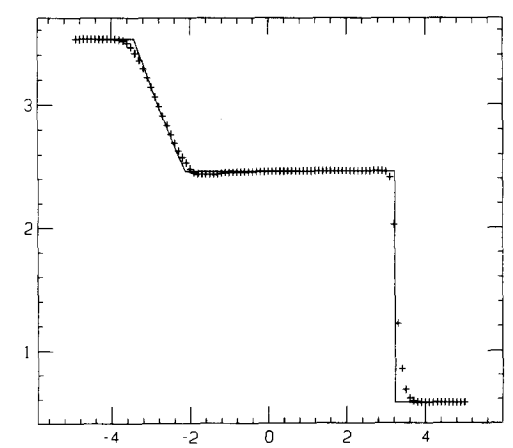

FIG. 18. Third-order, characteristicwise limiter (4.2b), pressure.

qualitatively similar. The results of characteristicwise limiters are in Figs. 7–18. As expected, we can see some wriggles in the componentwise version (and the wriggles become more severe with the third-order scheme) and good behaviors in the characteristicwise version. For Figs. 1–18 we use $M_2 = 0$ in (2.15b). Because of the simple structure of the solutions, we do not see the advantage of the third-order scheme over the second-order one. In fact we observe more smearing in the third-order case, due to a stronger limiting. Relaxing the limiting by taking $M_2 = 50$ improves the smearing at the price of slight over- and under-shoots. Compare Figs. 19 and 20 with 13 and 16.

We remark that the contact discontinuities and the corners of rarefaction waves are smeared more than the shocks. Some artificial compression or "sub-cell resolution" [6, 16] should help.

EXAMPLE 2. We consider the interaction of blast waves of the Euler equation (4.1) with

$$\mathbf{u}(x,0) = \begin{cases} \mathbf{u}_L, & 0 \leqslant x < 0.1, \\ \mathbf{u}_M, & 0.1 \leqslant x < 0.9, \\ \mathbf{u}_R, & 0.9 \leqslant x < 1, \end{cases} \tag{4.3}$$

where $\rho_L = \rho_M = \rho_R = 1$, $q_L = q_M = q_R = 0$, $p_L = 10^3$, $p_M = 10^{-2}$, $p_R = 10^2$. A reflecting boundary condition is applied to both ends. See [18, 5].

The results are in Figs. 21–24. The solid line is a converged solution taken from [16]. We see that the pictures are satisfactory, except for the above-mentioned smearing of contact discontinuities, which seems more serious for the third-order scheme. For this problem we have not noticed any significant difference for taking M_2 in (2.15b) from 0 to 300. During the collision any finite M_2 may underestimate the second derivative.

EXAMPLE 3. To demonstrate the advantage of higher order methods, we use the Euler equation (4.1) with initial condition

$$\begin{aligned} &(\rho_L, q_L, p_L) = (3.857143, 2.629369, 10.333333), && \text{when} \quad x < -4, \\ &(\rho_R, q_R, p_R) = (1 + 0.2\sin(5x), 0, 1), && \text{when} \quad x \geqslant -4. \end{aligned} \tag{4.4}$$

This example was used in [16]. It contains both shocks and fine structures in smooth regions—a simple model for shock-turbulence interactions. Our results are shown in Figs. 25–30. The solid lines are taken from [16] and can be regarded as a converged solution. We can see that the third-order TVD schemes (Fig. 25) and the TVB (Fig. 26) scheme with $M_2 = 30$ in (2.15b) give poor results with 400 cells, while the third-order TVB scheme with $M_2 = 300$ gives good results even with only 100 cells (Figs. 27–29), performing actually better than the third-order ENO schemes in [16]. Comparing Fig. 30 with Fig. 28, we can see that the third-order scheme performs much better in this case than the second-order one.

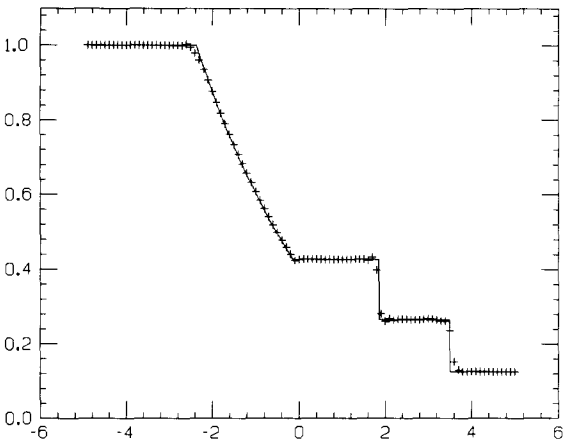

FIG. 19. Third-order, characteristicwise limiter (4.2a), density, $M_2 = 50$.

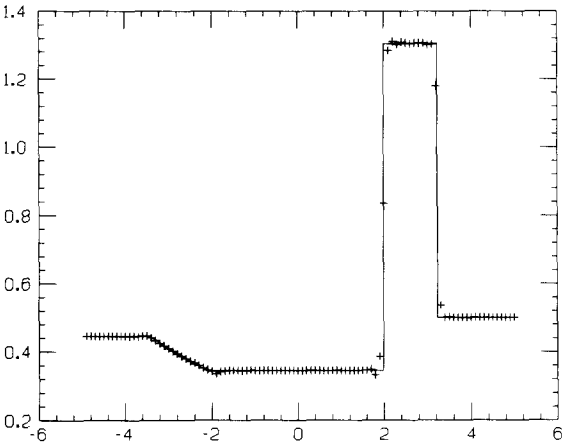

FIG. 20. Third-order, characteristicwise limiter (4.2b), density, $M_2 = 50$.

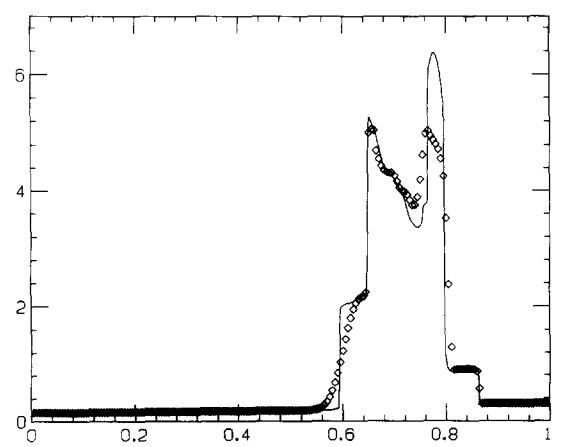

FIG. 21. Second-order, characteristicwise limiter, 200 points, density.

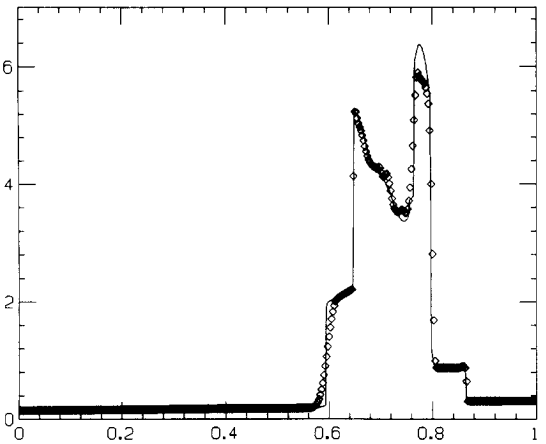

FIG. 22. Second-order, characteristicwise limiter, 400 points, density.

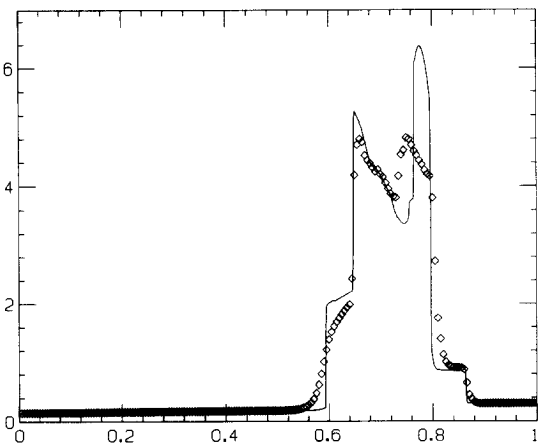

FIG. 23. Third-order, characteristicwise limiter, 200 points, density.

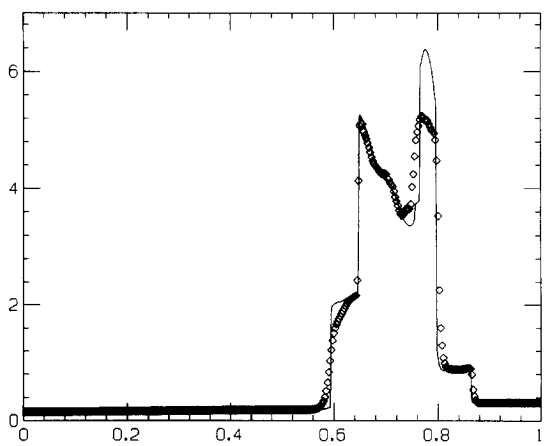

FIG. 24. Third-order, characteristicwise limiter, 400 points, density.

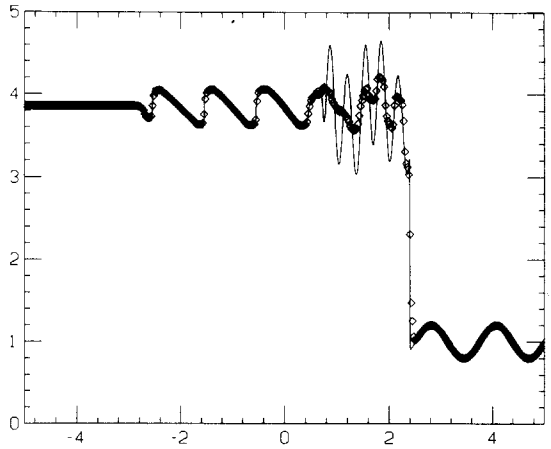

FIG. 25. Third-order TVD, (4.4), 400 points, density.

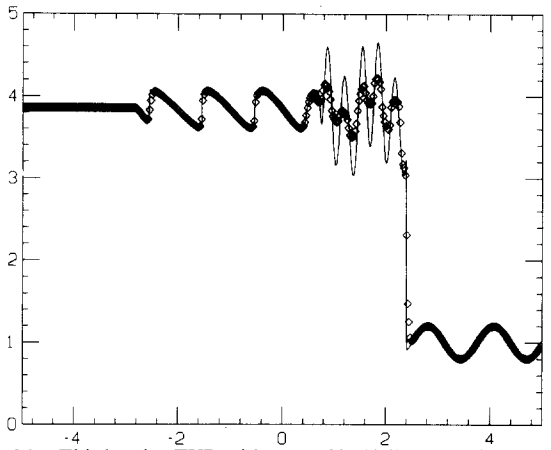

FIG. 26. Third-order TVB with $M_2 = 30$, (4.4), 400 points, density.

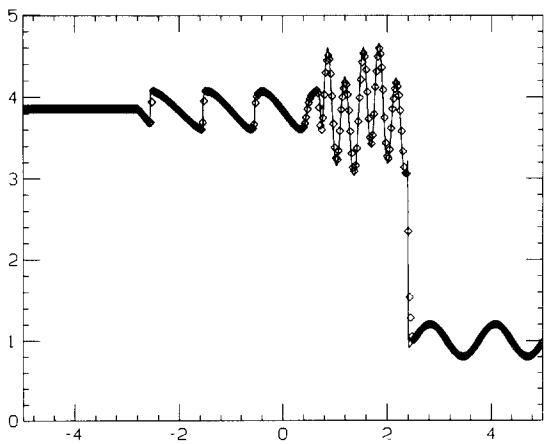

FIG. 27. Third-order TVB with $M_2 = 300$, (4.4), 400 points, density.

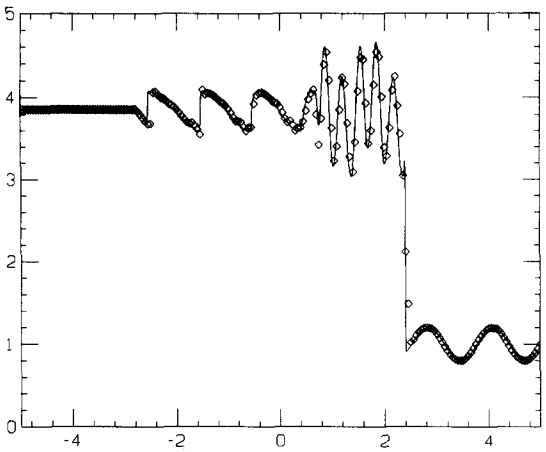

FIG. 28. Third-order TVB with $M_2 = 300$, (4.4), 200 points, density.

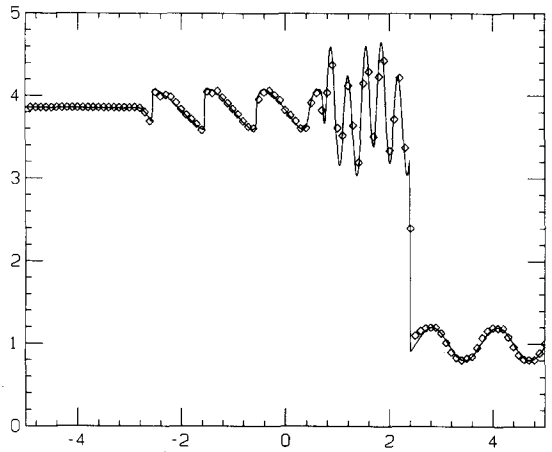

FIG. 29. Third-order TVB with $M_2 = 300$, (4.4), 100 points, density.

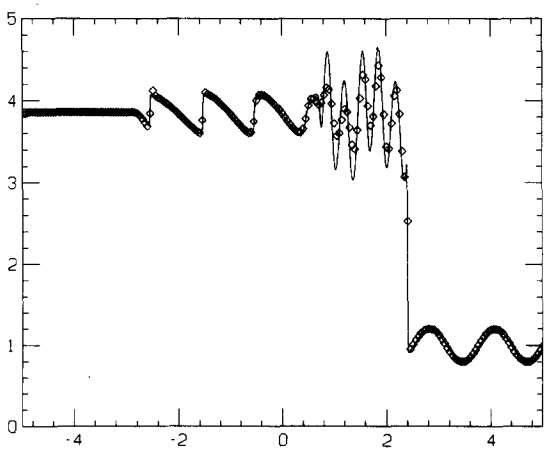

FIG. 30. Second-order TVB with $M_2 = 300$, (4.4), 200 points, density.

We point out here that unlike the scalar case [4] where M_2 can be estimated from the initial conditions, for nonlinear systems M_2 has to be ajusted according to the solutions. If M_2 is adjusted adequately, the higher order schemes can perform extremely well.

EXAMPLE 4. We use our second-order and third-order methods (2.8)–(2.10), with the boundary treatment (3.3), to solve the equation

$$\begin{pmatrix} u \\ v \end{pmatrix}_t = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}_x, \tag{4.5}$$

with the initial boundary conditions

$$\begin{aligned} &u(x, 0) = v(x, 0) = \sin 2\pi x, \qquad 0 \leqslant x \leqslant 1, \\ &u(0, t) = -v(0, t), \qquad v(1, t) = -u(1, t), \end{aligned} \tag{4.6a}$$

and

$$\begin{cases} v(x, 0) = \begin{cases} 1, & \text{if } \frac{1}{3} \leqslant x \leqslant \frac{2}{3}, \\ 0, & \text{otherwise,} \end{cases} \\ u(x, 0) \equiv 0, \qquad x \geqslant 0, \\ u(0, t) = v(0, t). \end{cases} \tag{4.6b}$$

The numerical errors at $t = 2$ for (4.5)–(4.6a) are listed in Table I. We can see that the boundary treatment works very well for smooth problems.

The solutions of (4.5)–(4.6b), at $t = 0.5$ and $t = 1.0$, are in Figs. 31–36. We can see that the boundary treatment is total variation stable. The smearing is again due to the fact that the equation is linear, and we expect improvements by artificial compressions or sub-cell corrections [6, 16].

Finally, let us point out that we only considered time dependent problems in this paper. We have not included the very important class of steady state problems,

TABLE I

		Second order				Third order			
	Δx	L_∞	r	L_1	r	L_∞	r	L_1	r
u	$\frac{1}{20}$	0.21×10^{-1}		0.13×10^{-1}		0.42×10^{-3}		0.17×10^{-3}	
	$\frac{1}{40}$	0.49×10^{-2}	2.08	0.32×10^{-2}	2.07	0.37×10^{-4}	3.51	0.17×10^{-4}	3.28
	$\frac{1}{80}$	0.12×10^{-2}	2.03	0.77×10^{-3}	2.03	0.38×10^{-5}	3.31	0.20×10^{-5}	3.15
v	$\frac{1}{20}$	0.21×10^{-1}		0.13×10^{-1}		0.29×10^{-3}		0.13×10^{-3}	
	$\frac{1}{40}$	0.49×10^{-2}	2.08	0.32×10^{-2}	2.07	0.35×10^{-4}	3.06	0.15×10^{-4}	3.14
	$\frac{1}{80}$	0.12×10^{-2}	2.03	0.77×10^{-3}	2.03	0.42×10^{-5}	3.07	0.18×10^{-5}	3.06

Note. L_∞: L_∞-error; L_1: L_1-error; r = numerical order of convergence.

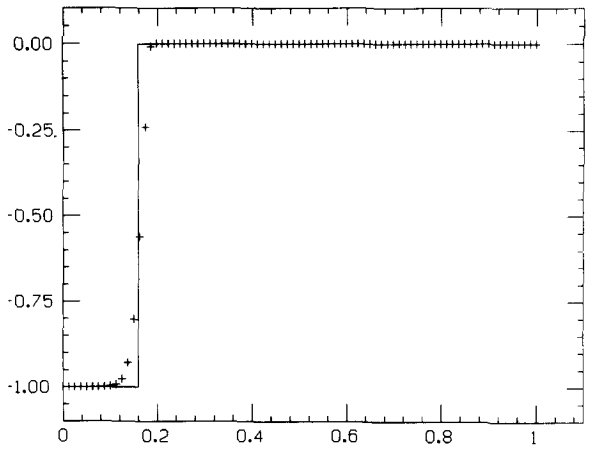

FIG. 31. Second order, $t = 0.5$, u.

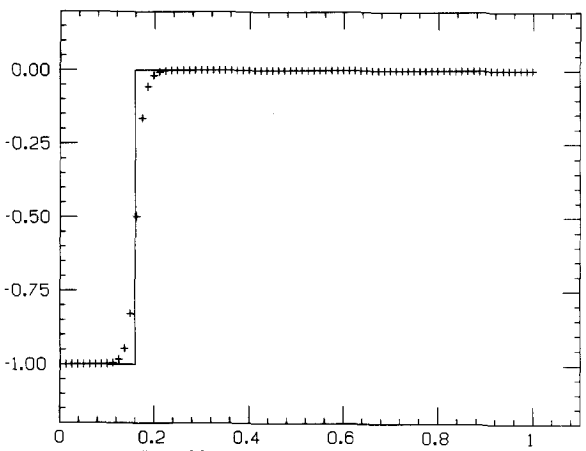

FIG. 32. Third order, $t = 0.5$, u.

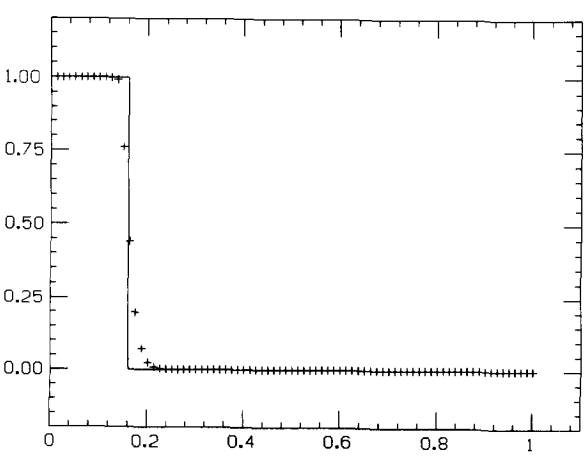

FIG. 33. Second order, $t = 0.5$, v.

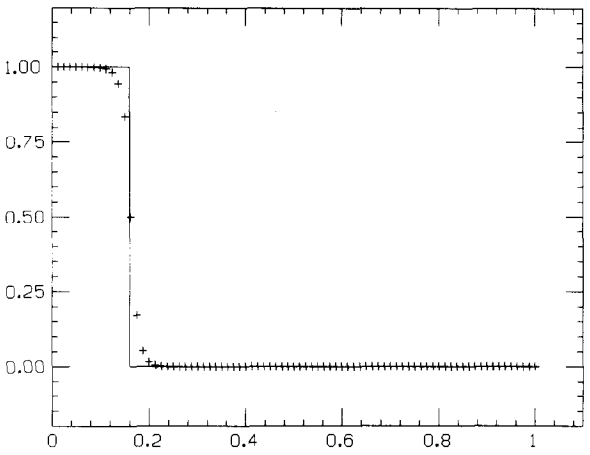

FIG. 34. Third order, $t = 0.5$, v.

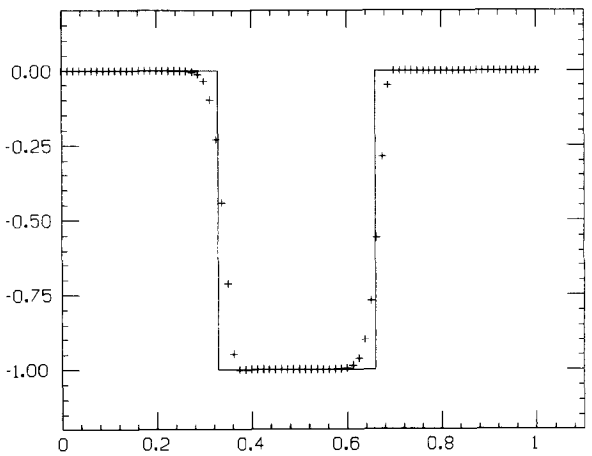

FIG. 35. Second order, $t = 1$, u.

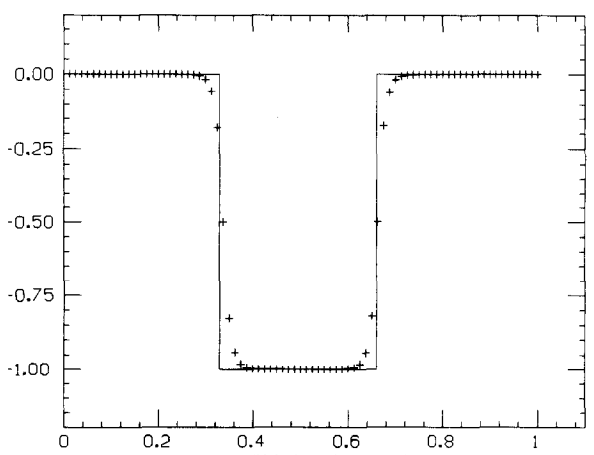

FIG. 36. Third order, $t = 1$, u.

mainly because we have used Runge–Kutta methods (2.9)–(2.10) which, with high temporal orders and relatively small *cfl* numbers, are more suitable for time dependent calculations.

5. Concluding Remarks

The extension of our earlier work about a class of TVB explicit discontinuous Galerkin methods [3, 4] to 1-dimensional systems seems to work very well. Theoretically, we have proven TVB for linear systems with either initial or initial-boundary conditions. Numerically, for the nonlinear Euler's equation of gas dynamics, we have obtained results which in most cases are comparable to the recent non-oscillatory finite difference methods [5, 16]. Two-dimensional computations using general triangulations constitute current research.

Acknowledgments

We thank the referees for some constructive comments on the first version of this paper. The work of Cockburn and Shu was initiated when both were in residence at the Institute of Mathematics and Its Applications, University of Minnesota. We thank IMA and its staff for support and hospitality.

References

1. G. Chavent and B. Cockburn, IMA Preprint Series No. 341, University of Minnesota, September 1987; M^2AN, in press.
2. G. Chavent and G. Salzano, *J. Comput. Phys.* **45**, 307 (1982).
3. B. Cockburn and C.-W. Shu, IMA Preprint Series No. 388, University of Minnesota, January 1988; in *Proceedings, First National Fluid Dynamics Congress, Cincinnati, August 1988*, in press.
4. B. Cockburn and C.-W. Shu, IMA Preprint Series No. 392, University of Minnesota, March 1988; *Math. Comput.* **52**, 411 (1989).
5. A. Harten, B. Engquist, S. Osher, and S. Chakravarthy, *J. Comput. Phys.* **71**, 231 (1987).
6. A. Harten, ICASE Report 87-56, NASA Langley Research Center, August 1987; *J. Comput. Phys.* **83**, 148 (1989).
7. T. Hughes and M. Mallet, *Finite Element Fluids* **6**, 339 (1985).
8. C. Johnson and J. Pitkaranta, *Math. Comput.* **46**, 1 (1986).
9. P. Lax, *Commun. Pure Appl. Math.* **7**, 159 (1954).
10. P. LeSaint and P. Raviart, in *Mathematical Aspects of Finite Element in Partial Differential Equations*, edited by C. DeBoor (Academic Press, New York, 1974), p. 89.
11. S. Osher, *SIAM J. Num. Anal.* **22**, 947 (1985).
12. P. Roe, *J. Comput. Phys.* **43**, 357 (1981).
13. C.-W. Shu, *Math. Comput.* **49**, 105 (1987).
14. C.-W. Shu, *Math. Comput.* **49**, 123 (1987).
15. C.-W. Shu and S. Osher, *J. Comput. Phys.* **77**, 439 (1988).
16. C.-W. Shu and S. Osher, ICASE Report 88-24, NASA Langley Research Center, April 1988; *J. Comput. Phys.* **83**, 32 (1989).
17. G. Sod, *J. Comput. Phys.* **27**, 1 (1978).
18. P. Woodward and P. Colella, *J. Comput. Phys.* **54**, 115 (1984).