SIAM J. SCI. STAT. COMPUT.
Vol. 9, No. 6, November 1988

TOTAL-VARIATION-DIMINISHING TIME DISCRETIZATIONS*

CHI-WANG SHU†

Abstract. In the computation of conservation laws $u_t + f(u)_x = 0$, total-variation-diminishing (TVD) schemes have been very successful. Many TVD schemes are of method-of-lines form (i.e., discretized in spatial variables only); hence time discretizations that keep TVD and have other properties (e.g., large CFL numbers for steady state calculations, or high-order accuracy for time-dependent problems) are desirable. In this paper we present a class of m-step Runge–Kutta-type TVD time discretizations with large CFL number m, suitable for steady state calculations, and a class of multilevel type TVD high-order time discretizations suitable for time-dependent problems. Some preliminary numerical results are also given.

Key words. conservation law, TVD scheme, time discretization

AMS(MOS) subject classifications. 65M05, 35L65, 65M10

1. Introduction.

Consider the hyperbolic conservation law

$$u_t + \sum_{i=1}^{d} f_i(u)_{x_i} = 0 \qquad (\text{or} = g(u, x, t), \text{ a source term}), \tag{1.1a}$$

$$u(x, 0) = u_0(x). \tag{1.1b}$$

Here $u = (u_1, \cdots, u_m)^T$, $x = (x_1, \cdots, x_d)$, and any real combination of the Jacobian matrices $\sum_{i=1}^{d} \xi_i (\partial f_i / \partial u)$ has m real eigenvalues and a complete set of eigenvectors.

On a computational grid $x_j = j \cdot \Delta x$, $t_n = n\Delta t$, we use u_j^n to denote the computed approximation to the exact solution $u(x_j, t_n)$ of (1.1).

As a guiding principle for scheme designing, usually rigorous analysis (stability, convergence) is only done for the scalar, one-dimensional nonlinear case ($d = m = 1$ in (1.1)). Although theory becomes extremely difficult for multidimensional systems, numerical experiments using direct field-by-field generalizations (e.g., [12]) of the scalar, one-dimensional TVD, TVB, ENO schemes usually give very good results (e.g., [3], [4], [6], [11]). This is because characteristic field decomposition is an effective approximate local decoupling of the system. In the following we shall confine our discussion to the scalar, one-dimensional case, unless otherwise stated.

As is well known, the solution to (1.1) may develop discontinuities (shocks, contact discontinuities, etc.) even if the initial condition $u_0(x)$ in (1.1b) is smooth. Traditional finite difference schemes, even if linearly stable, often give poor results when discontinuities are present. A very successful class of schemes for solving (1.1) is the class of total-variation-diminishing (TVD) conservative schemes, i.e., schemes

$$u_j^{n+1} = u_j^n - \lambda (\hat{f}_{j+1/2}^n - \hat{f}_{j-1/2}^n), \tag{1.2}$$

which are TVD:

$$TV(u^{n+1}) \leqq TV(u^n) \tag{1.3}$$

under the total-variation definition

$$TV(u) = \sum_j |u_{j+1} - u_j|. \tag{1.4}$$

* Received by the editors June 1, 1987; accepted for publication (in revised form) February 16, 1988.
† Institute for Mathematics and Its Applications, University of Minnesota, Minneapolis, Minnesota 55455. Present address, Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912.

In (1.2), the numerical flux $\hat{f}_{j+1/2}$ is defined by

$$\hat{f}_{j+1/2}=\hat{f}(u_{j-k+1},\cdots,u_{j+k}), \tag{1.5}$$

which is Lipschitz continuous in all its arguments, and satisfies the consistency condition

$$\hat{f}(u,\cdots,u)=f(u). \tag{1.6}$$

We can also consider TVB (total-variation-bounded) schemes, that satisfy

$$TV(u^n)\leqq B \tag{1.7}$$

for some fixed $B>0$ and all possible n and Δt such that $n\Delta t\leqq T$.

TVD or TVB schemes have the advantage of high-order accuracy in smooth regions, while resolving discontinuities without spurious oscillations. Another major advantage of TVD or TVB schemes is that there is a convergent (in L_1^{local}) subsequence as $\Delta x\to 0$ to a weak solution of (1.1). If an additional entropy condition, which implies uniqueness of weak solution to (1.1), is satisfied, then the scheme is convergent. For details, see, e.g., [3], [4], [10], [11], [13].

TVD schemes do have a disadvantage of local degeneracy to first-order accuracy at nonsonic critical points, (see, e.g., [10]). Now we already have remedies (TVB modifications) to overcome this difficulty [13].

The more recently developed essentially nonoscillatory (ENO) schemes [5], [6], [16], share many advantages with and usually perform better than TVD or TVB schemes, because they use an adaptive stencil trying to obtain information from the smoothest regions. Although it is still not proven rigorously that ENO schemes are TVB (however, see [15]), numerically ENO schemes are extremely stable.

We will use the standard notation

$$\Delta_+ u_j=u_{j+1}-u_j;\qquad \Delta_- u_j=u_j-u_{j-1}.$$

A semidiscrete (method of lines) scheme to (1.1) of conservation form is a system of ordinary differential equations (ODEs)

$$\frac{\partial}{\partial t}u_j=-\frac{1}{\Delta x}(\hat{f}_{j+1/2}-\hat{f}_{j-1/2}) \tag{1.8}$$

where the numerical flux $\hat{f}_{j+1/2}$ is defined by (1.5)-(1.6).

We also consider Euler forward time discretization (1.2) of (1.8), $\lambda=\Delta t/\Delta x$. $\lambda \max|f'(u)|$ is called the CFL number.

Method-of-lines is a common practice in solving time-dependent partial differential equations (PDEs). Usually method-of-lines schemes are much simpler than fully discrete schemes. See, e.g., [11], [13], [16] in which TVD, TVB, and ENO method-of-lines schemes are given with very high spatial orders. It remains to discretize in time. A common practice is to use some ODE solver (Runge–Kutta or multilevel); however, it seems that only linear stability analysis is available in the literature. For our purposes we hope to keep the TVD property because linear stability is not enough for convergence in the presence of discontinuities, and a stronger nonlinear stability such as TVD is needed. In § 2 of this paper we consider steady state calculations in which time accuracy is not important, but a large CFL number is desirable. We present a class of Runge–Kutta m-step TVD time discretizations that has CFL number m. In § 3 we consider time-dependent problems and present a class of high-order multilevel type TVD time discretizations. In § 4 we include some preliminary numerical results using the method in § 2.

2. Steady state calculations—a class of Runge–Kutta type TVD schemes with large CFL numbers. For steady state calculations (i.e., to solve for a time-independent

solution of (1.1)), we do not care about time accuracy; hence the simple Euler forward version (1.2) of (1.8) can be used and will give high-order spatial accuracy in steady states. Unfortunately, the CFL restriction

$$\lambda \max |f'(u)| \leqq \lambda_0 \tag{2.1}$$

is usually severe, e.g., for the third-order β-scheme in [11], $\lambda_0 = \frac{2}{5}$. This makes the convergence to steady states very slow and costly. In practice people use Runge-Kutta type schemes with a larger CFL number. A special class of such schemes has been used by Jameson [7] and Turkel [17], and analyzed for linear stability under CFL restriction less than m for m-step methods (e.g., for a 4-step method the CFL number is 2.8). Motivated by their work, we use different parameters and different ways of freezing nonlinear parts, and prove the resulting schemes are TVD under CFL number m for m-step methods.

Many TVD schemes (1.2) (e.g., all the schemes in [11]) can be written in the following form:

$$u_j^{n+1} = u_j^n - \lambda(-C_{j+1/2}\Delta_+ u_j^n + D_{j-1/2}\Delta_- u_j^n) \tag{2.2}$$

where

$$C_{j+1/2} \geqq 0, \tag{2.3a}$$

$$D_{j+1/2} \geqq 0, \tag{2.3b}$$

$$\lambda(C_{j+1/2} + D_{j+1/2}) \leqq 1. \tag{2.3c}$$

Here $C_{j+1/2}$ and $D_{j+1/2}$ are complicated nonlinear functions of $u_{j-k+1}, \cdots, u_{j+k}$.

If we decompose $f(u)$ into $f(u) = f^+(u) + f^-(u)$, where $(f^+)' \geqq 0$, $(f^-)' \leqq 0$, then C and D in (2.2) can be considered as approximations to $(f^\pm)'$. Hence the decomposition in (2.2) (upwind-downwind decomposition) is very natural. For example, the ENO schemes in [16] can be easily written in the form (2.2).

Now we define our Runge-Kutta-type scheme as follows:

$$u_j^{(k)} = u_j^{(0)} - \alpha_k \lambda[-C_{j+1/2}^{(0)}\Delta_+ u_j^{(k-1)} + D_{j-1/2}^{(0)}\Delta_- u_j^{(k-1)}] \tag{2.4}$$

for $k = 1, 2, \cdots, m$, with

$$u_j^{(0)} = u_j^n; \qquad u_j^{(m)} = u_j^{n+1}. \tag{2.5}$$

Notice that the scheme (2.4)-(2.5) is of the same form as the ones in [7] and [17], except the way of freezing nonlinear parts.

Since we only need first-order accuracy in time for steady state calculations, which is guaranteed by $\alpha_m = 1$, we have complete freedom in choosing $\alpha_1, \cdots, \alpha_{m-1}$ to optimize CFL restrictions. Based on linear stability considerations, a scheme in [7] uses, for $m = 4$, $\alpha_i = 1/(5-i)$, $i = 1, 2, 3, 4$. We choose

$$\alpha_i = \frac{i}{m(m-i+1)}, \qquad i = 1, 2, \cdots, m \tag{2.6}$$

and have the following proposition.

PROPOSITION 2.1. *The scheme* (2.4)-(2.6) *is* TVD *under the* CFL *condition*

$$\lambda(C_{j+1/2}^{(0)} + D_{j+1/2}^{(0)}) \leqq m \tag{2.7}$$

if the building block (2.2)-(2.3) *is* TVD *under the* CFL *condition* (2.3c).

Proof. Let the operator $L^{(0)}$ be defined as

$$L^{(0)} = \lambda C_{j+1/2}^{(0)} T_+ + \lambda D_{j-1/2}^{(0)} T_- - \lambda(C_{j+1/2}^{(0)} + D_{j-1/2}^{(0)}) \tag{2.8a}$$

where T_+ and T_- are the shift operators

$$T_+ u_j = u_{j+1}, \qquad T_- u_j = u_{j-1}, \tag{2.8b}$$

i.e.,

$$\begin{aligned}(L^{(0)}(u))_j &= \lambda C^{(0)}_{j+1/2} u_{j+1} + \lambda D^{(0)}_{j-1/2} u_{j-1} - \lambda (C^{(0)}_{j+1/2} + D^{(0)}_{j-1/2}) u_j \\ &= \lambda C^{(0)}_{j+1/2} \Delta_+ u_j - \lambda D^{(0)}_{j-1/2} \Delta_- u_j.\end{aligned}$$

Notice that $L^{(0)}$ is a linear operator.

Then (2.4) can be rewritten as

$$u^{(k)} = u^{(0)} + \alpha_k L^{(0)}(u^{(k-1)}), \qquad k = 1, \cdots, m. \tag{2.9}$$

We can easily prove by induction that

$$u^{(k)} = \left(\sum_{j=0}^{k} \left(\prod_{i=k-j+1}^{k} \alpha_i \right) (L^{(0)})^j \right) u^{(0)}, \qquad k = 1, \cdots, m. \tag{2.10}$$

Hence, by the definition of α_i in (2.6), we have

$$\begin{aligned}u^{n+1} = u^{(m)} &= \left[\sum_{j=0}^{m} \left(\prod_{i=m-j+1}^{m} \alpha_i \right) (L^{(0)})^j \right] u^{(0)} \\ &= \left[\sum_{j=0}^{m} \left(\prod_{i=m-j+1}^{m} \frac{i}{m(m-i+1)} \right) (L^{(0)})^j \right] u^{(0)} \\ &= \left[\sum_{j=0}^{m} \binom{m}{j} \frac{1}{m^j} (L^{(0)})^j \right] u^{(0)} \\ &= \left(I + \frac{L^{(0)}}{m} \right)^m u^{(0)}.\end{aligned}$$

Now, under the CFL condition (2.7), $I + L^{(0)}/m$ is a TVD operator, and so is $(I + L^{(0)}/m)^m$. This finishes the proof. □

Remark 2.1. If we use upwind differencing in (2.2), then the CFL condition (2.7) is optimal in the sense that the domain of dependency of the scheme already coincides with that of the differential equation.

Remark 2.2. Since many method-of-lines TVD, TVB, and ENO schemes (e.g., in [11], [13], [16]) and their Euler forward versions can be written in the form (2.2), scheme (2.4)-(2.6) is widely applicable.

Remark 2.3. Notice that in each cycle, the costly nonlinear terms Cs and Ds are only computed once at the beginning and then frozen. This greatly reduces the computational cost. In practical computations, m can be chosen as a function of the residue $\|u^n - u^{n-1}\|$, and can vary from level to level. At the beginning, when the residue $\|u^n - u^{n-1}\|$ is large, we use a smaller m; near steady state, when $\|u^n - u^{n-1}\|$ is small, we may use a larger m. This will lead to faster convergence to steady state solutions, and also avoid sticking to limit cycles or converging to physically unstable steady solutions, according to our preliminary computational experience with scalar problems. An example illustrating this is a boundary value problem for a Burgers equation with forcing term [1]. Two steady state solutions (both satisfy the entropy condition) exist; one of them is physically unstable. It was illustrated in [1] that implicit methods with large CFL numbers may converge to the wrong solution. We have tried our scheme starting from the wrong solution (the worst case). For fixed, big m ($m \geqq 6$) it converges to the wrong solution. However for variable m depending on the residue, it converges very fast to the correct solution.

Remark 2.4. For nonlinear systems ($m>1$ in (1.1)), C and D in (2.2) are matrices obtained via field-by-field decompositions. We do not carry out the details here but indicate that the generalization from scalar, one-dimensional scheme (2.4)-(2.6) to multidimensional systems is very natural and straightforward. Work is currently under way with colleagues to apply scheme (2.4)-(2.6) to two-dimensional Euler equations in gas dynamics under general curvilinear coordinates. Since within each cycle only matrix-vector multiplications are involved, the savings in cost mentioned in Remark 2.3 is even more significant for systems, especially on vector machines.

Remark 2.5. For preliminary numerical results (scalar case only) see § 4.

3. Time-dependent problems—a class of TVD multilevel type high-order time discretizations. For time-dependent problems, we need time accuracy as well. One way is still to use a high-order ODE solver (Runge-Kutta type or multilevel type); another way is via the Lax-Wendroff procedure, i.e., by using $u_t=-f_x$, $u_{tt}=(f'f_x)_x$, $\cdots$, $u_j^{n+1}=u_j^n+\Delta t(u_t)_j^n+\Delta t^2/2(u_{tt})_j^n+\cdots$, then discretizing the spatial derivatives. In the former case, again, only linear stability analysis is available in the literature, and for the latter case TVD is proven only for the second-order case (e.g., [3]). We now present a class of multilevel-type high-order time discretizations and prove they are TVD. Besides simplicity they also have the following two important advantages: they are easily generalizable to equations with forcing terms, and easily generalizable to multi-dimensional problems.

Let

$$L^{(1)}(u)_j=-\lambda(\hat{f}_{j+1/2}-\hat{f}_{j-1/2}) \tag{3.1}$$

and write scheme (1.2) as

$$u_j^{n+1}=u_j^n+L^{(1)}(u)_j. \tag{3.2}$$

Assume the method-of-lines scheme is rth order in space, i.e.,

$$L^{(1)}(u)=\Delta t(-f(u))_x+O(\Delta x^{r+1}). \tag{3.3}$$

By using the TVD schemes in [11] and the TVB modifications in [13], we can have schemes for r up to 15 in (3.3). We can also use ENO schemes in [16].

For simplicity we only consider TVD schemes (3.2). All the following results are valid if (3.2) is TVB (just change TVD to TVB and make minor adjustments in the proof).

Assume (3.2)-(3.3) is TVD for some CFL condition

$$\lambda\leqq\lambda_0. \tag{3.4}$$

Also, assume

$$u_j^{n+1}=u_j^n+L^{(-1)}(u)_j, \tag{3.5}$$

$$L^{(-1)}(u)=\Delta t(f(u))_x+O(\Delta x^{r+1}) \tag{3.6}$$

is TVD under CFL condition (3.4). (This is just the same TVD scheme applied to $u_t=(f(u))_x$ instead of to the original equation (1.1). Here $L^{(-1)}$ does not mean the inverse of $L^{(1)}$, it is just a handy notation. See Remark 3.2 below.)

We construct our time discretization as follows:

$$u_j^{n+1}=\sum_{k=0}^{m}[\alpha_k u_j^{n-k}+\beta_k(\text{sign}\,(\beta_k)L^{(\text{sign}(\beta_k))}(u^{n-k})_j)] \tag{3.7}$$

with

$$\alpha_k\geqq 0,\qquad k=0,1,\cdots,m \tag{3.8}$$

and $\alpha_k=0$ only if $\beta_k=0$.

PROPOSITION 3.1. *Scheme* (3.1)-(3.8) *is*

(a) TVD *if*

$$\sum_{k=0}^{m} \alpha_k = 1 \tag{3.9}$$

under the CFL *condition*

$$\lambda \leqq \lambda_0 \min_k \left(\frac{\alpha_k}{|\beta_k|} \right); \tag{3.10}$$

(b) *s-order accurate in time and space* $(1 \leqq s \leqq r)$ *if* (3.9) *and*

$$\beta_0 - \sum_{k=1}^{m} (k\alpha_k - \beta_k) = 1, \tag{3.11a}$$

$$(-1)^l \sum_{k=1}^{m} k^{l-1}(k\alpha_k - l\beta_k) = 1, \qquad l = 2, 3, \cdots, s \tag{3.11b}$$

are satisfied.

Proof. (a) Rewrite scheme (3.7) as

$$u_j^{n+1} = \sum_{k=0}^{m} \alpha_k u_j^{(k)} \tag{3.12}$$

with $u_j^{(k)}$ defined by

$$u_j^{(k)} = u_j^{n-k} + \frac{|\beta_k|}{\alpha_k} L^{(\operatorname{sign}(\beta_k))}(u^{n-k})_j. \tag{3.13}$$

Since (3.2) and (3.5) are both TVD under CFL restriction (3.4), (3.13) is TVD under CFL restriction (3.10), i.e.,

$$TV(u^{(k)}) \leqq TV(u^{n-k}). \tag{3.14}$$

Convexity of (3.12) now easily leads to

$$\begin{aligned} TV(u^{n+1}) &\leqq \sum_{k=0}^{m} \alpha_k TV(u^{(k)}) \\ &\leqq \sum_{k=0}^{m} \alpha_k TV(u^{n-k}) \leqq \max_{0 \leqq k \leqq m} TV(u^{n-k}). \end{aligned} \tag{3.15}$$

This leads to

$$TV(u^n) \leqq \max_{0 \leqq k \leqq m} TV(u^k). \tag{3.16}$$

Hence part (a) is proved.

(b) The proof of part (b) follows easily from Taylor expansions (see, e.g., [8]). □

COROLLARY 3.1. *For any rth order in space* TVD *scheme* (3.2)-(3.5), *there exist globally rth-order accurate* (*in time and space*) TVD *schemes of the form* (3.7) *with positive* CFL *numbers* (3.10).

Proof. Take $s = m + 1 = r$ in (3.11) and rewrite it as

$$\beta_0 + \sum_{k=1}^{r-1} \beta_k = 1 + \sum_{k=1}^{r-1} k\alpha_k, \tag{3.17a}$$

$$\sum_{k=1}^{r-1} k^{l-1}\beta_k = \frac{1}{l}\left[(-1)^{l-1} + \sum_{k=1}^{r-1} k^l \alpha_k \right], \qquad l = 2, 3, \cdots, r. \tag{3.17b}$$

This is an $r \times r$ linear system for $\beta_0, \beta_1, \cdots, \beta_{r-1}$ and the coefficient matrix is nonsingular (its determinant is $(r-1)!\, V(1, 2, \cdots, r-1) = (r-1)! \prod_{1 \leqq j < i \leqq r-1} (i-j) \neq 0$,

where $V(1, 2, \cdots, r-1)$ is the Vandermonde determinant). So for any given $\alpha_1, \cdots, \alpha_{r-1}$ there is a unique solution for $(\beta_0, \beta_1, \cdots, \beta_{r-1})$. If we pick any positive $\alpha_0, \cdots, \alpha_{r-1}$ satisfying (3.9), e.g., $\alpha_0 = \cdots = \alpha_{r-1} = 1/r$, we will get $(\beta_0, \beta_1, \cdots, \beta_{r-1})$ from (3.17), and the scheme (3.7) is rth order, TVD, with CFL number (3.10) greater than zero. (The only possibility for CFL number to be zero is when $\alpha_k = 0$ and the corresponding $\beta_k \neq 0$.) □

Remark 3.1. Corollary 3.1 provides a relatively easy way to find an rth-order TVD time discretization: we just need to solve an $r \times r$ linear system to get the coefficients for the scheme. Usually we choose $\alpha_0, \cdots, \alpha_{r-1}$ to get an optimal CFL number in (3.10). The approach in the corollary is of course not unique. We may put some of the α's into the unknowns of (3.17) to reduce the number of levels needed; or we may use more levels, trying to get a better CFL number; or we may require β's to be also nonnegative in order to avoid the usage of $L^{(-1)}$ (this will reduce the work of each timestep by one half).

Remark 3.2. The use of $L^{(-1)}$ in (3.6) is important. Without it, for the negative β's in (3.7), we are actually solving a nonlinear hyperbolic equation (1.1) backward, which is a well-known disaster at the differential equation level. (If the scheme is upwind, not using $L^{(-1)}$ would be equivalent to using a downwind scheme for negative β's, which is unstable in any sense.)

Remark 3.3. Jointly with Stanley Osher, we presented a class of high-order Runge–Kutta-type TVD time discretizations in [16]. Runge–Kutta-type methods have the advantage of being self-starting and requiring smaller storage. However for very high-order methods (order $\geqq 5$) multilevel methods may be simpler (see Table 1 and Corollary 3.1).

Some of the schemes (3.7) are listed in Table 1. The list is by no means exhaustive, nor are the schemes listed the optimal of the corresponding order. But from the list we can already see that we can get reasonable CFL numbers for third- and fourth-order schemes, and we can increase the CFL number by going to more levels.

In [13], TVB method-of-lines schemes were implemented by the TVD high-order time discretizations (3.7). We refer the reader to [13] for numerical results of these schemes.

TABLE 1
Coefficients for some of the high-order TVD *time discretizations* (3.7).

m	Order	CFL#	α_i	β_i
1	2	0.5	$\frac{4}{5}, \frac{1}{5}$	$\frac{8}{5}, \frac{-2}{5}$
2	2	0.5	$\frac{3}{4}, 0, \frac{1}{4}$	$\frac{3}{2}, 0, 0$
3	2	0.67	$\frac{8}{9}, 0, 0, \frac{1}{9}$	$\frac{4}{3}, 0, 0, 0$
3	2	0.75	$\frac{16}{17}, 0, 0, \frac{1}{17}$	$\frac{64}{51}, 0, 0, -\frac{4}{51}$
2	3	0.27	$\frac{4}{7}, \frac{2}{7}, \frac{1}{7}$	$\frac{25}{12}, -\frac{20}{21}, \frac{37}{84}$
3	3	0.33	$\frac{16}{27}, 0, 0, \frac{11}{27}$	$\frac{16}{9}, 0, 0, \frac{44}{27}$
4	3	0.5	$\frac{25}{32}, 0, 0, 0, \frac{7}{32}$	$\frac{25}{16}, 0, 0, 0, \frac{5}{16}$
5	3	0.57	$\frac{108}{125}, 0, 0, 0, 0, \frac{17}{125}$	$\frac{36}{25}, 0, 0, 0, 0, \frac{6}{25}$
3	4	0.15	$\frac{29}{72}, \frac{7}{24}, \frac{1}{4}, \frac{1}{18}$	$\frac{481}{192}, -\frac{1055}{576}, \frac{937}{576}, -\frac{197}{576}$
5	4	0.245	$\frac{747}{1280}, 0, 0, 0, \frac{81}{256}, \frac{1}{10}$	$\frac{237}{128}, 0, 0, 0, \frac{165}{128}, -\frac{3}{8}$
4	5	0.077	$\frac{1}{4}, \frac{1}{4}, \frac{7}{24}, \frac{1}{6}, \frac{1}{24}$	$\frac{185}{64}, -\frac{851}{288}, \frac{91}{24}, -\frac{151}{96}, \frac{199}{576}$
5	5	0.1298	$\frac{7}{20}, \frac{3}{10}, \frac{4}{15}, 0, \frac{7}{120}, \frac{1}{40}$	$\frac{291201}{108000}, -\frac{198401}{86400}, \frac{88063}{43200}, 0, -\frac{17969}{43200}, \frac{73061}{432000}$

Besides simplicity, we should also point out two other important advantages of the scheme (3.7). One is that the scheme as well as the theory can be easily generalized to (1.1) with a forcing term $g(u, x, t)$. Notice that if the forcing term g depends on u, then it is very messy to generalize the Lax-Wendroff type schemes for (1.1), and usually the TVD or TVB properties of the original scheme may not be preserved. But for scheme (3.7), we do not have such difficulties.

Let

$$u_j^{n+1} = u_j^n + \tilde{L}^{(\pm 1)}(u)_j^n \tag{3.18}$$

with

$$\tilde{L}^{(\pm 1)}(u)_j^n = L^{(\pm 1)}(u)_j^n \pm \Delta t g(u_j^n, x_j, t^n), \tag{3.19}$$

where $L^{(\pm 1)}(u)$ satisfy (3.3) and (3.6).

With mild restrictions on g (e.g., Lipschitz continuity in u and x), we can easily prove (3.18) is TVB if (3.2) and (3.5) are both TVD (or TVB).

The generalization of scheme (3.7) is just to replace L by $\tilde{L}$:

$$u_j^{n+1} = \sum_{k=0}^{m} (\alpha_k u_j^{n-k} + \beta_k (\operatorname{sign}(\beta_k) \tilde{L}^{(\operatorname{sign}(\beta_k))}(u^{n-k})_j)). \tag{3.20}$$

Repeating the proof of Proposition 3.1, we get

PROPOSITION 3.2. *Scheme* (3.20), *approximating* (1.1) *with forcing term g, is s-order accurate in time and space* $(1 \leqq s \leqq r)$ *and* TVB, *under the conditions* (3.1)-(3.6), (3.8)-(3.11), *and* (3.18)-(3.19). □

Another important advantage of scheme (3.7) is that it can be easily generalized to multidimensional problems. ($d > 1$ in (1.1).) For simplicity of notation only, we assume Δx is the same for all x_l directions. In each x_l direction we apply (3.3) and (3.6):

$$\pm L_{x_l}^{(\pm 1)}(u) = \Delta t(-f_l(u))_{x_l} + O(\Delta x^{r+1}). \tag{3.21}$$

Also, denote $J = (j_1, \cdots, j_d)$. The d-dimensional version of scheme (3.7) is

$$u_J^{n+1} = \sum_{k=0}^{m} \left(\alpha_k u_J^{n-k} + \beta_k \left(\operatorname{sign}(\beta_k) \sum_{l=1}^{d} L_{x_l}^{(\operatorname{sign}(\beta_k))}(u^{n-k})_J \right) \right). \tag{3.22}$$

By (3.21), we have

$$\pm \sum_{l=1}^{d} L_{x_l}^{(\pm 1)}(u) = \Delta t \sum_{l=1}^{d} (-f_l(u))_{x_l} + O(\Delta x^{r+1}) = \Delta t \cdot u_t + O(\Delta t^{r+1}) \tag{3.23}$$

if u is an exact solution of (1.1). Hence, we still have Proposition 3.3.

PROPOSITION 3.3. *Scheme* (3.22) *is s-order accurate in time and space* $(1 \leqq s \leqq r)$ *if* (3.9) *and* (3.11) *are satisfied.* □

Thus all the schemes in Table 1 can be used in multidimensional calculations.

What about the stability of the scheme (3.22)? We can at least have a maximum principle for u_J^n in (3.22) if the CFL number in (3.10) is divided by $2d$.

PROPOSITION 3.4. *The solution* u_J^n *of the scheme* (3.22) *is uniformly bounded under the* CFL *restriction*

$$\lambda \leqq \frac{1}{2d} \lambda_0 \min_k \left(\frac{\alpha_k}{|\beta_k|} \right). \tag{3.24}$$

Proof. Write (3.22) as

$$u_J^{n+1} = \sum_{k=0}^{m} \sum_{l=1}^{d} \frac{\alpha_k}{d} u_J^{(k,l)} \tag{3.25}$$

with $u_J^{(k,l)}$ defined by

$$u_J^{(k,l)} = u_J^{n-k} + \frac{d|\beta_k|}{\alpha_k} L_{x_l}^{(\mathrm{sign}(\beta_k))}(u^{n-k})_J. \tag{3.26}$$

We can also write it as (see (2.2)–(2.3)):

$$u_J^{(k,l)} = u_J^{n-k} + \frac{d|\beta_k|\lambda}{\alpha_k\lambda_0}(\lambda_0(-C_{j_l+1/2}^{(k,l)}\Delta_+^{(x_{j_l})}u_J^{n-k} + D_{j_l-1/2}^{(k,l)}\Delta_-^{(x_{j_l})}u_J^{n-k}) \tag{3.27}$$

with

$$C_{j_l+1/2}^{(k,l)} \geqq 0, \qquad D_{j_l+1/2}^{(k,l)} \geqq 0, \tag{3.28a}$$

$$\lambda_0(C_{j_l+1/2}^{(k,l)} + D_{j_l+1/2}^{(k,l)}) \leqq 1. \tag{3.28b}$$

Under the CFL condition (3.24), we have

$$\frac{d|\beta_k|\lambda}{\alpha_k\lambda_0} \leqq \frac{1}{2}. \tag{3.29}$$

Hence

$$1 - \frac{d|\beta_k|\lambda}{\alpha_k\lambda_0}(\lambda_0 C_{j_l+1/2}^{(k,l)} + \lambda_0 D_{j_l-1/2}^{(k,l)}) \geqq 1 - \frac{1}{2}(1+1) = 0.$$

So, if we denote $1_{j_l} = (0, \cdots, 0, 1, 0, \cdots, 0)$, where the 1 is at the j_lth position, we have

$$\begin{aligned} |u_J^{(k,l)}| &\leqq \left(1 - \frac{d|\beta_k|\lambda}{\alpha_k\lambda_0}(\lambda_0 C_{j_l+1/2}^{(k,l)} + \lambda_0 D_{j_l-1/2}^{(k,l)})\right)|u_J^{n-k}| \\ &\quad + \frac{d|\beta_k|\lambda}{\alpha_k\lambda_0}(\lambda_0 C_{j_l+1/2}^{(k,l)}|u_{J+1_{j_l}}^{n-k}| + \lambda_0 D_{j_l-1/2}^{(k,l)}|u_{J-1_{j_l}}^{n-k}|) \\ &\leqq \max_J |u_J^{n-k}|. \end{aligned}$$

Plugging this into (3.25) and noticing that

$$\sum_{k=0}^{m}\sum_{l=1}^{d}\frac{\alpha_k}{d} = 1$$

we get

$$\max_J |u_J^{n+1}| \leqq \max_{0\leqq k\leqq m}\max_J |u_J^{n-k}|.$$

This trivially implies

$$\max_J |u_J^n| \leqq \max_{0\leqq k\leqq m}\max_J |u_J^n|.$$

Proposition 3.4 is thus proved. □

Remark 3.4. It is possible to double the CFL number (3.24) by using the conservation form of the scheme (1.2) (see, e.g., [4], [14]). The factor $1/d$ in (3.24) is due to the definition of λ_0 in (3.28b). If a one-dimensional CFL restriction is $\lambda \max |f'(u)| \leqq 1$, then for d-dimensional this says $\lambda \max |f_l'(u)| \leqq 1/d$ for each l, or equivalently, $\lambda \sum_{l=1}^{d} \max |f_l'(u)| \leqq 1$, which is natural.

Proposition 3.4 gives a stability result of the scheme (3.22) in terms of the maximum norm. This is certainly no worse than any linear stability results for problems with discontinuities. Maximum norm boundedness removes the danger of large overshoots

or undershoots, but it does not provide enough information to imply convergence. On the other hand, it is not so easy to get a total-variation bound for (3.22) that guarantees convergence. At present, there seems to be no scheme of local stencil with order of accuracy higher than one, which is proven to be TVD in $d \geqq 2$ dimensions (however, see [15]). Recently, Goodman and LeVeque [2] obtained a rather discouraging result: they proved that in $d = 2$, a TVD scheme under the total-variation definition

$$TV(u) = \sum_{j_1} \sum_{j_2} (\Delta x_2 |\Delta_+^{x_1} u_{j_1 j_2}| + \Delta x_1 |\Delta_+^{x_2} u_{j_1 j_2}|) \tag{3.30}$$

is at most first-order accurate. Although their result does not rule out the possibility of the existence of TVB high-order multidimensional schemes, numerical experiments do indicate a total-variation increase when using a one-dimensional TVD scheme and dimension splitting. Usually the total-variation increase is caused by distortions of level curves rather than by oscillations; hence (3.30) may not be the best choice for the definition of total-variation in two-dimension.

4. Preliminary numerical results. We present some preliminary numerical results for the method in § 2 (scalar case only). Numbers are sometimes written in exponential forms, e.g., $4.2-1$ means 4.2×10^{-1}. (See Tables 2–4.)

Example 1. We use the Runge–Kutta-type TVD schemes in § 2 equipped with the third-order TVD scheme in [11] (the β-scheme there). Also see [13].

The following problems are solved:

$$\begin{aligned} &u_t + u_x = -e^{-x}, \qquad 0 \leqq x \leqq 1, \\ &u(0, t) = 1, \qquad u(x, 0) = 1, \end{aligned} \tag{4.1}$$

$$\begin{aligned} &u_t + u_x = \pi \cos \pi x, \qquad -1 \leqq x \leqq 1, \\ &u(-1, t) = 0, \qquad u(x, 0) = 0. \end{aligned} \tag{4.2}$$

Notice that the boundary conditions given are well posed. The steady state solutions are e^{-x} and $\sin \pi x$, respectively. The equations are linear, but the schemes we use are nonlinear.

For numerical boundary conditions, we use the TVB boundary treatment in [14], which is a combination of extrapolations, fixed boundary conditions at inflow (in these problems the left boundary), and letting the scheme itself compute the outflow. The boundary treatment is proved to be TVB in [14].

First we compute for steady state with $m = 1$ in (2.4) (which is just Euler forward in time), then we use a variable m according to the following rule: $m = 2, 4$, or 6 when the L_1-residue $\|u^n - u^{n-1}\|_1$ is >1, in $[0.01, 0.1]$ or <0.01, respectively. The stopping criterion is $\|u^n - u^{n-1}\|_1 < 10^{-4} \Delta t$.

The results are in Table 2.

Notice that the Runge–Kutta-type method does save the computation cost significantly. Because in each cycle, the costly nonlinear terms C's and D's are computed only once.

Example 2. The Runge–Kutta-type schemes in § 2 are used to solve a nonlinear singular perturbation problem [9]:

$$\begin{aligned} &\varepsilon y'' - (\tfrac{1}{2} y^2)' - y = 0, \qquad 0 \leqq x \leqq 1, \\ &y(0) = A, \qquad y(1) = B \end{aligned} \tag{4.3}$$

Here ε is a small positive number. We take $\varepsilon = 10^{-6}$.

TABLE 2
Example 1

Equation	Δx	L_1-error	L_∞-error	$m=1$ steps	$m=2-4-6$ cycles/steps
4.1	1/10	4.2-5	2.1-4	321	54/264
	1/20	6.6-6	3.4-5	343	64/314
4.2	1/20	8.9-4	9.9-3	490	179/886
	1/40	1.4-4	2.3-3	750	170/822

The exact solutions to (4.3) are (see [9]):

(i) For $A=1$, $B=0.5$; $y(x)\equiv 0$ (boundary layers at $x=0$ and $x=1$).

(ii) For $A=0.1$, $B=-2.0$; $y(x)=-x-1$ (boundary layer at $x=0$).

(iii) For $A=0.25$, $B=-0.4$;

$$y(x)=\begin{cases}-x+0.25, & 0\leqq x<0.25,\\ 0, & 0.25\leqq x\leqq 0.6,\\ -x+0.6, & 0.6<x\leqq 1.\end{cases}$$

(Continuous but not differentiable at $x=0.25$ and $x=0.6$.)

(iv) For $A=0.75$, $B=-0.5$;

$$y(x)=\begin{cases}0.75-x, & 0\leqq x<0.625,\\ -x+0.5, & 0.625<x\leqq 1.\end{cases}$$

(Shock at $x=0.625$.)

In [9], problem (4.3) is solved by going to steady state using the scheme

$$u_j^{n+1}=u_j^n-\lambda\Delta_- h_{j+1/2}-\Delta t u_j^n+\frac{\varepsilon}{\Delta x}\lambda\Delta_+\Delta_- u_j^n \tag{4.4}$$

where $h_{j+1/2}$ is a two-point, first-order monotone flux. Convergence as $n\to\infty$, independently of ε and independently of u^0, is proved.

A natural extension is to consider replacing $h_{j+1/2}$ in (4.4) by a higher-order TVD flux $\hat{f}_{j+1/2}$. We again use the third-order β-scheme in [11] and run (4.4) (Euler forward in time) to steady state. The convergence is slow but we do get better accuracy than the first-order scheme in steady state solutions. The result for $\Delta x=1/20$ is in Table 3.

Further computations using $\Delta x=1/40$ verifies third-order accuracy in steady state solutions away from boundary layers and discontinuities. But the convergence is very slow. For the worst case $A=0.75$, $B=-0.5$ (with an inner shock), the Euler forward

TABLE 3
Example 2
$\varepsilon=1-6$; $\Delta x=1/20$; *initial guess* $u_j^0\equiv A$, $j\leqq N-1$; $u_N^0=B$.

		L_1-error		Number of iterations	
A	B	1st order	3rd order	1st order	3rd order
−1.0	0.5	0	5.5-5	135	485
0.1	−2.0	7.0-3	8.1-4	135	193
0.25	−0.4	1.6-2	3.3-3	135	242
0.75	−0.5	1.3-2	3.6-3	135	450

TABLE 4
Example 2
k = the number of cycles needed for converging to steady state.

m	1	2	3	4	5	6
k	>2000	182	135	112	99	90

procedure does not converge even after $k = 2000$ steps. On the other hand, the Runge-Kutta-type TVD schemes (2.4)-(2.6) (with $m \geqq 2$) do this very well. See Table 4 for the comparisons.

This example again shows that our Runge-Kutta type schemes (2.4)-(2.6) are very efficient for computing steady state solutions.

Acknowledgments. The author expresses his sincere thanks to Professor Stanley Osher for his valuable help and suggestions, and to an unknown referee for helpful suggestions on the first version of this paper, especially the problems discussed in Remark 2.3.

REFERENCES

[1] P. EMBID, J. GOODMAN, AND A. MAJDA, *Multiple steady states for* 1-D *transonic flow*, SIAM J. Sci. Statist. Comput., 5 (1984), pp. 21-41.

[2] J. GOODMAN AND R. LEVEQUE, *On the accuracy of stable schemes for* 2D *scalar conservation laws*, Math. Comp., 45 (1985), pp. 15-21.

[3] A. HARTEN, *High resolution schemes for hyperbolic conservation laws*, J. Comput. Phys., 49 (1983), pp. 357-393.

[4] ———, *On a class of high resolution total-variation-stable finite difference schemes*, SIAM J. Numer. Anal., 21 (1984), pp. 1-23.

[5] A. HARTEN AND S. OSHER, *Uniformly high order accurate non-oscillatory schemes*, I, SIAM J. Numer. Anal., 24 (1987), pp. 279-309.

[6] A. HARTEN, B. ENGQUIST, S. OSHER, AND S. CHAKRAVARTHY, *Uniformly high order accurate essentially non-oscillatory schemes*, III, ICASE Report 86-22, ICASE NASA-Langley Research Center, Hampton, VA, 1986; J. Comput. Phys., 71 (1987), pp. 231-303.

[7] A. JAMESON, *A Non-Oscillatory Shock Capturing Scheme Using Flux Limited Dissipation*, AMS Lectures in Applied Mathematics 22, Providence, RI, 1985, pp. 345-370.

[8] J. D. LAMBERT, *Computational Methods in Ordinary Differential Equations*, John Wiley, New York, 1973.

[9] S. OSHER, *Nonlinear singular perturbation problems and one sided difference schemes*, SIAM J. Numer. Anal., 18 (1981), pp. 129-144.

[10] S. OSHER AND S. CHAKRAVARTHY, *High resolution schemes and the entropy condition*, SIAM J. Numer. Anal., 21 (1984), pp. 955-984.

[11] ———, *Very high order accurate* TVD *schemes*, ICASE Report #84-44, ICASE NASA-Langley Research Center, Hampton, VA, 1984; IMA Volumes in Mathematics and its Applications 2, Springer-Verlag, Berlin, New York, 1986, pp. 229-274.

[12] P. ROE, *Approximate Riemann solvers, parameter vectors, and difference schemes*, J. Comput. Phys., 43 (1981), pp. 357-372.

[13] C. SHU, TVB *Uniformly high order schemes for conservation laws*, Math. Comp., 49 (1987), pp. 105-121.

[14] ———, TVB *boundary treatment for numerical solutions of conservation laws*, Math. Comp., 49 (1987), pp. 123-134.

[15] ———, TVD *properties of a class of modified* ENO *schemes for scalar conservation laws*, IMA Preprint Series, #308, University of Minnesota, Minneapolis, MN, 1987.

[16] C. SHU AND S. OSHER, *Efficient implementation of essentially non-oscillatory shock capturing schemes*, ICASE Report 87-33, ICASE NASA-Langley Research Center, Hampton, VA, 1987; J. Comp. Phys., to appear.

[17] E. TURKEL, *Acceleration to a steady state for the Euler equations*, ICASE Report #84-32, ICASE NASA-Langley Research Center, Hampton, VA, 1984; *Numerical Methods for the Euler Equations of Fluid Dynamics*, F. Angrand, A. Dervieux, J. A. Desideri, and R. Glowinski, eds., Society for Industrial and Applied Mathematics, Philadelphia, PA, 1985, pp. 281-311.