Differential Equations and Computational Simulations III
J. Graef, R. Shivaji, B. Soni & J. Zhu (Editors)
Electronic Journal of Differential Equations, Conference 01, 1997, pp. 171–191.
ISSN: 1072-6691. URL: http://ejde.math.swt.edu or http://ejde.math.unt.edu
ftp 147.26.103.110 or 129.120.3.113 (login: ftp)

ENTROPY CONSISTENT, TVD METHODS WITH HIGH ACCURACY FOR CONSERVATION LAWS

Xuefeng Li

Abstract

The Godunov method for conservation laws produces numerical solutions that are total-variation diminishing (TVD) and converge to weak solutions which satisfy the entropy condition (Entropy Consistency), but the method is only first order accurate. Many second and higher order accurate Godunov–type methods have been developed by various researchers. Although these high order methods perform very well numerically, convergence and entropy-consistency has not been proven, maybe due to the highly nonlinear approach. In this paper, we develop a new class of Godunov–type methods that are TVD, converge to weak solutions of conservation laws, and satisfy the entropy condition. The error produced by these methods are theoretically controllable by the choice the piecewise constant functions used in the numerical approximation. Numerical experiments confirm that our methods produce numerical solutions that are comparable to those produced by higher order methods, while maintaining all the good characteristics of the Godunov method.

1. Introduction

A new class of numerical methods for nonlinear systems of hyperbolic conservation laws,

$$\partial_t u + \partial_x f(u) = 0, \quad (x,t) \in (-\infty, +\infty) \times (t_0, +\infty), \tag{1.1a}$$

$$u(x,t_0) = u_0(x), \quad x \in (-\infty, +\infty), \tag{1.1b}$$

is presented in this paper. The solution vector u has m components and is a function of two variables x and t. That is, $u = u(x,t) = (u_1, \dots, u_m)^T$, $f(u) = (f_1(u), \dots, f_m(u))^T$, and matrix $\partial_u f$ has m distinct real eigenvalues.

It is well known that a typical solution of (1.1) develops jump discontinuities (called shocks) in finite time [Lax, 1973]. Thus solution at large exists only in a weak sense. A weak solution of (1.1) is defined as a function $u(x,t)$ that satisfies

$$\int_{t_0}^{+\infty} \int_{-\infty}^{+\infty} [(\partial_t \phi)u + (\partial_x \phi) f(u)]\, dx\, dt + \int_{-\infty}^{+\infty} \phi(x,t_0) u_0(x)\, dx = 0, \tag{1.2}$$

for all $\phi(x,t) \in C^\infty$ with compact support in the half plane $(-\infty, +\infty) \times (t_0, +\infty)$.

1991 *Subject Classification:* 65C20, 65M12, 65M06.
Key words and phrases: Conservation Laws, Godunov Method, Entropy Condition, Convergence, High Accuracy.
©1998 Southwest Texas State University and University of North Texas.
Published November 12, 1998.

Since weak solutions of (1.1) are not unique [Lax, 1973], additional conditions must be used to identify the physically relevant solution, or the entropy solution $u(x,t)$ (piecewise continuous) that satisfies the entropy conditions

$$\partial_t U(u) + \partial_x F(u) \le 0, \quad (x,t) \in (-\infty, +\infty) \times (t_0, +\infty), \tag{1.3a}$$

in weak sense, or

$$-\int_{t_0}^{+\infty}\int_{-\infty}^{+\infty} [(\partial_t \phi)U(u) + (\partial_x \phi)F(u)]\, dx\, dt - \int_{-\infty}^{+\infty} \phi(x,t_0)U(u_0(x))\, dx \le 0, \tag{1.3b}$$

for all nonnegative $\phi(x,t) \in C^\infty$ with compact support in the half plane $(-\infty,+\infty)\times(t_0,+\infty)$. In addition, across a discontinuity of $u(x,t)$ with the speed of propagation being S, there must be

$$F(u_r) - F(u_l) - S[U(u_r) - U(u_l)] \le 0\,. \tag{1.3c}$$

Here, u_l and u_r are the states of u on the left and right of the discontinuity of $u(x,t)$; $U(u)$ is the entropy function of (1.1) and U is convex, that is,

$$U(\frac{1}{2}(p+q)) \le \frac{1}{2}(U(p) + U(q))\,. \tag{1.4}$$

$F(u)$ is the entropy flux of (1.1) and satisfies $\partial_u U \partial_u f = \partial_u F$. The existence of entropy U and entropy flux F of (1.1) is assumed here. For most conservation laws, this assumption is indeed true [Harten, 1983]. In the case of scalar conservation laws, the entropy condition (1.3) ensures the uniqueness of weak solution of (1.1) in the class of piecewise continuous functions as stated in the following theorem.

Theorem 1.1 [Oleinik, 1957; Lax, 1973] (Uniqueness). *A weak piecewise continuous solution $u(x,t)$ of a scalar conservation law in the form of (1.1) is unique in the class of piecewise continuous functions (finitely many discontinuities inside any compact set in x–t plane), if and only if $u(x,t)$ satisfies (1.3).*

Let $\{x_{j+\frac{1}{2}}\}$ be a set of grid points on the $x-$axis. Denote $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$ by I_j. Define grid size h to be $h = \sup_j |x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}|$. Let $x_j = \frac{1}{2}(x_{j-\frac{1}{2}} + x_{j+\frac{1}{2}})$, $\Delta x_j = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$, $\Delta x_{j+\frac{1}{2}} = x_{j+1} - x_j$. Let $\Delta t_n = \tau$ be the time step, i.e., $t_{n+1} = t_n + \Delta t_n = t_n + \tau$ $(n = 0, 1, \ldots)$. For ease of exposition, equal spacing in x is assumed from now on. That is, $\Delta x_j = \Delta x_{j+\frac{1}{2}} \equiv \Delta x \equiv h$. Denote τ/h by λ. If we define the averaged value of $u(x,t)$ over I_j by u_j^n, i.e.,

$$u_j^n = \frac{1}{h}\int_{I_j} u(x,t_n)dx, \tag{1.5}$$

it can then be shown using Green's theorem over the rectangle $I_j \times [t_n, t_{n+1}]$ that $\{u_j^n\}$ satisfy the following relations:

$$u_j^{n+1} = u_j^n - \lambda[f_{j+\frac{1}{2}}^n - f_{j-\frac{1}{2}}^n], \tag{1.6a}$$

$$f_{j+\frac{1}{2}}^n = \frac{1}{\tau}\int_{t_n}^{t_n+\tau} f(u(x_{j+\frac{1}{2}}, t))dt. \tag{1.6b}$$

In this paper, a new class of numerical methods is developed based on (1.6). This new class of highly accurate numerical methods are convergent, and the limits of their numerical solutions satisfy the entropy condition of (1.3b). And their numerical accuracy are comparable to those of high order methods.

A general discussion of numerical methods based on (1.6) is presented in section 2. Approximations of functions using piecewise constant functions are discussed in section 3. The formulation of the new methods and further remarks are given in section 4. Numerical implementation and tests are presented in section 5 in comparison with results from Godunov method. It can be shown that the newly developed first order methods produce numerical solutions with sharp resolution seen in those produced by high order methods.

2. Godunov and Godunov Type methods

When a numerical solution to (1.1) is to be computed, it is often the discrete averaged values of $u(x,t)$ defined in (1.5) that are being calculated, using relations (1.6) where $\{u_j^0\}$ are initialized by the initial function $u_0(x)$ in (1.5) and $u(x,t_0) = u_0(x)$. Many numerical methods are different only in the ways the integral in (1.6b) is approximated. Evaluation of the integral in (1.6b) often involves the evaluation of Riemann problems, or generalized Riemann problems of (1.1). A Riemann problem of (1.1) is defined as:

$$\begin{aligned} \partial_t u + \partial_x f(u) = 0, &\quad (x,t) \in (-\infty,+\infty)\times(t_0,+\infty), \\ u(x,t_0) = u_l, &\quad x < x_0, \\ u(x,t_0) = u_r, &\quad x > x_0. \end{aligned} \tag{2.1}$$

That is a conservation law with the initial function as a step function. The solution of (2.1) is self–similar with respect to the point (x_0,t_0) and is denoted by $R((x-x_0)/(t-t_0);u_l,u_r)$. And a generalized Riemann problem of (1.1) is defined as:

$$\begin{aligned} \partial_t u + \partial_x f(u) = 0, &\quad (x,t) \in (-\infty,+\infty)\times(t_0,+\infty), \\ u(x,t_0) = u_l(x), &\quad x < x_0, \\ u(x,t_0) = u_r(x), &\quad x > x_0. \end{aligned} \tag{2.2}$$

The solution of a generalized Riemann problem is denoted by $G(x,t;x_0,t_0,u_l(\cdot),u_r(\cdot))$.

A difference scheme is said to be consistent with conservation law (1.1) (called a conservative scheme) if it can be represented in the following format

$$\begin{aligned} v_j^{n+1} &= v_j^n - \lambda[f^n_{j+\frac{1}{2}} - f^n_{j-\frac{1}{2}}], \\ f^n_{j+\frac{1}{2}} &= g(v^n_{j-l+1},\ldots,v^n_{j+l}), \\ g(u,\ldots,u) &\equiv f(u). \end{aligned} \tag{2.3}$$

A difference scheme is said to be consistent with the entropy condition (1.3a) if

$$\begin{aligned} U_j^{n+1} &\le U_j^n - \lambda[F^n_{j+\frac{1}{2}} - F^n_{j-\frac{1}{2}}], \\ U_j^n &= U(v_j^n), \\ F^n_{j+\frac{1}{2}} &= g(v^n_{j-l+1},\ldots,v^n_{j+l}), \\ g(u,\ldots,u) &\equiv F(u). \end{aligned} \tag{2.4}$$

The lattice function $\{v_j^n\}$ is usually extended to continuous values of x, t by setting:

$$v_h(x,t) = v_j^n, \quad (x,t) \in I_j \times [t_n, t_{n+1}). \tag{2.5}$$

A difference scheme is said to be total variation (TV) stable if the total variation in x of $v_h(x,t)$,

$$TV(v_h(\cdot,t)) \equiv TV(v^n) = \sum_j |v_{j+1}^n - v_j^n|, \tag{2.6}$$

is uniformly bounded in t and h; here n is the integer part of t/τ. A difference scheme is said to be total variation diminishing (TVD) if:

$$TV(v^{n+1}) \le TV(v^n). \tag{2.7}$$

Here are two important theorems concerning the above consistent difference schemes.

Theorem 2.1 [Lax, Wendroff, 1960; Harten, Lax, Van Leer, 1983]. *Suppose a difference scheme is conservative and consistent with the entropy condition (1.3a) in the form of (2.3). Let $\{v_j^n\}$ be a solution of (2.3), with initial values $v_j^0 = u_j^0$ as defined in (1.5). Let $v_h(x,t)$ be the extended function of $\{v_j^n\}$ as defined in (2.5). Suppose that for some sequence $h_k \to 0^+$, $\tau_k/h_k = \lambda$ =constant, the limit:*

$$\lim_{h_k \to 0^+} v_{h_k}(x,t) = u(x,t), \tag{2.8}$$

exists in the sense of bounded, L^1_{loc} convergence. Then the limit $u(x,t)$ satisfies the weak form (1.2) of the conservation law, and the weak form (1.3b) of the entropy condition.

Theorem 2.2 [Harten, 1984]. *Suppose the difference scheme (2.3) is conservative and TV stable. Then the scheme produces numerical solution with a convergent subsequence whose limit (in the sense of bounded, L^1_{loc}) is a weak solution of (1.1).*

In the case of scalar conservation laws, Theorem 2.1 ensures the uniqueness of the limit solution when the entropy condition is satisfied, while Theorem 2.2 guarantees the existence of a limit solution if the difference scheme is TV stable. Therefore, *a conservative difference scheme is convergent and its limit function is the unique solution of (1.1) if the scheme is TV stable and entropy consistent. The goal of this paper is to develop a new class of difference methods that are TV stable and entropy consistent. Furthermore, they produce numerical solutions with accuracy comparable to those produced by high order numerical methods.*

Godunov [1959], Van Leer (MUSCL) [1979], Colella (MUSCL) [1985], Colella and Woodward (PPM) [1982], Harten, Engquist, Osher and Chakravarthy (ENO) [1987] developed numerical methods for solving (1.1) which use relations (1.6) in the numerical approximations. In particular, Godunov method uses piecewise constant functions (constant over each interval I_j with possible jumps at $\{x_{j+\frac{1}{2}}\}$) to approximate $u(x,t_n)$; thus the integral in (1.6b) can be evaluated exactly by solving Riemann problems of (1.1) at $\{x_{j+\frac{1}{2}}\}$. In the MUSCL (PPM) scheme, piecewise linear (quadratic) functions (linear (quadratic) over each interval I_j with possible

jumps at $\{x_{j+\frac{1}{2}}\}$) are used to approximate $u(x,t_n)$; the integral in (1.6b) is approximated using the trapezoidal rule. In the ENO schemes, piecewise polynomial functions (N-th degree polynomial over each interval I_j with possible jumps at $\{x_{j+\frac{1}{2}}\}$) are used to approximate $u(x,t_n)$; the integral in (1.6b) is approximated using an appropriate K-point numerical quadrature. All of those methods can be formulated in the following fashion:

$$v_j^{n+1} = v_j^n - \lambda[f_{j+\frac{1}{2}}^n - f_{j-\frac{1}{2}}^n], \tag{2.9a}$$

$$f_{j+\frac{1}{2}}^n = \sum_k \alpha_k f(v_{j+\frac{1}{2}}^{n,\gamma_k}), \tag{2.9b}$$

$$\lambda = \frac{\tau}{h} \le \frac{\theta}{\Lambda}, \tag{2.9c}$$

where (2.9b) is an appropriate numerical quadrature approximating the integral in (1.6b); θ is a positive constant less than 1, and it is called the CFL number; α_k and γ_k are coefficients of the numerical quadrature; $\{v_{j+\frac{1}{2}}^{n,\gamma_k}\}$ are values of u needed in the quadrature; Λ is the largest eigenvalue of the matrix $\partial_u f$ in absolute value, which represents the maximum speed of propagation of discontinuities of $u(x,t)$.

Equations listed in (2.9) represent a class of recursive algorithms for solving (1.1). A major task of these algorithms is the evaluation of $\{v_{j+\frac{1}{2}}^{n,\gamma_k}\}$, values of $u(x,t)$ needed in the quadrature in (2.9b). Assuming $u(x,t_n)$ is approximated by $u_j^n(x)$ over interval I_j. Then

$$v_{j+\frac{1}{2}}^{n,\gamma_k} = G(x_{j+\frac{1}{2}}, \gamma_k\tau; x_{j+\frac{1}{2}}, t_n, u_j^n(\cdot), u_{j+1}^n(\cdot)), \tag{2.10}$$

as indicated in (2.2).

In the Godunov method, $\{u_j^n(x)\}$ are chosen to be constants. That is, $u(x,t_n)$ is approximated using a piecewise constant function

$$P_n(x) = v_j^n = \frac{1}{x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}} \int_{I_j} u(x,t_n)\,dx \quad \text{for } x \in I_j\,.$$

The piecewise constant function $P_{n+1}(x)$ approximating $u(x,t_{n+1})$ is computed by averaging the exact solutions of (1.1) over each I_j, using $P_n(x)$ as the initial function. This exact solution consists of a sequence of Riemann solutions at $\{x_{j+\frac{1}{2}}\}$. Using Green's theorem over the rectangle $I_j \times [t_n, t_{n+1}]$, one obtains

$$\int_{I_j} P_n(x)dx - \int_{I_j} u(x,t_{n+1})dx - \int_{t_n}^{t_n+\tau} f(u(x_{j+\frac{1}{2}},t))dt + \int_{t_n}^{t_n+\tau} f(u(x_{j-\frac{1}{2}},t))dt = 0. \tag{2.11}$$

Because the initial value $P_n(x) = v_j^n,\ x \in I_j$, is piecewise constant,

$$u(x_{j+\frac{1}{2}},t) = R((x - x_{j+\frac{1}{2}})/(t-t_n); v_j^n, v_{j+1}^n)|_{x=x_{j+\frac{1}{2}}} = R(0; v_j^n, v_{j+1}^n) = \text{constant}, \tag{2.12}$$

provided $\tau \leq \frac{\theta}{\Lambda}h$, which ensures that shocks from $x_{j-\frac{1}{2}}$ and $x_{j+1+\frac{1}{2}}$ will not reach $x_{j+\frac{1}{2}}$. Therefore, the approximated averaged value of $u(x, t_{n+1})$ over I_j satisfies:

$$\begin{aligned} v_j^{n+1} &= \frac{1}{\Delta x}\int_{I_j} u(x, t_{n+1})\, dx \\ &= v_j^n - \lambda[f(R(0; v_j^n, v_{j+1}^n)) - f(R(0; v_{j-1}^n, v_j^n))], \end{aligned} \tag{2.13}$$

which indeed can be formulated in terms of (2.9):

$$\begin{aligned} v_j^{n+1} &= v_j^n - \lambda[f_{j+\frac{1}{2}}^n - f_{j-\frac{1}{2}}^n], \\ f_{j+\frac{1}{2}}^n &= f(R(0; v_j^n, v_{j+1}^n)), \\ \lambda &= \frac{\tau}{h} \leq \frac{\theta}{\Lambda}, \end{aligned} \tag{2.14}$$

where the numerical quadrature used here is just the rectangle rule.

One concludes that $\{v_j^{n+1}\}$ is computed from $\{v_j^n\}$ through the use of two types of operations: Riemann Problem Solver (in (2.12)) and Integral Averaging (in (2.13)). The two processes are TV non–increasing in the case of scalar conservation laws. Therefore, Godunov method is TV stable (it is actually total variation diminishing). For a convex entropy $U(u)$ and entropy flux $F(u)$,

$$\frac{1}{\Delta x}\int_{I_j} U(u(x, t_{n+1}))dx \leq U(v_j^n) - \lambda[F(R(0; v_j^n, v_{j+1}^n)) - F(R(0; v_{j-1}^n, v_j^n))] \tag{2.15}$$

because $u(x, t_{n+1})$ is the unique solution of (1.1) with initial value $P_n(x)$ (Theorem 1.1). Due to Jensen's inequality for convex functions,

$$\begin{aligned} \frac{1}{\Delta x}\int_{I_j} U(u(x, t_{n+1}))dx &\geq U(\frac{1}{\Delta x}\int_{I_j} u(x, t_{n+1})dx) \\ &= U(v_j^{n+1}) = U_j^{n+1}. \end{aligned} \tag{2.16}$$

One thus derives that:

$$U_j^{n+1} \leq U_j^n - \lambda[F_{j+\frac{1}{2}}^n - F_{j-\frac{1}{2}}^n], \tag{2.17}$$

which indicates that Godunov method is also entropy consistent.

The Godunov method enjoys the advantages of being TVD and entropy consistent because $u_j^n(x)$ is constant, which also results in the method being just first order accurate. On the other hand, $u_j^n(x)$ is a polynomial in x, in methods like MUSCL, PPM, ENO, thus these methods enjoy the advantage of being highly accurate but lacking the stability property (such as TV stability) and entropy consistency property. TV stability and entropy consistency of those high order schemes are yet to be further studied [Vila, 1989].

The goal of this paper is to develop a class of highly accurate methods that use piecewise constant functions for approximations. Because they use piecewise constant functions for approximations, they enjoy all advantages the Godunov method does: TV stability and entropy consistency.

Let's first consider approximating a function $f(x)$ using piecewise constant functions in the next section.

3. Approximation using Piecewise Constant Functions

The following theorem states results concerning approximating a function using piecewise constant function over a partition of an interval.

Theorem 3.1. *Let $f(x)$ be a function over interval $[x_l, x_r]$. Given an integer N, let $x_l = x_1 < x_2 < \ldots < x_N < x_{N+1} = x_r$ be a partition of interval $[x_l, x_r]$, and $P_N(x)$ be a piecewise constant function which equals b_i for $x \in [x_i, x_{i+1}) \equiv I_i$. Let $\delta_i = x_{i+1} - x_i$, and the L^2-error be:*

$$E_0 = \sqrt{\frac{1}{x_r - x_l}\int_{x_l}^{x_r} |f(x) - P_N(x)|^2 dx} = \sqrt{\frac{1}{x_r - x_l}\sum_{i=1}^{N}\int_{I_i} |f(x) - b_i|^2 dx}. \quad (3.1)$$

Then the piecewise constant function that minimizes the above L^2-error E_0 satisfies:

$$b_i = \frac{1}{\delta_i}\int_{I_i} f(x)dx \quad (3.2a)$$

$$\Big[\frac{1}{\delta_i}\int_{I_i} f(x)dx - \frac{1}{\delta_{i+1}}\int_{I_{i+1}} f(x)dx\Big]\times$$
$$\Big[\frac{1}{\delta_i}\int_{I_i} f(x)dx + \frac{1}{\delta_{i+1}}\int_{I_{i+1}} f(x)dx - 2f(x_i)\Big] = 0 \quad (3.2b)$$
$$for\ i = 2, 3, \ldots, N$$

Proof: By setting $\partial E_0^2/\partial b_i = 0$, one obtains the only solution of $b_i = \frac{1}{\delta_i}\int_{I_i} f(x)dx$.

By setting $\partial E_0^2/\partial x_{i+1} = 0$, one obtains

$$(f(x_i) - b_i)^2 - (f(x_i) - b_{i+1})^2 = 0, \quad (3.3)$$

or

$$(b_{i+1} - b_i)(2f(x_i) - b_i - b_{i+1}) = 0. \quad (3.4)$$

Substitute the result in (3.2a) into (3.4), one obtains the result in (3.2b). ♢

Corollary 3.2. *If $f(x)$ in Theorem 3.1 is strictly monotone over interval $[x_l, x_r]$, equation (3.2b) is equivalent to the following equation:*

$$\frac{1}{\delta_i}\int_{I_i} f(x)dx + \frac{1}{\delta_{i+1}}\int_{I_{i+1}} f(x)dx = 2f(x_i). \quad (3.5)$$

Proof: Because $f(x)$ is strictly monotone, $\frac{1}{\delta_i}\int_{I_i} f(x)\,dx \neq \frac{1}{\delta_{i+1}}\int_{I_{i+1}} f(x)\,dx$. ♢

Corollary 3.3. *If $f(x)$ in Theorem 3.1 is linear over interval $[x_l, x_r]$, that is, $f(x) = mx + b$, $m \neq 0$, $x \in [x_l, x_r]$, equation (3.2b) is equivalent to the following equation*

$$x_i - x_{i-1} = x_{i+1} - x_i, \quad for\ i = 2, 3, \ldots, N \quad (3.6)$$

which indicates an equally spaced partition. Furthermore, the minimum L^2-error achieved with the equally spaced partition is (where $h = x_r - x_l$)

$$\min_{all\ partitions} E_0 = \frac{|m|}{\sqrt{12}N}(x_r - x_l) = \frac{|m|}{\sqrt{12}N}h. \tag{3.7}$$

Proof: Because $f(x) = mx + b$,

$$b_i = \frac{1}{2}m(x_i + x_{i+1})(x_{i+1} - x_i) + b(x_{i+1} - x_i) \tag{3.8a}$$

$$2f(x_i) = 2mx_i + 2b. \tag{3.8b}$$

Because linear functions are strictly monotone, (3.2b) is equivalent to (3.5), substitute (3.8) into (3.5) results in the following:

$$\frac{1}{2}m(x_{i-1} + x_i) + b + \frac{1}{2}m(x_i + x_{i+1}) + b = 2mx_i + 2b, \tag{3.9}$$

which means:

$$m(x_i - x_{i-1}) = m(x_{i+1} - x_i). \tag{3.10}$$

Equation (3.7) can be obtained now with simple algebraic computation and the derivation is omitted here. ♢

Note: When $f(x)$ in Theorem 3.1 is nonlinear, equation (3.2b) is a nonlinear equation satisfied by the partition $\{x_2, x_3, \ldots, x_N\}$ for $i = 2, 3, \ldots, N$. The optimal partition satisfying (3.2b) is not equally spaced. In fact, if $f(x) = ax^2 + bx + c$, $a \neq 0$, an equally spaced partition $\{x_2, x_3, \ldots, x_N\}$ of interval $[x_l, x_r]$ results in the following:

$$2a(\frac{2}{3}a(x_{i-1} + x_i + x_{i+1}) + b)\delta^2 = 0, \tag{3.11}$$

where $\delta = \frac{1}{N}(x_r - x_l) = \frac{h}{N}$. That means an equally spaced partition is a solution of (3.2b) up to a first order error term. That is indeed the best this methodology can achieve because piecewise constant function on an equally spaced partition is used for the approximation.

Let's now consider approximating a general function using a linear function. The results will be used later for the construction of the new methods.

Theorem 3.4. *Let $f(x)$ be a function over interval $[x_l, x_r]$. Then the linear function that minimizes the following L^2-error.*

$$E_1 = \sqrt{\frac{1}{x_r - x_l}\int_{x_l}^{x_r} |f(x) - L(x)|^2 dx}, \tag{3.12}$$

where $L(x)$ is any linear function over interval $[x_l, x_r]$, is the function $L(x) = m(x - c) + b$ ($c = \frac{1}{2}(x_l + x_r)$, midpoint of interval $[x_l, x_r]$) with:

$$m = \frac{12}{(x_r - x_l)^3}\int_{x_l}^{x_r} f(x)(x - c)dx, \tag{3.13a}$$

$$b = \overline{f}(x_l, x_r) = \frac{1}{x_r - x_l}\int_{x_l}^{x_r} f(x)dx, \tag{3.13b}$$

and the minimum L^2-error achieved is (where $h = x_r - x_l$):

$$\min_{\textit{linear functions}} E_1 = \sqrt{\frac{1}{x_r - x_l}\int_{x_l}^{x_r} [f(x) - (m(x-c)+b)]^2 dx}$$
$$= \frac{|f''(c)|}{8\sqrt{5}}h^2 + O(h^3) \tag{3.14}$$

Proof: Since $L(x)$ is a linear function over $[x_l, x_r]$, let's assume $L(x) = p(x-c)+q$. Therefore,

$$E_1^2 = \frac{1}{x_r - x_l}\int_{x_l}^{x_r} (f(x) - p(x-c) - q)^2 dx. \tag{3.15}$$

By setting $\partial E_1^2/\partial p = 0$, and $\partial E_1^2/\partial q = 0$, one obtains

$$\int_{x_l}^{x_r} (f(x) - p(x-c) - q)(x-c)dx = 0 \tag{3.16a}$$
$$\int_{x_l}^{x_r} (f(x) - p(x-c) - q)dx = 0. \tag{3.16b}$$

Because c is the midpoint of the interval $[x_l, x_r]$, $\int_{x_l}^{x_r}(x-c)dx = 0$. Relations in (3.13) are thus derived.

Using Taylor expansion, $f(x)$ can be expressed as

$$f(x) = f(c) + f'(c)(x-c) + \frac{1}{2}f''(c)(x-c)^2$$
$$+ \frac{1}{6}f'''(c)(x-c)^3 + \frac{1}{24}f^{(4)}(c)(x-c)^4 + \frac{1}{120}f^{(5)}(\eta)(x-c)^5, \tag{3.17}$$

where η is a certain number in interval (x_l, x_r). Substitute the right side of (3.17) into (3.13), one obtains:

$$m = f'(c) + \frac{1}{40}f'''(c)h^2 + \frac{1}{4480}f^{(5)}(\eta_1)h^4, \tag{3.18a}$$
$$b = f(c) + \frac{1}{24}f''(c)h^2 + \frac{1}{1920}f^{(4)}(\eta_2)h^4 \tag{3.18b}$$

where η_1, η_2 are two other numbers in interval (x_l, x_r). Substitute the right sides of (3.17), (3.18) into (3.12), one obtains:

$$E_1^2 = \frac{1}{x_r - x_l}\int_{x_l}^{x_r} |f(x) - (m(x-c)+b)|^2 dx$$
$$= \frac{1}{x_r - x_l}\int_{x_l}^{x_r} [-\frac{1}{24}f''(c)h^2 + \frac{1}{2}f''(c)(x-c)^2 + O(h^3)]^2 dx \tag{3.19}$$
$$= \frac{(f''(c))^2}{320}h^4 + O(h^5)$$

Therefore,

$$E_1 = \sqrt{\frac{(f''(c))^2}{320}h^4(1+O(h))} = \frac{|f''(c)|}{8\sqrt{5}}h^2 + O(h^3). \tag{3.20}$$

♢

Results from Corollary 3.3 and Theorem 3.4 indicate while linear function approximation results in second order accuracy, piecewise linear function approximation results in first order accuracy but also inversely proportional to N, the number of constants used in the approximation. This leads to the following key result of this paper.

Theorem 3.5 (Error estimate of approximation using piecewise constant function). *Let $f(x)$ be a function defined on $[x_l, x_r]$, $c = \frac{1}{2}(x_l + x_r)$, $h = x_r - x_l$ and $L(x) = m(x-c)+b$ be the linear function minimizing the L^2-error between $f(x)$ and all linear functions on interval $[x_l, x_r]$, as defined in Theorem 3.1. Given an integer N, let $P_N(x)$ be a piecewise constant function over the equally spaced partition $x_l = x_1 < x_2 < \ldots < x_N < x_{N+1} = x_r$, which equals $b_i = \frac{1}{x_{i+1}-x_i}\int_{x_i}^{x_{i+1}} L(x)dx = L(c_i)$, where $c_i = \frac{1}{2}(x_i + x_{i+1})$. Then the L^2-error:*

$$E_2 = \sqrt{\frac{1}{h}\int_{x_l}^{x_r} |f(x) - P_N(x)|^2 dx} \tag{3.21}$$

satisfies:

$$E_2 \le [\frac{|f''(c)|}{8\sqrt{5}} + \frac{|f'(c)|}{\sqrt{12}}]\ h^2 + O(h^3), \tag{3.22}$$

provided:

$$\frac{1}{N} \le h. \tag{3.23}$$

That indicates that piecewise linear function approximation can actually achieve second order accuracy.

Proof: Due to the triangle inequality for norms,

$$\begin{aligned}
E_2 &= \sqrt{\frac{1}{h}\int_{x_l}^{x_r} |f(x) - P_N(x)|^2 dx} && (3.24a)\\
&\le \sqrt{\frac{1}{h}\int_{x_l}^{x_r} |f(x) - L(x)|^2 dx} + \sqrt{\frac{1}{h}\int_{x_l}^{x_r} |L(x) - P_N(x)|^2 dx} && (3.24b)\\
&= \frac{|f''(c)|}{8\sqrt{5}}h^2 + O(h^3) + \frac{|m|}{\sqrt{12N}}h && (3.24c)\\
&\le \frac{|f''(c)|}{8\sqrt{5}}h^2 + \frac{|m|}{\sqrt{12}}h^2 + O(h^3) && (3.24d)\\
&\le \frac{|f''(c)|}{8\sqrt{5}}h^2 + \frac{|f'(c)|}{\sqrt{12}}h^2 + \frac{|f'''(c)|}{40\sqrt{12}}h^4 + O(h^6) + O(h^3) && (3.24e)\\
&= [\frac{|f''(c)|}{8\sqrt{5}} + \frac{|f'(c)|}{\sqrt{12}}]\ h^2 + O(h^3). && (3.24f)
\end{aligned}$$

Here, (3.7) and (3.14b) are used in getting (3.24c), (3.23) is used in getting (3.24d), (3.18a) is used in getting (3.24e). ♢

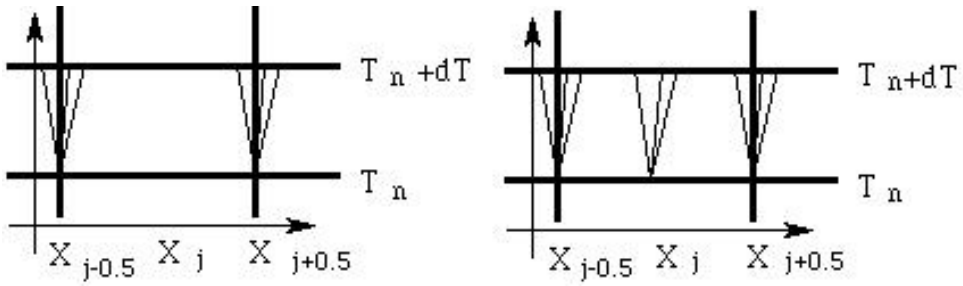

Fig. 4.1. Jumps in Godunov Method and the new Method when N=2

4. Formulation of the New Methods

The result of theorem 3.5 as well as the proof of it suggested the methodology to be used for high accurate approximations of general functions using piecewise constant functions. When a function $f(x)$ is to be approximated, the optimal linear approximation in L^2 norm is constructed. Based on the linear approximation, a piecewise constant function is constructed for the approximation of $f(x)$. The constructed piecewise constant function is then used in the recursive formula outlined in (2.9) for the computation of numerical solutions to conservation laws of (1.1).

The new methods are developed following the steps listed below:

(i) Averaged values $\{v_j^0\}$ of $u(x,t)$ are computed from initial value of $u(x,t)$;

(ii) Based on the averaged values $\{v_j^n\}$ of $u(x,t)$ computed in the previous iteration, a piecewise linear function is constructed which is linear over each interval I_j, that approximates $u(x,t_n)$ with second order accuracy $O(h^2)$ (h = the length of interval I_j), as indicated in theorem 3.4;

(iii) Next, a piecewise constant function is constructed as defined in Theorem 3.5, which has N constants over each interval I_j and the N constants are equally spaced. This piecewise constant function approximates the linear function constructed in (ii) with $O(h^2)$ error provided $1/N = O(h)$, as indicated in Corollary 3.3;

(iv) The exact solution to conservation law (1.1) at time $t = t_n + \tau$ using the piecewise constant function constructed in (iii) as the initial value function is then computed by solving generalized Riemann problems at all interfaces $\{x_{j+\frac{1}{2}}\}$, as indicated in (2.10) where $u_j^n(x)$ = the N–piece constant function constructed in (iii). That is the approximate solution of $u(x,t)$ at $t = t_n + \tau = t_{n+1}$;

(v) Averaged values $\{v_j^{n+1}\}$ of $u(x,t)$ at $t = t_{n+1}$ are then computed from results in (iv). The iteration continues again from (ii).

Remark (a) Step (ii) is the normal procedure for all second order accurate Godunov type methods. The slope of the linear function over I_j is constructed according to (3.13a). Using (3.18a), (where $x_l = x_{j-\frac{1}{2}}$, $x_r = x_{j+\frac{1}{2}}$, and $x_r - x_l = h$ and $c = \frac{1}{2}(x_l + x_r)$), one can see that:

$$m = f'(c) \tag{4.1}$$

results in only an $O(h^2)$ error in the approximation of the slope of the linear function, which is tolerable in all second order accurate numerical methods.

Therefore, the slope of the linear function over I_j is constructed using the smallest between the forward and backward difference quotients,

$$\partial_x u(x,t_n) \approx \begin{cases} (v^n_{j+1}-v^n_j)/h, & \text{if } |v^n_{j+1}-v^n_j| < |v^n_j - v^n_{j-1}| \\ (v^n_j - v^n_{j-1})/h, & \text{if } |v^n_j - v^n_{j-1}| < |v^n_{j+1}-v^n_j| \\ 0, & \text{if } v^n_j \text{ is a local extrema} \end{cases} \tag{4.2}$$

sometimes the centered difference quotient, $(v^n_{j+1} - v^n_{j-1})/(2h)$ is also considered. That is to ensure TV stability for the difference method. The setting used in (4.2) is to avoid creating new local extrema in linear function constructions when v^n_j is a local extremum among v^n_j and $v^n_{j\pm1}$, which is also a normal procedure among many Godunov–type methods.

Remark (b) If piecewise linear functions were used for approximating $u(x,t)$, possible jump discontinuities are introduced at each $x_{j+\frac{1}{2}}$, which resolve into the solutions of generalized Riemann problems whose numerical solutions are not readily available in general, neither can it provide insight of TV stability property of the numerical solutions. Thus piecewise constant functions are used for the actual approximations. As indicated in theorem 3.5, piecewise constant functions can result in second order accuracy provided $1/N = O(h)$.

Theorem 3.4 states that $u(x,t_n)$ can be approximated by

$$L_j(x) = \partial_x u(x_j,t_n)(x-x_j) + v^n_j$$

with $O(h^2)$ error. Theorem 3.5 states that over the equally spaced partition on I_j:

$$x_{j,1} = x_{j-\frac{1}{2}},\ x_{j,i+1} = x_{j,i} + h/N,\ c_{j,i} = \frac{1}{2}(x_{j,i} + x_{j,i+1}),\ i = 1,2,\ldots,N \tag{4.3}$$

the piecewise constant function:

$$P_{j,N}(x) = L_j(c_{j,i}), \quad \text{for} x \in [x_{j,i}, x_{j,i+1}) \tag{4.4}$$

approximates $L_j(x)$ with error $|\partial_x u(x_j,t_n)|h/(\sqrt{12}N)$.

Because N constants are used to approximate $u(x,t_n)$, there are $N+1$ possible jump discontinuities over each I_j (see fig. 4.1 for the case of $N=2$ where the jump discontinuities resolve into three Riemann solutions) which resolve into Riemann solutions and their numerical solutions are readily available. To avoid shock interactions from neighboring Riemann problems, the time step τ must be chosen so that:

$$\lambda \le \frac{1}{2N}\frac{\theta}{\Lambda}. \tag{4.5}$$

As a matter of fact, since only the averaged values of $u(x,t_{n+1})$ on I_j are to be computed, using Green's theorem over the rectangle $I_j \times [t_n, t_n+\tau]$, one obtains:

$$\int_{I_j} P_{j,N}(x)dx - \int_{I_j} u(x,t_{n+1})dx - \int_{t_n}^{t_n+\tau} f(R(0; P_{j,N}(c_{j,N}), P_{j+1,N}(c_{j+1,1})))dt + \int_{t_n}^{t_n+\tau} f(R(0; P_{j-1,N}(c_{j-1,N}), P_{j,N}(c_{j,1})))dt = 0. \tag{4.6}$$

Therefore, the averaged value of $u(x, t_{n+1})$ over I_j can be evaluated by solving just two Riemann problems,

$$R(0; P_{j,N}(c_{j,N}), P_{j+1,N}(c_{j+1,1})) \text{ and}$$
$$R(0; P_{j-1,N}(c_{j-1,N}), P_{j,N}(c_{j,1})),$$

instead of all $N+1$ Riemann problems, and

$$v_j^{n+1} = \frac{1}{h}\int_{I_j} u(x, t_{n+1})dx \tag{4.7a}$$
$$= \frac{1}{h}\int_{I_j} P_{j,N}(x)dx - \frac{\tau}{h}[f(R(0; P_{j,N}(c_{j,N}), P_{j+1,N}(c_{j+1,1})))$$
$$- f(R(0; P_{j-1,N}(c_{j-1,N}), P_{j,N}(c_{j,1})))] \tag{4.7b}$$
$$= v_j^n - \lambda[f_{j+\frac{1}{2}}^n - f_{j-\frac{1}{2}}^n], \tag{4.7c}$$

where

$$f_{j+\frac{1}{2}}^n = f(R(0; P_{j,N}(c_{j,N}), P_{j+1,N}(c_{j+1,1}))), \tag{4.8a}$$
$$P_{j,N}(c_{j,N}) = L_j(x_{j+\frac{1}{2}} - \frac{1}{2}h/N), \tag{4.8b}$$
$$P_{j+1,N}(c_{j+1,1}) = L_{j+1}(x_{j+\frac{1}{2}} + \frac{1}{2}h/N). \tag{4.8c}$$

Because only Riemann solutions at $\{x_{j+\frac{1}{2}}\}$ are needed in (4.7), formulas in (4.7) remain valid so long as shocks from inside the interval I_j do not reach the boundary of the interval I_j. That means (4.7) remains valid for

$$\lambda \le \frac{1}{N}\frac{\theta}{\Lambda}. \tag{4.9}$$

Remark (c) Compare (4.9) to (2.9c), one sees that the time steps in the new methods are $\frac{1}{N}$ of those in other methods described in (2.9). In other words, a new method must carry out N times more iterations than a method of (2.9) in order to get a numerical solution to (1.1) at a specific time T [$T/(\tau/N) = N(T/\tau)$ steps compared to just T/τ steps]. That seems to be the price for achieving high accuracy using piecewise constant functions for approximations.

On the other hand, since Λ is the largest speed of propagation of shock waves over the entire computational domain, in an interval I_j where the shock speed is smaller, requirement (4.9) is too excessive (should have been $\lambda \le \frac{1}{N}\frac{\theta}{\Lambda_j}$ where $\Lambda_j < \Lambda$, and Λ_j is the maximum speed of propagation of discontinuities of $u(x,t)$ in interval I_j). And in an interval I_j where the shock speed does reach the maximum, the slope of the linear function $L_j(x)$, according to (4.2c), is set to zero to avoid creating new local extrema. That means the N constants are identical, or, only one constant, instead of N different constants, is used to approximate $u(x,t)$ in I_j, and once again requirement (4.9) is too excessive (should have been $\lambda \le \frac{\theta}{\Lambda}$). Therefore, in actual numerical computations, one can use time steps slightly larger than those required

by (4.9). The benefit of using larger time steps in numerical computations is most visible in a region where $u(x,t)$ is smooth (such as inside a rarefaction wave); the drawback is degradation in accuracy in an area near the fastest shock wave, where one constant, instead of N constants, will have to be used for approximations. Thus (4.9) is modified to become:

$$\lambda \le \frac{\zeta}{N}\frac{\theta}{\Lambda}, \tag{4.10}$$

where $\zeta \ge 1$ is a pre-chosen constant.

Remark (d). Since numerical solution over I_j is to be computed, I_j itself is quite small in practice, and it is not recommended to further divide I_j into too many subintervals due to round off errors. As indicated by numerical examples shown later, the mere use of $N = 2$ can produce solutions with satisfactory resolution where restriction (4.10) becomes

$$\lambda \le \frac{\zeta}{2}\frac{\theta}{\Lambda}. \tag{4.11}$$

ζ can be chosen between 1.5 and 2.0 without introducing too much degradation in numerical solutions, which makes the method in case of $N = 2$ very efficient yet accurate.

Remark (e) Restriction (4.10) determines the number of computations (addition, subtraction, multiplication, division and logical comparison) required in a new method for computing approximation to $u(x,T)$. Assuming there are m grid points in the computational domain, the number of Riemann problems to be solved on each time step is m. Thus the total number of computations required for the Godunov method and a new method to compute approximations to $u(x,T)$ is $G_c(T) = O(m\omega\Lambda T/(\theta h))$ and $N_c(T) = O(m\omega\Lambda NT/(\zeta\theta h))+O(L)$, respectively. Here, ω is the number of computations used in solving a Riemann problem; $O(L)$ is the number of computations used for the construction of linear functions described in Theorem 3.4, 3.5 and step (ii) of section 4, and $O(L) = O(m\kappa\Lambda NT/(\zeta\theta h))$ where κ is a constant integer related to the number of computations required for the construction of a linear function over one subinterval I_j, which is described in (4.2). That also indicates that $O(L)$ depends on the number of local extrema in $u(x,t)$, or, $O(L)$ is problem–dependent. Therefore, the relative difference of the number of computations in the two methods is $R_c(T) = [N_c(T) - G_c(T)]/G_c(T) = O(N/\zeta - 1 + (\kappa/\omega)(N/\zeta))$. And one can see that $R_c(T)$ is proportional to the ratio N/ζ, which can be chosen by a user. When $\zeta = N$, $R_c(T) = O(\kappa/\omega)$, which is usually quite small because ω is usually large (the number of computations, i.e., addition, subtraction, multiplication, division and logical comparison, required to compute a solution to a Riemann problem). Numerical tests in section 5 confirm that the new methods indeed out–perform the Godunov methods in terms of accuracy and efficiency.

Remark (f) It will be shown later that the new methods are TVD and entropy consistent. Therefore, the new methods are different from other high order accurate methods. In addition, They also enjoy another advantage over other high order accurate methods. In general, a high order accurate method relies on numerical approximations of derivatives of the exact solution to achieve high order accuracy. On the other hand, the current new methods do not make use of any differentiability information about the exact solution. Thus in a region where the exact solution is only continuous but not differentiable, such as inside a rarefaction wave of gas

dynamics, the new methods should perform equally well as in other regions (where $u(x,t)$ is differentiable), as indicated in fig. 5.1–5.3.

Remark (g) The new methods are different from other high order accurate methods as discussed in remark (f). They are also different from the first order accurate Godunov method in several ways, as discussed below.

Firstly, they are different in terms of grid size and accuracy. In order to achieve a pre-chosen accuracy $\varepsilon \ll 1$, the Godunov method, being first order accurate, must be applied using a grid $\{x_{j+\frac{1}{2}}\}$ with grid size being $\sup_j |x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}| = O(\varepsilon)$; whereas the new methods, being second order accurate, only need a grid $\{y_{j+\frac{1}{2}}\}$ with grid size $\sup_j |y_{j+\frac{1}{2}} - y_{j-\frac{1}{2}}| = O(\sqrt{\varepsilon})$. The savings in the amount of computer memory used is tremendous, especially in solving three dimensional problems with operator splitting techniques.

Secondly, they are different in terms of the number of Riemann problems being solved per time step. Over a fixed region with length L in the computational domain, the number of grid points falling into that region for the Godunov method and the new methods are $O(L/\varepsilon)$ and $O(L/\sqrt{\varepsilon})$, respectively. That translates into solving $O(L/\varepsilon)$ number of Riemann problems for the Godunov method as compared to solving $O(L/\sqrt{\varepsilon})$ number of Riemann problems for the new methods. The ratio between them is $1 : O(1/\sqrt{\varepsilon})$. Because solving Riemann problems, whether exactly or approximately, requires a lot of numerical computation, the savings in using the new methods is again tremendous.

Thirdly, high order accurate methods can achieve accuracy that a first order accurate method may not. When a floating point number is sufficiently small, it is treated by digital computers as a 'zero', though it is not exactly zero. Such a small number is referred to as the machine epsilon. A typical machine epsilon is in the range of $10^{-8} \sim 10^{-16}$. Denote a machine epsilon by $\bigcirc$. That is, $\bigcirc = 0$(on a digital computer). In rare situations, where a pre-chosen accuracy $\varepsilon \ll 1$ is close to the machine epsilon $\bigcirc$ (which is computer–dependent), that is, $\varepsilon \approx \bigcirc$, then the implementation of the Godunov method on such a computer becomes unpredictable because the grid size $\sup_j |x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}| = O(\varepsilon) \approx \bigcirc$. On the other hand, the implementation of the new methods on such a computer remains normally functional, because the grid size $\sup_j |y_{j+\frac{1}{2}} - y_{j-\frac{1}{2}}| = O(\sqrt{\varepsilon}) \gg \varepsilon \approx \bigcirc$.

In light of results from Theorems 3.1, 3.4 and 3.5, as well as the above remarks, a class of new Godunov type numerical methods is formulated as follows:

Given an integer N in $\{1, 2, \ldots\}$ put $v_j^0 = u_j^0$. Then for for $n = 1, 2, \ldots$ put

$$v_j^{n+1} = v_j^n - \frac{\tau}{\Delta x_j}[f_{j+\frac{1}{2}}^n - f_{j-\frac{1}{2}}^n] \tag{4.12a}$$

$$f_{j+\frac{1}{2}}^n = f(R(0; v_{j,r}^n, v_{j+1,l}^n)) \tag{4.12b}$$

$$v_{j,l}^n = L_j(x_{j-\frac{1}{2}} + \frac{1}{2}\Delta_N) \tag{4.12c}$$

$$v_{j,r}^n = L_j(x_{j+\frac{1}{2}} - \frac{1}{2}\Delta_N) \tag{4.12d}$$

$$L_j(x) = s_j(x - x_j) + v_j^n \tag{4.12e}$$

$$\Delta_N = \frac{x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}}{N} \ (= \frac{h}{N} \text{inuniformgrids}) \tag{4.12f}$$

$$s_j = \frac{v^n_{j+1} - v^n_j}{x_{j+1} - x_j} \text{ if } |\frac{v^n_{j+1} - v^n_j}{x_{j+1} - x_j}| < |\frac{v^n_j - v^n_{j-1}}{x_j - x_{j-1}}| \tag{4.12g}$$

$$= \frac{v^n_j - v^n_{j-1}}{x_j - x_{j-1}} \text{ if } |\frac{v^n_j - v^n_{j-1}}{x_j - x_{j-1}}| < |\frac{v^n_{j+1} - v^n_j}{x_{j+1} - x_j}| \tag{4.12h}$$

$$= 0 \text{ if } v^n_j \text{ is a local extrema among } v^n_j \text{and } v^n_{j\pm1} \tag{4.12i}$$

$$\max_j \frac{\tau}{\Delta x_j} \le \frac{\zeta}{N}\frac{\theta}{\Lambda} \ (\frac{\tau}{g} \le \frac{\zeta}{N}\frac{\theta}{\Lambda} \text{ in uniform grids}) \tag{4.12j}$$

$$\zeta = \text{constant} \ge 1. \tag{4.12k}$$

Notice that (4.12) becomes (2.13) when $N = 1$ and $\zeta = 1$. One can also see that (4.12) can be used on non–uniform grids without much difficulty.

The above algorithms can be applied to systems of conservation laws with just minor modification in (4.12). That is, slopes of all components of $u(x,t)$ must be set to zero when a local extrema occurs in any ONE component of $u(x,t)$ in interval I_j. That is to ensure that Jensen's inequality can be used for $\{v^n_j\}$ as described in (2.15) to obtain entropy consistency.

The following theorems summarize the conclusions of this paper.

Theorem 4.1. *For scalar conservation laws, a numerical method defined in (4.12) is TV stable for* $\zeta = 1$.

Theorem 4.2. *A numerical method defined in (4.12) is entropy consistent for* $\zeta = 1$.

Theorem 4.3. *For a scalar conservation law and* $\zeta = 1$, *a numerical method defined in (4.12) produces numerical solution with a convergent subsequence whose limit function is the unique weak solution of the scalar conservation law.*

Proof of Theorem 4.1 Steps $(4.12g - i)$ ensure that no new local extrema is created when piecewise constant functions are constructed. That means the process of constructing piecewise constant functions is total variation non–increasing. As indicated earlier, (4.12) produces $\{v^{n+1}_j\}$ from $\{v^n_j\}$ by first solving Riemann problems, then averaging solutions of those from Riemann problems. The total variation of a solution of a Riemann problem $R((x-x_0)/(t-t_0); u_l, u_r)$ equals $|u_r - u_l|$ for a scalar conservation law. Because interaction of Riemann solutions is not allowed due to $(4.12k)$, the process of resolving Riemann problems is total variation non–increasing. The process of averaging is also total variation non–increasing. Therefore, (4.12) is TV stable. ♢

Proof of Theorem 4.2 The argument is the same as the one used to prove that Godunov method is entropy consistent [(2.15–17)], because just like Godunov method, (4.12) also uses piecewise constant functions for approximations. ♢

Proof of Theorem 4.3 It follows directly from Theorems 4.1, 4.2, 2.1, 2.2. ♢

5. Numerical examples

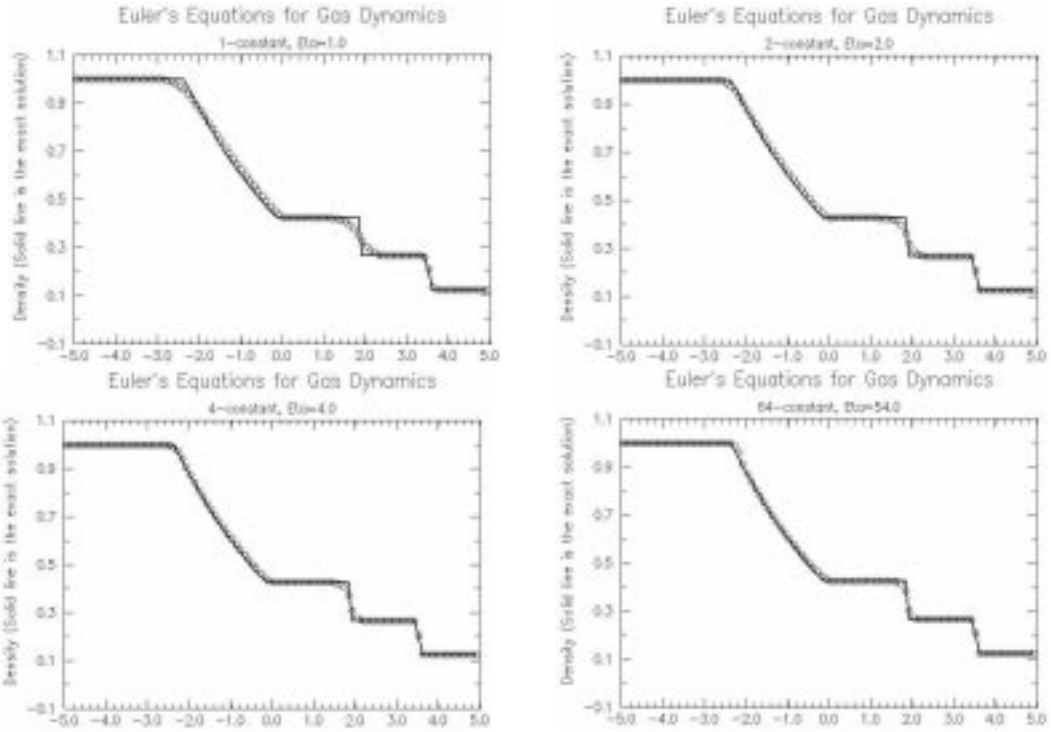

Fig. 5.1. Density profiles of test #1 (solid lines are the exact solutions)

Two tests of a system of conservation laws are presented here, where their exact solutions are available for error analysis. They are for a system of conservation laws which is the Euler's equations for gas dynamics in the form of (1.1) where $u = (\rho, \rho q, e)^T$, $f(u) = (\rho q, \rho q^2 + p, q(e + p))^T$. Here, ρ is density, p is pressure, q is velocity, $e = \rho\epsilon + \frac{1}{2}\rho q^2$ is the total energy per unit volume, and $\epsilon = p/[(\gamma - 1)\rho]$ is the internal energy per unit mass, where γ is the ratio of specific heats (a constant greater than 1).

The first test problem [Sod, 1978] has the initial values:

$$(q, p, \rho) = \begin{cases} (0, 1, 1), & \text{if } x < 0, \\ (0, 0.1, 0.125), & \text{if } x > 0. \end{cases}$$

The solution to this problem consists of one rarefaction wave traveling to the left, one shock wave traveling to the right, and a contact discontinuity staying in between. The density profile in the exact solution is monotone decreasing where there is a small (weak) jump in density at the contact discontinuity. Numerical solution at time $T = 2$ is reported. Density profiles are shown in figures 5.1. The number of constants used in approximation is indicated in the figures where when ONE constant is used for approximation, it is the first order Godunov method.

The second test problem [Harten, Engquist, Osher and Chakravarthy, 1987] has initial values:

$$(q, p, \rho) = \begin{cases} (0.698, 3.528, 0.445), & \text{if } x < 0, \\ (0, 0.571, 0.5), & \text{if } x > 0. \end{cases}$$

The solution to this problem consists of one rarefaction wave traveling to the left, one shock wave traveling to the right, and a contact discontinuity staying in between.

Table 5.1. Error Chart for Test Problem #1 ($dhx = 0.05000$)			
in $\mathcal{L}^1$ Norm			
# of Constants used	Norm	Absolute Error	Relative Error
1	21.16886	0.36453	1.72201%
2	21.16886	0.18274	0.86327%
4	21.16886	0.09882	0.46684%
64	21.16886	0.09305	0.43958%
2nd Order Method	*21.16886*	*0.13497*	*0.63759%*
in $\mathcal{L}^2$ Norm			
1	5.50185	0.14527	2.64044%
2	5.50185	0.08472	1.53980%
4	5.50185	0.05116	0.92979%
64	5.50185	0.05623	1.02203%
2nd Order Method	*5.50185*	*0.06356*	*1.15518%*
in $\mathcal{L}^\infty$ Norm			
1	3.50000	0.38778	11.07947%
2	3.50000	0.26011	7.43167%
4	3.50000	0.17894	5.11253%
64	3.50000	0.23941	6.84039%
2nd Order Method	*3.50000*	*0.17336*	*4.95301%*

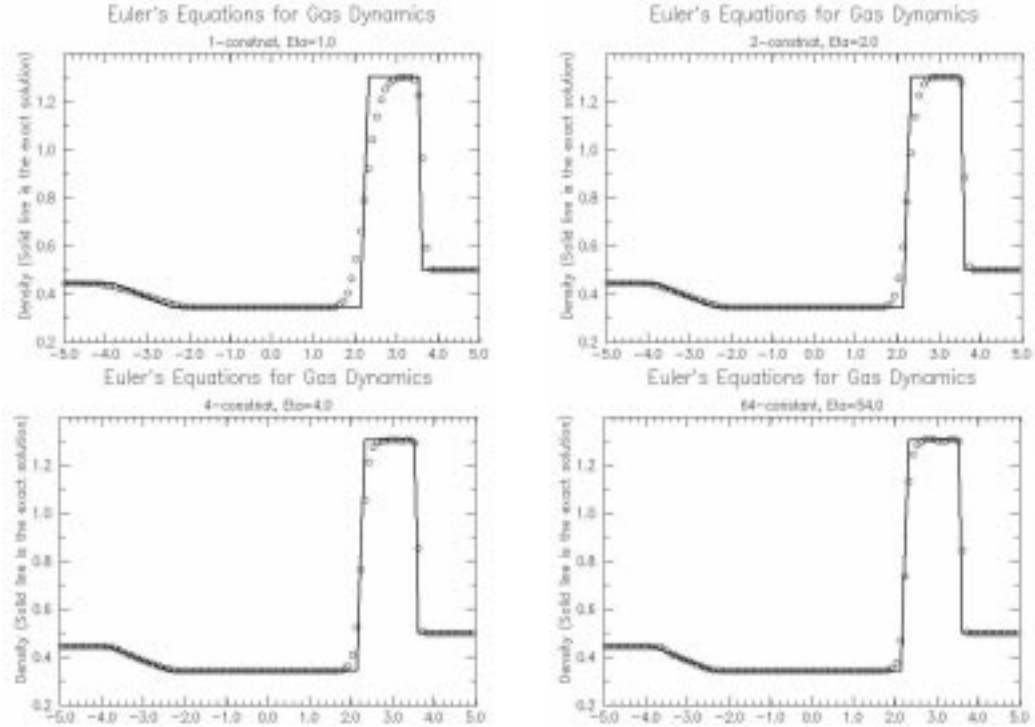

Fig. 5.2. Density profiles of test #2 (solid lines are the exact solutions)

Different from the previous test problem, the density profile of this problem has a built–up in the center part of the solution. Therefore, this test problem provides a different scenario to test all numerical methods. Numerical solution at time $T = 1.445$ is reported. Density profiles are shown in figures 5.2, and pressure profiles are shown in figures 5.3. The number of constants used in approximation is indicated in

Table 5.2. Error Chart for Test Problem #2 ($dhx = 0.05000$)			
in $\mathcal{L}^1$ Norm			
# of Constants used	Norm	Absolute Error	Relative Error
1	75.84542	1.77505	2.34035%
2	75.84542	1.02258	1.34825%
4	75.84542	0.71428	0.94176%
64	75.84542	0.67377	0.88834%
2nd Order Method	*75.84542*	*1.29140*	*1.70276%*
in $\mathcal{L}^2$ Norm			
1	21.77542	1.02382	4.70173%
2	21.77542	0.74414	3.41734%
4	21.77542	0.64455	2.96001%
64	21.77542	0.61964	2.84561%
2nd Order Method	*21.77542*	*0.81321*	*3.73455%*
in $\mathcal{L}^\infty$ Norm			
1	10.98673	3.35377	30.52561%
2	10.98673	3.09033	28.12782%
4	10.98673	3.01107	27.40644%
64	10.98673	2.93118	26.67929%
2nd Order Method	*10.98673*	*3.91365*	*35.62160%*

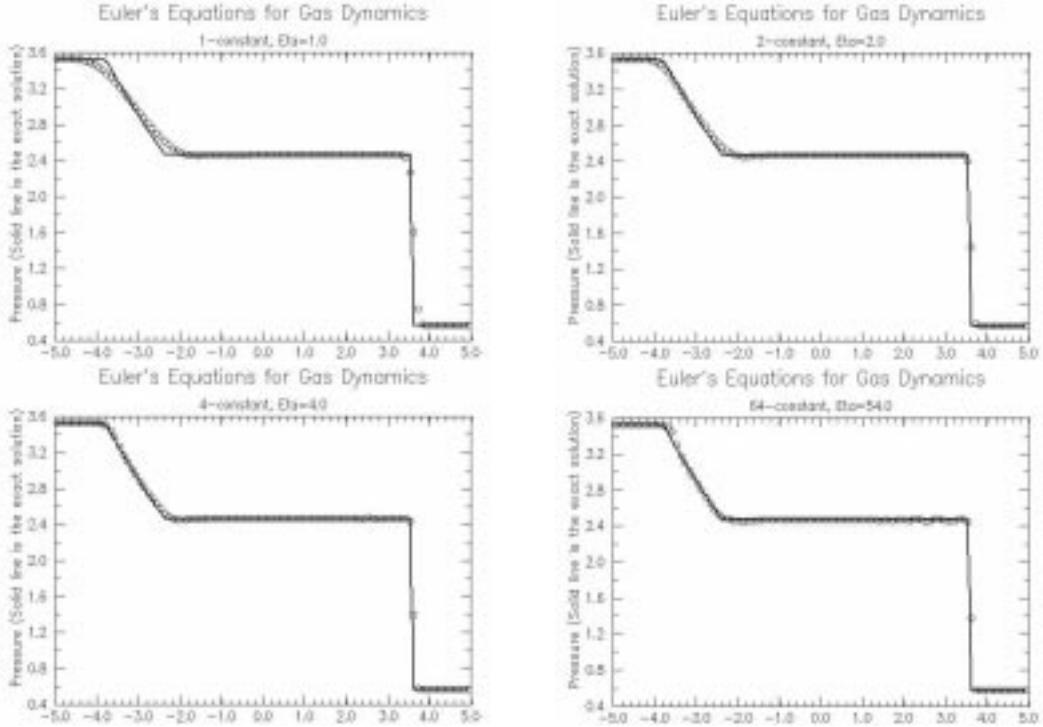

Fig. 5.3. Pressure profiles of test #2 (solid lines are the exact solutions)

the figures where when ONE constant is used for approximation, it is the first order Godunov method.

The above profiles confirm that the newly developed methods indeed are able to produce numerical solutions with better resolutions than those produced by a

Table 5.3. Time Comparison Chart for Test Problem #1			
Relative $\mathcal{L}^1$ error $\leq 1.0\%$			
# of Constants used	# of grids used	Ratio N/ζ	CPU Time
1	500	1	14.0
2	200	1	4.1
4	120	1	1.7
8	120	1	1.7
16	120	16/14	2.2
64	120	64/54	2.2
2nd Order Method	*120*	*N/A*	*2.0*

Table 5.4. Time Comparison Chart for Test Problem #1			
Relative $\mathcal{L}^2$ error $\leq 2.0\%$			
# of Constants used	# of grids used	Ratio N/ζ	CPU Time
1	410	1	9.9
2	140	1	2.0
4	100	1	1.1
8	100	1	1.2
16	100	16/14	1.4
64	80	64/54	0.9
2nd Order Method	*120*	*N/A*	*2.0*

first order method. The improvement is most significant in regions where rarefaction waves exist. These profiles are comparable to those presented in the paper by Harten, Engquist, Osher and Chakravarthy (ENO) [1987].

Errors between the numerical solution and the exact solution at the indicated times are computed in $\mathcal{L}^1$, $\mathcal{L}^2$ and $\mathcal{L}^\infty$ norms. They are reported in table 5.1 and 5.2 for the two test problems. Similar errors from a particular implementation of a second order Godunov–type method [Li, 1994] are also reported in the tables for comparison purpose. It can be seen that the current methods indeed produce numerical solutions comparable to those produced by higher order methods which is also apparent from the pictures in fig. 5.1–5.3. The results are much better than those produced by a first order conservative method.

The above test problems were run on an IBM RS/6000 model 550 computer running AIX which is highly efficient in floating point number computations. Table 5.3 records the CPU time used by the different methods which are required to produce numerical solutions with a relative $\mathcal{L}^1$ error less than or equal to 1.0%, while Table 5.4 records the CPU time used by the different methods which are required to produce numerical solutions with a relative $\mathcal{L}^2$ error less than or equal to 2.0%. One concludes from the two tables that **(a)**, to achieve a pre–determined accuracy, the new methods are much more efficient than the first order Godunov method in terms of CPU time and number of grid points used, and they are comparable to a particular high order accurate Godunov–type method, though the exact savings in CPU time are computer–dependent and norm–dependent; **(b)**, using a pre–chosen number of grid points for numerical approximations, the new methods produce numerical solutions with much better accuracy than the one produced by the first order Godunov method, and they are also comparable to the one produced by a particular

high order Godunov–type method.

A major problem that hinders the efficiency of the new methods is the restriction in $(4.12j)$, where a time step must be chosen to be smaller than the one used in other Godunov–type methods as indicated in $(2.9c)$. Otherwise the savings would have been much greater. The author [Li, 1998] is working on an implicit version of those new methods which will ease up the strict restriction in $(4.12j)$, resulting in greater savings over other Godunov–type methods. Implementation of the current new methods to conservation laws in multiple space dimensions may also be investigated in the future.

Acknowledgment The author thanks the referees for suggesting the inclusion of comparison on the computing time among different methods which greatly enhanced the quality of this manuscript.

References

[1] Colella, P. and P. R. Woodward (1985). The Piecewise-Parabolic Method(PPM) for Gas-Dynamics, *J. Comp. Physics*, 54, 174–201.

[2] Godunov, S. (1959). Finite Difference Method For Numerical Computation of Discontinuous Solutions of the Equations of Fluid Dynamics, *Mat. Sbornik*, 47(89), Number 3, page 271.

[3] Harten, A. (1983). On the Symmetric Form of Systems of Conservation Laws with Entropy, *J. Comp. Physics*, 49, 151–164.

[4] Harten, A., P. D. Lax and B. Van Leer (1983). On Upstream Differencing and Godunov–type Schemes for Hyperbolic Conservation Laws, *SIAM Review*, 25, 35–61.

[5] Harten, A. (1984). On a Class of High Resolution Total–Variation–Stable Finite–Difference Schemes, *SIAM J. Numer. Anal.*, 21, 1–23.

[6] Harten, A., B. Engquist, S. Osher and S. Chakravarthy (1987). Uniformly High Order Accurate Essentially Non-oscillatory Schemes, III, *J. Comp. Physics*, 71, 231–303.

[7] Lax, P. D. (1973). *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*, Society for Industrial and Applied Mathematics.

[8] Lax, P. D., B. Wendroff (1960). Systems of Conservation Laws, *Comm. Pure Appl. Math.*, 13, 217–237.

[9] Li, X. (1994). A Numerical Method for Systems of Hyperbolic Conservation Laws with Single Stencil Reconstructions, *Applied Mathematics and Computation*, 65, 125–140.

[10] Li, X. (1998). Entropy Consistent, Implicit TVD Methods with High Order Accuracy for Conservation Laws, *working paper*.

[11] Oleinik, O. A. (1957). On Discontinuous Solutions of Nonlinear Differential Equations, Uspekhi Mat. Nauk, Vol. 12, pp.3–73. English translation, Amer. Math. Soc. Trans., Ser. 2, No. 26, pp. 95–172.

[12] Sod, G. A. (1978). A Survey of Several Finite Difference Methods for Systems

of Nonlinear Hyperbolic Conservation Laws, *J. Comp. Physics*, 27, 1–31.

[13] Van Leer, B. (1979). Towards the Ultimate Conservative Differences Scheme, V. A Second Order Sequel to Godunov's Methods, *J. Comp. Physics*, 32, 101–136.

[14] Vila, J. P. (1989). An Analysis of a Class of Second–Order Accurate Godunov–type Schemes, *SIAM J. Numer. Anal.*, 26, 830–853.

Xuefeng Li
Department of Mathematics and Computer Science, Loyola University
6363 St. Charles Avenue, New Orleans, LA 70118, USA.
E-mail address: Li@Loyno.edu