A Uniformly Third Order Accurate Scheme for Genuinely Nonlinear Conservation Laws

Xu-Dong Liu* Eitan Tadmor† Tamir Tassa‡

February 25, 2004

Abstract

The entropy condition and the total variation boundedness of weak solutions of convex scalar conservation laws are enforced by Lip^+-stability which the physical solutions satisfy. The first order Godunov and Lax-Friedrichs schemes, and the second order Maxmod scheme are consistent with this Lip^+-stability and are, therefore, entropy convergent. In this paper, a uniformly third order accurate scheme is introduced, which is consistent with the Lip^+-stability and, hence, is entropy convergent. Error estimates, both global and local, are obtained. Numerical experiments on systems of conservation laws demonstrate excellent results and sharp resolution of shock discontinuities. A Fortran subroutine of the reconstruction procedure is spelled out in the end of the paper.

*Courant Institute of Mathematical Sciences, 251 Mercer Street, NY, NY 10012; xliu@cims.nyu.edu. This author was supported by NSF Grant DMS-9112654 and ONR Grant #N00014-91-J-1034.

†School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978 Israel; tadmor@math.tau.ac.il. This author was supported by ONR Grant #N0014-91-J-1343.

‡Department of Mathematics, University of California, Los Angeles, CA 90024; tassa@math.ucla.edu. This author was supported by ONR Grant #N00014-92-J-1890.

Keywords: Hyperbolic conservation laws, high order accuracy, Lip^+-stability, TVB, entropy.

AMS(MUS) subject classification: Primary 65M10; Secondary 65M05.

Proofs sent to: Tamir Tassa, School of Mathematical Sciences, Tel Aviv University, Tel-Aviv 69978 Israel.

1 Introduction

In this paper we introduce a uniformly third order accurate Godunov-type scheme for the approximate solution of genuinely nonlinear scalar conservation laws,

$$u_t + f(u)_x = 0, \qquad f''(u) \neq 0 \ . \tag{1.1}$$

We prove the convergence of the scheme to the physically relevant entropy solutions, obtain error estimates – global, as well as local ones – and demonstrate the performance of the scheme in both scalar and system problems.

In §2 we review briefly the main results concerning the one-sided-Lipschitz-stability of genuinely nonlinear conservation laws and Godunov-type schemes. This type of stability, which in the convex case is called Lip^+-stability, identifies the entropy solutions of (1.1) and, therefore, guarantees uniqueness. A similar Lip^+-consistency, if obeyed by a Godunov-type scheme, implies the total variation boundedness of the scheme (and, hence, the existence of a convergent subsequence) as well as its entropy consistency (which implies the convergence of the whole sequence of approximate solutions to the unique exact entropy solution). In other words, establishing the Lip^+-consistency of a scheme is equivalent to the usual two-step recipe of convergence proofs – total variation boundedness and entropy-consistency.

It is shown that the Lip^+-consistency of a Godunov-type scheme depends solely on the properties of the associated reconstruction procedure. Proposition 2.3 provides us with two easy-to-check conditions on the reconstruction which, if satisfied, imply the Lip^+-consistency of the scheme and, consequently, its convergence.

With this in mind, we proceed to §3 where we describe a reconstruction procedure for the convex case that satisfies those conditions (a similar version of this reconstruction exists in the concave case). This reconstruction uses, as its building blocks, quadratic interpolants. Such interpolants have two advantages over linear ones, as used in the so-called MUSCL schemes. First, using quadratic polynomials we obtain third order accuracy in the formal sense, i.e., whenever the solution is C^3-smooth. The second advantage is manifested in smaller spurious oscillations: as Lip^+-consistency prohibits the piecewise polynomial reconstruction to produce increasing jump discontinuities, using rigid linear interpolants results in strong under- and overshoots. Replacing the linear interpolants by quadratic ones, we gain flexibility and significantly reduce the the inevitable under- and overshoots.

In §4 we concentrate on scalar convex conservation laws and, based on the convergence rate analysis for Godunov-type schemes that was introduced in [9], obtain global and local error estimates for our scheme.

Finally, §5 is devoted to numerical experiments. In §5.1 we test the performance of our scheme on scalar convex problems. We provide error tables that reflect the high order accuracy of the scheme, as well as graphs of the approximate and exact solutions that demonstrate the sharp resolution of shocks. In §5.2 we deal with Euler equations of gas dynamics. This is a 3-system in which one field is convex, another is concave and the third one is neither. In order to resolve the genuinely nonlinear fields, we use either the convex or the concave version of our scheme, while the last field is resolved by using the third order non-oscillatory reconstruction procedure that was introduced in [8].

2 An overview of Godunov-type schemes and their Lip^+-stability

In this section we review some of the basic results concerning the Lip^+-stability of scalar convex conservation laws,

$$u_t + f(u)_x = 0, \qquad f''(u) \geq \alpha > 0 \ , \tag{2.1}$$

and their numerical approximate solutions. It is well known [1, 10] that the physically relevant weak solutions of (2.1) are identified by the following Lip^+-stability,

$$\|u(\cdot,t)\|_{Lip^+} \leq (\|u(\cdot,0)\|^{-1}_{Lip^+} + \alpha t)^{-1} \quad \forall t \geq 0 \tag{2.2}$$

where

$$\|v\|_{Lip^+} := \sup_{x \neq y} \left(\frac{v(x) - v(y)}{x - y} \right)_+ . \tag{2.3}$$

In other words, this Lip^+-stability may serve in the convex case as an alternative entropy condition in order to single out the admissible entropy solutions of (2.1). This property implies that $\|u(\cdot,t)\|_{Lip^+}$ is always finite at $t > 0$ and bounded by $\frac{1}{\alpha t}$, even if the initial value is Lip^+-unbounded. Hence, any increasing discontinuity in the initial data is immediately spread out and becomes a rarefaction wave. Only decreasing jump discontinuities – shock waves – may appear at $t > 0$. Therefore, the Lax entropy condition for convex conservation laws [7] is enforced by the Lip^+-stability. Another crucial implication of the Lip^+-stability is the Total Variation Boundedness (TVB) of the solution at $t > 0$ [1].

Let $u^h(x,t)$ be an approximate solution of (2.1) obtained by a numerical scheme on the grid $\{I_j\} \otimes \{T_n\}$, where $\{I_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]\}_j$ is an h-uniform partition of $\mathcal{R}_x$, $h = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$, and $\{T_n = [t_n, t_{n+1}]\}_{n \geq 0}$ is a Δt-uniform partition of $\mathcal{R}^+_t$, $\Delta t = t_{n+1} - t_n$.

Letting $\bar{u}^n_j$ denote the cell-average, $\bar{u}^n_j := \frac{1}{h} \int\limits_{I_j} u^h(x, t_n)\, dx$, we define the *discrete* Lip^+*-seminorm* of $u^h(\cdot, t^n)$ as

$$p^n = \sup_j \left(\frac{\bar{u}^n_{j+1} - \bar{u}^n_j}{h} \right)_+ . \tag{2.4}$$

The numerical scheme is called Lip^+-consistent if

$$p^n \leq (\|u^h(\cdot,0)\|^{-1}_{Lip^+} + \alpha t_n)^{-1}. \tag{2.5}$$

This discrete Lip^+-decay condition is the analogous of (2.2). Some numerical schemes fail to satisfy (2.5) and satisfy the weaker estimate

$$p^n \leq \|u^h(\cdot,0)\|_{Lip^+}. \tag{2.6}$$

Such schemes are called weakly Lip^+-consistent.

The following Proposition asserts that the Lip^+-consistency of a scheme implies its convergence to the entropy solution (see [1] for a proof).

Proposition 2.1 *Let the initial data $u(x,0)$ be either periodic or constant for large values of $|x|$ and assume that one of the following conditions hold:*

(i) the scheme is Lip^+-consistent, has a uniformly bounded speed of propagation and $u(x,0)$ has a finite number of increasing jumps; or

(ii) the scheme is weakly Lip^+-consistent and $\|u(\cdot,0)\|_{Lip^+} < \infty$.

Then the scheme is TVB, entropy-consistent and, therefore, converges to the exact entropy solution.

In the remainder of this section we concentrate on Godunov-type schemes [17],

$$u^h(\cdot,t^{n+1}) = E(\Delta t)RAu^h(\cdot,t_n), \quad n \geq 0, \tag{2.7}$$

where A stands for the averaging operator,

$$(Av)(x) = \frac{1}{h}\int_{I_j} v(\xi)d\xi \quad \forall x \in I_j \ , \tag{2.8}$$

R is a piecewise-polynomial reconstruction of the cell averages and $E(\Delta t)$ is the exact evolution operator. We shall use henceforth the notation $R(\cdot,t_n)$ for the reconstructed solution at time t_n, i.e., $R(\cdot,t_n) = RAu^h(\cdot,t_n)$.

In the following proposition it is shown that if the reconstruction R does not increase the discrete Lip^+-seminorm, the scheme is Lip^+-consistent (see also [1, Proposition 4] and [4]).

Proposition 2.2 *A Godunov-type scheme is Lip^+-consistent if*

$$\|R(\cdot,t_n)\|_{Lip^+} \leq p^n = \sup_j \left(\frac{\bar{u}^n_{j+1} - \bar{u}^n_j}{h}\right)_+ . \tag{2.9}$$

Proof. The proof is based on two simple observations:
(a) the exact evolution operator $E(\Delta t)$ satisfies, in view of (2.2),

$$\|E(\Delta t)R(\cdot,t_n)\|_{Lip^+} \le (\|R(\cdot,t_n)\|^{-1}_{Lip^+} + \alpha\Delta t)^{-1} \ ; \tag{2.10}$$

(b) the discrete Lip^+-seminorm, (2.4), is bounded by the regular one,

$$p^n \le \|u^h(\cdot,t_n)\|_{Lip^+} \ . \tag{2.11}$$

Using (2.10), (2.9) and (2.11), in this order, we conclude by (2.7) that

$$\begin{aligned} \|u^h(\cdot,t_{n+1})\|_{Lip^+} = \|E(\Delta t)R(\cdot,t_n)\|_{Lip^+} \le (\|R(\cdot,t_n)\|^{-1}_{Lip^+} + \alpha\Delta t)^{-1} \le \\ \le ((p^n)^{-1} + \alpha\Delta t)^{-1} \le (\|u^h(\cdot,t_n)\|^{-1}_{Lip^+} + \alpha\Delta t)^{-1} \ . \end{aligned} \tag{2.12}$$

Finally, inequality (2.12) implies, by induction on n, the desired Lip^+-consistency,

$$p^n \le \|u^h(\cdot,t_n)\|_{Lip^+} \le (\|u^h(\cdot,0)\|^{-1}_{Lip^+} + \alpha t_n)^{-1} \ .$$

□

The main idea of Proposition 2.2 is that the Lip^+-consistency of a Godunov-type scheme depends solely on the properties of its associated reconstruction operator, R. Since any Godunov-type scheme has a finite speed of propagation (as $E(\Delta t)$ does and R is a *local* operator), this Lip^+-consistency implies, in view of Proposition 2.1, the convergence of the scheme to the exact entropy solution.

We conclude this section with a straightforward proposition that provides us with two easy-to-verify conditions on R that are equivalent to the Lip^+-stability condition (2.9):

Proposition 2.3 *If the reconstructed solution at $t = t_n$, $R(\cdot,t_n)$, has only decreasing jump discontinuities,*

$$R(x_{j+\frac{1}{2}}-,t_n) \ge R(x_{j+\frac{1}{2}}+,t_n) \qquad \forall j \tag{2.13}$$

and

$$\sup_{x_{j-\frac{1}{2}}<x<x_{j+\frac{1}{2}}} R_x(x,t_n) \le p^n \qquad \forall j \ , \tag{2.14}$$

then $R(\cdot,t_n)$ satisfies (2.9). Consequently, the corresponding Godunov-type scheme is Lip^+-consistent and, therefore, converges to the exact entropy solution.

As an example, we consider the maxmod scheme which is identified by the linear reconstruction

$$R(x,t_n)\Big|_{I_j} = \bar{u}^n_j + s^n_j \cdot (x - x_j) \qquad s^n_j = \frac{1}{h}\max(\bar{u}^n_{j+1} - \bar{u}^n_j, \bar{u}^n_j - \bar{u}^n_{j-1})$$

(x_j denotes henceforth the center of the cell I_j). This scheme is formally second order accurate and satisfies the Lip^+-stability conditions (2.13)–(2.14). However, it permits spurious under- and overshoots and, therefore, is not recommended for practical calculations.

We introduce below a simple third order reconstruction which, like the max-mode reconstruction, satisfies the Lip^+-stability conditions (2.13)–(2.14), but – owing to the use of quadratic interpolants rather than linear ones – creates significantly smaller under- and overshoots.

3 Third order reconstruction procedure

We start by describing the building blocks of our reconstruction: these are polynomials of degree 2 or less, which are conservative in the sense that they preserve the cell averages.

First, we introduce the quadratic polynomials, $P_j(x)$, that interpolate the cell averages $\bar{u}_{j-1}, \bar{u}_j$ and $\bar{u}_{j+1}$ (we omit henceforth the time superscript n) in the cell-average sense, i.e.,

$$\frac{1}{h}\int\limits_{I_i} P_j(x)\,dx = \bar{u}_i, \qquad i = j-1, j, j+1.$$

These polynomials take the explicit form

$$P_j(x) = c_0 + c_1(x - x_j) + c_2(x - x_j)^2 \ , \tag{3.1}$$

where

$$c_0 = \bar{u}_j - \frac{\Delta^2 \bar{u}_j}{24} \quad , \quad c_1 = \frac{\Delta \bar{u}_{j-1} + \Delta \bar{u}_j}{2h} \quad , \quad c_2 = \frac{\Delta^2 \bar{u}_j}{2h^2} \ ,$$

and $\Delta \bar{u}_j = \bar{u}_{j+1} - \bar{u}_j$, $\Delta^2 \bar{u}_j = \bar{u}_{j+1} - 2\bar{u}_j + \bar{u}_{j-1}$. In addition to that, we define the linear polynomials

$$L_j^-(x) = \bar{u}_j + \frac{\Delta \bar{u}_{j-1}}{h}(x - x_j) \quad \text{and} \quad L_j^+(x) = \bar{u}_j + \frac{\Delta \bar{u}_j}{h}(x - x_j) \ . \tag{3.2}$$

We are looking for a reconstruction, $R_j(x) = R(x)\Big|_{I_j}$, based on the above building blocks, $P_j(x)$ and $L_j^{\pm}(x)$, that has the following desired properties:

($\mathcal{P}1$) Conservation, i.e., $\int_{I_j} R_j(x)dx = \bar{u}_j$ for all j;

($\mathcal{P}2$) Third order (formal) accuracy, i.e., $R_j(x) - u^h(x) = \mathcal{O}(h^3)$, whenever $u^h(x)$ is C^3-smooth;

($\mathcal{P}3$) Lip^+-stability, namely, conditions (2.13)–(2.14) must be satisfied.

The conservation requirement ($\mathcal{P}1$) dictates the following set of acceptable values for $R_j(x)$: $P_i(x)$, $j-1 \le i \le j+1$ or $L_j^{\pm}(x)$. The accuracy requirement ($\mathcal{P}2$) narrows down our choices of acceptable reconstructions on I_j to only $P_i(x)$,

$j-1 \leq i \leq j+1$, which satisfy $P_i(x) - u^h(x)\Big|_{I_j} = \mathcal{O}(h^3)$, unless u^h_{xx} vanishes in the neighborhood of I_j, in which case also $L^{\pm}_j(x)$ are locally third order accurate and, hence, acceptable. Therefore, we need to use the remaining degrees of freedom in order to enforce the *Lip*$^+$-stability conditions, (2.13)–(2.14). The resulting reconstruction is described below:

Case 1. If $\Delta^2 \bar{u}_j \leq 0$ and $\Delta^2 \bar{u}_{j-1} \leq 0$,

$$R_j(x) = \begin{cases} P_j(x) & \text{if } P_{j-1}(x_{j-\frac{1}{2}}) \geq P_j(x_{j-\frac{1}{2}}) \,, \\ P_{j-1}(x) & \text{otherwise.} \end{cases} \tag{3.3}$$

Case 2. If $\Delta^2 \bar{u}_j \leq 0$ and $\Delta^2 \bar{u}_{j-1} \geq 0$,

$$R_j(x) = L^-_j(x) \quad \text{and} \quad R_{j-1}(x) = L^+_{j-1}(x) \,. \tag{3.4}$$

Case 3. If $\Delta^2 \bar{u}_j \geq 0$ and $\Delta^2 \bar{u}_{j+1} \geq 0$,

$$R_j(x) = \begin{cases} P_j(x) & \text{if } P_{j+1}(x_{j-\frac{1}{2}}) \geq P_j(x_{j-\frac{1}{2}}) \,, \\ P_{j+1}(x) & \text{otherwise.} \end{cases} \tag{3.5}$$

Case 4. If $\Delta^2 \bar{u}_j \geq 0$ and $\Delta^2 \bar{u}_{j+1} \leq 0$,

$$R_j(x) = L^+_j(x) \quad \text{and} \quad R_{j+1}(x) = L^-_{j+1}(x) \,. \tag{3.6}$$

Proposition 3.1 *The above described reconstruction satisfies the Lip$^+$-stability conditions (2.13) and (2.14).*

The proof of Proposition is given below. Next, we may state the main result of this section that follows directly from Propositions 2.3 and 3.

Corollary 3.1 *The formally third order accurate Godunov-type scheme that corresponds to the above reconstruction is Lip$^+$-consistent in the convex case and converges to the exact entropy solution.*

Remark. It is straightforward to figure out the reconstruction procedure for concave scalar conservation laws ($f''(u) \leq -\alpha < 0$), for which the one-sided-Lipschitz-stability takes the form (compare to (2.2))

$$\|u(\cdot,t)\|_{Lip^-} \leq (\|u(\cdot,0)\|^{-1}_{Lip^-} + \alpha t)^{-1} \quad \forall t \geq 0,$$

where

$$\|v\|_{Lip^-} = \sup_{x \neq y} \left(\frac{v(x) - v(y)}{x - y} \right)_- , \qquad a_- = -\min(a, 0).$$

Proof of Proposition 3. We observe that

$$P_j'(x) \leq \frac{\Delta \bar{u}_{j-1}}{h} \quad \forall x \in [x_{j-\frac{1}{2}}, +\infty) \quad \text{if } \Delta^2 \bar{u}_j \leq 0 \tag{3.7}$$

and

$$P_j'(x) \leq \frac{\Delta \bar{u}_{j+1}}{h} \quad \forall x \in (-\infty, x_{j+\frac{1}{2}}] \quad \text{if } \Delta^2 \bar{u}_j \geq 0 \ . \tag{3.8}$$

The above inequalities enable us to prove that (2.14) holds in each of the four cases: In *Case 1*, (3.7) implies that

$$P_{j-1}'(x) \leq \frac{\Delta \bar{u}_{j-2}}{h} \quad \text{and} \quad P_j'(x) \leq \frac{\Delta \bar{u}_{j-1}}{h} \qquad \forall x \in I_j \ .$$

Consequently,

$$R_j'(x) \leq \frac{1}{h} \max(\Delta \bar{u}_{j-2}, \Delta \bar{u}_{j-1}) \leq p^n \ .$$

In *Case 2*,

$$R_j'(x) := \frac{\Delta \bar{u}_{j-1}}{h} \leq p^n \ .$$

Cases 3 and *4* are treated similarly. Therefore, our reconstruction satisfies condition (2.14).

Hence, it remains to prove (2.13). We start by obtaining inequalities for the reconstruction R_j at the end points of the cell in which it is defined, $x_{j-\frac{1}{2}}$ and $x_{j+\frac{1}{2}}$. In *Case 1* R_j is chosen to be either P_{j-1} or P_j, whichever attains the smaller value at the left end point $x_{j-\frac{1}{2}}$, (3.3). Hence,

$$R_j(x_{j-\frac{1}{2}}) \leq \min(P_{j-1}(x_{j-\frac{1}{2}}), P_j(x_{j-\frac{1}{2}})) \ . \tag{3.9}$$

Since it may be verified that in this case

$$P_{j-1}(x_{j-\frac{1}{2}}) \leq P_j(x_{j-\frac{1}{2}}) \quad \text{iff} \quad P_{j-1}(x_{j+\frac{1}{2}}) \geq P_j(x_{j+\frac{1}{2}}) \ ,$$

we conclude that

$$R_j(x_{j+\frac{1}{2}}) \geq \max(P_{j-1}(x_{j+\frac{1}{2}}), P_j(x_{j+\frac{1}{2}})) \ . \tag{3.10}$$

Case 3 is similar. Here, R_j is chosen to be either P_j or P_{j+1}, whichever attains the smaller value at the left end point $x_{j-\frac{1}{2}}$, (3.5). Hence,

$$R_j(x_{j-\frac{1}{2}}) \leq \min(P_j(x_{j-\frac{1}{2}}), P_{j+1}(x_{j-\frac{1}{2}})) \ . \tag{3.11}$$

Since in this case

$$P_j(x_{j-\frac{1}{2}}) \leq P_{j+1}(x_{j-\frac{1}{2}}) \quad \text{iff} \quad P_j(x_{j+\frac{1}{2}}) \geq P_{j+1}(x_{j+\frac{1}{2}}) \ ,$$

we conclude that

$$R_j(x_{j+\frac{1}{2}}) \geq \max(P_j(x_{j+\frac{1}{2}}), P_{j+1}(x_{j+\frac{1}{2}})) \ . \tag{3.12}$$

In *Case 2* it is easy to see that

$$R_j(x_{j-\frac{1}{2}}) = R_{j-1}(x_{j-\frac{1}{2}}) \ , \tag{3.13}$$

while, since $\Delta^2 \bar{u}_j \le 0$,

$$R_j(x_{j+\frac{1}{2}}) \ge P_j(x_{j+\frac{1}{2}}) \ . \tag{3.14}$$

Finally, in *Case 4*,

$$R_j(x_{j-\frac{1}{2}}) \le P_j(x_{j-\frac{1}{2}}) \tag{3.15}$$

because $\Delta^2 \bar{u}_j \ge 0$, and

$$R_j(x_{j+\frac{1}{2}}) = R_{j+1}(x_{j+\frac{1}{2}}) \ . \tag{3.16}$$

We claim that the desired inequality (2.13) now follows from (3.9)–(3.16). In order to show that, we concentrate on two adjacent intervals, say I_j and I_{j+1}, and prove that

$$R_j(x_{j+\frac{1}{2}}) \ge R_{j+1}(x_{j+\frac{1}{2}}) \ . \tag{3.17}$$

There are nine possible combinations of cases in the two intervals:

◇ *1 & 1* (namely, in I_j we have *Case 1* and likewise in I_{j+1}). Here, (3.10) and (3.9) (for $j+1$) imply that

$$R_j(x_{j+\frac{1}{2}}) \ge P_j(x_{j+\frac{1}{2}}) \ge R_{j+1}(x_{j+\frac{1}{2}}) \ .$$

◇ *1 & 3*. In this case $\Delta^2 \bar{u}_j \le 0$ while $\Delta^2 \bar{u}_{j+1} \ge 0$. Since

$$P_j(x_{j+\frac{1}{2}}) = -\frac{\Delta^2 \bar{u}_j}{6} + \frac{\bar{u}_j + \bar{u}_{j+1}}{2} \quad \text{and} \quad P_{j+1}(x_{j+\frac{1}{2}}) = -\frac{\Delta^2 \bar{u}_{j+1}}{6} + \frac{\bar{u}_j + \bar{u}_{j+1}}{2} \ ,$$

it follows that

$$P_j(x_{j+\frac{1}{2}}) \ge P_{j+1}(x_{j+\frac{1}{2}}) \ . \tag{3.18}$$

Hence, by (3.10) and (3.11), we get the desired inequality (3.17):

$$R_j(x_{j+\frac{1}{2}}) \ge P_j(x_{j+\frac{1}{2}}) \ge P_{j+1}(x_{j+\frac{1}{2}}) \ge R_{j+1}(x_{j+\frac{1}{2}}) \ .$$

◇ *1 & 4*. Here, like in the previous case, (3.18) still holds and (3.17) follows from (3.10) and (3.15).

◇ *2 & 1*. We use (3.14) and (3.9).

◇ *2 & 3*. We use (3.14), (3.18) and (3.11).

◇ *2 & 4*. We use (3.14), (3.18) and (3.15).

◇ *3 & 3*. We use (3.12) and (3.11).

◇ *3 & 4*. We use (3.12) and (3.15).

◇ *4 & 2*. We use (3.16).

This proves (3.17) and, thus, completes the proof of the proposition. □

4 Convergence rate estimates for the scalar problem

In this section we obtain global and local convergence rate estimates for the scheme, based on the convergence rate analysis for Godunov type schemes in [9]. Let us recall briefly the main results of that paper:

Let $u = u(x,t)$ be the exact entropy solution of (2.1) subject to a rarefaction-free initial data, $\|u(\cdot,0)\|_{Lip^+} < \infty$, and let $u^h = u^h(x,t)$ be a Godunov-type approximate solution for that problem,

$$u^h(\cdot,t_{n+1}) = E(\Delta t)Pu^h(\cdot,t_n) \qquad n \geq 0 \ . \tag{4.1}$$

Here, P is some projection onto a subspace of piecewise h-grid functions and $E(\Delta t)$ denotes, as before, the exact evolution operator. Then if the scheme is weakly Lip^+-consistent, (2.6), and Lip'-consistent[1] in the sense that

$$\|Pv - v\|_{Lip^+} \leq O(h^2)\|v\|_{BV} \quad \forall v \in BV \ , \tag{4.2}$$

the following error estimates hold [9, Theorem 2.3]:

$$\|u^h(\cdot,t) - u(\cdot,t)\|_{W^{s,p}} \leq O(h^{\frac{1-sp}{2p}}) \ , \qquad -1 \leq s \leq \frac{1}{p} \ , \quad 1 \leq p \leq \infty \ . \tag{4.3}$$

In addition, we may post-process the approximate solution by means of mollification, $\tilde{u}^h := \psi_h * u^h$, so that wherever the exact solution is C^r-smooth, we have

$$\|\tilde{u}^h - u\|_{L^\infty_{loc}} \leq O(h^{\frac{r}{r+2}}) \ . \tag{4.4}$$

In other words, we may recover the exact value $u(x,t)$ from $u^h(\cdot,t)$ to within an error as close to $O(h)$ as the local smoothness permits. Since our scheme is formally third order accurate and the numerical evidence support this formal accuracy (see in §5), this local first order accuracy is not sharp. However, in the absence of a more exhaustive local convergence rate analysis, this is the best result we can furnish.

The projection P in our scheme takes the Godunov-type form $P = RA$ where $A = A_h$ is the cell averaging operator, (2.8), and R is the piecewise quadratic reconstruction procedure that was described in §3. In the following proposition we prove the Lip'-consistency of that projection:

Proposition 4.1 *The projection $P = RA$ satisfies estimate (4.2).*

Proof. We decompose the error term in (4.2) as follows:

$$\|Pv - v\|_{Lip^+} \leq \|R\bar{v} - \bar{v}\|_{Lip^+} + \|\bar{v} - v\|_{Lip^+} \quad , \quad \bar{v} := Av \ . \tag{4.5}$$

[1] The Lip'-norm is defined for functions w of zero average, $\int_{\mathcal{R}} w = 0$, as the L_2-dual norm of $\|\phi\|_{Lip} = \text{ess sup}_{x \neq y} \ |(\phi(x) - \phi(y))/(x-y)|$.

Since, in view of [9, Proposition 3.1],

$$\|\bar{v} - v\|_{Lip^+} \leq O(h^2)\|v\|_{BV} \ , \tag{4.6}$$

we only have to prove that

$$\|R\bar{v} - \bar{v}\|_{Lip^+} \leq O(h^2)\|v\|_{BV} \ . \tag{4.7}$$

To this end we recall that if w is a function for which the distance between two successive zeroes of its primitive $W(x) = \int_{-\infty}^{x} w(\xi)d\xi$ is h at the most, then

$$\|w\|_{Lip^+} \leq h \cdot \|w\|_{L^1} \tag{4.8}$$

[9, Lemma A.1]. Since R is a conservative reconstruction, $AR\bar{v} = \bar{v}$, the primitive of $R\bar{v} - \bar{v}$ vanishes at the grid points $x_{j+\frac{1}{2}}$ for all j. Hence, by (4.8),

$$\|R\bar{v} - \bar{v}\|_{Lip^+} \leq h \cdot \|R\bar{v} - \bar{v}\|_{L^1} \ . \tag{4.9}$$

In the cell I_j, $R\bar{v}$ may be either one of the quadratic polynomials $\{P_i(x) : \ j-1 \leq i \leq j+1\}$, (3.1), or one of the linear functions $L_j^{\pm}(x)$, (3.2). Considering each of these possibilities we conclude that

$$\|R\bar{v} - \bar{v}\|_{L_\infty(I_j)} \leq C \cdot (|\Delta\bar{v}_{j-2}| + |\Delta\bar{v}_{j-1}| + |\Delta\bar{v}_j| + |\Delta\bar{v}_{j+1}|) \tag{4.10}$$

for some constant C. Hence,

$$\|R\bar{v} - \bar{v}\|_{L^1} \leq h \cdot \sum_j \|R\bar{v} - \bar{v}\|_{L_\infty(I_j)} \leq 4Ch \sum_j |\Delta\bar{v}_j| = 4Ch\|\bar{v}\|_{BV} \ . \tag{4.11}$$

Finally, (4.9) and (4.11) imply that

$$\|R\bar{v} - \bar{v}\|_{Lip^+} \leq 4Ch^2\|\bar{v}\|_{BV} \leq O(h^2)\|v\|_{BV} \ . \tag{4.12}$$

This proves (4.7) and the proof is therefore completed. $\square$

In Proposition 4.1 we established the Lip'-consistency of our scheme under the assumption that at each time step we solve the equation *exactly* with piecewise quadratic initial data; i.e., $E(\Delta t)$ in (4.1) was assumed to be the exact evolution operator. However, in practice we employ an approximate solver, $\tilde{E}(\Delta t)$, and that yields an additional Lip'-error. In order that this *practical* version of the scheme will still satisfy error estimates (4.3) and (4.4), the approximate solver should satisfy

$$|(A_h\tilde{E}(\Delta t) - A_hE(\Delta t))u^h(\cdot, t_n)| \leq O(h^2) \ , \tag{4.13}$$

[9, Proposition 3.5]. Indeed, our choice of the second order Lax-Wendroff approximate solver satisfies (4.13). Hence, we conclude that our *formally* third order scheme satisfies the global error estimates (4.3) as well as the local one (4.4).

5 Numerical experiments

5.1 Scalar convex problems

In this subsection we test our scheme on Burgers' equation. In the time evolution steps we replace the exact solution operator $E(\Delta t)$ with the approximate Lax-Wendroff solver, see [5] for details.

Example 1. We solve Burgers' equation with periodic boundary condition

$$\begin{array}{ll} u_t + (\frac{1}{2}u^2)_x = 0 & -1 \le x \le 1 \\ u(x,0) = u_0(x) & u_0(x) \text{ periodic with period 2.} \end{array}$$

When $u_0(x) = 1 + \frac{1}{2}sin(\pi x)$ the exact solution is smooth up to $t = \frac{2}{\pi}$ and then it develops a moving shock which interacts with a rarefaction wave. Table 1 lists the errors at time $t = 0.3$ with $\Delta t/h = 0.66$. We note that the L_1 and L_∞ convergence rates are close to 3.

TABLE 1

l	L_1 error	L_1 order	L_∞ error	L_∞ order
40	4.0790225910620D-04		8.2980813630651D-04	
80	6.9342852473318D-05	2.55	1.0831851524740D-04	2.94
160	9.9659315021962D-06	2.80	1.2665172799076D-05	3.10
320	1.5514357547895D-06	2.68	1.7406302790235D-06	2.86

Figure 5.1 depicts the solution at $t = 1.1$ and demonstrates the excellent behavior of the scheme. The shock resolution is optimal as the strong shock is captured in only one cell.

Figure 5.2 depicts the solution that corresponds to the initial value

$$u_0(x) = \begin{cases} \frac{1}{2} + x & -\frac{1}{2} \le x \le \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} .$$

Also here we obtain excellent results: the strong shock is resolved in one cell and there is no waggle around the rarefaction wave.

While the previous initial values were rarefaction-free, $\|u(\cdot,0)\|_{Lip^+} < \infty$, the next initial value,

$$u_0(x) = \begin{cases} 2 & -\frac{1}{2} \le x \le \frac{1}{2} \\ 1 & \text{otherwise.} \end{cases}$$

includes an initial rarefaction, i.e., $\|u(\cdot,0)\|_{Lip^+} = +\infty$. The numerical solution that corresponds to this initial value is shown in Figures 5.3-5.5. Here we see waggles at the two fronts of the rarefaction wave, in similarity to the second order Lip^+-consistent scheme [1]. As can be seen, those waggles are vanishing in L_1 as $h \to 0$.

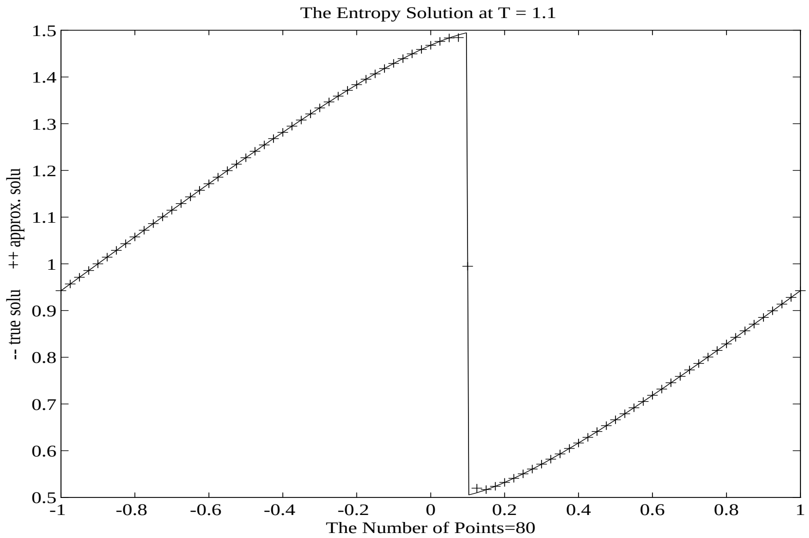

Figure 5.1: $\Delta t/h = 0.66$

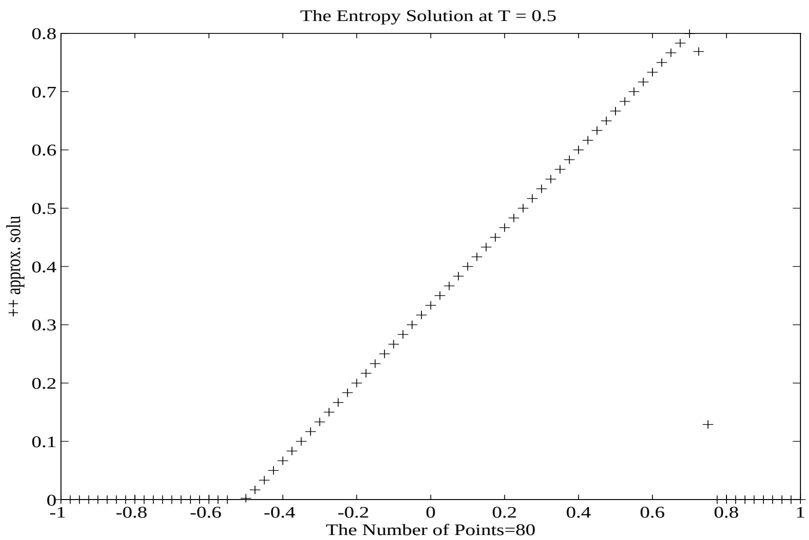

Figure 5.2: $\Delta t/h = 0.9$

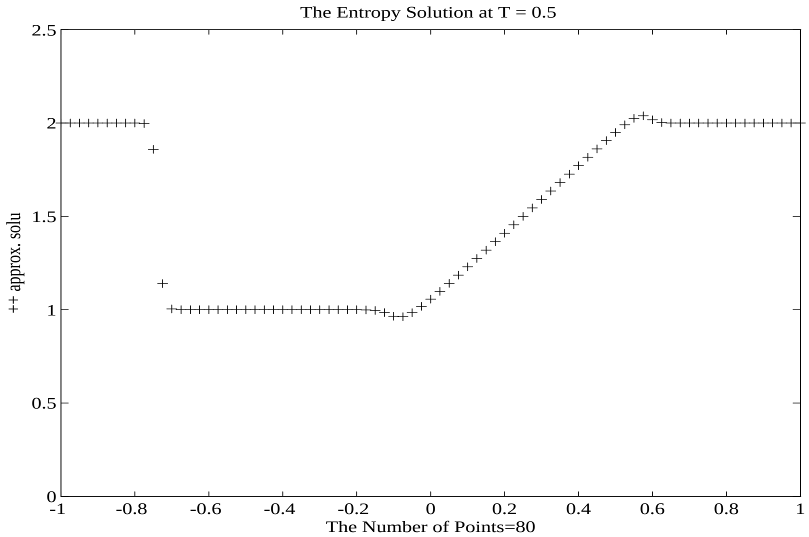

Figure 5.3: $\Delta t/h = 0.45$

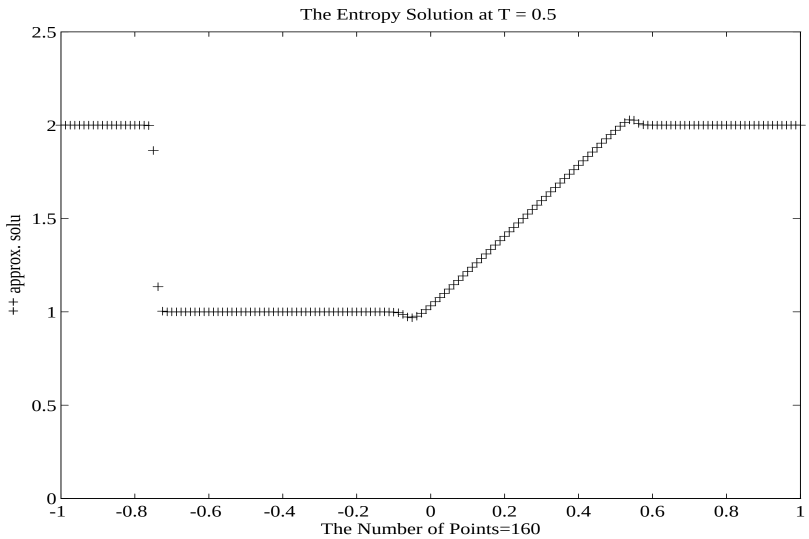

Figure 5.4: $\Delta t/h = 0.45$

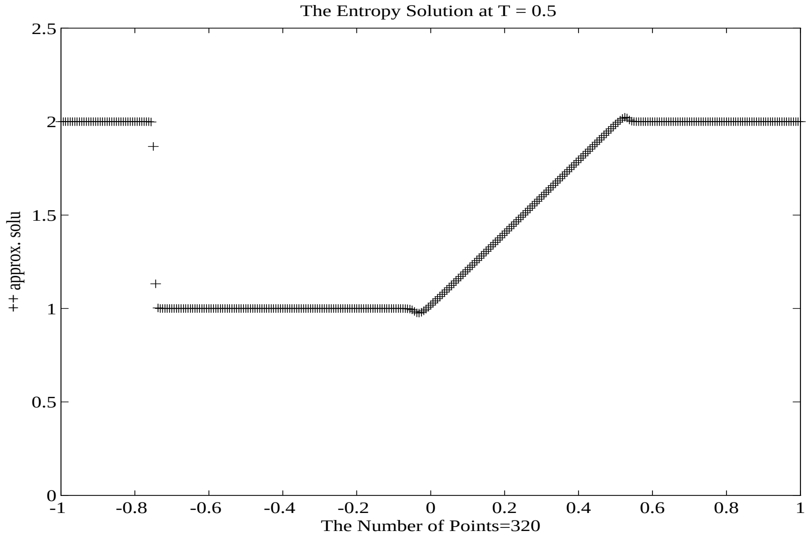

Figure 5.5: $\Delta t/h = 0.45$

Example 2. Next, we examine the convergence of the numerical solution to the stationary solution. We perform this test on the following periodic initial value problem with source term that was suggested by Van Leer (private communication) for entropy convergence test [18],

$$
\begin{array}{ll}
u_t + (\frac{1}{2}u^2)_x = -sin(\pi x), & -1 \leq x \leq 1 \\
u(x,0) = \left\{ \begin{array}{cr} 1 & -1 \leq x < 0 \\ -1 & 0 \leq x \leq 1 \end{array} \right. & .
\end{array}
$$

The stationary solution in this case is given by

$$
u(x,\infty) = \left\{ \begin{array}{ll} -\sqrt{\frac{2}{\pi}(1+cos(\pi x))} & 0 \leq x \leq 1 \\ +\sqrt{\frac{2}{\pi}(1+cos(\pi x))} & -1 \leq x \leq 0. \end{array} \right.
$$

The reconstructed solution at time $t = 0$ is plotted in Figure 5.6. There are both overshoot and undershoot in the reconstruction solution around $x = 0$. However the increasing discontinuity of the initial data immediately spreads out by the overshoot and undershoot into a rarefaction wave. Using a step size of $h = 1/8$, $N = 16$ cells and $dt/h = 0.64$, we obtain the solution shown in Figure 5.7. Adopting the L_1-Cauchy criterion for convergence to the stationary solution,

$$
\sum_{i=1}^{N} \mid u_i^n - u_i^{n+1} \mid < 10^{-3} \ ,
$$

the scheme converged after 21 time steps.

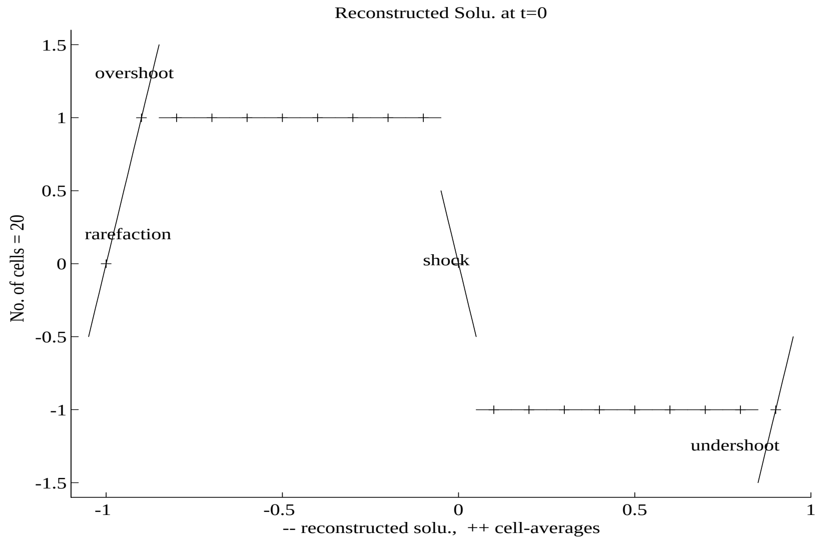

Figure 5.6:

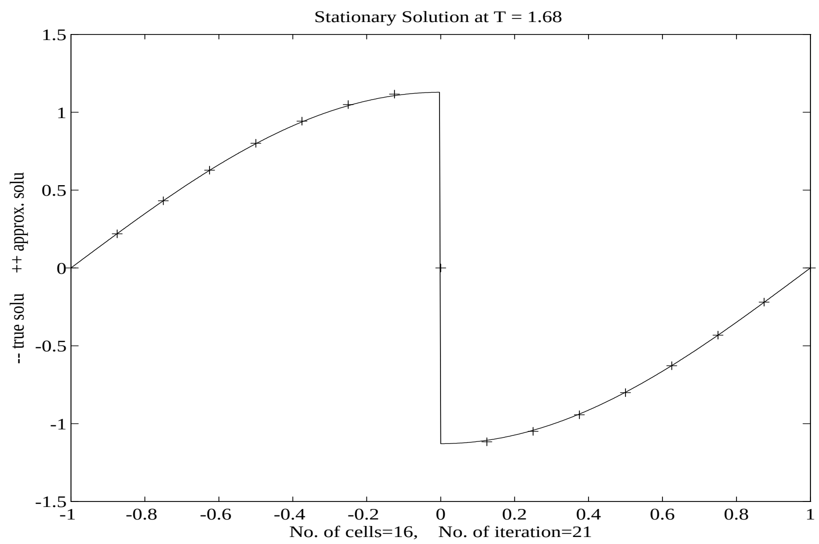

Figure 5.7: $\Delta t/h = 0.64$

5.2 Euler Equations of Gas Dynamics

Here, we apply our scheme to the Euler equations of gas dynamics for a polytropic gas,

$$\begin{array}{l} u_t + f(u)_x = 0 \\ u = (\rho, m, E)^T \\ f(u) = (m, \rho q^2 + P, q(E+P))^T \\ P = (\gamma - 1)(E - \frac{1}{2}\rho q^2) \\ m = \rho q \quad . \end{array}$$

As in [5], we set $\gamma = 1.4$ in all of our computations herein.

The Jacobian matrix $A(u) = \partial f / \partial u$ is

$$\begin{pmatrix} 0 & 1 & 0 \\ \frac{-m^2}{\rho^2} + \frac{m^2(\gamma-1)}{2\rho^2} & \frac{2m}{\rho} - \frac{m(\gamma-1)}{\rho} & \gamma - 1 \\ \frac{-m(E+(E-m^2/(2\rho))(\gamma-1)}{\rho^2} + \frac{m^3(\gamma-1)}{2\rho^3} & \frac{(E+(E-m^2/(2\rho))(\gamma-1)}{\rho} - \frac{m^2(\gamma-1)}{\rho^2} & \frac{m\gamma}{\rho} \end{pmatrix}.$$

The eigenvalues of $A(u)$ are

$$\lambda_1 = q - c \quad , \quad \lambda_2 = q \quad , \quad \lambda_3 = q + c \ ,$$

where $c = (\gamma P/\rho)^{1/2}$ is the speed of sound. The corresponding right-eigenvectors are

$$r_1(u) = \begin{pmatrix} 1 \\ q - c \\ H - qc \end{pmatrix} \ , \ r_2(u) = \begin{pmatrix} 1 \\ q \\ \frac{1}{2}q^2 \end{pmatrix} \ , \ r_3(u) = \begin{pmatrix} 1 \\ q + c \\ H + qc \end{pmatrix},$$

where $H = (E+P)/\rho$ is the enthalpy. The corresponding left-eigenvectors are

$$\begin{array}{l} l_1(u) = \frac{1}{2}(b_2 + q/c, -b_1 q - 1/c, b_1) \\ l_2(u) = (1 - b_2, b_1 q, -b_1) \\ l_3(u) = \frac{1}{2}(b_2 - q/c, -b_1 q + 1/c, b_1), \end{array}$$

where

$$\begin{array}{l} b_1 = (\gamma - 1)/c^2 \\ b_2 = \frac{1}{2}q^2 b_1 \quad . \end{array}$$

Since

$$\Delta_u \lambda_1 \cdot r_1 < 0 \tag{5.1}$$

$$\Delta_u \lambda_2 \cdot r_2 = 0 \tag{5.2}$$

$$\Delta_u \lambda_3 \cdot r_3 > 0 \tag{5.3}$$

we might use the concave reconstruction procedure for the first field w_1 of locally defined characteristic variables $w = (w_1, w_2, w_3)^T$ (where $w_i = l_i(\hat{u}) \cdot u, i = 1, 2, 3$), the non-oscillatory reconstruction procedure as in [8] for the second field w_2, and the convex reconstruction procedure for the third field w_3. For details of the locally defined characteristic variables, see [5].

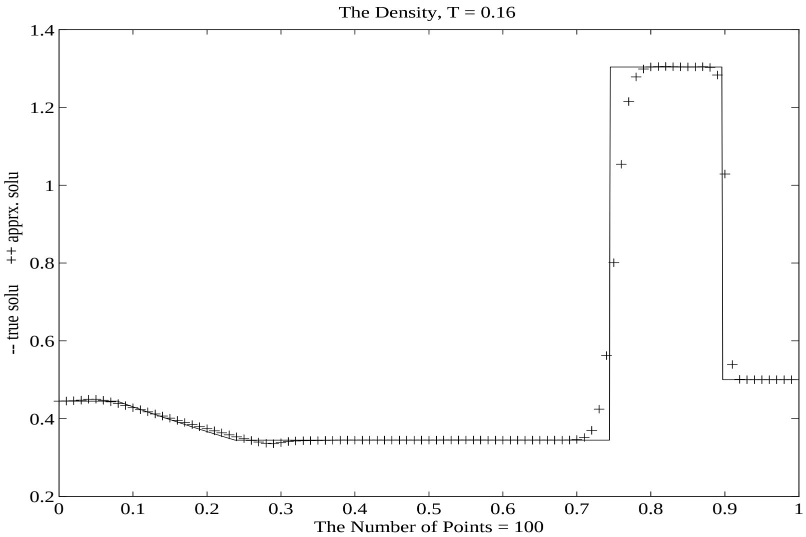

Figure 5.8: $\Delta t/h = 0.2$

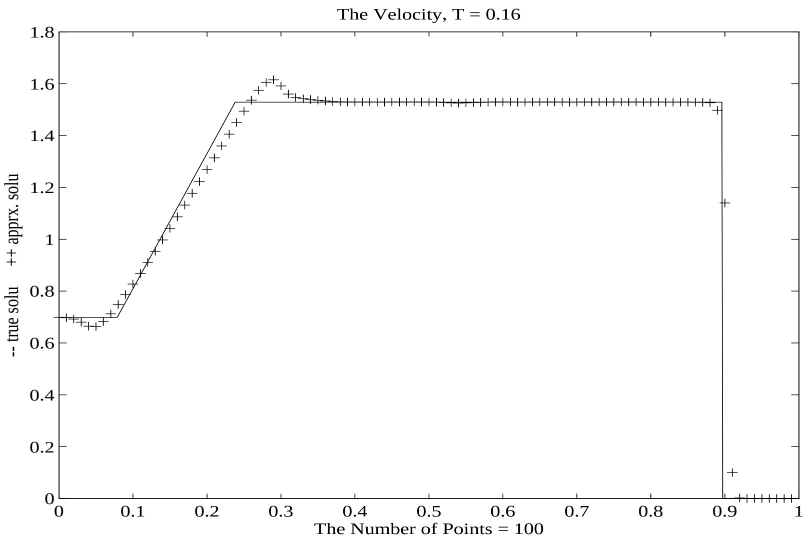

Figure 5.9: $\Delta t/h = 0.2$

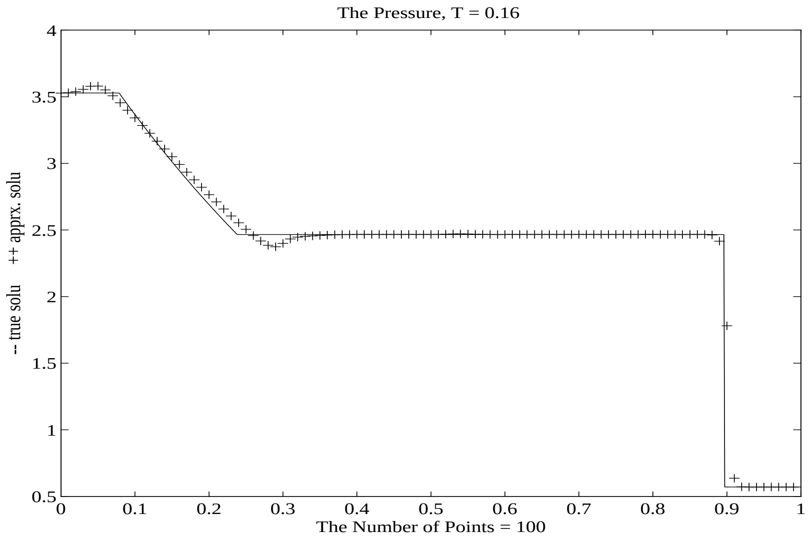

Figure 5.10: $\Delta t/h = 0.2$

Example 3. We consider the Riemann problem

$$u_0(x) = \begin{cases} u_l & 0 \le x < 0.5 \\ u_r & 0.5 \le x \le 1 \end{cases} .$$

Two sets of initial date are used: the first was proposed by Sod in [14],

$$(\rho_l, m_l, E_l) = (1, 0, 2.5) \quad , \quad (\rho_r, m_r, E_r) = (0.125, 0, 0.25) .$$

The other was used by Lax in [6],

$$(\rho_l, m_l, E_l) = (0.445, 0.311, 8.928) \quad , \quad (\rho_r, m_r, E_r) = (0.5, 0, 1.4275) .$$

We use the Roe flux [13] **without** entropy fix formed by Roe's average as the numerical flux; for details see [5]. The numerical results for Lax tube problem are shown in Figures 5.8-5.10. Some waggles are present around the rarefaction wave, in similarity to the scalar cases.

To avoid the waggles, we may use the non-oscillatory reconstruction procedure in [8] to reconstruct the first field w_1 of locally defined characteristic variables w. By doing so, we obtain excellent numerical results for both tube problems: the results are comparable with ENO 4th order scheme, [5], where an improvement is obtained in the profile for the density in Lax tube problem (in both problems we used only 100 cells). See Figures 5.11-5.13 for Sod tube problem and Figures 5.14-5.16 for Lax tube problem.

The reader is referred to the Appendix, where we spell out the FORTRAN subroutine of our reconstruction procedure for convex (and concave) fluxes.

Remark. The scheme works very well for both shock tube problems. However, the overshoot and the undershoot of the reconstruction step might lead to negative

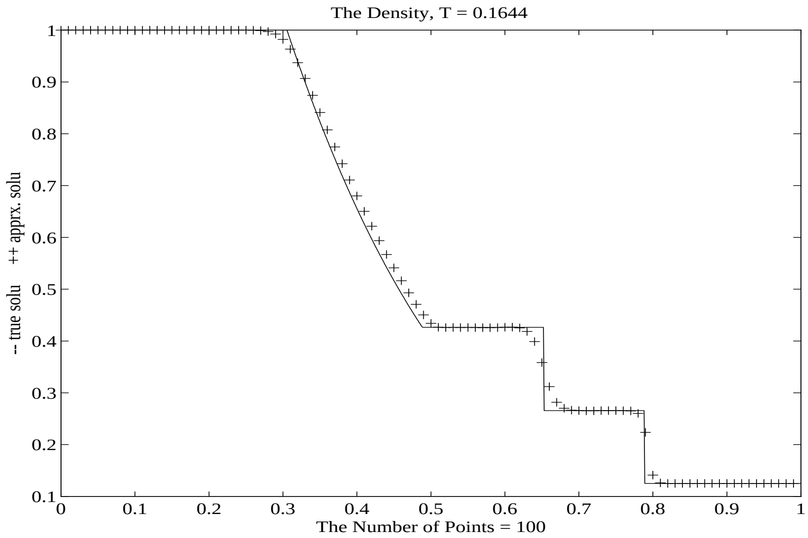

Figure 5.11: $\Delta t/h = 0.2$

pressure values in the numerical solution of Euler equations, and that stops the scheme. Another artifact of the over- and undershoots is the appearance of waggles around the rarefaction wave. These difficulties are currently under further investigation.

Acknowledgment. The first author is grateful to Professor Peter D. Lax for his careful reading of the manuscript, his helpful advice, support, and forbearance. Being novice to this field, the first author is in no position to present a meaningful historical review, however the author would like to appreciate the works of Lax [7], Brenier and Osher [1], Osher [11, 12], Tadmor [16] and [15], Engquist and Osher [3], Goodman and LeVeque [4], Crandall and Majda [2], and some others listed in the following references. All of them had great influence on the research reported in this paper.

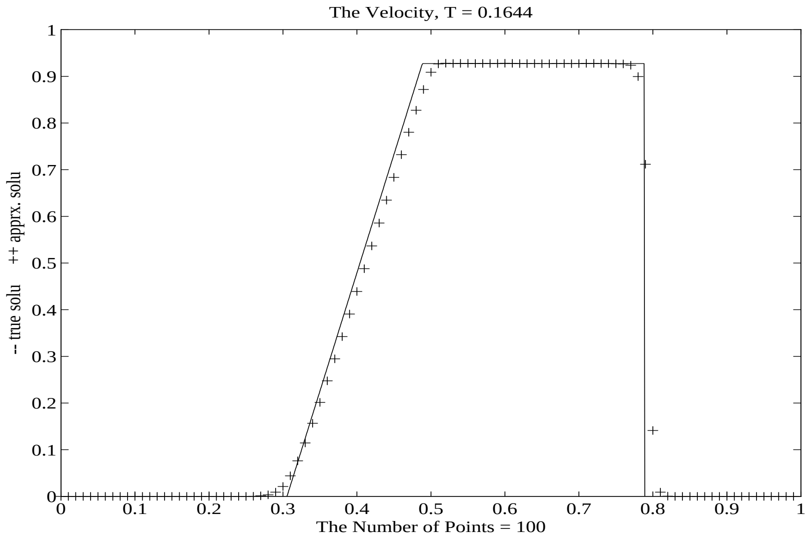

Figure 5.12: $\Delta t/h = 0.2$

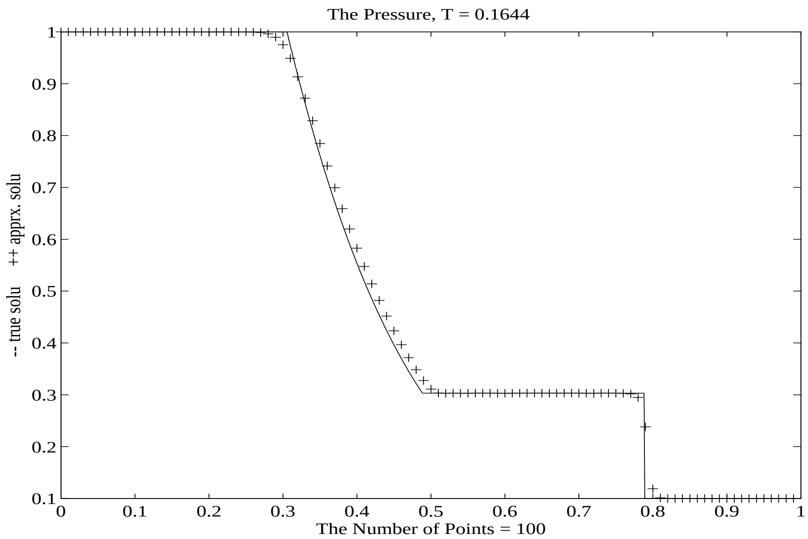

Figure 5.13: $\Delta t/h = 0.2$

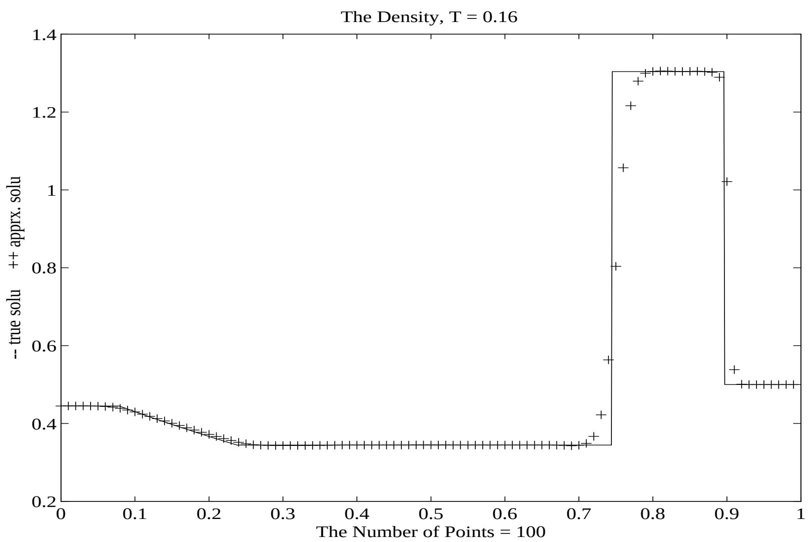

Figure 5.14: $\Delta t/h = 0.2$

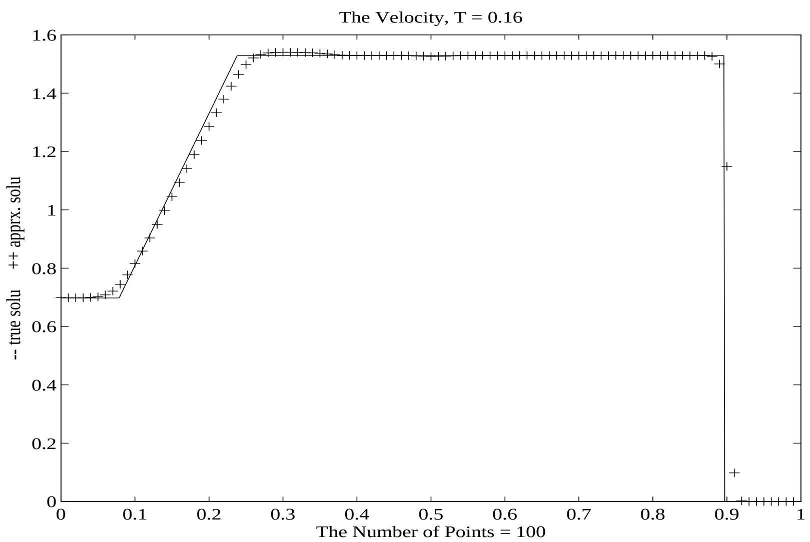

Figure 5.15: $\Delta t/h = 0.2$

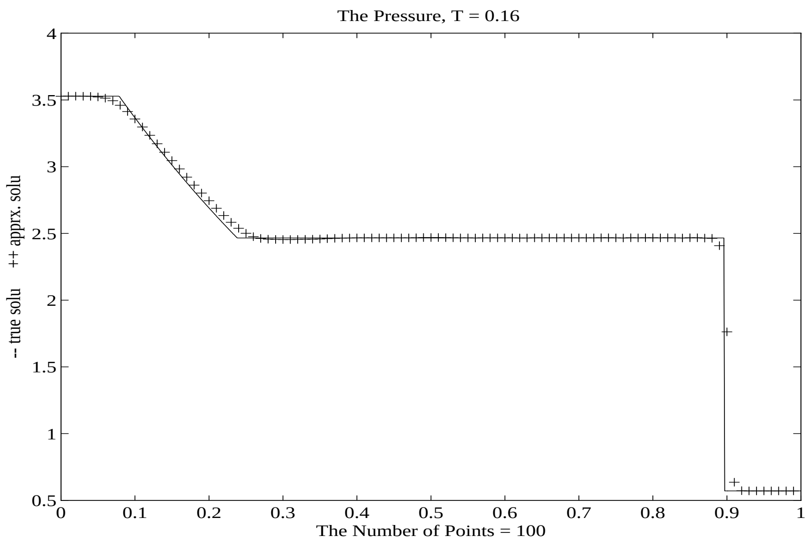

Figure 5.16: $\Delta t/h = 0.2$

6 Appendix - The Reconstruction Procedure

The Fortran code below performs the reconstruction in a typical cell I_j. The variables $u(k)$, $1 \leq k \leq 5$, denote the cell-averages on the five cells around I_j, namely I_k, $j-2 \leq k \leq j+2$. The variables ul and ur are left and right end point values, respectively, that are obtained after the reconstruction; i.e., $u_l = R_j(x_{j-\frac{1}{2}})$ and $u_r = R_j(x_{j+\frac{1}{2}})$. Hence, the reconstructed quadratic polynomial is characterized by the three values ul, $u(3)$ and ur.

The code may be easily modified to handle concave fluxes. Just replace all *.lt.*, *.le.* and *.ge.* to *.gt.*, *.ge.*, and *.le.*, respectively.

```
subroutine entropy_convex(ul,ur,u)
implicit real*8 (a-h,o-z)
dimension u(5)
dd1=u(3)-2.0d00*u(2)+u(1)
dd2=u(4)-2.0d00*u(3)+u(2)
dd3=u(5)-2.0d00*u(4)+u(3)
if(dd2.lt.0.0d00) then
  if(dd1.le.0.0d00) then
    ur1=( 2.0d00*u(3)+5.0d00*u(2)-u(1))/6.0d00
    ul2=(-u(4)+5.0d00*u(3)+2.0d00*u(2))/6.0d00
    if(ul2.le.ur1) then
      ul=ul2
      ur=(2.0d00*u(4)+5.0d00*u(3)-u(2))/6.0d00
    else
      ur=(11.0d00*u(3)-7.0d00*u(2)+2.0d00*u(1))/6.0d00
      ul=ur1
```

```
    end if
  else
    ul=0.5d00*(u(3)+u(2))
    ur=1.5d00*u(3)-0.5d00*u(2)
  end if
else
  if(dd3.ge.0.0d00) then
    ur1=( 2.0d00*u(4)+5.0d00*u(3)-u(2))/6.0d00
    ul2=(-u(5)+5.0d00*u(4)+2.0d00*u(3))/6.0d00
    if(ul2.le.ur1) then
      ul=(-u(4)+5.0d00*u(3)+2.0d00*u(2))/6.0d00
      ur=ur1
    else
      ur=ul2
      ul=(11.0d00*u(3)-7.0d00*u(4)+2.0d00*u(5))/6.0d00
    end if
  else
    ur=0.5d00*(u(3)+u(4))
    ul=1.5d00*u(3)-0.5d00*u(4)
  end if
end if
return
end
```

References

[1] Brenier, Y. and Osher, S., "The Discrete One-Sided Lipschitz Condition For Convex Scalar Conservation Laws," *SIAM J. Numer. Anal.*, Vol. 25, No. 1, 1988, pp. 8-23.

[2] Crandall, M. and Majda, A., "Monotone Difference Approximations for Scalar Conservation Laws," *Math. of Comp.* Vol. 34, 1980, pp. 1-21.

[3] Engquist, B. and Osher, S., "Stable and Entropy Satisfying Approximations for Transonic Flow Calculations," *Math. Comp.*, Vol. 34, No. 149, 1980, pp.45-75.

[4] Goodman, J.B. and LeVeque, R.J., "A Geometric Approach to High Resolution TVD Schemes," *SIAM J. Numer. Anal.*, Vol. 25 pp. 268-284, 1988.

[5] Harten, A., Engquist, B., Osher, S. and Chakravarthy, S., "Uniformly High Order Accurate Essentially Non-Oscillatory Schemes III," *J. Comp. Phys.*, Vol. 71, 1987, pp. 231-303; also *ICASE Report* No. 86-22, April 1986.

[6] Lax, P.D., "Weak Solutions of Nonlinear Hyperbolic Equations and Their Numerical Computation," *Commun. Pure Appl. Math.* 7, pp. 159-193(1954).

[7] Lax, P.D., "Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves," CMBS Regional Conference Series in Applied Mathematics, SIAM, Philadephia, 1972

[8] Liu, X.D. and Osher, S., "Nonoscillatory High Order Accurate Self-Similar Maximum Principle Satisfying Shock Capturing Schemes I", to appear on *SIAM Numer. Anal.*

[9] Nessyahu, H., Tadmor, E. and Tassa, T., "The Convergence Rate of Godunov Type Schemes," *SIAM J. Numer. Anal.*, Vol. 31, No. 1, pp. 1-16, Feb. 1994.

[10] Oleinik, O., "Discontinuous solutions of nonlinear differential equations," *Uspekhi Mat. Nauk.*, 12 (1957), pp 3-73. (in Russian.) *Amer. Math. soc. Transl.*, 26, pp. 95-172. (in English.).

[11] Osher, S., "Riemann Solvers, the Entropy Condition, and Difference Approximations," *SIAM J. Numer. Anal.*, Vol. 21, No. 2, pp. 217-235, 1984.

[12] Osher, S., "Convergence of Generalized MUSCL Schemes," *SIAM J. Numer. Anal.*, Vol. 22, No. 5, 1985, pp. 947-961.

[13] Roe, P.L., "Approximate Riemann Solver, Parameter Vectors and Difference Schemes," *Math. of Comp.*, Vol. 43, pp. 357-372 (1981).

[14] Sod, G., "A Survey of Several Finite Difference Methods for Systems of Nonlinear Hyperbolic Conservation Laws," *J. Comput. Phys.* Vol. 27, No. 1, pp. 1-31(1978).

[15] Tadmor, E., "The Large-time Behavior of the Scalar, Genuinely Nonlinear Lax-Friedrichs Scheme," *Math. of Comp.*, Vol. 43, 1984, pp. 353-368.

[16] Tadmor, E., "Numerical Viscosity and the Entropy Condition for Conservative Difference Schemes," *Math. Comp.*, 43, pp. 369-381(1984).

[17] Van Leer B., " Towards the Ultimate Conservative Difference Scheme II. Monotonicity and Conservation Combined in a Second Order Scheme," *J. Comput. Phys.*, Vol. 14, pp. 361-376 (1974).

[18] Van Leer B., "On the Relation Between the Upwind-Differencing Schemes of Godunov, Engquist-Osher and Roe," *SIAM J. Sci. Stat. Comp.*, Vol. 5, No. 1, pp. 1-20, March 1984.