Local error analysis for approximate solutions of hyperbolic conservation laws

Smadar Karni [a] and Alexander Kurganov [b]

[a] *Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1109, USA*
E-mail: karni@math.lsa.umich.edu
[b] *Department of Mathematics, Tulane University, New Orleans, LA 70118, USA*
E-mail: kurganov@math.tulane.edu

We consider approximate solutions to nonlinear hyperbolic conservation laws. If the exact solution is unavailable, the truncation error may be the only quantitative measure for the quality of the approximation. We propose a new way of estimating the local truncation error, through the use of localized test-functions. In the convex scalar case, they can be converted into L_{loc}^{∞} estimates, following the *Lip'* convergence theory developed by Tadmor et al. Comparisons between the local truncation error and the L_{loc}^{∞}-error show remarkably similar behavior. Numerical results are presented for the convex scalar case, where the theory is valid, as well as for nonconvex scalar examples and the Euler equations of gas dynamics. The local truncation error has proved a reliable smoothness indicator and has been implemented in adaptive algorithms in [Karni, Kurganov and Petrova, J. Comput. Phys. 178 (2002) 323–341].

Keywords: conservation laws, weak solutions, truncation error, error estimates

AMS subject classification: 65M15, 35L65

1. Introduction

We consider the system of one-dimensional hyperbolic conservation laws subject to the compactly supported (or periodic) initial data,

$$u_t + f(u)_x = 0, \quad u(x, t = 0) = u_0(x). \tag{1.1}$$

Such systems arise in many different applications, including fluid mechanics, astrophysics, meteorology, semiconductors and others – thus, solving (1.1) is of a great practical importance.

System (1.1) admits the (finite time) formation of shock discontinuities, even for infinitely smooth initial data. This makes the theory of (1.1) difficult and the design of numerical methods challenging (a detailed review of a variety of modern methods and approaches can be found in, e.g., [1,3,6,9]).

Beyond the time of shock formation, classical solutions cease to exist and the solution of (1.1) is extended as a weak solution in the sense of distributions.

Definition 1.1. $u(x, t)$ is a weak solution of the IVP (1.1) if it satisfies

$$-\int_{t=0}^{\infty} \int_X \left\{ u(x, t)\phi_t(x, t) + f\big(u(x, t)\big)\phi_x(x, t) \right\} \mathrm{d}x\,\mathrm{d}t + \int_X u(x, 0)\phi(x, 0)\,\mathrm{d}x = 0 \quad (1.2)$$

for all test-functions $\phi(x, t) \in C_0^{\infty}(X \times [0, \infty[)$.

Weak solutions are not unique and an additional criterion, known as an entropy condition, is used to select the unique physically relevant solution (see, e.g., [14]).

In this paper, we focus on the problem of measuring local errors in numerical solutions. Since solutions of nonlinear conservation laws are generally nonsmooth, standard methods of convergence rate analysis, based on the Taylor expansions (see, e.g., [12]), are invalid. Hitherto, there is no general convergence theory available, even for *scalar* hyperbolic conservation law (1.1).

Our work falls within the theoretical framework developed by Tadmor [17] for general vanishing viscosity approximations to strictly convex scalar conservation laws ($f'' \geqslant \alpha > 0$). In this theory, the errors are measured in the *weak Lip'*-norm:

$$\big\| w(x, t) \big\|_{Lip'} := \sup_{\phi \in C_0^{\infty}} \frac{|(w - \overline{w}, \phi)|}{\|\phi(x, t)\|_{Lip}} = \sup_{\phi \in C_0^{\infty}: \ \|\phi(x,t)\|_{Lip}=1} \big| (w - \overline{w}, \phi) \big|, \qquad (1.3)$$

where

$$\overline{w} := \frac{1}{\mathrm{supp}(w)} \int_{\mathrm{supp}(w)} w(x, t)\,\mathrm{d}x\,\mathrm{d}t, \quad \text{and}$$

$$\big\| \phi(x, t) \big\|_{Lip} := ess \sup_{(x,t) \neq (y,\tau)} \frac{|\phi(x, t) - \phi(y, \tau)|}{|x - y| + |t - \tau|}.$$

Let $u^{\Delta}(x, t)$ be an approximate solution to (1.1), depending on a small parameter Δ. A method is said to be *Lip'*-consistent if $u^{\Delta}(x, t)$ satisfies

$$\big\| u_t^{\Delta} + f\big(u^{\Delta}\big)_x \big\|_{Lip'} + \big\| u_0(x) - u^{\Delta}(x, 0) \big\|_{Lip'} = \mathrm{O}\big(\Delta^p\big), \quad p > 0, \qquad (1.4)$$

where the first term on the left-hand side is the *Lip'* truncation error.[1]

A-priori truncation error analysis, [10,17], shows that in the convex scalar case, *Lip'*-consistency together with one-sided stability imply convergence of the approximate solution to the unique entropy solution of (1.1). This approach was extended in [10,11] for various approximations including a number of first-order finite difference

---

[1] Throughout this paper, we follow the theoretical framework and notational convention of [10,11,17], where $\| u_t^{\Delta} + f(u^{\Delta})_x \|_{Lip'}$ is referred to as the *Lip'* truncation error, while, in fact, it should be referred to as the *Lip'* residual.

schemes and spectral viscosity approximations. The analogous convergence rate results for second-order Godunov-type schemes were obtained in [7].

Alternatively, error estimates can be obtained using a-posteriori truncation error analysis following the approach by Süli [16] developed for the discontinuous Galerkin finite element method.

In either approach, the rate of convergence, measured in the *Lip'*-norm, is of the same order as the *Lip'*-size of truncation error (provided that the approximation of the initial condition in (1.4) is sufficiently accurate). Total variation boundedness of the approximate solution then enables one to convert *Lip'* error estimates to both *global*, $W^s(L^p)$-, and *local* (all the way up to the discontinuity, consult [18,19]) L^∞_{loc} convergence rate estimates. The latter are of particular usefulness in applications.

In the present work, we propose a new method for *practical measurement* of the local *Lip'* truncation errors by using a basis of locally supported test-functions. Our particular choice of such test-functions is the localized quadratic B-splines. A global, compactly supported test-function may then be approximated by means of the local test-functions, and thus, in the scalar convex case the global *Lip'* truncation error can be obtained from the local ones.

In particular, we show that the local truncation error is of order $O(\Delta^{r+2})$ away from discontinuities, and $O(\Delta)$ near shocks, where r is the formal order of accuracy of the method. This difference of several orders of magnitude between smooth and nonsmooth regions provides an effective tool for identifying nonsmooth parts of the solution, which was successfully utilized in scheme and mesh adaption algorithms in [5].

In the case of a system of conservation laws, no rigorous error estimates can be obtained. However, one may still compute the local *Lip'* truncation error and use it as an error indicator and a tool to identify nonsmooth regions.

The paper is organized as follows. In section 2 we derive the formula for computing local *Lip'* truncation errors. In section 3 we utilize this result to establish the error estimates for both smooth and nonsmooth solutions. A variety of numerical examples are presented in section 4.

2. Computation of the local errors

Let $(x_j := j\Delta x, t^n := n\Delta t)$ be a uniform grid, and u_j^n be approximate values of $u(x_j, t^n)$, obtained by a conservative method. We denote by $u^\Delta(x, t)$ the corresponding piecewise constant approximation,

$$u^\Delta(x, t) := u_j^n, \quad \text{if } (x, t) \in [x_{j-1/2}, x_{j+1/2}] \times \left[t^{n-1/2}, t^{n+1/2}\right], \qquad (2.1)$$

where $x_{j\pm1/2} := x_j \pm \Delta x/2$ and $t^{n\pm1/2} := t^n \pm \Delta t/2$.

Our goal is to obtain a measure of how good this approximation is by computing how much it fails to locally satisfy (1.2). To this end, we compute the (x, t)-integral in (1.2) using *locally supported* test-functions,

$$\varphi_j^n(x, t) = B_j(x)B^n(t). \qquad (2.2)$$

There are various possible choices of such C_0^1 test-functions. In the following derivation, we have used local quadratic B-splines (see, e.g., [2]). We thus have

$$B_j(x) := \begin{cases} \dfrac{1}{2}\left(\dfrac{x - x_{j-3/2}}{\Delta x}\right)^2, & \text{if } x_{j-3/2} \leqslant x \leqslant x_{j-1/2}, \\[2mm] \dfrac{3}{4} - \left(\dfrac{x - x_j}{\Delta x}\right)^2, & \text{if } x_{j-1/2} \leqslant x \leqslant x_{j+1/2}, \\[2mm] \dfrac{1}{2}\left(\dfrac{x - x_{j+3/2}}{\Delta x}\right)^2, & \text{if } x_{j+1/2} \leqslant x \leqslant x_{j+3/2}, \\[2mm] 0, & \text{otherwise}; \end{cases} \tag{2.3}$$

$$B^n(t) := \begin{cases} \dfrac{1}{2}\left(\dfrac{t - t^{n-3/2}}{\Delta t}\right)^2, & \text{if } t^{n-3/2} \leqslant t \leqslant t^{n-1/2}, \\[2mm] \dfrac{3}{4} - \left(\dfrac{t - t^n}{\Delta t}\right)^2, & \text{if } t^{n-1/2} \leqslant t \leqslant t^{n+1/2}, \\[2mm] \dfrac{1}{2}\left(\dfrac{t - t^{n+3/2}}{\Delta t}\right)^2, & \text{if } t^{n+1/2} \leqslant t \leqslant t^{n+3/2}, \\[2mm] 0, & \text{otherwise}. \end{cases} \tag{2.4}$$

We then plug (2.1) and (2.2) into the (x, t)-integral in (1.2). A straightforward computation yields the *weak form of the truncation error*,

$$\begin{aligned} E_j^n &:= -\int_{t^{n-3/2}}^{t^{n+3/2}} \int_{x_{j-3/2}}^{x_{j+3/2}} \left\{ u^\Delta(x, t)\left[\varphi_j^n(x, t)\right]_t + f\left(u^\Delta(x, t)\right)\left[\varphi_j^n(x, t)\right]_x \right\} \mathrm{d}x \, \mathrm{d}t \\ &= \frac{1}{12}\Big\{ \big[u_{j+1}^{n+1} - u_{j+1}^{n-1} + 4\big(u_j^{n+1} - u_j^{n-1}\big) + u_{j-1}^{n+1} - u_{j-1}^{n-1}\big]\Delta x \\ &\quad + \big[f\big(u_{j+1}^{n+1}\big) - f\big(u_{j-1}^{n+1}\big) + 4\big(f\big(u_{j+1}^n\big) - f\big(u_{j-1}^n\big)\big) + f\big(u_{j+1}^{n-1}\big) - f\big(u_{j-1}^{n-1}\big)\big]\Delta t \Big\}. \end{aligned} \tag{2.5}$$

Remarks. Given a discrete approximate solution (2.1), the corresponding weak form of the truncation error E_j^n can be computed straightforwardly by inserting the values of $\{u_j^n\}$ into (2.5) both for scalar equations and for systems.

3.   Error etimates

In this section, we derive error estimates for one-dimensional scalar conservation laws using the weak local truncation errors E_j^n, (2.5). Throughout the section, we will assume that

($\mathcal{H}_1$). The approximate solution u^Δ is computed by an rth-order stable scheme,[2] and hence it converges toward the exact entropy solution.

Smooth case. We begin by considering the simplest case of the unique, sufficiently smooth solution of (1.1). Then, the weak form of the truncation error E_j^n may be estimated using Taylor expansions about (x_j, t^n). Let $\widetilde{u}$ be the exact solution of the *modified equation*,

$$\widetilde{u}_t + f(\widetilde{u})_x = RHS,$$

where RHS is O(Δ^r), $\Delta := \max(\Delta x, \Delta t)$. The computed solution, $\{u_j^n\}$, is then an $(r+1)$st-order approximation of $\widetilde{u}$, and the substitution of $\widetilde{u}$ into (2.5) gives

$$
\begin{aligned}
E_j^n = \frac{1}{12} &\Big\{ \big[\widetilde{u}(x_{j+1}, t^{n+1}) - \widetilde{u}(x_{j+1}, t^{n-1}) + 4(\widetilde{u}(x_j, t^{n+1}) - \widetilde{u}(x_j, t^{n-1})) \\
&+ \widetilde{u}(x_{j-1}, t^{n+1}) - \widetilde{u}(x_{j-1}, t^{n-1})\big]\Delta x + \big[f(\widetilde{u}(x_{j+1}, t^{n+1})) - f(\widetilde{u}(x_{j-1}, t^{n+1})) \\
&+ 4(f(\widetilde{u}(x_{j+1}, t^n)) - f(\widetilde{u}(x_{j-1}, t^n))) + f(\widetilde{u}(x_{j+1}, t^{n-1})) \\
&- f(\widetilde{u}(x_{j-1}, t^{n-1}))\big]\Delta t \Big\} + \mathrm{O}(\Delta^{r+2}) \\
= \Delta x \Delta t &\big[\widetilde{u}_t + f(\widetilde{u})_x\big] + \frac{(\Delta x)^3 \Delta t}{6}\big[\widetilde{u}_t + f(\widetilde{u})_x\big]_{xx} + \frac{\Delta x(\Delta t)^3}{6}\big[\widetilde{u}_t + f(\widetilde{u})_x\big]_{tt} \\
&+ \mathrm{O}(\Delta^{r+2}) + \mathrm{O}(\Delta^5) = \mathrm{O}(\Delta^{r+2}) + \mathrm{O}(\Delta^5).
\end{aligned}
\tag{3.1}
$$

Remarks. Estimate (3.1) indicates that in the smooth case, the weak local truncation error is dominated by the accuracy of the method, provided that $r \leqslant 3$, that is, the method is first-, second- or third-order accurate. For higher-order methods, estimate (3.1) is still valid but will be dominated by the O(Δ^5) term, which is a consequence of the choice of quadratic B-splines as local test-functions. In this case, different local test-functions (for example, higher degree B-splines with a larger support) would be needed to reduce the B-splines approximation error.

Nonsmooth case. Let us now consider the general case of a nonsmooth weak solution. We assume that the exact solution of (1.1) is piecewise smooth, and that it has finite number of shocks and rarefaction waves, the case commonly encountered in practice. We also assume that our approximate solution u^Δ is bounded by the constant depending only on the initial data.

Consider the IVP (1.1) for $0 \leqslant t \leqslant T$, where T is finite. We split the domain $X \times [0, T]$ into two parts according to the smoothness of the exact solution: let $u(x, t)$ be smooth in D_1, and let D_2 be a unity of small neighborhoods of discontinuity curves with $D_1 \oplus D_2 = X \times [0, T]$, see figure 1.

---

[2] When a smooth solution is considered, it is enough to assume that the scheme is linearly stable (e.g., L^∞-, L^1- or L^2-stable). In a generic, nonsmooth case, the lip^+-stability [10,11] or, say, the BV-stability together with the discrete entropy inequality are assumed to be satisfied.
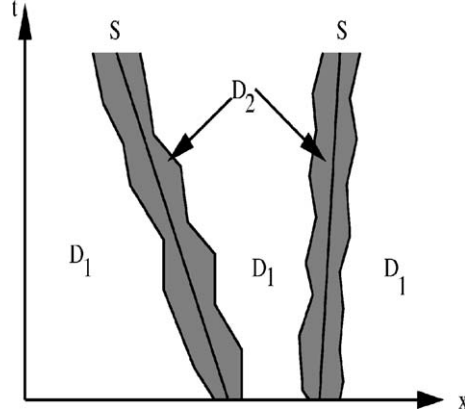
Figure 1. Smooth and nonsmooth domains.

Throughout the remainder of the section, we will assume that

($\mathcal{H}_2$) The size of domain D_2 is $O(\Delta)$, and that outside of D_2 the computed solution u^Δ manifests an order α behavior, namely,

$$\max_{D_1} \left| u^\Delta - u \right| = O(\Delta^\alpha), \tag{3.2}$$

where $0 < \alpha \leqslant r$.

Remarks. In general, the value of α depends on the order of the discretization scheme r and on how the domain is split into smooth and nonsmooth regions. Notice that in a generic case of a "reasonable", convergent scheme of order $r \geqslant 1$, one may choose D_2 to be sufficiently large to ensure $\alpha \geqslant 1$. We also note that for computed shock waves of width of order $O(\Delta)$, if the domains D_1 and D_2 are chosen accordingly, one expects $\alpha \approx r$.

We will treat the smooth and nonsmooth regions separately. If the solution is smooth in the neighborhood of (x_j, t^n), we may still use the Taylor expansions, to obtain the estimate similar to (3.1):

$$\left| E_j^n \right| \leqslant C\Delta^{\alpha+2} + O(\Delta^5), \quad \text{if } (x_j, t^n) \in D_1. \tag{3.3}$$

For the points located near the discontinuities or rarefaction "tips", we can only use the boundedness of the solution. Therefore, from (2.5) we obtain,

$$\left| E_j^n \right| = O(\Delta), \quad \text{if } (x_j, t^n) \in D_2. \tag{3.4}$$

Next, we need to estimate the *Lip′ truncation error* in order to establish the *Lip′*-consistency of the approximate solution, (1.4). Let $\phi(x, t)$ be a $C_0^\infty(X \times [0, T])$ test-function, such that $\|\phi\|_{Lip} = 1$. First, note that for the exact solution

$$\int_{t=0}^{T} \int_{X} \left(u_t + f(u)_x\right)\phi \, dx \, dt = 0. \tag{3.5}$$

For an approximate solution, $u^\Delta(x, t)$, we use (3.5) and integration by parts to obtain

$$\int_{t=0}^{T} \int_{X} \left(u_t^\Delta + f\left(u^\Delta\right)_x\right)\phi \, dx \, dt = \int_{t=0}^{T} \int_{X} \left(\left[u^\Delta - u\right]_t + \left[f\left(u^\Delta\right) - f(u)\right]_x\right)\phi \, dx \, dt$$

$$= \int_{X} \left[u^\Delta(x, 0) - u_0(x)\right]\phi \, dx - \int_{t=0}^{T} \int_{X} \left(\left[u^\Delta - u\right]\phi_t + \left[f\left(u^\Delta\right) - f(u)\right]\phi_x\right) dx \, dt$$

$$=: I - II. \tag{3.6}$$

The first integral on the right-hand side of (3.6), I, is small provided the approximation of the initial data is sufficiently accurate. We thus consider integral II, and split the integration into two contributions – over smooth and nonsmooth parts of the solution,

$$II = \iint_{D_1} \left(\left[u^\Delta - u\right]\phi_t + \left[f\left(u^\Delta\right) - f(u)\right]\phi_x\right) dx \, dt$$

$$+ \iint_{D_2} \left(\left[u^\Delta - u\right]\phi_t + \left[f\left(u^\Delta\right) - f(u)\right]\phi_x\right) dx \, dt =: IIa + IIb. \tag{3.7}$$

Our goal is to establish bounds on IIa and IIb. To estimate IIa, we interpolate $\phi(x, t)$ together with its derivatives by the localized B-splines, (2.2)–(2.4). For the quadratic B-splines one has (see, e.g., [2])

$$\phi(x, t) = \sum_j \sum_n c_j^n \varphi_j^n(x, t) + O\left(\Delta^3\right),$$

$$\phi_x(x, t) = \sum_j \sum_n c_j^n \frac{\partial \varphi_j^n}{\partial x}(x, t) + O\left(\Delta^2\right), \tag{3.8}$$

$$\phi_t(x, t) = \sum_j \sum_n c_j^n \frac{\partial \varphi_j^n}{\partial t}(x, t) + O\left(\Delta^2\right),$$

where c_j^n are independent of the grid scale Δ. Then, using (2.5), (3.2), (3.3) and (3.8) we may estimate IIa as follows,

$$|IIa| = \left| \sum_j \sum_n c_j^n \int_{t^{n-3/2}}^{t^{n+3/2}} \int_{x_{j-3/2}}^{x_{j+3/2}} \left(\left[u^\Delta - u\right]\left(\varphi_j^n\right)_t \right. \right.$$

$$\left. + \left[f\left(u^\Delta\right) - f(u)\right]\left(\varphi_j^n\right)_x\right) dx \, dt \Bigg| + O\left(\Delta^{\alpha+2}\right)$$

$$\leqslant \sum_j \sum_n \left|c_j^n\right|\left|E_j^n\right| + O\left(\Delta^{\alpha+2}\right). \tag{3.9}$$

Using hypothesis $(\mathcal{H}_2)$, the boundedness of u, ϕ_x and ϕ_t, and the assumed boundedness of u^Δ, we obtain the estimate on IIb:

$$|IIb| \leqslant |D_2| \cdot \max_{(x,t)\in D_2} \left(\left|u^\Delta\right| + |u| + \left|f\left(u^\Delta\right)\right| + |f(u)|\right) \cdot \|\phi\|_{Lip} = O(\Delta). \qquad (3.10)$$

Finally, we use (3.6), (3.7), (3.9), (3.10), and the definition of the global Lip'-norm, (1.3), to estimate the Lip' truncation error,

$$
\begin{aligned}
\left\|u_t^\Delta + f\left(u^\Delta\right)_x\right\|_{Lip'} &= \sup_{\|\phi(x,t)\|_{Lip}=1} \left| \int_{t=0}^{T} \int_X \left(u_t^\Delta + f\left(u^\Delta\right)_x\right)\phi \, dx \, dt \right| \\
&\leqslant \sup_{\{c_j^n\}:\, \|\phi(x,t)\|_{Lip}=1} \left[\sum_j \sum_n \left|c_j^n\right|\left|E_j^n\right| \right] + O\left(\Delta^{\alpha+2}\right) + O(\Delta) \\
&\leqslant \sup_{\{c_j^n\}:\, \|\phi(x,t)\|_{Lip}=1} \left[\sum_j \sum_n \left|c_j^n\right|\left(O\left(\Delta^{\alpha+2}\right) + O\left(\Delta^5\right)\right) \right] + O(\Delta) \\
&= O\left(N^2\right)\left(O\left(\Delta^{\alpha+2}\right) + O\left(\Delta^5\right)\right) + O(\Delta) = O\left(\Delta^\alpha\right) + O(\Delta), \quad (3.11)
\end{aligned}
$$

where $N = O(1/\Delta)$ is the number of grid points.

Remarks. Estimate (3.11) demonstrates that the optimal estimate of the Lip' truncation error is limited by the convergence rate in the nonsmooth region, and is of order $O(\Delta)$. In principle, this limitation can be alleviated only if the shock location is known to better than $O(\Delta)$ accuracy.

We summarize these results in the following theorem.

Theorem 3.1. Let $\varphi_j^n(x, t) \in C_0^1$ be the set of locally supported B-splines (2.2)–(2.4), E_j^n be the weak local truncation error (2.5), and $u^\Delta(x, t)$ be an approximate solution to (1.1) of formal order $r \leqslant 3$, satisfying hypotheses $(\mathcal{H}_1)$ and $(\mathcal{H}_2)$, and $\Delta := \max(\Delta x, \Delta t)$. Then

(i) if the exact solution of (1.1) is smooth

$$
\begin{aligned}
\left|E_j^n\right| &= O\left(\Delta^{r+2}\right), \quad \forall j, n \\
\left\|u_t^\Delta + f\left(u^\Delta\right)_x\right\|_{Lip'} &\leqslant O\left(\Delta^r\right);
\end{aligned}
$$

(ii) if $u(x, t)$ is nonsmooth

$$
\begin{aligned}
\left|E_j^n\right| &= O\left(\Delta^{\alpha+2}\right), \quad \text{away from discontinuities,} \\
\left|E_j^n\right| &= O(\Delta), \quad \text{near discontinuities,} \\
\left\|u_t^\Delta + f\left(u^\Delta\right)_x\right\|_{Lip'} &= O\left(\Delta^\alpha\right) + O(\Delta),
\end{aligned}
$$

where α is the assumed order of accuracy in D_1.

Using the results in [10,11,17] for the convex scalar case with rarefaction-free initial data, *Lip′*-consistency (3.11) together with one-sided stability of the numerical method imply *Lip′*-convergence of the same order. Note that this technique was developed for obtaining *a-priori* error estimates, while our truncation error computation is an *a-posteriori* computation.

In the framework of a-posteriori convergence analysis (for example, see [16]), the error is obtained directly in terms of the a-posteriori truncation error, and no appeal to the stability of the method is made. Let us denote the a-posteriori *discrete Lip$^+$*-seminorm of the approximate solution $u^\Delta(x, t)$ by

$$\max_n \left\| u^\Delta\left(\cdot, t^n\right) \right\|_{lip^+} := \max_{j,n} \left(\frac{u_{j+1}^n - u_j^n}{\Delta x} \right)_+ \equiv K_\Delta. \qquad (3.12)$$

We then obtain the following *a-posteriori Lip′* error estimate,

$$\left\| u(\cdot, t) - u^\Delta(\cdot, t) \right\|_{Lip'(X)} \leqslant C(K_\Delta)\Delta \max_{j,n} \left| E_j^n \right|, \qquad (3.13)$$

which is valid with the constant C depending on the discrete *lip$^+$*-seminorm of the approximate solution.

Remarks.

1. If the approximate solution is also *BV*-stable, then the estimate (3.13) can be converted both into a *global $W^s(L^p)$* with $-1 \leqslant s \leqslant 1/p$, $1 \leqslant p \leqslant \infty$ [10,17], and into a *local L_{loc}^∞* error estimates (all the way up to the discontinuity) [18,19].

2. For rarefaction-free initial data, if the approximate solution is *lip$^+$*-stable then the constant $C(K_\Delta)$ in (3.13) is of order O(1), and the *Lip′* convergence rate is of the same order as the *Lip′* truncation error. Otherwise, $C(K_\Delta)$ may get very large, and the estimate (3.13), while still valid, may become totally useless.

4. Numerical results

In this section, we present examples for the convex scalar case, for which the error estimates in section 3 hold, as well as for the nonconvex scalar case and for systems. We compute the weak form of the truncation error (2.5) for the first- and second-order Roe-type upwind schemes [13] and the first-, second- and third-order semi-discrete central schemes in [8]. Convergence rates in smooth and nonsmooth subregions are checked, and are generally in excellent agreement with our local convergence theory.

Within the *Lip′* convergence theory, *Lip′* consistency together with (discrete) *Lip$^+$* stability imply global *Lip′* convergence, which can then be converted into local L_{loc}^∞ convergence. However, it does not provide a direct relation between the local truncation error and the local error. In particular, smallness of the local truncation error does not ensure smallness of the local error, which, in fact, is not obvious [4]. The potential of the local truncation error as an error indicator is demonstrated by comparison with the actual pointwise L_{loc}^∞-error (see example 1).

In the convex scalar case, one unexpected finding is that the local convergence rate near nonlinear discontinuities is of order $O(\Delta)$, as predicted by the theory, while near linear discontinuities the order appears to be somewhat larger – between $O(\Delta)$ and $O(\Delta^2)$.

While the *Lip′* convergence theory and the error estimates in section 3 are not valid for systems of conservation laws, the weak form of the truncation error (2.5) can still be computed. In section 4.2, we present an example with the Euler equations of gas dynamics. It is interesting to note, that the convergence rates near the shock and the contact discontinuities are in agreement with the nonlinear and linear (section 4.1) scalar examples.

In the following, we use the abbreviations LTE and LE to denote the weak form of the truncation error, (2.5), and the $L_{\mathrm{loc}}^{\infty}$-error. We also use the notation U1, U2, C1, C2 and C3 for the first-, second- and third-order upwind and central schemes, respectively.

4.1. Scalar equations

Example 1 (Burgers equation). We start with Burgers equation subject to a periodic boundary conditions,

$$\begin{cases} u_t + \left(\dfrac{u^2}{2}\right)_x = 0, & x \in [0, 2], \\ u(x, 0) = \sin(\pi x). \end{cases}$$

The computations were performed on a sequence of grids with $N = 40, 80, 160, 320, 640$ and 1280 points. The results are shown at time $T = 1$, which is beyond the time of shock formation ($t_* = 1/\pi$). The local convergence rates are computed over the smooth subregion $x \in [0.4, 0.6]$, and over the entire interval.

Figures 2–4 show the computed solution, LTE and LE, respectively. We note that the region where the LTE is significantly larger is in good agreement with the shock location. As predicted by the theory, this region becomes narrower with mesh refinement, and the amplitude of the LTE decays. Convergence rates for the various schemes are summarized in tables 1–3, and show excellent agreement with the theoretical predictions – $O(\Delta)$ near the shock, and $O(\Delta^{r+2})$ in the the smooth subregion.

We note that the difference in size of the LTE between the smooth and nonsmooth regions is of several orders of magnitude – $O(\Delta)$ compared with $O(\Delta^{r+2})$ – which is

Table 1
Burgers equation – maximal LTE in the smooth subregion [0.4, 0.6].

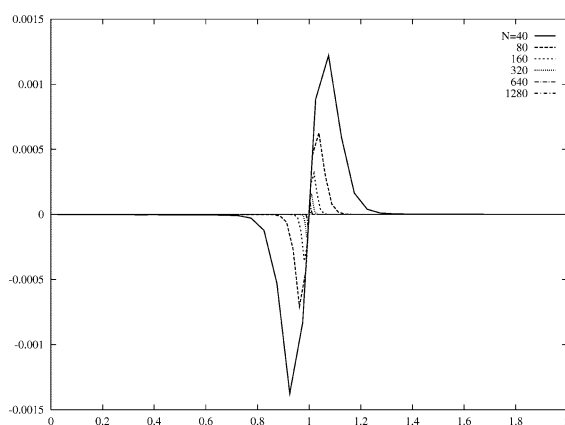| N	U1	Rate	U2	Rate	C1	Rate	C2	Rate	C3	Rrate
40	3.905e−06	–	8.132e−07	–	2.105e−05	–	8.430e−07	–	3.967e−07	–
80	4.854e−07	3.01	1.220e−08	6.06	3.156e−07	6.06	1.953e−09	8.75	5.245e−08	2.92
160	6.097e−08	2.99	7.196e−10	4.08	5.105e−08	2.63	1.293e−10	3.92	4.645e−12	13.5
320	7.644e−09	3.00	4.482e−11	4.00	6.925e−09	2.88	8.237e−12	3.97	1.396e−13	5.06
640	9.571e−10	3.00	2.797e−12	4.00	8.960e−10	2.95	5.286e−13	3.96	4.405e−15	4.99
1280	1.197e−10	3.00	1.747e−13	4.00	1.138e−11	2.98	3.330e−14	3.99	1.387e−16	4.99
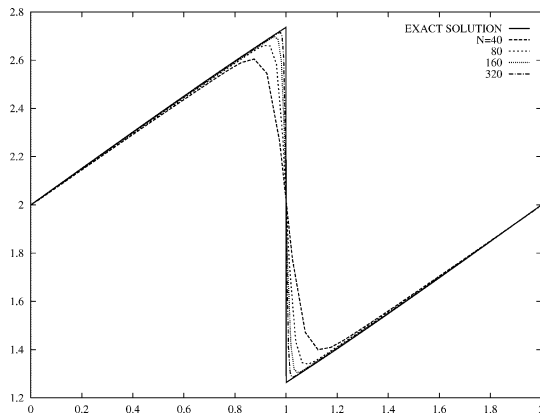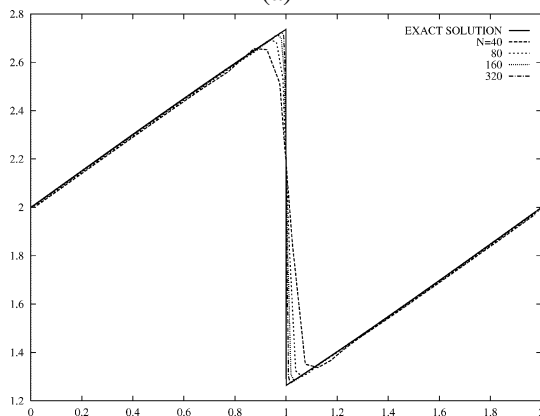
Figure 2. Solution (a) by 1-order upwind scheme, (b) by 2-order central scheme.

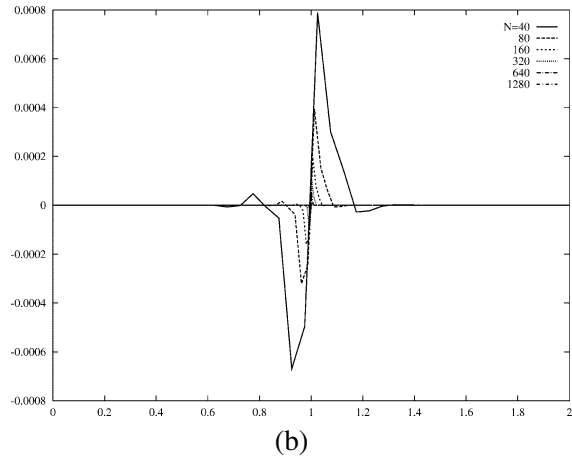Figure 3. LTE, (a) 1-order upwind scheme, (b) 2-order central scheme.

(b)

Figure 3. (Continued).

(a)

(b)
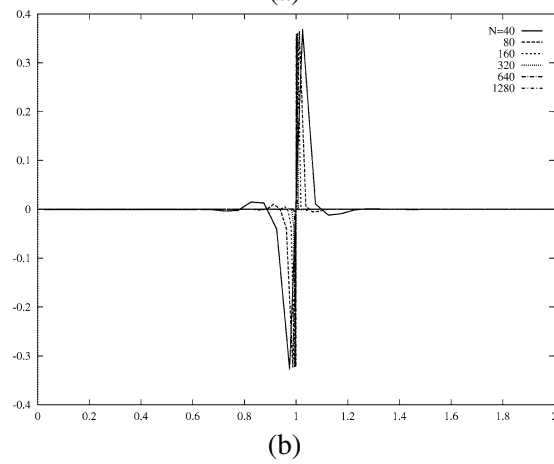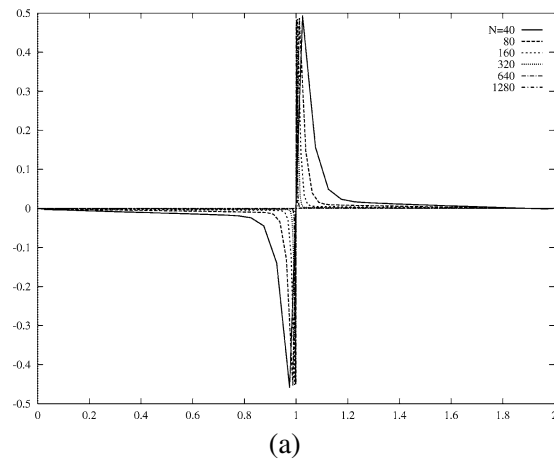
Figure 4. LE, (a) 1-order upwind scheme, (b) 2-order central scheme.

Table 2
Burgers equation – maximal LE in the smooth subregion [0.4, 0.6].

N	U1	Rate	U2	Rate	C1	Rate	C2	Rate	C3	Rate
40	1.376e−02	–	1.391e−03	–	1.301e−02	–	7.299e−04	–	3.164e−04	–
80	6.957e−03	0.98	3.465e−04	2.01	5.359e−03	1.28	1.997e−04	1.87	5.829e−05	2.44
160	3.500e−03	0.99	8.684e−05	2.00	3.279e−03	0.71	5.235e−05	1.93	1.463e−06	5.32
320	1.756e−03	1.00	2.170e−05	2.00	1.774e−03	0.89	1.340e−05	1.97	1.858e−07	2.98
640	8.794e−04	1.00	5.420e−06	2.00	9.188e−04	0.95	3.391e−06	1.98	2.342e−08	2.99
1280	4.401e−04	1.00	1.357e−06	2.00	4.672e−04	0.98	8.530e−07	1.99	2.940e−09	2.99

Table 3
Burgers equation – maximal LTE in the entire interval [0, 2].

N	U1	Rate	U2	Rate	C1	Rate	C2	Rate	C3	Rate
40	1.376e−03	–	1.173e−03	–	3.830e−04	–	7.891e−04	–	7.743e−04	–
80	7.072e−04	0.96	5.714e−04	1.03	2.200e−04	0.80	3.938e−04	1.00	3.825e−04	1.02
160	3.581e−04	0.98	2.811e−04	1.02	1.175e−04	0.90	1.958e−04	1.01	1.778e−04	1.11
320	1.802e−04	0.99	1.398e−04	1.01	6.038e−05	0.96	9.763e−05	1.00	8.356e−05	1.09
640	9.038e−05	1.00	6.971e−05	1.00	3.059e−05	0.98	4.869e−05	1.00	3.808e−05	1.13
1280	4.526e−05	1.00	3.484e−05	1.00	1.539e−05	0.99	2.434e−05	1.00	1.857e−05	1.04

reflected in the size of the LTE in the respective regions (for example, $\sim 10^{-4}$ compared with $\sim 10^{-10}$, for the second-order schemes with $N = 160$). This suggest that the LTE might be a useful tool in identifying regions of large errors and flagging them for mesh refinement, as supported by the numerical results presented in [5]. We would also like to point out the excellent agreement between the LTE and the LE (note the different scale).

Figure 5 displays a comparison of the LTE for the aforementioned schemes. We note that the width of the shock layer decreases with the order of the scheme, which is clearly reflected by the behavior of the LTE. We also note that the amplitude of the LTE near the shock is not an indication of the quality of the approximation. In fact, the magnitude of the LTE may even be larger for higher resolution schemes.

Example 2 (Linear advection equation). We next compute the LTE for the linear advection equation,

$$u_t + u_x = 0, \quad x \in [0, 2],$$

with the discontinuous initial data,

$$u(x, 0) = \begin{cases} 2, & x < 0.5, \\ -2, & x > 0.5, \end{cases}$$

and Dirichlet boundary conditions.

The LTE and convergence rates at time $T = 1$ are summarized in table 4. An unexpected result is that the LTE converges with the rate between $O(\Delta)$ and $O(\Delta^2)$, as opposed to $O(\Delta)$ in the nonlinear case (see table 3). It is worth noting that this difference
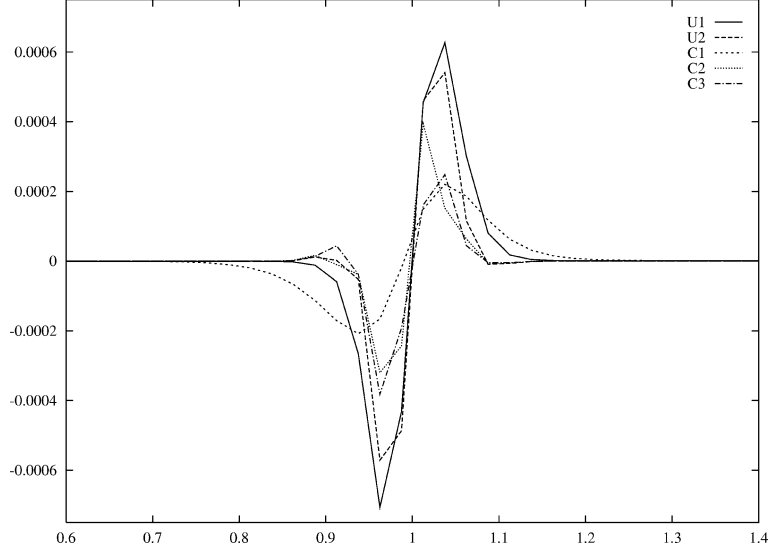
Figure 5. LTE – comparison of different schemes with $N = 80$ (zoom).

Table 4
Linear advection equation – maximal LTE.

N	U1	Rate	U2	Rate	C1	Rate	C2	Rate	C3	Rate
40	3.103e−04	–	2.702e−04	–	1.436e−04	–	9.728e−05	–	1.298e−04	–
80	1.102e−04	1.49	7.032e−05	1.94	3.804e−05	1.92	9.222e−05	0.08	1.726e−05	2.91
160	3.032e−05	1.86	1.936e−05	1.86	9.686e−06	1.97	1.163e−05	2.99	7.968e−06	1.12
320	7.595e−06	2.00	6.708e−06	1.53	2.453e−06	1.98	9.810e−06	0.25	2.976e−06	1.42
640	1.897e−06	2.00	2.463e−06	1.45	6.181e−07	1.99	4.924e−06	0.99	9.598e−07	1.63
1280	4.741e−07	2.00	9.088e−07	1.44	1.554e−07	1.99	1.346e−06	1.87	2.286e−07	2.07

in convergence rates is also observed in the case of the system of Euler equations – $O(\Delta)$ near a shock, and somewhat larger order of up to $O(\Delta^2)$ near a contact discontinuity (see section 4.2).

Example 3 (Buckley–Leverett equation). In this example, we solve the Riemann problem for the Buckley–Leverett equation,

$$
\begin{cases}
u_t + \left[\dfrac{u^2}{u^2 + (1-u)^2} \bigl(1 - k(1-u)^2\bigr) \right]_x = 0, & x \in [0, 1], \\[2mm]
u(x, 0) = \begin{cases}
0, & 0 \leqslant x < 1 - \dfrac{1}{\sqrt{2}}, \\[2mm]
1, & 1 - \dfrac{1}{\sqrt{2}} \leqslant x \leqslant 1,
\end{cases}
\end{cases}
$$

with gravitation ($k = 5$) and without gravitation ($k = 0$).
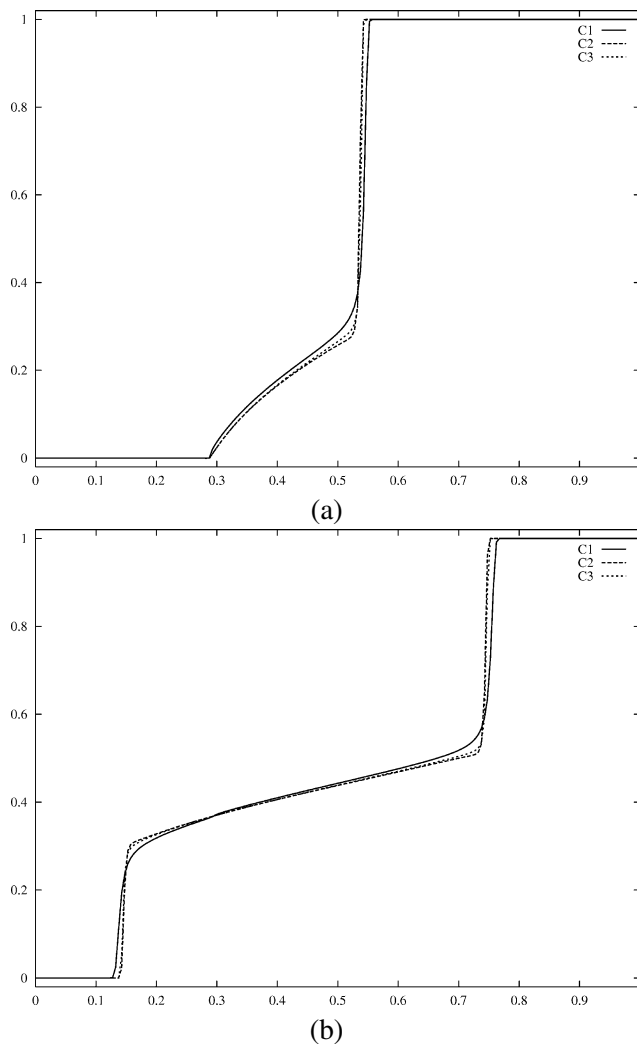
Figure 6. Buckley–Leverett solution, (a) $k = 0$, (b) $k = 5$.

In this nonconvex case, the solution consists of shock-rarefaction ($k = 0$) or shock-rarefaction-shock ($k = 5$) wave configurations. The numerical solutions at time $T = 2$, computed by the central schemes C1, C2 and C3 with $N = 200$ are shown in figure 6.

The corresponding LTEs are displayed in figure 7. As can be observed in these figures, the LTE clearly identifies all the shock regions, figures 7(a), (b). A zoom at the corner of the rarefaction wave (figure 7(c)) reveals that the LTE there, while significantly smaller than at the shock, is still large compared to the smooth regions. The LTE might thus be useful in identifying sharp rarefaction corners, particularly in high-order schemes (see also the Euler system in section 4.2).
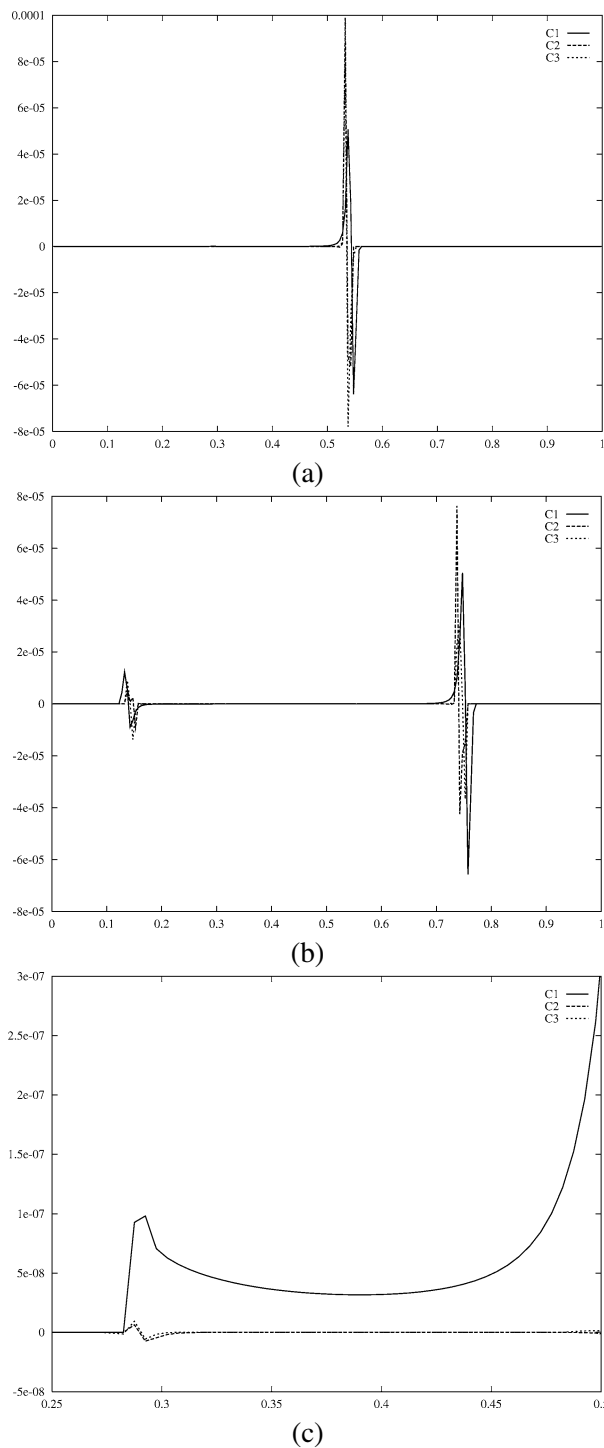
Figure 7. Buckley–Leverett, (a) $k = 0$, LTE; (b) $k = 5$, LTE; (c) $k = 0$, LTE – zoom.
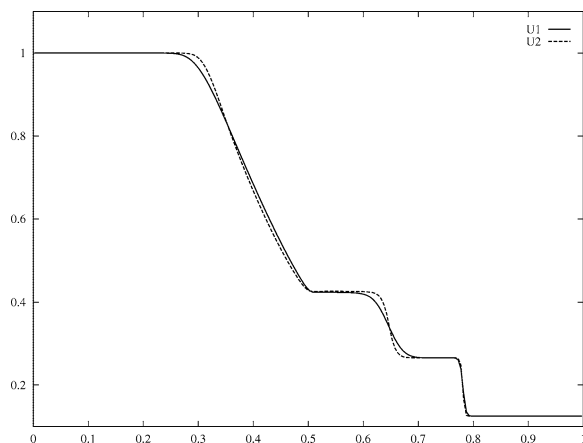
4.2. Euler equations of gas dynamics

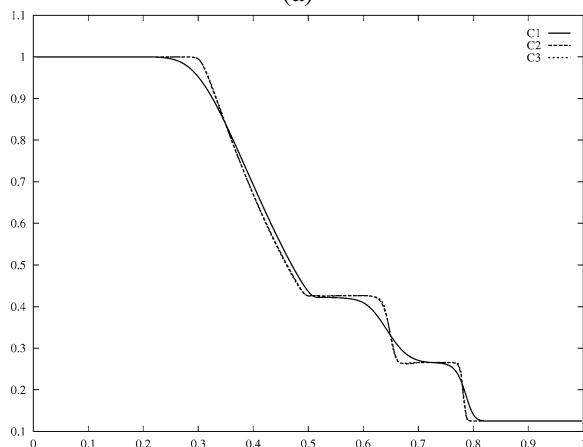In this section, we compute the LTE for the Euler system,

$$\frac{\partial}{\partial t} \begin{bmatrix} \rho \\ m \\ E \end{bmatrix} + \frac{\partial}{\partial x} \begin{bmatrix} m \\ \rho u^2 + p \\ u(E + p) \end{bmatrix} = 0, \quad p = (\gamma - 1) \cdot \left(E - \frac{1}{2}\rho u^2 \right),$$

where ρ, u, $m = \rho u$, p and E are the density, velocity, momentum, pressure and total energy, respectively. The initial data corresponds to the shock-tube problem proposed by Sod [15],

$$(\rho, u, p)(x, 0) = \begin{cases} (1.000, 0, 1.0), & x < 0.5, \\ (0.125, 0, 0.1), & x > 0.5. \end{cases}$$

(a)

(b)

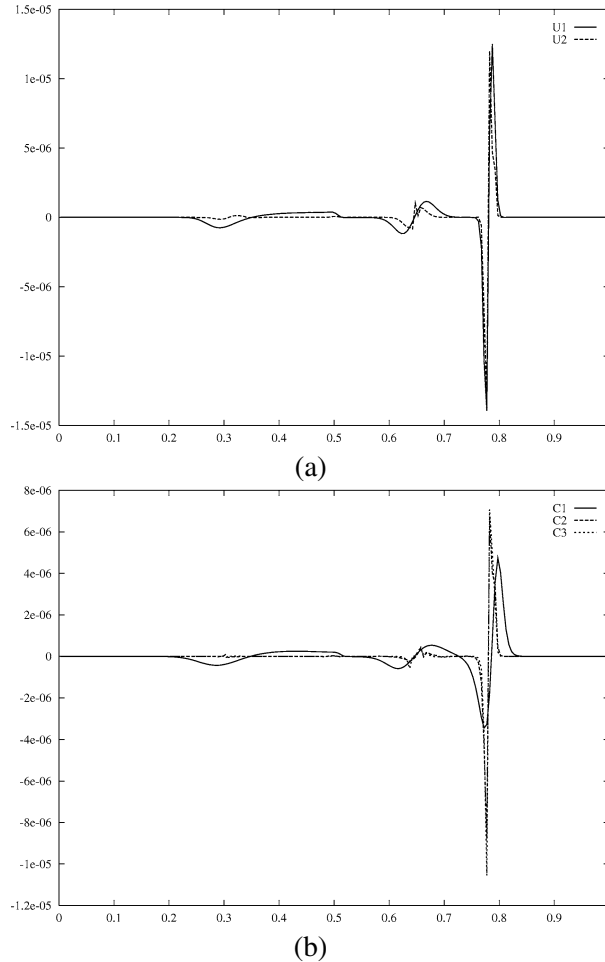Figure 8. Density (a) by upwind schemes, (b) by central schemes.

Figure 9. Density LTE (a) by upwind schemes, (b) by central schemes.

The density and the corresponding LTE for $N = 200$ are displayed at time $T = 0.16$ in figures 8, 9. The convergence rates of the LTE, measured near the rarefaction corner, in the smooth part of the rarefaction wave, near the contact discontinuity and near the shock, are shown in tables 5–8, respectively.

In agreement with the theory, which is restricted for scalar equations, the convergence rate in the smooth subregion appears to be $O(\Delta^{r+2})$, and only $O(\Delta)$ near the shock. One can also observe that the convergence rate in the neighborhood of the (linear) contact wave is between $O(\Delta)$ and $O(\Delta^2)$, similarly to the linear advection case.

Again we observe the significant difference in the order of magnitude of the LTE between the smooth and nonsmooth regions (for example, $\sim 10^{-5}$, $\sim 10^{-6} \sim 10^{-7}$ and $\sim 10^{-9}$ near the shock, contact, rarefaction corner and smooth regions, respectively, for the second-order schemes with $N = 200$).

Table 5
Euler equation of gas dynamics, maximal density LTE near the rarefaction corner [0.25, 0.35].

N	U1	Rate	U2	Rate	C1	Rate	C2	Rate	C3	Rate
100	3.931e−06	–	8.478e−07	–	2.173e−06	–	5.912e−07	–	1.385e−07	–
200	7.524e−07	2.39	1.380e−07	2.62	4.302e−07	2.34	8.515e−08	2.80	6.713e−08	1.04
400	1.403e−07	2.42	2.225e−08	2.63	8.334e−08	2.37	5.399e−09	3.98	1.051e−08	2.68
800	2.589e−08	2.44	3.554e−09	2.65	1.585e−08	2.39	2.476e−09	1.12	1.223e−09	3.10
1600	4.706e−09	2.46	5.607e−10	2.66	2.963e−09	2.42	1.236e−09	1.00	6.263e−10	0.97
3200	8.481e−10	2.47	8.861e−11	2.66	5.462e−10	2.44	8.748e−11	3.82	1.392e−10	2.17
6400	1.512e−10	2.48	1.394e−11	2.67	9.960e−11	2.46	2.935e−11	1.58	1.867e−11	2.90

Table 6
Euler equation of gas dynamics, maximal density LTE in the smooth subregion [0.35, 0.45].

N	U1	Rate	U2	Rate	C1	Rate	C2	Rate	C3	Rate
100	2.389e−06	–	1.818e−07	–	1.694e−06	–	2.037e−08	–	3.371e−08	–
200	3.416e−07	2.81	7.908e−09	4.52	2.467e−07	2.78	1.372e−09	3.89	9.266e−10	5.19
400	4.525e−08	2.92	6.633e−10	3.58	3.503e−08	2.82	9.344e−11	3.88	1.060e−10	3.13
800	5.807e−09	2.96	3.802e−11	4.12	4.909e−09	2.84	6.485e−12	3.85	7.633e−12	3.80
1600	7.403e−10	2.97	2.060e−12	4.21	6.775e−10	2.86	4.393e−13	3.88	2.223e−13	5.10
3200	9.376e−11	2.98	1.156e−13	4.16	9.212e−11	2.88	2.929e−14	3.91	1.047e−14	4.41
6400	1.181e−11	2.99	6.761e−15	4.10	1.236e−11	2.90	1.917e−15	3.93	4.795e−16	4.45

Table 7
Euler equation of gas dynamics, maximal density LTE near the contact discontinuity [0.6, 0.7].

N	U1	Rate	U2	Rate	C1	Rate	C2	Rate	C3	Rate
100	4.729e−06	–	3.328e−06	–	2.310e−06	–	1.103e−06	–	1.006e−06	–
200	1.155e−06	2.03	1.028e−06	1.70	5.875e−07	1.98	4.295e−07	1.36	5.381e−07	0.90
400	2.899e−07	1.99	2.206e−07	2.22	1.497e−07	1.97	1.793e−07	1.26	1.336e−07	2.01
800	7.285e−08	1.99	8.418e−08	1.39	3.800e−08	1.98	4.689e−08	1.94	2.398e−08	2.48
1600	1.834e−08	1.99	2.629e−08	1.68	9.621e−09	1.98	1.493e−08	1.65	6.426e−09	1.90
3200	4.623e−09	1.99	2.831e−09	3.22	2.425e−09	1.99	6.923e−09	1.11	2.099e−09	1.61
6400	1.163e−09	1.99	1.160e−09	1.29	6.104e−10	1.99	3.074e−09	1.17	5.070e−10	2.05

Table 8
Euler equation of gas dynamics, maximal density LTE in the entire interval [0, 1].

N	U1	Rate	U2	Rate	C1	Rate	C2	Rate	C3	Rate
100	2.712e−05	–	2.466e−05	–	1.005e−05	–	1.687e−05	–	1.935e−05	–
200	1.395e−05	0.96	1.314e−05	0.91	4.737e−06	1.09	8.793e−06	0.94	1.055e−05	0.88
400	7.354e−06	0.92	7.235e−06	0.86	2.308e−06	1.04	4.541e−06	0.95	5.874e−06	0.84
800	3.904e−06	0.91	4.279e−06	0.76	1.126e−06	1.04	2.208e−06	1.04	2.980e−06	0.98
1600	1.911e−06	1.03	2.289e−06	0.90	5.702e−07	0.98	9.052e−07	1.29	1.120e−06	1.41
3200	9.097e−07	1.07	8.741e−07	1.39	2.872e−07	0.99	5.600e−07	0.69	7.058e−07	0.67
6400	4.814e−07	0.92	4.917e−07	0.83	1.413e−07	1.02	2.836e−07	0.98	3.464e−07	1.03

Acknowledgements

References

[1] B. Cockburn, C. Johnson, C.-W. Shu and E. Tadmor, *Advanced Numerical Approximation of Nonlinear Hyperbolic Equations*, C.I.M.E. Course in Cetraro, Italy, June 1997, ed. A. Quarteroni, Lecture Notes in Mathematics, Vol. 1697 (Springer, New York, 1998).

[2] C. de Boor, *A Practical Guide to Splines*, Applied Mathematical Sciences, Vol. 27 (Springer, New York, 1978).

[3] E. Godlewski and P.-A. Raviart, *Numerical Approximation of Hyperbolic Systems of Conservation Laws* (Springer, New York, 1996).

[4] P. Houston and E. Süli, Local a posteriori error indicators for hyperbolic problems, Technical Report NA-97/14, Oxford University Computing Laboratory (1997), `http://web.comlab.ox.ac.uk/oucl/publications/natr/na-97-14.html`.

[5] S. Karni, A. Kurganov and G. Petrova, A smoothness indicator for adaptive algorithms for hyperbolic systems, J. Comput. Phys. 178 (2002) 323–341.

[6] D. Kröner, *Numerical Schemes for Conservation Laws* (Wiley, Chichester, 1997).

[7] A. Kurganov, Conservation laws: stability of numerical approximations and nonlinear regularization, Ph.D. thesis, Tel-Aviv University, Israel (1997).

[8] A. Kurganov, S. Noelle and G. Petrova, Semi-discrete central-upwind scheme for hyperbolic conservation laws and Hamilton–Jacobi equations, SIAM J. Sci. Comput. 23 (2001) 707–740.

[9] R. LeVeque, *Finite Volume Methods for Hyperbolic Problems*, Cambridge Texts in Applied Mathematics (Cambridge Univ. Press, Cambridge, 2002).

[10] H. Nessyahu and E. Tadmor, The convergence rate of approximate solutions for nonlinear scalar conservation laws, SIAM J. Numer. Anal. 29 (1992) 1505–1519.

[11] H. Nessyahu, E. Tadmor and T. Tassa, The convergence rate of Godunov type schemes, SIAM J. Numer. Anal. 31 (1994) 1–16.

[12] R. Richtmyer and K.W. Morton, *Difference Methods for Initial-Value Problems*, 2nd ed. (Interscience, New York, 1967).

[13] P. L. Roe, Approximate Riemann solvers, parameter vectors and difference schemes, J. Comput. Phys. 43 (1981) 357–372.

[14] J. Smoller, *Shock Waves and Reaction Diffusion-Equations*, 2nd ed., Grundleheren Series, Vol. 258 (Springer, New York, 1994).

[15] G. Sod, A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws, J. Comput. Phys. 22 (1978) 1–31.

[16] E. Süli, A posteriori error analysis and adaptivity for finite element approximations of hyperbolic problems, Lecture Notes in Computer Science and Engineering, eds. D. Kröner, M. Ohlberger and C. Rhode, Vol. 5 (Springer, New York, 1998); also, Technical Report NA-97/21, Oxford University Computing Laboratory (1997); `http://web.comlab.ox.ac.uk/oucl/publications/natr/na-97-21.html`.

[17] E. Tadmor, Local error estimates for discontinuous solutions of nonlinear hyperbolic equations, SIAM J. Numer. Anal. 28 (1991) 891–906.

[18] E. Tadmor and T. Tang, Pointwise convergence rate for nonlinear conservation laws, in: *Proc. of the 7th Internat. Conf. on Hyperbolic Problems*, Zürich, 1998, eds. M. Fey and R. Jeltsch, International Series of Numerical Mathematics, Vol. 130 (Birkhäuser, Basel, 1999) pp. 925–934.

[19] E. Tadmor and T. Tang, Pointwise error estimates for scalar conservation laws with piecewise smooth solutions, SIAM J. Numer. Anal. 36 (1999) 1739–1756.