COMMUN. MATH. SCI. © 2015 International Press
Vol. 13, No. 6, pp. 1501–1514

MUSCL RECONSTRUCTION AND HAAR WAVELETS[*]

LAURENT GOSSE[†]

Abstract. MUSCL extensions (*Monotone Upstream-centered Schemes for Conservation Laws*) of the Godunov numerical scheme for scalar conservation laws are shown to admit a rather simple reformulation when recast in the formalism of the Haar multi-resolution analysis of $L^2(\mathbb{R})$. By pursuing this wavelet reformulation, a seemingly new MUSCL-WB scheme is derived for advection-reaction equations which is stable for a Courant number up to 1 (instead of roughly $\frac{1}{2}$). However these high-order reconstructions aren't likely to improve the handling of delicate nonlinear wave interactions in the involved case of systems of Conservation/Balance laws.

Key words. Godunov scheme, Haar wavelets, multi-resolution analysis, MUSCL reconstruction, second-order resolution (SOR), slope-limiter, wave interactions, well-balanced (WB) scheme.

AMS subject classifications. 65M06, 65T60, 35Q35.

1. Introduction

The goal of this text is to recast the widely-used MUSCL high-order schemes for computing the entropy solution of a one-dimensional convex scalar conservation law,

$$\partial_t u + \partial_x f(u) = 0, \qquad u(t=0,\cdot) = u_0 \in L^1 \cap BV(\mathbb{R}), \qquad (t,x) \in \mathbb{R}_*^+ \times \mathbb{R}, \tag{1.1}$$

into the formalism of a multi-resolution analysis of $L^2(\mathbb{R})$ derived from the Haar wavelet. For convenience, we shall always work with a Cartesian uniform computational grid, determined by a space-step Δx and a time-step Δt satisfying the standard homogeneous CFL restriction. Let $J \in \mathbb{Z}$ be fixed, we select in a first stage:

$$\Delta x = 2^{-J}, \qquad \max|f'(u)|\Delta t \le \frac{\Delta x}{2} = 2^{-J-1}.$$

1.1. The standard Godunov scheme.

By defining $C_k = (x_{k-\frac{1}{2}}, x_{k+\frac{1}{2}})$ as the generic computational cell of width Δx centered on $x_k = k\Delta x$, $k \in \mathbb{Z}$, one may apply the Divergence Theorem on any rectangle $C_k \times (t^n, t^{n+1})$ in order to derive a mass-preserving numerical scheme for (1.1):

$$\int_{C_k} u(t^{n+1},x)\mathrm{d}x = \int_{C_k} u(t^n,x)\mathrm{d}x - \int_{t^n}^{t^{n+1}} f\left(u(\tau, x_{k+\frac{1}{2}})\right) - f\left(u(\tau, x_{k-\frac{1}{2}})\right)\mathrm{d}\tau.$$

This is equivalent to writing down the weak formulation of (1.1) with test-functions being the indicator functions of C_k, denoted $\chi(C_k)$. Hereafter, we use the standard notation $u_k^n = \int_{C_k} u(t^n,x)\frac{\mathrm{d}x}{\Delta x}$. Yet the observation leading to the Godunov scheme is the following: in case $u(t^n,\cdot)$ is constant on each computational cell C_k, then the boundary flux terms can be explicitly computed by resolving all the discontinuities, that is to say, Riemann problems at both interfaces $x_{k\pm\frac{1}{2}}$. Moreover, since Riemann fans $\omega(\frac{x}{t}; u^L, u^R)$ display a self-similar structure, one has a nice simplification,

$$\int_{t^n}^{t^{n+1}} f\left(u(\tau, x_{k+\frac{1}{2}})\right)\mathrm{d}\tau = \Delta t \cdot f\left(\omega(0; u_k^n, u_{k+1}^n)\right). \tag{1.2}$$

---

[*] Received: February 25, 2014; accepted (in revised form): September 14, 2014. Communicated by Shi Jin.

[†] IAC–CNR "Mauro Picone", Via dei Taurini 19, 00185 Roma, Italy (l.gosse@ba.iac.cnr.it).

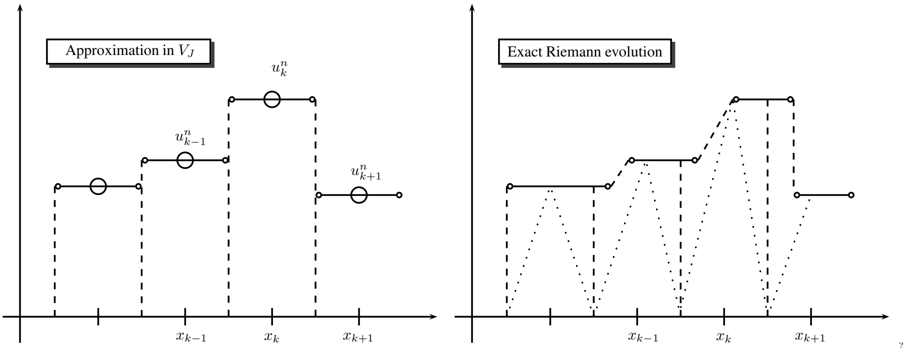

FIG. 1.1. *Piecewise constant approximation (left) and exact Riemann fans (right) in (1.4).*

When seeking an explicit time-marching algorithm, one may want to get rid of the Riemann solution ω, and it can be shown that (1.2) defines a smooth and consistent *numerical flux function* denoted by $F:\mathbb{R}^2\to\mathbb{R}$,

$$\forall u,v\in\mathbb{R}^2,\qquad F(u,v)=f(\omega(0;u,v))=\begin{cases}\min\limits_{u\le\xi\le v} f(\xi) & \text{if } u\le v\\ \max\limits_{v\le\xi\le u} f(\xi) & \text{if } u>v.\end{cases}\tag{1.3}$$

Now, let's consider another formulation of this numerical scheme: denote by $\mathcal{P}_J$ the L^2-projector onto the space of piecewise-constant functions on the computational grid, and $\mathcal{E}_J(t)$ the exact Riemann evolution operator[1] at each interface point $x_{k+\frac{1}{2}}=(k+\frac{1}{2})\Delta x=(k+\frac{1}{2})2^{-J}$ of the grid, the Godunov approximation reads:

$$\forall n\in\mathbb{N},\qquad u^{\Delta x}(t^n,\cdot)=[\mathcal{P}_J\circ\mathcal{E}_J(\Delta t)]^n\mathcal{P}_J(u_0).\tag{1.4}$$

Godunov wipes all the details at a finer scale than the grid by layered local averaging.

1.2. Scaling function and the Multi-Resolution formalism.

We recall the definition of a *Multi-Resolution Analysis* (MRA, [25]) as $L^1\cap BV(\mathbb{R})\subset L^2(\mathbb{R})$.

DEFINITION 1.1. *A sequence of nested (scale-limited) subspaces* $V_j\subset L^2(\mathbb{R})$ *is called a* **Multi-Resolution Analysis** *of* $L^2(\mathbb{R})$ *if* $\{0\}\subset\cdots\subset V_{-1}\subset V_0\subset V_1\subset\cdots\subset L^2(\mathbb{R})$. *Moreover, the following properties must hold:*

- *for all* $f\in L^2(\mathbb{R})$, $\|\mathcal{P}_jf-f\|_{L^2}\to 0$ *as* $j\to+\infty$ *also,* $\mathcal{P}_jf\to 0$ *as* $j\to-\infty$.
- *if* $f(x)\in V_j$, *then* $f(\frac{x}{2})\in V_{j-1}$ *and for all* $k\in\mathbb{Z}$, $f(x-2^jk)\in V_j$.
- *there exists a shift-invariant orthonormal base of* V_0 *given by the* **scaling function** $\varphi_k(x)=\varphi(x-k)$ *for* $k\in\mathbb{Z}$.

Hence $\mathcal{P}_j$ stands for the orthogonal projector onto the subspace V_j. Wavelet spaces W_j are defined as the **orthogonal complement** of V_j in V_{j+1}: $V_{j+1}=V_j\oplus W_j$. From φ_k, the base of V_0, one deduces a base of V_j, $j\in\mathbb{Z}$, by simple dilatation,

$$\forall k\in\mathbb{Z},\qquad \varphi_{j,k}(x)=\sqrt{2^j}\varphi_k(2^jx)=\sqrt{2^j}\varphi(2^jx-k).\tag{1.5}$$

---

[1] Actually the scale index J isn't indispensable and one may denote $\mathcal{E}$ as the exact (entropy) solution operator: however, we shall keep on displaying J (or $J+1$) hereafter for ease of reading.

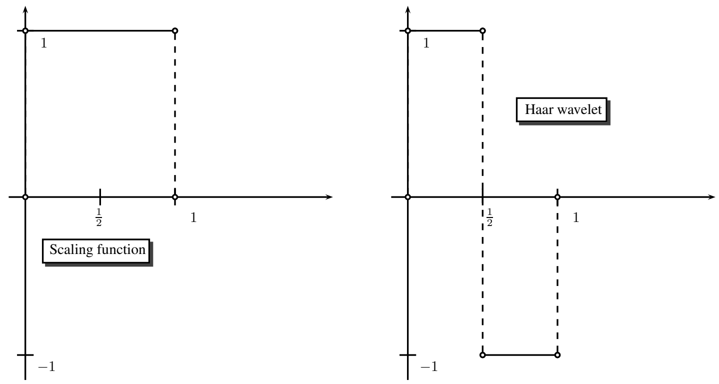

FIG. 1.2. *Haar's scaling function (left) and wavelet (right).*

Thus, the orthogonal projection of f onto the **scale-limited subspace** V_j reads:

$$\mathcal{P}_j f=\sum_{k\in\mathbb{Z}}<f,\varphi_{j,k}>\varphi_{j,k},\qquad <f,\varphi_{j,k}>=\int_{\mathbb{R}}f(x)\varphi_{j,k}(x)\mathrm{d}x,\tag{1.6}$$

which is the best approximation of f in V_j in the least-squares sense. In the context of applications to the Godunov scheme (1.4), there exists a scaling function (hence a MRA) of particular interest, which is part of the "Haar system" (see Figure 1.2),

$$\varphi(x+\frac{1}{2})=\chi([0,1]),\qquad \psi(x+\frac{1}{2})=\chi\left([0,\frac{1}{2})\right)-\chi\left((\frac{1}{2},1]\right).\tag{1.7}$$

The shift factor of $\frac{1}{2}$ is necessary in order to fit with the notation of (1.4), that is, to ensure that the Haar scaling functions match the indicator of each computational cell C_k. A simple observation is that, thanks to the definition (1.7), for a computational grid for which $\Delta x=2^{-J}$, the Godunov approximation reads now:

$$\forall n\in\mathbb{N},\qquad u^{\Delta x}(t^n,\cdot)=\frac{1}{\sqrt{2^J}}\sum_{k\in\mathbb{Z}}u_k^n\ \varphi_{J,k},\tag{1.8}$$

where the initial coefficients are obtained through:

$$u_k^{n=0}=\frac{1}{\Delta x}\int_{C_k}u_0(x)\mathrm{d}x=\sqrt{2^J}\int_{\mathbb{R}}u_0(x)\sqrt{2^J}\varphi\left(2^Jx-k\right)\mathrm{d}x=2^{\frac{J}{2}}<u_0,\varphi_{J,k}>.$$

1.3. Main theorem and outline.

Having at hand the expressions (1.7) of both the Haar father and mother wavelets, we see that MUSCL reconstructions rewrite as a set of (mother) wavelet corrections:

THEOREM 1.2. *Let $\Delta x=2^{-J}$ be the grid's parameter and $u_0\in L^1\cap BV(\mathbb{R})$ be Cauchy data for (1.1). For any TVD-admissible slope-limiter function $\phi:\mathbb{R}\to[0,2]$, let R_ϕ stand for the associated MUSCL reconstruction (2.2), then $R_\phi:V_J\to V_{J+1}$ and*

$$R_\phi\circ\mathcal{P}_J(u_0)=2^{-\frac{J}{2}}\left(\sum_{k\in\mathbb{Z}}u_k^0\ \varphi_{J,k}-\sum_{k\in\mathbb{Z}}\phi(r_k^0)\frac{u_{k+1}^0-u_k^0}{2}\psi_{J,k}\right),\quad r_k^0=\frac{u_k^0-u_{k-1}^0}{u_{k+1}^0-u_k^0}.\tag{1.9}$$

Accordingly, $Id - R_\phi \circ \mathcal{P}_J$ *maps* V_J *into* W_J *and there is a "back-projection" property:*

$$\mathcal{P}_J \circ R_\phi \circ \mathcal{P}_J = \mathcal{P}_J, \qquad L^2(\mathbb{R}) \to V_J. \tag{1.10}$$

Both equations (1.9) and (1.10) imply that for any limiter ϕ, MUSCL reconstructions induce only a "fluctuation component" in W_J, so it can't recover the type of sub-grid details (in the elementary Riemann fans) which are discarded in the Godunov averaging step, like the ones displayed for instance on the right part of Figure 1.1. It is possible to devise local projectors furnishing exact solutions at certain times; however, it doesn't seem possible to recast them in this "Haar wavelet framework" because they result from an interpolation between $\mathcal{P}_J$ and the random sampling of Glimm, see [13]. It is important to remember that, even if R_ϕ generates an approximation in V_{J+1}, usual MUSCL schemes still use the Riemann evolution operator $\mathcal{E}_J$, that is to say, the (new) discontinuities located in x_k aren't resolved (see Figure 2.2):

$$\forall n \in \mathbb{N}, \qquad u^{\Delta x,\phi}(t^n,\cdot) = [\mathcal{P}_J \circ \mathcal{E}_J(\Delta t) \circ R_\phi]^n \mathcal{P}_J(u_0). \tag{1.11}$$

In Section 2, we prove the Main Theorem and in Section 3, advection-reaction equations and some issues raised by interaction of waves for systems [1, 12, 14] are studied. Finally, in Appendix A, some facts about "evolutionary errors" for discretizations based on the *Method of Lines* are recalled following mainly [9, 26, 29, 30, 39, 40].

REMARK 1.3. Hereafter, ϕ stands for a slope-limiter. One may set up a flux-limiter instead, but showing an analogy with wavelet formalism would be more difficult. Generally, the term "flux-limiter" is used when it acts directly on fluxes, and "slope-limiter", when it acts just on states. Both have the same mathematical form and have the effect of limiting the solution's gradient near shocks or local extrema.

2. Proof of the main theorem

MUSCL-based numerical schemes extend the idea of using a linear piecewise approximation to each cell by using *slope limited* left and right extrapolated states[2]. With the Godunov flux (1.3), they yield second-order resolution (SOR), Total-Variation Diminishing (TVD) time-marching processes after some approximations.

2.1. MUSCL reconstruction as an extrapolation process.

Let's recall how the "extrapolated states" are derived: for any indexes $k,n \in \mathbb{Z} \times \mathbb{N}$, the Godunov averaging furnished an approximate (formally first-order) value $u_k^n \simeq u(t^n,x_k)$, from which a piecewise-linear reconstruction is deduced in each cell,

$$v_k^n : C_k \to \mathbb{R}, \qquad v_k^n(x) = u_k^n + (x - x_k)\sigma_k^n. \tag{2.1}$$

A first way to proceed is by analogy with Lax–Wendroff second-order schemes with $f(u) = u$: so, a convenient definition of the local slopes reads,

$$\sigma_k^n = \frac{u_{k+1}^n - u_k^n}{\Delta x}\phi(r_k^n), \qquad r_k^n = \frac{u_k^n - u_{k-1}^n}{u_{k+1}^n - u_k^n}.$$

The slope-limiter ϕ must satisfy several constraints in order to ensure both the TVD and second-order accuracy, see [35].

---

[2] The classical Donoho–Stark criterion suggests that, as its scaling function is discontinuous, performing scale-limited extrapolation in the Haar multi-resolution spaces may be unstable [10, 17]. Hereafter, extrapolation refers to *piecewise polynomial* extrapolation, and not to *scale-limited* one.

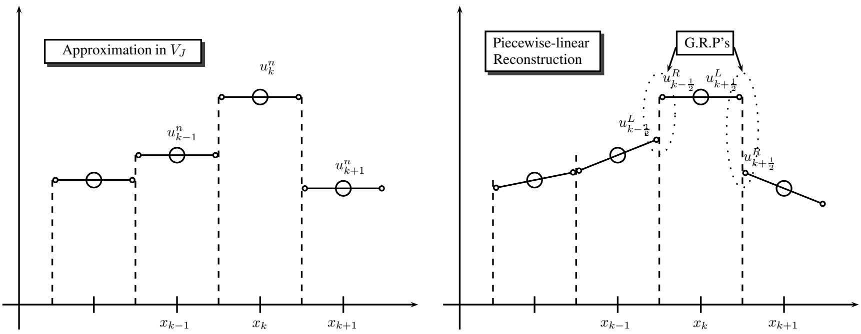

FIG. 2.1. *Piecewise-linear reconstruction leading to Generalized Riemann Problems.*

Similar to the construction of the Godunov scheme, we must now resolve the interfacial discontinuities separating linear polynomials instead of constant states: this is usually called a *Generalized Riemann Problem* (see [36] pages 427–9 and Figure 2.1). The issue is, quoting Osher *"obtaining the exact solution to this nonlinear initial value problem with piecewise linear initial data is a nontrivial business"* (see also [2, 3]): in particular, the simplification (1.2) is lost along with the self-similarity property, except for the astute derivation presented in [16]. So in the vast majority of cases, the MUSCL algorithm limits itself to solving the usual (self-similar) *Riemann problems with extrapolated states* at each interface $x_{k\pm\frac{1}{2}}$ of the grid with $\Delta x = 2^{-J}$, hence $\mathcal{E}_J(\Delta t)$. For the advection equation, i.e. $f(u) = a \cdot u$, the global error generated by MINMOD reconstructions was analyzed in [28]: a convergence rate slightly greater than $\frac{1}{2}$ was obtained with weak solutions.

REMARK 2.1. Another way to motivate MUSCL piecewise-linear reconstructions is to work out the ODE system obtained by semi-discretization in space (the "Method of Lines", evoked in [24]) in order to obtain a Local (space-) Truncation Error in Δx^2 for smooth exact solutions u: see our Appendix A and Verwer's papers [29, 30, 39, 40].

2.2. The Haar wavelet fluctuation. As we explained in the former subsection, a more correct representation of the MUSCL algorithm is displayed in Figure 2.2, where there remains only a set of extrapolated states $u^{L/R}_{k-\frac{1}{2}}$ and the corresponding (usual, self-similar) Riemann problems:

$$\forall k \in \mathbb{Z}, \qquad u^R_{k-\frac{1}{2}} = u^n_k - \phi(r^n_k)\frac{u^n_{k+1} - u^n_k}{2}, \qquad u^L_{k+\frac{1}{2}} = u^n_k + \phi(r^n_k)\frac{u^n_{k+1} - u^n_k}{2}. \tag{2.2}$$

Since the local reconstructions (2.1) are odd in the $x - x_k$ variable, it is now obvious that we can rewrite the states (2.2) by means of the Haar wavelet ψ. More precisely, given (1.8) as the Godunov approximation at time t^n in V_J, these states read:

$$u^R_{k-\frac{1}{2}} = u^n_k - \phi(r^n_k)\frac{u^n_{k+1} - u^n_k}{2}\psi(-\frac{1}{2}), \qquad u^L_{k+\frac{1}{2}} = u^n_k - \phi(r^n_k)\frac{u^n_{k+1} - u^n_k}{2}\psi(\frac{1}{2}).$$

Since W_J is the orthogonal complement of V_J in V_{J+1}, the MUSCL extrapolated states furnish a piecewise-constant approximation in the finer scale-limited subspace,

$$2^{-\frac{J}{2}}\left(u^n_k \ \varphi_{J,k}(x) - \phi(r^n_k)\frac{u^n_{k+1} - u^n_k}{2}\psi_{J,k}(x)\right) \in V_{J+1} = V_J \oplus W_J, \tag{2.3}$$

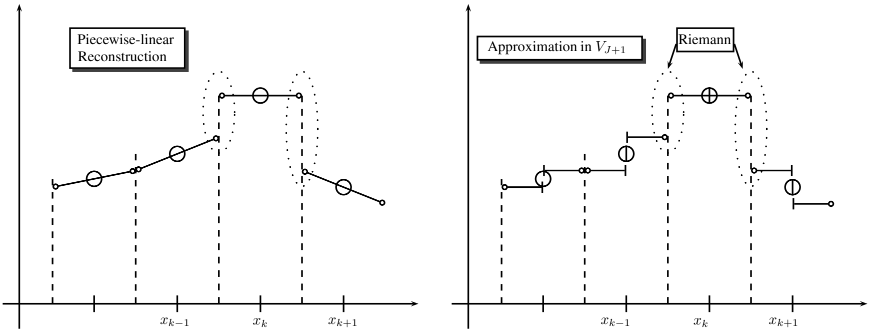

FIG. 2.2. *MUSCL Piecewise-linear reconstruction and approximation in* V_{J+1}.

because Haar wavelets satisfy the following relation:

$$2^{-\frac{J}{2}}\psi_{J,k}(x_{k\pm\frac{1}{2}}) = \psi\left(2^J(k\Delta x \pm \frac{\Delta x}{2}) - k\right) = \psi\left(2^J(k\cdot 2^{-J} \pm 2^{-J-1}) - k\right) = \psi(\pm\frac{1}{2}).$$

A consequence of the formulation (2.2) is that one has $\frac{1}{2}(u^R_{k-\frac{1}{2}} + u^L_{k+\frac{1}{2}}) = u^n_k$ in all the cells C_k. However, such a symmetry should occur for a transonic rarefaction wave.

2.3. The back-projection (conservation) property. Property (1.10) is a direct consequence of the simple observation:

$$\forall k,k' \in \mathbb{Z}^2, \qquad \int_{\mathbb{R}} \varphi_{J,k}(x)\psi_{J,k'}(x)\mathrm{d}x = 0.$$

Indeed, thanks to definition (1.6) and expression (2.3), one sees that by linearity of the integral, the former orthogonality property implies

$$\forall k \in \mathbb{Z}, \qquad \left\langle u^n_k\ \varphi_{J,k} - \phi(r^n_k)\frac{u^n_{k+1} - u^n_k}{2}\psi_{J,k}, \varphi_{J,k}\right\rangle = < u^n_k\ \varphi_{J,k}, \varphi_{J,k} > = u^n_k.$$

This completes the proof of the Main Theorem 1.2. □

3. Inhomogeneous equations, comments and outlook

Looking at Figure 1.1, one sees that the reconstruction on Figure 2.2 cannot yield an improvement in terms of elementary wave interactions. Indeed, in order to be compatible with the sub-grid details which are discarded in the averaging for the case of Figure 1.1, the extrapolation process should address the right-half of the computational cells only. But such a reconstruction process wouldn't belong to W_J: instead, it would have a component in V_J and the "back-projection" (1.10) wouldn't hold (see [23]).

3.1. A new MUSCL-WB scheme for advection-reaction. A case where G.R.P.'s arising from piecewise-linear reconstructions as in Figure 2.1 is when $f(u) = au$, with $a > 0$ taken for convenience: formula (1.2) modifies into,

$$\int_{t^n}^{t^{n+1}} a\cdot v^n_k\left(x_{k+\frac{1}{2}} - a(\tau - t^n)\right)\mathrm{d}\tau = \underbrace{a\Delta t\left(u^n_k + \frac{\sigma^n_k\Delta x}{2}\right)}_{usual\ MUSCL\ flux} - \underbrace{a^2\frac{\sigma^n_k\Delta t^2}{2}}_{correction\ GRP}, \tag{3.1}$$

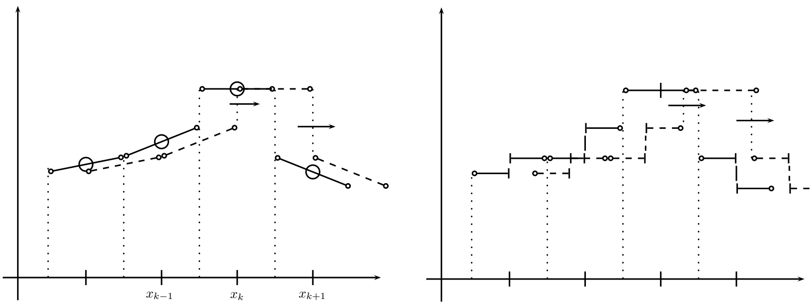

FIG. 3.1. *Derivation of numerical fluxes: GRP's (left) and approximation in V_{J+1} (right).*

so numerical fluxes do depend on time (see (5.3b) in [20]). In contrast, when one simply substitutes G.R.P.'s with the usual self-similar Riemann problems, such a phenomenon doesn't show up, but the CFL number must be lowered to a value around $\frac{1}{2}$ (see [35]).

Yet, according to Figure 3.1, our reconstruction in V_{J+1} is different because it involves new discontinuities at the center of each cell, x_k. Hence it appears logical to replace the Riemann solver $\mathcal{E}_J$ which handles only the discontinuities at interfaces $x_{k\pm\frac{1}{2}}$ by $\mathcal{E}_{J+1}$ processing jumps at both interfaces and center. The resulting scheme,

$$\tilde{u}^{\Delta x,\phi}(t^n,\cdot) = [\mathcal{P}_J \circ \mathcal{E}_{J+1}(\Delta t) \circ R_\phi]^n \mathcal{P}_J(u_0), \qquad \phi(r) = \frac{r+|r|}{1+|r|}, \tag{3.2}$$

involves numerical fluxes still derived by modifying formula (1.2), see Figure 3.2:

$$a\int_{t^n}^{t^{n+1}} u_k^n - 2^{\frac{J}{2}}\psi_{J,k}\big(x_{k+\frac{1}{2}} - a(\tau - t^n)\big)\frac{\sigma_k^n \Delta x}{2}\mathrm{d}\tau \overset{def}{=} a\Delta t\left(\frac{u_{k+1}^n + u_k^n}{2} - \tilde{Q}_{j+\frac{1}{2}}^n \frac{u_{k+1}^n - u_k^n}{2}\right)$$
$$= a\Delta t \cdot u_k^n + \frac{\sigma_k^n \Delta x}{2}\left(\min(a\Delta t, \frac{\Delta x}{2}) - \max(0, a\Delta t - \frac{\Delta x}{2})\right).$$

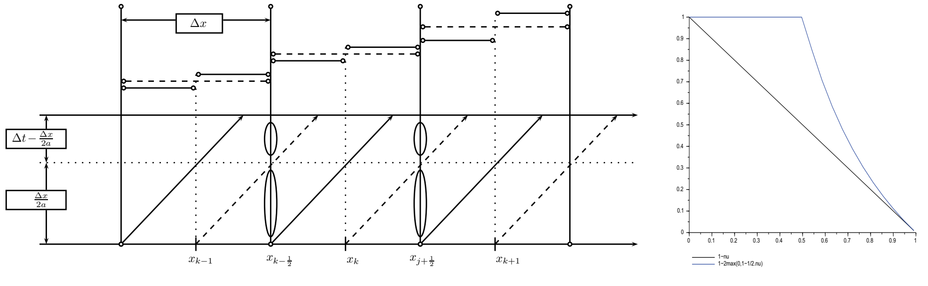

FIG. 3.2. *Illustration of scheme (3.2) (left), and deviation w.r.t. exact GRP fluxes (right).*

A slope-limiter ϕ is indispensable because its numerical viscosity is only,

$$\tilde{Q}_{j+\frac{1}{2}}^n = 1 - \phi(r_j^n)\left(\min(1, \frac{1}{2\nu}) - \max(0, 1 - \frac{1}{2\nu})\right), \qquad \nu = \frac{a\Delta t}{\Delta x}, \tag{3.3}$$
$$= 1 + \phi(r_j^n)\left(2\max(0, 1 - \frac{1}{2\nu}) - 1\right), \quad \left(\text{as } 1 - \min(1, \frac{1}{2\nu}) = \max(0, 1 - \frac{1}{2\nu})\right).$$

One may compare it to $Q^n_{j+\frac{1}{2}}$, the one associated to the numerical flux (3.1):

$$Q^n_{j+\frac{1}{2}}=1-\phi(r^n_j)(1-\nu), \qquad \tilde{Q}^n_{j+\frac{1}{2}}=1-\phi(r^n_j)\left(1-2\max(0,1-\frac{1}{2\nu})\right),$$

see Figure 3.2. All in all, this yields the following (and seemingly new) discretization,

$$u^{n+1}_k=u^n_k-\frac{a\Delta t}{\Delta x}\left[(u^n_k-u^n_{k-1})+\left(2\min(1,\frac{1}{2\nu})-1\right)\frac{\Delta x(\sigma^n_k-\sigma^n_{k-1})}{2}\right], \tag{3.4}$$

which rewrites simply, $u^{n+1}_k=u^n_k-\frac{a\Delta t}{\Delta x}\left(\tilde{u}^n_{k+\frac{1}{2}}-\tilde{u}^n_{k-\frac{1}{2}}\right)$, after having defined,

$$\tilde{u}^n_{k+\frac{1}{2}}=(1-\alpha^n_k)u^n_k+\alpha^n_k u^n_{k+1}, \qquad \alpha^n_k=\left(\min(1,\frac{1}{2\nu})-\frac{1}{2}\right)\phi(r^n_k)\in[0,1].$$

Below we display numerical results for $a=1$, $u_0(x)=\sin^3(2\pi x)$, and 2^6 points in $x\in(0,1)$ with periodic boundary conditions. To prevent unaesthetic staircases in the numerical solution, we sampled initial data on the grid $\{\frac{\Delta x}{2},\frac{3\Delta x}{2},\dots,1-\frac{\Delta x}{2}\}$: results are displayed on the left of Figure 3.3 with $\Delta t=0.7\Delta x$. On its right, one can see the dependence of the time-growth of the L^1-error with respect to the Courant number; even if the error gets big when it goes lower than 0.8, at least it stops growing after a certain time. Since our MUSCL scheme isn't restricted by low CFL numbers, they are less vulnerable to numerical diffusion's bad effects.

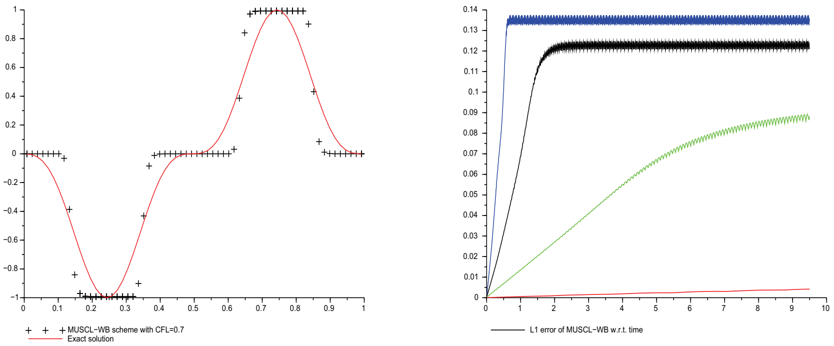

FIG. 3.3. *Advection with CFL=0.7 (left) and L^1-errors (right) for CFL=0.55, 0.7, 0.85, 0.99.*

REMARK 3.1. For $\nu=1$, the scheme (3.4) yields $u^{n+1}_k=u^n_{k-1}$, so it is exact. If $\nu\le\frac{1}{2}$, it reduces to the usual second-order MUSCL scheme. In contrast, for $\frac{1}{2}<\nu<1$,

$$r^n_k=\frac{u^n_k-u^n_{k-1}}{u^n_{k+1}-u^n_k}=1-\frac{u^n_{k+1}-2u^n_k+u^n_{k-1}}{u^n_{k+1}-u^n_k}=1+O(\Delta x), \qquad \text{for } u(t^n,\cdot) \text{ smooth},$$

so $\phi(r^n_k)=\phi(1)+O(\Delta x)$ because ϕ is Lipschitz-continuous and $\phi(1)=1$ [35]. Yet,

$$\begin{aligned}\tilde{u}^n_{k+\frac{1}{2}} &= u^n_k+\tfrac{1}{2}\left(1-(2-\tfrac{1}{\nu})\right)(u^n_{k+1}-u^n_k)(1+O(\Delta x))\\ &= \tfrac{1}{2}(u^n_{k+1}+u^n_k)+O(\Delta x^2)-\left(1-\tfrac{1}{\nu}\right)(u^n_{k+1}-u^n_k)(1+O(\Delta x))\\ &= \tfrac{1}{2}(u^n_{k+1}+u^n_k)+\left(\tfrac{1}{\nu}-1\right)O(\Delta x)+O(\Delta x^2).\end{aligned}$$

According to (A.5), the L.T.E. is $O(\Delta x)$ for $\nu>\frac{1}{2}$ despite the weak viscosity (3.3).

Next, when considering an inhomogeneous equation of the type $\partial_t u + \partial_x u = k(x)u$, the scheme (3.2) can match the WB framework [18] where one solves a “lifted equation”,

$$\partial_t u + \partial_x u - u\partial_x a = 0, \qquad \partial_t a = 0, \qquad (\text{because} \quad \partial_x a(x) = k(x)),$$

which induces a solver $\tilde{\mathcal{E}}$ now including a “standing wave” locally rendering the source,

$$\forall n \in \mathbb{N}, \qquad \tilde{u}^{\Delta x,\phi,WB}(t^n,\cdot) = \Big[\mathcal{P}_J \circ \tilde{\mathcal{E}}_{J+1}(\Delta t) \circ R_\phi\Big]^n \mathcal{P}_J(u_0).$$

In the same manner as for (3.4), we get a simple expression of the resulting scheme,

$$u_k^{n+1} = u_k^n - \frac{\Delta t}{\Delta x}\left(\tilde{u}^n_{k+\frac{1}{2}} - \tilde{u}^n_{k-\frac{1}{2}} \cdot \exp\left(a(x_k) - a(x_{k-1})\right)\right).$$

The exact solution reads $u(t,x) = u_0(x-t)\exp(a(x-t) - a(x))$: on the left of Figure 3.4, our WB scheme is set up for $k(x) \equiv -\frac{1}{2}$ and compared to it with 2^6 grid points and CFL=0.7 at $t = 9.5$.

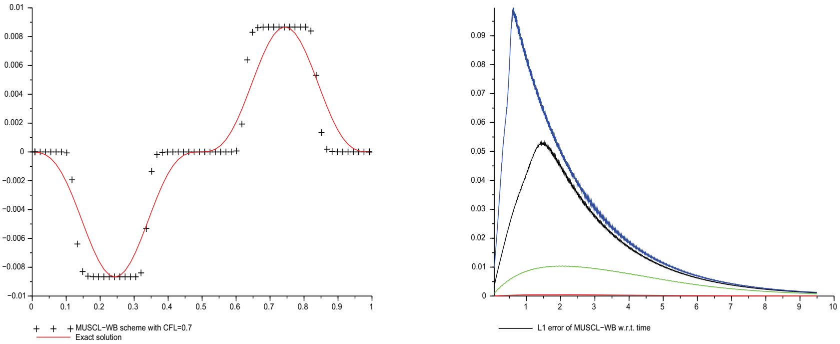

FIG. 3.4. *Exponential decay, CFL*=0.7 *(left);* L^1*-errors (right) for CFL*=0.55, 0.7, 0.85, 0.99.

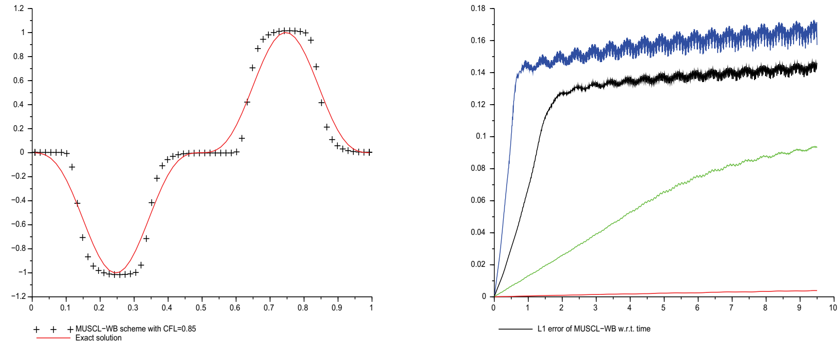

FIG. 3.5. *Oscillating* $k(x)$ *with CFL*=0.85 *(left);* L^1*-errors (right) for CFL*=0.55, 0.7, 0.85, 0.99.

The function $a(x)$ is discretized according to the grid corresponding to V_J, so it jumps only at interfaces $x_{k\pm\frac{1}{2}}$: this means that the discontinuities in x_k are resolved with the homogeneous Riemann solver. A more accurate scheme would be produced if $a(x)$ is sampled on the finer scale of V_{J+1}. In this case, even the discontinuities in x_k's are resolved with the WB Riemann solver. Such a scheme would be well suited for source terms containing an oscillating coefficient $k(x)$: for instance, the rather delicate case where $k(x)=\frac{\cos(4\pi x)}{2}$ is presented in Figure 3.5 on the previous page. Such a MUSCL-WB scheme may be useful for kinetic models involving slow particles [18].

3.2. Wave-interactions and Engquist–Sjogreen counter-example. Now we switch to the more involved case of nonlinear systems of conservation laws: left apart the Temple class, shock curves aren't straight lines in the Hugoniot space. A first problem materializes because intermediate points resulting from the numerical "viscous smearing" of jumps generally do not belong to these curves. This creates spurious (small) waves of other characteristic families in the numerical solution: see Figure 3.6 and [1, 22, 31]. Clearly MUSCL reconstructions cannot noticeably improve this situation which occurs mainly for large discontinuities.

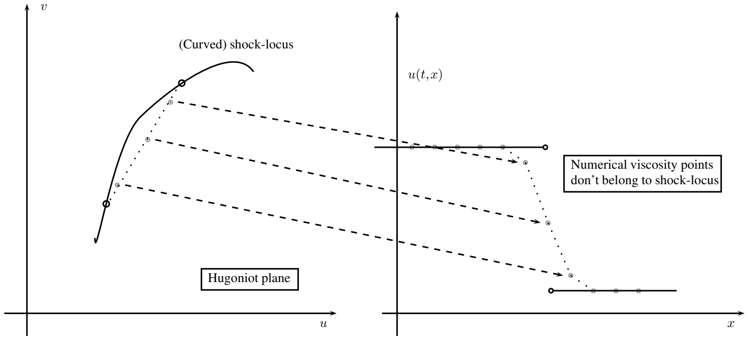

FIG. 3.6. *Numerical viscosity and its effects for systems with curved shock loci.*

About numerical wave interactions, P. L. Roe writes: *It is natural still to feel some anxiety about replacing the discrete jump conditions (Rankine–Hugoniot relations), that hold across an infinitesimally thin shock, with a 'smeared-out' statement of conservation. It seems likely that such a strategy will lead to some sort of unavoidable error. Currently, rather delicate computations of the interaction between a strong shock and weak acoustic waves are not successful unless the shock is either represented as an explicit discontinuity or else the grid spacing is greatly reduced in its vicinity* ([21], page 15).

SOR can fail when Glimm's interaction potential is positive like in a p-system, an interaction between a 1-shock and a 2-rarefaction. Along with numerical smearing of the profile, the local truncation error can reduce from second-order to first-order in the vicinity of the shock wave. When interaction occurs (see Figure 3.7), all the variables are coupled, and the formal accuracy of the outgoing rarefaction wave may not be second-order: this is the essence of what appears in [12], and later in [11, 32, 33].

3.3. Multi-dimensional issues. Most of the existing 2D MUSCL schemes do not completely fit in our Haar wavelet formalism as the scaling function and mother wavelets read $\varphi^{2D}(x,y)=\varphi(x)\varphi(y)$ and

$$\psi^{2D,vert}(x,y)=\varphi(x)\psi(y), \quad \psi^{2D,hor}(x,y)=\psi(x)\varphi(y), \quad \psi^{2D,diag}(x,y)=\psi(x)\psi(y).$$

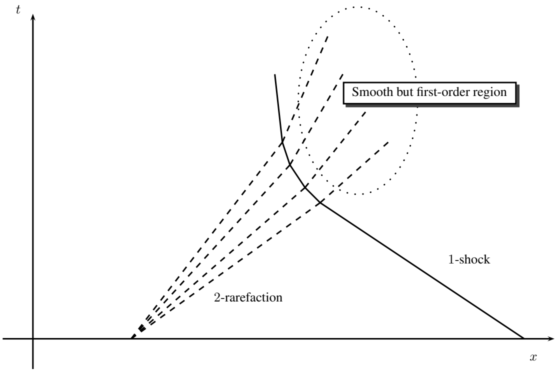

FIG. 3.7. *Nonlinear 2×2 interaction and loss of SOR in a smooth region.*

Hence it perceives a diagonal direction. The issues of numerical wave interactions still exist in 2D, see [7, 32, 38]. Multi-resolution schemes were studied in [8] (also [19, 4]).

Appendix A. Formal analysis of the evolutionary error. Hereafter we follow the conventions of Cullen and Morton [9] in order to shed some light onto the various mechanisms of error creation/propagation (see also [5, 15, 40]).

A.1. Semi-discretization in space (Method of Lines). Let a Cauchy problem for a given partial differential operator $\mathcal{L}$ be,

$$\partial_t u = \mathcal{L}u, \qquad u(t=0,\cdot)=u_0. \tag{A.1}$$

For $\Delta x = 2^{-J}$ fixed and the corresponding griding of the real line, a finite-differences approximation of $\mathcal{L}$ acting on $\Delta x\cdot\mathbb{Z}$ is denoted by $\mathcal{L}_J$, so (A.1) reduces to an (infinite) differential system (*Method of Lines*, Ch. 17 in [24]), with $\tilde{u}(t,\cdot)\in\ell^\infty(\mathbb{Z})$, say:

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{u} = \mathcal{L}_J\tilde{u}, \qquad \tilde{u}(t=0,\cdot)=\mathcal{P}_J u_0, \tag{A.2}$$

for which one can legitimately wonder about the global error $u-\tilde{u}$ at each time $t>0$.

- one "triangulates" $u(t,\cdot)-\tilde{u}(t,\cdot)$ by inserting $\mathcal{P}_J u(t,\cdot)$,

  $$u-\tilde{u} = (Id-\mathcal{P}_J)u + (\mathcal{P}_J u - \tilde{u}) := a_J + e_J,$$

  where a_J is purely an approximation error, which belongs to the wavelet subspace $\cup_{j\ge J} W_j$. On the contrary, e_J stands for an evolutionary error, which may accumulate in time, and satisfies a differential equation,

  $$\frac{\mathrm{d}}{\mathrm{d}t}e_J = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}_J u - \frac{\mathrm{d}}{\mathrm{d}t}\tilde{u} = \mathcal{P}_J\mathcal{L}u - \mathcal{L}_J\tilde{u}. \tag{A.3}$$

- Triangulating again, one gets $\frac{\mathrm{d}e_J}{\mathrm{d}t} = (\mathcal{P}_J\mathcal{L}u - \mathcal{L}_J\mathcal{P}_J u) + (\mathcal{L}_J\mathcal{P}_J u - \mathcal{L}_J\tilde{u})$, so

  $$\frac{\mathrm{d}}{\mathrm{d}t}e_J + (\mathcal{L}_J\tilde{u} - \mathcal{L}_J\mathcal{P}_J u) = (\mathcal{P}_J\mathcal{L}u - \mathcal{L}_J\mathcal{P}_J u) := L.T.E.,$$

  and by substituting $\tilde{u}$ by $\mathcal{P}_J u - e_J$, we get finally:

  $$\frac{\mathrm{d}}{\mathrm{d}t}e_J + \big[\mathcal{L}_J(\mathcal{P}_J u - e_J) - \mathcal{L}_J\mathcal{P}_J u\big] = L.T.E., \quad \text{(Local Truncation Error).} \tag{A.4}$$

  Hence, the L.T.E. is just a source term inside the differential Equation (A.4) governing the scheme's evolutionary error; this was noted in [26, 29, 40].

In case both (A.1) and its (consistent) discrete approximation $\mathcal{L}_J$, are dissipative ("contractive" [39, 30], "strongly stable" in a terminology of [26]) in some norm, this source term is responsible for most of the error e_J; if, on the contrary, (A.1) happens to be accretive, for instance if $\|u(t)-v(t)\| \leq K\|u_0-v_0\|$ with $K>1$ like in Bressan–Glimm's theory of strictly hyperbolic systems of conservation laws [6], then both $\mathcal{L}_J$ and the L.T.E. can contribute to the increase of the evolutionary error, see again [40].

Remark A.1. If the approximation $\mathcal{L}_J$ is linear, then (A.4) simplifies into,

$$\forall t>0, \qquad \frac{\mathrm{d}}{\mathrm{d}t} e_J(t) = \mathcal{L}_J e_J(t) + \tau_u(t),$$

where $\tau_u(t)$ stands for the L.T.E. related to (x-derivatives of) the exact solution $u(t,\cdot)$ to (A.1) at time t. Duhamel's principle yields an expression of the evolutionary error,

$$e_J(t) = \exp(t\cdot\mathcal{L}_J)\left(e_J(t=0) + \int_0^t \exp(-s\cdot\mathcal{L}_J)\tau_u(s)\mathrm{d}s\right).$$

Quantities like $\exp(t\cdot\mathcal{L}_J)$ are usually estimated by "logarithmic norms", see e.g. [30].

A.2. Local Truncation Error (LTE) and second-order accuracy. Second-order accuracy in space for 1D scalar conservation laws (or linear advection equations) was studied in [27] (see also [21, 36]). These equations are dissipative in L^1, so the former analysis yielding (A.4) indicates that the local truncation error is probably the main source of evolutionary error. For $\mathcal{L}u = -\partial_x f(u)$, it reads:

$$\forall k\in\mathbb{Z}, \quad \mathcal{P}_J\mathcal{L}u(t,x_k) = -\frac{1}{\Delta x}\int_{x_{k-\frac12}}^{x_{k+\frac12}} \partial_x f(u)\mathrm{d}x = -\frac{f(u(t,x_{k+\frac12})) - f(u(t,x_{k-\frac12}))}{\Delta x},$$

by exact integration of the conservation law (1.1). Now, since high-order accuracy is only concerned with smooth exact solutions u, one approximates this expression with a second-order mid-point rule by taking advantage of $x_{k+\frac12} = \frac{x_{k+1}+x_k}{2}$,

$$\mathcal{P}_J\mathcal{L}u(t,x_k) = \frac{f\left(\frac{u(t,x_{k+1})+u(t,x_k)}{2}\right) - f\left(\frac{u(t,x_k)+u(t,x_{k-1})}{2}\right) + O(\Delta x^2)}{\Delta x},$$

and so, the L.T.E. is the difference between this approximation and the numerical scheme $\mathcal{L}_J$ applied to the piecewise constant projection of the exact solution, $\mathcal{P}_J u$. Since $\mathcal{L}_J$ needs to be conservative and consistent with $\mathcal{L}$, we assume it is given by a (smooth) numerical flux which reads, in standard notation,

$$\tilde{F}_{k+\frac12} = F(u^L_{k+\frac12}, u^R_{k+\frac12}), \qquad \mathcal{L}_J\mathcal{P}_J u(t,x_k) = \frac{\tilde{F}_{k+\frac12}(t) - \tilde{F}_{k-\frac12}(t)}{\Delta x},$$

where $u^{L/R}_{k+\frac12}$ are obtained from the set of cell-centered values $\mathcal{P}_J u$ by means of a reconstruction like (2.2) and F is, for instance, the exact Godunov flux (1.3). Hence,

$$L.T.E. = \frac{[f\left(\frac{u(t,x_{k+1})+u(t,x_k)}{2}\right) - \tilde{F}_{k+\frac12}] - [f\left(\frac{u(t,x_k)+u(t,x_{k-1})}{2}\right) - \tilde{F}_{k-\frac12}]}{\Delta x}.$$

As the CFL condition imposes $\Delta t = O(\Delta x)$, second-order accuracy asks for,

$$\left| f\left(\frac{u(t,x_{k+1})+u(t,x_k)}{2}\right) - \tilde{F}_{k+\frac12}(t)\right| = O(\Delta x^2),$$

which, by the smoothness of the flux functions, reduces simply to,

$$\forall t,k \in \mathbb{R}^+ \times \mathbb{Z}, \qquad \left| u^{L/R}_{k+\frac{1}{2}}(t) - \frac{u(t,x_{k+1}) + u(t,x_k)}{2} \right| = O(\Delta x^2). \tag{A.5}$$

And this meets with the definition used by Osher (see Lemma 2.1, page 953 in [27]) and Sjogreen (see Theorem 3.9 in [34], page 47). A slightly different derivation of a second-order scheme for smooth solutions is given in [5] (page 53), essentially by keeping the term $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}_J u$ in (A.3) inside the expression of the L.T.E as follows:

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}_J u(t,\cdot) &= \lim_{\Delta t \to 0} \left(\tfrac{\mathcal{P}_J u(t+\Delta t,\cdot) - \mathcal{P}_J u(t,\cdot)}{\Delta t} \right) \\ &= -\tfrac{F\left(u(t,\cdot+\Delta x), u(t,\cdot)\right) - F\left(u(t,\cdot), u(t,\cdot-\Delta x)\right)}{\Delta x}, \end{aligned}$$

where F is the exact flux defined in (1.3). The L.T.E. is now defined like,

$$\forall k \in \mathbb{Z}, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}_J u(t,x_k) - \mathcal{L}_J \mathcal{P}_J u(t,x_k) = -\frac{\mathcal{F}_{k+\frac{1}{2}}(t) - \mathcal{F}_{k-\frac{1}{2}}(t)}{\Delta x},$$

where $\mathcal{F}_{k+\frac{1}{2}}(t) = F\big(u(t,x_k+\Delta x), u(t,x_k)\big) - F_{k+\frac{1}{2}}(t)$. The scheme induced by the numerical flux $F_{k+\frac{1}{2}}$ is called second-order in space as soon as, for any smooth exact solution $u(t,\cdot)$, $\mathcal{F}_{k+\frac{1}{2}}$ is a quadratic quantity (possibly depending on $|\partial_{xx} u(t,\cdot)|$),

$$\forall t \geq 0, \qquad |\mathcal{F}_{k+\frac{1}{2}}(t)| = O(\Delta x^2). \tag{A.6}$$

Clearly, both criteria pick up variants of the (unstable) "centered scheme" which is second-order, but unstable because it lets the total variation increase strongly: despite having small L.T.E., the evolutionary error quickly grows with the left-hand side of (A.3). MUSCL reconstructions, involving a slope limiter, allow us to keep both L.T.E. and other terms,in the O.D.E. (A.3) governing e_J rather small (in smooth regions).

REFERENCES

[1] M. Arora and P.L. Roe, *On postshock oscillations due to capturing schemes in unsteady flows*, J. Comput. Phys., 130, 25–40, 1997.

[2] M. Ben-Artzi and J. Falcovitz, *Generalized Riemann problems in computational fluid dynamics*, Cambridge Monographs on Appl. Comput. Math., 11, 2003.

[3] C. Berthon, C. Sarazin, and R. Turpault, *Space-time generalized Riemann problem solvers of Order k for linear advection with unrestricted time step*, J. Sci. Comput., 55, 268–308, 2013.

[4] B.L. Bihari and A. Harten, *Multiresoution schemes for the numerical solution of 2D conservation laws*, SIAM J.Sci. Comput., 18, 315–354, 1997.

[5] F. Bouchut, *Nonlinear Stability of Finite Volume Methods for Hyperbolic Conservation Laws, and Well-balanced Schemes for Sources*, Frontiers in Mathematics series, Birkhäuser, 2004. ISBN 3-7643-6665-6

[6] A. Bressan, *Hyperbolic Systems of Conservation Laws – The One-Dimensional Cauchy Problem*, Oxford Lecture Series Math. Appl., 20, 2000.

[7] Mark H. Carpenter and Jay H. Casper, *Accuracy of shock capturing in two spatial dimensions*, AIAA J., 37, 1072–1079, 1999.

[8] Albert Cohen, Sidi Mahmoud Kaber, Siegfried Müller, and Marie. Postel, *Fully adaptive multiresolution finite volume schemes for conservation laws*, Math. Comput., 72, 183–225, 2003.

[9] M.J.P. Cullen and K.W. Morton, *Analysis of evolutionary error in finite element and other methods*, J. Comput. Phys., 34, 245–267, 1980.

[10] D.L. Donoho and P.B. Stark, *Uncertainty principles and signal recovery*, SIAM J. Appl. Math., 49, 906–931, 1989.

[11] G. Efraimsson and G. Kreiss, *A remark on numerical errors downstream of slightly viscous shocks*, SIAM J. Numer. Anal., 36, 853–863, 1999.

[12] B. Engquist and B. Sjögreen, *The convergence rate of finite difference schemes in the presence of shocks*, SIAM J. Numer. Anal., 35, 2464–2485, 1998.

[13] H. Gilquin, *Une famille de schémas numériques T.V.D. pour les lois de conservation hyperboliques*, RAIRO - Model. Math. Anal. Num., 20, 429–460, 1986.

[14] J. Glimm, *The interaction of nonlinear hyperbolic waves*, Commun. Pure Appl. Math., 41, 569–590, 1988.

[15] Sonia M. Gomes, *Unified overview of wavelet-based methods for differential equations*, Proc. SPIE 3078, Wavelet Appl. IV, 730, April 3, 1997. doi:10.1117/12.271758

[16] J.B. Goodman and R.J. LeVeque, *A geometric approach to high resolution TVD schemes*, SIAM J. Numer. Anal., 25, 268–284, 1988.

[17] L. Gosse, *A Donoho–Stark criterion for stable signal recovery in discrete wavelet subspaces*, J. Comput. Appl. Math., 235, 5024–5039, 2011.

[18] L. Gosse, *Computing Qualitatively Correct Approximations of Balance Laws*, Springer, 2013. ISBN 978-88-470-2891-3

[19] A. Harten, *Discrete multi-resolution analysis and generalized wavelets*, Appl. Numer. Math., 12, 153–192, 1993.

[20] A. Harten and S. Osher, *Uniformly high-order accurate non-oscillatory schemes*, SIAM J. Numer. Anal., 24(2), 279–309, 1987. (paper available in [21])

[21] M. Hussaini, Bram Leer, and John Rosendale (eds.), *Upwind and High-Resolution Schemes*, (English) Berlin: Springer, 1997.

[22] Shi Jin and Jian-Guo Liu, *The effects of numerical viscosities. I. Slowly moving shocks*, J. Comput. Phys., 126, 373–389, 1996.

[23] F. Lagoutiere, *Non-dissipative entropy satisfying discontinuous reconstruction schemes for hyperbolic conservation laws*, preprint, www.math.u-psud.fr/˜ lagoutie/publi.html.

[24] R.J. LeVeque, *Numerical methods for conservation laws*, ETH Zurich, Birkhauser 1992.

[25] S. Mallat, *A Vavelet Tour of Signal Processing*, Academic Press, 1998.

[26] Keith W. Morton, *On the analysis of finite volume methods for evolutionary problems*, SIAM J. Numer. Anal., 35, 2195–2222, 1998.

[27] S. Osher, *Convergence of generalized MUSCL schemes*, SIAM J. Numer. Anal., 22, 947–961, 1985. (paper available in [21])

[28] B. Popov and O. Trifonov, *Order of convergence of second order schemes based on the MINMOD limiter*, Math. of Comput., 75, 1735–1753, 2006.

[29] J.M. Sanz-Serna and J.G. Verwer, *Convergence analysis of one-step schemes in the method of lines*, Appl. Math. Comput., 31, 183–196, 1989.

[30] J.M. Sanz-Serna and J.G. Verwer, *Stability and convergence at the PDE/stiff ODE interface*, Appl. Numer. Math., 5, 117–132, 1989.

[31] M. Siklosi, B. Batzorig, and G. Kreiss, *An investigation of the internal structure of shock profiles for shock capturing schemes*, J. Comput. Appl. Math., 201, 8–29, 2007.

[32] M. Siklosi and G. Efraimsson, *Analysis of first order errors in shock calculations in two space dimensions*, SIAM J. Numer. Anal., 43, 672–685, 2005.

[33] M. Siklosi and G. Kreiss, *Elimination of first order errors in time dependent shock calculations*, SIAM J. Numer. Anal., 41, 2131–2148, 2003.

[34] B. Sjogreen, *Lecture notes*, at www.math.fsu.edu/˜ sussman/Bjorn_Sjogreen_Notes.pdf

[35] P.K. Sweby, *High resolution schemes using flux limiters for hyperbolic conservation laws*, SIAM J. Numer. Anal., 21, 995–1011, 1984.

[36] E.F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*, Third Edition, Springer, 2009.

[37] B. Van Leer, *Towards the ultimate conservative difference scheme, V.A second order sequel to Godunov's method*, J. Comput. Phys., 32, 101–136, 1979.

[38] B. Van Leer, *Upwind and high-resolution methods for compressible flow: from donor cell to residual–distribution schemes*, Commun. Comput. Phys., 1, 192–206, 2006.

[39] J.G. Verwer, *Contractivity in locally one-dimensional splitting methods*, Numer. Math., 44, 247–259, 1984.

[40] J.G. Verwer and J.M. Sanz-Serna, *Convergence of method of lines approximations to partial differential equations*, Computing, 33, 297–313, 1984.