Available online at www.sciencedirect.com

SCIENCE DIRECT®

C. R. Acad. Sci. Paris, Ser. I 338 (2004) 549–554

COMPTES RENDUS MATHEMATIQUE

Partial Differential Equations/Numerical Analysis

A non-dissipative entropic scheme for convex scalar equations via discontinuous cell-reconstruction

Frédéric Lagoutière

Laboratoire Jacques-Louis-Lions, Université Pierre-et-Marie-Curie, boîte courrier 187, 75252 Paris cedex 05, France

Received 25 September 2003; accepted after revision 27 January 2004

Presented by Olivier Pironneau

Abstract

We present a finite volume non-dissipative but entropic scheme for convex scalar equations based on a discontinuous reconstruction of the solution in each cell of the mesh. This discontinuous representation of the numerical solution in each cell is done satisfying the L^∞-norm, Total Variation and entropy decreasing properties. This allows us to prove the convergence towards the unique entropy solution. Numerical computations are reported, showing the non-dissipative behavior of the algorithm. ***To cite this article: F. Lagoutière, C. R. Acad. Sci. Paris, Ser. I 338 (2004).***

Résumé

Un schéma entropique non dissipatif par reconstruction discontinue pour des équations aux dérivées partielles scalaires. Nous étudions un schéma de type volumes finis non dissipatif mais entropique pour les équations scalaires à flux convexes. Ce schéma est basé sur une reconstruction de la solution numérique sous forme discontinue à l'intérieur de chaque maille, cette reconstruction étant effectuée en veillant à la décroissance de la norme L^∞, de la variation totale et de l'entropie de la solution discrète. Ceci permet de montrer la convergence de la solution numérique vers l'unique solution entropique. Nous présentons quelques résultats numériques afin de mettre en évidence le caractère non dissipatif de l'algorithme. ***Pour citer cet article : F. Lagoutière, C. R. Acad. Sci. Paris, Ser. I 338 (2004).***

Version française abrégée

Nous proposons ici un nouvel algorithme pour la résolution approchée d'équations scalaires à flux convexes. Cet algorithme, dans l'esprit du schéma MUSCL de Van Leer (cf. [14,15]) et des schémas d'ordre élevé de Goodman et LeVeque (dans [9]), est basé sur une reconstruction de la solution discrète dans chaque maille.

Cependant, à la différence de la plupart des schémas de reconstruction, notre but n'étant pas d'obtenir un schéma d'ordre élevé mais plutôt d'approcher des solutions discontinues, nous reconstruisons cette solution sous la forme d'une fonction constante par morceaux *dans chaque maille*, en plaçant dans chaque maille *une* discontinuité. Le

E-mail address: lagoutie@ann.jussieu.fr (F. Lagoutière).

1631-073X/$ – see front matter © 2004 Académie des sciences. Published by Elsevier SAS. All rights reserved.
doi:10.1016/j.crma.2004.01.024

schéma résulte donc à chaque pas de temps de la reconstruction de la solution sous la forme évoquée ci-dessus (à partir d'une fonction constante par maille), de la résolution exacte de l'équation considérée avec comme donnée initiale la reconstruction, puis de sa projection sur le maillage.

Nous mettons en évidence des conditions sur cette reconstruction sous lesquelles la solution discrète est à variation totale décroissante. La reconstruction que nous proposons alors est celle qui maximise la variation totale de la solution reconstruite à l'intérieur de chaque maille. Nous montrons que le schéma qui en résulte est, pour l'équation d'advection linéaire, équivalent au *schéma décentré aval sous contraintes amont* (voir [6] et [7]), ou encore au limiteur Ultra-Bee (cf. [13]). Ce schéma n'est pas entropique, il génère dans le cas non linéaire des « chocs de détente ». Une modification entropique de celui-ci a été proposée dans [10] mais elle est source de diffusion numérique.

Le formalisme ici développé présente l'avantage d'être plus exploitable : il permet en particulier d'exhiber de manière très simple de nouvelles contraintes sur la reconstruction discontinue qui assurent la décroissance (au sens large) de l'entropie. L'hypothèse de convexité du flux de l'équation aux dérivées partielles permet la démonstration de la convergence de la solution numérique vers l'unique solution entropique de l'équation. Les résultats numériques obtenus avec ce schéma, dont certains sont donnés à la fin de la Note, montrent que ce schéma est « non dissipatif ». Une étude détaillée et une généralisation des techniques introduites ici seront proposées dans [11].

1. Introduction

We are here concerned with the finite volume approximation of scalar conservation laws of the form

$$\partial_t u + \partial_x f(u) = 0, \quad t \in \mathbb{R}^+,\ x \in \mathbb{R}, \tag{1}$$

with initial condition $u(0,x) = u^0(x) \in BV(\mathbb{R})$ and where $f \in \mathcal{C}^2(\mathbb{R})$ is convex. We furthermore are especially interested in *entropy* solutions of (1), i.e. weak solutions satisfying $\partial_t S(u) + \partial_x G(u) \leqslant 0$ for every entropy–entropy flux pair (S,G), i.e. every pair of $C^1(\mathbb{R})$ functions (S,G) such that S is convex and $G' = S' f'$. It is known that there exists a unique entropy weak solution to (1).

Given a cell size $\Delta x > 0$ and a time increment $\Delta t > 0$, the finite volume framework leads to solve $u_j^0 = \frac{1}{\Delta x}\int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} u^0(x)\,\mathrm{d}x\ \forall j \in \mathbb{Z}$ and

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x}\big(f_{j+1/2}^n - f_{j-1/2}^n\big) \quad \forall j \in \mathbb{Z},\ \forall n \in \mathbb{N}, \tag{2}$$

where the numerical solution $u(t,x) = \sum_{n\in\mathbb{N}}\sum_{j\in\mathbb{Z}} u_j^n \mathbb{1}_{[n\Delta t,(n+1)\Delta t[}(t)\mathbb{1}_{[(j-1/2)\Delta x,(j+1/2)\Delta x[}(x)$ is ought to converge towards the exact entropy solution to (1) as Δx and Δt converge to 0. Choosing a formula to compute $(f_{j+1/2}^n)_{j\in\mathbb{Z}}$ with the help of $(u_j^n)_{j\in\mathbb{Z}}$ is choosing an explicit numerical scheme. As for [1] or [6], the goal is here to find non-dissipative fluxes.

2. A reconstruction scheme

The principle of reconstruction schemes is to consider the constant-in-cell approximated solution as the projection on the grid of a more detailed function for which it is possible to solve, exactly or approximately, Eq. (1). Traditionally, this is a way to obtain second order schemes ([2,5,9,15]) using a piecewise affine reconstruction, or high-order schemes (cf. [4,12]) using piecewise polynomials (of desired order) reconstructions. We here propose to use a discontinuous-in-cell reconstruction. This in general will not lead to a high order scheme, but, as will be seen, to a scheme able to compute sharp discontinuities, both in linear and non-linear cases. For simplicity reasons, we restrict ourselves to a reconstruction with only one discontinuity per cell and decide that the reconstructed function

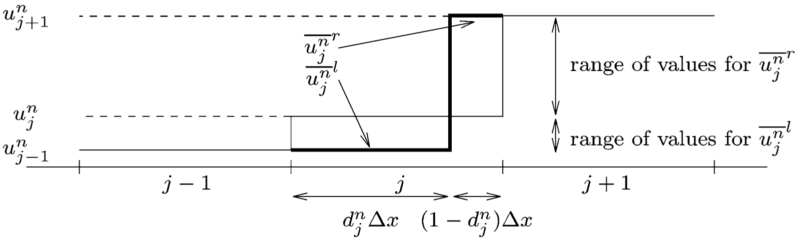

Fig. 1. A discontinuous reconstruction satisfying conservativity and TVD requirements.

will be constant on each side of the discontinuity. Each time step of the algorithm consists in a reconstruction procedure (starting from a constant-in-cell data) followed by an *exact resolution* of Eq. (1) with the so-reconstructed initial condition, and finally a projection of this solution on the grid. Let us assume that the second step, exact resolution of the equation with piecewise constant initial condition, is handleable, and let us focus on the reconstruction step.

Given the projected value u^n_j, we first want to reconstruct it as a function taking the value $\overline{u^n_j}^l$ on an interval of length $d^n_j\,\Delta x \in [0, \Delta x]$ on the left side of the cell, and the value $\overline{u^n_j}^r$ on an interval of length $(1-d^n_j)\Delta x \in [0, \Delta x]$ on the right side of the cell (see Fig. 1). We denote by $\overline{u^n}: \mathbb{R} \to \mathbb{R}$ the reconstructed solution:

$$\overline{u^n}(x) = \begin{cases} \overline{u^n_j}^l \text{ if and only if } x \in \left[(j-1/2)\Delta x, (j-1/2+d^n_j)\Delta x\right[, \\ \overline{u^n_j}^r \text{ if and only if } x \in \left[(j-1/2+d^n_j)\Delta x, (j+1/2)\Delta x\right[, \end{cases} \quad j \in \mathbb{Z}. \tag{3}$$

Let $\bar{u}^n(t,x)$ be the exact solution of (1) with initial condition (3). We now give the formal definition of the numerical fluxes:

$$f^n_{j+1/2} = \frac{1}{\Delta t}\int_0^{\Delta t} f\left(\bar{u}^n\left(s, (j+1/2)\Delta x\right)\right) \mathrm{d}s, \quad j \in \mathbb{Z}, \tag{4}$$

where d^n_j, $\overline{u^n_j}^l$ and $\overline{u^n_j}^r$ are to be determined.

The most usual convergence theorems are valid under three main hypothesis: the conservativity of the scheme, the L^∞-stability: $\exists C \in \mathbb{R}$ independent on n and Δt such that

$$\max_{j\in\mathbb{Z}} \left|u^n_j\right| \leqslant C, \tag{5}$$

and the Total Variation Diminishing (TVD) property:

$$\sum_{j\in\mathbb{Z}} \left|u^{n+1}_{j+1} - u^{n+1}_j\right| \leqslant \sum_{j\in\mathbb{Z}} \left|u^n_{j+1} - u^n_j\right| \quad \forall n \in \mathbb{N} \tag{6}$$

at every time step n. We only consider reconstructions that satisfy these properties.

The conservativity assumption naturally leads to

$$d^n_j \overline{u^n_j}^l + \left(1-d^n_j\right)\overline{u^n_j}^r = u^n_j, \tag{7}$$

for the reason that the second step (exact resolution) is conservative.

In order to satisfy the L^∞-stability and the TVD property, let us state the following preliminary result, *which does not involve the convexity of the flux f.*

Lemma 2.1. *Assume there is no sonic point in the computational domain*: $f'(u) > 0\ \forall u \in [\min_j(u_j^n), \max_j(u_j^n)]$. *Assume the reconstruction operation is conservative* (*Eq.* (7)) *and such that*

$$\overline{u_j^n}^{\,l} \in \left[u_{j-1}^n, u_j^n\right] \quad \textit{and} \quad \overline{u_j^n}^{\,r} \in \left[u_j^n, u_{j+1}^n\right] \quad \forall j \in \mathbb{Z} \quad (\textit{see Fig. } 1). \tag{8}$$

Then, the scheme (2)–(4) *is* L^∞*-stable* (*Eq.* (5)) *and TVD* (*Eq.* (6)).

Recall that the second step is usually done under the CFL condition $\sup_{j\in\mathbb{Z}} f'(u_j^n)\Delta t \leqslant \Delta x$.

This result is not obvious because under the only hypothesis (8), the total variation of the reconstructed solution can be larger than the one before reconstruction (up to three times greater). An interesting point is that this result holds for any flux function $f \in \mathcal{C}^2(\mathbb{R})$. The proof is given in [11].

The Riemann problems that are to be solved after the reconstruction are separated by lengths $d_j^n\Delta x$, $(1-d_j^n)\Delta x$ and not Δx: the schemes here discussed are workable assuming the interaction between three Riemann problems is exactly computable (which is the case for advection; for Burgers' equations, we will use a slight simplification: see Remark 1). Otherwise, it should be of great interest to use an approximate solver such as the 'transport-collapse' method, see [3,2].

Note that Eq. (8) together with conservativity implies that

$$\begin{array}{ccc} \left(u_j^n - u_{j-1}^n\right)\left(u_{j+1}^n - u_j^n\right) \leqslant 0 & \Longrightarrow & \overline{u_j^n}^{\,l} = \overline{u_j^n}^{\,r} = u_j^n \\ \text{(presence of a local extremum)} & & \text{(there is namely no reconstruction).} \end{array}$$

Following a well-known theorem (see [8] for example), a consequence of Lemma 2.1 is that the numerical solution converges (in $L^\infty([0,T], L^1_{\mathrm{loc}}(\mathbb{R}))$) towards a weak solution of (1) provided the reconstruction is locally Lipschitz continuous.

We now will choose a way to reconstruct the solution in each cell, satisfying hypothesis of Lemma 2.1. To loose the minimum of detail of the numerical solution, we first propose to *maximize* the total variation *in each cell*, that is to say, in the case where u_j^n is not a local extremum, to define the reconstructed solution as

$$\overline{u_j^n}^{\,l} = u_{j-1}^n \quad \text{and} \quad \overline{u_j^n}^{\,r} = u_{j+1}^n \quad \text{(see Fig. 1).} \tag{9}$$

According to Lemma 2.1, this gives a TVD scheme. Moreover, away from local extrema, this is compatible with the conservativity assumption: it suffices to define d_j^n as

$$d_j^n = \frac{u_j^n - u_{j+1}^n}{u_{j-1}^n - u_{j+1}^n} \in [0,1] \tag{10}$$

(let us repeat that on a local extremum, $\overline{u_j^n}^{\,l} = \overline{u_j^n}^{\,r} = u_j^n$). This reconstruction is, of course, Lipschitz continuous and, according to the classical result already mentioned, the algorithm converges towards a weak solution.

Proposition 2.2. *In the case where f is linear, the discontinuous reconstruction algorithm defined by* (2)–(4), (9), (10) *is equivalent to the anti-dissipative algorithm called 'limited downwind scheme' in* [6,7], *which is equivalent to the Ultra-Bee limiter of* [13].

It has already been shown in the cited references that the limited downwind scheme is not entropic. In the non-linear case, (2)–(4), (9), (10) is neither. Its formalism fortunately allows us to deal easily with entropy inequalities. Indeed, it is possible to reconstruct the solution adding a new constraint that ensures that the entropy of the reconstructed solution will not exceed the entropy of the solution at the preceding time step *before* the L^2-projection on the grid. Assuming that the flux f is convex, a sufficient condition to recover the entropy solution is to ask only for *one* numerical entropy inequality, provided the considered entropy is strictly convex (cf. [8]). For example, for $f(u) = u^2/2$ (Burgers' equation), it suffices to take into account entropy inequalities for the entropy $S(u) = u^2/2$

with entropy flux $G(u) = u^3/3$. Let us denote s_j^n the entropy of the exact solution before projection on the grid at the preceding time step: $s_j^n = \frac{1}{\Delta x}\int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} S(\bar{u}^{n-1}(\Delta t, y))\,\mathrm{d}y$. We precisely want to ensure

$$d_j^n \frac{(\overline{u_j^n}^l)^2}{2} + \left(1 - d_j^n\right)\frac{(\overline{u_j^n}^r)^2}{2} \leqslant s_j^n. \tag{11}$$

Let us now resume: the new proposed algorithm is based on the following four computations:

- given the constant-in-cell approximate solution at time step n, compute the piecewise constant in each cell reconstructed solution such that it verifies (6), (7) and (11);
- compute the associated exact solution at time $n+1$ (eventually satisfying the CFL condition);
- compute the entropy in each cell (entropy associated to the exact solution, to be used in the further reconstruction);
- project the exact solution on the mesh.

Remark 1. As already pointed out, to compute the entropy s_j^n of the solution with the reconstructed initial condition may imply to deal with the interaction of 3 waves per cell edge under the classical CFL condition in the genuine nonlinear case. For the numerical results presented in Section 3, we use a simplification of this idealized algorithm: we compute the entropy of the detailed (exact) solution only in the case when the three involved waves do not interact. In the opposite case (let us assume these waves interact in cell j), we replace s_j^n by $S(u_j^n)$: this leads to not to reconstruct the solution at the following time step in cell j. This simplification is showed on numerical results not to be dissipative.

The three constraints (7), (6) and (11) are of course compatible: at least, if no reconstruction is done (we keep the projection on the grid), we recover the Godunov algorithm which is conservative, TVD and entropic (thanks to Jensen's inequality). Following the idea already used, we now maximize the total variation on each cell under the three constraints (7), (6) and (11). This maximization problem has of course an explicit solution: no iterative procedure is needed to solve it (see [11]). Let us just precise that the reconstructed solution verifying the three constraints has a d_j^n exactly defined by (10): in other words, the entropy constraint (11) does not affect the value of d_j^n. The so-reconstructed solution in the cell j is a Lipschitz continuous function of u_{j-1}^n, u_j^n and u_{j+1}^n, so that we can state

Theorem 2.3. *Assume $f'(u) \neq 0$ $\forall u \in [\min_j(u_j^n), \max_j(u_j^n)]$. Then, the discontinuous reconstruction scheme* (2), (4) *that maximizes the total variation in each cell under constraints* (7), (8) *and* (11) *is convergent towards the unique entropy solution of* (1).

Let us lastly mention that non-convex fluxes as well as sonic points will be treated in [11].

3. Numerical results

We now present a few numerical results that will show the non-dissipative behavior of the studied scheme, focusing on Burgers' equation. We present some results for different Riemann problems and a regular initial condition leading to the formation of a shock (avoiding sonic points). We in each case assume the initial condition is periodic with period 1. The first problem (Fig. 2) is a double Riemann problem where we can follow the evolution of a shock and the formation of a rarefaction wave (and compare Godunov's scheme with the present one). What is worth noticing is:

- the behavior of the new scheme on the shock: this discontinuity is perfectly resolved (in only one cell);

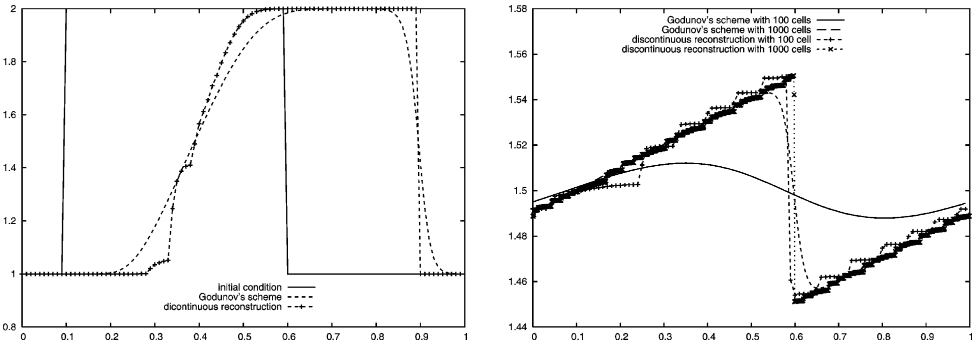

Fig. 2. Riemann problems; $t = 0.2$ on the left, $t = 10$ on the right.

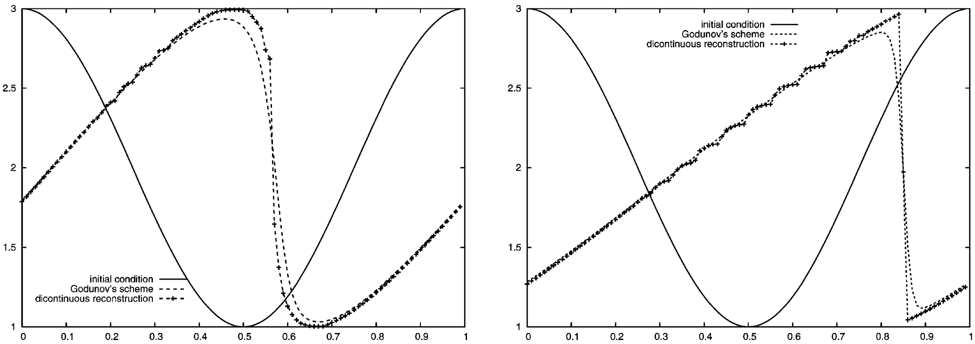

Fig. 3. Cosine initial condition; $t = 1/2\pi$ on the left, $t = 0.4$ on the right.

– the behavior on the rarefaction wave, where some staircases appear; when refining the mesh, the size of these staircases goes to 0, in accordance with the convergence result (see results in long time on the right, with 100 and 1000 cells).

Fig. 3 shows the evolution of the same two schemes when initial condition is a cosine. Theory predicts a shock at time $t = 1/2\pi$. This is well computed with the new scheme, and the Godunov's scheme cannot detect it (dissipative behavior).

References

[1] F. Bouchut, An anti-diffusive entropy scheme for monotone scalar conservation laws, J. Sci. Comput., in press.
[2] F. Bouchut, Ch. Bourdarias, B. Perthame, Math. Comp. 65 (216) (1996) 1439–1461.
[3] Y. Brenier, SIAM J. Numer. Anal. 21 (6) (1984) 1013–1037.
[4] P. Collela, P.R. Woodward, J. Comp. Phys. 54 (1984) 174–201.
[5] F. Coquel, P.G. LeFloch, Numer. Math. 74 (1996) 1–33.
[6] B. Després, F. Lagoutière, C. R. Acad. Sci. Paris, Ser. I 328 (10) (1999) 939–944.
[7] B. Després, F. Lagoutière, J. Sci. Comput. 16 (4) (2001) 479–524.
[8] E. Godlewski, P.-A. Raviart, Hyperbolic Systems of Conservation Laws, Ellipses, 1991.
[9] J.J. Goodman, R.J. LeVeque, SIAM J. Numer. Anal. 25 (2) (1988) 268–284.
[10] F. Lagoutière, CEMRACS 1999, Orsay, in: ESAIM Proc., vol. 10, SMAI, Paris, 1999, pp. 183–199 (electronic).
[11] F. Lagoutière, in preparation.
[12] P.G. LeFloch, J.M. Mercier, C. Rohde, SIAM J. Numer. Anal. 40 (5) (2002) 1968–1992 (electronic).
[13] E.F. Toro, Riemann Solvers and Numerical Methods for Fluid Dynamics, Springer-Verlag, 1997.
[14] B. Van Leer, J. Comput. Phys. 23 (1977) 263–275.
[15] B. Van Leer, J. Comput. Phys. 32 (1979) 101–136.