On the reservoir technique convergence - Part I

Stéphane Labbé *, Emmanuel Lorin †

December 8, 2006

Abstract

This paper is devoted to the analysis of flux schemes coupled with the reservoir technique for hyperbolic equations and linear hyperbolic systems of conservation laws [2], [1]. We prove in some specifical configurations the long time convergence of the reservoir technique and its TVD property. We use in particular, the formalism introduced by Labbé in [12]. The proofs are based on a precise study of the reservoir technique treatment of shock and rarefaction waves.

1 Introduction

This paper is devoted to the convergence of the reservoir technique introduced in [2]. This method allows to avoid or to reduce drastically the numerical diffusion for hyperbolic systems of conservation laws. This technique is based on the fact that for a scalar linear advection equation a flux scheme [8] is stable and non-diffusive at CFL (Courant-Friedrich-Lewy ondition) equal to 1. Let us indeed consider the advection problem, with $a > 0$: $\partial_t u + a\partial_x u = 0,\ x \in \mathbb{R},\ t \geqslant 0$, and a conservative explicit upwind scheme $\forall n \geqslant 0, \forall j \in \mathbb{Z},\ u_j^{n+1} = u_j^n - a\Delta t/\Delta x(u_j^n - u_{j-1}^n)$ approaching this equation. It is well known that this scheme is stable under the CFL criterion: $\lambda = a\frac{\Delta t}{\Delta x} \in]0, 1]$, and that taking $\Delta t = \Delta x/a$ ($\lambda = 1$) leads to $u_j^{n+1} = u_{j-1}^n$ which corresponds the exact propagation of the solution at the discrete level. However taking $\Delta t = \Delta x/a$ is very restrictive (if Δx or a is not constant in space for example) and we must constraint our scheme to behave as if $\lambda = 1$ although the time step is smaller than $\Delta t^* = \Delta x/a$ in general. This is precisely the goal of the reservoir scheme. To understand the principle we consider the case where $\Delta t = \Delta t^*/k$. The idea consists in waiting k time steps before updating the solution. More generally, let us take several different time steps Δt_i (not necessarily equal), we introduce R_j the reservoir associated to the j^{th} cell (initialized to 0) and c_j a positive real CFL counter (also initialized to 0). At each time step, we fill up R_j with the local current numerical flux difference

*Département de Mathématiques, Université d'Orsay, 91405 Orsay (France)
†Centre de Recherche en Mathématiques, Montréal (Canada)

$R_j \leftarrow R_j - a\Delta t_i/\Delta x(u_j - u_{j-1})$ and increment the CFL counter $c_j \leftarrow c_j + a\Delta t_i/\Delta x$. Once c_j reaches 1, we update the solution and reinitialize both the reservoir and the counter.

$$u_j \leftarrow u_j + R_j,\ c_j \leftarrow 0,\ R_j \leftarrow 0.$$

In the advection equation case we recover the CFL= 1. Indeed the reservoir technique for one-dimensional linear equation behaves as :

$$u_j^{n+1} = u_j^n - a\sum_{i=1}^{m} \frac{\Delta t_i}{\Delta x}(u_j^n - u_{j-1}^n) = u_{j-1}^n.$$

In this formula n denotes the real time step and i denotes the local iteration between two time steps such that: $a\sum_{i=1}^{i=m} \Delta t_i/\Delta x = 1$.
Extending this idea to hyperbolic systems of conservation laws we only update the numerical solution for each discretization point when the local CFL counter reaches 1. More precisely, at each space point and for each characteristic field we attach a counter that measures the local CFL number and a vectorial reservoir that stores the numerical flux. This technique has been succesfully applied to nonlinear hyperbolic systems of conservation laws [2], [3] and is currently developped for multidimensional equations, see [4]. The reservoir technique is not linked to a particular scheme but is applicable to many kind of flux schemes [8] with approximate Riemann solvers as Roe [14], VFFC [9], Colella-Glaz [6] or with exact Riemann solvers as Godunov [11]. Nevertheless, it is coupled with the Colella-Glaz scheme [6] that the reservoir technique gives the most impressive results [3]. Indeed, the results obtained with the Colella-Glaz scheme are excellent compared with higher order classical schemes as ENO [15], WAF [5] for example, as we can see in [13]. In this paper we prove the convergence of the reservoir technique coupled with flux schemes [8] approximating one-dimensional scalar hyperbolic of conservation laws, for some specifical initial data. In fact, we also extend some parts of this proof to multidimensional flux schemes coupled with the reservoir techniques. The whole difficulty for multidimensional schemes is reported to the convergence of a perturbation operator will not be discussed in details.
In order to prove the convergence of the reservoir technique we will introduce the formalism introduced by Labbé in [12]. It gives a rigorous and a clear framework to define the consistency, the stability and the convergence of numerical methods. Indeed this formalism based on categories, consider very large classes of numerical methods as finite volume, finite differences, finite element or spectral methods. The idea is to use precisely sufficiently general objects and spaces in order to bring the convergence of any numerical method to the proof of relatively simple properties. The main difficulty is then to introduce the correct discrete operators and discrete spaces. Then we study precisely the reservoir technique behavior for shock and rarefaction waves solutions. We are then able to prove its long time convergence for shock waves and its order 2 convergence for rarefaction waves. The reservoir technique main specificity comes from the fact that as for Glimm's scheme, the local error in space, is not cumulated at all times, but can be totally (shock waves) or partially (rarefaction waves) cancelled after some iterations. With the same principle, the TVD property is proved. We then prove for linear hyperbolic systems of conservation laws, the long time convergence of

the reservoir technique. Under arithmetical conditions on the characteristic velocities, we can even prove that the reservoir technique captures again the exact solution at the discrete level. The paper is organized as follows. In section 2. we give some generalities on the convergence of finite volume schemes and give an abstract framework in order to simplify the proofs of convergence. Then, in section 3. we prove the convergence of the reservoir technique coupled with flux schemes for nonlinear hyperbolic equations of conservation laws for particular initial data and its TVD property. The case of linear systems is then studied in section 4. Finally in the last concluding section we briefly present the reservoir technique for hyperbolic systems of conservation laws. The proof of convergence for nonlinear systems will be presented in a forthcoming paper.

2 On the convergence of a finite volume scheme coupled with the reservoir technique

In this part we consider finite volume schemes approximating multi-dimensional hyperbolic systems of conservation laws. We introduce some general concepts allowing to simplify the study of convergence of discrete operators. To this end we propose to follow the main principles of [12].

2.1 Notations and assumptions

In this section we will present all the elements and framework necessary the prove the convergence of the reservoir scheme. The building of these elements is fully explained in [12].
Let us consider, the semi-group $(S(t))_{t\in(R)^+}$ of domain W. In the following we focus on the case $W = SBV(\Omega)$ where Ω is an open subset of $\mathbb{R}$. Let us consider $W_{\underline{h}}$ a family of finite dimensional vector spaces indexed by $\underline{h}$, where $\underline{h}$ denotes a decreasing and vanishing sequel of strictly positive real numbers. Objects $A_{\underline{h}}$ will designate sequels indexed by the elements of the sequel $\underline{h}$. We suppose that, for all n, p in $\mathbb{N}$ such that $n < p$ one has $W_{h_n} \subset W_{h_p}$; furthermore, we set for all n in $\mathbb{N}$, the applications P_{h_n} and $P^\star_{h_n}$ such that

(i) $\forall u \in W_{h_n},\ P_{h_n}(u) \in L(W, W_{h_n}),\ P^\star_{h_n}(u) \in L(W_{h_n}, W)$, where $L(A, B)$ designates the linear continuous applications from A onto B.

(ii) $P_{\underline{h}}$ and $P^\star_{\underline{h}}$ are uniformly bounded families of operators.

(iii) $P_{h_n} \circ P^\star_{h_n} = Id_{W_{h_n}}$.

(iv) $\forall u \in W,\ \lim_{n\to\infty} \|u - P^\star_{h_n} \circ P_{h_n}(u)\|_W = 0$.

We will denote by $\mathcal{M}(W, W_{\underline{h}})$ a cell of discretization built on sequels $P_{\underline{h}}$ and $P^\star_{\underline{h}}$. Then, for every T in $\mathbb{R}^+$, we set $k = (k_i)_{i\in\{0,\dots,N\}}$ a sequel of strictly positive real numbers which

summation is equal to T. Hence, we set, for every i in $\{0, ..., N\}$ and every n in $\mathbb{N}$: $S_{i,h_n} = P_{h_n} \circ S(k_i) \circ P^{\star}_{h_n}$. A discretization of $S(T)$ will be $S_{h_n}(T)$ defined by:

$$S_{h_n}(T) = \bigcirc_{i\in\{0,...,N\}} S_{i,h_n}.$$

Eventually, we will introduce approximations of this discretization in order take into account that an exact computation of $S(k_i)$ is not always possible.

With the assumptions done above on the behavior of the cell of discretization $\mathcal{M}(W, W_{\underline{h}})$, the convergence of the operator $S_{h_n}(T)$ to $S(T)$ when n goes to infinity is guaranteed by the continuity of the elements of the family $(S(t))_{t>0}$.

In the remainder of the section, we will consider $W = SBV(\Omega)$. For every n in $\mathbb{N}$, we set $\Omega_{h_n} = (\omega_i)_{i\in I_n}$, family of connected open subsets of Ω such that:

$$\forall (i,j) \in I_n, i \neq j,\ \omega_i \cap \omega_j = \emptyset,$$

and

$$\bar{\Omega} = \bigcup_{i\in I_n} \bar{\omega_i}.$$

Then we choose particular families $P_{\underline{h}}$ and $P^{\star}_{\underline{h}}$ defined for every n in $\mathbb{N}$, by:

$$\forall u \in W,\ P_{h_n}(u) = \sum_{i\in I_n} \frac{1}{|\omega_i|} \int_{\omega_i} u(x)dx,$$

and $P^{\star}_{h_n}$ in the canonical injection of W_{h_n} into $L^2(\Omega)$.

For the discretization class $\mathcal{C}$ there is consistency, with the equation if and only if for all $u \in L^2(\Omega)$:

$$\lim_{n\to+\infty} \|P^{*}_{N,h_n} \circ \mathcal{P}_{h_n} \circ P_{0,h_n}(u) - \mathcal{P}(u)\|_{0,\Omega} = 0,$$

with

$$\mathcal{P}_{h_n} = \bigcirc_{l=1}^{l=N} S_{k_l,h_n} = P_{N,h_n} \circ (\bigcirc_{l=1}^{l=N} (S_{k_l} \circ \mathbf{id}_{k_{l-1},h_n})) \circ P^{*}_{0,h_n},\ \mathcal{P} = S_T = \bigcirc_{l=1}^{l=N} S_{k_l}.$$

with

$$P_{l,h_n}(u)(\mathbf{x},t) = \sum_l \frac{1}{k_l} \int_{t^l}^{t^{l+1}} \left(\sum_{i\in\mathcal{I}} \left(\frac{1}{|h_n|} \int_{\omega_i} u(\mathbf{x}',t')d\mathbf{x}' \right) \xi_{\omega_i}(\mathbf{x}) \right) dt' \zeta_{[t^l,t^{l+1}]}(t) = (u^i_j)_{j\in\mathcal{I},i\in[1,N]}.$$

with

$$\xi_i(\mathbf{x}) = \begin{cases} 1, \text{ if } \mathbf{x} \in \omega_i, \\ 0, \text{ else.} \end{cases}$$

and

$$\zeta_l(t) = \begin{cases} 1, \text{ if } t \in [t^l, t^{l+1}], \\ 0, \text{ else.} \end{cases}$$

In space we then define at time $t = t^i$ by :

$$P^*_{l,h_n}(v) = v, \ \forall v \in W_{l,h_n} := \{w \in W^x_{l,h_n}, w_{\zeta_{[t^l,t^{l+1}]}}(t) \in P^0(\mathbb{R})\}.$$

With the previous definitions :

$$\lim_{h\to 0} \|\mathrm{id}_h u_0 - u_0\|_{L^2(\Omega)} = 0.$$

and for all $i \in [1, N]$ there exists $C_{k_i} > 0$ such that :

$$\|S_{k_i} u\|_{L^2(\Omega)} \leqslant (1 + C_{k_i} k_i)\|u\|_{L^2(\Omega)}.$$

$$\forall (u, v) \in L^2(\Omega), \ \exists C'_{k_i} / \|S_{k_i} u - S_{k_i} v\|_{L^2(\Omega)}.$$

2.2 Generalities on the multi-dimensional finite volume schemes convergence

We denote by S_T, $\bigcirc_{i=1}^N P_h \circ S_{k_i} \circ P_h^*$ with $(S_{k_i})_{i\in\{1,\dots,N\}}$ the exact Riemann solvers. Similarly, we denote by $\tilde{S}_{T,h} = \bigcirc_{i=1}^N P_h \circ \tilde{S}_{k_i} \circ P_h^*$ with $(\tilde{S}_{k_i})_{i\in\{1,\dots,N\}}$ the approximate Riemann solvers. The first point to solve is the convergence to zero of:

$$\lim_{h\to 0} \|P_h^* \circ \left(\bigcirc_{i=1}^{i=N} S_{k_i} \circ \mathrm{id}_h\right) \circ P_h u_0 - \bigcirc_{i=1}^{i=N} S_{k_i} u_0\|_{L^2(\Omega)} = 0, \ \forall u_0. \tag{1}$$

We introduce the projection error r_h, defined by

$$r_h u := \mathrm{id}_h u - u$$

with $\|r_h u\|_{L^2(\Omega)} \leqslant \varepsilon(h)\|u\|_{L^2(\Omega)}$, for all $u \in W_h$with $\varepsilon(h) \to 0$ when $h \to 0$.
We then have :

$$\begin{aligned} \mathrm{id}_h \circ S_{k_1} \circ \mathrm{id}_h u_0 &= \mathrm{id}_h \circ S_{k_1}(u_0 + \mathrm{id}_h u_0 - u_0) \\ &= \mathrm{id}_h \circ S_{k_1}(u_0 + r_h u_0). \end{aligned}$$

Then

$$\begin{aligned} \mathrm{id}_h \circ S_{k_1} \circ \mathrm{id}_h u_0 - S_{k_1} u_0 &= S_{k_1}(u_0 + r_h u_0) + r_h \circ S_{k_1}(u_0 + r_h u_0) - S_{k_1} u_0 \\ &= S_{k_1}(u_0 + r_h u_0) - S_{k_1} u_0 + r_h \circ S_{k_1}(u_0 + r_h u_0). \end{aligned}$$

But as there exists a non-negative real C such that

$$\|r_h \circ S_{k_1}(u_0 + r_h u_0)\|_{L^2(\Omega)} \leqslant \varepsilon(h)(1 + C_{k_1})(1 + \varepsilon(h))\|u_0\|_{L^2(\Omega)},$$

there exists a non-negative real C' such that

$$\|S_{k_1} \circ \mathrm{id}_h u_0 - S_{k_1} u_0\|_{L^2(\Omega)} \leqslant C' \|\mathrm{id}_h u_0 - u_0\|_{L^2(\Omega)} \leqslant \varepsilon(h) C' \|u_0\|_{L^2}(\Omega).$$

We introduce δ_N defined by

$$\delta_N u_0 := \mathrm{id}_h \circ S_{k_N} \circ \left(\bigcirc_{i=1}^{N} (\mathrm{id}_h \circ S_{k_i})\right) \circ \mathrm{id}_h u_0 - S_{k_N} \circ \bigcirc_{i=1}^{N-1} S_{k_i} u_0$$

Then

$$\begin{aligned}
\delta_N u_0 &= S_{K_N} \circ \bigcirc_{i=1}^{N-1} \delta_{k_i} u_0 + r_h \circ S_{k_N} \circ \left(\bigcirc_{i=1}^{N} \mathrm{id}_h \circ S_{k_i} u_0\right) \circ \mathrm{id}_h u_0 - S_{K_N} \bigcirc_{i=1}^{N-1} u_0 \\
&= S_{K_N} \circ [\bigcirc_{i=1}^{N-1} u_0 + \delta_{N-1} u_0] + r_h \circ S_{k_N} \circ \left(\bigcirc_{i=1}^{N} \mathrm{id}_h \circ S_{k_i}\right) u_0 + r_h \circ S_{K_N} \bigcirc_{i=1}^{N-1} S_{k_i} \circ \mathrm{id}_h u_0.
\end{aligned}$$

That then gives :

$$\begin{aligned}
\|\delta_N u_0\|_{L^2(\Omega)} &\leqslant C' \|\delta_{N-1} u_0\|_{L^2(\Omega)} + \varepsilon(h)(1 + Ck_N) \|(\bigcirc_{i=1}^{N-1} \mathrm{id}_h \circ S_{k_i}) \circ \mathrm{id}_h u_0\| \\
&\leqslant C' \left(C' \delta_{N-2} u_0\|_{L^2(\Omega)} + \varepsilon(h) \Pi_{i=1}^{N-1} (1 + Ck_i)(1 + \varepsilon(h))^{N-1} \|u_0\|_{L^2(\Omega)}\right) \\
&\quad + \varepsilon(h) \Pi_{i=1}^{N} (1 + Ck_i)(1 + \varepsilon(h))^{N} \|u_0\|_{L^2(\Omega)}. \\
&\leqslant (C')^{N-1} |\delta_1 u_0\|_{L^2(\Omega)} + \left(\sum_{j=0}^{N-2} \varepsilon(h) \Pi_{i=1}^{N-j} (1 + Ck_i)(1 + \varepsilon(h))^{N-j}\right) \|u_0\|_{L^2(\Omega)}.
\end{aligned}$$

By induction we finally find :

$$\|\delta_N u_0\|_{L^2(\Omega)} \leqslant \varepsilon(h) \left(\sum_{j=0}^{N} (C')^j \Pi_{k=1}^{N-j} (1 + Ck_i)(1 + \varepsilon(h))^{N-j}\right) \|u_0\|_{L^2(\Omega)}.$$

We now set $k_* := \inf_i k_i$ and $k^* = \sup_i k_i$ so that

$$\|\delta_N u_0\|_{L^2(\Omega)} \leqslant \varepsilon(h) \left(\sum_{j=0}^{N} (C')^j \Pi_{k=1}^{N-j} (1 + Ck^*)(1 + \varepsilon(h))^{N-j}\right) \|u_0\|_{L^2(\Omega)}.$$

And as $Nk_* \leqslant T \leqslant Nk^*$ we have :

$$\|\delta_N u_0\|_{L^2(\Omega)} \leqslant \varepsilon(h) [\left(N(\sup(1, C')^N (1 + Ck^*)^N (1 + \varepsilon(h))^N + (C')^N\right] \|u_0\|_{L^2(\Omega)}.$$

That is

$$\|\delta_N u_0\|_{L^2(\Omega)} \leqslant \varepsilon(h) e^{T \log C''/k_*} \left[1 + \frac{T}{k_*} e^{\frac{T}{k_*} \log((1+Ck^*)(1+\varepsilon(h)))}\right] \|u_0\|_{L^2(\Omega)}.$$

That induces (1). So that we can enounce :

Theorem 2.1 *Let S be the exact semi-group of hyperbolic system of conservation laws, P_h and P_h^* on W_h defined above. Denoting by $(k_n)_{n\in\{1,...,N\}}$ with $T = \sum_{i=1}^{N} k_i$ and if for all $i \in \{1, ..., N\}$ and all $u \in D(S_T)$, if $\|S_{k_i}u\| \leqslant (1 + Ck_i)\|u\|$ then*

$$\lim_{h\to 0} \|P_h^* \circ \big(\bigcirc_{i=1}^{i=N} S_{k_i} \circ id_h\big) \circ P_h - \bigcirc_{i=1}^{i=N} S_{k_i}\|_{L^2(\Omega)} = 0.$$

Recall that in the previous theorem S denotes the exact operator. We now have to prove the second step towards convergence that is :

$$\|S_{k_i}u_0 - \tilde{S}_{k_i}u_0\|_{L^2(\Omega)} \leqslant C_{k_i}h\|u_0\|_{L^2(\Omega)}.$$

With

$$\tilde{S}_{k_i} := S_{k_i} + \sigma_{k_i}$$

where σ_{k_i} denotes a perturbation of the exact semi-group Riemann solver S_{k_i}. That is $\tilde{S}_{k_i}$ is a continuous operator based on the approximate Riemann solver used to solve the hyperbolic system of consevation laws. And we have

$$P_h \circ \tilde{S}_{h,T} \circ P_{h,T} = \bigcirc_{i=1}^{N}(\mathrm{id}_h \circ \tilde{S}_{k_i}) \circ \mathrm{id}_h = \bigcirc_{i=1}^{N}\mathrm{id}_h \circ (S_{k_i} + \sigma_{k_i}) \circ \mathrm{id}_h.$$

We suppose

$$\|\sigma_{k_i}u\|_{L^2(\Omega)} \leqslant C(k_i, h)\|u\|_{L^2(\Omega)}.$$

The convergence is then obtained provided that $\|\sigma_{k_i}u\|_{L^2(\Omega)} \leqslant C_{k_i}h\|u_0\|_{L^2(\Omega)}$. To obtain this kind of estimate it is necessary to study precisely the perturbation semi-group as the convergence study is reported on a non-discrete operator.
This approach is then valid in very general situations (2-D, 3-D, systems), however in the one-dimensional case we are able to describe more precisely the behavior (convergence, precision, *etc*) of the reservoir technique, due to the good knwoledge of one-dimensional hyperbolic systems of conservation laws solutions. In the following, we will then propose a more *ad hoc* approach to complete the proof.

2.3 Flux schemes and reservoir technique

We consider first order finite volume schemes based on Riemann solvers [8]:

Definition 2.1 *An explicit conservative three-points finite volume scheme with Riemann solver writes :*

$$u_j^{n+1} = u_j^n - \frac{\Delta t^n}{\Delta x_j}\big(g(u_j^n, u_{j+1}^n) - g(u_{j-1}^n, u_j^n)\big),\ j \in \mathbb{Z},\ n \geqslant 0, \tag{2}$$

where the interfacial numerical flux is denoted by $g(u, v)$ and is computed with an exact or approximate Riemann solver.

Among these finite volume schemes we will consider in this paper, flux schemes:

Definition 2.2 *A flux scheme is a finite volume scheme that writes:*

$$u_j^{n+1} = u_j^n - \frac{\Delta t^n}{\Delta x_j}(g(u_j^n, u_{j+1}^n) - g(u_{j-1}^n, u_j^n))$$

where g is the numerical flux given by:

$$g(s,t) = \frac{1}{2}(f(s) + f(t)) - \frac{1}{2}\Lambda(s,t)(f(t) - f(s)), \tag{3}$$

and Λ is a sign matrix [8], [9].

Recall that flux schemes are first order convergent schemes under a classical CFL stability condition [8]. In the reservoir technique framework, we consider two kinds of flux schemes: those with a numerical flux computed using the Colella-Glaz Riemann solver and those using a linearized Riemann solver (VFFC or Roe for instance). The main difference between these two approaches comes from the fact that, the Colella-Glaz approach consists in solving Riemann problems at the mesh interfaces using a nonlinear solver and an original flux difference decomposition, in opposite with linear flux schemes that use linearized Riemann solvers. Note that in the sequel the space step will be supposed to be constant, $\Delta x_j = \Delta x$ for all $j \in \mathbb{Z}$. In the paper we will denote by Δt the maximum of all Δt^i, where Δt^i is the time step at iteration i:

$$\Delta t = \max_{0 \leqslant i \leqslant 0} \Delta t^i. \tag{4}$$

Main goal: consider an initial data $u^0 \in L^\infty(\mathbb{R}) \cap L^1(\mathbb{R}) \cap BV(\mathbb{R})$. As usual we will consider weak discontinuities. As the time and space steps are related by a stability condition (typically with a positive constant c, $\Delta t/\Delta x \leqslant c$), we expect to prove that there exists $C > 0$ such that

$$\|v(\cdot, t^n) - u^n\|_{L^1} \leqslant \|v(\cdot, 0) - u^0\|_{L^1} + Ct^n\Delta t, \quad \forall n \in \mathbb{N}.$$

In fact we can even prove the existence of two constants $C_1 > 0$ and C_2 independent of t^n such that:

$$\|v(\cdot, t^n) - u^n\|_{L^1} \leqslant \|v(\cdot, 0) - u^0\|_{L^1} + C_1\Delta t + C_2 t^n(\Delta t)^2, \quad \forall n \in \mathbb{N},$$

where v denotes the exact solution of the continuous problem and u^n its reservoir approximation. Note that this property is very close to Lagoutière and Després estimate for their antidiffusive numerical scheme presented in [7]. This allows us in particular to prove a long-time convergence.

Suppose that at time $t = 0$ the initial data is given by u^0. Then by definition of the reservoir technique, for all $j \in \mathbb{Z}$ there exists p_j in $\mathbb{N}$ such that $c_j^{p_j-1} < 1$ and $c_j^{p_j} = 1$ with:

$$\begin{cases} u_j^l = u_j^0, \ 0 \leqslant l < p_j, \\ \\ u_j^{p_j} = u_j^0 - \sum_{l=0}^{p_j-1} \dfrac{\Delta t^l}{\Delta x}\big(g(u_j^l, u_{j+1}^l) - g(u_{j-1}^l, u_j^l)\big), \ j \in \mathbb{Z}. \end{cases}$$

That involves that the reservoir technique combined with a flux scheme based on exact or approximate Riemann solvers consists in choosing a particular local time step equals to $\sum_{l=1}^{p_j-1} \Delta t^l$.
For a classical flux scheme the solution w after p_j iterations in the cell $j \in \mathbb{Z}$ from 0 to t^{p_j} is given by:

$$\begin{cases} w_j^1 = u_j^0 - \dfrac{\Delta t^0}{\Delta x}\big(g(u_j^0, u_{j+1}^0) - g(u_{j-1}^0, u_j^0)\big), \\ \\ w_j^l = w_j^{l-1} - \dfrac{\Delta t^{l-1}}{\Delta x}\big(g(w_j^{l-1}, w_{j+1}^{l-1}) - g(w_{j-1}^{l-1}, w_j^{l-1})\big), \ \forall l \leqslant p_j. \end{cases}$$

That is:

$$w_j^{p_j} = u_j^0 - \sum_{l=0}^{p_j-1} \frac{\Delta t^l}{\Delta x}\big(g(w_j^l, w_{j+1}^l) - g(w_{j-1}^l, w_j^l)\big), \ j \in \mathbb{Z}.$$

For a finite volume scheme approximating a system of conservation laws, let us recall that the approximate solution $(w^n)_n$ is such that there exists $\tilde{C} > 0$:

$$\|v(\cdot, t^n) - w^n\|_{L^1} \leqslant \|v(\cdot, 0) - w^0\|_{L^1} + \tilde{C} t^n \Delta t, \quad \forall n \in \mathbb{N}.$$

Error estimates for the reservoir technique combined with a flux scheme can unfortunately not be deduced easily from the error estimate for the considered flux scheme. Indeed using the same notations than above we have after p_j iterations:

$$\|w_j^{p_j} - u_j^{p_j}\|_{L^1} = \Big\| \sum_{k=0}^{p_j-1} \frac{\Delta t^k}{\Delta x}\big(g(u_j^k, u_{j+1}^k) - g(u_{j-1}^k, u_j^k)\big)\Big\|_{L^1}. \tag{5}$$

We can deduce that there exists a positive constant D such that the error $\|w_j^l - u_j^l\|_{L^1}$ is always strictly less the $D\Delta t$ (recall that Δt and Δx are linked by a stability condition). This is however not sufficient to deduce the convergence.

3 Scalar equation

3.1 Numerical scheme

We consider a nonlinear scalar conservation law

$$\begin{cases} u_t + f(u)_x = 0, \ x \in \mathbb{R}, \ t \geqslant 0, \\ u(x, 0) = u_0(x) \in BV(\mathbb{R}) \cap L^1(\mathbb{R}) \cap L^\infty(\mathbb{R}), \ x \in \mathbb{R}, \end{cases}$$

with f a regular strictly convex[1] function such that $f(0) = 0$. We discretize the solution on an uniform mesh with a space step Δx. As usual we search for an approximation u_j^n of the solution $u(t^n, j\Delta x)$. The initial data is given by: $u^0 = \int u_0(x)dx/\Delta x$. In this case, the classical upwind scheme writes:

$$u_j^{n+1} = u_j^n - \frac{\Delta t_n}{\Delta x}\left(f_{j+\frac{1}{2}}^n - f_{j-\frac{1}{2}}^n\right), j \in \mathbb{Z}, n \geqslant 1,$$

where $f_{j+\frac{1}{2}}^n$ is the interfacial upwinded flux :

$$f_{j+\frac{1}{2}}^n = \frac{f(u_j^n) + f(u_{j+1}^n)}{2} - \frac{1}{2}\mathrm{sgn}(f'(u_{j+1/2}^n))\,(f(u_{j+1}^n) - f(u_j^n)),$$

with $\Delta t^n = t^{n+1} - t^n$ and $u_{j+1/2}^n$ is an approximate value of the solution at the interface. In the framework of one-dimensional nonlinear scalar equations two kinds of waves can appear: rarefaction and shock waves. At each interface $j - 1/2$ between cells $j - 1$ and j the Lax entropy condition states that $f'(u_{j-1}^n) > f'(u_j^n)$ generates an entropy shock wave whereas $f'(u_j^n) > f'(u_{j-1}^n)$ generates a rarefaction wave. Let us also recall that from the Rankine-Hugoniot condition, a shock wave at the interface $j - 1/2$ has a speed equal to:

$$\sigma_{j-1/2}^n = \frac{f(u_j^n) - f(u_{j-1}^n)}{u_j^n - u_{j-1}^n}.$$

Using this interfacial speed, we propose a method to update reservoirs and counters, upwinding the interfacial numerical flux. Let us denote by $\lambda_-^n = f'(u_{j-1}^n)$ and by $\lambda_+^n = f'(u_j^n)$ the left and right characteristic speeds. The idea is to constrain the time step according to the largest speed in the wave. For shock waves, it is given by the Rankine-Hugoniot conditions whereas for rarefaction fans it is given by the largest speed λ_-^n or λ_+^n. Moreover, in the case of sonic points ($\lambda_-^n \leqslant 0 \leqslant \lambda_+^n$), we must split the wave in two parts, one going to the left and the other one going to the right.

-Right shock wave ($\lambda_+^n \leqslant \lambda_-^n$ and $\sigma_{j-1/2}^n > 0$)

$$\begin{pmatrix} u_j^{n+1} \\ c_j^{n+1} \\ R_j^{n+1} \end{pmatrix} = \begin{cases} \left(u_j^n, c_j^n + \dfrac{|\sigma_{j-1/2}^n|\Delta t_n}{\Delta x}, R_j^n - \dfrac{\Delta t_n}{\Delta x}\big(f(u_j^n) - f(u_{j-1}^n)\big)\right)^T, \text{ if } c_j^n + \dfrac{|\sigma_{j-1/2}^n|\Delta t_n}{\Delta x} < 1, \\ \left(u_j^n + R_j^n - \dfrac{\Delta t_n}{\Delta x}\big(f(u_j^n) - f(u_{j-1}^n)\big), 0, 0\right)^T, \text{ if } c_j^n + \dfrac{|\sigma_{j-1/2}^n|\Delta t_n}{\Delta x} = 1. \end{cases}$$

-Left shock wave ($\lambda_+^n \leqslant \lambda_-^n$ and $\sigma_{j-1/2}^n < 0$)

$$\begin{pmatrix} u_j^{n+1} \\ c_j^{n+1} \\ R_j^{n+1} \end{pmatrix} = \begin{cases} \left(u_j^n, c_j^n + \dfrac{|\sigma_{j-1/2}^n|\Delta t_n}{\Delta x}, R_j^n - \dfrac{\Delta t_n}{\Delta x}\big(f(u_{j+1}^n) - f(u_j^n)\big)\right)^T, \text{ if } c_j^n + \dfrac{|\sigma_{j-1/2}^n|\Delta t_n}{\Delta x} < 1, \\ \left(u_j^n + R_j^n - \dfrac{\Delta t_n}{\Delta x}\big(f(u_{j+1}^n) - f(u_j^n)\big), 0, 0\right)^T, \text{ if } c_j^n + \dfrac{|\sigma_{j-1/2}^n|\Delta t_n}{\Delta x} = 1. \end{cases}$$

[1]It is well known that non convex fluxes lead to much more complex Riemann problems.

-Right rarefaction wave ($\lambda_+^n \geqslant \lambda_-^n \geqslant 0$)

$$\begin{pmatrix} u_j^{n+1} \\ c_j^{n+1} \\ R_j^{n+1} \end{pmatrix} = \begin{cases} \left(u_j^n, c_j^n + \dfrac{|\lambda_+^n|\Delta t_n}{\Delta x}, R_j^n - \dfrac{\Delta t_n}{\Delta x}\big(f(u_j^n) - f(u_{j-1}^n)\big)\right)^T, \text{ if } c_j^n + \dfrac{|\lambda_+^n|\Delta t_n}{\Delta x} < 1, \\ \left(u_j^n + R_j^n - \dfrac{\Delta t_n}{\Delta x}\big(f(u_j^n) - f(u_{j-1}^n)\big), 0, 0\right)^T, \text{ if } c_j^n + \dfrac{|\lambda_+^n|\Delta t_n}{\Delta x} = 1. \end{cases}$$

-Left rarefaction wave ($0 \geqslant \lambda_+^n \geqslant \lambda_-^n$)

$$\begin{pmatrix} u_j^{n+1} \\ c_j^{n+1} \\ R_j^{n+1} \end{pmatrix} = \begin{cases} \left(u_j^n, c_j^n + \dfrac{|\lambda_-^n|\Delta t_n}{\Delta x}, R_j^n - \dfrac{\Delta t_n}{\Delta x}\big(f(u_{j+1}^n) - f(u_j^n)\big)\right)^T, \text{ if } c_j^n + \dfrac{|\lambda_-^n|\Delta t_n}{\Delta x} < 1, \\ \left(u_j^n + R_j^n - \dfrac{\Delta t_n}{\Delta x}\big(f(u_{j+1}^n) - f(u_j^n)\big), 0, 0\right)^T, \text{ if } c_j^n + \dfrac{|\lambda_-^n|\Delta t_n}{\Delta x} = 1. \end{cases}$$

For sonic points the scheme is described in [3].

3.2 Numerical analysis

In this section, we then study the convergence of the reservoir technique for nonlinear scalar equations:

$$u_t + f(u)_x = 0, \qquad f \in \mathcal{C}^2(\mathbb{R} \times \mathbb{R}^+, \mathbb{R}) \text{ strictly convex}, \;\; u_0 \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R}) \cap BV(\mathbb{R}). \quad (6)$$

We consider solutions without wave interactions. This "non-interaction" assumption is reasonable for a large class of initial data at least for a finite time T. Let us first precise the produced error during the reservoir technique computation. The following analysis is valid for all discrete initial data $(u_j^0)_{j\in\mathbb{Z}}$ defined as:

$$u^0 = \sum_{j\in\mathbb{Z}} u_j^0 \mathbf{1}_{\omega_j},$$

but in practice as we consider non-interacting waves we can also consider in this paper step-like initial data, such that for $j_0 \in \mathbb{Z}$:

$$u^0 = \sum_{j\leqslant j_0} u_l \mathbf{1}_{\omega_j} + \sum_{j>j_0} u_r \mathbf{1}_{\omega_j}.$$

Theorem 3.1 *The reservoir scheme for 1D scalar hyperbolic equation of conservation laws with a strictly convex flux, converges to the entropy solution. More precisely, there exist $C > 0$ and $D > 0$ such that:*

$$\|u^n - v(\cdot, t^n)\|_{L^1} \leqslant \|u^0 - v(\cdot, 0)\|_{L^1} + CTV(u^0)\Delta t + DTV(u^0)t^n(\Delta t)^2.$$

Proof. As there exist two kinds of non-trivial waves for such equations, shock and rarefaction waves, we propose to study the two situations and we split the proof in two parts.

Step 1. Shock waves.
Consider a time iteration n_0 such that all the counters are equal to zero (ex: $n_0 = 0$). Then for all $n \in \mathbb{N}^*$ and $j \in \mathbb{Z}$, while $c_j^{n_0+n} < 1$, we freeze the physical flux associated to the cell j at time t^{n_0}. We will by convention study the case of a right-entropy shock with a discontinuity located in $x_{j_0-1/2}$ at time t^{n_0} and of velocity $\sigma_{j_0-1/2}^{n_0}$. Now as $\mathrm{sgn}(f')$ is positive, we update the solution as follows (for $n = 1$):

$$u_j^{n_0+1} = u_j^{n_0} - \frac{\Delta t^{n_0}}{\Delta x}\big(f(u_j^{n_0}) - f(u_{j-1}^{n_0})\big)1_{c_j^{n_0+1}=1}.$$

More generally, for every n and j in $\mathbb{Z}$ such that $c_j^{n_0+n} < 1$ one has

$$\begin{aligned} u_j^{n_0+n} &= u_j^{n_0+n-1} - \frac{\Delta t^{n_0+n-1}}{\Delta x}\big(f(u_j^{n_0}) - f(u_{j-1}^{n_0})\big)1_{c_j^{n_0+n-1}=1} \\ &= u_j^{n_0} - \textstyle\sum_{i=1}^{n-1} \dfrac{\Delta t^{n_0+i}}{\Delta x}\big(f(u_j^{n_0}) - f(u_{j-1}^{n_0})\big)1_{c_j^{n_0+i-1}=1}. \end{aligned} \tag{7}$$

Recall that in this case the time step is chosen such that:

$$\Delta t^n = \min_j \Big((1 - c_j^n)\frac{\Delta x}{|\sigma_{j-1/2}^{n_0}|}\Big).$$

We can then prove:

Lemma 3.1 *Let us suppose that the exact solution v is a shock wave with a front located in $j_0\Delta x$ at time t^{n_0}. Then it exists $j' \in \mathbb{Z}$ and $n' > n_0$ such that*

$$|v(x_{j'}, t^{n'}) - u_{j'}^{n'}| = |v(x_{j_0}, t^{n_0}) - u_{j_0}^{n_0}|,$$

with

$$|v(x_{j'}, t^k) - u_{j'}^k| = |v(x_j, t^{n_0}) - u_j^{n_0}| + \mathcal{O}(\Delta t),\ \forall k,\ n_0 \leqslant k < n'.$$

We add that $j' = j_0 + sign(\sigma_{j_0-1/2}^{n_0})$.

Proof of the lemma. The shock is then supposed to be a right-entropy shock. An important point in (7) is the fact that as f' is increasing; then while $c_{j_0}^{n_0+n} < 1$ one has $c_{j_0-1}^{n_0+n} < 1$. So that the corresponding physical flux $f(u_{j_0-1}^{n_0})$ is not updated. Remark also that $f(u_{j_0}^{n_0}) - f(u_{j_0-1}^{n_0}) = \sigma_{j_0-1/2}^{n_0}\big(u_{j_0}^{n_0} - u_{j_0-1}^{n_0}\big)$ by the Rankine-Hugoniot jump condition.

Let us now denote by k_{j_0} the index such that: $c_{j_0}^{n_0+k_{j_0}} = 1$, that is $\sum_{l=1}^{k_{j_0}} \dfrac{\Delta t^{n_0+k_{j_0}}}{\Delta x}\sigma_{j_0-1/2}^{n_0} = 1$. Then

$$u_{j_0}^{n_0+k_{j_0}} = u_{j_0}^{n_0} - \sum_{l=1}^{k_{j_0}} \frac{\Delta t^{n_0+k_{j_0}}}{\Delta x}\sigma_{j_0-1/2}^{n_0}\big(u_{j_0}^{n_0} - u_{j_0-1}^{n_0}\big) = u_{j_0-1}^{n_0}.$$

At time $t^{n_0+k_{j_0}}$ the exact solution is obviously given by:

$$v(x_{j_0}, t^{n_0+k_{j_0}}) = v(x_{j_0-1}, t^{n_0}).$$

Then:

$$\sum_{j\in\mathbb{Z}} |v(x_j, t^{n_0+k_{j_0}}) - u_j^{n_0+k_{j_0}}| = \sum_{j\in\mathbb{Z}} |v(x_j, t^{n_0}) - u_j^{n_0}|.$$

That is in the current cell and for the time $t^{n_0+k_{j_0}}$ the reservoir-technique combined with the flux scheme has captured the exact solution. By induction we deduce the lemma and by symmetry the result is valid for left entropy-shocks. $\square$

Now, for all $j \in \mathbb{Z}$ and for all time $(t^{n_{0;j_i}})_{i\in\mathbb{N}}$ such that $c_j^{n_{0;j_i}} = 1$ the numerical solution will be exact in the cell j. That is:

$$|v(x_j, t^{n_{0;j_i}}) - u_j^{n_{0;j_i}}| = |v(x_j, t^{n_0}) - u_j^0|.$$

However, between times $t^{n_{0;j_i}}$ and $t^{n_{0;j_{i+1}}}$ because of (5), we have the following estimate:

$$|v(x_j, t^k) - u_j^k| = |v(x_j, 0) - u_j^0| + \mathcal{O}(\Delta t), \quad \forall k \in \{n_{0;j_i}+1, \cdots, n_{0;j_{i+1}}-1\},$$

as the solution in the cell j has been frozen until the local CFL number is equal to 1. However the error is not cumulated at each iteration. This results allows to prove that the global error stays at order 1 but in each cell there exists some times such that the error is zero: then using that u^0 belongs to BV and the previous lemma, for *shock waves*, there exists a real constant $C_1 > 0$ such that:

$$\|v(\cdot, t^n) - u^n\|_{L^1} \leqslant \|v(\cdot, 0) - u^0\|_{L^1} + C_1 TV(u^0)\Delta t \quad \forall n \in \mathbb{N}.$$

That is $S_h(t^{N_i})u^0 = S(t^{N_i})u^0 + \mathcal{O}(\Delta t)$.

Step 2. Rarefaction waves.
We suppose now that the exact solution of the continuous problem is a right-rarefaction wave. We again suppose for the sake of simplicity that $f' > 0$.
Then for all $n \in \mathbb{N}$, and $j \in \mathbb{Z}$, $\lambda_j^n \leqslant \lambda_{j+1}^n$ (entropy condition), where $\lambda_j^n = f'(u_j^n)$.
Consider n_0 such that the j^{th} counter is equal to zero (ex: $n_0 = 0$). Then for $n \in \mathbb{N}$, while $c_j^{n_0+n} < 1$ and by definition of the reservoir scheme:

$$u_j^{n_0+n} = u_j^{n_0} \quad \text{and} \quad R_j^{n_0+n} = R_j^{n_0+n-1} - \frac{\Delta t^{n_0+n}}{\Delta x}\big(f(u_j^{n_0+n-1}) - f(u_{j-1}^{n_0+n-1})\big).$$

But as $c_j^{n_0+n-1} < 1$ then $u_j^{n_0+n-1} = u_j^{n_0}$ and there exists k_j such that

$$u_j^{n_0+k_j} = u_j^{n_0} - \sum_{l=0}^{k_j-1} \frac{\Delta t^{n_0+l}}{\Delta x}\Big(f(u_j^{n_0}) - f(u_{j-1}^{n_0+l})\Big).$$

It has to be noticed that even if $\lambda_{j-1}^n \leqslant \lambda_j^n$ between t^{n_0} and $t^{n_0+k_j}$ the solution in the cell $j-1$ can have been updated (excepted if $c_{j-1}^{n_0} = 0.$). However $\lambda_{j-1}^n \leqslant \lambda_j^n$ implies necessarily that the solution in the cell $j-1$ has been updated at most one time. Then:

$$u_j^{n_0+k_j} = u_j^{n_0} - \sum_{l=0}^{k_j-1} \frac{\Delta t^{n_0+l}}{\Delta x}\Big(f(u_j^{n_0}) - f(u_{j-1}^{n_0+l})\Big).$$

with

$$u_{j-1}^{n_0+k_j-1} = \begin{cases} u_{j-1}^{n_0}, \quad \text{if } c_{j-1}^{n_0+k} < 1, \ \forall k \in \{1, \cdots, k_j - 1\}, \\ \\ u_{j-1}^{n_0} + \mathcal{O}(\Delta t), \quad \text{if } \exists k \in \{1, \cdots, k_j - 1\} \ / \ c_{j-1}^{n_0+k} = 1. \end{cases}$$

For weak enough initial data $|u_j^0 - u_{j-1}^0| = \mathcal{O}(\Delta x)$, for all j in $\mathbb{Z}$, and by regularity of f we then have:

$$\begin{aligned} u_j^{n_0+k_j} = \ & u_j^{n_0} - \textstyle\sum_{l=1}^{k_j} \frac{\Delta t^{n_0+l}}{\Delta x}\Big(\lambda_j^{n_0}(u_j^{n_0} - u_{j-1}^{n_0}) + \frac{f''(u_j^{n_0})}{2}(u_j^{n_0} - u_{j-1}^{n_0})^2 + \\ & \mathcal{O}\big((u_j^{n_0} - u_{j-1}^{n_0})\Delta t\big) + \mathcal{O}\big((u_j^{n_0} - u_{j-1}^{n_0})^3\big)\Big). \end{aligned} \tag{8}$$

That is, using that $c_j^{n_0+k_j} = 1$,

$$u_j^{n_0+k_j} = u_{j-1}^{n_0} + \alpha_j^{n_0} \frac{f''(u_j^{n_0})}{2}\big(u_{j-1}^{n_0} - u_j^{n_0}\big)^2 + \mathcal{O}\big((u_j^{n_0} - u_{j-1}^{n_0})\Delta t\big) + \mathcal{O}\big((u_j^{n_0} - u_{j-1}^{n_0})^3\big),$$

with

$$\alpha_j^{n_0} := \sum_{l=1}^{k_j} \frac{\Delta t^{n_0+l}}{\Delta x}.$$

Remark. Note that a rarefaction wave $\phi(x/t)$ solution of the scalar equation is such that $f'(\phi(x/t)) - x/t = 0$, that is in the discrete case, $f'(u_j^{n_0})\Delta t/\Delta x = 1$. This corresponds to $c_j^{n_0+k_j} = \sum_{l=1}^{k_j} \frac{\Delta t^{n_0+l}}{\Delta x}\lambda_j^{n_0} = 1$.
The following lemma will be useful to prove the convergence.

Lemma 3.2 *Suppose that the exact solution is a rarefaction wave, then for all $j \in \mathbb{Z}$ and for all n, it exists $n' > n$, such that*

$$|v(x_j, t^{n'}) - u_j^{n'}| = |v(x_j, 0) - u_j^0| + \mathcal{O}\big((\Delta t)^2\big),$$

with

$$|v(x_j, t^k) - u_j^k| = |v(x_j, 0) - u_j^0| + \mathcal{O}(\Delta t), \qquad \forall k, \qquad n \leqslant k < n'.$$

Proof of the lemma. By Taylor expansion at time $t^{n_0+k_j}$ we have

$$u(j\Delta x, t^{n_0+k_j}) = u(j\Delta x, t^{n_0}) + \sum_{l=1}^{k_j} \Delta t^{n_0+l} u_t(j\Delta x, t^{n_0}) + \mathcal{O}\big((\sum_{l=1}^{k_j} \Delta t^{n_0+l})^2\big).$$

Using the conservation law $u_t = -f(u)_x = -u_x f'(u)$ we get

$$u(j\Delta x, t^{n_0+k_j}) = u(j\Delta x, t^{n_0}) - \sum_{l=1}^{k_j} \partial_x u(j\Delta x, t^{n_0}) f'\big(u(j\Delta x, t^{n_0})\big) + \mathcal{O}\big((\sum_{l=1}^{k_j} \Delta t^{n_0+l})^2\big).$$

We can rewrite the previous equation:

$$\begin{aligned} u(j\Delta x, t^{n_0+k_j}) = \;& u(j\Delta x, t^{n_0}) + \textstyle\sum_{l=1}^{k_j} \Delta t^{n_0+l} \dfrac{u(j\Delta x, t^{n_0}) - u((j-1)\Delta x, t^{n_0})}{\Delta x} f'\big(u(j\Delta x, t^{n_0})\big) + \\ & \mathcal{O}\big((\textstyle\sum_{l=1}^{k_j} \Delta t^{n_0+l})^2\big) + \mathcal{O}\big(\Delta x \textstyle\sum_{l=1}^{k_j} \Delta t^{n_0+l}\big). \end{aligned}$$

If we denote by $\phi(x/t)$ the exact self-similar continuous rarefaction wave, we have then

$$u(j\Delta x, t^{n_0+k_j}) = \phi\big(\frac{j\Delta x}{t^{n_0+k_j}}\big)$$

and then

$$\begin{aligned} \phi\big(\frac{j\Delta x}{t^{n_0+k_j}}\big) = \;& \phi\big(\frac{j\Delta x}{t^{n_0}}\big) + \textstyle\sum_{l=1}^{k_j} \dfrac{\Delta t^{n_0+l}}{\Delta x} f'\Big(\phi\big(\frac{j\Delta x}{t^{n_0}}\big)\Big)\Big[\phi\big(\frac{j\Delta x}{t^{n_0}}\big) - \phi\big(\frac{(j-1)\Delta x}{t^{n_0}}\big)\Big] + \\ & \mathcal{O}\big((\textstyle\sum_{l=1}^{k_j} \Delta t^{n_0+l})^2\big) + \mathcal{O}\big(\Delta x \textstyle\sum_{l=1}^{k_j} \Delta t^{n_0+l}\big). \end{aligned}$$

Then we recover the expression (8) at order 2:

$$|v(x_j, t^{n_0+k_j}) - u_j^{n_0+k_j}| = |v(x_j, t^{n_0}) - u_j^{n_0}| + \mathcal{O}\big((\Delta t)^2\big).$$

That means that for the cell j, there exists a time such that the error in this cell is reduced to the second order. This result has been proved for a particular time t^{n_0} and can be obviously extended to any positive time, by induction. This concludes on the lemma proof. □

Again for all $j \in \mathbb{Z}$ and for all time $(t^{n_{0;j_i}})_{i\in\mathbb{N}}$ such that $c_j^{n_{0;j_i}} = 1$, the numerical solution will be exact at order 2. That is:

$$|v(x_j, t^{n_{0;j_i}}) - u_j^{n_{0;j_i}}| = |v(x_j, 0) - u_j^0| + \mathcal{O}\big((\Delta t)^2\big).$$

However, between $t^{n_{0;j_i}}$ and $t^{n_{0;j_{i+1}}}$ because of (5), we have the following estimate:

$$|v(x_j, t^k) - u_j^k| = |v(x_j, 0) - u_j^0| + \mathcal{O}(\Delta t), \quad \forall k \in \{n_{0;j_i} + 1, \cdots, n_{0;j_{i+1}} - 1\},$$

There is then no order one error accumulations during the time process. So that for $u^0 \in BV$ such that the solution is a rarefaction wave there exists two real constants C_2 and C_3 such that:

$$\|u^n - v(\cdot, t^n)\|_{L^1} \leqslant \|u^0 - v(\cdot, 0)\|_{L^1} + C_2 TV(u^0)\Delta t + C_3 TV(u^0) t^n (\Delta t)^2.$$

Step 3. Conclusion.
As the solution of a hyperbolic equation of conservation law is a combination of shock and rarefaction waves and as we consider non-interacting waves we can deduce that there exist $C > 0$ and $D > 0$ such that:

$$\|u^n - v(\cdot, t^n)\|_{L^1} \leqslant \|u^0 - v(\cdot, 0)\|_{L^1} + CTV(u^0)\Delta t + DTV(u^0) t^n (\Delta t)^2.$$

i.e.

$$S_h(t_n)u^n = S(t_n)u^n + CTV(u^0)\Delta t + DTV(u^0) t^n (\Delta t)^2.$$

That is, for step-like initial data the reservoir scheme convergences to the entropy solution. □

We now prove that for scalar equations the combination of flux schemes and reservoirs is total variation diminushing (TVD).

Theorem 3.2 *The reservoir technique for scalar hyperbolic equations of conservation laws without wave interaction, and a strictly convex flux, is TVD:*

$$TV(u^{n+1}) = \sum_{j\in\mathbb{Z}} |u_j^{n+1} - u_{j-1}^{n+1}| \leqslant \sum_{j\in\mathbb{Z}} |u_j^n - u_{j-1}^n| = TV(u^n), \ \forall n \in \mathbb{N}.$$

Proof. Recall first that under CFL condition flux schemes are TVD [10]. If we denote by w^n a flux scheme solution and we suppose that at each iteration the time step is chosen such that the CFL number is equal to 1, we then have:

$$TV(u^{n+1}) \leqslant TV(u^n), \ \forall n \in \mathbb{N}.$$

Now denoting by $(n_{0;j})_j$ some indices such that $c_j^{n_{0;j}} = 1$, the reservoir scheme is defined in the cell j and for n greater than $n_{0;j}$, by:

$$u_j^{n+1} = \begin{cases} u_j^{n_{0;j}}, \quad \text{if } c_j^{n+1} < 1, \\ \\ u_j^{n_{0;j}} - \sum_{l=0}^{n-n_{0;j}} \dfrac{\Delta t^{n_{0;j}+l}}{\Delta x}\big(g(u_j^{n_{0;j}}, u_{j+1}^{n+l}) - g(u_{j-1}^{n+l}, u_j^{n_{0;j}})\big), \quad \text{if } \ c_j^{n+1} = 1. \end{cases}$$

Then if we sum over j

$$\begin{aligned} \sum_{j\in\mathbb{Z}} |u_j^{n+1} - u_{j-1}^{n+1}| = \ & \sum_{j\in\mathbb{Z}} \Big| u_j^{n_{0;j}} - \sum_{l=0}^{n-n_{0;j}} \frac{\Delta t^{n_{0;j}+l}}{\Delta x}(\cdots) 1_{c_j^{n+1}=1} \\ & - u_{j-1}^{n_{0;j-1}} + \sum_{l=0}^{n-n_{0;j-1}} \frac{\Delta t^{n_{0;j-1}+l}}{\Delta x}(\cdots) 1_{c_{j-1}^{n+1}=1} \Big|. \end{aligned}$$

As we again do not consider waves interactions only three situtations can occur:

Step 1. Shock waves
We have seen that the solution is updated only when the local CFL number reaches 1; this corresponds to an exact shock propagation. Then the total variation is then exactly conserved in this case.

Step 2. Constant states
Trivially the total variation is zero.

Step 3. Rarefaction waves
This situation is more complicated. We will prove for right-rarefaction waves (by symmetry the result will be valid for left rarefaction waves), that the total variation diminushing using an inductive argument and the reservoirs and counters features.
From time $t = 0$ to t^1 we have:

$$u_j^1 = \begin{cases} u_j^0, \ \text{ if } \ c_j^1 < 1, \\ \\ u_j^0 - \dfrac{\Delta t^0}{\Delta x}\big(f(u_j^0) - f(u_{j-1}^0)\big), \ \text{ if } \ c_j^1 = 1. \end{cases}$$

and

$$u_{j-1}^1 = \begin{cases} u_{j-1}^0, \ \text{ if } \ c_{j-1}^1 < 1, \\ \\ u_{j-1}^0 - \dfrac{\Delta t^0}{\Delta x}\big(f(u_{j-1}^0) - f(u_{j-2}^0)\big), \ \text{ if } \ c_{j-1}^1 = 1. \end{cases}$$

Then, as f is increasing (recall that we consider right-rarefaction waves) and using the fact that the time step is supposed to be chosen such that the local CFL number is less or equal to 1, we check easily that for right-rarefaction waves:

$$|u_j^1 - u_{j-1}^1| \leqslant |u_j^0 - u_{j-1}^0|,$$

so that:

$$TV(u^1) = \sum_{j\in\mathbb{Z}} |u_j^1 - u_{j-1}^1| \leqslant \sum_{j\in\mathbb{Z}} |u_j^0 - u_{j-1}^0| = TV(u^0).$$

Suppose now that at time t^n:

$$TV(u^k) \leqslant TV(u^{k-1}), \ 1 \leqslant k \leqslant n.$$

At time t^{n+1} we have the following equations:

$$\begin{cases} u_j^{n+1} = u_j^n - \sum_{l=0}^{n-n_{0;j}} \dfrac{\Delta t^{n_{0;j}+l}}{\Delta x}\big(f(u_j^{n_{0;j}}) - f(u_{j-1}^{n_{0;j}+l})\big)1_{c_j^{n+1}=1}, \\ \\ u_{j-1}^{n+1} = u_{j-1}^n - \sum_{l=0}^{n-n_{0;j-1}} \dfrac{\Delta t^{n_{0;j-1}+l}}{\Delta x}\big(f(u_{j-1}^{n_{0;j-1}}) - f(u_{j-2}^{n_{0;j-1}+l})\big)1_{c_{j-1}^{n+1}=1}. \end{cases}$$

Then

$$|u_j^{n+1} - u_{j-1}^{n+1}| = \Big|u_j^n - u_{j-1}^n - \sum_{l=0}^{n-n_{0;j}} \frac{\Delta t^{n_{0;j}+l}}{\Delta x}\big(f(u_j^{n_{0;j}}) - f(u_{j-1}^{n_{0;j}+l})\big)1_{c_j^{n+1}=1} +$$

$$\sum_{l=0}^{n-n_{0;j-1}} \frac{\Delta t^{n_{0;j-1}+l}}{\Delta x}\big(f(u_{j-1}^{n_{0;j-1}}) - f(u_{j-2}^{n_{0;j-1}+l})\big)1_{c_{j-1}^{n+1}=1}\Big|.$$

Four situations can occur (corresponding to four kinds of cells):

- If $c_j^{n+1} < 1, \;\; c_{j-1}^{n+1} < 1$ then

  $$u_j^{n+1} = u_j^n = u_j^{n_{0;j}}, \;\; u_{j-1}^{n+1} = u_{j-1}^n = u_{j-1}^{n_{0;j-1}}.$$

  In such cells j we have:

  $$|u_j^{n+1} - u_{j-1}^{n+1}| = |u_j^n - u_{j-1}^n|.$$

  For such cells there is no variation increasing.

- If $c_j^{n+1} = 1, \;\; c_{j-1}^{n+1} < 1$ then

  $$|u_j^{n+1} - u_{j-1}^{n+1}| = \Big|u_j^n - u_{j-1}^n - \sum_{l=0}^{n-n_{0;j}} \frac{\Delta t^{n_{0;j}+l}}{\Delta x}\big(f(u_j^{n_{0;j}}) - f(u_{j-1}^{n_{0;j}+l})\big)\Big|.$$

  Then as f is increasing (right-rarefaction waves are considered) and by entropy condition the term inside the sum is positive and then:

  $$-\sum_{l=0}^{n-n_{0;j}} \frac{\Delta t^{n_{0;j}+l}}{\Delta x}\big(f(u_j^{n_{0;j}}) - f(u_{j-1}^{n_{0;j}+l})\big) \leqslant 0.$$

  We have (under CFL condition):

  $$|u_j^{n+1} - u_{j-1}^{n+1}| \leqslant |u_j^n - u_{j-1}^n|.$$

- If now $c_j^{n+1} < 1, \;\; c_{j-1}^{n+1} = 1$, then

  $$|u_j^{n+1} - u_{j-1}^{n+1}| = \Big|u_j^n - u_{j-1}^n + \sum_{l=0}^{n-n_{0;j-1}} \frac{\Delta t^{n_{0;j-1}+l}}{\Delta x}\big(f(u_{j-1}^{n_{0;j-1}}) - f(u_{j-2}^{n_{0;j-1}+l})\big)\Big|.$$

  And again the term inside the sum is positive, so that:

  $$|u_j^{n+1} - u_{j-1}^{n+1}| = |u_j^n - u_{j-1}^n + \sum_{l=0}^{n-n_{0;j-1}} \frac{\Delta t^{n_{0;j-1}+l}}{\Delta x}\big(f(u_{j-1}^{n_{0;j-1}}) - f(u_{j-2}^{n_{0;j-1}+l})\big)|.$$

  $$\geqslant |u_j^n - u_{j-1}^n|.$$

We can then not conclude directly on the total variation. However in the cell $j-2$:

$$u_{j-2}^{n+1} = u_{j-2}^{n} - \sum_{l=0}^{n-n_{0;j-2}} \frac{\Delta t^{n_{0;j-2}+l}}{\Delta x}\big(f(u_{j-2}^{n_{0;j-2}}) - f(u_{j-3}^{n_{0;j-2}+l})\big)1_{c_{j-2}^{n+1}=1}.$$

Now let denote by k, with $k \geqslant 2$ the first cell such that $c_{j-k'}^{n+1} < 1$ with $c_{j-k}^{n+1} = 1$ for $k' \in \{1, \cdots, k-1\}$. To simplify the presentation let us suppose that $k = 2$. Then if $c_{j-2}^{n+1} < 1$:

$$\begin{aligned} |u_{j-1}^{n+1} - u_{j-2}^{n+1}| = & \left|u_{j-1}^{n} - u_{j-2}^{n} - \textstyle\sum_{l=0}^{n-n_{0;j-1}} \frac{\Delta t^{n_{0;j-1}+l}}{\Delta x}\big(f(u_{j-1}^{n_{0;j-1}}) - f(u_{j-2}^{n_{0;j-1}+l})\big)\right| \\ & \leqslant |u_{j-1}^{n} - u_{j-2}^{n}|. \end{aligned}$$

Defining A by:

$$A = u_{j-1}^{n} - u_{j-2}^{n} - \sum_{l=0}^{n-n_{0;j-1}} \frac{\Delta t^{n_{0;j-1}+l}}{\Delta x}\big(f(u_{j-1}^{n_{0;j-1}}) - f(u_{j-2}^{n_{0;j-1}+l})\big),$$

we obtain by Taylor expansion:

$$\begin{aligned} A = & \quad u_{j-1}^{n} - u_{j-2}^{n} - \textstyle\sum_{l=0}^{n-n_{0;j-1}} \frac{\Delta t^{n_{0;j-1}+l}}{\Delta x}\Big(f(u_{j-1}^{n_{0;j-1}}) - f(u_{j-1}^{n_{0;j-1}+l}) - \\ & f'(u_{j-1}^{n_{0;j-1}+l})(u_{j-2}^{n_{0;j-1}+l} - u_{j-1}^{n_{0;j-1}+l}) - f^{(2)}(u_{j-1}^{n_{0;j-1}+l})(u_{j-2}^{n_{0;j-1}+l} - u_{j-1}^{n_{0;j-1}+l})^2/2 + \\ & \mathcal{O}\big((u_{j-2}^{n_{0;j-1}+l} - u_{j-1}^{n_{0;j-1}+l})^3\big)\Big). \end{aligned}$$

So using the fact that

$$\sum_{l=0}^{n-n_{0;j-1}} \lambda_{j-1}^{n_{0;j-1}+l} \frac{\Delta t^{n_{0;j-1}+l}}{\Delta x} = 1,$$

we deduce that

$$A = \frac{1}{2} f^{(2)}(u_{j-1}^{n_{0;j-1}+l})(u_{j-2}^{n_{0;j-1}+l} - u_{j-1}^{n_{0;j-1}+l})^2 + \mathcal{O}\big((u_{j-2}^{n_{0;j-1}+l} - u_{j-1}^{n_{0;j-1}+l})^3\big) \geqslant 0.$$

Then we can deduce that:

$$|u_{j-1}^{n+1} - u_{j-2}^{n+1}| + |u_{j}^{n+1} - u_{j-1}^{n+1}| \leqslant |u_{j-1}^{n} - u_{j-2}^{n}| + |u_{j}^{n} - u_{j-1}^{n}|.$$

As:

$$|u_{j}^{n+1} - u_{j-1}^{n+1}| = \left|u_{j}^{n} - u_{j-1}^{n} + \textstyle\sum_{l=0}^{n-n_{0;j-1}} \frac{\Delta t^{n_{0;j-1}+l}}{\Delta x}\big(f(u_{j-1}^{n_{0;j-1}}) - f(u_{j-2}^{n_{0;j-1}+l})\big)\right|,$$

$$|u_{j-1}^{n+1} - u_{j-2}^{n+1}| = \left|u_{j-1}^{n} - u_{j-2}^{n} - \textstyle\sum_{l=0}^{n-n_{0;j-1}} \frac{\Delta t^{n_{0;j-1}+l}}{\Delta x}\big(f(u_{j-1}^{n_{0;j-1}}) - f(u_{j-2}^{n_{0;j-1}+l})\big)\right|.$$

In the situation $k \geqslant 3$ with $c_{j-k}^{n+1} < 1$ with $c_{j-k'}^{n+1} = 1$ for $k' \in \{1, \cdots, k-1\}$ we would have:

$$\sum_{l=0}^{k} |u_{j-l}^{n+1} - u_{j-l-1}^{n+1}| \leqslant \sum_{l=0}^{k} |u_{j-l}^{n} - u_{j-l-1}^{n}|.$$

- The last case corresponds to $c_j^{n+1} = 1, \;\; c_{j-1}^{n+1} = 1$. That is:

$$\begin{cases} u_j^{n+1} = u_j^n - \sum_{l=0}^{n-n_{0;j}} \dfrac{\Delta t^{n_{0;j}+l}}{\Delta x}\big(f(u_j^{n_{0;j}}) - f(u_{j-1}^{n_{0;j}+l})\big), \\ \\ u_{j-1}^{n+1} = u_{j-1}^n - \sum_{l=0}^{n-n_{0;j-1}} \dfrac{\Delta t^{n_{0;j-1}+l}}{\Delta x}\big(f(u_{j-1}^{n_{0;j-1}}) - f(u_{j-2}^{n_{0;j-1}+l})\big). \end{cases}$$

But as f' is strictly positive that necessarily involves that $n_{0;j-1} < n_{0;j}$ and we can expect that:

$$|u_j^{n+1} - u_{j-1}^{n+1}| \geqslant |u_j^n - u_{j-1}^n|.$$

But as above (third points) we prove that there exists a k such that:

$$\sum_{l=0}^{k} |u_{j-l}^{n+1} - u_{j-l-1}^{n+1}| \leqslant \sum_{l=0}^{k} |u_{j-l}^{n} - u_{j-l-1}^{n}|, \; \forall k \leqslant n.$$

So finally globally for rarefaction waves we have:

$$TV(u^{n+1}) = \sum_{j\in\mathbb{Z}} |u_j^{n+1} - u_{j-1}^{n+1}| \leqslant \sum_{j\in\mathbb{Z}} |u_j^n - u_{j-1}^n| = TV(u^n).$$

□

4 Linear systems

In this section we study the convergence of the reservoir technique approximating linear systems of conservation laws without waves interaction.

4.1 Reservoir scheme for linear systems

Consider the linear hyperbolic system with m equations:

$$\begin{cases} \partial_t V \; + \; \partial_x(AV) \; = \; 0, \; x \in \mathbb{R}, \; t \geqslant 0, \\ V(x,0) = V_0(x), \end{cases} \tag{9}$$

with $V = (v_1, \cdots, v_m)$, $A \in \mathcal{M}_m(\mathbb{R})$ diagonalizable in $\mathbb{R}$ with eigenvalues $(\lambda_1, \lambda_2, \cdots, \lambda_m)$, and corresponding right eigenvectors $R = col(r_1, \cdots, r_m)$, associated to the initial data

$V_0 \in BV(\mathbb{R})$. Note that the change of variables $W = R^{-1}V = (w_1, \cdots, w_m)$, satisfies the following diagonal system of m independent advection equations:

$$\left[\begin{array}{rcl} \partial_t w_1 \ + \ \lambda_1 \partial_x w_1 & = & 0, \\ \vdots & & \\ \partial_t w_m \ + \ \lambda_m \partial_x w_m & = & 0. \end{array} \right.$$

Let us introduce the flux $F(V) = AV$. We are looking for an approximation V_j^n of $V(t^n, j\Delta x)$ in (9). The initial data is given by: $V_j^0 = V_0(j\Delta x)$, $j \in \mathbb{Z}$. Then the classical upwind scheme may be written in this case:

$$V_j^{n+1} = V_j^n - \frac{\Delta t_n}{\Delta x} \left(\Phi_{j+\frac{1}{2}}^n - \Phi_{j-\frac{1}{2}}^n \right), j \in \mathbb{Z}, n \geqslant 1,$$

where $\Delta t_n = t^{n+1} - t^n$ and $\Phi_{j+\frac{1}{2}}^n$ is the interfacial upwinded flux :

$$\Phi_{j+\frac{1}{2}}^n = \frac{F(V_j^n) + F(V_{j+1}^n)}{2} - \frac{1}{2}\mathrm{sgn}(A)\,(F(V_{j+1}^n) - F(V_j^n)).$$

Here, $\mathrm{sgn}(\cdot)$ is the sign matrix given by $\mathrm{sgn}(A) = R\ \mathrm{diag}_{i=1,\cdots,p}(\mathrm{sgn}(\lambda_i))R^{-1}$. We easily see that the scheme can also be written as:

$$W_j^{n+1} = W_j^n - \frac{\Delta t_n}{\Delta x}\, R^{-1} \left(\Phi_{j+\frac{1}{2}}^n - \Phi_{j-\frac{1}{2}}^n \right),$$

with $W_j^n = R^{-1}V_j^n$ or componentwise,

$$\left[\begin{array}{rcl} w_{k,j}^{n+1} & = & w_{k,j}^n - \alpha_k \left(w_{k,j+1}^n - w_{k,j}^n \right), \ \alpha_k \leqslant 0, \\ w_{k,j}^{n+1} & = & w_{k,j}^n - \alpha_k \left(w_{k,j}^n - w_{k,j-1}^n \right), \ \alpha_k > 0, \end{array} \right.$$

with $\alpha_k = \lambda_k\, \Delta t_n/\Delta x$, $k = 1, \cdots, m$. It is well-known that this scheme is ℓ^2-stable under the CFL condition:

$$\alpha \ \stackrel{\mathrm{def}}{:=} \ \frac{\Delta t_n}{\Delta x} \max(|\lambda_1|, \cdots, |\lambda_m|), \text{ satisfies } \alpha \leqslant 1. \tag{10}$$

If Δt_n is chosen such that $\alpha = 1$, the propagation is perfectly solved on the components w_k for which $\alpha_k = 1$ and from the analysis done in the introduction, some numerical diffusion will be produced on the components w_j for which $\alpha_j < 1$. To obtain a CFL equal to one on each characteristic field, let us consider different time steps Δt_n (not necessarily equal). We introduce m vectorial reservoirs $R_{1;j}, \cdots, R_{m;j} \in \mathbb{R}^m$ associated to the cell j and CFL counters $c_{k;j} \in [0,1]$, $k = 1, \cdots, m$, initialized to zero[2]. We denote by $V_{\mathrm{R}}^k(V, W)$ the solution of the Riemann problem with left state V and right state W which lies between the k^{th} and the

[2]Note that for each characteristic field these quantities are constant in space and the index j is then not necessary, but we keep this notation for the forthcoming sections.

$(k+1)^{\text{th}}$ wave with $V_{\mathrm{R}}^0(V,W) := V$, and $V_{\mathrm{R}}^m(V,W) := W$ by convention. For convenience we set in the following

$$\mathcal{C}_{k;j}^n = c_{k;j}^n + |\lambda_k| \frac{\Delta t_n}{\Delta x}.$$

At each time step we fill up the reservoirs $R_{k;j}$ with the current numerical flux difference upwinding with respect to the sign of λ_k. More precisely, we take for $\lambda_k < 0$

$$\begin{pmatrix} \tilde{V}_{k;j}^{n+1} \\ c_{k;j}^{n+1} \\ R_{k;j}^{n+1} \end{pmatrix} = \begin{cases} \begin{pmatrix} 0 \\ c_{k;j}^n + |\lambda_k| \dfrac{\Delta t_n}{\Delta x} \\ R_{k;j}^n - \dfrac{\Delta t_n}{\Delta x}\big(F(V_{\mathrm{R}}^k(V_j^n, V_{j+1}^n)) - F(V_{\mathrm{R}}^{k-1}(V_j^n, V_{j+1}^n))\big) \end{pmatrix}, & \text{if } \mathcal{C}_{k;j}^n < 1, \\ \begin{pmatrix} R_{k;j}^n - \dfrac{\Delta t_n}{\Delta x}\big(F(V_{\mathrm{R}}^k(V_j^n, V_{j+1}^n)) - F(V_{\mathrm{R}}^{k-1}(V_j^n, V_{j+1}^n))\big) \\ 0 \\ 0 \end{pmatrix}, & \text{if } \mathcal{C}_{k;j}^n = 1. \end{cases} \tag{11}$$

whereas, for $\lambda_k > 0$, we upwind on the right side

$$\begin{pmatrix} \tilde{V}_{k;j}^{n+1} \\ c_{k;j}^{n+1} \\ R_{k;j}^{n+1} \end{pmatrix} = \begin{cases} \begin{pmatrix} 0 \\ c_{k;j}^n + |\lambda_k| \dfrac{\Delta t_n}{\Delta x} \\ R_{k;j}^n - \dfrac{\Delta t_n}{\Delta x}\big(F(V_{\mathrm{R}}^k(V_{j-1}^n, V_j^n)) - F(V_{\mathrm{R}}^{k-1}(V_{j-1}^n, V_j^n))\big) \end{pmatrix}, & \text{if } \mathcal{C}_{k;j}^n < 1, \\ \begin{pmatrix} R_{k;j}^n - \dfrac{\Delta t_n}{\Delta x}\big(F(V_{\mathrm{R}}^k(V_{j-1}^n, V_j^n)) - F(V_{\mathrm{R}}^{k-1}(V_{j-1}^n, V_j^n))\big) \\ 0 \\ 0 \end{pmatrix}, & \text{if } \mathcal{C}_{k;j}^n = 1. \end{cases} \tag{12}$$

Eventually, we update the solution by taking

$$V_j^{n+1} = V_j^n + \sum_{k=1}^m \tilde{V}_{k;j}^{n+1}.$$

As before, the timestep Δt_n must be chosen according to the classical stability condition and as big as possible. This leads to the natural choice

$$\Delta t_n = \min_{j,k} \big((1 - c_{k;j}^n) \frac{\Delta x}{|\lambda_k|}\big). \tag{13}$$

4.2 Numerical analysis

We then consider

$$V_t + AV_x = 0, \quad V(x,0) = V_0(x), \ V_0 \in L^1 \cap L^\infty \cap BV. \tag{14}$$

with $A \in M_m(\mathbb{R})$ diagonalizable in $\mathbb{R}$ such that $\lambda_1 < \lambda_2 < \cdots < \lambda_m$. The initial data is given by:

$$V_j^0 = \begin{cases} V_l := \sum_{i=1}^m \alpha_i \mathrm{r}_i, & \text{if } j \leqslant j_0, \\ \\ V_r := \sum_{i=1}^m \beta_i \mathrm{r}_i, & \text{if } j > j_0. \end{cases}$$

Recall that an explicit flux scheme approximating a linear system of conservation law, is convergent (in particular stable) under a CFL condition $\max_{i\in\{1,\cdots,m\}} |\lambda_i|\Delta t/\Delta x \leqslant 1$. Again we will use the fact that the reservoir technique combined with a flux scheme is equivalent to choose particular time steps.

Theorem 4.1 *A flux scheme combined with the reservoir technique approaching a linear hyperbolic system of conservation laws is convergent and there exists $C > 0$ such that, at all time t^n:*

$$S_h(t^n)\mathbf{V}^0 = S(t^n)\mathbf{V}^0 + C\Delta t, \qquad \|V(\cdot,t^n) - V^n\|_{L^1} \leqslant \|V(\cdot,0) - V^0\|_{L^1} + C\Delta t,$$

where S_h denotes the associated discrete semi-group and S the exact one.
Moreover for rational eigenvalues and for all time t^n, there exists a time $t^{n'} \geqslant t^n$ such that the discrete solution is exact at the discrete level, that is:

$$S_h(t^{n'})\mathbf{V}^0 = S(t^{n'})\mathbf{V}^0, \qquad \|V(\cdot,t^{n'}) - V^{n'}\|_{L^1} = \|V(\cdot,0) - V^0\|_{L^1}.$$

Proof. We first enounce an important lemma:

Lemma 4.1 $\forall j \in \mathbb{Z}, \forall l \in \{1,\cdots,m\}, \ \forall n > 0, \ \exists n' > n,$

$$\sum_j |V_l(x_j,t^{n'}) - V_{j;l}^{n'}| = \sum_j |V_l(x_j,0) - V_{j;l}^0|,$$

with

$$\sum_j |V_l(x_j,t^k) - V_{j;l}^k| = \sum_j |V_l(x_j,0) - V_{j;l}^0| + \mathcal{O}(\Delta t), \qquad \forall k, \qquad n \leqslant k < n'.$$

or differently speaking $\forall j \in \mathbb{Z}, \ \forall l \in \{1,\cdots,m\}, \ \forall n_{0;j}$ such that $c_j^{n_{0;j}} = 0, \ \exists k_j > 0$, such that $c_j^{n_{0;j}+k_j} = 1$, and

$$\sum_j |V_l(x_j,t^{n_{0;j}+k_j}) - V_{l;j}^{n_{0;j}+k_j}| = \sum_j |V_l(x_j,0) - V_{j;l}^0|,$$

with

$$\sum_j |V_l(x_j,t^{n_{0;j}+k}) - V_{j;l}^{n_{0;j}+k}| = \sum_j |V_l(x_j,0) - V_{j;l}^0| + \mathcal{O}(\Delta t), \qquad \forall k, \qquad n \leqslant k < k_j.$$

Proof of lemma 4.1. In the proof, we suppose that $\lambda_i \geqslant 0$, for all $1 \leqslant i \leqslant m$ and that $\lambda_m/\lambda_1 \leqslant 2$. These assumptions do not delete any theoritical difficulty, but allows us to simplify the notations. In particular the second one ensures us that for all N in $\mathbb{N}$, from the time iterations $N \times m$ to $(N+1) \times m$ each counter (recall that there are m counters) is updated one and only one time, what then simplifies the indexations.
Suppose that the current time step index is equal to n with[3] $c^n_{k;j} = 0$ for all k in $\{1, \cdots m\}$ and all j in $\mathbb{Z}$ (such a n exists, for instance $n = 0$). We consider an initial data given by (4.2). Then at each time step we update the solution in each cell the following way:

$$V^{n+1}_{j+1} = V^n_{j+1} - \frac{\Delta t^n}{\Delta x} A\Big(V^{m+1}_{\mathrm{R}}(V^n_j, V^n_{j+1}) - V^m_{\mathrm{R}}(V^n_j, V^n_{j+1})\Big) = V^n_{j+1} - \frac{\Delta t^n}{\Delta x} A\Big(V^n_{j+1} - V^m_{\mathrm{R}}(V^n_j, V^n_{j+1})\Big).$$

Diagonalizing A it is always possible to suppose that A is diagonal and given by $A = \mathrm{diag}(\lambda_1, \cdots, \lambda_m)$. First, Δt^n is chosen such that $c^{n+1}_{m;j+1} = \lambda_m \Delta t^n/\Delta x = 1$ and as

$$V^n_{j+1} - V^m_{\mathrm{R}}(V^n_j, V^n_{j+1}) = \big(\beta_m - \alpha_m\big)\mathbf{r}_m,$$

then

$$V^{n+1}_{j+1} = V^n_{j+1} - \big(\beta_m - \alpha_m\big)\mathbf{r}_m = \sum_{i=1}^{m-1} \beta_i \mathbf{r}_i + \alpha_m \mathbf{r}_m.$$

This corresponds to an exact propagation at the discrete level for the m^{th} wave, that is without numerical diffusion.
Now, for the $(m-p)^{\text{th}}$ characteristic, with $p \leqslant m-1$:

$$V^{n+p+1}_{j+1} = V^{n+p}_{j+1} - \textstyle\sum_{i=0}^{p} \dfrac{\Delta t^{n+i}}{\Delta x} A\Big(V^{m-p+1}_{\mathrm{R}}(V^{n+p}_j, V^{n+p}_{j+1}) - V^{m-p}_{\mathrm{R}}(V^{n+p}_j, V^{n+p}_{j+1})\Big),$$

with

$$V^{m-p+1}_{\mathrm{R}}(V^{n+p}_j, V^{n+p}_{j+1}) - V^{m-p}_{\mathrm{R}}(V^{n+p}_j, V^{n+p}_{j+1}) = \big(\beta_{m-p+1} - \alpha_{m-p+1}\big)\mathbf{r}_{m-p+1}.$$

This time $c^{n+p}_{m-p;j+1} = \lambda_{m-p} \sum_{i=0}^{p} \Delta t^{n+i}/\Delta x = 1$. Now, as $(\mathbf{r}_i)_{i\in\{1,\cdots,m\}}$ are the eigenvectors of A:

$$V^{n+p+1}_{j+1} = V^{n+p}_{j+1} - \sum_{i=0}^{p} \frac{\Delta t^{n+i}}{\Delta x} \lambda_{m-p}\big(\beta_{m-i+1} - \alpha_{m-i+1}\big)\mathbf{r}_{m-i+1} = \sum_{i=p}^{m} \alpha_i \mathbf{r}_i + \sum_{i=1}^{p-1} \beta_i \mathbf{r}_i.$$

That then proves that for each component $p \leqslant m$ the propagation of the p^{th} wave is exact. Finally by induction, we can sum-up the process by:

$$V^{n+m}_{j+1} = V^n_{j+1} - \sum_{i=1}^{m} \frac{\Delta t^{n+i}}{\Delta x} \lambda_{m-i+1}\big(\beta_{m-i+1} - \alpha_{m-i+1}\big)\mathbf{r}_{m-i+1} = V^n_j.$$

[3]for example $n = 0$.

This process can then easily be extended to all time allowing to deduce the lemma, where $V_{j;l}^n$ denotes the l^{th} solution component in cell j at time t^n. $\square$

This lemma ensures us that at time t, for each characteristic wave there exists a time $t' \geqslant t$ such that the propagation is exact at time t'. However this does not allow us to conclude that: at all time t there exists a time $t' \geqslant t$ such that at time t' the discrete solution is exact (without numerical diffusion), that is the propagation is exact for all characteristic waves all together.

Remark If for all $i \in \{1, \cdots, m-1\}$ $\lambda_{i+1} = 2^i \lambda_1$, then the discrete solution of the reservoir scheme is exact every 2^m iterations. More precisely the propagation of the l^{th} characteristic wave, $l \in \{1, \cdots, m\}$, of the discrete solution is exact every 2^l iterations.

To generalize this idea let us enounce the following lemma:

Lemma 4.2 *Suppose that $\lambda_i \in \mathbb{Z}^*$, for all $i \in \{1, \cdots, m\}$ and denote by $P_{1,\cdots,m}$ the following least common multiple:*

$$P_{1,\cdots,m} = lcm(|\lambda_1|, \cdots, |\lambda_m|).$$

Then the discrete solution is exact every $P_{1,\cdots,m}$ iterations.

Proof. Let us denote by t^n the current time, such that all counters are zero (for example $n = 0$). We then define $(r_i)_{i\{1,\cdots,m\}}$:

$$r_i = \Big|\frac{\lambda_{i+1}}{\lambda_i}\Big| \in \mathbb{Q}^*, \qquad \forall i \in \{1, \cdots, m\},$$

and by δt_i (CFL=1 for the component i):

$$\delta t_i = \frac{\Delta x}{|\lambda_i|}, \qquad \forall i \in \{1, \cdots, m\}.$$

Note that $\delta t_i = r_i \delta t_{i+1}$ and recall that at times $t^n + N\delta t_i$, $N \in \mathbb{N}$, the propagation of the i^{th} wave is exact.
As λ_m is the fastest velocity $\Delta t^{n+1} = \delta t_m = \Delta x/|\lambda_m|$.

- If $r_{m-1} \in \mathbb{N}^*$ at iterations $[n, n + r_{m-1}[$ the m^{th} reservoir is the only one to be updated, with zero diffusion on the associated characteristic wave. Then by definition of the scheme at time $t^{n+r_{m-1}}$ the m^{th} and $(m-1)^{\text{th}}$ reservoirs are updated and the diffusion for the corresponding characteristic waves is zero.

- If now $r_{m-1} \in \mathbb{Q}^* - \mathbb{N}$ and denoting by $P_{m,m-1} = lcm(|\lambda_m|, |\lambda_{m-1}|)$ then we can assert that at times $t^n + NP_{m,m-1}\delta t_m$, $N \in \mathbb{N}^*$ the propagation is exact for these two characteristic waves.

Using similar arguments, for all k and l with $1 \leqslant k < l < m$, the propagation of the discrete k^{th} and l^{th} characteristic waves are exact at times $t^n + NP_{l,k}\delta t_l$ with $P_{k,l} = lcm(|\lambda_k|, |\lambda_l|)$ and $N \in \mathbb{N}^*$. By induction on the wave indices, we easily deduce that at times $t^n + NP_{1,\cdots,m}\delta t_m$, $N \in \mathbb{N}^*$ so that the propagation is exact for all characteristic waves and the discrete solution is exact at the discrete level.□

In fact we can even easily extend this results to rational eigenvalues.

Lemma 4.3 *Supposing the eigenvalues rational, then for all time iteration n there exists $n' \geqslant n$ such that the discrete solution is exact at time $t^{n'}$.*

Proof. Let us write λ_i as

$$\lambda_i = p_i/q_i, \qquad i \in \{1, \cdots, m\}, \qquad p_i,\ q_i \text{ relatively prime integers,}$$

and denote by $R_{1,\cdots,m}$ the following least common multiple:

$$R_{1,\cdots,m} = lcm(|p_1|, \cdots, |p_m|).$$

With the same arguments than above we prove that the discrete solution is exact at the discrete level at least every $R_{1,\cdots,m}$ iterations. For all n, it is sufficient to define n' as a multiple of $R_{1,\cdots,m}$ greater than n. □

We can then deduce from lemma 4.1 that for all $n \in \mathbb{N}$ there exists $C > 0$ such that:

$$S_h(t^n)\mathbf{V}^0 = S(t^n)\mathbf{V}^0 + C\Delta t, \quad \|V(\cdot, t^n) - V^n\|_{L^1} \leqslant \|V(\cdot, 0) - V^0\|_{L^1} + C\Delta t.$$

And from lemma 4.3 we can add that for rational eigenvalues, there exist $\tilde{n}$ such that $c^{\tilde{n}}_{k;j} = 1$ for all k in $\{1, \cdots, m\}$ and:

$$S_h(t^{\tilde{n}})\mathbf{V}^0 = S(t^{\tilde{n}})\mathbf{V}^0, \quad \|V(\cdot, t^{\tilde{n}}) - V^{\tilde{n}}\|_{L^1} = \|V(\cdot, 0) - V^0\|_{L^1}.$$

□

5 Conclusion - Nonlinear systems

We have proved for some specifical initial data, the convergence and some important features of the reservoir technique for hyperbolic equations and linear system of conservation laws. For non interacting waves solutions, we were able to prove the long time convergence of the reservoir technique under technical assumptions. In conclusion, we briefly present in this section the reservoir technique for hyperbolic systems of conservation laws. Its convergence, more technical, will be proved in a forthcoming paper. Let Ω be an open subset of $\mathbb{R}^m$ and

$f : \Omega \to \mathbb{R}^m$ be a sufficiently smooth Lipschitz function. We consider the following nonlinear system of conservation laws:

$$\partial_t V + \partial_x f(V) = 0, \qquad x \in \mathbb{R}, \qquad t > 0, \tag{15}$$

where $V = (V_1, \cdots, V_m)^T$ and $f(V) = (f_1(V), \cdots, f_m(V))^T$ and we set

$$A(V) = \left(\frac{\partial f_i}{\partial x_j}(V) \right)_{1 \leqslant i,j \leqslant m}$$

the jacobian matrix of f. We assume the system to be hyperbolic, that is $A(V)$ is diagonalizable in $\mathbb{R}$. We call

$$\lambda_1(V) \leqslant \lambda_2(V) \leqslant \cdots \leqslant \lambda_m(V)$$

the m eigenvalues of $A(V)$ and $r_k(V)$ (resp. $l_k(V)$) the right (resp. left) eigenvectors of $A(V)$ associated to $\lambda_k(V)$.

For such equations we need to define characteristics velocities for each interface $j - 1/2$ and each characteristic field k, by $\alpha^n_{k;j-1/2}$. This quantity denotes the velocity of the current wave. Typically, $\alpha^n_{k;j-1/2}$ will be given by a Riemann solver. We again set

$$\mathcal{C}^{n+1}_{k;j} = \mathcal{C}^n_{k;j} + |\alpha^n_{k;j-1/2}| \frac{\Delta t^n}{\Delta x}.$$

Namely, we assume that a Riemann solver (exact or approximate) gives us m waves with related speeds, $\alpha_k(V, W)$ and $m + 1$ states $\left(V^k_{\mathrm{R}}(V, W) \right)_k$ for any left (resp. right) state V (resp. W). More precisely, for $\alpha^n_{k;j-1/2} < 0$

$$\begin{pmatrix} \tilde{V}^{n+1}_{k;j} \\ c^{n+1}_{k;j} \\ R^{n+1}_{k;j} \end{pmatrix} = \begin{cases} \begin{pmatrix} 0 \\ c^n_{k;j} + |\alpha^n_{k;j-1/2}| \dfrac{\Delta t^n}{\Delta x} \\ R^n_{k;j} - \dfrac{\Delta t^n}{\Delta x} \big(F(V^k_{\mathrm{R}}(V^n_j, V^n_{j+1})) - F(V^{k-1}_{\mathrm{R}}(V^n_j, V^n_{j+1})) \big) \end{pmatrix}, \text{ if } \mathcal{C}^{n+1}_{k;j} < 1, \\ \begin{pmatrix} R^n_{k;j} - \dfrac{\Delta t^n}{\Delta x} \big(F(V^k_{\mathrm{R}}(V^n_j, V^n_{j+1})) - F(V^{k-1}_{\mathrm{R}}(V^n_j, V^n_{j+1})) \big) \\ 0 \\ 0 \end{pmatrix}, \text{ if } \mathcal{C}^{n+1}_{k;j} = 1, \end{cases} \tag{16}$$

whereas, for $\alpha^n_{k;j-1/2} > 0$, we upwind on the right side

$$\begin{pmatrix} \tilde{V}^{n+1}_{k;j} \\ c^{n+1}_{k;j} \\ R^{n+1}_{k;j} \end{pmatrix} = \begin{cases} \begin{pmatrix} 0 \\ c^n_{k;j} + |\alpha^n_{k;j-1/2}| \dfrac{\Delta t^n}{\Delta x} \\ R^n_{k;j} - \dfrac{\Delta t^n}{\Delta x}\big(F(V^k_{\mathrm{R}}(V^n_{j-1}, V^n_j)) - F(V^{k-1}_{\mathrm{R}}(V^n_{j-1}, V^n_j))\big) \end{pmatrix}, \text{ if } \mathcal{C}^{n+1}_{k;j} < 1, \\ \begin{pmatrix} R^n_{k;j} - \dfrac{\Delta t^n}{\Delta x}(F(V^k_{\mathrm{R}}(V^n_{j-1}, V^n_j)) - F(V^{k-1}_{\mathrm{R}}(V^n_{j-1}, V^n_j)) \\ 0 \\ 0 \end{pmatrix}, \text{ if } \mathcal{C}^{n+1}_{k;j} = 1. \end{cases} \tag{17}$$

The reservoir method just consists in computing the values $\alpha^n_{l;j-1/2}$ according to the corresponding wave speed given by the Riemann solver and the corresponding eigenvalue of the Jacobian matrix of the physical flux on left and right states. We do that in the spirit given in section 3 for each wave separately. Namely, denoting by $\lambda^n_{l;j} = \lambda_l(V^n_j)$, we define $\alpha^n_{l;j-1/2}$ as

$$\forall l \in \{1, \cdots, m\},\ \forall j \in \mathbb{Z} \begin{cases} \alpha^n_{l;j-1/2} = \lambda^n_{l;j-1} \text{ if } 0 \leqslant \lambda^n_{l;j-1} \leqslant \lambda^n_{l;j}, \text{ for right-rarefaction waves,} \\ \alpha^n_{l;j-1/2} = \lambda^n_{l;j} \text{ if } \lambda^n_{l;j-1} \leqslant \lambda^n_{l;j} \leqslant 0, \quad \text{for left-rarefaction waves,} \\ \alpha^n_{l;j-1/2} = \sigma^n_{l;j-1/2}, \text{ for shock waves, } (\lambda^n_{l;j-1} > \sigma^n_{l;j-1/2} > \lambda^n_{l;j}). \\ \text{For sonic points, see the remark.} \end{cases} \tag{18}$$

Again, the solution is updated the following manner

$$V^{n+1}_j = V^n_j + \sum_{k=1}^{m} \tilde{V}^{n+1}_{k;j}.$$

To justify the interest of such a method we give some numerical results obtained with it on classical benchmarks on Euler equations [13]. Details can be found in [3].
We will finally prove in part II:

Theorem 5.1 *The reservoir-Colella-Glaz scheme for hyperbolic systems of conservation laws is a convergent entropy scheme and is exact at the discrete level for some special configurations.*

From sections 3. and 4., we can directly state that the above result is true for shock waves solutions (on each characteristic field).

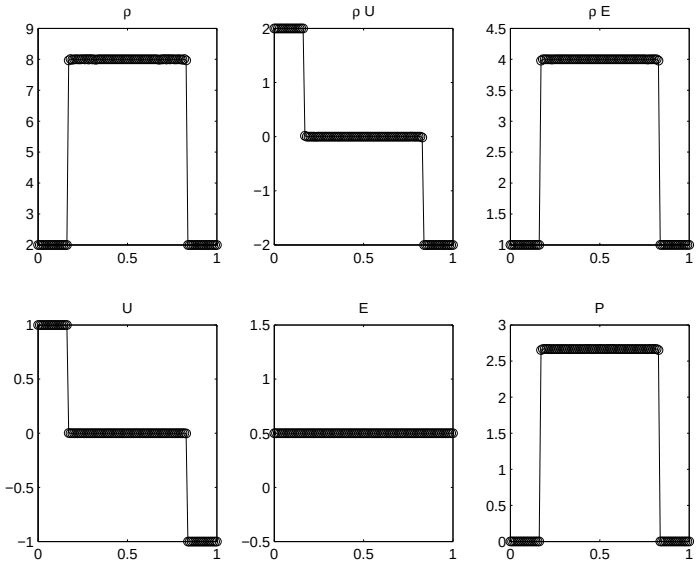

Figure 1: Noh benchmark with Colella-Glaz reservoir at final time

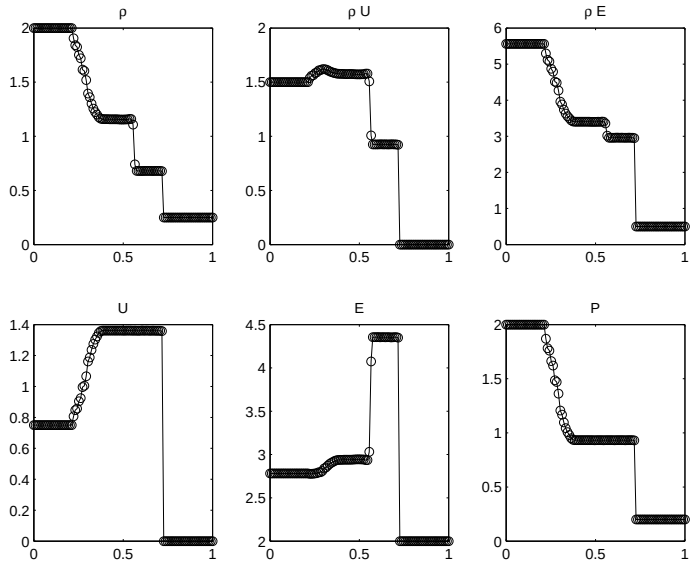

Figure 2: Sod Wendroff tube with Colella-Glaz reservoir

References

[1] F. Alouges, F. De Vuyst, G. Le Coq, and E. Lorin. The reservoir scheme for systems of conservation laws. In *Finite volumes for complex applications, III (Porquerolles, 2002)*, pages 247–254 (electronic). Lab. Anal. Topol. Probab. CNRS, Marseille, 2002.

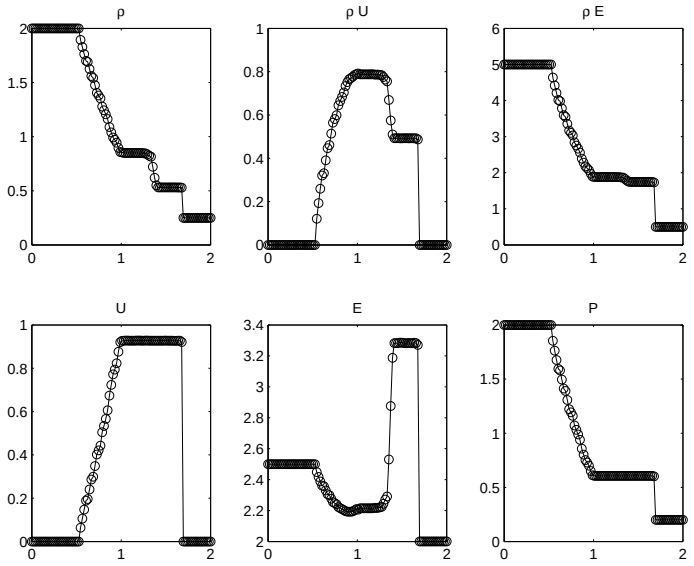

Figure 3: Sod tube with Colella-Glaz reservoir

[2] F. Alouges, F. De Vuyst, G. Le Coq, and E. Lorin. Un procédé de réduction de la diffusion numérique des schémas à différence de flux d'ordre un pour les systèmes hyperboliques non linéaires. *C. R. Math. Acad. Sci. Paris*, 335(7):627–632, 2002.

[3] F. Alouges, F. De Vuyst, G. Le Coq, and E. Lorin. The reservoir technique: a way to make godunov-type schemes zero or very low diffusive. application to colella-glaz. *In revision. Preprint CMLA 2005-16*, 2006.

[4] Le Coq Gérard Alouges, François and Emmanuel Lorin. Two-dimensional extension of the reservoir technique for some linear advection equations. *To appear in J. of Sci. Comput.*, 2006.

[5] S. J. Billett and E. F. Toro. On WAF-type schemes for multidimensional hyperbolic conservation laws. *J. Comput. Phys.*, 130(1):1–24, 1997.

[6] P. Colella and H. M. Glaz. Efficient solution algorithms for the Riemann problem for real gases. *J. Comput. Phys.*, 59(2):264–289, 1985.

[7] B. Després and F. Lagoutière. Contact discontinuity capturing schemes for linear advection and compressible gas dynamics. *J. Sci. Comput.*, 16(4):479–524 (2002), 2001.

[8] J.-M. Ghidaglia. Numerical computation of 3D two phase flow by finite volumes methods using flux schemes. In *Finite volumes for complex applications II*, pages 69–94. Hermes Sci. Publ., Paris, 1999.

[9] J.-M. Ghidaglia, A. Kumbaro, and G. Le Coq. Une méthode "volumes finis" à flux caractéristiques pour la résolution numérique des systèmes hyperboliques de lois de conservation. *C. R. Acad. Sci. Paris Sér. I Math.*, 322(10):981–988, 1996.

[10] E. Godlewski and P.-A. Raviart. *Hyperbolic systems of conservation laws*, volume 3/4 of *Mathématiques & Applications (Paris) [Mathematics and Applications]*. Ellipses, Paris, 1991.

[11] S. K. Godunov. A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Mat. Sb. (N.S.)*, 47 (89):271–306, 1959.

[12] S. Labbé. *Preprint, Orsay University: www.math.u-psud.fr*, 2006.

[13] R. J. LeVeque and H. C. Yee. A study of numerical methods for hyperbolic conservation laws with stiff source terms. *J. Comput. Phys.*, 86(1):187–210, 1990.

[14] P. L. Roe. Approximate Riemann solvers, parameter vectors, and difference schemes. *J. Comput. Phys.*, 43(2):357–372, 1981.

[15] C.-W. Shu and S. Osher. Efficient implementation of essentially nonoscillatory shock-capturing schemes. *J. Comput. Phys.*, 77(2):439–471, 1988.