Entropic sub-cell shock capturing schemes via Jin-Xin relaxation and Glimm front sampling

Frédéric Coquel[*], Shi Jin[†], Jian-Guo Liu[‡] and Li Wang[§]

May 4, 2015

Abstract

We introduce a sub-cell shock capturing method for scalar conservation laws built upon the Jin-Xin relaxation framework. Here, sub-cell shock capturing is achieved thanks to an original defect measure correction technique. The proposed correction exactly restores entropy shock solutions of the exact Riemann problem while otherwise, it produces monotone and entropy satisfying approximate self-similar solutions. These solutions are then sampled thanks to the Glimm's random choice method to advance the method in time. The resulting scheme combines the simplicity of the Jin-Xin relaxation method with the resolution of the Glimm's scheme in the capture of discontinuities. The strong benefit of using defect measure corrections over usual sub-cell shock capturing methods stays in the property that the scheme can be easily made entropy satisfying with respect to infinitely many entropy pairs. Consequently and under a classical CFL condition, the method is proved to converge to the unique entropy weak solution of the Cauchy problem for general non-linear flux functions.

AMS subject classification : 35L65, 35L67, 65M12, 76M12, 76M45
Keywords : Conservation laws, Relaxation, Defect measure, Riemann problem, Shock capturing methods, Cell entropy condition, Glimm's random choice

1 Introduction

Modern high resolution shock capturing methods for nonlinear hyperbolic systems of conservation laws contain two ingredients: building blocks (Godunov type upwind schemes

[*] CNRS and Centre de Mathématiques Appliquées UMR 7641, École Polytechnique, Route de Saclay, 91128 Palaiseau Cedex, France (frederic.coquel@cmap.polytechnique.fr)

[†] Department of Mathematics, University of Wisconsin-Madison, 480 Lincoln Drive, Madison, WI 53706, USA (jin@math.wisc.edu)

[‡] Department of Physics and Department of Mathematics, Duke University, Durham, NC 27708, USA (Jian-Guo.Liu@duke.edu)

[§] Department of Mathematics, University of California Los Angeles, 520 Portola Plaza, Los Angeles, CA 90095-1555, USA (liwang@math.ucla.edu)

based on exact or approximate Riemann solvers, Lax-Friedrichs type central schemes, kinetic schemes, etc.) [14] and reconstructions that hybridize higher order interpolations in smooth part of the solution and first order methods around discontinuities–shocks and contact discontinuities– (total-variation-diminishing (TVD), essentially-non-oscillatory (ENO) or weighted essentially-non-oscillatory (WENO), Discontinuous Galerkin, *etc*) [27], [28]. These methods have been very successfully applied to many fluid flow problems, magnetohydrodynamics, reacting flows (see [6] and the references therein). Analyses of these methods, on the other hand, are much less developed and are mostly available only for scalar problems.

In these high resolution methods, due to the use of first order methods near discontinuities, the shocks and contact discontinuities, are smeared out across few grid points. Such smearing is not an issue for most inviscid flow calculations, however, there are many problems where the smearing due to numerical viscosities can cause significant pitfalls which lead to polluted or even unphysical numerical solutions. For examples, in multiphase flows, the smeared numerical solutions across the interfaces between the two phases correspond to unphysical phases [18]; in phase transition problems, such as van der Waals flows [29], smeared solutions enter the elliptic regions which are unstable [18]; in the computation of stiff reacting flows [8], the artificially smeared temperature profiles incorrectly trigger chemical reactions which lead to unphysical detonations that propagate with incorrect speeds (see [2] for a correction based on a random projection method). Numerical viscosities are also blamed for numerical oscillations behind slowly moving shocks [15], artificial wall heating [23], and the carbuncle phenomena [25]. Indeed, it contributes to numerical instability in Lax-Friedrichs and Godunov schemes for nonlinear hyperbolic systems [1].

This paper aims at developing a one-dimensional shock capturing method that captures shocks sharply–without numerical smearing– and establishing an entropic convergence theory of this method for scalar conservation laws. The method combines the Jin-Xin relaxation approximation [16] with Dirac measure and Glimm sampling [10]. Thanks to the linear convection of the Jin-Xin relaxation, the Riemann invariants are linear which can be easily inverted and the entropy property satisfied by the scalar conservation laws can be lifted to the relaxation system. We design a specific Dirac measure which allows us to obtain the total-variation-diminishing property and cell entropy condition for both square [24] and Kružkov entropies [19]. The Glimm sampling gives a sharp shock. We refer to [13], [12] for a related sampling strategy based on Roe's approximate solvers and to [3] where a Suliciu solver for the p-system is advocated. In [4], mixed hyperbolic-elliptic Euler equations are solved within the frame of a Sulicu method but using a deterministic front tracking technique while a Glimm front sampling could have be used as well.

Here, we provide a theoretical foundation for this approach, namely the method indeed converges to the entropic solution of the scalar conservation law for general non-linear fluxes.

Numerically we only use Dirac measure and Glimm sampling near the shock. Elsewhere standard high resolution mechanism, such as higher order TVD or ENO/WENO reconstruction can still be used to offer better numerical accuracy.

There were other efforts focused on obtaining sharp shocks numerically. One is the front tracking method which relies on solving Riemann problems exactly. Within the framework of shock capturing methods, which is the approach in this paper, Harten [11] introduced the subcell method, which creates an intermediate state based on the conservation property.

However, conservation itself only guarantees the capturing of a weak solution according to the celebrated Lax-Wendroff theorem, it does not prevent the formation of entropy violating shocks. Our approach always produces entropic shocks.

Like the subcell method, our approach will also encounter major challenges when extended to nonlinear systems and higher dimensions. This will be a subject of future research.

The next section is a follow up to the Introduction. We further motivate the introduction of a Dirac measure correction to the classical numerical application of the Jin-Xin relaxation framework. For that purpose, we recall known mathematical properties of this convenient framework while introducing the required notations. We are then in a position to shade light in the design principle of the Dirac measure correction we develop hereafter in the paper. Its format is given at the end of the section.

2 Relaxation defect measures and their numerical application

We consider the Cauchy problem for a non-linear scalar conservation law

$$\begin{cases} \partial_t u + \partial_x f(u) = 0, \ t > 0, \ x \in \mathbb{R}, \\ u(t,0) = u_0(x), \end{cases} \tag{2.1}$$

supplemented with the following entropy selection principle

$$\partial_t \mathcal{U}(u) + \partial_x \mathcal{F}(u) \leq 0. \tag{2.2}$$

Here we assume a smooth flux function $f \in \mathcal{C}^2(\mathbb{R})$. Inequality (2.2) has to be satisfied in the sense of the distributions for all smooth convex functions $\mathcal{U}(u)$ with $\mathcal{F}'(u) = \mathcal{U}'(u) f'(u)$. In (2.1), the initial data u_0 is chosen in $L^\infty(\mathbb{R}) \cap \mathrm{BV}(\mathbb{R})$. It is well-known (see [26] for instance) that the Cauchy problem (2.1)–(2.2) admits a unique entropy weak solution, the so-called Kružkov solution. In [16], Jin and Xin proposed to approximate this solution by the solution of the following relaxation system

$$\begin{cases} \partial_t u^\epsilon + \partial_x v^\epsilon = 0, & \text{(2.3a)} \\ \partial_t v^\epsilon + a^2 \partial_x u^\epsilon = -\dfrac{1}{\epsilon}(v^\epsilon - f(u^\epsilon)), & \text{(2.3b)} \end{cases}$$

with well-prepared initial data

$$u(0,x) = u_0(x), \ v(0,x) = v_0(x) = f(u_0(x)). \tag{2.4}$$

Here $\epsilon > 0$ denotes a small relaxation time. For any given fixed $\epsilon > 0$, existence and uniqueness of a solution (u^ϵ, v^ϵ) can be established (see for instance [5], [22]). Under the sub-characteristic condition

$$\sup |f'(u)| < a, \tag{2.5}$$

for all the u under consideration, the sequence $\{u^\epsilon, v^\epsilon\}_{\epsilon>0}$ is shown to converge strongly as $\epsilon \to 0^+$ in $\mathcal{C}((0,\infty), L^1_{loc}(\mathbb{R}))$ to $(u, f(u))$ with u being the Kružkov solution of (2.1) (see [5],

[22] for a precise statement). In particular, this result applies to any initial data u_0 under the form

$$u_0(x) = u_L + (u_R - u_L)H(x), \quad x \in \mathbb{R}, \tag{2.6}$$

where H denotes the Heaviside function. In (2.6) the constant states u_L and u_R satisfy

$$-\sigma(u_L, u_R)(u_R - u_L) + f(u_R) - f(u_L) = 0, \tag{2.7}$$

and for all entropy pairs $(\mathcal{U}, \mathcal{F})$

$$-\sigma(u_L, u_R)\left(\mathcal{U}(u_R) - \mathcal{U}(u_L)\right) + \mathcal{F}(u_R) - \mathcal{F}(u_L) \leq 0. \tag{2.8}$$

This initial data defines a Riemann problem for (2.1) that gives rise to an entropy shock solution

$$u(t,x) = u_L + (u_R - u_L)H(x - \sigma(u_L, u_R)t), \quad t > 0, x \in \mathbb{R}. \tag{2.9}$$

Under the stability condition (2.5), the well-prepared initial data (2.4) for (2.3) built from u_0 in (2.6) thus gives rise to a family of solutions $\{(u^\epsilon, v^\epsilon)\}_{\epsilon>0}$ which converges as ϵ goes to zero to $(u, v \equiv f(u))$ where u is given by (2.9). It can be easily shown that the following limit holds in the sense of the distributions

$$\begin{aligned} \lim_{\epsilon\to 0} \frac{1}{\epsilon}(f(u^\epsilon) - v^\epsilon) &= \Big\{ -\sigma(u_L, u_R)(f(u_R) - f(u_L)) + a^2(u_R - u_L) \Big\} \delta_{x-\sigma(u_L,u_R)t} \\ &= (a^2 - \sigma^2(u_L, u_R))(u_R - u_L)\delta_{x-\sigma(u_L,u_R)t}. \end{aligned} \tag{2.10}$$

Hence the limit pair (u, v) solves again in the sense of the distributions the following system involving a measure source term :

$$\begin{cases} \partial_t u + \partial_x v = 0, \\ \partial_t v + a^2 \partial_x u = (a^2 - \sigma^2(u_L, u_R))(u_R - u_L)\delta_{x-\sigma(u_L,u_R)t}, \end{cases} \tag{2.11}$$

with initial data :

$$u_0(x) = u_L + (u_R - u_L)H(x), \quad v_0(x) := f(u_0(x)) = f(u_L) + (f(u_R) - f(u_L))H(x), \quad x \in \mathbb{R}, \tag{2.12}$$

where we have used the definition of the Heaviside function in the last equality. In the sequel, the measure source term entering (2.11) is referred to as a *relaxation defect measure*.

At this level, it is crucial to observe that despite the Cauchy problem (2.3) with the Riemann data (2.12) does not admit a self-similar solution (u^ϵ, v^ϵ) for any given fixed $\epsilon > 0$, the limit PDE model (2.11) does by contrast admit the self similar solution

$$\begin{aligned} u(t,x) &= u_L + (u_R - u_L)H(x - \sigma(u_L, u_R)t), \quad t > 0,\ x \in \mathbb{R}, \\ v(t,x) &= f(u_L) + (f(u_R) - f(u_L))H(x - \sigma(u_L, u_R)t), \end{aligned} \tag{2.13}$$

where the u-component is nothing but the entropy satisfying shock solution (2.9) of (2.1)-(2.6).

With this in mind, let us briefly revisit the widely used application of the relaxation model (2.3) to the numerical approximation of the Kružkov solution of (2.1). An operator splitting strategy is generally promoted to circumvent the lack of self-similar solutions. Covering $\mathbb{R}_t^+$

by a collection of small time steps, one thus solves in each time step first the homogeneous Cauchy problem

$$\begin{cases} \partial_t u + \partial_x v = 0, & \text{(2.14a)} \\ \partial_t v + a^2 \partial_x u = 0, & \text{(2.14b)} \end{cases}$$

with appropriate initial data, and then the following singular ODE problem

$$d_t u^\epsilon = 0, \quad d_t v^\epsilon = -\frac{1}{\epsilon}(v^\epsilon - f(u^\epsilon)), \quad \text{in the limit } \epsilon \to 0+, \tag{2.15}$$

again with appropriate data. Of course, the first step now allows for self-similar solutions. Such solutions are just made of a single intermediate state $(u^\star, v^\star)$ separated by two discontinuities propagating with speed $-a$ and $+a$ respectively. But this step yields a poor resolution of the shock solutions of the original conservation law (2.1). Actually and under the mandatory stability condition (2.5), it can be seen [14] that the intermediate value $u^\star$ in the self-similar solution of (2.14)–(2.12) coincides with the space averaging of the exact self-similar solution (2.9) whose fan is bordered by the two waves $-a$ and $+a$. Put in other words, exact shock waves are inherently smeared from the very first step of the procedure. The rough splitting performed on the relaxation PDEs (2.3) is responsible for that. Too little from the relaxation mechanism has been accounted for in the first step. In order to involve those mechanisms in a deeper manner at the first step, we again argue of the limit of the singular source term $(f(u^\epsilon) - v^\epsilon)/\epsilon$ in the limit $\epsilon \to 0^+$. Formally speaking and for general well-prepared initial data (2.4), the limit under consideration can be split in two contributions. A first singular part is a Radon measure $\mathcal{M}_{t,x}$, made of the sum of all the relaxation defect measures concentrated on the shocks in the limit solution $u(t,x)$. A second smooth contribution comes from the smooth part of the Kružkov solution and reads $\partial_t f(u) + a^2 \partial_x u$. Motivated by this natural decomposition, we propose a new splitting procedure involving in the first step the singular first part $\mathcal{M}_{t,,x}$ while the second step is devoted to handle the smooth second part. The first step then consists in solving

$$\begin{cases} \partial_t u + \partial_x v = 0, & \text{(2.16a)} \\ \partial_t v + a^2 \partial_x u = \mathcal{M}_{t,x}. & \text{(2.16b)} \end{cases}$$

and then

$$d_t u^\epsilon = 0, \quad d_t v^\epsilon = -\frac{1}{\epsilon}(v^\epsilon - f(u^\epsilon)), \quad \text{in the limit } \epsilon \to 0+, \tag{2.17}$$

with appropriate initial data. Let us stress that the second step cannot develop relaxation defect measures. With this respect, we really have performed a consistent splitting of the PDE model (2.3) in the limit $\epsilon \to 0^+$. Then we underline that the Cauchy problem (2.16) can be solved by a succession of non interacting Riemann problems of the form (2.11) , once the Radon measure $\mathcal{M}_{t,x}$ is conveniently approximated.

Let us now describe the main building principle for relevant approximations of $\mathcal{M}_{t,x}$. Approximations are to be performed locally for Riemann problems under the generic form

$$\begin{cases} \partial_t u + \partial_x v = 0, & \text{(2.18a)} \\ \partial_t v + a^2 \partial_x u = m(u_L, u_R)\delta_{x-\sigma(u_L,u_R)t}, & \text{(2.18b)} \end{cases}$$

with well-prepared initial data

$$u(0,x)=u_0(x)=\left\{\begin{array}{l} u_L,\ x<0,\\ u_R,\ x>0,\end{array}\right. \qquad v(0,x)=v_0(x)=\left\{\begin{array}{l} f(u_L),\ x<0,\\ f(u_R),\ x>0.\end{array}\right. \tag{2.19}$$

In (2.18), $m(u_L,u_R)$ refers to the mass of a Dirac measure concentrated at $x=\sigma(u_L,u_R)t$ where $\sigma(u_L,u_R)$ plays the role of a velocity to be defined under the natural condition

$$|\sigma(u_L,u_R)|<a. \tag{2.20}$$

Both the mass m and velocity σ are to be defined depending on the states u_L,u_R to meet suitable properties in the solution of the Cauchy problem (2.18)–(2.19). But whatever the precise definitions are, the solution we seek for is clearly self-similar. To condense the notations, $\mathbb{U}=(u,v)^T\in\mathbb{R}^2$ refers to the unknown in (2.18). The case of an identically zero mass $m(u_L,u_R)=0$ boils down to a Riemann solution for the 2×2 homogeneous linear system (2.14). It is generically made of three constant states $\mathbb{U}_L$, $\mathbb{U}^\star$ and $\mathbb{U}_R$ separated by two waves propagating with speed $-a$ and $+a$ respectively. For a non-zero mass, easy considerations on the weak form of the PDEs (2.18) reveal the existence of an intermediate discontinuity propagating with speed $\sigma(u_L,u_R)$ (see indeed the condition (2.20)) which separates, in the wave span, two inner states denoted $\mathbb{U}_L^\star$ and $\mathbb{U}_R^\star$. Across the intermediate discontinuity, these two states have to satisfy the following jump conditions :

$$\begin{array}{ll} -\sigma(u_L,u_R)(u_R^\star-u_L^\star)+(v_R^\star-v_L^\star) & =0,\\ -\sigma(u_L,u_R)(v_R^\star-v_L^\star)+a^2(u_R^\star-u_L^\star) & =m(u_L,u_R).\end{array} \tag{2.21}$$

We propose to define the mass $m(u_L,u_R)$ and the velocity $\sigma(u_L,u_R)$ in order to preserve some of the essential properties of the exact solution of the Riemann problem

$$\left\{\begin{array}{l} \partial_t u+\partial_x f(u)=0,\\ u(0,x)=\left\{\begin{array}{l} u_L,\ x<0,\\ u_R,\ x>0.\end{array}\right.\end{array}\right. \tag{2.22}$$

Properties to be preserved include the monotonicity property in the self-similar variable $\xi=x/t$, some consistency requirement with the entropy inequalities (2.2) and an exactness property regarding discontinuous solutions of (2.22). For pairs of states (u_L,u_R) satisfying (2.7)–(2.8), the exactness requirement we propose amounts to define the mass $m(u_L,u_R)$ and velocity $\sigma(u_L,u_R)$ in the Cauchy problem (2.18)–(2.19) so that its self-similar solution $\mathbb{U}(\xi,u_L,u_R)$ reduces component-wise to (2.13). Doing so, we naturally recover the definition of the relaxation defect measure stated in (2.11).

Lemma 1. *Given any pair of states (u_L,u_R) verifying the Rankine-Hugoniot condition (2.7) and the entropy inequalities (2.8), define the velocity*

$$\sigma(u_L,u_R)=\frac{f(u_L)-f(u_R)}{u_L-u_R},\ \ u_L\neq u_R;\quad \sigma(u_L,u_R)=f'(u_L)=f'(u_R),\ \ \textit{otherwise}, \tag{2.23}$$

and the mass

$$m_s(u_L,u_R)=\left(a^2-\sigma^2(u_L,u_R)\right)(u_R-u_L). \tag{2.24}$$

Then the $\mathbb{U}(\xi,u_L,u_R)$ of the Riemann problem (2.18)–(2.19) is given by the self-similar function (2.13).

Proof. One has to prove to that the self-similar function (2.13) is a solution of the Riemann problem (2.18)–(2.19) with intermediate states $\mathbb{U}_L^\star = \mathbb{U}_L$ and $\mathbb{U}_R^\star = \mathbb{U}_R$ as soon as the velocity and mass are prescribed according to (2.23) and (2.24). In such a case the jumps in the waves with speed $-a$ and $+a$ have to be trivial. One thus has to ckeck that the jump conditions (2.21) across the intermediate wave are satisfied with $\mathbb{U}_L^\star = \mathbb{U}_L$ and $\mathbb{U}_R^\star = \mathbb{U}_R$. But, clearly

$$-\sigma(u_L,u_R)(u_R^\star - u_L^\star) + (v_R^\star - v_L^\star) = -\sigma(u_L,u_R)(u_R - u_L) + f(u_R) - f(u_L) = 0, \quad (2.25)$$

by the definition of $\sigma(u_L,u_R)$ while the mass $m_s(u_L,u_R)$ has to prescribed to meet the identity :

$$-\sigma(u_L,u_R)(v_R^\star - v_L^\star) + a^2(u_R^\star - u_L^\star) = -\sigma^2(u_L,u_R)(u_R - u_L) + a^2(u_R - u_L) := m_s(u_L,u_R). \quad (2.26)$$

This concludes the proof. □

Obviously the property that the pair (u_L,u_R) under consideration verifies the entropy condition(s) stated in (2.8) plays no role in the fact that (2.13) actually solves (2.18)–(2.19). In the definition of the mass $m_s(u_L,u_R)$, only the satisfaction of the jump condition (2.7) matters. But in the present scalar setting, it is always possible to define the velocity $\sigma(u_L,u_R)$ so as to meet (2.7) for any given pair of states (u_L,u_R). Then choosing (2.24) to define $m(u_L,u_R)$ for arbitrary pairs would systematically result in a solution given by (2.13). In other words, we would merely end up with a Roe scheme which is known [14] to be entropy violating in the approximation of the solutions of (2.22).

We thus propose to modulate the definition of the mass in (2.24) by looking for a monitoring factor $\theta(u_L,u_R)$, namely a suitable real valued mapping $\theta : (u_L,u_R) \in \mathbb{R}^2 \to \theta(u_L,u_R) \in \mathbb{R}$, to define

$$m(u_L,u_R) = \theta(u_L,u_R)\left(a^2 - \sigma^2(u_L,u_R)\right)(u_R - u_L), \quad (2.27)$$

with $\sigma(u_L,u_R)$ given by (2.23). Clearly, $\theta(u_L,u_R)$ acts as an anti-diffusive parameter, allowing for a continuous shift from the Lax Friedrichs scheme when $\theta(u_L,u_R) = 0$ to the Roe scheme for $\theta(u_L,u_R) = 1$.

2.1 Design principle of approximate defect measures

Let us briefly put forward the design principle we propose for relevant anti-diffusive laws $\theta(u_L,u_R)$.

In that aim, let us first consider the case of a strictly convex flux function $f(u)$. A first obvious choice for the anti-diffusive law $\theta(u_L,u_R)$ would be :

$$\theta(u_L,u_R) = \begin{cases} 1, & u_L > u_R, \\ 0, & \text{otherwise}, \end{cases} \quad (2.28)$$

since the situation $u_L > u_R$ yields an entropy satisfying shock solution, while the converse gives rise to a rarefaction. Actually we will prove that more anti-diffuse choices for $\theta(u_L,u_R)$ can be performed while still allowing for convergence to the Kružkov solution. In particular, we will prove that $\theta(u_L,u_R)$ can be set close to 1 (at the order $\mathcal{O}(\Delta x)$ with $\Delta x > 0$ the space step) in the smooth part of the approximate solution. Hence rarefaction waves in the discrete solution can be handled with values asymptotically close to 1 and not to 0 as advocated in (2.28). The derivation of relevant anti-diffusive laws θ essentially relies on a consistency requirement with the entropy inequalities (2.2). In the case of a genuinely non-linear flux $f(u)$, it is known after Panov [24] that a single strictly convex entropy pair suffices to select the Kružkov solution of (2.1). θ-laws are derived accordingly on the ground of a single entropy pair.

The situation of a general non-linear flux function is more involved. First and clearly, the obvious choice (2.28) no longer applies. Then, infinitely many entropy pairs are required to single out the Kružkov solution. We are thus led to design θ-laws accordingly when asking consistency with all the Kružkov entropy pairs. We thus have to handle infinitely many entropy pairs in the design of the anti-diffusive laws.

Our consistency condition with the entropy inequalities (2.2) is built from the relaxation entropy pairs associated with the Jin-Xin's model (2.18). As established in [5] (see also [22]), any given smooth convex entropy pair $(\mathcal{U}, \mathcal{F})$ for (2.1) can be suitably lift to give rise to a relaxation entropy pair for (2.18), we denote (Φ, Ψ) in the sequel. Under the sub-characteristic condition (2.5), the relaxation mechanism in (2.18) can be shown to be dissipative with respect to any of those relaxation entropy pairs. More precisely, an invariant domain for the solutions of (2.18) may be built from (2.5). Convexity and dissipative properties for any pair (Φ, Ψ) are proved to hold true within that invariant domain. These crucial properties are generically lost outside of the invariant domain. It is thus of central importance to keep invariant the aforementioned domain for the solution of the Riemann problems (2.18) involving an approximation of the exact defect measure $\mathcal{M}_{t,x}$ in the limiting PDE model (2.16). This requirement will be fairly easily achieved from the choice (2.27), allowing us in turn to enforce consistency with the entropy inequalities (2.2).

2.2 Organization of the paper

In section 3, the Riemann problem (2.18)–(2.19) with a defect measure correction given by (2.27) is solved for a general pair of states (u_L, u_R). A central property due to the choice (2.27) is then revealed in the characteristic variables $(v - av\ , u + av)$. Indeed, this property allows to prove in section 4 that preserving the entropy invariant domain for self-similar solutions of (2.18) is achieved by requiring the anti-diffusive law $\theta(u_L, u_R)$ to take values in $[0, 1]$. Equivalence of the reported invariance property with a monotonicity property for the u-component of the solution (2.18) is then established. As a consequence, uniform sup-norm and BV estimates are inferred, allowing us to prove the convergence of the Jin-Xin relaxation solver with defect measure correction to a weak solution of (2.1). To enforce the entropy condition with the expected Kružkov solution, we then ask in subsection 4.2 that the discontinuity induced by the approximate defect measure in (2.18) is entropy satisfying with respect to relaxation entropy pairs (Φ, Ψ). This requirement further reduces the admissible graph of relevant anti-dissipative law $\theta : (u_L, u_R) \in \mathbb{R}^2 \to [0, \Theta(u_L, u_R)]$, where the positive real number $\Theta(u_L, u_R)$ denotes some optimal upper-bound. $\Theta(u_L, u_R)$ is in general strictly less than 1 for arbitrary pairs (u_L, u_R) but turns to be equal to 1 for the pairs satisfying (2.8). In other words, exact capture of entropy shock solutions is thus assured. In subsection 4.4, analysis is performed for a strictly convex flux on the ground of a single entropy inequality. Analysis is extended in subsection 4.4 to general flux functions, involving the whole Kružkov entropy pairs family. In both settings, the optimal upper-bound $\Theta(u_L, u_R)$ is given an explicit form whose evaluation turns to be fairly simple. In section 5, two numerical methods for approximating the Kružkov solution of (2.1)–(2.2) are introduced. The convergence of the corresponding families of approximate solutions is established in section 6. At last, we propose some numerical results to assess the central importance of designing optimal anti-diffusive law $\Theta(u_L, u_R)$ according to infinitely many entropy pairs in the frame of a flux function without genuine nonlinearities.

3 Relaxation Riemann problem with defect measure correction

Consider a pair of real numbers (u_L, u_R) and a constant positive velocity a prescribed under the sub-characteristic condition

$$\sup_{u\in\lfloor u_L,u_R\rceil} |f'(u)| < a, \tag{3.1}$$

where for any given pair of real numbers (a, b), $\lfloor a, b\rceil$ denotes in the sequel the interval $[\min(a, b), \max(a, b)]$. Motivated by the Introduction, we first give the precise form of the self-similar solution of the Riemann problem (2.18)–(2.19).

Proposition 2. *For a general pair of states (u_L, u_R), define the velocity $\sigma(u_L, u_R)$ according to (2.23) and consider a mass $m(u_L, u_R)$ under the form (2.27) for some given mapping $\theta : (u_L, u_R) \in \mathbb{R}^2 \to \theta(u_L, u_R) \in \mathbb{R}$. Then the solution $\mathbb{U}(.; u_L, u_R)$ of the Riemann problem (2.18)–(2.19) is generically made of four constants states $\mathbb{U}_L$, $\mathbb{U}^\star_L(\theta; u_L, u_R)$, $\mathbb{U}^\star_R(\theta; u_L, u_R)$ and $\mathbb{U}_R$ separated by three discontinuities propagating with speed $-a$, $\sigma(u_L, u_R)$ and $+a$ respectively. Defining*

$$u^\star = \frac{1}{2}(u_L + u_R) - \frac{1}{2a}\big(f(u_R) - f(u_L)\big), \quad v^\star = \frac{1}{2}\big(f(u_R) + f(u_L)\big) - \frac{a}{2}(u_R - u_L), \tag{3.2}$$

the intermediate state $\mathbb{U}^\star_L(\theta; u_L, u_R)$ reads component wise

$$\begin{aligned} u^\star_L(\theta; u_L, u_R) &= u^\star - \tfrac{1}{2a}\theta(u_L, u_R)(a - \sigma(u_L, u_R))(u_R - u_L), \\ v^\star_L(\theta; u_L, u_R) &= v^\star + \tfrac{1}{2}\theta(u_L, u_R)(a - \sigma(u_L, u_R))(u_R - u_L), \end{aligned} \tag{3.3}$$

while $\mathbb{U}^\star_R(\theta; u_L, u_R)$ is given by

$$\begin{aligned} u^\star_R(\theta; u_L, u_R) &= u^\star + \tfrac{1}{2a}\theta(u_L, u_R)(a + \sigma(u_L, u_R))(u_R - u_L), \\ v^\star_R(\theta; u_L, u_R) &= v^\star + \tfrac{1}{2}\theta(u_L, u_R)(a + \sigma(u_L, u_R))(u_R - u_L). \end{aligned} \tag{3.4}$$

Proof. One has to determine each of the two components in the intermediate states $\mathbb{U}_L^\star(\theta; u_L, u_R)$ and $\mathbb{U}_R^\star(\theta; u_L, u_R)$. The two jump conditions (2.21) at the intermediate discontinuity are supplemented with Rankine-Hugoniot relations for the waves propagating with speed $-a$ and $+a$ respectively

$$a(u_L^\star - u_L) + (v_L^\star - f(u_L)) = 0, \quad -a(u_R - u_R^\star) + (f(u_R) - v_R^\star) = 0. \tag{3.5}$$

The resulting 4×4 linear system is easily seen to be uniquely solvable for any given mass m provided that $|\sigma(u_L, u_R)| \neq a$, which holds in view of the sub-characteristic condition (2.20) satisfied with the choice (2.23). With little abuse in the notations, the components of the intermediate states expressed for a general value of the mass read

$$\begin{aligned} u_L^\star(m) &= u^\star - \tfrac{m}{2a(a+\sigma)}, & v_L^\star(m) &= v^\star + \tfrac{m}{2(a+\sigma)}, \\ u_R^\star(m) &= u^\star + \tfrac{m}{2a(a-\sigma)}, & v_R^\star(m) &= v^\star + \tfrac{m}{2(a-\sigma)}, \end{aligned} \tag{3.6}$$

with $u^\star$ and $v^\star$ given in (3.2). The required expressions (3.3)–(3.4) readily follow plugging the particular form (2.27) for the mass under consideration. □

Observe that the state $\mathbb{U}^\star \equiv (u^\star, v^\star)$ defined from (3.2) is nothing but the intermediate state involved in the classical solution for the homogeneous Riemann problem (2.18)–(2.19), *i.e.* with $m(u_L, u_R) = 0$. Let us then underline that the proposed formulas for the intermediate states $\mathbb{U}_L^\star(\theta; u_L, u_R)$ and $\mathbb{U}_R^\star(\theta; u_L, u_R)$ are well-behaved if the mapping $\theta(u_L, u_R)$ stays bounded for all pairs of states (u_L, u_R) in $\mathbb{R}^2$. The next sections devoted to the derivation of monitoring weight functions $\theta(u_L, u_R)$ will show that relevant mappings naturally keep their values in the interval $[0, 1]$.

A central property due to the choice (2.27) for defining the mass $m(u_L, u_R)$ is revealed when re-formulating the two intermediate states thanks to the characteristic variables

$$r^\pm = v \pm au. \tag{3.7}$$

Corollary 3. *Under the assumptions of Proposition 2, let us re-express the intermediate states $\mathbb{U}_L^\star(\theta; u_L, u_R)$ and $\mathbb{U}_R^\star(\theta; u_L, u_R)$ in the characteristic variables*

$$r_L^{\pm\star}(\theta) = v_L^\star(\theta; u_L, u_R) \pm a u_L^\star(\theta; u_L, u_R), \quad r_R^{\pm\star}(\theta) = v_R^\star(\theta; u_L, u_R) \pm a u_R^\star(\theta; u_L, u_R). \tag{3.8}$$

Then $r_L^{-\star}(\theta)$ and $r_R^{+\star}(\theta)$ can be equivalently rewritten as linear combinations in θ of $r_L^\pm = f(u_L) \pm au_L$ and $r_R^\pm = f(u_R) \pm au_R$ according to

$$\begin{aligned} r_L^{-\star}(\theta) &= \theta(u_L, u_R) r_L^- + \big(1 - \theta(u_L, u_R)\big) r_R^-, \\ r_R^{+\star}(\theta) &= \big(1 - \theta(u_L, u_R)\big) r_L^+ + \theta(u_L, u_R) r_R^+, \end{aligned} \tag{3.9}$$

while we have

$$r_R^{-\star}(\theta) = r_R^-, \quad r_L^{+\star}(\theta) = r_L^+. \tag{3.10}$$

Proof. Denoting $r^{+\star} = v^\star + au^\star$ with $u^\star$, $v^\star$ defined in (3.2) yields for any given value of θ the following identities

$$r_L^{+\star}(\theta) = r^{+\star}, \quad r_R^{+\star}(\theta) = r^{+\star} + \theta(a+\sigma)(u_R - u_L). \tag{3.11}$$

But it is easily seen that $r^{+\star} = f(u_L) + au_L = r_L^+$ so that we have $r_L^{+\star}(\theta) = r_L^+$ and $r_R^{+\star}(0) = r^{+\star} = r_L^+$ while $r_R^{+\star}(1) = f(u_L) + au_L + a(u_R - u_L) + (f(u_R) - f(u_L)) = r_R^+$. Because (3.11) is linear in θ, we readily infer

$$r_R^{+\star}(\theta) = (1-\theta)r_R^{+\star}(0) + \theta r_R^{+\star}(1) = (1-\theta)r_L^+ + \theta r_R^+. \tag{3.12}$$

The first linear combination in (3.9) follows similarly, noticing successively that $r_R^{-\star}(\theta) = r^{-\star} = r_R^-$, $r_L^{-\star}(\theta) = r^{-\star} + \theta(a-\sigma)(u_R - u_L)$ with $r_L^{-\star}(0) = r_R^-$ and $r_L^{-\star}(1) = r_L^-$. □

As already claimed, relevant mappings $\theta(u_L, u_R)$ will be seen to keep values in the interval $[0,1]$. With this respect, the linear combinations stated in (3.9) are nothing but convex decompositions of the left and right data expressed in terms of the characteristic variables (3.7). Such convex decompositions will be crucial in the design of relevant mappings $\theta(u_L, u_R)$ according to forthcoming consistency conditions with the entropy inequalities (2.2).

4 Design of non-linearly stable mappings $\theta(u_L, u_R)$

This section studies the monitoring mapping $\theta(u_L, u_R)$ in (2.27) for general pairs of states (u_L, u_R) so that the solution of the Riemann problem with defect measure correction (2.18)–(2.19) obeys linear and non-linear stability properties.

4.1 Monotonicity preservation

The main result of this paragraph is

Proposition 4. *For a general pair of states (u_L, u_R), define the velocity $\sigma(u_L, u_R)$ according to (2.23) and consider a mass $m(u_L, u_R)$ under the form (2.27). Then under the sub-characteristic condition (3.1), the u-component of the Riemann solution $\mathbb{U}(.; u_L, u_R)$ of the problem (2.18)–(2.19) satisfies the following monotonicity preserving properties*

$$TV(u(\cdot; u_L, u_R)) = |u_R - u_L| \tag{4.1}$$

if and only if

$$0 \le \theta(u_L, u_R) \le 1. \tag{4.2}$$

As a consequence and under (4.2), we have

$$\min(u_L, u_R) \le u(\cdot; u_L, u_R) \le \max(u_L, u_R), \tag{4.3}$$

furthermore :

$$|v(\cdot, u_L, u_R)| \le a \max(|u_L|, |u_R|), \quad TV\big(v(\cdot; u_L, u_R)\big) \le a|u_L - u_R|. \tag{4.4}$$

Proof. We will consider the case $u_L < u_R$, the reverse situation follows similar steps. If $\theta(u_L, u_R)$ is defined so that the following ordering is valid

$$u_L \leq u_L^\star(\theta; u_L, u_R) \leq u_R^\star(\theta; u_L, u_R) \leq u_R, \tag{4.5}$$

then the total variation estimate stated in (4.1) is guaranteed but is clearly violated otherwise. The maximum principle in (4.1) is also valid provided that (4.5) holds. Using the definitions of these intermediate values in (3.3)-(3.4) and $u^\star$ given in (3.2), one gets

$$\begin{array}{ll} u_L^\star(\theta) - u_L & = (u^\star - u_L) - \theta(u_L, u_R)\frac{a-\sigma}{2a}(u_R - u_L) = \big(1 - \theta(u_L, u_R)\big)\frac{a-\sigma}{2a}(u_R - u_L), \\ u_R^\star(\theta) - u_L^\star(\theta) & = \theta(u_L, u_R)(u_R - u_L), \\ u_R - u_R^\star(\theta) & = (u_R - u^\star) - \theta(u_L, u_R)\frac{a+\sigma}{2a}(u_R - u_L) = \big(1 - \theta(u_L, u_R)\big)\frac{a+\sigma}{2a}(u_R - u_L), \end{array} \tag{4.6}$$

so that under the sub-characteristic condition (2.20) inherited from (3.1), the proposed ordering (4.5) holds true if and only if the weight $\theta(u_L, u_R)$ verifies (4.2). Regarding the sup-norm estimate for the second component $v(.; u_L, u_R)$, we notice that

$$\text{sign}(u_R - u_L)v_L^\star(\theta) = \text{sgn}(u_R - u_L)v^\star + \frac{\theta}{2}(a - \sigma)|u_R - u_L|,$$

which is an increasing function of $\theta \in [0, 1]$:

$$\begin{array}{rl} \text{sign}(u_R - u_L)v^\star \leq \text{sign}(u_R - u_L)v_L^\star(\theta) \leq & \text{sign}(u_R - u_L)\big(v^\star + \frac{1}{2}(a - \sigma)(u_R - u_L)\big), \\ & = \text{sign}(u_R - u_L)f(u_L). \end{array} \tag{4.7}$$

Assuming without loss of generality $f(0) = 0$, we infer under the sub-characteristic condition (3.1)

$$|v_L^\star(\theta)| \leq \max(|f(u_L)|, |f(u_R)|) \leq a \max(|u_L|, |u_R|). \tag{4.8}$$

The same estimate holds true for $|v_R^\star(\theta)|$. At last, the total variation of $v(.; u_L, u_R)$ reads

$$\begin{array}{ll} \text{TV}\big(v(\cdot; u_L, u_R)\big) & = |v_L^\star(\theta) - v_L| + |v_R^\star(\theta) - v_L^\star(\theta)| + |v_R - v_R^\star(\theta)| \\ & = \big((1 - \theta)a + \theta|\sigma|\big)|u_R - u_L| \\ & \leq a|u_R - u_L|, \end{array}$$

where we have used the definitions of the intermediate states (3.2)–(3.4) and the sub-characteristic condition (2.20). □

As is well-known, the solution $u(.; u_L, u_R)$ of the Riemann problem (2.22)–(2.2) satisfies the *a priori* estimates (4.1). It seems thus natural to require the u-component of $\mathbb{U}(.; u_L, u_R)$ to satisfy the same estimates. This in turn equivalently asks $\theta(u_L, u_R)$ to satisfy the condition (4.2). The corresponding mappings will be referred as to *monotonicity preserving* in the sequel. Let us stress that the condition (4.2) actually implies a stronger property for $\mathbb{U}(.; u_L, u_R)$. To clarify this issue, it is convenient to adopt a slightly broader standpoint. After Chen-Levermore and Liu [5], Natalini [22], define the following two functions

$$h_\pm(u) = f(u) \pm au, \quad u \in \lfloor u_L, u_R \rceil, \tag{4.9}$$

and consider the compact intervals $\mathcal{I}_- = h_-(\lfloor u_L, u_R \rceil)$ and $\mathcal{I}_+ = h_+(\lfloor u_L, u_R \rceil)$. Under the sub-characteristic condition (3.1), the inverse functions $h_\pm^{-1} : r \in \mathcal{I}_\pm \to h_\pm^{-1}(r) \in \lfloor u_L, u_R \rceil$ are well-defined with the property that h_+^{-1} (respectively h_-^{-1}) is increasing (resp. decreasing)

$$\frac{d}{dr} h_+^{-1}(r) = \frac{1}{a + f'(h_+^{-1}(r))} > 0, \quad \frac{d}{dr} h_-^{-1}(r) = -\frac{1}{a - f'(h_-^{-1}(r))} < 0. \tag{4.10}$$

Equipped with these notations, let us built the following compact domain of $\mathbb{R}^2$ from the interval $\lfloor u_L, u_R \rceil$

$$\mathcal{D}(\lfloor u_L, u_R \rceil) \equiv \{\mathbb{U} = (u, v) \in \mathbb{R}^2;\ r_-(\mathbb{U}) = v - au \in \mathcal{I}_- \text{ and } r_+(\mathbb{U}) = v + au \in \mathcal{I}_+\}. \tag{4.11}$$

Of central importance in the sequel, the domain (4.11) can be shown to stay invariant by the Jin-Xin relaxation model under the sub-characteristic condition (3.1) (see [5], [22]). Namely given a well-prepared initial data $\mathbb{U}_0 = (u_0, v_0 = f(u_0))$ with $u_0(x) \in \lfloor u_L, u_R \rceil$ for a.e. $x \in \mathbb{R}$,then for any given relaxation time $\epsilon > 0$, the unique solution $\mathbb{U}^\epsilon$ of the relaxation Cauchy problem stays in $\mathcal{D}(\lfloor u_L, u_R \rceil)$. In particular, the solution of Riemann problem of the homogeneous system (2.18), *i.e.* with $m(u_L, u_R) = 0$, satisfies this invariance property. The reported invariance property turns to be crucial in the dissipative convex lift of the convex entropy pairs $(\mathcal{U}, \mathcal{F})$ for (2.22). The following consequence of the condition (4.2) shows that such a property extends to the corresponding solutions of the Riemann problem of the system with defect measure correction.

Corollary 5. *Given a pair of states (u_L, u_R), assume the sub-characteristic condition (3.1). Then the solution $\mathbb{U}(., u_L, u_R)$ of (2.18)–(2.19) for a given $\theta(u_L, u_R)$ stays in $\mathcal{D}(\lfloor u_L, u_R \rceil)$ (4.11) if and only if the monotonicity preserving condition (4.2) is satisfied.*

Proof. This statement is a direct consequence of Corollary 3. Indeed and under the sub-characteristic condition (3.1), r_L^-, r_R^- (respectively r_L^+, r_R^+) are nothing but the boundaries of the interval $\mathcal{I}_-$ (resp. $\mathcal{I}_+$) in view of the monotonicity properties (4.10). Keeping the domain $\mathcal{D}(\lfloor u_L, u_R \rceil)$ invariant is thus equivalent to require that the characteristic variables $r_L^{-\star}(\theta; u_L, u_R)$, $r_R^{-\star}(\theta; u_L, u_R)$ are convex combinations of these two boundaries (resp. $r_L^{+\star}(\theta; u_L, u_R)$, $r_R^{+\star}(\theta; u_L, u_R)$). According to (3.9), such a property is met if and only if the monotonicity preserving condition (4.2) holds true. □

4.2 Entropy consistency requirements

In this section, we propose and analyze entropy-like conditions to further limit the graph of monotonicity preserving mappings $\theta(u_L, u_R)$. The proposed limitation has to permit the expected value $\theta(u_L, u_R) = 1$ for pairs of states (u_L, u_R) that satisfy entropy inequalities (2.8). As already underlined, the entropy consistency condition we promote will concern a single entropy pair in the case of a genuinely non-linear flux function $f(u)$, corresponding to the choice

$$\mathcal{U}(u) = \frac{u^2}{2}, \quad \mathcal{F}(u) = \int_0^u v f'(v) dv, \tag{4.12}$$

while asking consistency with the Kružkov family of entropy pairs in the case of a general non-linear flux

$$\mathcal{U}_k = |u - k|, \quad \mathcal{F}_k(u) = \operatorname{sign}(u - k)\big(f(u) - f(k)\big), \quad k \in \mathbb{R}. \tag{4.13}$$

Our entropy consistency requirement relies on the extension proposed in [5],[22] of entropy pairs for the Jin-Xin relaxation system from convex entropy pairs of the scalar conservation law (2.22). Their design principle is of importance hereafter and we briefly revisit it. Given any interval of the form $\lfloor u_L, u_R \rceil$, the proposed extension is performed over the compact domain $\mathcal{D}(\lfloor u_L, u_R \rceil)$ defined in (4.11). In [5], [22], suitable properties for the proposed lift actually follow from the invariance property of $\mathcal{D}(\lfloor u_L, u_R \rceil)$ under the sub-characteristic condition (3.1). Such an invariance property indeed guarantees the monotonicity properties (4.10) of the functions $h^{\pm}$ defined in (4.9) for states $\mathbb{U}$ in $\mathcal{D}(\lfloor u_L, u_R \rceil)$. Let us stress that in the present setting, those properties are equivalently preserved under the monotonicity preserving condition (4.2) as put forward in Corollary 5. Given an entropy pair $(\mathcal{U}, \mathcal{F})$ for the scalar law (2.22), one seeks for an entropy pair (Φ, Ψ) for the Jin-Xin relaxation equations which is well defined over $\mathcal{D}(\lfloor u_L, u_R \rceil)$ and which coincides with $(\mathcal{U}, \mathcal{F})$ at equilibrium, namely

$$\Phi(u, f(u)) = \mathcal{U}(u), \;\; \Psi(u, f(u)) = \mathcal{F}(u), \quad \text{for all } u \in \lfloor u_L, u_R \rceil. \tag{4.14}$$

General entropy pairs for the (homogeneous) Jin-Xin relaxation equations write under the form

$$\begin{aligned} \Phi(\mathbb{U}) &= \varphi^+(r^+(\mathbb{U})) + \varphi^-(r^-(\mathbb{U})), \\ \Psi(\mathbb{U}) &= a\big(\varphi^+(r^+(\mathbb{U})\big) - \varphi^-(r^-(\mathbb{U}))\big), \end{aligned} \tag{4.15}$$

with $r^{\pm}(\mathbb{U}) = v \pm au$ for arbitrary pairs of functions (φ^-, φ^+). The consistency requirement (4.14) is therefore met if and only if

$$\varphi^-(h_-(u)) = \frac{1}{2}\Big(\mathcal{U}(u) - \frac{1}{a}\mathcal{F}(u)), \quad \varphi^+(h_+(u)) = \frac{1}{2}\Big(\mathcal{U}(u) + \frac{1}{a}\mathcal{F}(u)\Big), \quad \text{for all } u \in \lfloor u_L, u_R \rceil, \tag{4.16}$$

where $h^{\pm}(u)$ denote the two functions introduced in (4.9). Observe that as a consequence the functions $\varphi^{\pm} : r \in \mathcal{I}_{\pm} \to \varphi^{\pm}(r) \in \mathbb{R}$ under consideration satisfy

$$\frac{d}{dr}\varphi^+(r) = \frac{1}{2a}\mathcal{U}'(h_+^{-1}(r)), \;\; \frac{d}{dr}\varphi^-(r) = -\frac{1}{2a}\mathcal{U}'(h_-^{-1}(r)), \tag{4.17}$$

where again $h_{\pm}^{-1} : r \in \mathcal{I}_{\pm} \to h_{\pm}^{-1}(r) \in \lfloor u_L, u_R \rceil$ are well-defined under the sub-characteristic condition (3.1). Due to the convexity of $\mathcal{U}(u)$, the reported monotonicity properties (4.10) of $h_{\pm}^{-1}$ then ensures the convexity of $\Phi(\mathbb{U})$ over the domain $\mathcal{D}(\lfloor u_L, u_R \rceil)$. Observe that the proposed definitions (4.17) for $\varphi^{\pm}$ are meaningful in the case of the piecewise smooth Kružkov entropies (4.13).

Equipped with (4.15)–(4.17), one then investigates the dissipative properties of the proposed convex extension (Φ, Ψ) with respect to the relaxation mechanisms involved in the Jin-Xin's model. It can be shown (see again [5], [22]) that provided the compact domain $\mathcal{D}(\lfloor u_L, u_R \rceil)$ stays invariant for the relaxation equations, then

$$\partial_v \Phi(u, v)(f(u) - v) \leq 0, \quad \text{for any given } \mathbb{U} = (u, v) \in \mathcal{D}(\lfloor u_L, u_R \rceil). \tag{4.18}$$

As a direct consequence and for all relaxation time $\epsilon > 0$, the solutions $\mathbb{U}^{\epsilon}$ of the Jin-Xin relaxation model with well-prepared initial data $\mathbb{U}_0$ taking values in $\mathcal{D}(\lfloor u_L, u_R \rceil)$ obey in the usual weak sense the entropy-like inequality

$$\partial_t \Phi(\mathbb{U}^{\epsilon}) + \partial_x \Psi(\mathbb{U}^{\epsilon}) = \frac{1}{\epsilon}\partial_v \Phi(\mathbb{U}^{\epsilon})(f(u^{\epsilon}) - v^{\epsilon}) \leq 0. \tag{4.19}$$

Indeed recall that $\mathbb{U}^\epsilon$ remains in the invariant region $\mathcal{D}(\lfloor u_L, u_R \rceil)$. With this in mind, let us examine the behavior of the relaxation entropy pair (Φ, Ψ) for the self similar solution $\mathbb{U}(.; u_L, u_R)$ of the Riemann problem (2.18)–(2.19). In view of Corollary 5, let us stress that this makes sense for defect measure corrections built from mappings $\theta(u_L, u_R)$, *i.e.* satisfying (4.2), since $\mathcal{D}(\lfloor u_L, u_R \rceil)$ is invariant for the self-similar solutions of (2.18). Recall first that $\mathbb{U}(.; u_L, u_R)$ stays constant except across three discontinuities. Concerning the two waves with speed $-a$ and $+a$, their linear degeneracy ensures [26] that any given additional entropy law is exactly preserved for weak solutions. Namely whatever the pair of states (u_L, u_R) are and the definition of the mapping θ under (4.2) is, one has

$$\begin{aligned} +a\big(\Phi(\mathbb{U}_L^\star(\theta; u_L, u_R)) - \Phi(\mathbb{U}_L)\big) + \Psi(\mathbb{U}_L^\star(\theta; u_L, u_R)) - \Psi(\mathbb{U}_L) &= 0, \\ -a\big(\Phi(\mathbb{U}_R) - \Phi(\mathbb{U}_R^\star(\theta; u_L, u_R))\big) + \Psi(\mathbb{U}_R) - \Psi(\mathbb{U}_R^\star(\theta; u_L, u_R)) &= 0. \end{aligned} \tag{4.20}$$

Here $\mathbb{U}_L^\star(\theta; u_L, u_R)$ and $\mathbb{U}_R^\star(\theta; u_L, u_R)$ denote the two intermediate states in (3.3)–(3.4) separated by the discontinuity propagating with speed $\sigma(u_L, u_R)$. At this discontinuity, the defect measure correction acts. The inequality (4.19) strongly suggests that the definition of the mapping $\theta(u_L, u_R)$ should satisfy for any given pair of states (u_L, u_R) the entropy like jump condition

$$\begin{aligned} \mathcal{E}\{\mathcal{U}\}(\theta; u_L, u_R) \quad &:= -\sigma(u_L, u_R)\big(\Phi(\mathbb{U}_R^\star(\theta; u_L, u_R)) - \Phi(\mathbb{U}_L^\star(\theta; u_L, u_R))\big) \\ &\quad + \Psi(\mathbb{U}_R^\star(\theta; u_L, u_R)) - \Psi(\mathbb{U}_L^\star(\theta; u_L, u_R)) \\ &\leq 0. \end{aligned} \tag{4.21}$$

These observations motivate the following

Definition 6. Given any convex entropy pair $(\mathcal{U}, \mathcal{F})$ (2.2) for the scalar conservation law (2.22) and its relaxation extension (Φ, Ψ) (4.15)–(4.17). Then the monotonicity preserving mapping θ in (2.27) is said to be consistent with $(\mathcal{U}, \mathcal{F})$ if for all pair of states (u_L, u_R) the relaxation entropy jump $\mathcal{E}\{\mathcal{U}\}(\theta; u_L, u_R)$ defined in (4.21) is non-positive.

Observe that choosing $\theta(u_L, u_R) = 1$ for special pairs (u_L, u_R) verifying (2.8) is allowed by the proposed condition. Indeed, Lemma 1 ensures that $\mathbb{U}_L^\star(1; u_L, u_R) = \mathbb{U}_L$ and $\mathbb{U}_R^\star(1; u_L, u_R) = \mathbb{U}_R$. The states $\mathbb{U}_L$, $\mathbb{U}_R$ being well-prepared (2.19), the consistency property (4.14) relating (Φ, Ψ) to $(\mathcal{U}, \mathcal{F})$ readily gives

$$\begin{aligned} &\mathcal{E}\{\mathcal{U}\}(1; u_L, u_R) \\ &= -\sigma(u_L, u_R)\big(\Phi(\mathbb{U}_R^\star(1; u_L, u_R)) - \Phi(\mathbb{U}_L^\star(1; u_L, u_R))\big) + \Psi(\mathbb{U}_R^\star(1; u_L, u_R)) - \Psi(\mathbb{U}_L^\star(1; u_L, u_R)) \\ &= -\sigma(u_L, u_R)\big(\Phi(\mathbb{U}_R) - \Phi(\mathbb{U}_L)\big) + \Psi(\mathbb{U}_R) - \Psi(\mathbb{U}_L) \\ &= -\sigma(u_L, u_R)\big(\mathcal{U}(u_R) - \mathcal{U}(u_L)\big) + \mathcal{F}(u_R) - \mathcal{F}(u_L) \\ &\leq 0, \end{aligned} \tag{4.22}$$

Hence, the entropy criterion proposed in (4.21) is automatically satisfied by pairs of interest. This grounds Definition 6 with respect to our main goal. For general pair of states (u_L, u_R), the proposed Definition will be used in connection with the following Lemma which states that the minimum in the v-variable of any strictly convex relaxation entropy $\Phi(u, v)$ lies on the equilibrium manifold. It thus restores the equilibrium entropy $\mathcal{U}(u)$.

Lemma 7. *Assume the sub-characteristic condition (3.1), one has for any given* $u \in \lfloor u_L, u_R \rceil$ *the following Gibb's principle:*

$$f(u) = argmin_v \Phi(u,v). \tag{4.23}$$

Proof. Let u be given in $\lfloor u_L, u_R \rceil$, then by convexity of $\mathcal{U}(u)$ solving in v the equation

$$\partial_v \Phi(u,v) = \frac{1}{2a}\Big(\mathcal{U}'(h_+^{-1}(v+au)) - \mathcal{U}'(h_-^{-1}(v-au))\Big) = 0,$$

is equivalent to

$$h_+^{-1}(v+au) - h_-^{-1}(v-au) = 0.$$

Under condition (3.1) and for all $(u,v) \in \mathcal{D}(\lfloor u_L, u_R \rceil)$, the function $G(v) = h_+^{-1}(v+au) - h_-^{-1}(v-au)$ is strictly increasing in v thanks to (4.10), thus the unique solution of $G(v) = 0$ is given by $v = f(u)$ since $h_+^{-1}(f(u)+au) = h_-^{-1}(f(u)-au) = u$. Then the identity $\mathcal{U}(u) = \Phi(u, f(u))$ gives the conclusion. □

4.3 Entropy consistency for a genuinely non-linear flux function

The main result of this section is

Theorem 8. *Consider the entropy pair* $(\mathcal{U}(u), \mathcal{F}(u))$ *(2.2) with* $\mathcal{U}(u) = u^2/2$ *and the associated relaxation entropy pair* (Φ, Ψ) *(4.15)-(4.16). Assume the sub-characteristic condition (3.1). Then the monotonicity preserving condition (4.2) and the entropy condition* $\mathcal{E}\{\mathcal{U}\}(\theta; u_L, u_R) \leq 0$ *stated in (4.21) are satisfied provided that* $\theta(u_L, u_R)$ *satisfies*

$$0 \leq \theta(u_L, u_R) \leq \Theta(u_L, u_R) \equiv \max(0, \min(1, 1 + \Gamma(u_L, u_R)), \tag{4.24}$$

where

$$\Gamma(u_L, u_R) = \left\{ \begin{array}{ll} -2\ \gamma(u_L, u_R)\dfrac{\big(-\sigma(\mathcal{U}(u_R) - \mathcal{U}(u_L)) + (\mathcal{F}(u_R) - \mathcal{F}(u_L))\big)}{|u_R - u_L|^2}, & u_L \neq u_R, \\ 0, & \textit{otherwise}, \end{array}\right. \tag{4.25}$$

with

$$\gamma(u_L, u_R) = \left\{ \begin{array}{ll} \dfrac{a - \max(|f'(u_L)|, |f'(u_R)|\big)}{\big(a^2 - \sigma^2(u_L, u_R)\big)}, & u_L \neq u_R, \\ 1/\big(a + |f'(u_L)|\big), & \textit{otherwise.} \end{array}\right. \tag{4.26}$$

Observe that for pairs with distinct states $u_L \neq u_R$, $\Gamma(u_L, u_R)$ in (4.25) is clearly well defined under the sub-characteristic condition (3.1). Notice that

$$\gamma(u_L, u_R) = \frac{a - |f'(u_L)|}{a^2 - f'^2(u_L)} + \mathcal{O}(|u_R - u_L|) = \frac{1}{a + |f'(u_L)|} + \mathcal{O}(|u_R - u_L|), \tag{4.27}$$

hence we recover (4.26) in the limit $|u_R - u_L| \to 0$. Observe that for the pairs (u_L, u_R) of interest, namely those verifying the entropy inequality (2.8), we get as expected $\Theta(u_L, u_R) = 1$ so that the accuracy requirement put forward in Lemma 1 can be met. Besides and as it is well-known, general pairs of states come with a cubic entropy rate (see for instance Godlewski-Raviart [9])

$$-\sigma(\mathcal{U}(u_R) - \mathcal{U}(u_L)) + (\mathcal{F}(u_R) - \mathcal{F}(u_L)) = \mathcal{O}(|u_R - u_L|^3). \tag{4.28}$$

We deduce

$$\Gamma(u_L, u_R) = \mathcal{O}(|u_R - u_L|). \tag{4.29}$$

Therefore, Θ is expected to stay close to unity in the smooth zones of the discrete solutions and to reach ultimately 1 as the mesh step Δx goes to zero in those regions.

Expressing the relaxation entropy pair (Φ, Ψ) in terms of the convex pair (φ^-, φ^+) according to (4.15), we first observe that the relaxation entropy jump $\mathcal{E}\{\mathcal{U}\}(\theta; u_L, u_R)$ in (4.21) equivalently reads

$$\mathcal{E}\{\mathcal{U}\}(\theta; u_L, u_R) = (a - \sigma(u_L, u_R)) \left[\varphi^+\right](\theta; u_L, u_R) - (a + \sigma(u_L, u_R)) \left[\varphi^-\right](\theta; u_L, u_R) \tag{4.30}$$

where we have set

$$[\varphi^-](\theta; u_L, u_R) = \varphi^- \left(r_R^{-\star}(\theta)\right) - \varphi^- \left(r_L^{-\star}(\theta)\right), \quad [\varphi^+](\theta; u_L, u_R) = \varphi^+ \left(r_R^{+\star}(\theta)\right) - \varphi^+ \left(r_L^{+\star}(\theta)\right) \tag{4.31}$$

using the characteristic variables $r_L^{\pm\star}(\theta)$ and $r_R^{\pm\star}(\theta)$ defined in (3.8), Corollary 3. Next, the identities (3.10) stated in the same Corollary imply that for all values of $\theta \in [0, 1]$:

$$\varphi^-\left(r_R^{-\star}(\theta)\right) = \varphi^-\left(r_R^-\right), \quad \varphi^-\left(r_L^{+\star}(\theta)\right) = \varphi^+\left(r_L^+\right). \tag{4.32}$$

Hence $\mathcal{E}\{\mathcal{U}\}(\theta; u_L, u_R)$ actually becomes

$$\begin{aligned} \mathcal{E}\{\mathcal{U}\}(\theta; u_L, u_R) \ &= (a - \sigma(u_L, u_R)) \left(\varphi^+ \left(r_R^{+\star}(\theta)\right) - \varphi^+ \left(r_L^+\right) \right) \\ &- (a + \sigma(u_L, u_R)) \left(\varphi^- \left(r_R^-\right) - \varphi^- \left(r_L^{-\star}(\theta)\right) \right). \end{aligned} \tag{4.33}$$

Further notice from the definition (4.17) of the derivatives $\{\varphi^\pm\}'(r)$ that the choice of the quadratic entropy $\mathcal{U}(u) = u^2/2$ with $\mathcal{U}"(u) = 1$ yields

$$\{\varphi^\pm\}''(r) = \frac{1}{2a\big(a \pm f'(h_\pm^{-1}(r))\big)}, \tag{4.34}$$

but to shade light in the forthcoming developments, we mostly keep $\{\varphi^\pm\}"(r)$ unspecified untill the end of this section. The proof of Theorem 8 relies on the following technical result essentially motivated by (4.33) and the convex combination (3.9) for $r_L^{-\star}(\theta)$ and $r_R^{+\star}(\theta)$, stated in Corollary 3.

Lemma 9. *For any given smooth function φ^+ and any given real number θ, the following identity holds of $r_R^{+\star}(\theta)$ defined in (3.8) :*

$$\begin{aligned} &\varphi^+(r_R^{+\star}(\theta)) = \Big\{\theta\varphi^+(r_R^+) + (1-\theta)\varphi^+(r_L^+)\Big\} \\ &-\theta(1-\theta) \ \int_0^1 \Big\{(1-\theta)\{\varphi^+\}''(r_R^+(s,\theta)) + \theta\{\varphi^+\}''(r_L^+(s,\theta))\Big\}(1-s)ds \ \Big(r_R^+ - r_L^+\Big)^2, \end{aligned} \tag{4.35}$$

where we have set

$$r_R^+(s,\theta) = sr_R^+ + (1-s)r_R^{+\star}(\theta), \quad r_L^+(s,\theta) = sr_L^+ + (1-s)r_R^{+\star}(\theta), \quad s \in [0,1]. \tag{4.36}$$

The following identity is also valid for $r_L^{-\star}(\theta)$ (3.8) for all θ and any given smooth function φ^-

$$\begin{aligned} &\varphi^-(r_L^{-\star}(\theta)) = \Big\{(1-\theta)\varphi^-(r_R^-) + \theta\varphi^-(r_L^-)\Big\} \\ &-\theta(1-\theta)\ \int_0^1 \Big\{\theta\{\varphi^-\}''(r_R^-(s,\theta)) + (1-\theta)\{\varphi^-\}''(r_L^-(s,\theta))\Big\}(1-s)ds\ \Big(r_R^- - r_L^-\Big)^2. \end{aligned} \tag{4.37}$$

where we have defined :

$$r_R^-(s,\theta) = sr_R^- + (1-s)r_L^{-\star}(\theta), \quad r_L^-(s,\theta) = sr_L^- + (1-s)r_L^{-\star}(\theta), \quad s \in [0,1]. \tag{4.38}$$

Proof. First observe the identity

$$\varphi^+(r_R^+) - \varphi^+(r_R^{+\star}(\theta)) = \{\varphi^+\}'(r_R^{+\star}(\theta))\Big(r_R^+ - r_R^{+\star}(\theta)\Big) + \int_{r_R^{+\star}(\theta)}^{r_R^+} \{\varphi^+\}''(r)\Big(r_R^+ - r\Big)dr, \tag{4.39}$$

together with

$$\varphi^+(r_L^+) - \varphi^+(r_R^{+\star}(\theta)) = \{\varphi^+\}'(r_R^{+\star}(\theta))\Big(r_L^+ - r_R^{+\star}(\theta)\Big) + \int_{r_R^{+\star}(\theta)}^{r_L^+} \{\varphi^+\}''(r)\Big(r_L^+ - r\Big)dr, \tag{4.40}$$

so as to infer from the definition of $r_R^{+\star}(\theta)$ in term of the convex decomposition (3.9) stated in Corollary 3

$$\begin{aligned} &\varphi^+(r_R^{+\star}(\theta)) - \Big\{\theta\varphi^+(r_R^+) + (1-\theta)\varphi^+(r_L^+)\Big\} = \\ &\qquad -\theta\int_{r_R^{+\star}(\theta)}^{r_R^+} \{\varphi^+\}''(r)\Big(r_R^+ - r\Big)dr - (1-\theta)\int_{r_R^{+\star}(\theta)}^{r_L^+} \{\varphi^+\}''(r)\Big(r_L^+ - r\Big)dr. \end{aligned} \tag{4.41}$$

Introducing $r_R^+(s,\theta) = sr_R^+ + (1-s)r_R^{+\star}(\theta)$ with $s \in [0,1]$, a convenient form of the first integral in (4.41) reads

$$\begin{aligned} \int_{r_R^{+\star}(\theta)}^{r_R^+} \{\varphi^+\}''(r)\Big(r_R^+ - r\Big)dr \quad &= \int_0^1 \{\varphi^+\}''(r_R^+(s,\theta))(1-s)ds\ \Big(r_R^+ - r_R^{+\star}(\theta)\Big)^2 \\ &= (1-\theta)^2\ \int_0^1 \{\varphi^+\}''(r_R^+(s,\theta))(1-s)ds\ \Big(r_R^+ - r_L^+\Big)^2, \end{aligned} \tag{4.42}$$

thanks again to the linear decomposition (3.9) of $r_R^{+\star}(\theta)$. Defining similarly $r_L^+(s,\theta) = sr_L^+ + (1-s)r_R^{+\star}(\theta)$ with $s \in [0,1]$, the second integral in (4.41) can be equivalently rewritten as

$$\int_{r_R^{+\star}(\theta)}^{r_L^+} \{\varphi^+\}''(r)\Big(r_L^+ - r\Big)dr = \theta^2\ \int_0^1 \{\varphi^+\}''(r_L^+(s,\theta))(1-s)ds\ \Big(r_R^+ - r_L^+\Big)^2. \tag{4.43}$$

Hence, the representation formula (4.41) becomes

$$\begin{aligned}&\varphi^+(r_R^{+\star}(\theta)) - \Big\{\theta\varphi^+(r_R^+) + (1-\theta)\varphi^+(r_L^+)\Big\} = \\ &-\theta(1-\theta)\ \int_0^1 \Big\{(1-\theta)\{\varphi^+\}''(r_R^+(s,\theta)) + \theta\{\varphi^+\}''(r_L^+(s,\theta))\Big\}(1-s)ds\ \Big(r_R^+ - r_L^+\Big)^2.\end{aligned} \tag{4.44}$$

This is nothing but the required identity (4.35). The companion formula (4.36) follows using similar steps that are left to the reader. □

We are in a position to prove Theorem 8.

Proof of Theorem 8. The representation formulas (4.35) and (4.37) that are at the core of the proof, exhibit a rather intricate nonlinear dependance in θ through the mappings $r_L^{\pm\star}(\theta, s)$ and $r_R^{\pm\star}(\theta, s)$ in (4.36)–(4.38). For the sake of simplicity, we aim at lowering such a dependance to a quadratic one when introducing suitable lower-bounds of the integral remainder in the Taylor like expansions (4.35)–(4.37). To that purpose, it clearly suffices to propose a common positive lower-bound, say $m^+(u_L, u_R)$, for $\{\varphi^+\}''(r_L^{+\star}(s,\theta))$ and $\{\varphi^+\}''(r_R^{+\star}(s,\theta))$ valid for all the θ and s under consideration to get

$$\theta(1-\theta)\int_0^1 \Big\{(1-\theta)\{\varphi^+\}''(r_R^+(s,\theta)) + \theta\{\varphi^+\}''(r_L^+(s,\theta))\Big\}(1-s)ds \geq \frac{1}{2}\theta(1-\theta)m^+(u_L, u_R). \tag{4.45}$$

A similar estimate clearly holds true for (4.37) adopting the same procedure with some positive lower bound $m^-(u_L, u_R)$. Here again for simplicity we promote a common lower-bound $m(u_L, u_R) = m^-(u_L, u_R) = m^+(u_L, u_R)$. As already reported, Corollary 3 ensures that monotonicity preserving mappings $\theta(u_L, u_R)$ make $r_L^+(s,\theta)$, $r_R^+(s,\theta)$ (respectively $r_L^-(s,\theta)$, $r_R^-(s,\theta)$) cover the range $\lfloor r_L^+, r_R^+\rceil$ (resp. $\lfloor r_L^-, r_R^-\rceil$) as s and θ jointly vary in $[0,1]$. Consequently, both $h_+^{-1}(r)$ and $h_-^{-1}(r)$ keep their values in $\lfloor u_L, u_R\rceil$ for all the r under consideration. The lower-bound m we seek for, must therefore satisfy

$$\min_{u\in\lfloor u_L, u_R\rceil}\Big(\frac{1}{2a(a+f'(u))}, \frac{1}{2a(a-f'(u))}\Big) \geq m(u_L, u_R). \tag{4.46}$$

But since the flux function f is assumed to be genuinely non-linear, the minimum in the left hand-side is achieved for $u = u_L$ or $u = u_R$ and we can thus choose

$$m(u_L, u_R) = \frac{1}{2a\big(a - \max(|f'(u_L)|, |f'(u_R)|)\big)}. \tag{4.47}$$

Plugging the proposed estimate in the representation formulas (4.35) and (4.37) clearly gives :

$$\begin{aligned}\varphi^+(r_R^{+\star}(\theta)) &\leq \Big\{\theta\varphi^+(r_R^+) + (1-\theta)\varphi^+(r_L^+)\Big\} - \frac{\theta(1-\theta)}{2}m(u_L, u_R)\ |r_R^+ - r_L^+|^2, \\ \varphi^-(r_L^{-\star}(\theta)) &\leq \Big\{(1-\theta)\varphi^-(r_R^-) + \theta\varphi^-(r_L^-)\Big\} - \frac{\theta(1-\theta)}{2}m(u_L, u_R)\ |r_R^- - r_L^-|^2,\end{aligned} \tag{4.48}$$

where $|r_R^- - r_L^-| = (a-\sigma)|u_R - u_L|$ and $|r_R^+ - r_L^+| = (a+\sigma)|u_R - u_L|$. We can therefore bound the relaxation entropy jump $\mathcal{E}\{\mathcal{U}\}(\theta, u_L, u_R)$ equivalently defined in (4.33) according to

$$\begin{aligned}
\mathcal{E}\{\mathcal{U}\}(\theta, u_L, u_R) &= (a-\sigma)\Big(\varphi^+\left(r_R^{+\star}(\theta)\right) - \varphi^+\left(r_L^+\right)\Big) + (a+\sigma)\Big(\varphi^-\left(r_L^{-\star}(\theta)\right) - \varphi^-\left(r_R^-\right)\Big) \\
&\leq \theta\Big\{(a-\sigma)\big(\varphi^+(r_R^+) - \varphi^+(r_L^+)\big) + (a+\sigma)\big(\varphi^-(r_R^-) - \varphi^-(r_L^-)\big)\Big\} \\
&-\frac{\theta(1-\theta)}{2}m(u_L, u_R)\ \Big\{(a-\sigma)|r_R^+ - r_L^+|^2 + (a+\sigma)|r_R^- - r_L^-|^2\Big\} \\
&= \theta\Big\{-\sigma(\mathcal{U}(u_R) - \mathcal{U}(u_L)) + (\mathcal{F}(u_R) - \mathcal{F}(u_L)\Big\} \\
&-\theta(1-\theta)a(a^2-\sigma^2)m(u_L, u_R)|u_R - u_L|^2
\end{aligned} \tag{4.49}$$

since following exactly the same steps as those developed to get (4.22) gives

$$\begin{aligned}
&(a-\sigma)\big(\varphi^+(r_R^+) - \varphi^+(r_L^+)\big) + (a+\sigma)\big(\varphi^-(r_R^-) - \varphi^-(r_L^-)\big) \\
&\qquad = -\sigma(\mathcal{U}(u_R) - \mathcal{U}(u_L)) + (\mathcal{F}(u_R) - \mathcal{F}(u_L)) \\
&\qquad = \mathcal{E}\{\mathcal{U}\}(1, u_L, u_R).
\end{aligned}$$

Hence the estimate (4.49) gives

$$\mathcal{E}\{\mathcal{U}\}(\theta, u_L, u_R) \leq \theta\Big(\mathcal{E}\{\mathcal{U}\}(1, u_L, u_R) - (1-\theta)A(u_L, u_R)\Big) \tag{4.50}$$

where we have set

$$A(u_L, u_R) = a(a^2 - \sigma^2(u_L, u_R))m(u_L, u_R)|u_R - u_L|^2. \tag{4.51}$$

Hence assuming $u_L \neq u_R$, (4.50) just reads :

$$\mathcal{E}\{\mathcal{U}\}(\theta, u_L, u_R) \leq A(u_L, u_R)\Big\{\theta\big(\theta - (1 + \Gamma(u_L, u_R))\big)\Big\} \tag{4.52}$$

with $\Gamma(u_L, u_R)$ defined in (4.25). Since $A(u_L, u_R) > 0$, it suffices to require $\theta\big(\theta - (1 + \Gamma(u_L, u_R))\big) \leq 0$ to ensure the expected entropy inequality $\mathcal{E}\{\mathcal{U}\}(\theta, u_L, u_R) \leq 0$ for the pair of states (u_L, u_R) under consideration. Enforcing as mandatory the monotonicity preserving condition $0 \leq \theta(u_L, u_R) \leq 1$ thus yields the condition (4.24). This concludes the proof.

Let us underline that the upper-bound (4.50) is sharp with respect to our main motivation. Indeed, it boils downs to the equality $\mathcal{E}\{\mathcal{U}\}(\theta = 1, u_L, u_R) = \mathcal{E}\{\mathcal{U}\}(1, u_L, u_R)$ and therefore it exactly preserves all the pairs (u_L, u_R) of interest, *i.e.* those satisfying $\mathcal{E}\{\mathcal{U}\}(1, u_L, u_R) \leq 0$.

4.4 Entropy consistency for a general flux function

To begin with, it is worth briefly recalling a few well-known facts about the Kružkov entropy criterion for selecting admissible pairs of states (u_L, u_R) that satisfy the Rankine-Hugoniot relation

$$-\sigma(u_L, u_R)(u_R - u_L) + (f(u_R) - f(u_L)) = 0. \tag{4.53}$$

Kružkov entropy inequalities read

$$-\sigma(u_L,u_R)\big(|u_R-k|-|u_L-k|\big)+\big(\mathrm{sign}(u_R-k)(f(u_R)-f(k))-\mathrm{sign}(u_L-k)(f(u_L)-f(k))\big)\leq 0, \tag{4.54}$$

for all $k\in\mathbb{R}$. To discard empty intervals from the discussion, we tacitly assume that the states in all the pairs under consideration are distinct, namely $u_L\neq u_R$. In (4.54), values of the parameter k outside of the interval $\lfloor u_L,u_R\rceil$ are easily seen to satisfy the Rankine-Hugoniot jump relation (4.53), so that only the values of k in $\lfloor u_L,u_R\rceil$ are entropy diminishing

$$-\sigma(u_L,u_R)\big(u_R+u_L-2k\big)+\big(f(u_R)+f(u_L)-2f(k)\big)\leq 0. \tag{4.55}$$

In view of (4.53), this requirement is equivalent to the so-called Oleinik inequalities :

$$\begin{aligned}\mathcal{K}(k;u_L,u_R)\ &:= -\sigma(u_L,u_R)\big(u_R-k\big)+\big(f(u_R)-f(k)\big)\\ &= -\sigma(u_L,u_R)\big(u_L-k\big)+\big(f(u_L)-f(k)\big)\leq 0,\quad k\in\lfloor u_L,u_R\rceil.\end{aligned} \tag{4.56}$$

The main result of this section is

Theorem 10. *Let us consider the Kružkov entropy pairs $(\mathcal{U}_k(u),\mathcal{F}_k(u))$ (4.13) with $k\in\lfloor u_L,u_R\rceil$ and the associated relaxation entropy pairs (Φ_k,Ψ_k) (4.15)-(4.16). Assume the sub-characteristic condition (3.1) and consider monotonicity preserving mappings $\theta(u_L,u_R)$ (4.2). Then the relaxation entropy jump $\mathcal{E}\{\mathcal{U}_k\}(\theta;u_L,u_R)$ in (4.21) stay non positive for all $k\in\lfloor u_L,u_R\rceil$ provided that $\theta(u_L,u_R)$ is chosen to satisfy*

$$0\leq\theta(u_L,u_R)\leq\Theta(u_L,u_R)=\min_{k\in\lfloor u_L,u_R\rceil}\Gamma_{\mathcal{K}}(k;u_L,u_R), \tag{4.57}$$

where for the states $u^\star(u_L,u_R)$ and $v^\star(u_L,u_R)$ defined in (3.2)

$$\Gamma_{\mathcal{K}}(k;u_L,u_R)=2\gamma(u_L,u_R)\begin{cases}-\Big(\dfrac{-\sigma(u_L,u_R)\big(u^\star(u_L,u_R)-k\big)+\big(v^\star(u_L,u_R)-f(k)\big)}{u_R-u_L}\Big), & \text{if } u_L\neq u_R,\\[2ex] \dfrac{a^2-\sigma^2(u_L,u_R)}{2a}=\dfrac{1}{2\gamma(u_L,u_R)}, & \text{otherwise},\end{cases} \tag{4.58}$$

with

$$\gamma(u_L,u_R)=\frac{a}{a^2-\sigma^2(u_L,u_R)}. \tag{4.59}$$

For any given pair of states (u_L,u_R), $\Theta(u_L,u_R)$ takes value in $(0,1)$ and there exists at least one minimizer $k(u_L,u_R)$ of $\Gamma_{\mathcal{K}}(k;u_L,u_R)$ in $\lfloor u_L,u_R\rceil$ with the property that

$$\Theta(u_L,u_R)=1 \text{ if } \mathcal{K}(k;u_L,u_R)\leq 0 \text{ for all } k\in\lfloor u_L,u_R\rceil,\quad \text{and } 0<\Theta(u_L,u_R)<1 \text{ otherwise.} \tag{4.60}$$

Observe that the function $\Gamma_{\mathcal{K}}(k;u_L,u_R)$ has close relationships with the function $\mathcal{K}(k;u_L,u_R)$ entering the Oleinik inequalities (4.56). With this respect, the optimal law $\Theta(u_L,u_R)$ (4.57)–(4.58) is nothing but a natural extension of the corresponding formula (4.24)–(4.25) derived in the frame of a genuinely non-linear flux function. At this stage, it is interesting to give a geometric interpretation of the extrema of $\Gamma_{\mathcal{K}}(k; u_L, u_R)$ in terms of the function $\mathcal{K}(k; u_L, u_R)$. All existing extrema of $\Gamma_{\mathcal{K}}(k; u_L, u_R)$ that are achieved for values $k_e(u_L, u_R)$ in $\lfloor u_L, u_R \rceil$ and that are distinct from u_L and u_R must clearly satisfy the property

$$f'(k_e(u_L, u_R)) = \sigma(u_L, u_R). \tag{4.61}$$

But since $\mathcal{K}(k = u_L; u_L, u_R) = \mathcal{K}(k = u_R; u_L, u_R) = 0$ from (4.53), the function $k \in \lfloor u_L, u_R \rceil \to \mathcal{K}(k; u_L, u_R)$ also necessarily admits at least one extremum in the proposed interval. In addition all the possible extrema when distinct from u_L and u_R must also verify the condition (4.61) from the definition (4.56). Hence, both $\mathcal{K}(k; u_L, u_R)$ and $\Gamma_{\mathcal{K}}(k; u_L, u_R)$ achieve their extrema in the reported open interval at the same locations. Let us stress that those observations actually give us a natural and simple algorithm for computing the optimal bound $\Theta(u_L, u_R)$ in (4.57). At last observe that in the limit $|u_R - u_l| \to 0$, we get $\Theta(u_L, u_R) = 1$. Anticipating the numerical application, this means that the method is asymptotically close (in terms of the mesh step Δx) to a Roe solver in the smooth parts of the discrete solution.

In order to prove Theorem 10, we define the following two functions of the parameter k

$$\mathcal{R}_-(k) = \frac{r_R^- - h_-(k)}{r_R^- - r_L^-}, \quad \mathcal{R}_+(k) = \frac{h_+(k) - r_L^+}{r_R^+ - r_L^+}, \quad k \in \lfloor u_L, u_R \rceil, \tag{4.62}$$

based on the characteristic variables $r^\pm = v \pm au$ and the invertible mappings $h_\pm$ defined in (4.9). Direct calculations that are left to the reader give from the definitions of $h_\pm(r)$, $r_L^\pm$ and $r_R^\pm$

$$\begin{aligned}&\frac{a^2 - \sigma^2(u_L, u_R)}{2a}(u_R - u_L)\Big(\mathcal{R}_-(k) + \mathcal{R}_+(k)\Big) \\ &\qquad\qquad = (f(k) - \sigma(u_L, u_R)k) - (v^\star(u_L, u_R) - \sigma(u_L, u_R)u^\star(u_L, u_R)).\end{aligned}$$

This formula clearly stays at the basis of the definition of $\Gamma_{\mathcal{K}}(u_L, u_R, k)$ in (4.58). Equipped with this identity, we claim the following statement equivalent to Theorem 10.

Theorem 11. *Under the assumptions of Theorem 10, the relaxation entropy jump $\mathcal{E}\{\mathcal{U}_k\}(\theta; u_L, u_R)$ in (4.21) stay non positive for all $k \in \lfloor u_L, u_R \rceil$ provided that $\theta(u_L, u_R)$ is chosen to satisfy*

$$0 \le \theta(u_L, u_R) \le \Theta(u_L, u_R) = \min_{k \in \lfloor u_L, u_R \rceil} \Big\{\mathcal{R}_-(k) + \mathcal{R}_+(k)\Big\}, \tag{4.63}$$

where $\Theta(u_L, u_R)$ takes value in $(0, 1)$. For any given pair of states (u_L, u_R), there exists at least one minimizer $k(u_L, u_R)$ of $\mathcal{R}_-(k) + \mathcal{R}_+(k)$ in $\lfloor u_L, u_R \rceil$ with the property that

$$\Theta(u_L, u_R) = 1 \text{ if } \mathcal{K}(k; u_L, u_R) \le 0 \text{ for all } k \in \lfloor u_L, u_R \rceil, \quad \text{and } 0 < \Theta(u_L, u_R) < 1 \text{ otherwise.} \tag{4.64}$$

The proof of this statement is postponed to the end of the section. In order to comment properties of the optimal choice $\Theta(u_L, u_R)$ in (4.63), let us observe the following easy properties for the functions $\mathcal{R}_\pm(k)$. First, $\mathcal{R}_-(k)$ and $\mathcal{R}_+(k)$ keep values in $[0, 1]$ as k varies in

$\lfloor u_L, u_R \rceil$ because $h_-(k)$ (respectively $h_+(k)$) covers $\lfloor r_R^-, r_L^- \rceil$ (resp. $\lfloor r_L^+, r_R^+ \rceil$). In addition, it is easily seen from the definitions of $h_\pm(k)$ that

$$\mathcal{R}_-(u_L) + \mathcal{R}_+(u_L) = \mathcal{R}_-(u_R) + \mathcal{R}_+(u_R) = 1. \tag{4.65}$$

Hence $\Theta(u_L, u_R)$ naturally keeps its values in the interval $[0, 1]$ and is thus automatically monotonicity preserving. Next and because of the identity (4.65), the mapping $k \in \lfloor u_L, u_R \rceil \to \mathcal{R}_-(k) + \mathcal{R}_+(k)$ has clearly at least one extremum. As a consequence of a forthcoming representation formula for the entropy jump $\mathcal{E}\{\mathcal{U}\}_k(\theta; u_L, u_R)$, we prove hereafter that all the existing extrema stay necessarily larger than 1 in the case the pair (u_L, u_R) under consideration obeys the Kruzkov's selection principle $\mathcal{K}(k; u_L, u_R) \leq 0$ for all $k \in \lfloor u_L, u_R \rceil$. Hence we get from (4.65) the expected value $\Theta(u_L, u_R) = 1$. For other pairs, it will be seen that there exists necessarily one local minimizer $k_m(u_L, u_R)$ with the property that $(\mathcal{R}_- + \mathcal{R}_+)(k_m(u_L, u_R)) < 1$. Entropy limitation is active and we have $0 < \Theta(u_L, u_R) < 1$. Provided that all the extrema of $\mathcal{K}(k; u_L, u_R)$ stay non-positive, namely the Kruzkov's entropy condition (4.56) is valid, then $(\mathcal{R}_- + \mathcal{R}_+)(k)$ stays larger than one and achieves from (4.65) the value 1 at the boundaries. If one minimum of $\mathcal{K}(k; u_L, u_R)$ turns positive then entropy violation takes place for the pair (u_L, u_R) under consideration and there exists one minimum of $(\mathcal{R}_- + \mathcal{R}_+)(k)$ with a value less than one. The proof of Theorem 11 relies on the following technical result.

Lemma 12. *Given a smooth enough entropy pair $(\mathcal{U}, \mathcal{F})$ (2.2) and the corresponding relaxation entropy pair (Φ, Ψ) (4.15)-(4.16). Consider monotonicity preserving mappings $\theta(u_L, u_R)$ (4.2). Let us define from the pair of state (u_L, u_R) the following affine functions of the Riemann invariants :*

$$r_-(z) = r_R^- + z(r_L^- - r_R^-), \quad r_+(z) = r_L^+ + z(r_R^+ - r_L^+), \quad z \in [0, 1]. \tag{4.66}$$

Then the relaxation entropy jump $\mathcal{E}\{U\}(\theta; u_L, u_R)$ in (4.21) equivalently reads

$$\mathcal{E}\{U\}(\theta; u_L, u_R) = \frac{a^2 - \sigma^2(u_L, u_R)}{2a}\big(u_R - u_L\big) \int_0^\theta \Big\{\mathcal{U}'\big((h_+^{-1}(r_+(z))\big) - \mathcal{U}'\big((h_-^{-1}(r_-(z))\big)\Big\} dz. \tag{4.67}$$

Proof. Let us re-express the entropy jump $\mathcal{E}\{U\}(\theta; u_L, u_R)$ for the pair (Φ, Ψ) in terms of the underlying convex pair (φ^-, φ^+) in (4.15) :

$$\mathcal{E}\{U\}(\theta) = (a - \sigma)\Big(\varphi^+(r_R^{+\star}(\theta)) - \varphi^+(r_L^{+\star}(\theta))\Big) - (a + \sigma)\Big(\varphi^-(r_R^{-\star}(\theta)) - \varphi^-(r_L^{-\star}(\theta))\Big) \tag{4.68}$$

where by construction from Corollary 3, one has for all $\theta \in [0, 1]$ the following convex decompositions

$$\begin{array}{ll} r_L^{-\star}(\theta) = r_R^- + \theta(r_L^- - r_R^-), & r_R^{-\star}(\theta) = r_R^-, \\ r_L^{+\star}(\theta) = r_L^+, & r_R^{+\star}(\theta) = r_L^+ + \theta(r_R^+ - r_L^+). \end{array} \tag{4.69}$$

We can thus rewrite (4.68) as follows

$$\begin{array}{rl} \mathcal{E}\{U\}(\theta) & = (a - \sigma) \int_0^1 \varphi^{+'}\big(r_L^+ + s\theta(r_R^+ - r_L^+)\big) ds \, \Big\{\theta(r_R^+ - r_L^+)\Big\} \\ & +(a + \sigma) \int_0^1 \varphi^{-'}\big(r_R^- + s\theta(r_L^- - r_R^-)\big) ds \, \Big\{\theta(r_L^- - r_R^-)\Big\} \end{array}$$

where $r_R^+ - r_L^+ = (a+\sigma)(u_R - u_L)$ and $r_L^- - r_R^- = (a-\sigma)(u_R-u_L)$. By construction $\varphi^{\pm'}(r) = \pm U'(h_\pm^{-1}(r))/(2a)$, the above identity just reads

$$\mathcal{E}\{U\}(\theta) = \frac{(a^2-\sigma^2)}{2a}\big(u_R-u_L\big)\Big\{\theta\int_0^1 \mathcal{U}'\big(h_+^{-1}(r_+(s\theta))\big)ds - \theta\int_0^1 \mathcal{U}'\big(h_-^{-1}(r_-(s\theta))\big)ds\Big\}, \tag{4.70}$$

where $r_\pm$ denote the affine functions introduced in (4.66) but evaluated in $z = s\theta$. A change of variable gives the conclusion. □

It is worth observing that (4.70) can be recast as

$$\mathcal{E}\{U\}(\theta) = \theta\frac{(a^2-\sigma^2)}{2a}\mathcal{G}(\theta), \quad \mathcal{G}(\theta) = \big(u_R-u_L\big)\Big\{\int_0^1 \mathcal{U}'\big(h_+^{-1}(r_+(s\theta))\big) - \mathcal{U}'\big(h_-^{-1}(r_-(s\theta))\big)ds\Big\}, \tag{4.71}$$

where from (4.66) and the definitions of $h_\pm^{-1}(r)$ in (4.9), one has

$$\mathcal{G}(0) = \big(u_R-u_L\big)\Big\{\mathcal{U}'\big(h_+^{-1}(r_+(0))\big) - \mathcal{U}'\big(h_-^{-1}(r_-(0))\big)\Big\} = -\big(u_R-u_L\big)\Big\{U'(u_R) - U'(u_L)\Big\} < 0, \tag{4.72}$$

when assuming a strictly entropy $\mathcal{U}(u)$. We thus infer that for any given smooth enough strictly convex entropy, the associated entropy jump in (4.71) can be made negative for small enough positive values of θ. This observation strongly reflects the property stated in Theorem 11 that making varying $k \in \lfloor u_L, u_R \rceil$ in the Kružkov entropy pairs actually allows to define a positive value $\Theta(u_L, u_R)$ for any given pair of states (u_L, u_R). To further shade light, let us recall that under the subcharacteristic's condition (3.1) the mapping $h_+^{-1}(r)$ strictly increases while $h_-^{-1}(r)$ strictly decreases, so that for any given smooth strictly convex entropy $\mathcal{U}(u)$ easy calculations ensure $\mathcal{G}'(\theta) > 0$ for all $\theta \in [0,1]$ in view of

$$\frac{d}{d\theta}r_-(s\theta) = s(r_L^- - r_R^-) = s(a-\sigma)(u_R-u_L), \quad \frac{d}{d\theta}r_-(s\theta) = s(r_R^+ - r_L^+) = s(a+\sigma)(u_R-u_L). \tag{4.73}$$

For entropy violating pairs of states (u_L, u_R) and a prescribed entropy $\mathcal{U}$, $\mathcal{G}(1)$ may achieve a strictly positive value in contrast to $\mathcal{G}(0) < 0$. But thanks to the strict monotonicity of the mapping $\mathcal{G}(\theta)$, there exists a unique value $\theta(\mathcal{U}; u_L, u_R)$ in $(0,1)$ depending on the entropy $\mathcal{U}$ under consideration so that $\mathcal{E}\{U\}(\theta)$ keeps negative values for all θ in $(0, \theta(\mathcal{U}; u_L, u_R))$ and positive values otherwise. The pair (u_L, u_R) being prescribed, the difficulty is then to derive a sharp positive lower bound $\theta(u_L, u_R)$ of the values $\theta(\mathcal{U}; u_L, u_R)$ when making varying all the (strictly convex) entropy pairs $\mathcal{U}$ for the sake of uniqueness. Theorem 11 precisely provides one with the best lower bound, namely $\Theta(u_L, u_R)$ in (4.63), for the Kružkov entropy pairs with $k \in \lfloor u_L, u_R \rceil$.

Let us specialize the above result to the Kružkov entropy family.

Lemma 13. *Consider monotonicity preserving mappings $\theta(u_L, u_R)$ (4.2). Then the Kružkov entropy jump (4.54) for any given $k \in \lfloor u_L, u_R \rceil$ writes*

$$\mathcal{E}\{\mathcal{U}_k\}(\theta; u_L, u_R) = -\frac{a^2 - \sigma^2(u_L, u_R)}{2a}|u_R - u_L|\Big\{\mathcal{R}_-(k) + \mathcal{R}_+(k) - |\theta - \mathcal{R}_-(k)| - |\theta - \mathcal{R}_+(k)|\Big\}. \tag{4.74}$$

Proof. Lemma 12 ensures that the relaxation entropy jump coming with the Kružkov entropy pair (4.13) writes

$$\mathcal{E}\{\mathcal{U}_k\}(\theta) = \frac{a^2-\sigma^2(u_L,u_R)}{2a}|u_R-u_L| \int_0^\theta sign(u_R-u_L)\Big\{\mathcal{U}_k'\big((h_+^{-1}(r_+(z))\big)-\mathcal{U}_k'\big((h_-^{-1}(r_-(z))\big)\Big\}dz, \tag{4.75}$$

where

$$\mathcal{U}_k'\big(h_-^{-1}(r_-(z))\big) = \left\{\begin{array}{ll} -1, & h_-^{-1}\big(r_-(z)\big) < k, \\ +1, & h_-^{-1}\big(r_-(z)\big) > k, \end{array}\right. , \quad \mathcal{U}_k'\big(h_+^{-1}(r_+(z))\big) = \left\{\begin{array}{ll} -1, & h_+^{-1}\big(r_+(z)\big) < k, \\ +1, & h_+^{-1}\big(r_+(z)\big) > k. \end{array}\right. \tag{4.76}$$

Recall that under the sub-characteristic condition (3.1), $h_-^{-1}(r)$ strictly decreases while $h_+^{-1}(r)$ strictly increases so that (4.76) read equivalently

$$\mathcal{U}_k'\big(h_-^{-1}(r_-(z))\big) = \left\{\begin{array}{ll} +1, & r_-(z) < h_-(k), \\ -1, & r_-(z) > h_-(k) \end{array}\right. , \quad \mathcal{U}_k'\big(h_+^{-1}(r_+(z))\big) = \left\{\begin{array}{ll} -1, & r_+(z) < h_+(k), \\ +1, & r_+(z) > h_+(k). \end{array}\right. \tag{4.77}$$

Easy calculations left to the reader, based on the sign of (u_R-u_L) with $r_R^+ - r_L^+ = (a-\sigma)(u_R-u_L)$ and $r_L^- - r_R^- = (a+\sigma)(u_R-u_L)$ then allow to recast (4.77) as follows

$$\begin{aligned} \operatorname{sign}(u_R-u_L)\mathcal{U}_k'\big(h_-^{-1}(r_-(z))\big) &= \left\{\begin{array}{ll} +1, & z < \mathcal{R}_-(k), \\ -1, & z > \mathcal{R}_-(k), \end{array}\right. \\ \operatorname{sign}(u_R-u_L)\mathcal{U}_k'\big(h_+^{-1}(r_+(z))\big) &= \left\{\begin{array}{ll} -1, & z < \mathcal{R}_+(k), \\ +1, & z > \mathcal{R}_+(k), \end{array}\right. \end{aligned} \tag{4.78}$$

with $\mathcal{R}_\pm(k)$ defined in (4.62). The proposed formulas then yield

$$\begin{aligned} \int_0^\theta \operatorname{sign}(u_R-u_L)\mathcal{U}_k'\big((h_-^{-1}(r_-(z))\big)dz &= (+1)\min\big(\theta,\mathcal{R}_-(k)\big) + (-1)\big(\theta-\mathcal{R}_-(k)\big)_+, \\ &= \mathcal{R}_-(k) - \big(\mathcal{R}_-(k)-\theta\big)_+ - \big(\theta-\mathcal{R}_-(k)\big)_+ \\ &= \mathcal{R}_-(k) - |\theta-\mathcal{R}_-(k)|, \end{aligned} \tag{4.79}$$

where we have used the identity $\min(a,b) = b - \big(b-a\big)_+$ with $\big(b-a\big)_+ = \max(0, b-a)$ for any given pair of real numbers (a,b). Similarly, one can infer

$$\int_0^\theta \operatorname{sign}(u_R-u_L)\mathcal{U}_k'\big((h_+^{-1}(r_+(z))\big)dz = -\Big(\mathcal{R}_+(k) - |\theta-\mathcal{R}_+(k)|\Big), \tag{4.80}$$

so that the required identity (4.74) follows from (4.75). □

As an immediate consequence, we infer the following important result

Corollary 14. *Given any pair of states (u_L,u_R) obeying the Kružkov entropy condition $\mathcal{K}(k;u_L,u_R) \leq 0$ for all $k \in \lfloor u_L,u_R\rceil$ in (4.56). Then*

$$\min_{k\in\lfloor u_L,u_R\rceil}(\mathcal{R}_-(k)+\mathcal{R}_+(k)) = 1. \tag{4.81}$$

If there exists $k_\star$ in $\lfloor u_L, u_R \rceil$ with the property $\mathcal{K}(k_\star; u_L, u_R) > 0$, namely the pair (u_L, u_R) is entropy violating, then

$$\min_{k \in \lfloor u_L, u_R \rceil} (\mathcal{R}_-(k) + \mathcal{R}_+(k)) < 1. \tag{4.82}$$

Proof. Assume an entropy satisfying pair (u_L, u_R). Then from Lemma 1, we have on the one hand from (4.22)

$$\mathcal{E}\{\mathcal{U}_k\}(\theta = 1) = \mathcal{K}(k; u_L, u_R) \leq 0, \quad \text{for all } k \text{ under consideration,} \tag{4.83}$$

while on the other hand, the representation formula (4.74) asserts that

$$\mathcal{E}\{\mathcal{U}_k\}(\theta = 1; u_L, u_R) = -\frac{a^2 - \sigma^2(u_L, u_R)}{2a} |u_R - u_L| \Big\{ \mathcal{R}_-(k) + \mathcal{R}_+(k) - |1 - \mathcal{R}_-(k)| - |1 - \mathcal{R}_+(k)| \Big\}. \tag{4.84}$$

We thus infer that

$$\mathcal{R}_-(k) + \mathcal{R}_+(k) \geq |1 - \mathcal{R}_-(k)| + |1 - \mathcal{R}_+(k)|, \tag{4.85}$$

but since both functions $\mathcal{R}_\pm(k)$ keep their values in $[0, 1]$, we get

$$2(\mathcal{R}_-(k) + \mathcal{R}_+(k)) \geq 2, \quad \text{for all } k \text{ under consideration.} \tag{4.86}$$

This gives nothing but the required estimate (4.81) since in view of (4.65), equality in the above upper-bound is achieved for $k = u_L$ and $k = u_R$. Next assume some $k_\star$ in $\lfloor u_L, u_R \rceil$ with the property $\mathcal{K}(k_\star; u_L, u_R) > 0$. Let us check that any given monotonicity preserving mapping $\theta(u_L, u_R)$ cannot achieve the value 1 for the pair (u_L, u_R) under consideration. Assuming there exists one such mapping then the above steps would apply to infer

$$\mathcal{R}_-(k_\star) + \mathcal{R}_+(k_\star) < |1 - \mathcal{R}_-(k_\star)| + |1 - \mathcal{R}_+(k_\star)|, \quad \textit{i.e. } \mathcal{R}_-(k_\star) + \mathcal{R}_+(k_\star) < 1, \tag{4.87}$$

and this would result in a contradiction with the condition $1 = \theta(u_L, u_R) \leq \Theta(u_L, u_R)$ in view of the definition (4.63) of $\Theta(u_L, u_R)$. As a consequence, no monotonicity preserving mapping can reach the value 1 for the pair under consideration and we necessarily have $\Theta(u_L, u_R) < 1$. □

We conclude this section proving Theorem 11.

Proof of Theorem 11 : we start assuming that the mappings $\theta(u_L, u_R)$ under consideration are monotonicity preserving and we will prove that the resulting conditions actually imply this property. Let us define

$$\mathcal{H}(k, \theta) \equiv \mathcal{R}_-(k) + \mathcal{R}_+(k) - |\theta - \mathcal{R}_-(k)| - |\theta - \mathcal{R}_+(k)| \tag{4.88}$$

Limiting the values of θ such that $\mathcal{E}\{\mathcal{U}_k\}(\theta; u_L, u_R) \leq 0$ for all $k \in \lfloor u_L, u_R \rceil$ is equivalently to find θ with the property

$$\mathcal{H}(k, \theta) \geq 0, \quad \text{for all } k \in \lfloor u_L, u_R \rceil. \tag{4.89}$$

The identity

$$\begin{aligned} \mathcal{H}(k, \theta) &= \min(\mathcal{R}_-(k), \mathcal{R}_+(k)) + \max(\mathcal{R}_-(k), \mathcal{R}_+(k)) \\ &\quad - |\theta - \min(\mathcal{R}_-(k), \mathcal{R}_+(k))| - |\theta - \max(\mathcal{R}_-(k), \mathcal{R}_+(k))| \end{aligned} \tag{4.90}$$

then yields

$$\mathcal{H}(k,\theta) = \begin{cases} 2\theta, & 0 \le \theta \le \min(\mathcal{R}_-(k), \mathcal{R}_+(k)), \\ 2\min(\mathcal{R}_-(k), \mathcal{R}_+(k)), & \min(\mathcal{R}_-(k), \mathcal{R}_+(k)) \le \theta \le \max(\mathcal{R}_-(k), \mathcal{R}_+(k)), \\ 2\big(\mathcal{R}_-(k) + \mathcal{R}_+(k) - \theta\big), & \max(\mathcal{R}_-(k), \mathcal{R}_+(k)) \le \theta. \end{cases} \tag{4.91}$$

Since by assumption $\theta \ge 0$ while $\mathcal{R}_-(k)$ and $\mathcal{R}_+(k)$ are known to keep non-negative values for all $k \in \lfloor u_L, u_R \rceil$, the condition (4.89) resumes to

$$\theta \le \mathcal{R}_-(k) + \mathcal{R}_+(k), \quad \text{for all } k \in \lfloor u_L, u_R \rceil \text{ such that } \max(\mathcal{R}_-(k), \mathcal{R}_+(k)) \le \theta. \tag{4.92}$$

As already discussed, a strengthened version, but essentially similar to our main motivation, is

$$\theta \le \mathcal{R}_-(k) + \mathcal{R}_+(k), \quad \text{for all } k \in \lfloor u_L, u_R \rceil. \tag{4.93}$$

To conclude, observe that fulfilling the condition (4.89) from the equivalent form (4.90) in fact requires $\theta \ge 0$ since otherwise (recall that $\mathcal{R}_\pm(k)$ keep non-negative values), one would draw a contradiction with

$$\mathcal{H}(k,\theta) = 2\theta \ge 0. \tag{4.94}$$

Then the condition (4.92) is again in order so that $\theta \le 1$ in view of the identities (4.65): indeed one must simultaneously verify $\theta \le (\mathcal{R}_- + \mathcal{R}_+)(u_L) = 1$ and $1 = \max(\mathcal{R}_-(u_L), \mathcal{R}_+(u_L)) \le \theta$. Requiring $\mathcal{E}\{\mathcal{U}_k\}(\theta; u_L, u_R) \le 0$ for all $k \in \lfloor u_L, u_R \rceil$ is thus equivalent to the monotonicity preserving condition (4.2). This ends the proof.

5 The numerical approximation procedure

This section describes first order numerical methods for approximating the Kružkov solutions of a scalar conservation law, built from the Riemann solver with defect measure correction we have derived in the first part of this paper. From now on, we tacitly assume that the monitoring mapping $\theta(u_L, u_R)$ involved in the defect measures is monotonicity preserving and consistent with the entropy requirement(s) we have put forward. Convergence of the family of approximate solutions to the Kružkov solution will be proved in the next section.

We propose hereafter two variants of finite volume methods built from Riemann problems involving defect measure corrections. The first numerical method stays in the very spirit of the Glimm's approach and is directly built from a sequence of non-interacting Riemann solutions whose values are sampled in each cell. The second method is more in the spirit of Godunov's method and relies before the sampling procedure on suitable local averagings of two neighboring Riemann solutions that avoid any of the discontinuities with speed σ to prevent them from smearing. Both strategies intend to restore at the discrete level the exactness property highlighted in Lemma 1. To this aim, a relevant choice for the monotonicity preserving and entropy satisfying mappings θ is given by $\Theta(u_L, u_R)$, namely either given by (4.24)–(4.26) in the case of genuinely non-linear flux function or by (4.63) for a general flux.

Introduce the spatial grid points $x_{j+\frac{1}{2}}$ with uniform mesh width $\Delta x = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$. The discrete time level t^n is also spaced uniformly with time step $\Delta t = t^{n+1} - t^n$ and satisfy the strict CFL condition

$$a\frac{\Delta t}{\Delta x} < \frac{1}{2}, \tag{5.1}$$

where the sub-characteristic condition is specified as follows

$$\sup_{|u|<\|u_0\|_{L^\infty(\mathbb{R})}} |f'(u)| < a. \tag{5.2}$$

The numerical approximate solution $\mathbb{U}^\alpha(t^n, x)$ is sought for as a piecewise constant function whose components are denoted by

$$u^\alpha_{\Delta x}(t^n, x) = u^n_j, \;\; v^\alpha_{\Delta x}(t^n, x) = v^n_j, \quad x_{j-\frac{1}{2}} < x < x_{j+\frac{1}{2}}, \tag{5.3}$$

Here α refers to the random sequence used in the Glimm's sampling procedure. The initial data is discretized in a well prepared manner

$$u^0_{\Delta x}(x) = \frac{1}{\Delta x}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u_0(x)dx, \quad v^0_{\Delta x}(x) = f(u^0_{\Delta x}(x)), \quad x \in (x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}), \; j \in \mathbb{Z}. \tag{5.4}$$

5.1 The first algorithm

Assuming the piecewise constant approximate solution $\mathbb{U}(t^n, x)$ to be known at time t^n, we propose to evolve it to the next time level t^{n+1} in three steps.

$\star$ *Step 1: $t^n \to t^{n+1,(1)} \equiv (n+1)\Delta t^=$, Riemann problems with defect measure correction.* Solve the Cauchy problem exactly in the slab $(t^n, t^n + \Delta t)$

$$\begin{cases} \partial_t u + \partial_x v = 0, \\ \partial_t v + a^2 \partial_x u = \mathcal{M}(u^\alpha_{\Delta x}(t^n, x), v^\alpha_{\Delta x}(t^n, x)), \end{cases} \tag{5.5}$$

with initial data

$$u(0, x) = u^\alpha_{\Delta x}(t^n, x), \;\; v(0, x) = v^\alpha_{\Delta x}(t^n, x). \tag{5.6}$$

Here $\mathcal{M}$ is a bounded Borel measure which collects all successive defect measure corrections, i.e.,

$$\begin{aligned} \mathcal{M}(u^\alpha_{\Delta x}(t^n, x), v^\alpha_{\Delta x}(t^n, x)) \quad &= \Theta(u^n_j, u^n_{j+1})\left(a^2 - \sigma^2(u^n_j, u^n_{j+1})\right) & (5.7) \\ &(u^n_{j+1} - u^n_j)\delta_{(x - x_{j+1/2}) - \sigma(u^n_j, u^n_{j+1})(t - t^n)} & (5.8) \end{aligned}$$

for $x \in (x_j, x_{j+1})$ and $t \in (t^n, t^n + \Delta t)$. Under the CFL condition (5.1), the exact solution of (5.5)-(5.6) is the gluing of a sequence of noninteracting self-similar solutions :

$$\left(\mathbb{U}^\alpha_{\Delta x}\right)^{(1)}(t, x) := \mathbb{U}\left(\frac{x - x_{j+\frac{1}{2}}}{t - t^n}; u_j, u_{j+1}\right); \quad x \in [x_j, x_{j+1}], \; t^n < t < t^{n+1}, \tag{5.9}$$

as defined in Lemma 2. Thus the solution at the first "intermediate" time reads

$$\left(\mathbb{U}^\alpha_{\Delta x}\right)^{(1)}\left(t^{n+1}, x\right) = \mathbb{U}\left(\frac{x - x_{j+\frac{1}{2}}}{\Delta t}; u^n_j, u^n_{j+1}\right), \;\; x \in [x_j, x_{j+1}]. \tag{5.10}$$

⋆ *Step 2: $t^{n+1,(1)} \to t^{n+1,(2)} \equiv (n+1)\Delta t^-$, Pointwise relaxation.* From the solution of Cauchy problem (5.5)-(5.6), define at the second step $t^{n+1,(2)}$ pointwisely for $x \in (x_{j-1/2}, x_{j+1/2})$

$$\begin{aligned} {u_{\Delta x}^{\alpha}}^{(2)}\left(t^{n+1}, x\right) &= {u_{\Delta x}^{\alpha}}^{(1)}\left(t^{n+1}, x\right), &(5.11)\\ {v_{\Delta x}^{\alpha}}^{(2)}\left(t^{n+1}, x\right) &= f({u_{\Delta x}^{\alpha}}^{(2)}\left(t^{n+1}, x\right)). &(5.12)\end{aligned}$$

⋆ *Step 3: $t^{n+1,(2)} \to t^{n+1,(3)} \equiv t^{n+1}$, Sampling.* Draw a random number α_n from an equidistributed sequence in $(0,1)$, we define in each cells a constant value $\mathbb{U}_j^{n+1}$ following the Glimm's sampling strategy

$$\begin{aligned} u_{\Delta x}^{\alpha}(t^{n+1}, x) &= {u_{\Delta x}^{\alpha}}^{(2)}\left(t^{n+1}, x_{j-\frac{1}{2}} + \alpha_n \Delta x\right), \quad x \in [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}], &(5.13)\\ v_{\Delta x}^{\alpha}(t^{n+1}, x) &= {v_{\Delta x}^{\alpha}}^{(2)}\left(t^{n+1}, x_{j-\frac{1}{2}} + \alpha_n \Delta x\right) = f(u_{\Delta x}^{\alpha}(t^{n+1}, x)). &(5.14)\end{aligned}$$

This concludes the description of the method.

We summarize the first algorithm as follows. To shorten the notations, let us set

$$\sigma_{j+\frac{1}{2}}^n = \sigma(u_j^n, u_{j+1}^n). \tag{5.15}$$

Being given the random number $\alpha_n \in (0,1)$, define in each cell $(x_{j-1/2}, x_{j+1/2})$

- Update u_j^{n+1} from the $\{u_j^n\}_{j\in\mathbb{Z}}$

$$u_j^{n+1} = \begin{cases} u_L^\star(\theta; u_{j-1}^n, u_j^n), & \alpha_n < \sigma_{j-\frac{1}{2}}^n \frac{\Delta t}{\Delta x},\\[2ex] u_R^\star(\theta; u_{j-1}^n, u_j^n), & \sigma_{j-\frac{1}{2}}^n \frac{\Delta t}{\Delta x} \le \alpha_n < a_{j-1/2}^n \frac{\Delta t}{\Delta x},\\[2ex] u_j^n, & a_{j-\frac{1}{2}}^n \frac{\Delta t}{\Delta x} \le \alpha_n < 1 - a_{j+\frac{1}{2}}^n \frac{\Delta t}{\Delta x},\\[2ex] u_L^\star(\theta; u_j^n, u_{j+1}^n), & 1 - a_{j+\frac{1}{2}}^n \frac{\Delta t}{\Delta x} \le \alpha_n < 1 + \sigma_{j+\frac{1}{2}}^n \frac{\Delta t}{\Delta x},\\[2ex] u_R^\star(\theta; u_j^n, u_{j+1}^n), & 1 + \sigma_{j+\frac{1}{2}}^n \frac{\Delta t}{\Delta x} \le \alpha_n, \end{cases} \tag{5.16}$$

  where $u_L^\star$, $u_R^\star$ are defined in (3.3)–(3.4),

- Update $v_j^{n+1} = f(u_j^{n+1})$.

5.2 The second algorithm

Given the piecewise constant approximate solution $\mathbb{U}(t^n, x)$ at time t^n, we propose to update it to the next time level t^{n+1} in four steps from the intermediate times $t^n \to t^{n+1,(1)}$ to $t^{n+1,(3)} \to t^{n+1,(4)} \equiv t^{n+1}$. Three of these steps are virtually kept unchanged from the first numerical algorithm but are performed at (possibly) distinct intermediate times. We do not repeat the details of those steps and we explicitly refer hereafter the reader to the formulas described in the first algorithm while we clear up each of the corresponding intermediate times.

Namely we first solve a sequence of non-interacting Riemann problems with defect measure corrections (5.6) to define ${u^\alpha_{\Delta x}}^{(1)}(t,x), {v^\alpha_{\Delta x}}^{(1)}(t,x)$ from $\mathbb{U}(t^n, x)$. As a second step, we propose to perform local averaging of ${u^\alpha_{\Delta x}}^{(1)}(t^{n+1},x)$ to define ${u^\alpha_{\Delta x}}^{(2)}(t^{n+1},x)$. In contrast to the usual Godunov's approach, two neighboring Riemann solutions $\mathbb{U}((x-x_{j-1/2})/\Delta t, u^n_{j-1}, u^n_j)$ and $\mathbb{U}((x-x_{j+1/2})/\Delta t, u^n_j, u^n_{j+1})$ with x in $(x_{j-1/2}, x_{j+1/2})$ are not averaged within the cell under consideration. Instead, local averagings of neighboring Riemann solutions are performed over distinct intervals of the form $(x^{n+1}_{j-\frac{1}{2}}, x^{n+1}_{j+\frac{1}{2}})$ with length $\Delta x^{n+1}_j = x^{n+1}_{j+\frac{1}{2}} - x^{n+1}_{j-\frac{1}{2}}$ and boundaries defined by

$$x^{n+1}_{j+\frac{1}{2}} = x_{j+1/2} + \sigma(u^n_j, u^n_{j+1})\Delta t. \tag{5.17}$$

Clearly $x^{n+1}_{j+\frac{1}{2}}$ is nothing but the location of the intermediate discontinuity in $\mathbb{U}((x-x_{j+1/2})/\Delta t, u^n_j, u^n_{j+1})$ propagating with speed $\sigma(u^n_j, u^n_{j+1})$ and is thus located at time $t^{n+1,(2)}$ either in $(x_{j-1/2}, x_{j+1/2})$ or in $(x_{j+1/2}, x_{j+3/2})$ depending on the sign of the velocity under consideration. The proposed local averagings are thus given by

$$u^{n+1,(2)}_j = \frac{1}{x^{n+1}_{j+\frac{1}{2}} - x^{n+1}_{j-\frac{1}{2}}} \int_{x^{n+1}_{j-\frac{1}{2}}}^{x^{n+1}_{j+\frac{1}{2}}} {u^\alpha_{\Delta x}}^{(1)}(t^{n+1}, x)dx, \quad j \in \mathbb{Z}. \tag{5.18}$$

This choice precisely avoids any of the intermediate waves in order to prevent them from numerical smearing. In contrast to the first algorithm, the discrete solution ${u^\alpha_{\Delta x}}^{(2)}(t^{n+1},x)$ entering the last step is no longer made of up to five constant states within $(x_{j-1/2}, x_{j+1/2})$ but only up to three in the situation $\sigma(u^n_{j-1}, u^n_j) > 0$ and $\sigma(u^n_j, u^n_{j+1}) < 0$. Observe that the averaging (5.18) can be given the following form

$$u^{n+1,(2)}_j = \frac{\Delta x}{\Delta x^{n+1}_j} u^n_j - \frac{\Delta t}{\Delta x^{n+1}_j}\big(g^n_{j+1/2} - g^n_{j-1/2}\big), \quad j \in \mathbb{Z}, \tag{5.19}$$

with $g^n_{j+1/2} = g(u^n_j, u^n_{j+1})$ given by the 2-point numerical flux function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by

$$g(u_L, u_R) = v^\star_R(\theta; u_L, u_R) - \sigma(u_L, u_R) u^\star_R(\theta; u_L, u_R), \quad (u_L, u_R) \in \mathbb{R}^2. \tag{5.20}$$

Observe that the proposed definition (5.20) indeed results in a conservative finite volume scheme (5.19) in view of the identity inferred from the first jump condition in (2.21)

$$v^\star_R(\theta; u_L, u_R) - \sigma(u_L, u_R) u^\star_R(\theta; u_L, u_R) = v^\star_L(\theta; u_L, u_R) - \sigma(u_L, u_R) u^\star_L(\theta; u_L, u_R). \tag{5.21}$$

In this second step and for technical simplicity, the v-component is conveniently locally averaged, mimicking the u-component

$$v^{n+1,(2)}_j = \frac{1}{x^{n+1}_{j+\frac{1}{2}} - x^{n+1}_{j-\frac{1}{2}}} \int_{x^{n+1}_{j-\frac{1}{2}}}^{x^{n+1}_{j+\frac{1}{2}}} {v^\alpha_{\Delta x}}^{(1)}(t^{n+1}, x)dx, \quad j \in \mathbb{Z}. \tag{5.22}$$

As a third step, we operate a pointwise relaxation (5.11) and get

$${v^\alpha_{\Delta x}}^{(3)}(t^{n+1}, x) = f({u^\alpha_{\Delta x}}^{(3)}(t^{n+1}, x)), \quad \text{with } {u^\alpha_{\Delta x}}^{(3)}(t^{n+1}, x) = {u^\alpha_{\Delta x}}^{(2)}(t^{n+1}, x).$$

Obviously and in practice, this third step makes useless the local averagings proposed for the v-component in (5.22). But again, adopting the rather formal step (5.22) turns to be convenient in the forthcoming analysis.

Within each cell $(x_{j-1/2}, x_{j+1/2})$, we derive the final update $u^{\alpha}_{\Delta x}(t^{n+1}, x)$ using a sampling procedure (5.13) performed on the piecewise constant function $u^{\alpha}_{\Delta x}{}^{(3)}(t^{n+1}, x)$. The pointwise relaxation step ensures $v^{\alpha}_{\Delta x}(t^{n+1}, x) = f(u^{\alpha}_{\Delta x}(t^{n+1}, x))$. This concludes the description of the method. We summarize the second algorithm as follows.

Being given the random number $\alpha_n \in (0,1)$, define in each cell $(x_{j-1/2}, x_{j+1/2})$

- Update u^{n+1}_j from the $\{u^{n+1,(2)}_j\}_{j\in\mathbb{Z}}$ with $u^{n+1,(2)}_j$ given in (5.19)

$$u^{n+1}_j = \begin{cases} u^{n+1,(2)}_{j-1}, & \alpha_n < \sigma^n_{j-\frac{1}{2}}\frac{\Delta t}{\Delta x}, \\ u^{n+1,(2)}_{j}, & \sigma^n_{j-\frac{1}{2}}\frac{\Delta t}{\Delta x} \leq \alpha_n < 1 + \sigma^n_{j+\frac{1}{2}}\frac{\Delta t}{\Delta x}, \\ u^{n+1,(2)}_{j+1}, & 1 + \sigma^n_{j+\frac{1}{2}}\frac{\Delta t}{\Delta x} \leq \alpha_n, \end{cases} \tag{5.23}$$

- Update $v^{n+1}_j = f(u^{n+1}_j)$.

6 Convergence to the Kružkov entropy weak solution

In this section, we prove for both the finite volume methods (5.1) and (5.2) that the family of discrete solutions $\{\mathbb{U}^{\alpha}_{\Delta x}\}_{\Delta x>0}$ converges as Δx goes to zero to $\mathbb{U} = \big(u, f(u)\big)$ where u is the Kružkov solution of the Cauchy problem for (2.1) with initial data $u_0 \in L^{\infty}(\mathbb{R}) \cap BV(\mathbb{R})$. The main result is as follows.

Theorem 15. *Given $u_0 \in L^{\infty}(\mathbb{R}) \cap BV(\mathbb{R})$. Assume the sub-characteristic condition (5.2) and the CFL condition (5.1). Assume that the mapping $\theta(u_L, u_R)$ is monotonicity preserving (4.2) and consistent with the entropy condition (4.21), namely with the quadratic entropy pair in the case of a genuinely non-linear flux and with the whole Kružkov family in the case of a general non-linear flux function. Then for almost any given sampling sequence $\alpha = (\alpha_1, \alpha_2, ...) \in (0,1)^{\mathbb{N}}$, the family of approximate solutions $\{u^{\alpha}_{\Delta x}\}_{\Delta x>0}$ given either by (5.1) or (5.2) converges in $L^{\infty}\big((0,T), L^1_{loc}(\mathbb{R})\big)$ for all $T > 0$ and a.e. as $\Delta x \to 0$ with $\frac{\Delta t}{\Delta x}$ kept fixed, to the Kružkov solution of the corresponding Cauchy problem (2.1).*

By almost any given sampling sequences, it is meant, as identified by T.P. Liu ,that relevant sequences have to be equi-distributed (see [26] for instance). The proof of this statement relies on the following first result.

Proposition 16. *Assume the sub-characteristic condition (5.2) and the CFL condition (5.1). Assume that the mapping $\theta(u_L, u_R)$ is monotonicity preserving. Given any sampling sequence $\alpha = (\alpha_1, \alpha_2, ...) \in (0,1)^{\mathbb{N}}$, the sequence of discrete solutions $(u^{\alpha}_{\Delta x}(t,x), v^{\alpha}_{\Delta x}(t,x))_{\Delta t>0}$ given* *either by (5.1) or (5.2) satisfies the following uniform in Δx a priori estimates for all time $t > 0$.*

$$(i) \quad \| u^{\alpha}_{\Delta x}(t,\cdot) \|_{L^\infty(\mathbb{R})} \leq \| u_0 \|_{L^\infty(\mathbb{R})}, \;\; \| v^{\alpha}_{\Delta x}(t,\cdot) \|_{L^\infty(\mathbb{R})} \leq a \, \| u_0 \|_{L^\infty(\mathbb{R})}, \tag{6.1}$$

$$(ii) \quad TV(u^{\alpha}_{\Delta x}(t,.)) \leq TV(u_0), \;\; TV(v^{\alpha}_{\Delta x}(t,.)) \leq a\,TV(u_0), \tag{6.2}$$

$$(iii) \int_{\mathbb{R}} \left| u^{\alpha}_{\Delta x}{}^{(1)}(t,x) - u^{\alpha}_{\Delta x}(t^n,x) \right| dx \leq a\,TV(u_0)(t-t^n), \;\; t^n \leq t \leq t^{n+1}, \tag{6.3}$$

$$(iv) \int_{\mathbb{R}} \left| v^{\alpha}_{\Delta x}{}^{(1)}(t,x) - f(u^{\alpha}_{\Delta x}(t^n,x)) \right| dx \leq 2a^2\,TV(u_0)(t-t^n), \;\; t^n \leq t \leq t^{n+1}. \tag{6.4}$$

Proof. The proposed estimates are established within the frame of the second algorithm (5.2). Their derivation concerning the simpler first method (5.1) follows from virtually identical steps. Details are left to the reader.

(i) The sup-norm estimate in (6.1) follows from the corresponding local maximum principle stated in (4.1), Theorem 4, which is valid in the first step:

$$\sup_{x_j \leq x \leq x_{j+1}} |u^{\alpha}_{\Delta x}{}^{(1)}(t,x)| = \sup_{x_j \leq x \leq x_{j+1}} \left| u\left(\frac{x - x_{j+\frac{1}{2}}}{t - t^n}; u^n_j, u^n_{j+1} \right) \right| \leq \max(|u^n_j|, |u^n_{j+1}|), \tag{6.5}$$

for all $j \in \mathbb{Z}$ and $t \in (t^n, t^{n+1})$, and as a consequence

$$\| u^{\alpha}_{\Delta x}{}^{(1)} \left(t^{n+1}, \cdot\right) \|_{L^\infty(\mathbb{R})} \leq \| u^{\alpha}_{\Delta x}(t^n, \cdot) \|_{L^\infty(\mathbb{R})} \,. \tag{6.6}$$

As is well-known, the local averagings involved in the second step diminishes the sup-norm

$$\| u^{\alpha}_{\Delta x}{}^{(2)} \left(t^{n+1}, \cdot\right) \|_{L^\infty(\mathbb{R})} \leq \| u^{\alpha}_{\Delta x}{}^{(1)} \left(t^{n+1}, \cdot\right) \|_{L^\infty(\mathbb{R})} \,. \tag{6.7}$$

The third step devoted to pointwise relaxation does not change the u-component of the discrete solution, and the sampling procedure in the last step decreases the sup-norm, so that

$$\| u^{\alpha}_{\Delta x} \left(t^{n+1}, \cdot\right) \|_{L^\infty(\mathbb{R})} \leq \| u^{\alpha}_{\Delta x}{}^{(3)} \left(t^{n+1}, \cdot\right) \|_{L^\infty(\mathbb{R})} \leq \| u^{\alpha}_{\Delta x}{}^{(2)} \left(t^{n+1}, \cdot\right) \|_{L^\infty(\mathbb{R})} \leq \| u^{\alpha}_{\Delta x} \left(t^n, \cdot\right) \|_{L^\infty(\mathbb{R})} \,. \tag{6.8}$$

This immediately implies the expected uniform sup-norm estimate in view of the definition (5.4) of the discrete initial data. The derivation of the companion sup-norm estimate for $v^{\alpha}_{\Delta x}(t, \cdot)$ starts from the local estimate (4.4)

$$\sup_{x_j \leq x \leq x_{j+1}} |v^{\alpha}_{\Delta x}{}^{(1)}(t,x)| = \sup_{x_j \leq x \leq x_{j+1}} \left| v\left(\frac{x - x_{j+\frac{1}{2}}}{t - t^n}; u^n_j, u^n_{j+1} \right) \right| \leq a \max(|u^n_j|, |u^n_{j+1}|), \tag{6.9}$$

for all $j \in \mathbb{Z}$ and $t \in (t^n, t^{n+1})$, so that

$$\| v^{\alpha}_{\Delta x}{}^{(1)} \left(t^{n+1}, \cdot\right) \|_{L^\infty(\mathbb{R})} \leq \; a \, \| u^{\alpha}_{\Delta x}(t^n, \cdot) \|_{L^\infty(\mathbb{R})} \,. \tag{6.10}$$

Then in the third step, $v^{\alpha}_{\Delta x}$ is set at equilibrium pointwisely in x, and we get from estimate (6.6)

$$\| v^{\alpha}_{\Delta x}{}^{(3)} \left(t^{n+1}, \cdot\right) \|_{L^\infty(\mathbb{R})} = \| f(u^{\alpha}_{\Delta x}{}^{(2)} \left(t^{n+1}, \cdot\right)) \|_{L^\infty(\mathbb{R})} \leq a \, \| u^{\alpha}_{\Delta x} \left(t^n, \cdot\right)) \|_{L^\infty(\mathbb{R})} \;\;. \tag{6.11}$$

At last the sampling procedure does not increase the sup-norm of $v^\alpha_{\Delta x}$ so that

$$\| \, v^\alpha_{\Delta x} \left(t^{n+1}, \cdot \right) \|_{L^\infty(\mathbb{R})} \leq a \, \| \, u^\alpha_{\Delta x} \left(t^n, \cdot \right)) \, \|_{L^\infty(\mathbb{R})} \leq a \, \| \, u_0 \, \|_{L^\infty(\mathbb{R})} \quad . \tag{6.12}$$

(ii) In view of the local total variation estimate stated in (4.1), the first step gives

$$\mathrm{TV}_{]x_j,x_{j+1}[} \left({u^\alpha_{\Delta x}}^{(1)}(t,\cdot) \right) = \mathrm{TV} \left(u(\cdot; u^n_j, u^n_{j+1}) \right) \leq |u^n_{j+1} - u^n_j|, \;\; t^n \leq t \leq t^{n+1}. \tag{6.13}$$

Under the CFL condition (5.1), the discrete solution $u^\alpha_{\Delta x}(t,x)$ stays continuous at $x = x_j$ keeping the constant value u^n_j for all $t \in (t^n, t^{n+1,(1)})$, we infer

$$\mathrm{TV} \left({u^\alpha_{\Delta x}}^{(1)}(t,\cdot) \right) = \sum_{j\in\mathbb{Z}} \mathrm{TV}_{]x_j,x_{j+1}[} \left({u^\alpha_{\Delta x}}^{(1)}(t,\cdot) \right) \leq \sum_{j\in\mathbb{Z}} \left| u^n_{j+1} - u^n_j \right| = \mathrm{TV} \left(u^\alpha_{\Delta x}(t^n,\cdot) \right). \tag{6.14}$$

In the second step, $u^\alpha_{\Delta x}$ is locally averaged and its total variation decreases :

$$\mathrm{TV} \left({u^\alpha_{\Delta x}}^{(2)}(t^{n+1},\cdot) \right) \leq \mathrm{TV} \left({u^\alpha_{\Delta x}}^{(1)}(t^{n+1},\cdot) \right) \leq \mathrm{TV} \left(u^\alpha_{\Delta x}(t^n,\cdot) \right). \tag{6.15}$$

In the third step, $u^\alpha_{\Delta x}$ is kept unchanged and at last, the sampling procedure clearly diminishes the total variation, an immediate recursion gives the required uniform total variation estimate again from the definition (5.4) of the discrete initial data

$$\mathrm{TV}(u^\alpha_{\Delta x}(t^{n+1},\cdot)) \leq \mathrm{TV} \left(u^\alpha_{\Delta x}(t^n,\cdot) \right) \leq \mathrm{TV}(u^0_{\Delta x}) \leq \mathrm{TV}(u_0). \tag{6.16}$$

The estimate for $v^\alpha_{\Delta x}(t,\cdot)$ is derived similarly starting from the local estimate (4.4) for each self-similar solution to infer

$$\mathrm{TV}_{]x_j,x_{j+1}[} \left({v^\alpha_{\Delta x}}^{(1)}(t,\cdot) \right) \leq a \mathrm{TV} \left(u(\cdot; u^n_j, u^n_{j+1}) \right), \;\; t^n \leq t \leq t^{n+1}, \tag{6.17}$$

so that

$$\mathrm{TV} \left({v^\alpha_{\Delta x}}^{(1)}(t,\cdot) \right) \leq a \mathrm{TV} \left(u^\alpha_{\Delta x}(t^n,\cdot) \right), \;\; t^n \leq t \leq t^{n+1}. \tag{6.18}$$

In the second step, $v^\alpha_{\Delta x}$ is locally averaged according to (5.22) hence

$$\mathrm{TV} \left({v^\alpha_{\Delta x}}^{(2)}(t,\cdot) \right) \leq a \mathrm{TV} \left(u^\alpha_{\Delta x}(t^n,\cdot) \right), \;\; t^n \leq t \leq t^{n+1}, \tag{6.19}$$

and is then set at equilibrium in the third step

$$\mathrm{TV} \left({v^\alpha_{\Delta x}}^{(3)}(t^{n+1},\cdot) \right) = \mathrm{TV} \left(f({u^\alpha_{\Delta x}}^{(2)}(t^{n+1},\cdot)) \right) \leq a \mathrm{TV} \left(u^\alpha_{\Delta x}(t^n,\cdot) \right). \tag{6.20}$$

At last, the sampling procedure diminishes the total variation

$$\mathrm{TV}(v^\alpha_{\Delta x}(t^{n+1},\cdot)) \leq a \mathrm{TV} \left(u^\alpha_{\Delta x}(t^n,\cdot) \right) \leq a \mathrm{TV}(u_0). \tag{6.21}$$

(iii) Observe from the first step in the method the following identity which holds in the sense of the Radon measures

$$\partial_t {u^\alpha_{\Delta x}}^{(1)}(t,x) = -\partial_x {v^\alpha_{\Delta x}}^{(1)}(t,x), \;\; t \in (t^n, t^{n+1}). \tag{6.22}$$

Under the CFL condition (5.1), the total variation of the Radon measure $\partial_t u^\alpha_{\Delta x}$ can be bounded from above by

$$|\partial_t {u^\alpha_{\Delta x}}^{(1)}(t,x)|(\mathbb{R}_x) = \mathrm{TV}({v^\alpha_{\Delta x}}^{(1)}(t,.)) \leq a\mathrm{TV}(u_0) \tag{6.23}$$

so that, one can infer for $t \in (t^n, t^{n+1})$

$$\begin{aligned} \int_{\mathbb{R}_x} \left| {u^\alpha_{\Delta x}}^{(1)}(t,x) - u^\alpha_{\Delta x}(t^n,x)\right| dx &= \int_{t^n}^t |\partial_t {u^\alpha_{\Delta x}}^{(1)}(s,x)|(\mathbb{R}_x) ds & (6.24)\\ &\leq aTV(u_0)(t-t_n). & (6.25)\end{aligned}$$

(iv) The equation involving the defect measure correction reads for $t \in (t^n, t^{n+1})$ and $x \in (x_j, x_{j+1})$

$$\partial_t {v^\alpha_{\Delta x}}^{(1)} = -a^2 \partial_x {u^\alpha_{\Delta x}}^{(1)} + m(u^n_j, u^n_{j+1}) \delta_{x - x_{j+1/2} - \sigma(u^n_j, u^n_{j+1})(t-t^n)}, \tag{6.26}$$

and the quantities involved in the above identity are again regarded as Radon measures. The total variation of the Radon measure $\partial_t v^\alpha_{\Delta x}$ can be bounded by

$$\begin{aligned} |\partial_t {v^\alpha_{\Delta x}}^{(1)}(t,x)|(x_j, x_{j+1}) &\leq a^2 |\partial_x {u^\alpha_{\Delta x}}^{(1)}|(x_j,x_{j+1}) + |m(u^n_j, u^n_{j+1})| \\ &\leq a^2 |u^n_{j+1} - u^n_j| + (a^2 - \sigma^2)\left|u^n_{j+1} - u^n_j\right| \leq 2a^2 \left|u^n_{j+1} - u^n_j\right|. \qquad (6.27)\end{aligned}$$

Therefore by summation, (6.27) becomes, under the CFL condition (5.1)

$$|\partial_t {v^\alpha_{\Delta x}}^{(1)}(t,x)|(\mathbb{R}_x) \leq 2a^2 \mathrm{TV}\left({u^\alpha_{\Delta x}}^{(1)}(t^n,\cdot)\right) \leq 2a^2 \mathrm{TV}(u_0). \tag{6.28}$$

We deduce that for $t^n \leq t \leq t^{n+1}$

$$\int_{\mathbb{R}_x} \left| {v^\alpha_{\Delta x}}^{(1)}(t,x) - v^\alpha_{\Delta x}(t^n,x)\right| dx = \int_{t^n}^t |\partial_t {v^\alpha_{\Delta x}}^{(1)}(s,x)|(\mathbb{R}_x) ds \leq 2a^2 \mathrm{TV}(u_0)(t-t^n), \tag{6.29}$$

where by construction $v^\alpha_{\Delta x}(t^n,x) = f(u^\alpha_{\Delta x}(t^n,x))$. This concludes the proof. □

This proposition immediately implies the following convergence result.

Corollary 17. *Given $u_0 \in L^\infty \cap BV(\mathbb{R})$, any $T > 0$, then under the assumptions of Proposition 16, there exists an extracted subsequence still denoted by $\{u^\alpha_{\Delta x}\}_{\Delta x > 0}$ which converges, as $\Delta x \to 0$ with $\Delta t/\Delta x$ kept constant, to a limit u^α in $L^\infty\big((0,T), L^1_{loc}(\mathbb{R})\big)$. In addition, the limit u^α belongs to $L^\infty(\mathbb{R}_+, L^\infty \cap BV(\mathbb{R}))$.*

Proof. This proof is rather classical from the uniform estimates stated in Proposition 16, and one can refer for instance to [9] (Theorems 3.3 and 3.4, Chapter 3). □

The above corollary guarantees the existence of a limit. We now characterize this limit, showing that it is indeed the unique entropy weak solution of the original Cauchy problem (2.1). The proof mainly relies on the relaxation entropy inequalities inherited from the first step shared by both methods (5.1) and (5.2). Those are conveniently localized within each of the following time-space domains :

$$\begin{aligned}\mathcal{D}_j^n = \Big\{(t,x) \in \mathbb{R}^+ \times \mathbb{R} /\ t \in (t^n, t^{n+1}), \quad & x_{j-1/2}^n(t) < x < x_{j+1/2}^n(t), \\ & x_{j+1/2}^n(t) = x_{j+1/2} + \sigma_{j+1/2}^n (t - t^n)\Big\}.\end{aligned} \tag{6.30}$$

Observe that $x_{j+1/2}^n(t^{n+1})$ coincides with $x_{j+1/2}^{n+1}$ defined in (5.17). We state :

Lemma 18. *Under the assumptions of Proposition 16, the approximate solutions given in the first step either by (5.1) or (5.2) satisfy the following relaxation entropy equalities in the sense of the distributions*

$$\partial_t \Phi({u_{\Delta x}^\alpha}^{(1)}, {v_{\Delta x}^\alpha}^{(1)}) + \partial_x \Psi({u_{\Delta x}^\alpha}^{(1)}, {v_{\Delta x}^\alpha}^{(1)}) = 0, \quad (t,x) \in \mathcal{D}_j^n, n \geq 0, j \in \mathbb{Z}. \tag{6.31}$$

Assume in addition that the mapping $\theta(u_L, u_R)$ is consistent with the entropy condition (4.21), namely with the quadratic entropy pair in the case of a genuinely non-linear flux and with the whole Kružkov entropy family in the case of a general non-linear flux function. Then the discrete solutions given either by (5.1) or (5.2) satisfy the corresponding entropy jump(s) at each boundary $x_{j+1/2}^n(t)$:

$$\begin{aligned}&-\sigma_{j+1/2}^n \Big(\Phi({u_{\Delta x}^\alpha}^{(1)}, {v_{\Delta x}^\alpha}^{(1)})(t, x_{j+1/2}^n(t)_+) - \Phi({u_{\Delta x}^\alpha}^{(1)}, {v_{\Delta x}^\alpha}^{(1)})(t, x_{j+1/2}^n(t)_-) \Big) \\ &+ \Big(\Psi({u_{\Delta x}^\alpha}^{(1)}, {v_{\Delta x}^\alpha}^{(1)})(t, x_{j+1/2}^n(t)_+) - \Psi({u_{\Delta x}^\alpha}^{(1)}, {v_{\Delta x}^\alpha}^{(1)})(t, x_{j+1/2}^n(t)_-) \Big) \leq 0, \quad t \in (t^n, t^{n+1}).\end{aligned} \tag{6.32}$$

Proof. Under the strict CFL condition (5.1), two neighboring Riemann solutions do not interact. We thus observe from the definition of each of the domain $\mathcal{D}_j^n$ that the solution $(u_{\Delta x}^\alpha, v_{\Delta x}^\alpha)$ is locally made of three constant states separated by the discontinuity lines $x_{j-1/2} + a(t - t^n)$ and $x_{j+1/2} - a(t - t^n)$. The property that the relaxation entropy is preserved across these two discontinuities (see indeed (4.20)) yields the expected equality (6.31). Next and for the mapping $\theta(u_L, u_R)$ under consideration, the jump inequality across each of the boundary $x_{j+1/2}^n(t)$ reads nothing but our entropy consistency requirement (4.21) stated in Definition 6. □

As a consequence, we get :

Proposition 19. *Assume the sub-characteristic condition (5.2) and the CFL condition (5.1). Given an entropy pair $(\mathcal{U}, \mathcal{F})$ (2.2) with $\mathcal{U}$ convex and its corresponding relaxation entropy pair (Φ, Ψ) (4.15)–(4.16). Assume that the mapping $\theta(u_L, u_R)$ is consistent with the entropy requirement (4.21) for all the pairs (u_L, u_R) under consideration. Then for any non-negative test function $\zeta \in C_0^1((0,\infty) \times \mathbb{R}_x)$, the approximate solutions $(u_{\Delta x}^\alpha, v_{\Delta x}^\alpha)$ given either by (5.1) or (5.2) satisfy*

$$\begin{aligned}\int_{x_{j-1/2}^{n+1}}^{x_{j+1/2}^{n+1}} \mathcal{U}\left(u_{\Delta x}^\alpha(t^{n+1}, x)\right) \zeta(t^{n+1}, x) dx - \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathcal{U}\left(u_{\Delta x}^\alpha(t^n, x)\right) \zeta(t^n, x) dx + \mathcal{G}_{j+1/2}^n - \mathcal{G}_{j-1/2}^n \\ - \iint_{\mathcal{D}_j^n} \Phi(u_{\Delta x}^\alpha, v_{\Delta x}^\alpha) \partial_t \zeta + \Psi(u_{\Delta x}^\alpha, v_{\Delta x}^\alpha) \partial_x \zeta dt dx \\ \leq (\mathcal{E}_A)_j^n (\Delta x, \alpha, \zeta) + (\mathcal{E}_S)_j^n (\Delta x, \alpha, \zeta), \quad n \geq 0,\ j \in \mathbb{Z}.\end{aligned} \tag{6.33}$$

Here, $\mathcal{G}^n_{j+1/2}$ stands for the time average of the right trace of the entropy flux along the boundary $x^n_{j+1/2}(t)$ and reads for both methods :

$$\mathcal{G}^n_{j+1/2} := \int_{t^n}^{t^{n+1}} \Big\{\Psi({u^\alpha_{\Delta x}}^{(1)}, {v^\alpha_{\Delta x}}^{(1)}) - \sigma^n_{j+1/2}\Phi({u^\alpha_{\Delta x}}^{(1)}, v^\alpha_{\Delta x}(1))\Big\}(t, x^n_{j+\frac{1}{2}}(t)_+)\zeta(t, x^n_{j+\frac{1}{2}}(t))dt. \tag{6.34}$$

Concerning the method (5.1), the error term $(\mathcal{E}_A)$ due to local averagings is identically zero while the error term $(\mathcal{E}_S)$ due to the sampling procedure is given by

$$\left(\mathcal{E}_S\right)^n_j(\Delta x, \alpha, \zeta) := \int_{x^{n+1}_{j-1/2}}^{x^{n+1}_{j+1/2}} \Big(\mathcal{U}\left(u^\alpha_{\Delta x}(t^{n+1}, x)\right) - \mathcal{U}\left({u^\alpha_{\Delta x}}^{(2)}(t^{n+1}, x)\right)\Big)\zeta(t^{n+1}, x)dx. \tag{6.35}$$

For the method (5.2), the error terms $(\mathcal{E}_A)$ and $(\mathcal{E}_S)$ respectively read

$$\left(\mathcal{E}_A\right)^n_j(\Delta x, \alpha, \zeta) := \int_{x^{n+1}_{j-1/2}}^{x^{n+1}_{j+1/2}} \Big(\Phi\left({u^\alpha_{\Delta x}}^{(2)}, {v^\alpha_{\Delta x}}^{(2)})(t^{n+1}, x)\right) - \Phi\left({u^\alpha_{\Delta x}}^{(1)}, {v^\alpha_{\Delta x}}^{(1)})(t^{n+1}, x)\right)\Big)\zeta(t^{n+1}, x)dx, \tag{6.36}$$

and

$$\left(\mathcal{E}_S\right)^n_j(\Delta x, \alpha, \zeta) := \int_{x^{n+1}_{j-1/2}}^{x^{n+1}_{j+1/2}} \Big(\mathcal{U}\left(u^\alpha_{\Delta x}(t^{n+1}, x)\right) - \mathcal{U}\left({u^\alpha_{\Delta x}}^{(3)}(t^{n+1}, x)\right)\Big)\zeta(t^{n+1}, x)dx. \tag{6.37}$$

Remark 20. In (6.33), the subscript (1) has been omitted in the notation of the (volume) integral over $\mathcal{D}^n_j$ since time discontinuities in the subsequent steps $t^{n+1,(1)}$, $t^{n+1,(2)}$, $t^{n+1,(3)}$ form a negligible set in the proposed Lebesgue integral.

Proof. The proposed inequality is proved for the second algorithm (5.2). Its derivation for the method (5.1) follows the same lines. Since again $(u^\alpha_{\Delta x}, v^\alpha_{\Delta x})$ is nothing but a piecewise constant solution of the entropy conservation law (6.31) over the domain $\mathcal{D}^n_j$, multiplying (6.31) by any given non-negative test function $\zeta \in C^1_0(\mathbb{R}^+_t \times \mathbb{R})$ and integrating over $(t, x) \in \mathcal{D}^n_j$ yield

$$\begin{aligned}
&\int_{x^{n+1}_{j-1/2}}^{x^{n+1}_{j+1/2}} \Phi({u^\alpha_{\Delta x}}^{(1)}, {v^\alpha_{\Delta x}}^{(1)})\left(t^{n+1}, x\right)\zeta(t^{n+1}, x)dx - \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \Phi(u^\alpha_{\Delta x}, v^\alpha_{\Delta x})(t^n, x)\zeta(t^n, x)dx \\
&\qquad + \int_{t^n}^{t^{n+1}} \Big\{\Psi({u^\alpha_{\Delta x}}^{(1)}, {v^\alpha_{\Delta x}}^{(1)}) - \sigma^n_{j+1/2}\Phi({u^\alpha_{\Delta x}}^{(1)}, {v^\alpha_{\Delta x}}^{(1)})\Big\}(t, x^n_{j+1/2}(t)_-)\zeta(t, x^n_{j+1/2}(t))dt \\
&\qquad - \int_{t^n}^{t^{n+1}} \Big\{\Psi({u^\alpha_{\Delta x}}^{(1)}, {v^\alpha_{\Delta x}}^{(1)}) - \sigma^n_{j-1/2}\Phi({u^\alpha_{\Delta x}}^{(1)}, {v^\alpha_{\Delta x}}^{(1)})\Big\}\zeta(t, x^n_{j-1/2}(t)_+)\zeta(t, x^n_{j-1/2}(t))dt \\
&\qquad - \iint_{\mathcal{D}^n_j} \Phi(u^\alpha_{\Delta x}, v^\alpha_{\Delta x})\partial_t\zeta + \Psi(u^\alpha_{\Delta x}, v^\alpha_{\Delta x})\partial_x\zeta dtdx = 0,
\end{aligned} \tag{6.38}$$

where the left and right traces at any given interface $x^n_{j+1/2}(t)$ are well defined since both $u^\alpha_{\Delta x}(t, .)$ and $v^\alpha_{\Delta x}(t, .)$ have uniformly bounded total variation in space. Using the definition

(6.34) of $\mathcal{G}^n_{j+1/2}$ evaluated on the right trace, inequality (6.38) can clearly be recast as

$$\int_{x^{n+1}_{j-1/2}}^{x^{n+1}_{j+1/2}} \Phi({u^\alpha_{\Delta x}}^{(1)}, {v^\alpha_{\Delta x}}^{(1)})\left(t^{n+1},x\right)\zeta(t^{n+1},x)dx - \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \Phi(u^\alpha_{\Delta x}, v^\alpha_{\Delta x})(t^n,x)\zeta(t^n,x)dx$$
$$+\mathcal{G}^n_{j+1/2} - \mathcal{G}^n_{j-1/2} - \iint_{\mathcal{D}^n_j} \Phi(u^\alpha_{\Delta x}, v^\alpha_{\Delta x})\partial_t\zeta + \Psi(u^\alpha_{\Delta x}, v^\alpha_{\Delta x})\partial_x\zeta dtdx = \mathcal{S}^n_j, \tag{6.39}$$

where in view of the jump inequality (6.32) established in Lemma 18, the right hand side is non-positive :

$$\begin{aligned}\mathcal{S}^n_j := \quad & \int_{t^n}^{t^{n+1}} \Big\{ -\sigma^n_{j+1/2}\Big(\Phi({u^\alpha_{\Delta x}}^{(1)}, {v^\alpha_{\Delta x}}^{(1)})(t, x^n_{j+1/2}(t)_+) - \Phi({u^\alpha_{\Delta x}}^{(1)}, {v^\alpha_{\Delta x}}^{(1)})(t, x^n_{j+1/2}(t)_-)\Big) \\ & +\Big(\Psi({u^\alpha_{\Delta x}}^{(1)}, {v^\alpha_{\Delta x}}^{(1)})(t, x^n_{j+1/2}(t)_+) - \Psi({u^\alpha_{\Delta x}}^{(1)}, {v^\alpha_{\Delta x}}^{(1)})(t, x^n_{j+1/2}(t)_-)\Big)\Big\}\zeta(t, x^n_{j+\frac{1}{2}}(t))dt \\ & \leq 0.\end{aligned} \tag{6.40}$$

Then by construction $v^\alpha_{\Delta x}(t^n,x) = f(u^\alpha_{\Delta x}(t^n,x))$ for all x in $(x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})$ so that the consistency condition (4.14) which links the entropy $\mathcal{U}$ to its relaxation lift Φ gives

$$\Phi(u^\alpha_{\Delta x}, v^\alpha_{\Delta x})(t^n,x) = \mathcal{U}(u^\alpha_{\Delta x}(t^n,x)),\ \ x \in (x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}),\ j \in \mathbb{Z}. \tag{6.41}$$

Hence in view of (6.39)–(6.40), we infer

$$\int_{x^{n+1}_{j-1/2}}^{x^{n+1}_{j+1/2}} \Phi({u^\alpha_{\Delta x}}^{(1)}, {v^\alpha_{\Delta x}}^{(1)})\left(t^{n+1},x\right)\zeta(t^{n+1},x)dx - \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathcal{U}(u^\alpha_{\Delta x}(t^n,x))\zeta(t^n,x)dx$$
$$+\mathcal{G}^n_{j+1/2} - \mathcal{G}^n_{j-1/2} - \iint_{\mathcal{D}^n_j} \Phi(u^\alpha_{\Delta x}, v^\alpha_{\Delta x})\partial_t\zeta + \Psi(u^\alpha_{\Delta x}, v^\alpha_{\Delta x})\partial_x\zeta dtdx \leq 0. \tag{6.42}$$

After the second step devoted to the local averagings (5.18)–(5.22), we thus deduce from (6.42) the following inequality

$$\int_{x^{n+1}_{j-1/2}}^{x^{n+1}_{j+1/2}} \Phi({u^\alpha_{\Delta x}}^{(2)}, {v^\alpha_{\Delta x}}^{(2)})(t^{n+1},x))\zeta(t^{n+1},x)dx - \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathcal{U}(u^\alpha_{\Delta x}(t^n,x))\zeta(t^n,x)dx$$
$$+\mathcal{G}^n_{j+1/2} - \mathcal{G}^n_{j+1/2} - \iint_{\mathcal{D}^n_j} \Phi(u^\alpha_{\Delta x}, v^\alpha_{\Delta x})\partial_t\zeta + \Psi(u^\alpha_{\Delta x}, v^\alpha_{\Delta x})\partial_x\zeta dtdx \tag{6.43}$$
$$\leq \left(\mathcal{E}_A\right)^n_j(\Delta x, \alpha, \zeta), \tag{6.44}$$

where $\left(\mathcal{E}_A\right)^n_j(\Delta x, \alpha, \zeta)$ denotes the local averaging error term defined in (6.36). Under the sub-characteristic condition (5.2), the Gibbs principle (4.23) established in Lemma 7 ensures that in the third step, the following inequality holds pointwisely in x

$$\mathcal{U}({u^\alpha_{\Delta x}}^{(3)})\left(t^{n+1},x\right) = \Phi\left({u^\alpha_{\Delta x}}^{(3)}, f({u^\alpha_{\Delta x}}^{(3)})\right)\left(t^{n+1},x\right) \leq \Phi({u^\alpha_{\Delta x}}^{(2)}, {v^\alpha_{\Delta x}}^{(2)})\left(t^{n+1},x\right), \tag{6.45}$$

where by construction ${u^\alpha_{\Delta x}}^{(3)}(t^{n+1},x) = {u^\alpha_{\Delta x}}^{(2)}(t^{n+1},x)$. The expected inequality (6.33) clearly holds true at the end of the last step devoted to the sampling procedure, with an additional error term given by (6.37). This concludes the proof. □

We are in a position to prove the convergence of the family of discrete solutions given either by (5.1) or (5.2) to the unique Kružkov solution of (2.1).

Proof of Theorem 15. For any given non-negative test function $\zeta \in C_0^1\left((0,\infty)\times\mathbb{R}_x\right)$, let us sum up the inequalities (6.33) for $j \in \mathbb{Z}$ to get

$$\begin{aligned}
&\int_{\mathbb{R}} \mathcal{U}(u_{\Delta x}^{\alpha})\left(t^{n+1},x\right)\zeta(t^{n+1},x)dx - \int_{\mathbb{R}} \mathcal{U}(u_{\Delta x}^{\alpha})(t^{n},x)\zeta(t^{n},x)dx \\
&- \int_{t^n}^{t^{n+1}}\int_{\mathbb{R}} \Phi(u_{\Delta x}^{\alpha},v_{\Delta x}^{\alpha})\partial_t\zeta + \Psi(u_{\Delta x}^{\alpha},v_{\Delta x}^{\alpha})\partial_x\zeta dtdx \leq \qquad (6.46)\\
&\qquad \sum_{j\in\mathbb{Z}}\Big(\left(\mathcal{E}_A\right)_j^n(\Delta x,\alpha,\zeta) + \left(\mathcal{E}_S\right)_j^n(\Delta x,\alpha,\zeta)\Big), \quad n \geq 0. \qquad (6.47)
\end{aligned}$$

Summing with respect to $n \in \mathbb{N}$ then yields

$$-\sum_{n\geq 0}\int_{t^n}^{t^{n+1}}\int_{\mathbb{R}} \Phi(u_{\Delta x}^{\alpha},v_{\Delta x}^{\alpha})\partial_t\zeta + \Psi(u_{\Delta x}^{\alpha},v_{\Delta x}^{\alpha})\partial_x\zeta dtdx - \int_{\mathbb{R}} \mathcal{U}(u_{\Delta x}^{\alpha})(0,x)\zeta(0,x)dx \leq \mathcal{E}_A + \mathcal{E}_S \tag{6.48}$$

where we have set

$$\mathcal{E}_A = \sum_{n\geq 0}\int_{\mathbb{R}}\Big(\Phi(\mathbb{U}_{\Delta x}^{\alpha}{}^{(2)}(t^{n+1},x)) - \Phi(\mathbb{U}_{\Delta x}^{\alpha}{}^{(1)}(t^{n+1},x))\Big)\zeta(t^{n+1},x)dx, \tag{6.49}$$

$$\mathcal{E}_S = \sum_{n\geq 0}\int_{\mathbb{R}}\Big(\mathcal{U}(u_{\Delta x}^{\alpha}(t^{n+1},x)) - \mathcal{U}(u_{\Delta x}^{\alpha}{}^{(3)}(t^{n+1},x))\Big)\zeta(t^{n+1},x)dx. \tag{6.50}$$

First, the dominated convergence theorem readily ensures from the definition of the discrete initial data

$$\int_{\mathbb{R}_x} \mathcal{U}(u_{\Delta x}^{0}(x))\zeta(0,x)dx \to \int_{\mathbb{R}_x} \mathcal{U}(u_0(x))\zeta(0,x)dx \quad \text{as } \Delta x \to 0. \tag{6.51}$$

Let us now prove that in the limit $\Delta x \to 0$ with $\Delta t/\Delta x$ kept constant

$$\sum_{n\geq 0}\int_{t^n}^{t^{n+1}}\int_{\mathbb{R}_x} \Phi(u_{\Delta x}^{\alpha},v_{\Delta x}^{\alpha})(t,x)\partial_t\zeta dtdx \to \int_{\mathbb{R}_t^+\times\mathbb{R}_x} \mathcal{U}(u^{\alpha})(t,x)\partial_t\zeta dtdx. \tag{6.52}$$

In that aim, we make use of the following triangle inequality

$$\begin{aligned}
&\left|\sum_{n\geq 0}\int_{t^n}^{t^{n+1}}\int_{\mathbb{R}_x}\Big(\Phi(u_{\Delta x}^{\alpha},v_{\Delta x}^{\alpha})(t,x) - \mathcal{U}(u^{\alpha})(t,x)\Big)\partial_t\zeta dtdx\right| \\
\leq\ & \sum_{n\geq 0}\int_{t^n}^{t^{n+1}}\int_{\mathbb{R}_x}\left|\Phi(u_{\Delta x}^{\alpha}(t,x),v_{\Delta x}^{\alpha}(t,x)) - \Phi\big(u_{\Delta x}^{\alpha}(t,x),f(u_{\Delta x}^{\alpha}(t^n,x))\big)\right||\partial_t\zeta|dtdx \\
&+\sum_{n\geq 0}\int_{t^n}^{t^{n+1}}\int_{\mathbb{R}_x}\left|\Phi\big(u_{\Delta x}^{\alpha}(t,x),f(u_{\Delta x}^{\alpha}(t^n,x))\big) - \Phi\big(u_{\Delta x}^{\alpha}(t,x),f(u_{\Delta x}^{\alpha}(t,x))\big)\right||\partial_t\zeta|dtdx \\
&+\sum_{n\geq 0}\int_{t^n}^{t^{n+1}}\int_{\mathbb{R}_x}\left|\Phi\big(u_{\Delta x}^{\alpha}(t,x),f(u_{\Delta x}^{\alpha}(t,x))\big) - \mathcal{U}(u^{\alpha})(t,x)\right||\partial_t\zeta|dtdx \\
:=\ & I_1 + I_2 + I_3.
\end{aligned} \tag{6.53}$$

Inequality (6.4) together with the sup norm estimate (6.1) yield

$$\begin{aligned}\int_{t^n}^{t^{n+1}} \int_{\mathbb{R}_x} \big|\Phi(u_{\Delta x}^{\alpha}, v_{\Delta x}^{\alpha})(t,x) \quad & -\Phi\big(u_{\Delta x}^{\alpha}(t,x), f(u_{\Delta x}^{\alpha}(t^n,x))\big)\big|\big|\partial_t \zeta dt\big| dx \\ & \leq C \parallel \partial_t \zeta \parallel_{L^\infty(\mathbb{R}_t^+ \times \mathbb{R}_x)} \int_{t^n}^{t^{n+1}} \int_{supp(\zeta(t,.))} |v_{\Delta x}^{\alpha}(t,x) - f(u_{\Delta x}^{\alpha}(t^n,x))| dx dt \\ & \leq C \int_{t^n}^{t^{n+1}} (s-t^n) ds \boldsymbol{I}_{supp(\zeta(t,.))} \\ & \leq C \Delta t^2 \boldsymbol{I}_{supp(\zeta(t,.))},\end{aligned} \tag{6.54}$$

where $\boldsymbol{I}_{supp(\zeta(t,.))}$ denotes the characteristic function of the test function $\zeta(t,.)$ at time t. Hence $\Delta t / \Delta x$ being kept constant, we infer

$$I_1 \leq C \Delta x, \tag{6.55}$$

where C is independent of Δx. Similarly we use (6.3) and (6.1) to get

$$I_2 \leq C \Delta x. \tag{6.56}$$

Concerning I_3, since the extracted subsequence $\{u_{\Delta x}^{\alpha}\}_{\Delta x > 0}$ is uniformly bounded in sup norm and converges to u^α in $L^\infty\big((0,T), L^1_{loc}(\mathbb{R})\big)$ for all $T > 0$ and a.e., the dominated convergence theorem applies to prove

$$\sum_{n \geq 0} \int_{t^n}^{t^{n+1}} \int_{\mathbb{R}_x} \Phi\left(u_{\Delta x}^{\alpha}, f(u_{\Delta x}^{\alpha})\right)(t,x) \partial_t \zeta dt dx = \int_{\mathbb{R}_t^+ \times \mathbb{R}_x} \mathcal{U}(u_{\Delta x}^{\alpha})(x,t) \partial_t \zeta dt dx \to \int_{\mathbb{R}_t^+ \times \mathbb{R}_x} \mathcal{U}(u^\alpha)(t,x) \partial_t \zeta dt dx \tag{6.57}$$

as $\Delta x \to 0$, so that I_3 vanishes in the reported limit. Exactly the same steps apply to show that

$$\sum_{n \geq 0} \int_{t^n}^{t^{n+1}} \int_{\mathbb{R}_x} \Psi\left(u_{\Delta x}^{\alpha}, f(u_{\Delta x}^{\alpha})\right)(t,x) \partial_x \zeta dt dx \to \int_{\mathbb{R}_t^+ \times \mathbb{R}_x} \mathcal{F}(u^\alpha) \partial_x \zeta dt dx, \text{ as } \Delta x \to 0. \tag{6.58}$$

Next, let us rewrite the averaging error term as follows

$$\mathcal{E}_A = \sum_{n \geq 0} \sum_{j \in \mathbb{Z}} (\mathcal{E}_A)_j^n, \quad (\mathcal{E}_A)_j^n = \int_{x_{j-1/2}^{n+1}}^{x_{j+1/2}^{n+1}} \left(\Phi\left({\mathbb{U}_{\Delta x}^{\alpha}}^{(2)}(t^{n+1},x)\right) - \Phi\left({\mathbb{U}_{\Delta x}^{\alpha}}^{(1)}(t^{n+1},x)\right)\right) \zeta(t^{n+1},x) dx. \tag{6.59}$$

Introducing the averaged quantity

$$\zeta_j^{n+1} = \frac{1}{x_{j+1/2}^{n+1} - x_{j-1/2}^{n+1}} \int_{x_{j-1/2}^{n+1}}^{x_{j+1/2}^{n+1}} \zeta(t^{n+1},x) dx, \tag{6.60}$$

with

$$||\zeta(t^{n+1},.) - \zeta_j^{n+1}||_{L^\infty((x_{j-1/2}^{n+1}, x_{j+1/2}^{n+1}))} \leq C \Delta x \boldsymbol{I}_{supp(\zeta)}, \tag{6.61}$$

the following identity holds

$$
\begin{aligned}
(\mathcal{E}_A)_j^n \quad &= \int_{x_{j-1/2}^{n+1}}^{x_{j+1/2}^{n+1}} \Big(\Phi\Big({\mathbb{U}_{\Delta x}^{\alpha}}^{(2)}(t^{n+1},x)\Big) - \Phi\Big({\mathbb{U}_{\Delta x}^{\alpha}}^{(1)}(t^{n+1},x)\Big)\Big)(\zeta(t^{n+1},x) - \zeta_j^{n+1})dx \qquad (6.62)\\
&+ \int_{x_{j-1/2}^{n+1}}^{x_{j+1/2}^{n+1}} \Big(\Phi\Big({\mathbb{U}_{\Delta x}^{\alpha}}^{(2)}(t^{n+1},x)\Big) - \Phi\Big({\mathbb{U}_{\Delta x}^{\alpha}}^{(1)}(t^{n+1},x)\Big)\, dx\ \zeta_j^{n+1} \qquad (6.63)\\
&:= (I_4)_j^{n+1} + (I_5)_j^{n+1}. \qquad (6.64)
\end{aligned}
$$

The convexity property of the entropy Φ ensures the following pointwise inequality

$$
\begin{aligned}
&\Phi\Big({\mathbb{U}_{\Delta x}^{\alpha}}^{(1)}(t^{n+1},x)\Big) - \Phi\Big({\mathbb{U}_{\Delta x}^{\alpha}}^{(2)}(t^{n+1},x)\Big)\\
&\qquad -\nabla\Phi\Big({\mathbb{U}_{\Delta x}^{\alpha}}^{(2)}(t^{n+1},x)\Big)\cdot\big({\mathbb{U}_{\Delta x}^{\alpha}}^{(1)}(t^{n+1},x) - {\mathbb{U}_{\Delta x}^{\alpha}}^{(2)}(t^{n+1},x)\big) \geq 0.
\end{aligned}
\tag{6.65}
$$

Here we make use of the local averaging procedure (5.18) together with the convenient definition (5.22) proposed in the second step, to get the identity

$$
{\mathbb{U}_{\Delta x}^{\alpha}}^{(2)}(t^{n+1},x) := \mathbb{U}_j^{n+1,(2)} = \frac{1}{x_{j+\frac{1}{2}}^{n+1} - x_{j-\frac{1}{2}}^{n+1}} \int_{x_{j-\frac{1}{2}}^{n+1}}^{x_{j+\frac{1}{2}}^{n+1}} {\mathbb{U}_{\Delta x}^{\alpha}}^{(1)}(t^{n+1},x)dx, \quad x \in (x_{j-\frac{1}{2}}^{n+1}, x_{j+\frac{1}{2}}^{n+1}).
\tag{6.66}
$$

We thus infer from (6.65) and then (6.66) the following bound

$$
(I_5)_j^{n+1} \leq \zeta_j^{n+1}\nabla\Phi(\mathbb{U}_j^{n+1,(2)}) \cdot \int_{x_{j-1/2}^{n+1}}^{x_{j+1/2}^{n+1}} \big(\mathbb{U}_j^{n+1,(2)} - {\mathbb{U}_{\Delta x}^{\alpha}}^{(1)}(t^{n+1},x)\big)dx = 0,
\tag{6.67}
$$

so that in view of the local error sup norm estimate stated in (6.61), we deduce

$$
(\mathcal{E}_A)_j^n \leq C\Delta x\ \int_{x_{j-1/2}^{n+1}}^{x_{j+1/2}^{n+1}} |{\mathbb{U}_{\Delta x}^{\alpha}}^{(2)}(t^{n+1},x) - {\mathbb{U}_{\Delta x}^{\alpha}}^{(1)}(t^{n+1},x)|dx\boldsymbol{I}_{supp(\zeta)}
\tag{6.68}
$$

Invoking the identity (6.66), we get from the uniform BV bounds (6.2) :

$$
\begin{aligned}
(\mathcal{E}_A)_j^n \quad &\leq \frac{C\Delta x\boldsymbol{I}_{supp(\zeta)}}{x_{j+1/2}^{n+1} - x_{j-1/2}^{n+1}} \int_{x_{j-1/2}^{n+1}}^{x_{j+1/2}^{n+1}} \int_{x_{j-1/2}^{n+1}}^{x_{j+1/2}^{n+1}} |{\mathbb{U}_{\Delta x}^{\alpha}}^{(1)}(t^{n+1},x) - {\mathbb{U}_{\Delta x}^{\alpha}}^{(1)}(t^{n+1},y)|dxdy \qquad (6.69)\\
&\leq TV_{\mathbb{R}}(\mathbb{U}_{\Delta x}^{\alpha}(t^{n+1},.))\frac{C\Delta x}{x_{j+1/2}^{n+1} - x_{j-1/2}^{n+1}} \int_{x_{j-1/2}^{n+1}}^{x_{j+1/2}^{n+1}} \int_{x_{j-1/2}^{n+1}}^{x_{j+1/2}^{n+1}} |x-y|dxdy\boldsymbol{I}_{supp(\zeta)} \qquad (6.70)\\
&\leq C\Delta x^3\boldsymbol{I}_{supp(\zeta)}, \qquad (6.71)
\end{aligned}
$$

so that with $\Delta t/\Delta x$ kept constant, we deduce that the averaging error term is non-positive as Δx goes to zero

$$
(\mathcal{E}_A) \leq C\Delta x \sum_{n\geq 0}\sum_{j\in\mathbb{Z}} \boldsymbol{I}_{supp(\zeta)}\Delta x\Delta t \leq C\Delta x.
\tag{6.72}
$$

To conclude, the overall sampling error $\mathcal{E}_S(\Delta x, \alpha, \zeta)$ can be shown to go to zero with Δx for almost any given sequence $\alpha \in \mathcal{A} = (0,1)^{\mathbb{N}}$, using exactly the same arguments as those developed by Glimm [10] in the convergence analysis of his scheme (see also Serre [26]). With Serre's notations, consider $d\nu(\alpha)$ the measure defined on the Borel sets of the space of sequences $\mathcal{A} = (0,1)^{\mathbb{N}}$, then the following estimate follows

$$\int_{\mathcal{A}} |\mathcal{E}_S(\Delta x, \alpha, \zeta)|^2 \, d\nu(\alpha) \leq C \sup_{t \geq 0} \mathrm{TV}\left(u^\alpha_{\Delta x}(t, \cdot)\right) \Delta x \leq C \Delta x \tag{6.73}$$

invoking the property that $u^\alpha_{\Delta x}(t,x)$ has uniformly bounded total variation for all $\alpha \in \mathcal{A} = (0,1)^{\mathbb{N}}$. We refer the reader to [26] (Lemma 5.4.2, Chapter 5) for a proof. The proposed estimate actually ensures that for any given test function $\zeta \in \mathcal{C}^1_0(\mathbb{R}^+_t \times \mathbb{R}_x)$, there exists a negligeable set $\mathcal{N}_\zeta \subset \mathcal{A}$ such that for all sequences in $\mathcal{A}/\ \mathcal{N}_\zeta$, the sampling error $\mathcal{E}(\Delta x, \alpha, \zeta)$ goes to zero with Δx. We can therefore conclude that the limit function u^α verifies for almost any given sampling sequence $\alpha \in \mathcal{A}$

$$\int_{\mathbb{R}^+_t \times \mathbb{R}_x} \Big(\mathcal{U}(u^\alpha) \partial_t \zeta + \mathcal{F}(u^\alpha) \partial_x \zeta \Big) dt dx + \int_{\mathbb{R}_x} \mathcal{U}(u_0) \zeta(0,x) dx \geq 0, \tag{6.74}$$

for any non-negative test function $\zeta \in C^1_c\left((0,\infty) \times \mathbb{R}_x\right)$. Again and in the case of a genuinely non-linear flux function, the proposed inequality holds for a single strictly convex entropy pair but after Panov [24], it suffices to observe that in addition u^α verifies by construction

$$\int_{\mathbb{R}^+_t \times \mathbb{R}_x} \Big(u^\alpha \partial_t \zeta + f(u^\alpha) \partial_x \zeta \Big) dt dx + \int_{\mathbb{R}_x} u_0 \zeta(0,x) dx = 0, \tag{6.75}$$

namely u^α is a weak solution which satisfies one entropy inequality (6.74): it necessarily coincides with the Kružkov solution. In the situation of a general non-linear flux function, the inequality (6.74) holds true for the whole Kružkov family which readily implies that u^α is nothing but the Kružkov solution of the Cauchy problem under consideration. □

7 Numerical examples

In this section we present numerical results to highlight the importance of handling infinitely many entropy pairs in the design of the anti-diffusive law $\Theta(u_L, u_R)$ for a flux function without genuine non-linearity. In that aim, we approximate the Kružkov solution of the initial value problem

$$\begin{aligned} &\partial_t u + \partial_x \left(\tfrac{u^3}{3}\right) = 0, \quad t > 0, x \in (0,1), \\ &u(0,x) = u_0(x) = \begin{cases} u_L = -1, & x < 0.5, \\ u_R = +1, & x > 0.5, \end{cases} \end{aligned} \tag{7.1}$$

with Neumann Boundary conditions. The exact solution of this Riemann problem is a compound wave made of a shock attached to a rarefaction wave, as depicted in the Figures displayed hereafter. The initial data in (7.1) is chosen so that the entropy jump for the quadratic entropy pair is zero

$$-\sigma(u_L, u_R)\Big(\frac{u_R^2}{2} - \frac{u_L^2}{2}\Big) + \Big(\frac{u_R^4}{4} - \frac{u_L^4}{4}\Big) = 0, \quad \sigma(u_L, u_R) = \frac{1}{3}. \tag{7.2}$$

(7.3)

Hence choosing the anti-diffusive law (4.24) designed for genuinely non-linear flux functions comes with $\Gamma(u_L, u_R) = 0$ in (4.25) so that the optimal value $\Theta(u_L, u_R)$ in (4.24) boils down to 1. With such a law any of the two methods (5.1) and (5.2) capture a weak solution made of a single discontinuity propagating with speed $\sigma(u_L, u_R) = 1/3$. This weak solution is entropy violating. It is therefore of central importance to promote the anti-diffusive law (4.63) to enforce for validity all the Kružkov entropy inequalities. Numerical results displayed below assess these issues.

The solution of the IBVP (7.1) is approximated using the Jin-Xin method with and without defect measure corrections to illustrate their relative performance. The method (5.2) based on local space averagings is promoted. The anti-diffusive law is first set to the optimal law (4.63) especially designed for general non-linear flux function. It is then set to (4.24) for our numerical purposes. In the calculations, we use the low variance Van der Corput sequence $\alpha \equiv \{\alpha_n\}_{n\geq 0}$ (see [13] for instance) defined by

$$\alpha_n = \sum_{k=0}^{m} i_k 2^{-(k+1)}, \quad \text{with } n = \sum_{k=0}^{m} i_k 2^k, \tag{7.4}$$

where the i_k represents the binary expansion of the integer $n = 1, 2, ...$ The first few elements of this sequence are

$$\begin{array}{llll} a1 = 0.5, & a2 = 0.25, & a3 = 0.75, & a4 = 0.125, \\ a5 = 0.625, & a6 = 0.375, & a7 = 0.875, & a8 = 0.0625. \end{array} \tag{7.5}$$

The number of points in space is taken to be 250 and the CFL condition is set at the value 0.45. Exact and discrete solutions for the Xin-Jin method without defect measure corrections are compared in Figure 1. Corresponding results for the Xin-Jin method with defect measure corrections based on the optimal law (4.63) are displayed in Figure 7. Observe the fairly good agreement achieved with the exact solution. Results obtained for the optimal law (4.24) are plotted in Figure 3. As expected, the method captures a wrong weak solution.

References

[1] P. Baiti, A. Bressan and H. K. Jenssen, *BV instability of the Godunov scheme*, Comm. Pure Appl. Math. 59, 1604–1638, 2006.

[2] W. Bao, S. Jin, *The random projection method for hyperbolic conservation laws with stiff reaction terms.* J. Comput. Phys. 163, no. 1, 216-248, 1996.

[3] Chalons C., Coquel F., *Modified Suliciu relaxation system and exact resolution of isolated shock waves*, Math. Models Methods Appl. Sci. (M3AS), Vol. 24, No. 5 937971, 2014.

[4] Chalons C., Coquel F., Engel P., Rohde C., *Fast relaxation solver for hyperbolic-elliptic phase-transition problems*, SIAM J. Sci. Comput. (SISC), vol 34(3), 1753-1776, 2012.

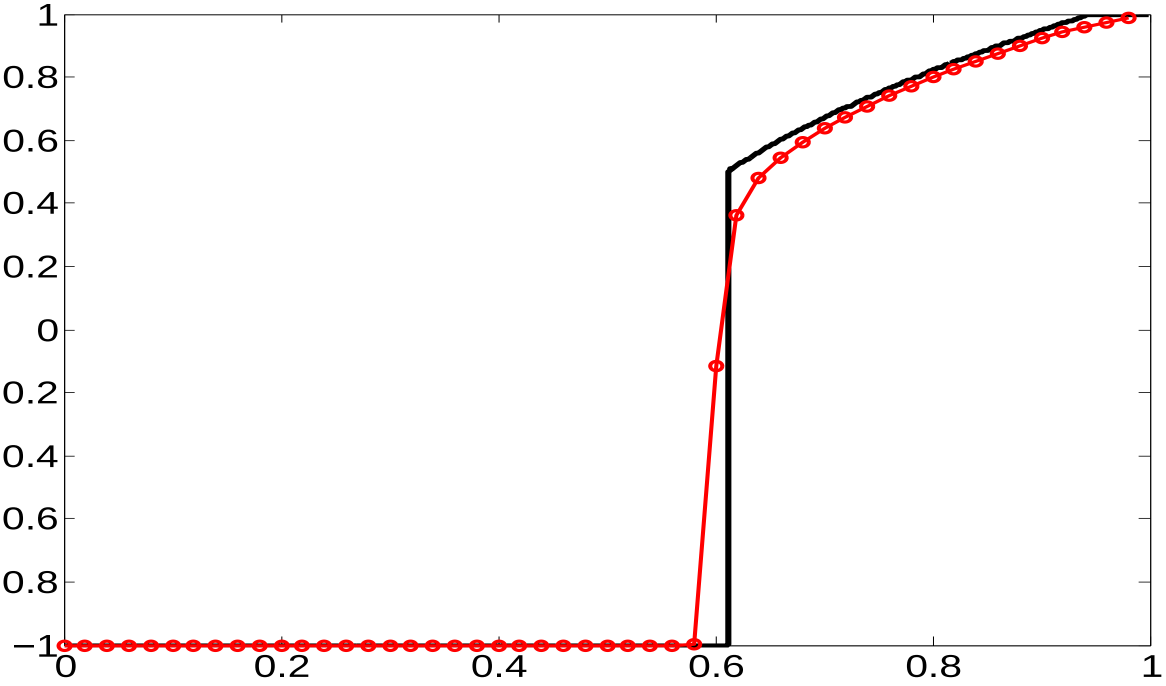

Figure 1: Xin-Jin method without defect measure corrections

[5] G.Q. Chen, C.D. Levermore and T.P. Liu, *Hyperbolic conservation laws with stiff relaxation terms and entropy,* Comm. Pure Appl. Math. 47, 787–830, 1994.

[6] J. Cheng, C.W. Shu, *Second order symmetry-preserving conservative Lagrangian scheme for compressible Euler equations in two-dimensional cylindrical coordinates.* J. Comput. Phys. 272, 245-265, 2014

[7] I.J. Chern, P. Colella, *A conservative front tracking method for hyperbolic conservation laws*, Preprint UCRL-97200, 1987.

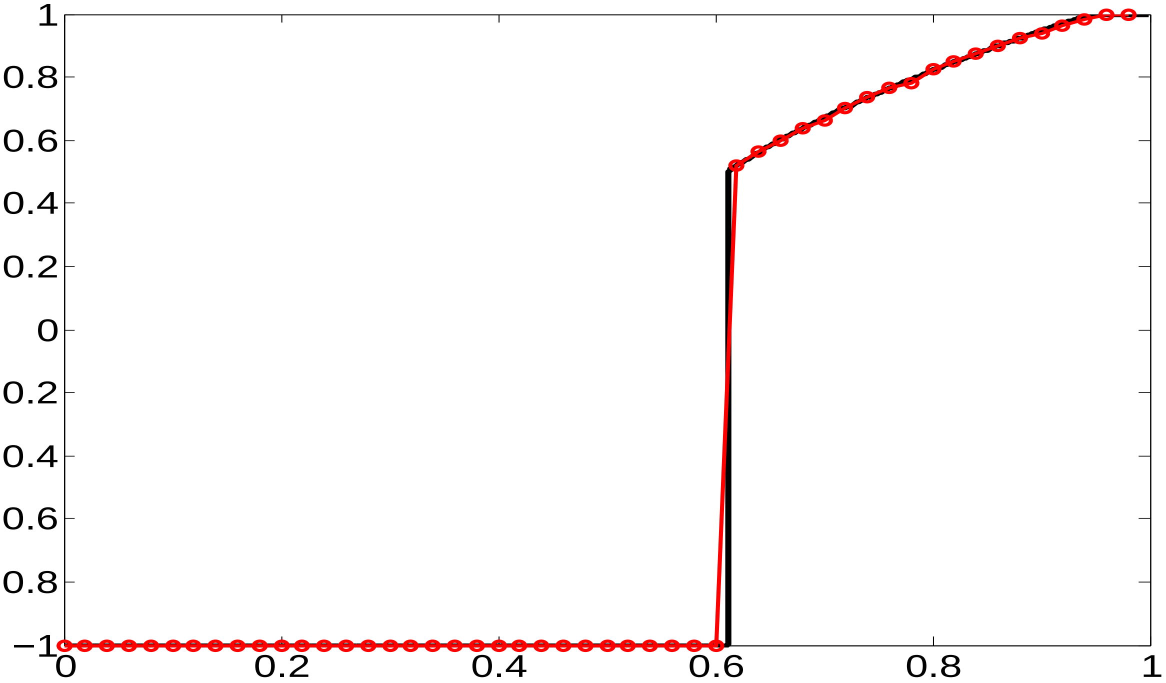

Figure 2: Xin-Jin method with defect measure corrections based on the anti-diffusive law (4.63) for a general non-linear flux

[8] P. Colella, A. Majda, V. Roytburd, *Theoretical and numerical structure for reacting shock waves.* SIAM J. Sci. Statist. Comput. 7 , no. 4, 1059-1080, 1986.

[9] E. Godlewski and P. A. Raviart, *Hyperbolic systems of conservation laws,* Math. and Appl. 3/4, Ellipses, Paris, 1991.

[10] J. Glimm, *Solutions in the large for nonlinear hyperbolic systems of equations*, Communications on Pure and Applied Mathematics, Vol XVIII, 697–715, 1965.

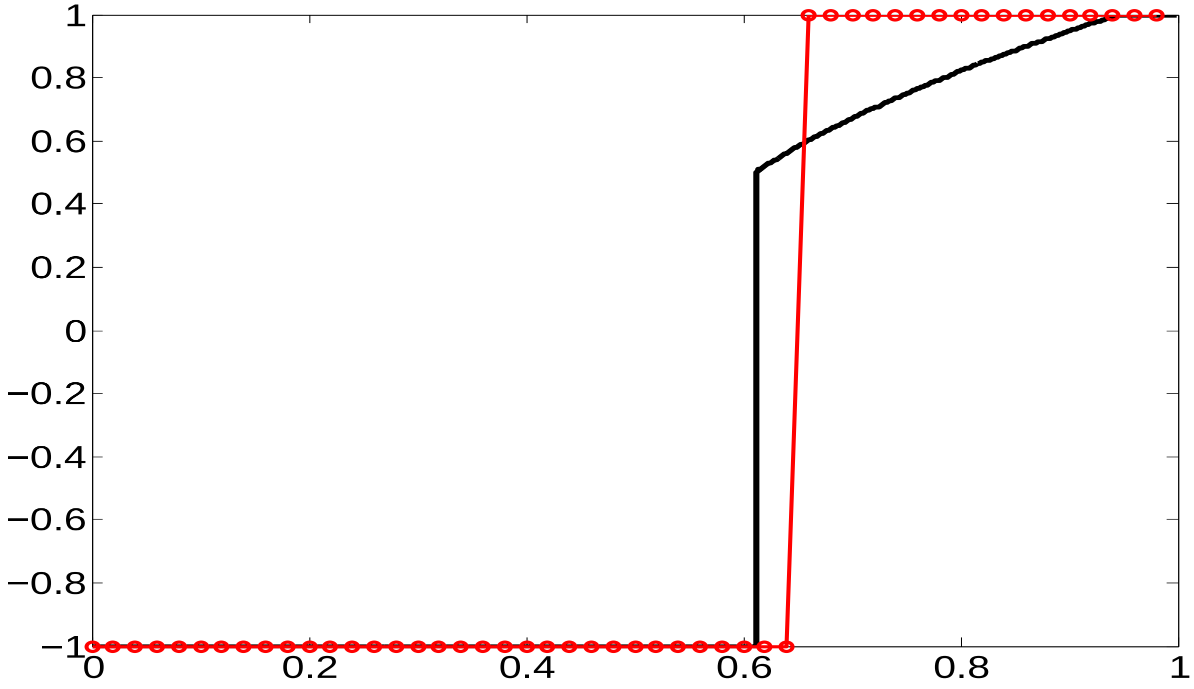

Figure 3: Xin-Jin method with defect measure corrections based on the anti-diffusive law (4.63) for a genuinely non-linear flux

[11] A. Harten, *ENO schemes with subcell resolution.* J. Comput. Phys., 83(1):148-184, 1989.

[12] A. Harten, J.M. Hyman, *Self adjusting grid Methods for one-dimensional hyperbolic conservation laws*, J. Comp. Phys., Vol. 50, No 2, 235–269, 1983.

[13] A. Harten, P.D. Lax, *A random choice finite difference schemes for hyperbolic conservation laws*, SIAM J. Numer. Anal., Vol 18, No 2, 289–315, 1981.

[14] A Harten, P Lax and B van Leer, *On upstream differencing and Godunov type methods for hyperbolic conservation laws* SIAM review. 25(1):35–61, 1983.

[15] S. Jin, J.G. Liu, *The effects of numerical viscosities. I. Slowly moving shocks.* J. Comput. Phys. 126, no. 2, 373-389, 1996

[16] S. Jin, Z. P. Xin, *The relaxation schemes for systems of conservation laws in arbitrary space dimensions,* Comm. Pure Appl. Math. 48, no.3, 235- - 276, 1995.

[17] S. Karni, *Multicomponent flow calculations by a consistent primitive algorithm.* J. Comput. Phys. 112, no. 1, 31-43, 1994.

[18] P. LeFloch, S. Mishra, *Numerical methods with controlled dissipation for small-scale dependent shocks*, Acta Numerica, 1–72, 2014.

[19] S. N. Kružkov, *First order quasilinear equations with several independent variables,* Mat. Sb. (N.S.) 81(123), 228 - 255, 1970.

[20] R.J. LeVeque, H.C. Yee, *Study of numerical methods for hyperbolic conservation laws with stiff source terms.* J. Comput. Phys. 86 (1990), no. 1, 187-210.

[21] T.P. Liu, *Hyperbolic conservation laws with relaxation,* Comm. Math. Phys. 108, 153 - 175, 1987.

[22] R. Natalini, *Convergence to equilibrium for the relaxation approximation of conservation laws,* Comm. Pure Appl. Math. 49, no. 8, 795 - 823, 1996.

[23] W. F. Noh, *Errors for calculations of strong shocks using an artificial viscosity and an artificial heat flux*, J. Comp. Phys., vol. 72, pp. 78-120, 1987.

[24] E. Yu. Panov, *Uniqueness of the Cauchy problem for a first order quasilinear equation with one admissible strictly convex entropy*, Mathematical Notes, Vol. 55, No 5, 517 – 525, 1994.

[25] J.J. Quirk, *A contribution to the great Riemann solver debate.* Internat. J. Numer. Methods Fluids 18, no. 6, 555-574, 1994.

[26] D. Serre, *Systems of Conservation Laws I, II*, Cambridge University Press, 2000.

[27] C.-W. Shu, *Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws*, Advanced Numerical Approximation of Nonlinear Hyperbolic Equations, Editor: A. Quarteroni, Lecture Notes in Mathematics, volume 1697, Springer, 1998.

[28] C.-W. Shu, *A brief survey on discontinuous Galerkin methods in computational fluid dynamics* , Advances in Mechanics, volume 43, pp.541-554, 2013.

[29] M. Slemrod, *Admissibility criteria for propagating phase boundaries in a van der Waals fluid.* Arch. Rational Mech. Anal. 81, no. 4, 301-315, 1983.