UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

Entropy Flux-Splittings for Hyperbolic Conservation Laws, Part I: General Framework

Gui-Qiang Chen
Philippe G. LeFloch

January 1995

CAM Report 95-1

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90024-1555

ENTROPY FLUX-SPLITTINGS FOR HYPERBOLIC CONSERVATION LAWS PART I: GENERAL FRAMEWORK

GUI-QIANG CHEN[1] AND PHILIPPE G. LEFLOCH[2]

CONTENTS

1. Introduction
2. Definitions and Examples
3. Convergence of Entropy Flux-Splitting Schemes
4. A Family of Entropy Flux-Splittings for the Euler Equations
5. Convergence of a Flux-Splitting Scheme for the Euler Equations
References
Appendix: Genuine Nonlinearity of the Flux-Splitting in Section 5

Preprint July 1994.
[1] Department of Mathematics, University of Chicago, Chicago, Illinois 60637.
E-mail: cheng@math.uchicago.edu.
[2] Department of Mathematics, University of Southern California, Los Angeles, CA 90024–1113.
E-mail: lefloch@cams.usc.edu
Mathematics Subject Classification. 35L65, 65M12.
Key words and phrases: conservation law, entropy discontinuous solution, flux-splitting, difference schemes, weak convergence.

ABSTRACT

This paper is devoted to the derivation of entropy satisfying flux-splittings associated with a system of conservation laws

$$\partial_t u + \partial_x f(u) = 0, \qquad u(x,t) \in \mathbf{R}^p,$$

endowed with a strictly convex entropy function $U : \mathbf{R}^p \to \mathbf{R}$. A decomposition of the form $f(u) = f^+(u) + f^-(u)$ is called a flux-splitting if $\pm\nabla f^\pm$ has only real and non-negative eigenvalues, which we denote by $\pm\nabla f^\pm \geq 0$. We say that it is an *entropy flux-splitting* if, furthermore, U is an entropy function for both

$$\partial_t v + \partial_x f^-(v) = 0, \qquad \text{and} \qquad \partial_t w + \partial_x f^+(w) = 0.$$

When $p \geq 2$, the notion depends upon the choice of the entropy U. We say that it is a *genuine flux-splitting* if $f^\pm(u) = f(u)$ and $f^\mp(u) = 0$ when $\pm\nabla f \geq \pm C$ for C large enough. Observe that this latter condition is weaker than the standard condition that $f^\pm(u) = f(u)$ and $f^\mp(u) = 0$ when $\pm\nabla f \geq 0$. We say that the splitting is a *diagonalizable flux-splitting* if the three matrices ∇f, ∇f^+, and ∇f^- have a common basis of eigenvectors.

Flux-splittings induce simple but useful difference schemes for the approximation of hyperbolic systems. We obtain in this paper a framework for the construction and analysis of these schemes. We prove that the schemes defined from an entropy flux-splitting satisfy a discrete cell entropy inequality. For the scheme generated from a diagonalizable splitting, the principle of bounded invariant regions applies to the scheme, and provides an a priori L^∞ estimate. Using compensated compactness arguments, we deduce the strong convergence of the flux-splitting schemes associated with strictly hyperbolic and genuinely nonlinear systems of two conservation laws.

This paper includes in a unified framework for the existing flux-splittings (Steger-Warming, van Leer, kinetic, Lax-Friedrichs type, etc). The approach leads to several new properties and characterizations of these splittings, as well as to a method for extending them to more general situations. A main result in this paper is the existence of a large family of genuine entropy flux-splittings for several significant examples: the (nonconvex) scalar conservation laws, the p-system, and the Euler system for an isentropic gas described by the so-called γ-law. The isentropic Euler system is particularly noteworthy. For this system, we obtain a family of splittings that satisfy the entropy inequality associated with the mechanical energy. We prove that there exists a unique genuine entropy flux-splitting that satisfies all the entropy inequalities. It is also the unique diagonalizable splitting associated with the isentropic Euler system. This splitting could also be derived by the so-called kinetic approximation. Finally, for this later splitting, we study the Riemann problem associated with the flux-functions f^+ and f^-, check the strict hyperbolicity and genuine nonlinearity of the corresponding hyperbolic systems, and prove the strong convergence of the scheme for $\gamma \in (1, 5/3]$.

Our companion paper [9] establishes the existence of entropy splittings for the isentropic 2×2, as well as full 3×3, systems of gas dynamics. The so-called real (i.e., not necessarily polytropic and perfect) gas is treated. Numerical results will be presented in a forthcoming publication.

1. Introduction

This paper and its companion [9] are devoted to the construction of entropy satisfying flux-splittings associated with the flux-function of any hyperbolic system of conservation laws. Flux-splittings are used in order to design the so-called flux-splitting finite difference schemes, and their high order accurate extensions. In this work, we propose a general framework which unifies the derivation and analysis of entropy satisfying flux-splittings. We intent to prove the convergence of the flux-splitting schemes to the entropy discontinuous solutions to the system of conservation laws.

Many flux-splittings has been derived in the literature for the Euler systems of gas dynamics and are currently used in fluid dynamics codes. In Steger-Warming[60] and van Leer[39], the derivation is heuristic and based on specific properties of the gas dynamics equations (such as the homogeneity), or on searching for the "simplest possible" splitting having a given a priori expression (e.g., polynomial function with minimum order). It is unknown if those splittings are consistent with the entropy criterion. In the works by Sanders-Prendergast, Pullin, Reitz, Kaniel, and their followers [56, 54, 55, 34], the splitting is built up in a rather natural manner from a kinetic formulation of the gas dynamics equations. It is known that some flux-splittings satisfy the entropy criterion (Deshpande [25], Perthame [52], Croisille-Delorme [20, 21]). The present paper continues the analysis of flux-splittings. The approach is intended to be rather general: we do not make a priori assumption on the hyperbolic system under consideration, nor on the form of the flux-splitting, and derive necessary and sufficient conditions for a splitting to exist and to be entropy satisfying. In this first paper, we explain how most of the classical flux-splittings can be derived in a simple and systematic manner, and in some cases characterized uniquely. The properties of general flux-splittings, and especially of those associated with isentropic Euler systems of gas dynamics are studied below. We also address the questions of stability and convergence of the corresponding schemes. We shall return to the gas dynamics systems in [9].

Let us consider a system of p conservation laws in one space variable

$$\partial_t u + \partial_x f(u) = 0, \qquad u(t,x) \in \mathcal{O}, \quad t > 0,\, x \in \mathbf{R}, \tag{1.1}$$

together with the Cauchy data

$$u(0,x) = u_0(x), \qquad x \in \mathbf{R}. \tag{1.2}$$

The subset $\mathcal{O} \subset \mathbf{R}^p$ is assumed to be convex and open, the flux-function $f : \mathcal{O} \to \mathbf{R}^p$ to be Lipschitz continuous at least, and the initial data $u_0 : \mathbf{R} \to \mathcal{O}$ to belong to the space $L^\infty_{\text{loc}}(\mathbf{R})^p$ of all measurable and bounded functions. We assume that the system (1.1) is endowed with a strictly convex entropy pair $(U, F) : \mathcal{O} \to \mathbf{R} \times \mathbf{R}$. In particular, for each compact subset $K \subset\subset \mathcal{O}$, one has

$$\nabla^2 U \geq C_K I \qquad \text{and} \qquad \nabla F = \nabla U^\top \nabla f, \tag{1.3}$$

where C_K is a positive constant depending on K, and each smooth solution to (1.1) satisfies the additional conservation law:

$$\partial_t U(u) + \partial_x F(u) = 0.$$

Here we have denoted by $\nabla^2 U$ the Hessian matrix of the scalar-valued function U, I the $p \times p$ identity matrix, ∇G the Jacobian matrix of a vector-valued function G, and $B^\top$ the transpose of a matrix B. As is well-known, the existence of an entropy pair implies that (1.1) is hyperbolic, but not necessarily strictly hyperbolic. This means that, for each $u \in \mathcal{O}$, the matrix $\nabla f(u)$ admits p real eigenvalues $\lambda_1(u) \le \lambda_2(u) \le \cdots \le \lambda_p(u)$, and a basis of right eigenvectors $r_1(u), r_2(u), \cdots, r_p(u)$.

We also recall that solutions to hyperbolic conservation laws in general are not smooth, and must be understood in the (weak) sense of distributions. The weak solutions in the space $L^\infty_{\mathrm{loc}}(\mathbf{R}_+ \times \mathbf{R})^p$ are not uniquely determined by their initial data. The so-called entropy condition is necessary to select the relevant solutions. Several formulations can be considered; in this paper, we refer to the Lax entropy inequality [38]:

$$\partial_t U(u) + \partial_x F(u) \le 0 \tag{1.4}$$

understood in the sense of distributions. We also refer to [37, 23, 59] for general background on hyperbolic equations.

We review first the background on the stability and convergence of finite difference schemes in conservative form associated with the Cauchy problem (1.1)–(1.2). Note that the rigorous convergence results below will be restricted to systems of two equations, while L^∞ stability estimates and entropy consistency will hold for general systems. Our results of L^∞ stability will be restricted to the systems endowed with invariant regions (cf. [11]). The entropy inequality (1.4) will be satisfied by the approximations at the discrete level, and the passage to the limit will be justified. We recall that the convergence of the Lax-Friedrichs' and Godunov's schemes applied to systems was established by DiPerna [27], and extended by Chen [5, 6] (also [26]). The L^∞ stability of the Godunov scheme for general systems was studied by Hoff in [32]. The entropy consistency for systems is addressed by Tadmor in [64]. Also see [10] for high order methods. The case of scalar equations (i.e. $p = 1$) has also received much attention. For recent progress on the convergence analysis of schemes applied to scalar problems, we refer the reader to [29, 30, 31, 33, 37, 42, 49, 50, 51, 62, 63] for the fundamental concepts and results on one-dimensional equations, and to [7, 12, 13, 14, 15, 16, 19, 61], for the extension to multidimensional equations. To our knowledge, the present paper represents the first attempt to address the convergence of a class of schemes for general hyperbolic systems.

The present work focuses on the class of schemes based on a flux-splitting, i.e. a decomposition of the flux-function f in the form

$$f(u) = f^+(u) + f^-(u), \tag{1.5}$$

where the functions $f^\pm$ are, at least, locally Lipschitz continuous and $\pm\nabla f^\pm(u)$ has only real and nonnegative eigenvalues. A scheme based on (1.5) is obtained by averaging together (cf. Section 3) the Godunov schemes associated with each two systems

$$\partial_t v + \partial_x f^-(v) = 0, \qquad \text{and} \qquad \partial_t w + \partial_x f^+(w) = 0. \tag{1.6}$$

Since all the eigenvalues of $\pm\nabla f^\pm(u)$ have a constant sign (independent of u and the index of the characteristic field), the Godunov scheme for each of the systems in (1.6) reduces to the upwind scheme. So (1.6) yields a simple, but efficient for numerical purposes, method in order to approximate of the discontinuous solutions to (1.1). Flux-splittings has been mathematically studied first on scalar conservation laws by Engquist-Osher [29]. The concept of flux-splitting has been also studied in detail in the paper by Harten-Lax-vanLeer [31]. (See Brenier [1, 2] for a related approach.)

This idea of splitting is actually well-spread in the literature. The first flux-splittings were derived for the gas dynamics equations by Sanders-Prendergast, then Steger-Warming and van Leer [56, 60, 39]. The results therein concern the polytropic and perfect gas, and were next extended to the so-called real gas (i.e. with no assumption on the equation of state of the gas) in [58, 66]. The dynamics of reactive gas was treated in [3] and [36]. We refer to the paper by Lerat [44] for the first mathematical analysis of the Steger-Warming scheme. Based on a kinetic interpretation of the gas dynamics equations, numerous flux-splittings were derived beginning with the works Sanders-Prendergast, Pullin, Reitz, Kaniel, Deshpande, etc, c.f. [56, 54, 55, 34, 25]. The first mathematical analysis of those schemes was given by Perthame [52, 53], who proved the entropy consistency of kinetic schemes for the gas dynamics system. See also [45, 46] for the mathematical analysis of the kinetic approximation at the continuous level. A related approach is considered by Bourdel-Delorme-Mazet [4], and Croisille-Delorme [21].

In this paper we attempt to provide a systematic method for the construction of flux-splittings associated with the system of conservation laws (1.1), and the analysis of the stability and convergence of the corresponding flux-splitting schemes. We say that (1.5) is an *entropy flux-splitting* if each of the systems in (1.6) is endowed U as an entropy function. This is actually equivalent to saying that $\nabla^2 U \, \nabla f^{\pm}$ are symmetric matrices. When $p \geq 2$, the notion depends upon the choice of the entropy U. We shall say that (1.5) is a *diagonalizable flux-splitting* if the matrices ∇f, and $\nabla f^{\pm}$ have a common basis of eigenvectors.

We say that (1.5) is a *genuine flux-splitting* if $f^{\pm}(u) = f(u)$ and $f^{\mp}(u) = 0$ when $\pm\nabla f \geq \pm C$ for C large enough. This condition is weaker than the classical condition used by Steger-Warming and van Leer, that $f^{\pm}(u) = f(u)$ and $f^{\mp}(u) = 0$ when $\pm\nabla f \geq 0$. A genuine splitting leads to an upwind scheme similar to the Godunov scheme (upwind is often an attractive property for numerical computations), while a non-genuine splitting leads to a scheme similar to the Lax-Friedrichs scheme, and generally contains more numerical diffusion than an upwinding scheme.

At first glance, these definitions may seem rather restrictive. Actually, one of the main results in this paper is the *existence of a family of entropy flux-splittings* for many systems occurring in physics. As we show it, diagonalizable splittings can be constructed for certain systems. For those splittings, the systems (1.1) and (1.6) have the same Riemann invariants, i.e. the same invariant regions.

We complete this introduction with a brief presentation of the results established in this paper.

Section 2 contains the general properties satisfied by an entropy flux-splitting. We show that any system of conservation laws admits a family of entropy flux-splittings, i.e. the Lax-Friedrichs type splittings as we call them. Namely, those splittings lead to a variant of the Lax-Friedrichs scheme (first suggested by Shu [57] for the computation of phase transitions). This splitting is not a genuine splitting, however.

A scalar conservation law admits a family of genuine flux-splittings. We recover the Engquist-Osher's splitting as an "extremal" case in this family. With regard to the entropy inequality (1.4), we observe that any flux-splitting is actually an *entropy* flux-splitting. In other words, in the special case of scalar equations, the condition $\pm\nabla f^{\pm}(u) \geq 0$ is enough to imply the consistency of the splitting scheme with the entropy inequality (1.4). The result does not hold for systems, and is related to the fact that any function is an entropy when $p = 1$. (An entropy must satisfy an –overdetermined, in general– set of compatibility relations when $p \geq 2$.)

We prove in Section 2 that the p-system of isentropic gas dynamics in Lagrangian coordinates admits a family of entropy flux-splittings. With any entropy flux-function to the p-system, we can associate an entropy flux-splitting. Note that the p-system does not admit any genuine flux-splitting: this is due to the fact that the two eigenvalues of the p-system have distinct signs.

Next the isentropic Euler system for the so-called γ–law gas is fully studied. We prove in Section 4 that this system admits a whole family of entropy flux-splittings associated with the mechanical energy function (which plays the role of an entropy function for that system). Moreover, among those decompositions, there exists a *unique* splitting which is consistent with *all* the entropy functions associated with the Euler system. As a matter of fact, we show that two entropy functions are enough to characterize a unique splitting. It is noteworthy that this splitting can also be characterized as being the only diagonalizable splitting. Our results also show that the class of the so-called kinetic splittings contains a single flux-splitting that satisfies all the entropy inequalities. This later splitting was first pointed out by Khobalatte-Perthame [35] in the (slightly different) situation of the full 3×3 system of gas dynamics. We refer to Chen-LeFloch [9] for the extension to the full 3×3 system of gas dynamics, and to arbitrary equations of state.

The main results of stability and convergence in this paper are as follows. In Section 3, we consider a strictly hyperbolic system of conservation laws, and prove that a scheme based on an entropy flux-splitting is consistent with the entropy inequality (1.4), and so can converge to only an entropy solution. We next restrict ourselves to a system of two equations endowed with bounded invariant regions, and show that the flux-splitting scheme is stable in the L^∞ norm if (1.5) is a diagonalizable splitting. From DiPerna's work [27] based on the compensated compactness method developed by Murat and Tartar [47, 48, 65], it follows that the entropy flux-splitting schemes converge in the strong L^1 sense to an entropy weak solution. For general flux-splittings, we use the notion of nonconservative product due to Dal Maso-LeFloch-Murat [24] and LeFloch-Liu [43], and derive a discrete cell entropy inequality in nonconservative form.

Section 5 is devoted to the proof of convergence of a flux-splitting scheme for the isentropic Euler system. The solutions can include the vacuum state, at which point the system becomes nonstrictly hyperbolic. The proof is based on the compactness framework due to DiPerna [27] and Chen [5, 6, 26]. For the analysis, we have to determine the regions where systems (1.6) are strictly hyperbolic and/or genuinely nonlinear in that case. The solution to the Riemann problem for both systems in (1.6) is shown to exist for arbitrary large jump in the initial data, and satisfy the Lax entropy inequalities.

Finally, let us recall that the flux-splitting schemes provide some alternative between the more classical Godunov and Lax-Friedrichs schemes. They combine the upwind feature of the Godunov scheme, without the complicated and costly resolution of Riemann problems. The numerical flux-functions of flux-splitting schemes are provided by explicit formulas, as is the case of the Lax-Friedrichs scheme, but still they contain limited numerical viscosity, and so allow a satisfactory computation of discontinuities. Furthermore, the numerical flux-function of a flux-splitting scheme can be of class C^1, and even sometimes of class C^2, while most schemes have only Lipschitz continuous flux-functions. This latter property is useful for the computation of steady state solutions using a time-dependent method.

The content of this paper was announced in [8].

2. Definitions and Examples

In this section, we introduce several definitions, and derive the main properties satisfied by an entropy flux-splitting. As an illustration to the definitions, the existence of such splittings is checked for the scalar conservation laws and the system of isentropic gas dynamics in Lagrangian coordinates.

2.1. Definitions and Main Properties. We consider the system of conservation laws (1.1) endowed with a strictly convex entropy pair (U, F) satisfying (1.3). Let us recall the following definition, which is classical.

Definition 2.1. *A* flux-splitting *for the system* (1.1) *is a decomposition of the form*

$$f(u) = f^+(u) + f^-(u), \qquad u \in \mathcal{O}, \tag{2.1a}$$

where the functions $f^\pm : \mathcal{O} \to \mathbf{R}^p$ are locally Lipschitz continuous, the Jacobian matrices $\nabla f^\pm$ have only real eigenvalues $\lambda_j^\pm, 1 \le j \le p$, and a basis of eigenvectors $r_j^\pm, 1 \le j \le p$, and satisfy:

$$\begin{aligned} &\lambda_1^-(u) \le \lambda_2^-(u) \le \cdots \le \lambda_p^-(u) \le 0 \\ \le &\lambda_1^+(u) \le \lambda_2^+(u) \le \cdots \le \lambda_p^+(u), \quad u \in \mathcal{O}. \end{aligned} \tag{2.1b}$$

We shall use the notation $\pm\nabla f^\pm \ge 0$ if the matrix $\pm\nabla f^\pm$ has only real and nonnegative eigenvalues, and possesses a basis formed by eigenvectors. The following definition contains the new concept which is at the origin of this paper. The definition involves the entropy function U of system (1.1) and the following two systems

$$\partial_t v + \partial_x f^+(v) = 0, \quad v \in \mathcal{O}, \tag{2.2a}$$

$$\partial_t w + \partial_x f^-(w) = 0, \quad w \in \mathcal{O}. \tag{2.2b}$$

Note in passing that Definition 2.1 implies that (2.2) are hyperbolic systems. Definition 2.2 is concerned with the entropy inequality (1.4).

Definition 2.2. *We say that* (2.1) *is an* entropy flux-splitting *for the system* (1.1) *if the function U is an entropy function for both systems* (2.2*a*) *and* (2.2*b*). *In other words, there must exist entropy flux-functions $F^{\pm} : \mathcal{O} \to \mathbf{R}$ such that all the smooth solutions v and w to* (2.2*a*) *and* (2.2*b*) *also satisfy*

$$\partial_t U(v) + \partial_x F^+(v) = 0, \tag{2.3a}$$

$$\partial_t U(w) + \partial_x F^-(w) = 0, \tag{2.3b}$$

respectively.

If the condition in Definition 2.2 holds for an arbitrary decomposition like (2.1a), then the systems (2.2) necessarily are hyperbolic, although the sign conditions (2.1b) need not hold. When necessary, we shall specify that (2.1) is an entropy splitting associated with *the* entropy U. For some systems, we shall be able to derive splittings that satisfy Definition 2.2 for *all* entropy functions. As we prove it below (Proposition 2.1), this is the case of the diagonalizable splittings, as we call them.

Definition 2.3. *We say that* (2.1) a diagonalizable *flux-splitting if the three matrices ∇f, ∇f^+, and ∇f^- have a common basis of eigenvectors.*

We introduce yet another definition.

Definition 2.4. *We say that* (2.1) *is a* genuine *flux-splitting if for each compact subset $K \subset\subset \mathcal{O}$, and for some positive constant C_K, and all u in K,*

$$\begin{aligned} &f^+(u) = f(u), \quad f^-(u) = 0, \quad \text{if} \quad \lambda_1(u) \geq C_K, \\ &f^+(u) = 0, \quad f^-(u) = f(u), \quad \text{if} \quad \lambda_p(u) \leq -C_K. \end{aligned} \tag{2.4}$$

Our primary interest in this paper concerns the genuine flux-splittings. Note that Definition 2.4 is of interest only in the case of systems whose eigenvalues change their sign, as is the case of the system of gas dynamics in Eulerian coordinates. A system, whose all eigenvalues are either non-negative or non-positive, admits a trivial splitting (Proposition 2.2). A system whose eigenvalues have a constant sign independent of u, but whose not all eigenvalues have the same sign can not admit a genuine splitting. This latter case arises with the system of gas dynamics in Lagrangian coordinates to be studied in Proposition 2.6. Proposition 2.5 below will give a full description of the genuine entropy flux-splittings associated with a scalar equation.

The following proposition provides a characterization of the entropy flux-splittings that will play a central role in this paper. From now on, we tacitly assume that the flux-function f and the entropy U admit locally bounded second order derivatives.

Proposition 2.1. *(i)– The decomposition* (2.1) *with $f^{\pm} \in W^{2,\infty}(\mathcal{O})^p$ is an entropy flux-splitting if and only if*

$$\nabla^2 U \nabla f^+ \ \text{ and } \ \nabla^2 U \nabla f^- \ \text{ are symmetric matrices.} \tag{2.5}$$

(ii)– In particular, a diagonalizable splitting is an entropy flux-splitting for any choice of entropy function U. Furthermore, the three systems in (1.1) *and* (2.2) *possess the same set of Riemann invariants.*

The proof of Proposition 2.1 is based on the following classical fact: since U is an entropy function to system (1.1), $\nabla F = \nabla U \nabla f$ is a gradient, and thus $\nabla^2 U \nabla f$ is a symmetric matrix. The latter condition is also a sufficient condition for U to be an entropy. This characterization of an entropy function can applied to each system in (2.2) as well.

We can prove that any system of conservation laws admits a large family of entropy flux-splittings. The proof relies directly on Proposition 2.1.

Proposition 2.2. *(i)– Let* (1.1) *be a system of conservation laws endowed with a strictly convex entropy function U. Consider the decomposition:*

$$f^+(u) = Af(u) + Bu, \quad f^-(u) = (I - A)f(u) - Bu, \qquad u \in \mathcal{O}, \tag{2.6}$$

where, for all $u \in \mathcal{O}$, the constant $p \times p$ matrices A and B satisfy

$$\nabla^2 U(u) A \nabla f(u), \quad \text{and} \quad \nabla^2 U(u) B \quad \text{are symmetric matrices,} \tag{2.7}$$

and

$$B + A\nabla f(u) \geq 0, \quad B + (A - I)\nabla f(u) \geq 0. \tag{2.8}$$

Then $(2.6) - (2.8)$ *defines an entropy flux-splitting for the system* (1.1).

(ii)– Conditions $(2.7) - (2.8)$ *are satisfied (at least on any compact subset of $\mathcal{O}$) with the choice:*

$$A = \frac{1}{2} I, \quad B = bI \quad \text{with } b \text{ positive large enough.} \tag{2.9}$$

In that case (2.6) *is a diagonalizable splitting.*

(iii)– Let (1.1) *be a system whose characteristic eigenvalues $\lambda_i(u)$ have a constant sign independent of both i and u, say $\lambda_i(u) \geq 0$. Consider the decomposition:*

$$f^+(u) = f(u), \quad \text{and} \quad f^-(u) = 0, \quad u \in \mathcal{O}. \tag{2.10}$$

Then (2.10) *defines a (trivial) entropy flux-splitting.*

The splittings found in Proposition 2.2 will be called the *Lax-Friedrichs type splittings*. The scheme designed from (2.6) (cf. Section 3) turns out to be a simple extension of the Lax-Friedrichs scheme. This generalization of the Lax-Friedrichs scheme was first proposed by Shu [57] for mixed (hyperbolic-elliptic) problems.

By contrast to the result in Proposition 2.2, we observe that an arbitrary system need not admit a *genuine* entropy flux-splitting. At the end of this section and in Section 3, we shall derive such genuine splittings. Note also that a given system admits several distinct splittings. However, it will be shown in Section 4 (cf. also [9]) that the system of gas dynamics admits a class of genuine entropy splittings, and that a unique splitting can be selected among those. The general definitions above assume the splitting to be Lipschitz continuous only, however smoother flux-splittings often exist as we shall see below.

2.2. Actual Derivation of Flux-Splittings. The actual construction of entropy flux-splittings is based on the characterization in Proposition 2.1. Let us comment upon the general splittings first, and the diagonalizable ones next.

Consider first the case of a symmetric system (1.1), i.e. that ∇f is symmetric. Then the function $U(u) = |u|^2/2$ is an entropy function. Since ∇f is a symmetric matrix, there exists a scalar function $\psi : \mathcal{O} \to \mathbf{R}$ such that: $f = \nabla\psi$. The entropy inequality (1.4) then becomes:

$$\partial_t\Big(\frac{|u|^2}{2}\Big) + \partial_x\big(u \cdot \nabla\psi(u) - \psi(u)\big) \leq 0. \tag{2.11}$$

With that choice of entropy, (2.5) is equivalent to saying that each matrix $\nabla f^{\pm}$ is symmetric. So we can search for $f^{\pm}$ in the form:

$$f^{\pm} = \nabla\psi^{\pm}, \ \psi^{+} + \psi^{-} = \psi, \qquad \text{with } \psi^{\pm} : \mathcal{O} \to \mathbf{R}. \tag{2.12}$$

It is easily checked that the sign conditions (2.1b) are equivalent to saying that:

$$\psi^{+} \text{ and } \psi^{-} \text{ are convex and concave functions, respectively.} \tag{2.13}$$

Therefore the question of the construction of flux-splittings is reduced to the question of finding a decomposition of the function ϕ as the sum of a convex function and a concave function. As we will see, this can be done geometrically in the scalar case. However, such a decomposition need not be possible for systems.

The case of non-symmetric systems is slightly more involved. We observe that any system can be put in symmetric form by using the so-called entropy variable. By Definition 2.2, the three systems under consideration have a common entropy variable. By setting $v = \phi(u) = \nabla U(u)$, the system (1.1) takes the form:

$$\partial_t g(v) + \partial_x h(v) = 0, \qquad v(t,x) \in \mathcal{O}', \tag{2.14}$$

where $\mathcal{O}' = \phi(\mathcal{O})$, $g(v) = \phi^{-1}(v)$, and $h(v) = f(\phi^{-1}(v))$. It can be checked that $\nabla_v g$ and $\nabla_v h$ are symmetrix matrices. In view of Proposition 2.1, the entropy flux-splittings associated with system (1.1) are those of the form:

$$h(v) = h^{+}(v) + h^{-}(v) \ \text{ with } \nabla_v h^{\pm}(v) \text{ symmetric.} \tag{2.15}$$

Therefore, a similar argument as above leads us to the following result:

Proposition 2.3. *The entropy flux-splittings* (2.1) *associated with a system of conservation laws* (1.1) *endowed with a strictly convex entropy function* U *are given by the formula:*

$$f^{+}(u) = \nabla_v\psi^{+}(v), \quad f^{-}(u) = \nabla_v\psi^{-}(v), \tag{2.16}$$

where we have set $v = \nabla_u U(u)$, the scalar-valued function $\psi(v)$ is defined by the relation $f(u) = \nabla_v \psi(v)$, and

$$\psi(v) = \psi^+(v) + \psi^-(v) \tag{2.17a}$$

is any decomposition of the function ψ in the sum of a "convex function" ψ^+ and a "concave function" ψ^-, in the following sense:

$$\pm \nabla_{vv}\psi^\pm \nabla_{uu} U \ \geq 0. \tag{2.17b}$$

Decompositions of the form (2.17) were first pointed out by Bourdel-Delorme-Mazet [4] in their work on the so-called K-diagonalisable systems, which indeed admit naturally an entropy flux-splitting in the form (2.17).

We now consider the derivation of *diagonalizable flux-splittings*. Without further restrictions, we shall set

$$r_j(u) = r_j^+(u) = r_j^-(u), \quad j = 1, 2, \cdots, p. \tag{2.18}$$

The eigenvalues of the systems (1.1) and (2.2) satisfy the relation:

$$\lambda_j(u) = \lambda_j^+(u) + \lambda_j^-(u), \quad u \in \mathcal{O},\ 1 \leq j \leq p. \tag{2.19}$$

We also set:

$$P = (r_1, r_2, \cdots, r_p), \qquad \Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_p). \tag{2.20}$$

In practical computations, we shall derive the diagonalizable splittings by searching for a decomposition of the diagonal matrix Λ in the form

$$\Lambda = \Lambda^+ + \Lambda^- \quad \text{where } \Lambda^\pm = \mathrm{diag}(\lambda_1^\pm, \lambda_2^\pm, \cdots, \lambda_p) \geq 0, \tag{2.21}$$

with the restriction:

$$P\Lambda^\pm P^{-1} = \nabla f^\pm \quad \text{for some functions } f^\pm : \mathbf{O} \to \mathbf{R}^p. \tag{2.22}$$

Condition (2.22) is equivalent to saying:

$$\nabla\big(P\Lambda^\pm P^{-1}\big) \quad \text{is a symmetric matrix.} \tag{2.22'}$$

Finally we comment upon the positivity conditions (2.1b). Throughout this paper, all matrices under consideration will have real eigenvalues and a full set of eigenvectors. This is the case of the Jacobian matrices $\nabla f^\pm$ in the systems (2.2) provided that the condition in Definition 2.2 holds. In the examples, we shall have to verify the positivity condition (2.1b). Observe first that the condition (2.1b), i.e. $\pm\nabla f^\pm \geq 0$, is *weaker* than the more classical requirement $\pm r^\top \nabla f^\pm r \geq 0$, for all $r \in \mathbf{R}^p$. The latter condition implies the former one, but the two conditions are distinct when applied to *non-symmetric* matrices, which is generally the case here. For further references, we state the following elementary lemma:

Lemma 2.4. *Let us restrict ourselves to matrices having real eigenvalues only.*
–(i) Suppose that p=2. Then the matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \textit{has non-negative eigenvalues}$$

if and only if:

$$\text{(2.23)} \qquad tr(A) = a + d \geq 0, \quad det(A) = ad - bc \geq 0.$$

–(ii) Suppose that p=3. Then the matrix

$$A = \begin{pmatrix} a_1 & b_1 & c_1 \\ b_2 & a_2 & d_1 \\ c_2 & d_2 & a_3 \end{pmatrix} \quad \textit{has non-negative eigenvalues}$$

if and only if:

$$\text{(2.24)} \qquad tr(A) \geq 0,\ a_1a_2 + a_2a_3 + a_3a_1 + b_1b_2 + c_1c_2 + d_1d_2 \geq 0,\ det(A) \geq 0.$$

By contrast, the condition $\pm r^\top A r \geq 0$, for all $r \in \mathbf{R}^p$, is equivalent to saying (when $p = 2$):

$$\text{(2.25)} \qquad \text{tr}(A) = a + d \geq 0, \quad \det(A + A^\top) = 4ad - (b + c)^2 \geq 0.$$

Condition (2.25) implies that a and d are non-negative, but this is not equivalent to the condition (2.23). For instance, the matrix $A = \begin{pmatrix} 3 & 2 \\ -1 & -1/2 \end{pmatrix}$ has two real positive, distinct eigenvalues: it satisfies the inequalities (2.23), but not (2.25).

Finally, we observe that, when a diagonalizable splitting does exist, the system (1.1) must satisfy very peculiar properties. For instance, consider the case $p = 2$, and suppose that a diagonalizable splitting exists. In view of Proposition 2.1, the systems (1.1) and (2.2) have a common set of Riemann invariant coordinates, say (w_1, w_2). In that situation, the entropy function U has to satisfy simultaneously the following three linear PDE's:

$$\text{(2.26)} \qquad (\lambda_2 - \lambda_1)\partial^2_{w_1 w_2} U + \partial_{w_1}\lambda_2 \partial_{w_2} U + \partial_{w_2}\lambda_1 \partial_{w_1} U = 0,$$

and

$$\text{(2.27)} \qquad (\lambda_2^\pm - \lambda_1^\pm)\partial^2_{w_1 w_2} U + \partial_{w_1}\lambda_2^\pm \partial_{w_2} U + \partial_{w_2}\lambda_1^\pm \partial_{w_1} U = 0.$$

In other words, the above PDE's have one common solution, at least. In the significant case of the gas dynamics system, we shall show that equation (2.26) is actually a *linear combination* of the equations in (2.27).

2.3. Examples of Entropy Flux-Splittings. We now turn to the study of two significant examples, the scalar conservation laws and the p-system. In both cases, we establish the existence of a large family of genuine entropy flux-splittings. We treat first the case of a scalar conservation law. A genuine splitting by definition satisfies

$$\begin{cases} \frac{df^+}{du}(u) = f'(u) & \text{if} \quad f'(u) \geq \lambda_0, \\ \frac{df^-}{du}(u) = f'(u) & \text{if} \quad f'(u) \leq -\lambda_0, \end{cases} \tag{2.28}$$

for some nonnegative constant λ_0. Non-genuine splittings correspond to the limiting value $\lambda_0 = \infty$. With the choice $\lambda_0 = 0$, (2.28) selects a *unique splitting*, namely:

$$\frac{df^\pm}{du}(u) = \frac{1}{2}(f'(u) \pm |f'(u)|),$$

and so

$$f^\pm(u) = \frac{1}{2}\int_0^u (f'(\bar{u}) \pm |f'(\bar{u})|)d\bar{u} + \frac{1}{2}f(0) \pm a. \tag{2.29}$$

We define the flux-functions up to a constant, and so we can drop the parameter a. When the function f is convex and achieves its minimum at the point $u = u_*$, we find

$$f^+(u) = \begin{cases} \frac{1}{2}f(u_*), & \text{for } u \leq u_*, \\ f(u) - \frac{1}{2}f(u_*), & \text{for } u \geq u_*, \end{cases} \tag{2.30a}$$

$$f^-(u) = \begin{cases} f(u) - \frac{1}{2}f(u_*), & \text{for } u \leq u_*, \\ \frac{1}{2}f(u_*), & \text{for } u \geq u_*. \end{cases} \tag{2.30b}$$

Formulas (2.29) and (2.30), which correspond to $\lambda_0 = 0$, are due to Engquist-Osher [29]. It follows from the results in [29] that (2.29) and (2.30) are indeed entropy flux-splittings. In fact, Definition 2.2 is easily seen to provide no restriction on the splittings when $p = 1$.

For arbitrary values of λ_0, we state the following result:

Proposition 2.5. *The flux-splittings of class $W^{1,\infty}$ associated with a scalar conservation law have the form:*

$$f^+(u) = g(u), \quad f^+(u) = f(u) - g(u), \tag{2.31}$$

where g is any function satisfying:

$$g \in W^{1,\infty}_{\text{loc}} \text{ with } g' \geq \max(0, f'), \text{ and } \lim_{u \to -\infty} g(u) = 0. \tag{2.32}$$

All these splittings are entropy flux-splittings. Moreover, (2.31)–(2.32) is a genuine splitting iff, for some given constant $\lambda_0 \geq 0$,

$$g'(u) = \max(0, f'(u)) \quad \text{when} \quad |f'(u)| \geq \lambda_0. \tag{2.33}$$

Furthermore, a scalar conservation law admits a unique genuine entropy flux-splitting satisfying condition (2.33) *with the extremal value of the parameter* λ_0, *i.e.* $\lambda_0 = 0$.

Geometrically, it is not hard to determine the functions g that satisfy the conditions (2.32)-(2.33). Consider the graph of the function f'. The graph of the function g' must lie above both the u-axis and the graph of f'. Moreover, the function $g'(u)$ coincides with $f'(u)$ as those points u where $f'(u)$ is larger than λ_0, and coincides with the u-axis when $-f'(u)$ is larger than λ_0. Note that g' is defined almost everywhere only. The function g' can be constructed to be piecewise continuous only. Note that the normalization $\lim_{u \to -\infty} g(u) = 0$ is chosen for definitness only. Finally, we observe that the numerical viscosity of the difference scheme associated with the flux-splittings in Proposition 2.5 is minimal exactly for the choice $\lambda_0 = 0$. This can be checked from the definition in [62], for instance.

Next we treat the case of the p-system, which is the isentropic system of gas dynamics in Lagrangian coordinates:

$$\partial_t w - \partial_x v = 0, \qquad \partial_t v + \partial_x p(w) = 0. \tag{2.34}$$

Here $w > 0$ and v represent the specific volume, and the velocity of the gas, respectively. The pressure $p = p(w)$ is a given function satisfying

$$p \in W^{2,\infty}_{loc}(\mathbf{R}) \quad \text{and} \quad \frac{dp}{dw} < 0, \tag{2.35}$$

which implies that (2.34) is a strictly hyperbolic system. We do not assume that the system is genuinely nonlinear. The system (2.34) has the form (1.1) with

$$u = (w, v), \quad \text{and} \quad f(u) = (-v, p(w)).$$

The eigenvalues and a set of eigenvectors are given by:

$$\begin{cases} \lambda_1(u) = -c(w) < 0 < \lambda_2(u) = c(w), \\ r_1(u) = (1, c(w)), \qquad r_2(u) = (1, -c(w)), \end{cases}$$

where $c(w) = \sqrt{-p'(w)}$ is the sound speed. We recall that (2.34) admits an infinite set of entropy functions U. They are obtained by solving the linear hyperbolic PDE:

$$\partial^2_{ww} U + p'(w) \partial^2_{vv} U = 0, \tag{2.36}$$

which can be done locally in the phase plane, at least. Among them is the physically meaningful entropy, the mechanical energy, (U_*, F_*) defined by

$$U_*(w, v) = W_*(w) + \frac{v^2}{2}, \quad F_*(w, v) = v\, p(w),$$

where $W_*(w) = -\int_0^w p(s)\, ds$ is the internal energy of the gas.

Proposition 2.6. *Let us consider the flux-splittings of class $W^{2,\infty}$ associated with the p-system of gas dynamics (2.34).*

(i) – The splittings consistent with the mechanical energy (U_*, F_*) are as follows:

$$f^+(w,v) = \begin{pmatrix} c(w)^{-2}\partial_w G(w,v) \\ \partial_v G(w,v) \end{pmatrix}, \quad f^-(w,v) = \begin{pmatrix} -v - c(w)^{-2}\partial_w G(w,v) \\ p(w) - \partial_v G(w,v) \end{pmatrix}, \tag{2.37}$$

where G is an arbitrary function of class $W^{2,\infty}$ satisfying the positivity conditions:

$$\begin{aligned} &\partial_w\big(c^{-2}\partial_w G\big) + \partial^2_{vv}G \geq 0, \\ &c^2\partial_w\big(c^{-2}\partial_w G\big)\partial^2_{vv}G - \max\big(|\partial^2_{wv}G|^2, |c^2 + \partial^2_{wv}G|^2\big) \leq 0. \end{aligned} \tag{2.38}$$

(ii) – A splitting consistent with the mechanical entropy U_*, and the dual mechanical entropy U_{**} defined by [28]:

$$U_{**}(w,v) = w\,v, \quad F_{**}(w,v) = -\frac{v^2}{2} + wp(w) + W_*(w) \tag{2.39}$$

is then consistent with all *the entropy functions associated with system (2.34). Those splittings are described by (2.37)–(2.38) where G is an* entropy flux *function to system (2.34), i.e. a solution to the linear hyperbolic PDE:*

$$\partial_w\big(c^{-2}\partial_w G\big) - \partial^2_{vv}G = 0. \tag{2.40}$$

(iii) – The class of diagonalizable splittings coincides with the class of splittings described in (ii).

Proposition 2.6 does not assume system (2.34) to be genuinely nonlinear. We recover the Lax-Friedrichs type splittings (2.6) with $A = aI$ and $B = abI$, a and b being constants, with the choice:

$$\begin{aligned} G(w,v) &= a\big(F_*(w,v) - bF_{**}(w,v)\big) \\ &= \frac{b}{2}v^2 + avp(w) - b\big(W_*(w) + wp(w)\big). \end{aligned} \tag{2.41}$$

and so:

$$f^+(u) = \begin{pmatrix} bw - av \\ bv + ap(w) \end{pmatrix}, \quad f^-(u) = -\begin{pmatrix} bw + (1-a)v \\ bv - (1-a)p(w) \end{pmatrix}.$$

Condition (2.38) holds if

$$b \geq \max\big(|a|, |1-a|\big)\max c(w), \tag{2.42}$$

which, for b large enough, is satisfied on sets of the form $\{(w,v) | c(w) \leq M\}$ ($M > 0$ constant).

Proof of Proposition 2.6. Using the entropy variables

$$\nabla U_* = (-p(w), v) = (c(w)^2, v),$$

it is not difficult to check the statement (i) from the result in Proposition 2.3. In order to prove (ii), we use the characterization (2.5). The splitting (2.37), which is consistent with the entropy U_*, is consistent with another entropy U if and only if:

$$U_{wv}\big(c^{-2}G_w\big)_w + U_{vv}UG_{vw} = U_{ww}c^{-2}G_{vw} + U_{vw}G_{vv}. \tag{2.43}$$

An entropy by definition satisfies:

$$\partial^2_{ww}U + p'(w)\partial^2_{vv}U = 0,$$

thus (2.43) is equivalent to:

$$U_{wv}\big\{G_{vv} - \big(c^{-2}G_w\big)_w\big\} = 0. \tag{2.44}$$

Clearly, it suffices one additional entropy such that $U_{wv} \neq 0$ in order to get the equation (2.40). For the dual mechanical entropy, one has: $U_{wv} = 1$.

We now discuss the derivation of the diagonalizable splittings. With the notation in (2.18)–(2.21), one gets

$$P(u) = \begin{pmatrix} 1 & 1 \\ c & -c \end{pmatrix}, \quad \text{so} \quad P(u)^{-1} = \frac{1}{2c}\begin{pmatrix} c & 1 \\ c & -1 \end{pmatrix},$$

and

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} \lambda_1^+ & 0 \\ 0 & \lambda_2^+ \end{pmatrix} + \begin{pmatrix} \lambda_1^- & 0 \\ 0 & \lambda_2^- \end{pmatrix} = \Lambda^+ + \Lambda^-.$$

Condition (2.22) here reads

$$P\Lambda^\pm P^{-1} = \frac{1}{2c}\begin{pmatrix} 1 & 1 \\ c & -c \end{pmatrix}\begin{pmatrix} \lambda_1^\pm & 0 \\ 0 & \lambda_2^\pm \end{pmatrix}\begin{pmatrix} c & 1 \\ c & -1 \end{pmatrix} = \nabla f^\pm$$

for some functions f^+ and f^-, i.e.,

$$\frac{1}{2c}\begin{pmatrix} c(\lambda_1^\pm + \lambda_2^\pm) & (\lambda_1^\pm - \lambda_2^\pm) \\ c^2(\lambda_1^\pm - \lambda_2^\pm) & c(\lambda_1^\pm + \lambda_2^\pm) \end{pmatrix} = \nabla f^\pm. \tag{2.45}$$

The coefficients of the matrix in (2.45) are supposed to be Lipschitz continuous at least, so (2.45) holds if and only if

$$\begin{aligned} &\partial_v\left(\lambda_1^\pm + \lambda_2^\pm\right) - \partial_w\left(\frac{1}{c}(\lambda_1^\pm - \lambda_2^\pm)\right) = 0, \\ &\partial_v\left(c(\lambda_1^\pm - \lambda_2^\pm)\right) - \partial_w\left(\lambda_1^\pm + \lambda_2^\pm\right) = 0. \end{aligned} \tag{2.46}$$

For convenience, we define auxilliary unknown functions $\alpha^\pm$ and $\beta^\pm$ by:

$$\lambda_2^\pm - \lambda_1^\pm = \alpha^\pm c, \quad \lambda_1^\pm + \lambda_2^\pm = \beta^\pm,$$

and system (2.46) takes the form:

$$\partial_w \alpha^{\pm} + \partial_v \beta^{\pm} = 0, \quad c(w)^2 \partial_v \alpha^{\pm} + \partial_w \beta^{\pm} = 0. \tag{2.47}$$

Equation (2.47) yield a second order linear hyperbolic PDE satisfied by the function $\alpha^{\pm}$

$$\partial^2_{ww} \alpha^{\pm} - \partial^2_{vv} \left(c(w)^2 \alpha^{\pm}\right) = 0, \tag{2.48}$$

which is the equation for an entropy to system (2.34), cf. (2.36). In view of (2.45), the gradient matrix of f is computed in term of the functions $\alpha^{\pm}$ and $\beta^{\pm}$:

$$\nabla f^{\pm} = \frac{1}{2} \begin{pmatrix} \beta\pm & -\alpha^{\pm} \\ -c^2 \alpha^{\pm} & \beta^{\pm} \end{pmatrix}. \tag{2.49}$$

Thus there exists a pair of functions (k, l) such that

$$f^{+}(w, v) = \begin{pmatrix} k(w, v) \\ -l(w, v) \end{pmatrix}.$$

Namely take k such that $\partial_w k = \beta^{+}/2$ and $\partial_v k = \alpha^{+}/2$, which is possible since $\partial_w \alpha^{\pm} + \partial_v \beta^{\pm} = 0$. Similarly, l satisfies: $\partial_w l = c^2 \alpha^{+}/2$ and $\partial_v l = -\beta^{+}/2$. Moreover, for the gradient of f^{+} to have the form (2.49), one need have

$$\partial_v l = -\partial_w k, \qquad \partial_w l = -c^2 \partial_v k,$$

which means that (k, l) is an entropy pair. Similarly, we can check that f^{+} therefore takes the form (2.37). The proof of Proposition 2.6 is complete. □

3. Convergence of Entropy Flux-Splittings Schemes

In this section we consider the finite difference schemes defined from a flux-splitting for system (1.1). We prove that a scheme built up from an entropy splitting satisfies a discrete cell entropy inequality, which has the usual conservative form. The scheme is stable in the L^{∞} norm provided (1.1) has bounded invariant regions and the flux-splitting is a diagonalizable splitting. The convergence follows by compensated compactness arguments for strictly hyperbolic and genuinely nonlinear systems of two equations. When the scheme is built up from an arbitrary flux-splitting, we derive a discrete cell entropy inequality which generally has a nonconservative form.

Let $\tau > 0$ and $h > 0$ be the time and space mesh lengths, respectively. For all integers $n \geq 0$ and j, set $\lambda = \tau/h$, $x_{j+\frac{1}{2}} = (j + \frac{1}{2})h$ and $t_n = n\tau$. We consider finite difference schemes in conservative form for the approximation of (1.1)–(1.4):

$$u_j^{n+1} = u_j^n - \lambda(g^n_{j+\frac{1}{2}} - g^n_{j-\frac{1}{2}}), \quad g^n_{j+\frac{1}{2}} = g(u^n_j, u^n_{j+1}) \tag{3.1}$$

for all integers $n \geq 0$ and j, with u_j^0 given by

$$u_j^0 = \frac{1}{h} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u_0(x) dx. \tag{3.2}$$

The numerical flux $g : \mathcal{O} \times \mathcal{O} \to \mathbf{R}^p$ is a locally Lipschitz continuous function. We always assume that g is consistent with the exact flux-function f, that is, $g(u, u) = f(u)$ for all u in $\mathcal{O}$. By definition, for a flux-splitting scheme, one has

$$g(v, w) = f^+(v) + f^-(w) \qquad \text{for all } v \text{ and } w \text{ in } \mathcal{O}, \tag{3.3}$$

where $f = f^+ + f^-$ is a flux-splitting for system (1.1).

The following decomposition of the scheme (3.1)–(3.3) will be useful in our analysis:

$$u_j^{n+1} = \frac{1}{2}(u_{j-1/2}^+ + u_{j+1/2}^-), \tag{3.4}$$

where

$$u_{j-1/2}^+ = u_j^n - 2\lambda(f^+(u_j^n) - f^+(u_{j-1}^n)), \tag{3.5a}$$

$$u_{j+1/2}^- = u_j^n - 2\lambda(f^-(u_{j+1}^n) - f^-(u_j^n)), \tag{3.5b}$$

for all integers $n \geq 0$ and j. Note that $u_{j-1/2}^+$ and $u_{j+1/2}^-$ are the approximations obtained from the initial data u_j^n, by applying the upwind scheme to systems (2.2a) and (2.2b), respectively. The stability of the scheme requires λ to satisfy a CFL condition:

$$2\lambda \max_{1 \leq i \leq p, u \in K} \left(|\lambda_i^+(u)|, |\lambda_i^-(u)|\right) < 1. \tag{3.6}$$

Decompositions of the form (3.4)-(3.5) were first introduced by Tadmor [62] in his work on the numerical viscosity of E-schemes, and used by Coquel-LeFloch [15] for deriving sharp estimates of the rate of entropy dissipation in an E-scheme.

We shall denote by $R^\pm(u_l, u_r)$ the solution to the Riemann problem:

$$\begin{aligned} &\partial_t u^\pm + \partial_x f^\pm(u^\pm) = 0, \\ &u^\pm(0, x) = \begin{cases} u_l & \text{for } x < 0, \\ u_r & \text{for } x > 0, \end{cases} \end{aligned} \tag{3.7}$$

where u_l and u_r are given in $\mathcal{O}$. It is assumed that, for arbitrary initial data in the set $\mathcal{O}$, the Riemann problem admits a unique solution satisfying the Lax entropy inequalities [38]. Throughout this section, the functions $f^\pm$ are assumed to be at least of class $W^{2,\infty}$. For definiteness, we assume that, for each u, the characteristic eigenvalues of the gradients $\nabla f^\pm(u)$ are either genuinely nonlinear or linearly degenerate and, moreover,

$$\lambda_j^\pm(u) = \lambda_j^\pm(u) \Longrightarrow \nabla\lambda_j^\pm(u) r_j^\pm(u) = 0. \tag{3.8}$$

In other words, if strict hyperbolicity fails at some point u, the corresponding characteristic fields have to be linearly degenerate at that point. This is indeed a sufficient condition for the Riemann problem (3.8) to have a unique solution in the class of piecewise smooth solutions composed of constant states separated by shock discontinuities (satisfying the Lax entropy condition), rarefaction fans, or contact discontinuities. We refer to Lax [38] for an explicit construction. In view of (2.1b), the Riemann solution $R^+(u_l, u_r)$ (respectively $R^-(u_l, u_r)$) contains non-negative (resp. nonpositive) wave speeds, only. In particular, we shall use:

$$\begin{aligned} f^+(R^+(u_l,u_r)(t,x=0)) &= f^+(u_l),\\ f^-(R^-(u_l,u_r)(t,x=0)) &= f^-(u_r). \end{aligned} \tag{3.9}$$

The assumption (3.8) above will be satisfied in the applications we have in mind (cf. Section 4).

We now define the approximate solutions u^h, u^h_+, and u^h_- associated with the flux-splitting scheme. In each elementary rectangle $\{(t,x) : x_j < x < x_{j+1},\, 0 \le t < \tau\}$, we consider the solution u^h_+ to the Riemann problem

$$\begin{aligned} &\partial_t u^h_+ + \partial_x f^+(u^h_+) = 0,\\ &u^h_+(0,x) = \begin{cases} u^0_j, & x < x_{j+1/2},\\ u^0_{j+1}, & x > x_{j+1/2}, \end{cases} \end{aligned}$$

and the solution u^h_- to

$$\begin{aligned} &\partial_t u^h_- + \partial_x f^-(u^h_-) = 0,\\ &u^h_-(0,x) = \begin{cases} u^0_j, & x < x_{j+1/2},\\ u^0_{j+1}, & x > x_{j+1/2}. \end{cases} \end{aligned}$$

The function $u^h(t)$, for $0 < t < \tau$, is defined by

$$u^h(t,x) = \frac{1}{2}\big(u^h_+(t,x) + u^h_+(t,x)\big). \tag{3.10}$$

In view of (3.9), one easily checks that

$$u^1_j = \frac{1}{h}\int_{x_{j-1/2}}^{x_{j+1/2}} u^h(\tau - 0, x)dx.$$

Suppose that $u^h(t)$ has been defined for all $t \le t_n$. Then we define u^h on the cell $\{(t,x) : x_j < x < x_{j+1},\, t_n < t < t_{n+1}\}$ by the formula above where u^h_+ and u^h_- are the solutions to the Riemann problems

$$\begin{aligned} &\partial_t u^h_+ + \partial_x f^+(u^h_+) = 0,\\ &u^h_+(t_n,x) = \begin{cases} u^n_j, & x < x_{j+1/2},\\ u^n_{j+1}, & x > x_{j+1/2}, \end{cases} \end{aligned} \tag{3.11}$$

and

$$\begin{aligned} &\partial_t u^h_- + \partial_x f^-(u^h_-) = 0,\\ &u^h_-(t_n,x) = \begin{cases} u^n_j, & x < x_{j+1/2},\\ u^n_{j+1}, & x > x_{j+1/2}, \end{cases} \end{aligned} \tag{3.12}$$

respectively. Again, with (3.10), we have

$$u_j^{n+1} = \frac{1}{h} \int_{x_{j-1/2}}^{x_{j+1/2}} u^h(t_{n+1} - 0, x) dx. \tag{3.13}$$

This completes the construction of the flux-splitting approximation. ¿From now on we assume that the conditions (3.6) and (3.7) are satisfied, and (3.3) is an entropy flux-splitting.

The following result concerns the entropy consistency.

Proposition 3.1. *Let* (3.3) *be defined from an entropy flux-splitting. Then the scheme* (3.1)–(3.3) *satisfies the local discrete entropy inequalities:*

$$U(u^{\pm}_{j+1/2}) - U(u_j^n) + 2\lambda\big(F^{\pm}(u_{j+1}^n) - F^{\pm}(u_j^n)\big) \le 0, \tag{3.14}$$

and thus:

$$U(u_j^{n+1}) - U(u_j^n) + \lambda\big(F^{+}(u_j^n) + F^{-}(u_{j+1}^n) - F^{+}(u_{j-1}^n) - F^{-}(u_j^n)\big) \le 0. \tag{3.15}$$

If the family (u^h) *is uniformly bounded in the* L^∞ *norm, and converges almost everywhere to an element* $u \in L^\infty$, *then* u *satisfies* (1.1), (1.2), *and* (1.4) *in the sense of distributions.*

Proof of Proposition 3.1. By assumption, U is an entropy function for both systems in (2.2): we denote by $F^{\pm}$ the corresponding entropy flux-functions. The Riemann solution $R^{\pm}(u_l, u_r)$, by construction, satisfies the Lax shock entropy inequalities. According to [38], it also satisfies the distributional entropy inequality:

$$\partial_t U(u^{\pm}) + \partial_x F^{\pm}(u^{\pm}) \le 0. \tag{3.16}$$

Indeed, thanks to the assumption (3.9), (3.16) holds even if strict hyperbolicity fails at some point. Inequality (3.14) follows by writing (3.16) with the functions $u^h_{\pm}$, and integrating it on the rectangles: $\{(t,x) : x_j < x < x_{j+1/2}, t_n < t < t_{n+1}\}$. Finally (3.15) is a consequence of (3.14) and Young's inequality for the convex function U. The passage to the limit into (3.15) is a classical matter by Lax-Wendroff theorem since (3.14) has the usual conservative form. □

Using the principle of invariant regions due to Chueh-Conley-Smoller [11], we can check the L^∞ stability of the scheme. As shown in Section 2, systems (2.2) and (1.1) have the same Riemann invariants, and invariant regions. Based on the decomposition (3.5), it is easy to see that:

Proposition 3.2. *Assume that system* (1.1) *admits a set of convex and bounded invariant regions, and* (3.3) *is based on a diagonalizable flux-splitting. Then the scheme* (3.1) – (3.3) *preserves the invariant domains. and so is stable in the* L^∞ *norm: if the initial data* u_0 *takes its values in a bounded invariant domain, then so does* u^h.

Remark 3.1. Partial results of L^∞ stability can be obtained for nondiagonalizable splittings, provided there exists a (not necessarily bounded in all directions) region left invariant by systems (2.2) and (1.1). This is actually often the case in the applications: for instance, the non-negative sign of the mass density in the systems of gas dynamics is usually an invariant.

Our main result of convergence is as follows.

Theorem 3.3. *Let (1.1) be a strictly hyperbolic system of two conservation laws with genuinely nonlinear characteristic fields. Assume that the system admits an entropy flux-splitting that satisfies the conditions in* Definition 2.2 *for all convex entropy functions. We also assume that the Riemann problem for systems (2.2) can be solved for large data, and the solutions satisfy the Lax entropy inequalities. Let u^h be the family of approximate solutions defined from the flux-splitting scheme (3.1) – (3.3). Suppose that either the splitting is a diagonalizable one, and system (1.1) admits a convex and bounded invariant region and u_0 takes its values in this domain, or instead u^h is uniformly bounded in L^∞. Then u^h converges in the strong L^p norm ($p < \infty$) to an entropy weak solution to the Cauchy problem (1.1) – (1.4).*

Proof of Theorem 3.3. In view of Proposition 3.2, u^h satisfies

$$\|u^h\|_{L^\infty(\mathbf{R}_+\times\mathbf{R})} \leq O(1), \tag{3.17}$$

where $O(1)$ is independent on h. Following DiPerna [27], we only need observe that, for any C^2 entropy-entropy flux pair (V, G), the dissipation measure satisfies:

$$\partial_t V(u^h) + \partial_x G(u^h) \in \text{ compact set of } \quad H^{-1}_{loc}, \tag{3.18}$$

and

$$\partial_t V(u^h) + \partial_x G(u^h) \leq E^h \to 0 \text{in the distrobutional sense as } h \to 0. \tag{3.19}$$

Indeed conditions (3.18) and (3.19) follow from the entropy inequalities derived in Proposition 3.1. Properties (3.17)–(3.19) together imply that u^h verifies the assumption of the compensated compactness framework for strictly hyperbolic systems established by DiPerna. According to the results in [27], u^h converges strongly to an entropy weak solution. The proof of Theorem 3.3 is completed. □

Remark 3.2. Theorem 3.3 applies to the example of the system of gas dynamics in Lagrangian coordinates, studied at the end of Section 2. It also applies to the system of gas dynamics in Eulerian coordinates provided that attention is restricted to solutions without vacuum points. Arbitrary solutions including possibly the vacuum are treated in the next section.

We now extend the result in Proposition 3.1 to arbitrary flux-splitting schemes. In that case, the entropy inequality has a nonconservative form and is derived by using the notion of nonconservative product due to Dal Maso-LeFloch-Murat [24], and LeFloch-Liu [43]. We introduce first some notation. Let $u_l = m_0^\pm, m_1^\pm, \cdots, m_p^\pm = u_r$ be the constant states associated with the Riemann solution $R^\pm(u_l, u_r)$. For each real- or vector- valued function a, the integral of a along the path naturally associated with $R^{\pm}(u_l, u_r)$ is defined by

$$\int_{R^{\pm}(u_l,u_r)} a = \sum_{k-\text{waves}} \int_0^1 a(m_k^{\pm}(s))ds, \tag{3.20}$$

where $m_k^{\pm}(s) = (1-s)m_{k-1}^{\pm} + s m_k^{\pm}$ if the wave connecting $m_{k-1}^{\pm}$ to $m_k^{\pm}$ is a shock or a contact discontinuity and, if it is a rarefaction wave, $m_k^{\pm}$ satisfies the following ordinary differential equation for some $C > 0$:

$$\frac{dm_k^{\pm}(s)}{ds} = C r_k(m_k^{\pm}(s)), \qquad m_k^{\pm}(0) = m_{k-1}^{\pm}, \quad m_k^{\pm}(1) = m_k^{\pm}. \tag{3.21}$$

Given a pair (u_l, u_r), we denote the corresponding path by $\phi(u_l, u_r)$. It is not hard to see that the family of paths ϕ satisfies the assumption made in [24]. Henceforth a nonconservative product $\alpha(u)\frac{du}{dx}$ of a composite function of a function of bounded variation u by the measure $\frac{du}{dx}$ is well defined as a locally bounded Borel measure. Following the notation in [24], we denote this measure by $\left[\alpha(u)\frac{du}{dx}\right]_{\phi}$. Observe that the product depends upon the choice of the family of paths. Using the paths described above based on the Riemann solution is essential in what follows. (Note that the so-called Volpert's product would not be sufficient for our purpose.)

Theorem 3.4. *Let* (1.1) *be a system of p conservation laws and* (3.1)–(3.3) *be a scheme based on an arbitrary flux-splitting for system* (1.1). *Suppose that the set $\mathcal{O}$ is small enough. Then the following local discrete entropy inequalities hold:*

$$U(u_{j+1/2}^{\pm}) - U(u_j^n) + 2\lambda \int_{R^{\pm}(u_j^n, u_{j+1}^n)} \nabla U \nabla f^{\pm} \le 0, \tag{3.22}$$

and thus

$$U(u_j^{n+1}) - U(u_j^n) + \lambda \int_{R^{+}(u_{j-1}^n, u_j^n)} \nabla U \nabla f^{+} + \lambda \int_{R^{-}(u_j^n, u_{j+1}^n)} \nabla U \nabla f^{-} \le 0, \tag{3.23}$$

Theorem 3.4 provides a weak form of the entropy inequality (3.4). Note that (3.23) has a nonconservative form and, in general, it is unclear whether (3.4) can be deduced from (3.23) even under the assumption of strong convergence of the scheme.

Remark 3.3. A similar inequality was proved in [50] for the so-called Osher-Solomon's scheme. Actually although this scheme is not quite a flux-splitting, it nevertheless shares many of their properties. Observe that the inequality in [50] has a conservative form however.

Proof of Theorem 3.4. The proof is similar to the one of Proposition 3.1. We have to prove that the solution $u^{\pm}$ to the Riemann problem $R^{\pm}(u_l, u_r)$ satisfies the following nonconservative version of the entropy inequality (3.4):

$$\partial_t U(u^{\pm}) + \left[\nabla U \nabla f^{\pm} \partial_x u\right]_{\phi} \le 0. \tag{3.24}$$

Inequalities (3.22) and (3.23) follow by integration of (3.24).

In view of the definition of the paths in (3.20), the inequality (3.24) clearly holds in the rarefaction fans since a rarefaction is a smooth solution to system (1.1). We only need to check (3.24) across a shock discontinuity. By computing its Taylor expansion, we now prove that

$$\Omega(u_r) = -\sigma(u_l, u_r)(U(u_r) - U(u_r)) - \int_0^1 \nabla U(u_s)\nabla f^{\pm}(u_s)(u_r - u_l)ds \leq 0, \tag{3.25}$$

provided that $|u_r - u_l|$ is small enough with u_l kept fixed, and the left and right states satisfy the Rankine-Hugoniot relation and the Lax entropy inequalities.

We have

$$\Omega(u_r) = \int_0^1 \nabla U(u(s))\big(-\sigma(u_l, u_r)I - \nabla f^{\pm}(u(s))\big)(u_r - u_l)ds$$

since $u'(s) = u_r - u_l$, where I is the unit matrix. The Rankine-Hugoniot condition yields the relation

$$\int_0^1 \big(-\sigma(u_l, u_r)I - \nabla f^{\pm}(u(s))\big)(u_r - u_l)ds,$$

so that

$$\Omega(u_r) = \int_0^1 \big(\nabla U(u(s)) - \nabla U(u_l)\big)\big(-\sigma(u_l, u_r)I - \nabla f^{\pm}(u(s))\big)(u_r - u_l)ds.$$

At this stage we make use of the *linearity* of the path $u(s)$ and, setting $u(s,\tau) = (1-\tau)u_l + \tau u(s)$, we observe that $u(s,\tau) - u_l = \tau s(u_r - u_l)$, thus

$$\Omega(u_r) = \int_0^1\int_0^1 (u_r - u_l)^{\top}\nabla^2 U(u(s,\tau))\big(-\sigma(u_l, u_r)I - \nabla f^{\pm}(u(s))\big)(u_r - u_l)\tau s\,ds\,d\tau.$$

Finally, after decomposition of $u_r - u_l$ on the basis of eigenvectors for $\nabla f^{\pm}$, specifically

$$u_r - u_l = \sum_{1 \leq k \leq p} \alpha_k(s) r_k^{\pm}(u(s)),$$

we get the following formula

(3.26)

$$\Omega(u_r) = \sum_{1\leq k\leq p} \iint r_k^{\pm}(u(s))^{\top}\nabla^2 U(u(s,\tau))\big(-\sigma(u_l, u_r) - \lambda_k^{\pm}(u(s))\big) r_k^{\pm}(u(s))\tau s\,ds\,d\tau,$$

where the integral is over $(0,1)\times(0,1)$. Consider first a genuinely nonlinear i–characteristic field. Then it is known after Lax [37] that:

$$u_r = u_l + \epsilon_i r_i^{\pm}(u_l) + O(\epsilon_i^2), \quad \sigma(u_l, u_r) = \lambda_i^{\pm}(u_l) + \epsilon_i \nabla\lambda_i^{\pm}(u_l) r_i^{\pm}(u_l) + O(\epsilon_i^2),$$

where $\epsilon_i \le 0$, so (3.26) gives

$$\Omega(u_r) = (\epsilon_i)^3 \big(\nabla\lambda_i^{\pm}(u_l) r_i^{\pm}(u_l)\big) r_i^{\pm}(u(s))^{\top} \nabla^2 U(u(s,\tau)) r_k^{\pm}(u(s)), \tag{3.27}$$

which is clearly non-positive.

The case of a linearly degenerate characteristic field is trivial, since then $\Omega(u_r)$ vanishes identically.

When two or more eigenvalues coincide, part of the integral in (3.26) vanishes while the other is nonpositive. The proof of (3.25) is complete. □

Finally we mention briefly that some results in this paper, in particular Theorem 3.4 remain valid for the class of schemes of the form [18]:

$$u_j^{n+1} = u_j^n - \lambda \int_{u_{j-1}^n}^{u_j^n} A^+(\bar{u}) d\bar{u} - \lambda \int_{u_j^n}^{u_{j+1}^n} A^-(\bar{u}) d\bar{u}, \tag{3.28}$$

where the matrix-valued functions $A^{\pm}$ satisfy

$$A^+(u) + A^-(u) = \nabla f(u), \quad \text{for all } u \quad \text{in } \mathcal{O}, \tag{3.29}$$

and the above integrals are performed along a given family of paths in $\mathbf{R}^p$, say $\phi(s; u_j^n, u_{j+1}^n)$, $s \in [0,1]$, for all (u_j^n, u_{j+1}^n). In other words, for instance,

$$\int_{u_j^n}^{u_{j+1}^n} A^+(\bar{u}) d\bar{u} = \int_0^1 A^+(\phi(s; u_j^n, u_{j+1}^n)) \partial_s \phi(s; u_j^n, u_{j+1}^n) ds.$$

Note that in this case the underlying systems have the following nonconservative form:

$$\partial_t v + A^+(v)\partial_x v = 0 \qquad \text{and} \qquad \partial_t w + A^-(w)\partial_x w = 0. \tag{3.30}$$

This idea of paths is due to [24], where the hyperbolic systems in nonconservative form are studied. This approach has been proposed and investigated recently by Coquel-Liou [18] for designing new schemes adapted to treat systems with linearly degenerate fields, as is the case for most physical systems.

Remark 3.4. 1) It would be interesting to extend the results in this paper to the high order accurate schemes based on a flux-splitting. Higher order accurate versions of the flux-splitting schemes can be constructed according to the so-called corrected antidiffusive flux approach. The entropy consistency can be proved by combining the results in the present paper and the techniques in [10, 14, 15]. Proving the L^∞ stability for high order schemes however seems difficult since the usual formulations destroy the invariant regions of the system. See [10] for a possible technique to deal with the difficulty.

2) Another open issue concerns the existence and the asymptotic stability of the discrete shock profiles associated with an entropy flux-splitting scheme.

3) One can use the Glimm scheme instead of the upwind scheme to approximate the systems (2.2a) and (2.2b) associated with flux-functions f^+ and f^-. This does not seem to be of pratical interest however. It would be interesting (at least from a theoretical standpoint) to search for a decomposition of the flux-function, similar to the one given here for difference schemes, which would simplify the analysis of the Glimm scheme.

4. A Family of Entropy Flux-Splittings for the Euler Equations

In this section we prove, Theorem 4.1, that the system of isentropic Euler equations admits a family of genuine entropy flux-splittings, associated with the mechanical energy, which plays the role of an entropy function here. The general result is illustrated in Proposition 4.2 with an example, that can be seen as an "entropy modification" of the van Leer splitting. Our formula is very similar to the formula in [39]: a polynomial form with a few terms only.

Next we prove in Theorem 4.3 that there exists a unique genuine entropy flux-splitting that satisfies *all* the entropy inequalities. Several properties of this splitting are pointed out, e.g. it is the only diagonalizable splitting. It is also the only splitting that satisfies two nontrivial entropies, at least. When $\gamma = 1+\frac{1}{2m+1}$, $m \geq 2$ integer, it is actually a polynomial function of the Mach number. Finally we check, Theorem 4.4, that the latter splitting can be derived as well by using the kinetic formulation proposed recently by Lions-Perthame-Tadmor [46].

In the present paper, we restrict ourselves to the case of a polytropic perfect gas with adiabatic exponent $1 < \gamma \leq 3$, refering to [9] for the treatment of the real gas case, and the system of non-isentropic gas dynamics.

Our results here, as well as in [9] on the real gas, are expected to extend to situations where no kinetic formulation is available. Furthermore, we can argue that our variant to van Leer splittings are fully explicit. Kinetic splittings are often given by integral formula, that can be integrated out for special values of the adiabatic exponent γ only. Numerical tests are in progress, and will be the subject of a future publication.

We consider the system of isentropic gas dynamics [22]:

$$\begin{aligned} &\partial_t \rho + \partial_x(\rho v) = 0, \\ &\partial_t(\rho v) + \partial_x(\rho v^2 + p(\rho)) = 0, \end{aligned} \tag{4.1}$$

where $\rho \geq 0$ and v are the density and the velocity of the gas, respectively. We restrict ourselves to the polytropic perfect gas:

$$p(\rho) = \frac{\rho^\gamma}{\gamma} \quad \text{with } \gamma \geq 1. \tag{4.2}$$

System (4.1) has the form (1.1) with $u = (\rho, \rho v)^\top$ and $f(u) = (\rho v, \rho v^2 + p(\rho))^\top$. The following notations will be useful: $m = \rho v$ (moment), $c = \sqrt{p'(\rho)} = \rho^\theta$ (sound speed), and $M = v/c$ (Mach number). Here $\theta = (\gamma - 1)/2$. Eigenvalues and a set of eigenvectors for system (4.1) are given by: $\lambda_1 = v - c$, $\lambda_2 = v + c$, $r_1 = (1, v-c)^\top$, and $r_2 = (1, v+c)^\top$. The flux-function can be written in the form:

$$f(u) = \begin{pmatrix} \rho c(\rho) a(M) \\ \rho c(\rho)^2 b(M) \end{pmatrix}, \; a(M) = M, \; b(M) = M^2 + \frac{1}{2\theta + 1}. \tag{4.3}$$

System (4.1) is strictly hyperbolic everywhere except at the points where $p'(\rho) = 0$, which occurs only at the vacuum state $\rho = 0$.

Our first objective is deriving a large family of entropy flux-splittings for system (4.1). For simplicity, we shall be mostly interested in tha case that the flux-functions $f^{\pm}$ have the form:

$$f^{\pm}(u) = \begin{pmatrix} \rho c(\rho) a^{\pm}(M) \\ \rho c(\rho)^2 b^{\pm}(M) \end{pmatrix}. \tag{4.4}$$

The actual restriction is that the functions a and b are independent of ρ. This choice is consistent with (4.3). For a real gas, the functions a, $a^{\pm}$, etc, also depend upon ρ as well, cf. [9]. For a γ law gas, one can construct indeed more general splittings than (4.4). Since

$$\begin{aligned} a^{+}(M) + a^{-}(M) &= M, \\ b^{+}(M) + b^{-}(M) &= M^2 + \frac{1}{2\theta + 1}, \end{aligned} \tag{4.5}$$

it is enough to determine one of the two sets of functions $(a^{\pm}, b^{\pm})$.

The following theorem describes all the possible entropy splittings. For definitness, we consider those consistent with the (physically meaningful) entropy U_*:

$$U_*(\rho, m) = \frac{m^2}{2\rho} + \frac{\rho^{\gamma}}{\gamma(\gamma - 1)}.$$

The corresponding entropy flux is chosen to be:

$$F_*(\rho, m) = m^3/2\rho^2 + m\rho^{\gamma-1}/((\gamma - 1)),$$

The result below can be easily extended to other entropy functions.

Theorem 4.1. *The system of isentropic Euler equation for a polytropic perfect gas admits a family of entropy flux-splittings $f = f^{+} + f^{-}$ associated with the mechanical energy U_*. The splittings take the form:*

$$f^{\pm}(\rho, m) = \nabla_{w,z} \psi^{\pm}(w, z), \tag{4.6}$$

where (w, z) denotes the entropy variables defined by:

$$w = u, \quad z = \frac{c(\rho)^2}{2\theta} - \frac{m^2}{2\rho^2}, \tag{4.7}$$

and the functions $\psi^{\pm}$ are of class C^2 and satisfy the consistency condition:

$$\psi^{+}(w, z) + \psi^{-}(w, z) = \psi(w, z) \equiv \frac{mc(\rho)^2}{2\theta + 1}, \tag{4.8a}$$

and the positivity condition:

(4.8b)

$$\pm \begin{pmatrix} w\partial_{ww}\psi^{\pm} + \partial_{wz}\psi^{\pm} & w\partial_{wz}\psi^{\pm} + \partial_{zz}\psi^{\pm} \\ ((\theta+1)w^2 + 2\theta z)\partial_{ww}\psi^{\pm} - w\partial_{wz}\psi^{\pm} & ((\theta+1)w^2 + 2\theta z)\partial_{wz}\psi^{\pm} - w\partial_{zz}\psi^{\pm} \end{pmatrix} \geq 0.$$

In the case of the splittings (4.4), the functions a^+ and b^+ (the formulas for a^- and b^- then are deduced from (4.5)) have the form:

$$\begin{aligned} a^+(M) &= (3\theta+1)k(M) - \theta M k'(M), \\ b^+(M) &= (3\theta+1)Mk(M) + (1-\theta M^2)k'(M) \end{aligned} \tag{4.9}$$

where $k : \mathbf{R} \to \mathbf{R}$ is any function of class $W^{2,\infty}$ satisfying (4.8b). The splitting is a genuine splitting if, moreover,

$$k(M) = \begin{cases} 0, & \text{for } M \le -M_*, \\ \frac{M}{2\theta+1}, & \text{for } M \ge M_*, \end{cases}$$

for some $M_* > 0$. Moreover, if the function k is of class C^{k+1} then the splitting is of class C^k.

Proof of Theorem 4.1. Using the entropy variables (4.7), system (4.1) takes the following symmetric form:

$$\partial_t \nabla_{w,z}\phi + \partial_x \nabla_{w,z}\psi = 0, \tag{4.10a}$$

with

$$\begin{aligned} \phi(w,z) &= \frac{1}{2\theta+1}\big(\theta w^2 + 2\theta z\big)^{\frac{2\theta+1}{2\theta}}, \\ \psi(w,z) &= \frac{w}{2\theta+1}\big(\theta w^2 + 2\theta z\big)^{\frac{2\theta+1}{2\theta}}. \end{aligned} \tag{4.10b}$$

Note that $\theta w^2 + 2\theta z = c^2$, so that:

$$\phi = \rho\frac{c(\rho)^2}{2\theta+1}, \quad \psi = m\frac{c(\rho)^2}{2\theta+1}.$$

By Proposition 2.3, the splittings have precisely the form (4.6), (4.8). The positivity requirement $\nabla_{\rho,m} f^\pm = \nabla_{\rho,m}(w,z)\nabla_{w,z} f^\pm \ge 0$ is equivalent to (4.8b) in view of:

$$\nabla_{\rho,m}(w,z) = \frac{1}{\rho}\begin{pmatrix} -w & 1 \\ (\theta+1)w^2 + 2\theta z & -w \end{pmatrix}.$$

The proof of Theorem 4.1 is completed. □

It can be checked that the classical Steger-Warming splitting, and the van Leer splitting are not entropy satisfying, even for another entropy than U_*. One can check that the function k in Theorem 4.1 can not be taken to be linear. However k can be taken to be quadratic, and one recovers a variant of Van Leer splitting.

Proposition 4.2. *Consider the system of isentropic gas dynamics for a γ law gas. The following formula defines a genuine entropy flux-splitting of class C^1 consistent with the entropy U_*:*

$$\begin{aligned} a^+(M) &= \alpha\big((\theta+1)M^2 + 2(2\theta+1)AM + (3\theta+1)A^2)\big) \\ b^+(M) &= \alpha\big((\theta+1)M^3 + 2(2\theta+1)AM^2 + (3\theta+1)A^2)M + 2M + 2A\big) \end{aligned} \tag{4.11}$$

with $1/\alpha = 4(2\theta+1)A$ and $A > 0$ is any number such that:

$$A^2 \ge \frac{8}{3\theta^2}. \tag{4.12}$$

Our next result concerns the existence and uniqueness of an entropy flux splitting for the Euler equations, that satisfies *all* the entropy inequalities.

Theorem 4.3. *1) – The system of isentropic gas dynamics (4.1) for a γ–law gas admits a unique genuine entropy flux-splitting $f(u) = f^+(u) + f^-(u)$, that satisfies all the entropy inequalities, and have the form (4.4).*
2) – Moreover this splitting admits a closed integral form (given after the statement of the theorem), and satisfies the property:

$$\begin{cases} f^+(u) = f(u), \quad f^-(u) = 0, & \text{if} \quad M = v/c \geq 1/\theta, \\ f^+(u) = 0, \quad f^-(u) = f(u), & \text{if} \quad M = v/c \leq -1/\theta. \end{cases} \tag{4.13}$$

When $1 < \gamma \leq 3$, the functions $(a^\pm, b^\pm)$ are of class $C^{\frac{\gamma+1}{2(\gamma-1)}}(\mathbf{R})^2$, and the systems associated with the flux-functions f^+ and f^- are strictly hyperbolic in the regions $\{M > -1/\theta\}$ and $\{M < 1/\theta\}$, respectively. When $\gamma = 1 + \frac{2}{2m+1}$, m being a positive integer, the restrictions of the functions $a^\pm$ and $b^\pm$ to the interval $(-1/\theta, 1/\theta)$ are polynomial functions of degrees $2m+2$ and $2m+3$, respectively.
3) – This splitting can also be characterized as the only splitting that satisfies –at least– the entropy inequalities associated with the two entropy functions:

$$U_*(\rho, m) = m^2/2\rho + \frac{\rho^\gamma}{\gamma(\gamma-1)}, \quad \text{and} \quad U_{**}(\rho, m) = \frac{m}{\rho}.$$

In other words two non trivial entropies are sufficient to select a unique splitting.
4) – This splitting is also the only splitting that is diagonalizable, i.e. such that the matrices $\nabla f^\pm$ have the same eigenvectors as ∇f. In other words the latter is the only splitting consistent with all the invariant regions associated with system (4.1).

As opposed to the splitting found in Proposition 4.2, the splitting above is given by an integral formula, that may be inconvenient in numerical computation. Moreover, even though one entropy is not enough to get a convergence proof, one entropy inequality is believed to be sufficient to ensure uniqueness of the solution. These two arguments argue in favor of the splitting in Proposition 4.2.

We shall see in the proof of Theorem 4.3 that the function $a^+(M)$ and $b^+(M)$ in Theorem 4.3 are given by:

$$a^+(M) = \begin{cases} 0, \quad \theta M \leq -1, \\ c_\theta M \int_{-1}^{\theta M} (1 - \frac{y}{\theta M})(1-y^2)^{\frac{1-\theta}{2\theta}} dy, \quad |\theta M| \leq 1, \\ M, \quad \theta M \geq 1. \end{cases} \tag{4.14}$$

and

$$b^+(M) = \begin{cases} 0, \quad \theta M \leq -1, \\ c_\theta \int_{-1}^{\theta M} (M^2 + 1 - \frac{\theta+1}{\theta^2} y^2)(1-y^2)^{\frac{1-\theta}{2\theta}} dy, \quad |\theta M| \leq 1, \\ M^2 + \frac{1}{2\theta+1}, \quad \theta M \geq 1. \end{cases} \tag{4.15}$$

with

$$c_\theta = \int_{-1}^{1} (1-y^2)^{\frac{1-\theta}{2\theta}} dy.$$

The expressions for the functions $a^-(M)$ and $b^-(M)$ are similar, and can be deduced from (4.5).

Proof of Theorem 4.3. The Jacobian matrix of $f^{\pm}(u)$ is computed by differentiation of the formula (4.4):

$$Df^{\pm} = \begin{pmatrix} c(\theta+1)(a^{\pm} - M\frac{da^{\pm}}{dM}) & \frac{da^{\pm}}{dM} \\ c^2\big((2\theta+1)b^{\pm} - (\theta+1)M\frac{db^{\pm}}{dM}\big) & c\frac{db^{\pm}}{dM} \end{pmatrix}. \tag{4.16}$$

Let U be an arbitrary entropy for system (4.1). The matrix $\nabla^2 U \nabla f^{\pm}$ is symmetric if and only if the coefficients a and b (we drop the indices from now on) satisfy the following first order differential equation:

$$\begin{aligned} \partial^2_{\rho\rho}U\,\frac{da}{dM} + c\,\partial^2_{\rho m}U\,\big((\theta+1)(M\frac{da}{dM} - a) + \frac{db}{dM}\big) \\ + c^2\,\partial^2_{mm}U\big((\theta+1)M\frac{db}{dM} - (2\theta+1)b\big) = 0. \end{aligned} \tag{4.17}$$

On the other hand, U, being an entropy, satisfies the second order hyperbolic equation:

$$\partial^2_{\rho\rho}U + 2\,c\,M\,\partial^2_{\rho m}U + c^2\,(M^2-1)\,\partial^2_{mm}U = 0. \tag{4.18}$$

Combining (4.17) and (4.18), we get

$$\begin{aligned} c\,\partial^2_{\rho m}U\,\big((\theta-1)M\frac{da}{dM} - (\theta+1)a + \frac{db}{dM}\big) \\ + c^2\,\partial^2_{mm}U\big((1-M^2)\frac{da}{dM} + (\theta+1)M\frac{db}{dM} - (2\theta+1)b\big) = 0. \end{aligned} \tag{4.19}$$

Plugging in (4.19) the Hessian matrix $\nabla^2 U_*$ of the mechanical energy:

$$\nabla^2 U_*(u) = \frac{1}{\rho}\begin{pmatrix} c^2(M^2+1) & -cM \\ -cM & 1 \end{pmatrix},$$

formula (4.19) becomes:

$$(1-\theta M^2)\frac{da}{dM} + (\theta+1)Ma + \theta M\frac{db}{dM} - (2\theta+1)b = 0. \tag{4.20}$$

Note in passing that, plugging (4.9) in (4.20), we see that (4.20) indeed hold for any choice of cuntion k. This provides us with an alternative way to derive (4.9).

Plugging now the entropy U_{**} in (4.19) and since

$$\nabla^2 U_{**}(u) = \frac{1}{\rho^2}\begin{pmatrix} 2\rho c M & -1 \\ -1 & 0 \end{pmatrix},$$

we have:

$$(1-M^2)\frac{da^{\pm}}{dM} + (1+\theta)M\frac{db^{\pm}}{dM} - (2\theta+1)b = 0. \tag{4.21}$$

Therefore, in view of (4.20) and (4.21), the functions a and b satisfy the following two equations:

$$(4.22)\qquad \begin{aligned}&(1-\theta)M\frac{da^\pm}{dM}+(1+\theta)a-\frac{db^\pm}{dM}=0,\\&(1-M^2)\frac{da^\pm}{dM}+(1+\theta)M\frac{db^\pm}{dM}-(2\theta+1)b=0.\end{aligned}$$

We now derive the conditions corresponding to the diagonalizable splittings. In view of (2.22), the flux-splitting have the form:

$$\nabla f^\pm = P\Lambda^\pm P^{-1}\equiv \begin{pmatrix}1 & 1\\ v-c & v+c\end{pmatrix}\begin{pmatrix}\lambda_1^\pm & 0\\ 0 & \lambda_2^\pm\end{pmatrix}\begin{pmatrix}1 & 1\\ v-c & v+c\end{pmatrix}^{-1},$$

that is,

$$(4.23)\quad \nabla f^\pm = \begin{pmatrix}(M+1)\lambda_1^\pm+(-M+1)\lambda_2^\pm & (-\lambda_1^\pm+\lambda_2^\pm)/c\\ (M^2-1)c\lambda_1^\pm+(-M^2+1)c\lambda_2^\pm & (-M+1)\lambda_1^\pm+(M+1)\lambda_2^\pm\end{pmatrix}.$$

¿From (4.4) and (4.23), we obtain a set of differential equations for the coefficients $\lambda_i^\pm$, $a^\pm$, and $b^\pm$. Using the new unknowns $\tilde\lambda_i^\pm\equiv\frac{1}{c}\lambda_i^\pm$, one has:

$$\begin{aligned}(k+1)a^\pm-(k+1)M\frac{da^\pm}{dM}&=(M+1)\tilde\lambda_1^\pm+(-M+1)\tilde\lambda_2^\pm,\\ \frac{da^\pm}{dM}&=-\tilde\lambda_1^\pm+\tilde\lambda_2^\pm,\\ (2k+1)b^\pm-(k+1)M\frac{db^\pm}{dM}&=(M^2-1)\tilde\lambda_1^\pm+(-M^2+1)\tilde\lambda_2^\pm,\\ \frac{db^\pm}{dM}&=(-M+1)\tilde\lambda_1^\pm+(M+1)\tilde\lambda_2^\pm.\end{aligned}$$

By elimination of the $\tilde\lambda_i^\pm$'s, and after some additional computation, we conclude that $a^\pm(M)$ and $b^\pm(M)$ must necessarily satisfy the same set of equations as before, cf. (4.22)! The proof of the part 3) of the theorem is completed.

We shall use below the following formula for the eigenvalues $\lambda_i^\pm$:

$$(4.24)\qquad \begin{aligned}\lambda_1^\pm&=c\Big((1-M)\frac{da^\pm}{dM}+\frac{db^\pm}{dM}\Big),\\ \lambda_i^\pm&=c\Big(-(1+M)\frac{da^\pm}{dM}+\frac{db^\pm}{dM}\Big).\end{aligned}$$

We now solve the system of partial differential equations (4.22). The functions $a^\pm$ can be determined first, and the functions $b^\pm$ be deduced from the $a^\pm$'s. Namely, $a^\pm$ solves the following second order P.D.E.:

$$(4.25)\qquad (1-M^2\theta^2)\frac{d^2a^+}{dM^2}(M)+\theta(\theta+1)M\frac{da^+}{dM}(M)-\theta(\theta+1)a^+=0,$$

and $b^+(M)$ satisfies a first order PDE system:

$$\begin{aligned} \frac{db^\pm}{dM} &= (1-\theta)M\frac{da^\pm}{dM} + (1+\theta)a, \\ (2\theta+1)b &= (1-\theta^2)\frac{da^\pm}{dM} + (1+\theta)^2 Ma. \end{aligned} \tag{4.26}$$

The first equation is automatically satisfied, and b^+ is an explicit function of a:

$$b^+(M) = \frac{1}{2\theta+1}\big\{(1-\theta^2M^2)\frac{da^+}{dM}(M) + (\theta+1)^2 Ma(M)\big\}. \tag{4.27}$$

We now give the construction of the function $a^+(M)$; the case of $a^-(M)$ is similar. We consider the domain $\Delta = \big\{M \,/\, |M| \le 1/\theta\big\}$, and in its complement Δ^c. We first solve the differential equation in the domain Δ^c, with the following condition at infinity:

$$a^+(M) \to M, \text{ resp. } 0, \quad \text{as } M \to \pm\infty. \tag{4.28}$$

It is a simple matter to check that a is then the trivial solution:

$$a^+(M) = a(M), \quad \text{for all } |M| \ge 1/\theta, \tag{4.29}$$

The condition at infinity (4.28) is thus sufficient to show that the splitting is a genuine flux-splitting. We next solve the system in Δ using suitable boundary conditions deduced from (4.29). Specifically, we have now to find a function a^+ defined in the domain Δ, that is a solution to (4.25), satisfies the positivity conditions

$$\lambda_i^- \le 0, \quad \text{and} \quad \lambda_i^+ \ge 0, \tag{4.30}$$

and the two boundary conditions at the boundary $\partial\Delta$:

$$a^+(-1/\theta) = 0, \quad a^+(1/\theta) = 1/\theta. \tag{4.31}$$

Equation (4.25) is a singular O.D.E.: the coefficient of the term d^2a^+/dM^2 vanishes at the end points $M = \pm 1/\theta$. Existence of a solution to this two-point boundary problem does not seem to be a standard matter. However it is not hard to show by a direct calculation that there is a unique solution to the boundary value problem (4.25), (4.31), given by the following formula:

$$a^+(M) = -\frac{2M\int_{-\frac{1}{\theta}}^{M} \frac{(1-\theta^2y^2)^{\frac{\theta+1}{2\theta}}}{y^2}dy}{\theta(\theta+1)\int_{-\frac{1}{\theta}}^{\frac{1}{\theta}}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}dy} \tag{4.32}$$

for all $M \ne 0$ in the interval $(-1/\theta, 1/\theta)$. After integration by parts in the above formula, we finally get:

(4.33)

$$a^+(M) = \begin{cases} 0, \quad M \le -\frac{1}{\theta}, \\ \frac{\big\{(1-\theta^2M^2)^{\frac{\theta+1}{2\theta}} + \theta(\theta+1)M\int_{-\frac{1}{\theta}}^{M}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}dy\big\}}{\theta(\theta+1)\int_{-\frac{1}{\theta}}^{\frac{1}{\theta}}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}dy}, & -\frac{1}{\theta} \le M \le \frac{1}{\theta}, \\ M, \quad M \ge \frac{1}{\theta}. \end{cases}$$

The corresponding function b^+ is:
(4.34)

$$b^+(M) = \begin{cases} 0, \quad M \le -\frac{1}{\theta}, \\ \dfrac{\left\{(\theta+1)^2 M(1-\theta^2M^2)^{\frac{\theta+1}{2\theta}} + \theta(\theta+1)(1+(2\theta+1)M^2)\int_{-\frac{1}{\theta}}^{M}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}\,dy\right\}}{\theta(\theta+1)(2\theta+1)\int_{-\frac{1}{\theta}}^{\frac{1}{\theta}}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}\,dy}, \quad -\frac{1}{\theta} \le M \le \frac{1}{\theta}, \\ M^2 + \frac{1}{2\theta+1}, \quad M \ge \frac{1}{\theta}. \end{cases}$$

It remains to show that the Jacobian matrix Df^+ is non-negative; the treatment of Df^- is similar. It suffices to consider the range of values $-1/\theta < M < 1/\theta$. We want to show that

$$Df^+(\rho, m) = \frac{D}{D(\rho, m)}\begin{pmatrix} a^+(M) \\ b^+(M) \end{pmatrix} > 0. \tag{4.35}$$

The eigenvalues of $Df^+(\rho, m)$ are given by:

$$\lambda_1^+ = c(\theta A + a^+ - \frac{da^+}{dM}), \quad \text{and} \quad \lambda_2^+ = c(\theta A + a^+ + \frac{da^+}{dM}), \tag{4.36}$$

where the function $A = A(M)$ is defined by

$$\begin{aligned} A(M) &= -M\frac{da^+}{dM}(M) + a^+(M) \\ &= \frac{(1-\theta^2M^2)^{\frac{\theta+1}{2\theta}}}{\theta(\theta+1)\int_{-\frac{1}{\theta}}^{\frac{1}{\theta}}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}\,dy}. \end{aligned} \tag{4.37}$$

We observe that

$$\theta A(M) + a^+(M) \pm \frac{da^+}{dM}(M) = \frac{g(M)}{\int_{-\frac{1}{\theta}}^{\frac{1}{\theta}}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}\,dy},$$

where we have defined the function $g = g(M)$ by

$$g(M) = (1-\theta^2M^2)^{\frac{1\pm\theta}{2\theta}} + \theta(M \pm 1)\int_{-\frac{1}{\theta}}^{M}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}\,dy.$$

Since the function g satisfies the conditions

$$\begin{aligned} &g(-\frac{1}{\theta}) = g'(-\frac{1}{\theta}) = 0, \\ &g'(M) = \theta(1-\theta)(1+\theta^2)(1-\theta^2M^2)^{\frac{1-3\theta}{2\theta}} \ge 0, \end{aligned}$$

we have

$$\lambda_1^+ > 0, \quad \lambda_2^+ > 0, \quad -\frac{1}{\theta} < M < \frac{1}{\theta},$$

which proves (4.35) for $-1/\theta < M < 1/\theta$.

Furthermore, in view of (4.36), we have

$$\lambda_2^+ - \lambda_1^+ = c\frac{da^+}{dM}(M) = \frac{\int_{-\frac{1}{\theta}}^{M}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}\,dy}{\int_{-\frac{1}{\theta}}^{\frac{1}{\theta}}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}\,dy} > 0, \quad -\frac{1}{\theta} < M < \frac{1}{\theta}.$$

It follows that the system associated with the flux-function f^+ is strictly hyperbolic for all $M > -1/\theta$. We can also check that Df^-, and the system with the flux f^- is strictly hyperbolic for all $M > -1/\theta$. This completes the proof of Theorem 4.3. □

Remark 4.1. 1)– The assumption made in the beginning of Section 3 in order to solve locally the Riemann problem is satisfied here.

2)– When $1 < \gamma \le \frac{5}{3}$ which is the typical range of values for γ in gas dynamics, we have (at least) $(a^\pm, b^\pm) \in C^2$. Steger-Warming's splitting is only Lipschitz continuous, while van Leer's splitting is of class C^1. The regularity of the splitting is a desirable property for the computation of steady state solutions based on a time-dependent scheme (e.g. van Leer [39] and Osher-Solomon [50]).

3)– When $\gamma = 1 + \frac{2}{2m+1}$, the functions $a^\pm(M)$ and $b^\pm(M)$ are polynomials, and our splitting is then similar to the splittings in [60, 40]. For instance, let us consider the case $m = 1$, i.e. $\gamma = \frac{5}{3}$. The formula (4.17) becomes:

$$a^+(M) = \begin{cases} 0, & \text{if } M \le -3, \\ -\frac{M^4}{108} + \frac{M^2}{2} + M + \frac{9}{4}, & \text{if } -3 < M < 3, \\ M, & \text{if } M \ge 3. \end{cases}$$

It is easy to check directly that $a^+ \in C^2$, but $a^+_{SW} \in \text{Lip}$, only.

Finally, we show that the splitting we derived in Theorem 4.3 could also be derived by the so-called kinetic approximation. We follows here the approach adopted by Perthame [52] for the systems of non-isentropic gas dynamics, and extend his idea to our case, i.e. the isentropic Euler system.

Theorem 4.4. *The kinetic flux-splitting that is associated with the kinetic formulation due to Lions-Perthame-Tadmor, coincides with the entropy flux-splitting consistent with all entropy functions derived in Theorem 4.3.*

Proof of Theorem 4.4. In view of the formulation in [46], and using the notation introduced in the beginning of this section, we obtain:

$$a^+(\rho, M) = \int_{\{\xi \ge 0\}} (\theta\xi + (1-\theta)v)\,\Big(\frac{c(\rho)^2}{\theta^2} - (\xi - v)^2\Big)_+^{\frac{1-\theta}{2\theta}}\,d\xi, \tag{4.38}$$

where, by definition, $z_+ = \max(0, z)$. The formula (4.41) can be written as:

$$a^+(\rho, M) = \rho^{\theta+1}\theta^{\frac{\theta-1}{\theta}} \int_M^\infty (M + \theta y)\,\Big(\frac{1}{\theta^2} - y^2\Big)_+^{\frac{1-\theta}{2\theta}}\,dy. \tag{4.39}$$

This is exactly the formula (4.36). Checking b is similar. The proof of Theorem 4.4 is completed. □

5. Convergence of a Flux-Splitting Scheme for the Euler Equations

In this section, we extend to the Euler equations the result of convergence of the flux-splitting schemes proved in Section 3.

Our objective is to prove the strong convergence of the difference scheme associated with the flux-splitting derived in Theorem 4.3, cf. formulas (4.22)-(4.23). Section 3 already guarantees the convergence of the scheme provided that the initial data takes its values in a small neighborhood of a given point away from vacuum. We are going to show that the convergence holds for arbitrary large initial data possibly containing the vacuum. For the analysis of the flux-splitting scheme, it is convenient to study first the Riemann problems associated with flux-functions f^+ and f^-, respectively.

Theorem 5.1. *The Riemann problem associated with the system with the flux-function f^+ given by (4.22) (respectively f^- given by (4.23)) and corresponding to arbitrary left and right states admit a global solution composed of two elementary waves that are either a shock wave or a rarefaction wave. Each wave has a non-negative speed (resp. non-positive speed), and each shock wave with speed σ connecting the left state (ρ_-, m_-) to the right state (ρ_-, m_-) satisfies the entropy inequality:*

$$\sigma(U(\rho_r, m_r) - U(\rho_l, m_l)) - (F(\rho_r, m_r) - F(\rho_l, m_l)) \geq 0, \tag{5.1}$$

for any convex entropy-entropy flux pair (U, F) of system (4.1). *Furthermore, the system admits bounded and convex regions that are invariant for the Riemann problem (i.e. if the Riemann data lies in such a region, so does the Riemann solution.)*

In particular, it follows from the proof of Theorem 5.1 that the systems with flux $f^\pm$ have *genuinely nonlinear* characteristic fields (except in the region where they vanish identically).

Proof of Theorem 5.1. For definitness, we treat the case of the function f^+, and for simplicity often write a for a^+, etc. First of all, since the system with flux f^+ has the same right eigenvectors as the original system (4.1), they have the same set of rarefaction curves. In particular, the rarefaction curve are globally defined and convex. In order to solve the Riemann problem, we only need prove that the shock curves exist globally and are convex (so that there will be a unique point of intersection), and that the entropy condition is satisfied all along the shock curves.

We consider the shock wave curves in the phase plane (ρ, m). Let (ρ_0, m_0) be the state on the left of a discontinuity traveling with speed σ. For definiteness, we only consider the case that (ρ_0, m_0) belongs to the region $\mathcal{D} = \{|M| < 1/\theta\}$, i.e.

$$1 + \theta M_0 > 0, \quad 1 - \theta M_0 > 0.$$

Other cases are trivial or much simpler.

The state (ρ, m) on the right of the discontinuity must satisfy the Rankine-Hugoniot relations

$$\begin{cases} \sigma(\rho - \rho_0) & = \frac{1}{2}\left(\rho^{\theta+1} a^+(\frac{m}{\rho^{\theta+1}}) - \rho_0^{\theta+1} a^+(\frac{m_0}{\rho_0^{\theta+1}})\right), \\ \sigma(m - m_0) & = \frac{1}{2}\left(\rho^{2\theta+1} b^+(\frac{m}{\rho^{\theta+1}}) - \rho_0^{2\theta+1} b^+(\frac{m_0}{\rho_0^{\theta+1}})\right), \end{cases} \tag{5.2}$$

and the Lax entropy condition ($i = 1, 2$):

$$\lambda_i^+(\rho, m) \leq \sigma_i^+ \leq \lambda_i^+(\rho_0, m_0). \tag{5.3}$$

We shall focus our attention on the part of the shock curve which lies inside the region $\mathcal{D}$, and we use the abreviated notation $(a, b, \lambda, \sigma) = (a^+, b^+, \lambda^+, \sigma^+)$. (Other regions and other cases for M_0 are similar or much simpler.)

In view of (5.2), we get the equation for the state on the right of the discontinuity

$$\frac{m - m_0}{\rho - \rho_0} = \frac{\rho^{2\theta+1} b(M) - \rho_0^{2\theta+1} b(M_0)}{\rho^{\theta+1} a(M) - \rho_0^{\theta+1} a(M_0)}, \tag{5.4}$$

or equivalently

$$\frac{t^{\theta+1} M - M_0}{t - 1} = \frac{t^{2\theta+1} b(M) - b(M_0)}{t^{\theta+1} a(M) - a(M_0)}, \tag{5.5}$$

where we have set $t = \rho/\rho_0$.

Let us check first that the equation (5.5) has two roots denoted by

$$M^\pm = M^\pm(t, M_0), \quad \text{for each value} \quad t > 0,\ t \neq 1. \tag{5.6}$$

Clearly, one has $M^\pm(1, M_0) = M_0$. It sufficies to prove that, for any $t > 0, t \neq 1$, there exists a unique $\bar{M}(t, M_0)$ such that

$$(M - \bar{M}) F_M(M, t; M_0) > 0, \quad M \neq \bar{M}, \tag{5.7}$$

where the function F is defined in view of (5.5) by

$$\begin{aligned} &F(M, t; M_0) \\ &= (t^{\theta+1} M - M_0)(t^{\theta+1} a(M) - a(M_0)) - (t - 1)(t^{2\theta+1} b(M) - b(M_0)). \end{aligned} \tag{5.8}$$

The first and second order derivatives of F with respect to the variable M are found to be:

$$\begin{aligned} &F_M(M, t; M_0) \\ &= t^{\theta+1}\left\{t^{\theta+1} a(M) - a(M_0) + a'(M)(t^{\theta+1} M - M_0) - (t - 1) t^\theta b'(M)\right\}, \end{aligned} \tag{5.9a}$$

$$\begin{aligned} F_{MM}(M, t; M_0) &= t^{\theta+1}\left\{2t^\theta a'(M) + (t^\theta(\theta t + 1 - \theta) M - M_0) a''(M)\right\} \\ &\equiv G(M, t; M_0). \end{aligned} \tag{5.9b}$$

Using the O.D.E. satisfied by the function a, we can compute explicitly the first order derivative of the function $G = F_{MM}$:

$$G_M(M, t; M_0) \equiv C_0 (1 - \theta^2 M^2)^{\frac{1-3\theta}{2\theta}} H(M, t; M_0), \tag{5.10}$$

where we have defined C_0 and H by

$$C_0 = \frac{2\theta(\theta+1)}{\theta(\theta+1)\int_{-\frac{1}{\theta}}^{\frac{1}{\theta}}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}dy},\tag{5.11}$$
$$H(M,t;M_0) = -\theta t^\theta(\theta t+1+\theta)M^2 + \theta(1-\theta)M_0 M + t^\theta(\theta t+3-\theta).$$

In the case $M_0 < 0$, we have

$$G_M(M,t;M_0) = C_0(1-\theta^2M^2)^{\frac{1-3\theta}{2\theta}}H(M,t;M_0) > 0, \quad \text{for all} \quad M \in [-\frac{1}{\theta}, 0],$$

and

$$G(-\frac{1}{\theta}, t; M_0) = 0,$$
$$G_M(M,t;M_0) > 0, \quad \text{for all} \quad M \geq 0.$$

Therefore, if $M_0 < 0$, we obtain

$$G(M,t;M_0) \geq 0.\tag{5.12}$$

In the case $M_0 > 0$, one easily checks that for all $M > \frac{M_0}{t^\theta(\theta t+3-\theta)}$:

$$F_{MM}(M,t;M_0) = t^{\theta+1}\{2t^\theta(a'(M)-Ma''(M))+(t^\theta(\theta t+3-\theta)M-M_0)a''(M)\} > 0,$$

and there exist two values M_1 and M_2 with $-\frac{1}{\theta} < M_1 < 0 < \frac{M_0}{t^\theta(\theta t+3-\theta)} < M_2 < \frac{1}{\theta}$, such that

$$G_M(M,t;M_0) > 0, \quad \text{for all} \quad M \in (M_1, M_2).\tag{5.13}$$

From (5.12), (5.13), and the relations

$$F_M(-\frac{1}{\theta}, t; M_0) = -t^{\theta+1}a(M_0) < 0,$$
$$F_M(M,t;M_0) \to \infty, \quad \text{as } M \to \infty,$$

we deduce that there exists some point $\bar{M} \in (-\frac{1}{\theta}, \infty)$ such that

$$(M-\bar{M})F_M(M,t;M_0) > 0.\tag{5.14}$$

This completes the proof of (5.7), and thus establishes the existence of two curves satisfying the Rankine-Hugoniot. In the following, we consider ρ_0 and M_0 as fixed and ρ (or equivalently $t = \rho/\rho_0$) as a parameter along the shock curve. The shock curves (actually half curves) are selected as follows:

– Along the 1–shock curve: $t > 1$, $M > M_0$,

– Along the 2–shock curve: $t > 1$, $M < M_0$

Next we prove that the shock curves are convex. It is convenient to show that the slope of the curve in (ρ, m) is a monotone function of t. By a lengthy but direct computation, we obtain the following expression for the derivative of the slope $(m-m_0)/(\rho-\rho_0)$:

$$\frac{d\left(\frac{m-m_0}{\rho-\rho_0}\right)}{d\rho}=\frac{t^{\theta+1}a'(M)}{\rho_0(t^{\theta+1}a(M)-a_0(M))^4F_M}\left\{t^{2\theta}-\left(\frac{t^\theta M-M_0}{t-1}\right)^2\right\}$$

where $M^\pm=M^\pm(t,M_0)$. One can check (see Appendix, Lemma 1) that

$$t^{2\theta}-\left(\frac{t^\theta M^\pm(t,M_0)-M_0}{t-1}\right)^2\neq 1,\quad \text{for all}\quad t\neq 1,$$

which implies that

$$\frac{d}{d\rho}\left(\frac{m-m_0}{\rho-\rho_0}\right)\neq 0. \tag{5.15}$$

Henceforth, the two shock curves are convex curves in the $\rho-m$ plane.

Furthermore, after a tedious but direct computation, the derivative of the shock speed $\sigma(\rho)$ along the shock curve is found to be:

$$\begin{aligned}2\sigma'(\rho)&=\frac{d}{d\rho}\left(\frac{\rho^{\theta+1}a(M)-\rho_0^{\theta+1}a(M_0)}{\rho-\rho_0}\right)\\&=\frac{\rho_0^{\theta-1}t^{\theta+1}}{F_M}\left\{\left(\frac{t^{\theta+1}a(M)-a(M_0)}{t-1}+t^\theta(\theta Ma'(M)-(\theta+1)a(M))\right)^2-t^{2\theta}(a'(M))^2\right\}.\end{aligned}$$

It can be checked (cf. Appendix, Lemma 2) that the term in the left hand side never vanishes for $t\neq 1$. In view of the Lax entropy condition (5.3), we thus obtain

(i)– along the 1-shock curve: $\rho>\rho_0,\quad \sigma'(\rho)\leq 0$,

(ii)– along 2-shock curve: $\rho<\rho_0,\quad \sigma'(\rho)\geq 0$.

Therefore the Riemann problem for the system with flux f_+ and arbitrary initial states admits a global solution that satisfies the entropy condition (5.1). This completes the proof of Theorem 5.1. □

Finally we state and prove the convergence of the flux-splitting scheme.

Theorem 5.2. *Consider the system of gas dynamics in Eulerian coordinates* (4.1) *with an initial data satisfying* $0\leq\rho^h\leq M$ *and* $0\leq|v_0(x)|\leq M$ *for some* M. *Let* (ρ^h,m^h) *be the approximate solutions given by the flux-splitting scheme* (3.1)-(3.3) *and* (4.22)-(4.23). *Then the sequence* (ρ^h,m^h) *converges strongly to an entropy weak solution to* (4.1).

Proof of Theorem 5.2. We use the compactness framework developed in [6,26]. It is sufficient to prove that the approximate solutions $u^h\equiv(\rho^h,m^h)$ satisfy the following three conditions:

(A_1) $0\leq\rho^h,|\frac{m^h}{\rho^h}|\leq M<\infty$,

(A_2) $\partial_tU(u^h)+\partial_xF(u^h)\in$ compact set of H^{-1}_{loc},

(A_3) $\partial_t U(\rho^h, m^h)+\partial_x F(\rho^h, m^h) \le E^h \to 0$ in the distributional sense as $h \to 0$, for any C^2 convex weak entropy pair (U, F).

The condition (A_1) is a consequence of the convex decomposition (3.4)–(3.6), and the property of the Riemann solutions. We recall that the Riemann problem admits bounded convex invariant regions.

In order to check the conditions (A_2) and (A_3), we study the dissipation measures $\partial_t U(u^h)+\partial_x F(u^h)$ for any C^2 convex weak entropy pair (U, F). For any $\phi \ge 0, \phi \in C_0^1(\Omega_T)$ with $\Omega_T = (-\infty, \infty) \times [0, T), T = mh$, we can use Green's formula and obtain:

$$
\begin{aligned}
&\iint_{\Omega_T} \left(U(u^h)\partial_t\phi + F(u^h)\partial_x\phi\right) dxdt + \int_{\mathbf{R}} \phi(0,x)U(u^h(0,x))dx \\
&= S^h(\phi) + L_1^h(\phi) + L_2^h(\phi) + M^h(\phi).
\end{aligned}
\tag{5.16}
$$

Here one has set:

$$
\begin{aligned}
S^h(\phi) &= \int_0^T \sum\{\sigma[U] - [F]\}\phi(x(t), t)dt, \\
L_1^h(\phi) &= \sum_{j,n} \phi_j^n \int_{x_{j-1/2}}^{x_{j+1/2}} (U(u_-^{t_n}) - U(u_j^{t_n}))dx, \\
L_2^h(\phi) &= \sum_{j,n} \int_{x_{j-1/2}}^{x_{j+1/2}} (U(u_-^{t_n}) - U(u_j^{t_n}))(\phi - \phi_j^n)dx, \\
M^h(\phi) &= \int_{\mathbf{R}} \phi(x,T)U(u^h(x,T))dx - \int_{\mathbf{R}} \phi(0,x)U(u^h(0,x))dx
\end{aligned}
$$

where $u_-^{t_n} = u^h(t_n - 0, x) = \left(u_+^h(t_n - 0, x) + u_-^h(t_n - 0, x)\right)/2$, $\phi_j^n = \phi(t_n, x_j)$, and the summation in $S^h(\phi)$ is taken (for each $t > 0$) over all shock waves in $u^h(t)$ with location $x(t)$.

Due to the finiteness of the propagation speed associated with the scheme, we can assume without loss of generality that the initial data $(\rho_0(x), m_0(x))$ has compact support. Say $u^h = (\rho^h, m^h)$ are compactly supported in the region Ω_T and

$$
\int_{-\infty}^{\infty} U_*(u_0(x))dx \le C < \infty.
$$

For simplicity in the notation, we drop the index h in $u^h(t,x)$. Let us substitute

$$
U = U_* = \frac{1}{2}\rho v^2 + \frac{1}{\gamma(\gamma-1)}\rho^\gamma, \quad F = F_* = \frac{1}{2}\rho v^3 + \frac{1}{\gamma - 1}\rho^\gamma v,
$$

and $\phi \equiv 1$ in the identity (5.16). We find

$$
\begin{aligned}
&\sum_{j,n} \int_{x_{-1/2}}^{x_{j+1/2}} (U_*(u_-^{t_n}) - U_*(u_j^{t_n}))dx + \int_0^T \sum\{\sigma[U_*] - [F_*]\}\phi(t, x(t))dt \\
&= \int_{\mathbf{R}} U_*(u^h(0,x))dx - \int_{\mathbf{R}} U_*(u^h(x,T))dx \\
&\le \int_{\mathbf{R}} U_*(u_0(0,x))dx \le C.
\end{aligned}
$$

Note that U_* is a convex entropy for both splitting systems and the original system. The Riemann solutions are constants in one of half cells $(x_{j-1}, x_{j-1/2}) \times (t_n, t_{n+1})$ and $x_{j-1/2}, x_j) \times (t_n, t_{n+1})$, therefore satisfies the entropy condition :

$$\sigma[U_*] - [F_*] \geq 0$$

across the shock waves. Moreover, using the averaging feature of the splitting scheme, we have

$$\sum_{j,n} \phi_j^n \int_{x_{j-1/2}}^{x_{j+1/2}} (U_*(u_-^{t_n}) - U_*(u_j^n))dx$$
$$= \sum_{j,n} \phi_j^n \int_{x_{j-1/2}}^{x_{j+1/2}} dx \int_0^1 (1-\theta)(u_-^{t_n} - u_j^n)^\top \nabla^2 U_*(u_j^n + \theta(u_-^{t_n} - u_j^n))(u_-^{t_n} - u_j^n) d\theta \geq 0.$$

We get:

$$\int_0^T \sum \{\sigma[U_*] - [F_*]\}\phi(t, x(t))dt \leq C,$$
$$\sum_{j,n} \phi_j^n \int_{x_{j-1/2}}^{x_{j+1/2}} dx \int_0^1 (1-\theta)(u_-^{t_n} - u_j^n)^\top \nabla^2 U_*(u_j^n + \theta(u_-^{t_n} - u_j^n))(u_-^{t_n} - u_j^n) d\theta \leq C.$$

In particular, we have

$$\sum_{j,n} \int_{x_{j-1/2}}^{x_{j+1/2}} |u_-^{t_n} - u_j^n|^2 dx \leq C.$$

Consider a function U, that is a convex entropy function for both systems (1.1) and (2.2). The Riemann solutions take constant values in one of the half cells $(x_{j-1}, x_{j-1/2}) \times (t_n, t_{n+1})$ and $(x_{j-1/2}, x_j) \times (t_n, t_{n+1})$, and satisfy the entropy condition. So, for any $\phi \geq 0, \phi \in C_0^1(\Omega_T)$, we get:

$$S^h(\phi) \geq 0,$$

and

$$\sum_{j,n} \phi_j^n \int_{x_{j-1/2}}^{x_{j+1/2}} (U(u_-^{t_n}) - U(u_j^n))dx$$
$$= \sum_{j,n} \phi_j^n \int_{x_{j-1/2}}^{x_{j+1/2}} dx \int_0^1 (1-\theta)(u_-^{t_n} - u_j^n)^\top \nabla^2 U(u_j^n + \theta(u_-^{t_n} - u_j^n))(u_-^{t_n} - u_j^n) d\theta \geq 0.$$

Moreover, for any $\phi \in C_0^\alpha(\Omega_T), \frac{1}{2} < \alpha < 1$, we have

$$\begin{aligned}
|L_2^h(\phi)| &\leq \sum_{j,n} \int_{x_{j-1/2}}^{x_{j+1/2}} |\phi(x, nh) - \phi_j^n||U(u_-^{t_n}) - U(u_j^n)|dx \\
&\leq h^{\alpha-1/2} \|\nabla U\|_{L^\infty} \|\phi\|_{C_0^\alpha} \left(\sum_{j,n} \int_{x_{j-1/2}}^{(j+1/2l)} |u_-^{t_n} - u_j^n|^2 dx \right)^{1/2} \\
&\leq Ch^{\alpha-1/2} \|\phi\|_{C_0^\alpha} \\
&\leq Ch^{\alpha-1/2} \|\phi\|_{W_0^{1,p}}, \quad p > \frac{n}{1-\alpha};
\end{aligned}$$

that is,

$$\|L_2^h(\phi)\|_{W^{-1,q_0}} \leq C\,h^{\alpha-1/2} \to 0,\ (h \to 0), \quad 1 < q_0 < \frac{n}{n-1+\alpha} < \frac{n}{n-1}.$$

We have

$$|M^h(\phi)| \leq C\|\phi\|_{C_0},$$

which implies:

$$M^h \in \text{ compact set of } W^{-1,q_0}.$$

Note that

$$E^h \equiv S^h + L_1^h + L_2^h + M^h \in \text{ bounded set of } W^{-1,q}, \quad 1 < p \leq \infty,$$

so that we have

$$E^h - L_2^h - M^h \in \text{ bounded set of } W^{-1,q_0},$$
$$E^h - L_2^h - M^h \geq 0 \quad \text{in } W^{-1,q_0}.$$

Using Murat's Lemma [47], i.e. that the embedding of the positive cone of W^{-1,q_0} in W^{-1,q_1} is completely continuous for all $q_1 < q_0$, we have

$$E^h - L_2^h - M^h \in \text{ compact set of } W^{-1,q_1}.$$

Therefore we obtain:

$$E^h \equiv S^h + L^h + M^h \in \text{ compact set of } W^{-1,q_1}.$$

Using a result by [26], we have that the embedding of a compact set of $W^{-1,q}$ and a bounded set of $W^{-1,r}$ is compact in H^{-1} for $1 < q \leq 2 < r < \infty$. From the above, we deduce that

$$S^h + L^h + M^h \in \text{ compact set of } H^{-1}_{loc}.$$

This proves the condition (A_2). The condition (A_3) can be obtained similarly as in [6].

This completes the proof of Theorem 5.2. □

Acknowledgements. This work was initiated when the first author visited the Courant Institute, and the second author was a Courant Instructor there from 1990 to 1992. The authors are grateful to Jonathan Goodman, Robert Kohn, Peter Lax and Cathleen Morawetz for their interest to this work, and motivating discussions. This work also benefited from a stimulating discussion with Constantine Dafermos, during a visit of the authors to Brown University. Part of the work of the second author was done while he was visiting the University of California at Los Angeles, where the hospitality of Stanley Osher was very much appreciated.

Gui-Qiang Chen's research was supported in part by NSF grant DMS-9207080, ONR grant N00014-91-J-1384, and by an Alfred P. Sloan Foundation Fellowship. Philippe LeFloch's research was supported in part by NSF grants DMS-92-09326 and DMS-94–09400, by ONR grant N00014-91-J-1384 when he visited the University of Chicago, and by NSF grant DMS-88-06731 when he was at Courant Institute.

References

[1] Brenier Y., *Résolution d'équations d'évolution quasilinéaires en dimension N d'espace à l'aide d'équations linéaires en dimension $N+1$*, J. Diff. Equa. 50, 1983, pp.375–390.

[2] Brenier Y., *Averged multivalued solutions for scalar conservation laws*, SIAM J. Numer. Anal. 21, 1984, pp.1013–1037.

[3] Abgrall R., *Généralisation du schéma d'Osher pour le calcul des écoulements de mélanges de gaz parfaits à concentrations variables et de gas réels*, La Recherche Aérospatiale 4, 1989, pp. 1–13.

[4] Bourdel F., Delorme P., and Mazet P.A., *Convexity in hyperbolic problems: Application to a discontinuous Galerkin method*, Proc. Intern. Conf. on Hyperbolic problems, Aachen (Germany), march 1988, Notes on Num. Fluid Mech. 24, Viewieg, Braunschweig, 1989, pp. 31–42.

[5] Chen, G.-Q., *Convergence of the Lax-Friedrichs scheme for isentropic gas dynamics III*, Acta Math. Sci. 6, 1986, pp. 75-120; (Chinese) 8, 1988, pp. 243-276.

[6] Chen, G.-Q., *The compensated compactness method and the system of isentropic gas dynamics*, Preprint MSRI-00527-91, Mathematical Sciences Research Institute, Berkeley, 1990.

[7] Chen G.-Q., Du Q., and Tadmor E., *Spectral viscosity approximation to multidimensional scalar conservation laws*, Math. of Comp., 1993.

[8] Chen G.-Q., and LeFloch P.G., *Entropy flux-splittings for hyperbolic conservation laws*, C. R. Acad. Sc. Paris 317, 1993, pp.761–766.

[9] Chen G.-Q., and LeFloch P.G., *Entropy flux-splittings for hyperbolic conservation laws. Part II: Gas dynamic equations*, in preparation, 1994.

[10] Chen G.-Q., and Liu J.-G., *Convergence of second-order schemes for isentropic gas dynamics*, to appear in Math. of Comp., 1993.

[11] Chueh, K., Conley, C., and Smoller, J., *Positively invariant regions for systems of nonlinear diffusion equations*, Indiana U. Math. J. 26, 1977, pp. 373-392.

[12] Cockburn B., Coquel F., and LeFloch P.G., *Convergence of the finite volume method for multidimensional conservation laws*, SIAM J. on Num. Anal., 1994.

[13] Cockburn B., Coquel F., and LeFloch P.G., *An error estimate for finite volume methods for multidimensional conservation laws*, Math. of Comp., 1994.

[15] Coquel F., and LeFloch P.G., *Convergence of finite difference schemes for conservation laws in several space variables: general theory*, SIAM J. on Numer. Anal. 30, 1993, pp. 675–700.

[16] Coquel F., and LeFloch P.G., *Convergence of finite difference schemes for scalar conservation laws in several space dimensions: the corrected antidiffusive flux approach*, Math. of Comp. 57, 1991, pp. 169–210.

[17] Coquel F., and Le Floch P.G., *Convergence of the finite volume method on arbitrary triangulations*, Preprint, Courant Institute, New York University, december 1991.

[18] Coquel F., and Liou M.-S., *Field by field hybrid upwind splitting methods*, ICASE Langley Report, 1993.

[19] Crandall M., and Majda A., *Monotone difference approximations for scalar conservation laws*, Math. of Comp. 34, 1980, pp. 1–21.

[20] Croisille J.P., *Contribution à l'étude théorique et à l'approximation de la dynamique des gaz*, Thesis Univ. Paris IV, 1990.

[21] Croisille J.P., and Delorme P., *Kinetic symmetrizations and pressure laws for the Euler equations*, Preprint Univ. Paris XI, 1991.

[22] Courant F., and Friedrichs K.O., Supersonic Flows and Shock Waves, Wiley Interscience, New York, 1948.

[23] Dafermos C. M., *Hyperbolic systems of conservation laws*. In: Systems of Nonlinear Partial Differential Equations, NATO Adv. Sci. Inst. Ser. C. Math. Phys. Sci. 111, 1983, pp. 25-70.

[24] Dal Maso G., LeFloch P.G., and Murat F., *Definition and stability of nonconservative products*, Preprint CMAP, Ecole Polytechnique (France). To appear in J. d'Analyse Math. (1994).
[25] Deshpande S.M., *A second-order accurate kinetic theory based method for inviscid compressible flows*, NASA Langley Report 2613, 1986 (unpublished).
[26] Ding X., Chen G.-Q., and Luo P., *Convergence of the Lax-Friedrichs scheme for isentropic gas dynamics I, II*, Acta Math. Sci. 5, 1985, pp. 483-540; (Chinese) 7, 1987 pp. 467-480; 8, 1988, pp. 64-91.
[27] DiPerna R.J., *Convergence of approximate solutions to conservation laws*, Arch. Rat. Mech. Anal. 82, 1983, pp. 27–70.
[28] DiPerna R.J., *Compensated compactness and genreal systems of conservation laws*, Trans. Amer. Math. Soc. 292, 1985, pp. 383–420.
[29] Engquist B., and Osher S.J., *One-sided difference approximations for nonlinear conservation laws*, Math. of Comp. 35, 1981, pp. 321–351.
[30] Harten A., Hyman J.M., and Lax P.D., *On finite-difference approximations and entropy conditions for shocks*, Comm. Pure Appl. Math. 29, 1976, pp. 297–322.
[31] Harten A., Lax P.D. and van Leer B., *On upstream differencing and Godunov-type schemes for hyperbolic conservation laws*, SIAM Rev. 25, 1983, pp. 35–61.
[32] Hoff, D., *Bounded invariant domains*, J. Diff. Equa.
[33] Johnson C., and Szepessy A., *Convergence of a finite element methods for a nonlinear hyperbolic conservation law*, Math. Comp. 49, 1988, pp. 427–444.
[35] Khobalatte, and Perthame B.,, Preprint, to appear in SIAM J. Numer. Anal.
[34] Kaniel S., *A kinetic model for the compressible flow equations*, Ind. Univ. Math. J. 37, 1988, pp. 537–563.
[36] Larrouturou B., *On upwind approximations of multidimensional multi-species flows*, in Proc. First European CFD Conf., pp. 117–126, Ch. Hirsh, J. Périaux, and E. Onate Ed., Elsevier, 1992.
[37] Lax P. D., *Hyperbolic systems of conservation laws II*, Comm. Pure Appl. Math. 10, 1957, pp. 537–566.
[38] Lax P. D., Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves, SIAM, Philadelphia (1973).
[39] van Leer B., *Flux-vector splitting for the Euler equations*, Lecture Notes in Physics, 170, Springer-Verlag, 1982, pp. 507-512.
[40] van Leer B., *On the relation between the upwind differencing schemes of Godunov, Engquist-Osher, and Roe*, SIAM J. Sta. Comp. 5, 1984, pp.1–20.
[41] LeFloch P.G., *Nonlinear hyperbolic systems in nonconservative form*, Comm. Part. Diff. Equ. 13, 1988, pp. 669–727.
[42] LeFloch P.G., and Liu J.G., *Entropy and monotonicity consistent (EMO) schemes for conservation laws*, Preprint CMAP-CNRS, Ecole Polytechnique, february 1994.
[43] LeFloch P.G., and Liu T.P., *Existence theory for nonconservative hyperbolic systems*, Forum Math. 5, 1993, pp. 261–280.
[44] Lerat A., *Propriété d'homogénéité et décomposition des flux en dynamique des gaz*, J. de Méc. Theor. et Appl. 2(2), 1983, pp. 185-213.
[45] Lions P.L., Perthame B., and Tadmor E., *Kinetic formulation for the multidimensional scalar equations.* Preprint, Ceremade, Univ. Paris IX-Dauphine, 1992.
[46] Lions P.L., Perthame B., and Tadmor E., *Kinetic formulation for the isentropic gas dynamics and p-system*, Preprint, Ceremade, Univ. Paris IX-Dauphine, 1993.
[47] Murat F., *Compacité par compensation*, Ann. Scuola Norm. Sup. Pisa Sci. Fis. Mat. 5, 1978, pp. 489–507.
[48] Murat F., *L'injection du cône positif de H^{-1} dans $W^{-1,q}$ est compacte pour tout $q < 2$*, J. Math. Pures Appl. 60, 1981, pp. 309-322.
[49] Osher S., *Riemann solvers, the entropy condition and difference approximations*, SIAM J. on Numer. Anal. 21, 1984, pp. 217–235.
[50] Osher S., and Solomon, F., *Upwind difference schemes for hyperbolic systems of conservation laws*, Math. of Comp. 38, 1982, pp. 339-374.

[51] Osher S., and Tadmor E., *On the convergence of difference approximations to scalar conservation laws*, Math. of Comp. 50, 1988, pp. 19–51.
[52] Perthame B., *Boltzmann-type schemes for gas dynamics and the entropy property*, SIAM J. on Numer. Anal. 27, 1990, pp. 1405–1421.
[53] Perthame B., *Second-order Boltzmann schemes for the compressible Euler equations in one or two space variables*, SIAM J. on Numer. Anal. 29, 1992, pp. 1–19.
[54] Pullin D.I., *Direct simulation method for compressible inviscid ideal-gas flow*, J. Comp. Phys 34, 1980, pp. 231–244.
[54] Reitz R.D., *One-dimensional compressible gas dynamics calculations using the Bolztmann equation*, J. Comp. Phys 42, 1981, pp. 108–123.
[56] Sanders R.H., and Prendergast K.H.,, Ap. J. 188, 1974.
[57] Shu C.W., *A numerical method for systems of conservation laws of mixed type admitting hyperbolic flux splitting*, J. of Comp. Phys. 100, 1992, pp. 424-429.
[58] Shuen J.S., Liou M.S., and van Leer B., *Inviscid flux-splitting algorithms for real gases with nonequilibrium chemistry*, J. of Comp. Phys. 90, 1990, pp. 371–395.
[59] Smoller J.S., Shock Waves and Reaction Diffusion Equations, Springer Verlag, New York, 1983.
[60] Steger J.L., and Warming R.F., *Flux vector splitting of the inviscid gas dynamics equations with application to finite-difference methods*, J. of Comp. Phys. 40, 1981, pp. 263-293.
[61] Szepessy A., *Convergence of a shock-capturing streamline diffusion finite element method for scalar conservation laws in two space dimensions*, Math. of Comp. 54, 1989, pp. 527–545.
[62] Tadmor E., *Numerical viscosity and the entropy condition for conservative difference schemes*, Math. of Comp. 43, 1984, pp. 369–381.
[63] Tadmor E., *Semi-discrete approximations to nonlinear systems of conservation laws; consistency and L^∞ imply convergence*, ICASE Report N. 88-41, 1988.
[64] Tadmor E. *The numerical viscosity of entropy stable schemes for systems of conservation laws*
[65] Tartar L., *The compensated compactness method applied to systems of conservation laws*, in Systems of Nonlinear P.D.E., J.M. Ball (ed.), NATO Series, C. Reidel publishing Col., 1983, pp. 263–285.
[66] Vinokur M., and Montagné J.-L., *Generalized flux-vector splitting and Roe average for an equilibrium real gas*, J. of Comp. Phys. 89, 1990, pp. 276–300.

Appendix: Genuine nonlinearity of the flux-splitting in Section 5

In this appendix, we give a proof of the following two lemmas whose results were used in Section 5. We use the notation introduced in the proof of Theorem 5.1.

Lemma 1. *For all $t \neq 1$, $M_0 \in (-1/\theta, 1/\theta)$, and $M = M^{\pm}(t, M_0) \in (-1/\theta, 1/\theta)$, we have*

$$t^{2\theta} - \left(\frac{t^\theta M - M_0}{t-1}\right)^2 \neq 0.$$

Lemma 2. *For all $t \neq 1$, $M_0 \in (-1/\theta, 1/\theta)$, and $M = M^{\pm}(t, M_0) \in (-1/\theta, 1/\theta)$, we have*

$$\left(\frac{t^{\theta+1}a(M) - a(M_0)}{t-1} + t^\theta(\theta M a'(M) - (\theta+1)a(M))\right)^2 - t^{2\theta}(a'(M))^2 \neq 0.$$

Proof of Lemma 1. It is equivalent to show that

$$g^{\pm}(t, M_0) \neq 0 \quad \text{for any } M, M_0 \in (-\frac{1}{\theta}, \frac{1}{\theta}), \quad t \neq 0,$$

where the function g is defined by:

$$g^{\pm}(t, M_0) = (M_0 \pm t^{\theta+1})(t^{\theta+1}a(M) - a(M_0)) - t^{2\theta+1}b(M) + b(M_0),$$

with $M = t^{-\theta}M_0 \pm (t-1)$.

Then

$$\begin{aligned} &g^{\pm}(1, M_0) = 0, \\ &g^{\pm}(0, M_0) = b(M_0) - M_0 a(M_0) > 0, \quad \text{for} \quad M, M_0 \in (-\frac{1}{\theta}, \frac{1}{\theta}). \end{aligned}$$

The t-derivative of the function g, $\partial_t g^{\pm}(t, M_0)$, is computed as follows:

$$\begin{aligned} &\partial_t g(t,M_0) \\ &= t^\theta\{\pm(\theta+1)[t^{\theta+1}a(M) - a(M_0)] + (\theta+1)(M_0 \pm t^{\theta+1})a(M) \\ &\qquad + (M_0 \pm t^{\theta+1})(-\theta t^{-\theta}M_0 \pm t)a'(M) \\ &\qquad - (2\theta+1)t^\theta b(M) - t^\theta b'(M)(-\theta t^{-\theta}M_0 \pm t)\} \\ &= t^{2\theta}\{\pm(\theta+1)[ta(M) - t^{-\theta}a(M_0)] \\ &\qquad + (\theta+1)(t^{-\theta}M_0 \pm t)a(M) \\ &\qquad (t^{-\theta}M_0 \pm t)(-\theta t^{-\theta}M_0 \pm t)a'(M) \\ &\qquad - (2\theta+1)b(M) - b'(M)(-\theta t^{-\theta}M_0 \pm t)\}. \end{aligned}$$

Since, by definition, $M \pm 1 = t^{-\theta}M_0 \pm t$, and $t^{-\theta}M_0 = M \pm (1-t)$, we obtain

$$\begin{aligned}
&t^{-2\theta}\partial_t g^{\pm}(t,M_0)\\
&= \pm(\theta+1)(ta(M) - t^{-\theta}a(M_0)) + (\theta+1)(M\pm 1)a(M)\\
&\quad + (M\pm 1)(-\theta M \pm \theta(t-1) \pm t)a'(M)\\
&\quad - (2\theta+1)b(M) - [-\theta M \pm \theta(t-1) \pm t]b'(M)\\
&= \pm(\theta+1)[ta(M) - t^{-\theta}a(M_0)] + (\theta+1)(M\pm 1)a(M)\\
&\quad + (M\pm 1)\,(-\theta(M\pm 1) \pm (\theta+1)t)\,a'(M)\\
&\quad - (2\theta+1)b(M) \pm (\theta(1\pm M) - (\theta+1)t)\,b'(M)
\end{aligned}$$

Note that

$$\begin{cases} (2\theta+1)b(M) - (\theta+1)Mb'(M) = (1-M^2)a'(M),\\ b'(M) = (1-\theta)Ma'(M) + (\theta+1)a(M),\end{cases}$$

and therefore

$$\begin{aligned}
(2\theta+1)b(M) &= (1-M^2)a'(M) + (\theta+1)M((1-\theta)Ma'(M) + (\theta+1)a(M))\\
&= (1-M^2+(1-\theta^2)M^2)a'(M) + (\theta+1)^2Ma(M)\\
&= (1-\theta^2M^2)a'(M) + (\theta+1)^2Ma(M).
\end{aligned}$$

We deduce that:

$$\begin{aligned}
&t^{-2\theta}\partial_t g^{\pm}(t,M_0)\\
&= \pm(\theta+1)(ta(M) - t^{-\theta}a(M_0)) + (\theta+1)(M\pm 1)a(M)\\
&\quad + (M\pm 1)(-\theta(M\pm 1) \pm (\theta+1)t)a'(M) - (1-\theta^2M^2)a'(M) - (\theta+1)^2Ma(M)\\
&\quad + [\theta a(M\pm 1) \mp (\theta+1)t][(1-\theta)Ma'(M) + (\theta+1)a(M)]\\
&= \pm(\theta+1)(ta(M) - t^{-\theta}a(M_0))\\
&\quad + a(M)[(\theta+1)(M\pm 1) - (\theta+1)^2M + \theta(\theta+1)(M\pm 1) \mp (\theta+1)62t]\\
&\quad + a'(M)[(M\pm 1)(-\theta(M\pm 1) \pm (\theta+1)t) - 1 + \theta^2M^2 + (1-\theta)M(\theta(M\pm 1) \mp (\theta+1)t)]\\
&= \pm(\theta+1)[ta(M) - t^{-\theta}a(M_0)] + a(M)(\theta+1)^2(M\pm 1 - M \mp t)\\
&\quad + a'(M)\{-1 + \theta^2M^2 \mp \theta(M\pm 1) + (\theta+1)t - \theta^2M(M\pm 1) \pm \theta M(\theta+1)t\}
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\frac{t^{-2\theta}}{\theta+1}\partial_t g^{\pm}(t,M_0)\\
&= \pm(ta(M) - t^{-\theta}a(M_0) - (\theta+1)(t-1)a(M)) + (t-1)(1\pm\theta M)a'(M)\\
&= \pm\{[(-\theta t+\theta+1)a(M) - t^{-\theta}a(M_0)] + (t-1)(1\pm\theta M)a'(M)\}\\
&= \pm t^{-\theta}\{-a(M_0) + t^{\theta}(-\theta t+\theta+1)A(M) + t^{\theta}a'(M)[M(-\theta t+\theta+1) \pm (t-1)(1\pm\theta M)]\}\\
&= \pm t^{-\theta}\{-a(M_0) + t^{\theta}(-\theta t+\theta+1)A(M) + t^{\theta}(M\pm(t-1))a'(M)\}\\
&\equiv \pm t^{-\theta}h^{\pm}(t,M_0),
\end{aligned}$$

since $a(M) = A(M) + Ma'(M)$. Here the function h is defined by:

$$h^{\pm}(t, M_0) = -a(M_0) + t^{\theta}(-\theta t + \theta + 1)a(M) \pm t^{\theta}(t-1)(1 \pm \theta M)a'(M),$$

with $M = t^{-\theta}M_0 \pm (t-1)$. The function h satisfies:

$$h^{\pm}(t, M_0)|_{t=1} = 0.$$

The t-derivative of the function M is:

$$tM_t = \pm(\theta+1)(t-1) - \theta M \pm 1,$$

so we find:

$$\begin{aligned}
&\partial_t h^{\pm}(t, M_0)\\
=&[\theta t^{\theta-1}(\theta+1-\theta t) - \theta t^{\theta}]a(M)\\
&+ \{t^{\theta}(\theta+1-\theta t)M_t + ((\theta+1)t^{\theta} - \theta t^{\theta-1})(\pm 1 + \theta M) + \theta t^{\theta}(t-1)M_t\}a'(M)\\
&\pm t^{\theta}(t-1)(1 \pm \theta M)M_t a''(M)\\
=&\theta(\theta+1)t^{\theta-1}(1-t)a(M)\\
&+ t^{\theta-1}\{[-\theta M \pm \theta(t-1) \pm t](\theta+1-\theta t+\theta t-\theta) + ((\theta+1)t-\theta)(\theta M \pm 1)\}a'(M)\\
&+ t^{\theta-1}(t-1)(1 \pm \theta M)(\mp\theta M + \theta(t-1) + t)a''(M)
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\frac{\partial_t h^{\pm}(t, M_0)}{t^{\theta-1}}\\
=&\theta(\theta+1)(1-t)a(M)\\
&+ [\theta(\theta+1)(t-1)M \pm 2((\theta+1)t-\theta)]a'(M)\\
&+ a''(M)(t-1)[1-\theta^2M^2 + (\theta+1)(t-1)(1 \pm \theta M)]\\
=&\theta(\theta+1)(1-t)A(M) - \frac{A'(M)}{M}(t-1)[1-\theta^2M^2 + (\theta+1)(t-1)(1 \pm \theta M)]\\
&+ a'(M)[\theta(\theta+1)(t-1)M \pm 2((\theta+1)t-\theta) + \theta(\theta+1)(1-t)M]\\
=&\pm 2((\theta+1)t-\theta)a'(M)\\
&\frac{A'(M)}{M}(t-1)\{-(1-\theta^2M^2) - (\theta+1)(t-1)(1 \pm \theta M) + 1 - \theta^2M^2\}\\
=&\pm 2((\theta+1)t-\theta)a'(M) + a''(M)(\theta+1)(t-1)^2(1 \pm \theta M)
\end{aligned}$$

by using

$$\begin{cases} a(M) = A(M) + Ma'(M), \\ A'(M) = -Ma''(M). \end{cases}$$

We shall prove now that

$$h^{+}(t, M_0) \neq 0, \quad \text{when} \quad t \neq 1,\ M, M_0 \in (-\frac{1}{\theta}, \frac{1}{\theta}),$$

$$h^{-}(t, M_0) = \begin{cases} < 0, & t > 1, \quad \text{or} \quad t < t_* \in (0, \frac{1+3\theta}{3(1+\theta)}), \\ > 0, & t \in (t_*, 1). \end{cases}$$

If this is proven, then we can conclude that

$$g^{\pm}(t, M_0) \neq 0, \quad \text{for any} \quad M, M_0 \in (-\frac{1}{\theta}, \frac{1}{\theta}), \quad t \neq 1.$$

Proof for the function: $h^+(t, M_0)$. We consider two cases depending on t.

Case 1: $(\theta+1)t - \theta \geq 0$. In this case we have:

$$\partial_t h^+(t, M_0) > 0,$$

for $t \neq 0$ and $M, M_0 \in (-\frac{1}{\theta}, \frac{1}{\theta})$. This implies that

$$h^+(t, M_0) > 0, \quad t > 1, \quad \text{for} \quad M, M_0 \in (-\frac{1}{\theta}, \frac{1}{\theta}),$$

and

$$h^+(t, M_0) < 0, \quad \text{for} \quad \frac{\theta}{\theta+1} \leq t < 1, \quad M, M_0 \in (-\frac{1}{\theta}, \frac{1}{\theta}).$$

Case 2: $(\theta+1)t - \theta \leq 0$.

We first consider the values $M \leq M_0$. We have:

$$\begin{aligned} h^+(t, M_0) =& [t^\theta(-\theta t + \theta + 1) - 1]a(M) + [a(M) - a(M_0)] \\ &+ \theta(\theta+1)t^\theta(t-1)(1+\theta M)\int_{-\frac{1}{\theta}}^{M} (1-\theta^2 y^2)^{\frac{1-\theta}{2\theta}}\, dy \\ <&0, \quad \text{for} \quad t \in (0, \frac{\theta}{\theta+1}),\ M, M_0 \in (-\frac{1}{\theta}, \frac{1}{\theta}), \end{aligned}$$

because

$$a(M) - a(M_0) \leq 0, \quad (\text{use} \quad a'(M) > 0),$$

and

$$\alpha(t) \equiv t^\theta(-\theta t + \theta + 1) - 1 < 0, \quad t \in (0, 1),$$

by using

$$\begin{cases} \alpha(1) = 0, \\ \alpha'(t) = \theta(\theta+1)t^{\theta-1}(1-t) > 0. \end{cases}$$

We next deal with the values $M > M_0$. We have

$$M_0 > \frac{t^\theta(1-t)}{1-t^\theta} \geq 0$$

because

$$M - M_0 = t^{-\theta}(1-t^\theta)(M_0 - \frac{t^\theta(1-t)}{1-t^\theta}) > 0.$$

We then observe that

$$\begin{aligned} h^+(t, M_0) =& t^\theta(-\theta t + \theta + 1)(1-\theta^2 M^2)^{\frac{1+\theta}{2\theta}} - (1-\theta^2 M_0^2)^{\frac{\theta+1}{2\theta}} \\ &+ 2\theta(\theta+1)t^\theta(t-1)\int_{-\frac{1}{\theta}}^{M} (1-\theta^2 y^2)^{\frac{1-\theta}{2\theta}}\, dy + \theta(\theta+1)M_0 \int_{M_0}^{M} (1-\theta^2 y^2)^{\frac{1-\theta}{2\theta}}\, dy \\ \equiv& p(M; M_0, t). \end{aligned}$$

But we have

$$\begin{aligned}\partial_M p(M;M_0,t) =&\theta(\theta+1)(1-\theta^2M^2)^{\frac{1-\theta}{2\theta}}\{-t^\theta(-\theta t+\theta+1)M+2t^\theta(t-1)+M_0\}\\ =&\theta(\theta+1)(1-\theta^2M^2)^{\frac{1-\theta}{2\theta}}\{-(-\theta t+\theta+1)(M_0+t^\theta(t-1))+2t^\theta(t-1)+M_0\}\\ =&-\theta(\theta+1)(1-t)(1-\theta^2M^2)^{\frac{1-\theta}{2\theta}}\{\theta M_0+\theta t^\theta(1-\theta+\theta t)\}\\ <&0, \quad \text{for} \quad t\in(0,\frac{\theta}{\theta+1}), \quad 0<M_0<M<\frac{1}{\theta}.\end{aligned}$$

Therefore, $p(M;M_0,t)$ is a monotone decreasing function of $M>M_0$ for each fixed $M_0>0$ and $t\in(0,\frac{\theta}{\theta+1})$. This implies

$$\begin{aligned}h^+(t,M_0)\le [t^\theta(-\theta t+\theta+1)-1](1-\theta^2M_0^2)^{\frac{1+\theta}{2\theta}}+2\theta(\theta+1)t^\theta(t-1)\int_{-\frac{1}{\theta}}^{M_0}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}\,dy\\ <0, \quad \text{for} \quad t\in(0,\frac{\theta}{\theta+1}), \quad 0<M_0<M<\frac{1}{\theta}.\end{aligned}$$

Proof for the function: $h^-(t,M_0)$. The proof is slightly more complicated than for the function h^+. We still distinguish between two cases.

Case 1: $t\ge 1$. We note that

$$h^-(t,M_0)=t^\theta(-\theta t+\theta+1)a(M)-a(M_0)-t^\theta(t-1)(1-\theta M)a'(M).$$

So we immediately obtain:

$$h^-(t,M_0)<0,$$

for $t\ge\frac{\theta+1}{\theta}$ and $M,M_0\in(-\frac{1}{\theta},\frac{1}{\theta})$.

It remains to treat the values t such that: $1<t<\frac{\theta+1}{\theta}$. We have

$$\begin{aligned}M-M_0=&t^{-\theta}(1-t^\theta)(M_0+\frac{t^\theta(1-t)}{1-t^\theta})\\ \le&-t^{-\theta}(t^\theta-1)(\frac{t^\theta(t-1)}{t^\theta-1}-\frac{1}{\theta})\\ <&0,\end{aligned}$$

using

$$\frac{t^\theta(t-1)}{t^\theta-1}+M_0\ge\lim_{t\to1}\frac{t^\theta(t-1)}{t^\theta-1}+M_0=\frac{1}{\theta}+M_0>0,$$

because

$$\frac{d(\frac{t^\theta(t-1)}{t^\theta-1})}{dt}=\frac{t^{\theta-1}[t^{\theta+1}-(\theta+1)t+\theta]}{(t^\theta-1)^2}>0, \quad \text{when } t>1.$$

Then

$$\begin{aligned}
h^-(t,M_0) =&[t^\theta(-\theta t+\theta+1)-1]a(M)+[a(M)-a(M_0)]\\
&-t^\theta(t-1)(1-\theta M)a'(M)\\
\le&[t^\theta(-\theta t+\theta+1)-1]a(M)-t^\theta(t-1)(1-\theta M)a'(M)\\
=&[t^\theta(-\theta t+\theta+1)-1](1-\theta^2M^2)^{\frac{\theta+1}{2\theta}}\\
&+\theta(\theta+1)\{[t^\theta(-\theta t+\theta+1)-1]M-t^\theta(t-1)(1-\theta M)\}\int_{-\frac{1}{\theta}}^{M}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}dy\\
<&-\theta(\theta+1)(t^\theta-1)[\frac{t^\theta(t-1)}{t^\theta-1}-M]\int_{-\frac{1}{\theta}}^{M}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}dy\\
<&0, \quad t\in(1,\frac{\theta+1}{\theta}), \quad M\in(-\frac{1}{\theta},\frac{1}{\theta}),
\end{aligned}$$

using

$$\frac{t^\theta(t-1)}{t^\theta-1}>\frac{1}{\theta}, \quad \text{for} \quad t>1.$$

Case 2: $0<t<1$, which is the last subcase.

For $M\le M_0$, that is,

$$M-M_0=t^{-\theta}(1-t^\theta)(M_0+\frac{t^\theta(1-t)}{1-t^\theta})\le 0,$$

which implies

$$M_0\le-\frac{t^\theta(1-t)}{1-t^\theta}<0,$$

and, when $M_0=-\frac{1}{\theta}$,

$$M\le-\frac{1}{\theta}, \quad \text{for} \quad t\in(0,1).$$

Now, for each fixed $t\in(0,1)$, we have

$$\begin{aligned}
\partial_{M_0}h^-(t,M_0) =&(-\theta t+\theta+1)a'(M)-(t-1)[(1-\theta M)a'(M)-\theta a(M)]-a'(M_0)\\
=&\theta(\theta+1)\{(1-t)(1-\theta M)(1-\theta^2M^2)^{\frac{1-\theta}{2\theta}}+\int_{M_0}^{M}(1-\theta^2y^2)^{\frac{1-\theta}{2\theta}}dy\}\\
\equiv&\theta(\theta+1)w(t,M_0),
\end{aligned}$$

using the relation $M=t^{-\theta}M_0+1-t$.

Note that

$$w(t,M)>0, \quad \text{for fixed} \quad t\in(0,1),$$

and

$$\begin{aligned}
t^\theta\partial_{M_0}w(t,M_0) =&t^\theta(1-\theta^2M_0^2)^{\frac{1-\theta}{2\theta}}+(1-\theta M)(1-\theta^2M^2)^{\frac{1-\theta}{2\theta}}[1+\theta M-\theta(1-\theta)M-\theta]\\
=&t^\theta(1-\theta^2M_0^2)^{\frac{1-\theta}{2\theta}}+(1-\theta M)(1-\theta^2M^2)^{\frac{1-\theta}{2\theta}}[\theta(1+\theta M)+1-2\theta]\\
>&0, \quad \text{for} -\frac{1}{\theta}<M<M_0\le 0.
\end{aligned}$$

This implies that

$$\partial_{M_0} h^-(t, M_0) > 0, \quad \text{for} \quad -\frac{1}{\theta} < M \le M_0 \in (-\frac{1}{\theta}, -\frac{t^\theta(1-t)}{1-t^\theta}), \quad t \in (0,1).$$

Therefore, we have

$$\begin{aligned} h^-(t, M_0) \ge & [t^\theta(-\theta t + \theta + 1) - 1](1 - \theta^2 M^2)^{\frac{\theta+1}{2\theta}} \\ & + 2\theta(\theta+1)(1-t)t^\theta \int_{-\frac{1}{\theta}}^{M} (1 - \theta^2 y^2)^{\frac{1-\theta}{2\theta}} dy \\ \equiv & \, r(t, M). \end{aligned}$$

Since

$$r(t, -\frac{1}{\theta}) = 0,$$

and

$$\begin{aligned} \partial_M r(t, M) = & \theta(\theta+1)(1-\theta^2 M^2)^{\frac{1-\theta}{2\theta}} \{2t^\theta(1-t) - M[t^\theta(-\theta t + \theta + 1) - 1]\} \\ > & 0, \end{aligned}$$

when

$$M > -\frac{2t^\theta(1-t)}{1 - t^\theta + \theta t^\theta(t-1)}.$$

for $t \in (0,1)$.

When

$$0 < t_0 \le t < 1,$$

with $3\theta(1-t_0)t_0^\theta + t_0^\theta - 1 = 0, t_0 \in (0,1)$, which is true when $t > \frac{1+3\theta}{3(\theta+1)}$, then

$$M > -\frac{1}{\theta} \ge -\frac{2t^\theta(1-t)}{1 - t^\theta + \theta t^\theta(t-1)}.$$

This is because, for $\beta(t) = 3\theta(1-t)t^\theta + t^\theta - 1$,

$$\begin{aligned} & \beta(1) = 0, \\ & \beta'(t) = \theta t^\theta(1 + 3\theta - 3(1+\theta)t) < 0, \quad \text{as} \quad t > \frac{1+3\theta}{3(1+\theta)}. \end{aligned}$$

Therefore, when $t \ge t_0 > \frac{1+3\theta}{3(\theta+1)}$,

$$h^-(t, M_0) > 0,$$

for $-\frac{1}{\theta} < M \le M_0 \le -\frac{t^\theta(1-t)}{1-t^\theta}$.

When $t \in (0, t_0)$, we have two subcases.

If $t \in (0, \frac{\theta}{\theta+1}]$, we have

$$\partial_t h(t, M_0) \ge 0,$$

for $M, M_0 \in (-\frac{1}{\theta}, \frac{1}{\theta})$.

If $t \in (\frac{\theta}{\theta+1}, t_0) \subset (\frac{\theta}{\theta+1}, t_0)$, then

$$\partial_t h^-(t, M_0) X(t, M_0) > 0,$$

where

$$X(t, M_0) = (\theta+1)\{-2a'(M) + \frac{(\theta+1)(t-1)^2(1-\theta M)(1-\theta^2 M^2)^{\frac{1-\theta}{2\theta}}}{(\theta+1)t-\theta}\}.$$

Then

$$\partial_t X(t, M_0) = a(s)x^2 + b(s)x + c(s),$$

where

$$\begin{aligned}
&s = (\theta+1)(1-t) \in ((\theta+1)(1-t_0), 1) \subset (\frac{2}{3}, 1),\\
&x = 1 + \theta M \in (0, 1 - \frac{2\theta t^\theta (1-t)}{1 - t^\theta + \theta t^\theta (t-1)}) \subset (0,1),\\
&a(s) = \theta s^2 - 2(\theta+1)s + 2(\theta+1),\\
&b(s) = s[s^3 - (3\theta+4)s^2 + (7\theta+9)s - 6(\theta+1)],\\
&c(s) = -(1-\theta)s^3(1+s),
\end{aligned}$$

One can prove that

$$\partial_t X(t, M_0) > 0,$$

for

$$\begin{aligned}
&t \in (\frac{\theta}{\theta+1}, t_0) \subset (\frac{\theta}{\theta+1}, \frac{1+3\theta}{3(1+\theta)}),\\
&M \in (-\frac{1}{\theta}, -\frac{2t^\theta(1-t)}{1-t^\theta+\theta t^\theta(t-1)}),\\
&M_0 \in (M, -\frac{t^\theta(t-1)}{t^\theta - 1}),
\end{aligned}$$

so we have

$$\partial_t h^-(t, M_0) > 0, \quad t \in (0, t_0).$$

Therefore, there exists a unique $t_* \in (0, t_0)$ such that

$$h^-(t, M_0)(t - t_*) > 0, \quad t \neq t_*, t \in (0, t_0),$$

which implies that

$$h^-(t, M_0)(t - t_*) > 0, \quad t \neq t_*, t \in (0, 1).$$

For $M > M_0$,

$$M - M_0 = t^{-\theta}(1 - t^\theta)(M_0 + \frac{t^\theta(1-t)}{1-t^\theta}) > 0,$$

which implies

$$M_0 > -\frac{t^\theta(1-t)}{1-t^\theta}.$$

First we note that

$$\begin{aligned} h^-(t, M_0) =& t^\theta(-\theta t + \theta + 1)(1-\theta^2 M^2)^{\frac{\theta+1}{2\theta}} - (1-\theta^2 M_0^2)^{\frac{\theta+1}{2\theta}} \\ &+ \theta(\theta+1)[M_0 + 2(1-t)t^\theta] \int_{-\frac{1}{\theta}}^{M} (1-\theta^2 y^2)^{\frac{1-\theta}{2\theta}} dy \\ &- \theta(\theta+1) M_0 \int_{-\frac{1}{\theta}}^{M_0} (1-\theta^2 y^2)^{\frac{1-\theta}{2\theta}} dy. \end{aligned}$$

and

$$\begin{aligned} \partial_{M_0} h^-(t, M_0) =& (-\theta t + \theta + 1) a'(M) - (t-1)[(1-\theta M)a'(M) - \theta a(M)] - a'(M_0) \\ =& \theta(\theta+1)\{(1-t)(1-\theta M)(1-\theta^2 M^2)^{\frac{1-\theta}{2\theta}} + \int_{M_0}^{M} (1-\theta^2 y^2)^{\frac{1-\theta}{2\theta}} dy\} \\ >& 0, \quad \text{for} \quad t \in (0,1), \quad -\frac{1}{\theta} < M_0 < M < \frac{1}{\theta}, \end{aligned}$$

using the relation $M = t^{-\theta} M_0 + 1 - t$.

We observe that

$$h^-(t, -\frac{t^\theta(1-t)}{1-t^\theta}) \geq 0,$$

from the first subcase.

Therefore, we obtain

$$h^-(t, M_0) > 0, \quad \text{for} \quad t \in (\frac{\theta}{\theta+1}, 1) \quad -\frac{t^\theta(1-t)}{1-t^\theta} < M_0 < M < -\frac{1}{\theta}.$$

We conclude that

$$h^-(t, M_0) \begin{cases} > 0, & t \in (0,1), \\ = 0, & t = 1, \\ < 0, & t \in (0, \infty), \end{cases}$$

when $M, M_0 \in (-\frac{1}{\theta}, \frac{1}{\theta})$.

This completes the proof of Lemma 1. □

Proof of Lemma 2. It is equivalent to show that a root $M = M(t, M_0)$ to one of the two functions $F_\pm$ defined by

$$F_\pm(M, t, M_0) \equiv -a(M_0) + t^\theta(-\theta t + \theta + 1) a(M) + t^\theta(t-1)(\pm 1 + \theta M) a'(M),$$

is not a root for the function relative to the Rankine-Hugoniot relation:

$$g(M, t, M_0) = \frac{t^{\theta+1} M - M_0}{t-1} (t^{\theta+1} a(M) - a(M_0)) - t^{2\theta+1} b(M) + b(M_0).$$

We are going to prove that for $t \neq 1$

$$\frac{\partial}{\partial t} g(M(t, M_0), t, M_0) \neq 0. \tag{A.1}$$

This is enough to conclude, since then the function g is monotone in t and only vanish at $t = 1$. We use the notation $g(t, M_0) = g(M(t, M_0), t, M_0)$. We shall need below the t-derivative of M, $M_t \equiv \partial M(t, M_0)/\partial t$:

(A.2)

$$t\, M_t\big\{a' + (t-1)(\theta M \pm 1)a''\big\} + (1-t)(1-\theta^2 M^2)a'' + \big(\theta M \pm 1 \pm (\theta+1)(t-1)\big)a' = 0.$$

The t-derivative of the function g is:

$$\begin{aligned} g_t(t, M_0) =& (t-1)^{-2}\big((\theta+1)t^\theta M(t-1) - t^{\theta+1}M + M_0\big)(t^{\theta+1}a(M) - a(M_0)) \\ &+ (t-1)^{-1}(t^{\theta+1}M - M_0)(\theta+1)t^\theta a(M) - (2\theta+1)t^{2\theta}b(M) \\ &+ (t-1)^{-1}M_t\big\{t^{\theta+1}(t^{\theta+1}a(M) - a(M_0)) \\ &+ (t^{\theta+1}M - M_0)t^{\theta+1}a'(M) - (t-1)t^{2\theta+1}b'(M)\big\}. \end{aligned}$$

Using the expressions for b and b' in terms of the functions a and a', we get after some additional calculation:

$$g_t = t^\theta (t-1)^{-1} a' \big\{tM_t \pm 1 + \theta M\big\}\big\{t^\theta M - M_0 \mp (t-1)t^\theta\big\}. \tag{A.3}$$

Next using in (A.3) the expression of M_t given by (A.2), it follows that:

(A.4)

$$g_t = t^\theta a'\big\{a' + (t-1)(\theta M \pm 1)a''\big\}^{-1}\big\{(1 \pm \theta M)a'' \pm (\theta+1)a'\big\}\big\{t^\theta M - M_0 \mp (t-1)t^\theta\big\}.$$

Let us now check that each of the three terms appearing in the right hand side of (A.4) does not vanish.

First we prove that:

$$(\theta+1)a' + (\pm 1 + \theta M)a'' > 0, \quad \text{for all } M \in (\frac{-1}{\theta}, \frac{1}{\theta}). \tag{A.5}$$

The "+" case is obvious since a' and a'' are positive. We only have to consider the function:

$$\phi(M) = (\theta+1)a' + (-1 + \theta M)a''.$$

Using the explicit formula for a'', we have

$$\phi'(M) = \big(1 - \theta^2 M^2\big)^{(1-3\theta)/(2\theta)}\big(-(2\theta+1)\theta^2 M^2 - (1-\theta)\theta M + 2\theta + 1\big).$$

So we have:

$$\phi''(M) < 0,\ \phi'(\frac{-1}{\theta}) > 0,\ \phi'(\frac{1}{\theta}) < 0,$$

and since

$$\phi(\frac{-1}{\theta}) = 0, \quad \phi(\frac{1}{\theta}) > 0,$$

the conclusion (A.5) follows.

We now prove that:

$$\text{(A.6)}\quad a' + (t-1)(\theta M \pm 1)a'' > 0, \qquad \text{for all } t \neq 1, \text{ and } M = M(t, M_0) \in (\frac{-1}{\theta}, \frac{1}{\theta}).$$

It is equivalent to prove that if M^1 is a root to

$$\text{(A.7)}\qquad a'(M^1) + (t-1)(\theta M^1 \pm 1)a''(M^1) = 0,$$

then the function $F_\pm(M^1(t, M_0), t, M_0)$ is monotone with respect to the variable t (and then vanishes only at $t = 1$). Indeed, making use of (A.7), we find (here $M = M^1$):

$$\frac{\partial}{\partial t} F_\pm(M^1(t, M_0), t, M_0) = a''\{(1-t)(1-\theta^2 M^2) - \big(\theta M \pm 1 \pm (\theta+1)(t-1)\big)(t-1)(\theta M \pm 1)\}$$

thus

$$\frac{\partial}{\partial t} F_\pm = (1-t)(1 \pm \theta M)\big(2 + (\theta+1)(t-1)\big) \neq 0,$$

since $t \geq 0$ and $0 < \theta < 1$. This completes the proof of (A.6).

We finally prove that

$$\text{(A.8)}\quad t^\theta M - M_0 \mp (t-1)t^\theta \neq 0, \qquad \text{for all } t \neq 1, \text{ and } M = M(t, M_0) \in (\frac{-1}{\theta}, \frac{1}{\theta}).$$

It is equivalent to prove that if M^2 solves the equation:

$$t^\theta M - M_0 \mp (t-1)t^\theta = 0,$$

then the function $F_\pm$ does not vanish, i.e.

$$-a(M_0) + t^\theta(-\theta t + \theta + 1)a(M^2) \pm t^\theta(t-1)(1 \pm \theta M^2)a'(M^2) \neq 0,$$

for $t \neq 1$ and $M_0, M \in (-\frac{1}{\theta}, \frac{1}{\theta})$. However, this is exactly what has been shown in the proof of Lemma 1 for $h^\pm(t, M_0)$. The proof of (A.8) is thus completed. This also completes the proof of Lemma 2.