Maps of Convex Sets and Invariant Regions
for Finite-Difference Systems of Conservation Laws

HERMANO FRID

Communicated by C. M. DAFERMOS

Abstract

General results about maps of convex sets in \( \mathbb{R}^n \) are proved. We outline their extensions to an infinite-dimensional context. Such extensions have applications in nonlinear analysis such as in the study of the invariance of convex sets under nonlinear maps. Here, we explore applications only in the finite-dimensional context. More specifically, we apply the general results to the problem of finding sufficient conditions for a region of the state space to be globally or locally invariant under finite-difference schemes applied to systems of conservation laws in several space variables. In particular, we establish a final characterization of the invariant regions under the Lax-Friedrichs scheme and also give sufficient conditions for the local invariance. Further, we give sufficient conditions for the global and local invariance of regions under flux-splitting finite-difference schemes. An example of the multi-dimensional Euler equations for non-isentropic gas dynamics is discussed.

1. Introduction

We are concerned with maps of convex sets. Among the general questions we want to address, the simplest ones may be posed in the following way. Given a closed convex set \( C \) in a Euclidean space \( E \) and a map \( f : C \rightarrow E \), can we find conditions on \( f \) and \( C \) such that (1) \((I + \varepsilon f)(C)\) is convex for \( \varepsilon > 0 \) sufficiently small, where \( I \) is the identity map; (2) \( f(C) \) itself is convex? In the first case, we should also give an upper bound for \( \varepsilon \), say \( \varepsilon_0 \), such that (1) is true for \( \varepsilon < \varepsilon_0 \). Our motivation for considering these questions is related to the problem of the existence of invariant regions under finite-difference schemes applied to systems of conservation laws in several space variables. However, they are of general interest in nonlinear analysis. For example, they are closely related to the problem of the invariance of convex sets under such maps. As will become clear from our discussion, with further natural conditions on the behavior of \( f \) on the boundary, (1) and (2) may lead to \((I + \varepsilon f)(C) \subseteq C \) and
$f(C) \subseteq C$ respectively. The latter, for instance, with $C$ bounded, allows a direct application of Brower’s fixed-point theorem, in the finite-dimensional context, or Leray-Schauder’s fixed-point theorem, in the infinite-dimensional context with $f$ compact, to show the existence of fixed points of the map $f$ in $C$. Here we will be mostly confined to the finite-dimensional case and the application to the existence of invariant regions for finite-difference schemes applied to systems of conservation laws in several space variables.

Invariant regions, when they exist, are important ingredients for ensuring stability of approximation schemes for systems of nonlinear differential equations modelling dynamical processes. They sometimes provide a priori bounds so that the approximate solutions may be constructed without “blowing up” in finite time, or by leaving out certain physical domains, in particular when the system itself is not defined in the whole state space. The situation in the discrete case is analogous to that in the continuous one. In the latter, generalizations of the maximum principle to systems of nonlinear diffusion equations, such as the one given by Chueh, Conley & Smoller [5], take the form of a characterization of (positively) invariant regions (see also [33]). For inviscid nonlinear hyperbolic systems of conservation laws in one space variable the characterization of positively invariant regions was studied by Hoff [16], Dafermos [6], Serre [32], Lions, Perthame & Tadmor [23] and Lions, Perthame & Souganidis [22], among others, the last two in connection with the study of the compactness of the solution operator. The analogous discrete case has also been discussed in [16, 17].

In [11] we studied invariant regions under Lax-Friedrichs schemes for nonlinear systems of conservation laws in several space variables. There, we obtained three sufficient conditions, labeled (C1), (C2) and (C3), which should be satisfied by the flux-vector functions of the system and the region itself in order to guarantee the invariance of the region under the referred scheme applied to the system. It was proved as well in [11] that the first two conditions were also necessary. Actually, therein, under some strong assumptions on $\partial_t \Omega_P$, we were able to prove that (C1), (C2) imply (C3) (see Lemma 1 of [11]). The first result about maps of convex sets in the present paper (see Theorem 3 below) asserts, in the case where $f$ is globally Lipschitz over the region, that the first two conditions (C1) and (C2) always imply (C3). As a consequence we find that conditions (C1) and (C2) completely characterize invariant regions under the Lax-Friedrichs scheme applied to (16). In particular, this gives a positive solution for a conjecture made in [17]. The more general statement in Theorem 3 is used in the study of local invariance of regions under the Lax-Friedrichs scheme.

Our second result about maps of convex sets (see Theorem 4 below) states, in particular, under (C1), (C2) and the additional condition that $\nabla f$ is symmetrizable over $\Omega$ and all its eigenvalues are positive, with some smoothness assumptions over $f$, that the image $f(\Omega)$ is itself convex. We use this result to obtain sufficient conditions for the invariance of regions under a wide class of so-called flux-splitting schemes. The more general statement in Theorem 4 is used in the study of local invariance of regions under these schemes.

As an application of the general results on (local) invariance of regions under finite-difference schemes, we consider the specific case of the multi-dimensional...
system of Euler equations in non-isentropic gas dynamics to obtain the local invariance of regions of the form \( S \geq r > 0 \), where \( S \) denotes the thermodynamic entropy of the gas.

Finally, we would like to mention that recently Serre [34] has found important applications of a result of the type shown in our Theorems 3 and 4 to the invariance of domains under semilinear and kinetic relaxation approximations, leading to the convergence of these approximations by means of the framework of compensated compactness developed by Tartar [37], Murat [25] and DiPerna [9]. Namely, Lemma 2.1 of [34] asserts that conditions (C1) and (C2) imply that \((I \pm \varepsilon f)(\overline{\Omega})\) is convex if \( \varepsilon < 1/\max\{\rho(\nabla f(u)) ; u \in \overline{\Omega}\} \), provided \( f \) is smooth, \( \nabla f(u) \) is diagonalizable and its eigenvalues are simple, and \( \overline{\Omega} \) is compact. Therefore, the second part of our Theorem 4, where \( \nabla f(u) \) is assumed to be diagonalizable everywhere in \( \Omega \), may be viewed as an extension of Serre's result since it does not assume that the eigenvalues of \( \nabla f(u) \) are simple and \( \overline{\Omega} \) is not assumed to be compact. Moreover, in Theorem 3, \( f \) is only assumed to be Lipschitz in \( \overline{\Omega} \) and smooth over the regular part of \( \partial \Omega \). These extensions are of concrete interest, in the applications studied here as well as in [34], because they allow the inclusion of the quadratic systems analyzed in [3], which present an umbilic point in the boundary of an invariant domain, the systems whose invariant domains are unbounded (e.g., \( p \)-systems), and systems for which \( f \) is not smooth on points of the singular part of the boundary of the invariant domain (e.g., the system of isentropic gas dynamics in Euler coordinates). On the other hand, Theorem 3 allows also the consideration of systems which change type (say, hyperbolic to elliptic) in certain regions, and which appear as prototypes for models of current use in applications such as, for instance, those studied in [13, 11].

The remainder of this paper is organized as follows. In Section 2 we prove the two results about maps of convex sets mentioned above, together with some corollaries. In Section 3 we expose the applications of Theorems 3 and 4 to the study of invariant regions for the Lax-Friedrichs and a large family of flux-splitting finite-difference schemes applied to (16). In Section 4 we study an example of the multi-dimensional system of Euler equations for non-isentropic gas dynamics in connection with our results concerning invariant regions for Lax-Friedrichs and flux-splitting finite-difference schemes.

2. Maps of convex sets

This section is devoted to the statement and proof of some results about maps of convex sets.

We recall that a subset \( \Omega \) of a real vectorspace \( X \) is called convex if, for every pair \( p, q \) of its points, it contains the entire segment \([p, q] = \{\theta p + (1 - \theta)q : 0 \leq \theta \leq 1\}\). The set \( H \) is called a supporting hyperplane of \( \Omega \subseteq X \) at the point \( p \in \Omega \) if \( p \in H \) and \( \Omega \) is entirely contained in one of the closed halfspaces bounded by \( H \). We say that \( \nu(\omega) \) is a vector in the outer normal cone of a convex set \( \Omega \) at \( \omega \in \partial \Omega \) if \( \nu(\omega) \) is orthogonal to a supporting hyperplane for \( \Omega \) at \( \omega \) and \( \omega + \nu(\omega) \) is separated from \( \Omega \) by the supporting hyperplane. We will frequently use, without explicit reference, the following basic fact about convex sets.
Theorem 1 (Minkowski [24], Brunn [2], Klee [19]). A closed set $C \subseteq \mathbb{R}^n$ is convex if and only if it possesses a supporting hyperplane at each of its boundary points.

If $C$ is a closed subset with non-empty interior in some real Hausdorff topological vectorspace $X$, $C$ is convex if and only if it possesses a supporting hyperplane at each of its boundary points.

We say that the subset $\Omega$ of the real topological vectorspace $X$ is locally convex at $p \in X$ if there exist a neighborhood $U$ of $p$ in $X$ and a convex set $C \subseteq X$ such that $\Omega \cap U = C \cap U$. For future reference we recall the following result.

Theorem 2 (Tietze [38], Klee [19]). Let $C$ be a closed connected subset of some real topological vectorspace. The subset $C$ is convex if and only if $C$ is locally convex at every point $p \in C$.

For many other facts about convex sets we refer to [1, 39, 30, 14] and the references therein.

Let $G_j : \mathbb{R}^n \to \mathbb{R}, j = 1, \ldots, N$, be smooth functions for which 0 is a regular value, and let

$$
\Omega = \bigcap_{j=1}^N \{ u : G_j(u) < 0 \},
$$

$$
S_j = \{ u : G_j(u) = 0 \} \cap \bigcap_{k \neq j} \{ u : G_k < 0 \}, \quad j = 1, \ldots, N,
$$

$$
\partial \Omega_s = \bigcup_{j=1}^N S_j.
$$

In what follows we agree to say that $f$ is smooth over a set $A \subseteq \mathbb{R}^n$ if $f$ is smooth on a neighborhood of each point of $A$. We now state and prove our first result about maps of convex sets.

Theorem 3. Let $\Omega$ and $\partial \Omega_s$ be as in (1)–(3), and $U \subseteq \mathbb{R}^n$ be an open connected set containing $\Omega \cup \partial \Omega_s$. Assume that $f : U \to \mathbb{R}^n$ is locally Lipschitz continuous and smooth over $\partial \Omega_s$. Suppose

\[\Omega\] is convex; \hspace{1cm} (C1)

\[Df_\omega(T_\omega(\partial \Omega_s)) \subseteq T_\omega(\partial \Omega_s), \text{ for all } \omega \in \partial \Omega_s,\] \hspace{1cm} (C2)

where $T_\omega(\partial \Omega_s)$ denotes the tangent space to $\partial \Omega_s$ at $\omega$.

Let $z_0 \in \Omega \cap U$ and $B \subseteq U$ be an open ball around $z_0$. Then, for any $M > M_0 = 2\Lip(f)B$, the sets $(M1 \pm f)(\Omega \cap B)$ are locally convex at all points of $(M1 \pm f)(\partial \Omega_s \cap B)$, respectively. In this case, given $\omega_0 \in \partial \Omega_s \cap B$, there exists a ball $B(\omega_0)$ such that, for any $u \in \Omega \cap B(\omega_0)$, $\omega \in \partial \Omega_s \cap B(\omega_0)$ the following
inequality holds:
\[
|\langle f(u) - f(\omega), v(\omega) \rangle| \leq M \langle u - v(\omega), v(\omega) \rangle,
\]
where \(v(\omega)\) is the unit outer normal of \(\partial \Omega_\omega\) at \(\omega\).

Moreover, if \(f\) is defined and (globally) Lipschitz over \(\overline{\Omega}\) then, for \(M \geq 2\text{Lip}(f)\), the sets \((M I \pm f)(\overline{\Omega} \cap B)\) are convex and inequality (4) holds for all \(u \in \overline{\Omega}\), \(\omega \in \partial \Omega\) and \(v(\omega)\) in the outer normal cone of \(\overline{\Omega}\) at \(\omega\).

**Proof.** Let \(M > M_0 = 2\text{Lip}(f)\) be given. We will show that \((M I \pm f)(\overline{\Omega} \cap B)\)
are locally convex at all points of \((M I \pm f)(\partial \Omega_\omega \cap B)\), respectively. We prove
the assertion for \(M I + f\); the one for \(M I - f\) follows replacing \(f\) by \(-f\). Suppose, on
the contrary, that \(V_M = (M I + f)(\overline{\Omega} \cap B)\) is not locally convex everywhere in
\((\partial V_M)_\omega \equiv (M I + f)(\partial \Omega_\omega \cap B)\). By definition, there must exist \(v_0 \in (\partial V_M)_\omega\), such
that \(B(v_0) \cap V_M\) is not convex for any open ball \(B(v_0)\) around \(v_0\), no matter how small. If we take \(B(v_0)\) small enough, \(B(v_0) \cap (\partial V_M)_\omega\) is the graph of a smooth
function, defined on some open set of \(\mathbb{R}^{n-1}\), whose Hessian is not non-
positive at some point corresponding to a certain \(v_1 \in B(v_0) \cap (\partial V_M)_\omega\), with \(B(v_0) \cap V_M\)
lying below the graph. For simplicity we take \(v_1 = v_0\).

Let \(u_0\) be the point in \(\partial \Omega_\omega\) satisfying \(v_0 = (M I + f)(u_0)\). Set
\[
g(u) = u + \frac{1}{M} (f(u) - f(u_0)).
\]
Then \(g(u_0) = u_0\), and for any open ball \(B(u_0)\) around \(u_0\), no matter how small,
\(B(u_0) \cap V\) is not convex when \(V = g(\overline{\Omega} \cap B)\), since \(V = T_M(V_M)\), where \(T_M\)
is the affine isomorphism given by \(T_M = 1/M(I - f(u_0))\). Clearly, \(u_0 \in \partial V =\)
g(\(\partial \Omega_\omega \cap B\)) and so \(\partial V \cap B(u_0)\) is a smooth submanifold for any sufficiently small
open ball \(B(u_0)\) around \(u_0\). We notice that \(g\) satisfies \(Dg_{\omega}(T_\omega(\partial \Omega_\omega)) \subseteq T_\omega(\partial \Omega_\omega)\)
for all \(\omega \in \partial \Omega_\omega\), as an immediate consequence of (C2). This means that, if \(v(\omega)\)
is the unit outer normal of \(\partial \Omega_\omega\) at \(\omega \in \partial \Omega_\omega \cap B\), then \(v(\omega)\) is also the unit outer
normal of \(\partial V\) at \(g(\omega)\). In fact, in this case \(v(\omega)\) must be an eigenvector of \(Dg_{\omega}\)
corresponding to a positive eigenvalue, by the choice of \(M\), and this in turn implies that \(v(\omega)\)
must be also outward directed with respect to \(V\). We can change the coordinates if necessary by setting \(u = (u_1, \ldots, u_n)\), \(u_0 = 0\), making the \(u_n\)-axis
have the same direction and orientation as \(v(u_0)\) and assuming that \(\nabla V = u_n\)-axis is the graph of a smooth function \(u_n = G(u_1, \ldots, u_{n-1})\) if \(B(u_0)\) is sufficiently
small. The coordinates \(u_1, \ldots, u_{n-1}\) can be taken such that \(\nabla^2 G(0) = 0\) and the \(u_1\)-axis is an eigendirection of \(\nabla^2 G(0), i = 1, \ldots, n - 1\), with the \(u_1\)-axis associated
with a positive eigenvalue. Let \(\Pi = \{u : u_2 = \cdots = u_{n-1} = 0\}\).

We may suppose \(\Pi \cap \partial V \cap B(u_0)\) to be parametrized by \(\alpha : [-\delta_0, \delta_0] \to \partial V,\)
with \(\alpha(0) = u_0\) and we set \(p = \alpha(-\delta), q = \alpha(\delta)\) for some \(0 < \delta < \delta_0\). If \(v(p)\)
and \(v(q)\) are the unit outer normal vectors of \(\partial V\) at \(p\) and \(q\) respectively, we must have (see Fig. 3)
\[
\langle v(p), q - p \rangle > 0, \quad \langle v(q), p - q \rangle > 0.
\]
On the other hand,
\[
\|\mathbf{u} - g(\mathbf{u})\| \leq \frac{M_0}{2M} \|\mathbf{u} - \mathbf{u}_0\| \leq \frac{M_0}{2M - M_0} \|g(\mathbf{u}) - g(\mathbf{u}_0)\|.
\]
from which it follows that
\[
\|g^{-1}(v) - v\| \leq \frac{M_0}{2M - M_0} \|v - \mathbf{u}_0\|.
\] (6)
where \(g^{-1}\) is the inverse of \(g\) restricted to \(B\). Now, since \(M_0/(2M - M_0) < 1\), (6) implies that, if \(\delta\) is sufficiently small, \(g^{-1}(p)\) and \(g^{-1}(q)\) both lie in the interior of opposite convex cones with vertex \(\mathbf{u}_0\) and axis parallel to \(\alpha'(0)\), in such a way that \(p\) and \(g^{-1}(p)\) stay in the interior of one sheet, while \(q\) and \(g^{-1}(q)\) are in the interior of the opposite sheet (see Fig. 1).

We assume for the moment that \(\nu(p)\) and \(\nu(q)\) are also parallel to \(\Pi\). This is not true in general, of course, but it is the case if the function \(G(u_1, \ldots, u_{n-1})\) above, with the coordinates \(u_1, \ldots, u_{n-1}\) chosen in that way, is quadratic. In this case, we have
\[
u(u_1) = \frac{1}{\sqrt{1 + 4\lambda_1^2u_1^2}}(-2\lambda_1u_1, 0, \ldots, 0, 1).
\]
We then have the diagram described in Fig. 2. The lines 1 and 3 are the intersections with \(\Pi\) of the hyperplanes orthogonal to \(p - q\), containing \(p\) and \(q\), respectively. The lines 2 and 4 are the intersections with \(\Pi\) of the hyperplanes orthogonal to \(g^{-1}(p) - g^{-1}(q)\), containing \(p\) and \(q\), respectively. Now, the convexity of \(\Omega\) implies that
\[
\langle \nu(p), g^{-1}(q) - g^{-1}(p) \rangle \leq 0, \quad \langle \nu(q), g^{-1}(p) - g^{-1}(q) \rangle \leq 0,
\]
where we used the fact that \(\nu(p)\) is also an outer unit normal vector of \(\partial \Omega\) at \(g^{-1}(p)\) and similarly for \(\nu(q)\) and \(g^{-1}(q)\). This means that \(\nu(p)\) and \(\nu(q)\) should
not point toward the interior of the strip bounded by the lines 2 and 4. But this is impossible because of (5); we have then arrived at a contradiction.

We now examine the general case, dropping the assumption that \(v(p)\) and \(v(q)\) are parallel to \(\Pi\). We then have that \(\partial V_s \cap B(u_0)\) may be expressed in the form

\[ u_n = \lambda_1 u_1^2 + \cdots + \lambda_{n-1} u_{n-1}^2 + O((u_1, \ldots, u_{n-1})^3), \]

again with \(\lambda_1 > 0\). We may further assume that the curve \(\alpha\) is parametrized by \(u_1\), so that, for \(v(u_1) = v(\alpha(u_1))\), we have

\[ v(u_1) = \frac{1}{\sqrt{1 + 4\lambda_1^2 u_1^2}} (-2\lambda_1 u_1, 0, \ldots, 0, 1) + O(|u_1|^2). \]

So the change in \(v(u_1)\) relatively to the quadratic case is \(O(|u_1|^2)\). On the other hand, for the cosine of the angle formed by \(v(u_1)\) and \(v(0)\) we have

\[ \cos(v(u_1)) = \frac{1}{\sqrt{1 + 4\lambda_1^2 u_1^2}} + O(|u_1|^2), \]

and so

\[ 1 - \cos(v(u_1)) = O(|u_1|^2). \]
Thus, for \( M > \{u_1\} \), we see that the change in the angle between the projection of \( v(u_1) \) in \( \Pi \) and \( v(0) \) is \( O(|u_1|) \). If \( \tilde{v}(q) \) is the projection of \( v(q) \) in \( \Pi \), we then have that the distance of \( q + \tilde{v}(q) \) to the hyperplane through \( q \) orthogonal to \( q - p \) is \( O(|\sin(v(u_1))|) = O(|u_1|) \). We also remark that the unit vector in the direction \( g^{-1}(q) - g^{-1}(p) \) belongs to the cone described above. Let us suppose, with no loss of generality, that the intersection with \( \Pi \) of the hyperplane through \( q \) orthogonal to \( g^{-1}(q) - g^{-1}(p) \) is as line 4 in Fig. 2. In particular, the distance of \( q + \tilde{v}(q) \) to this hyperplane can be no smaller than \( O(|u_1|) \), for \( |u_1| \) small. So, since \( q + \tilde{v}(q) \) is at a distance of order \( O(|u_1|^2) \) from \( q + \tilde{v}(q) \) they both should belong to the same halfspaces determined by the two hyperplanes through \( q \) orthogonal to \( q - p \) and \( g^{-1}(q) - g^{-1}(p) \) respectively, for \( |u_1| \) sufficiently small. Hence, for \( \delta \) sufficiently small we must have the same pattern as that for the case where \( v(p) \) and \( v(q) \) are parallel to \( \Pi \) and so we also get a contradiction in the general case. The assertion about the local convexity of \( (M_1 \pm f)(\Omega \cap B) \) at all points of \( (M_1 \pm f)(\partial \Omega_0 \cap B) \) is then proved.

Concerning the inequality (4), let \( \omega_0 \in \partial \Omega_0 \cap B \). We take \( B(\omega_0) \) such that \( (M_1 \pm f)(B(\omega_0) \cap \Omega) \) is contained in a convex set \( C \subseteq (M_1 \pm f)(\partial \Omega \cap B) \); we have proved that this is possible. We observe that if \( v(\omega) \) is the outer normal of \( \partial \Omega_0 \) at \( \omega \in \partial \Omega_0 \), then \( v(\omega) \) is the outer normal of \( (M_1 \pm f)(\partial \Omega_0 \cap B) \) at \( (M_1 \pm f)(\omega) \). But the convexity of \( C \) implies that, for all \( u \in B(\omega_0) \cap \Omega \) and \( \omega \in \partial \Omega_0 \cap B(\omega_0) \),

\[
\langle (M_1 \pm f)(u) - (M_1 \pm f)(\omega), v(\omega) \rangle \leq 0,
\]

which gives

\[
-M\langle \omega - u, v(\omega) \rangle \leq \langle f(u) - f(\omega), v(\omega) \rangle \leq M\langle \omega - u, v(\omega) \rangle.
\]

Now the convexity of \( \Omega \) implies that \( \langle \omega - u, v(\omega) \rangle \geq 0 \) and so we get

\[
|\langle f(u) - f(\omega), v(\omega) \rangle| \leq M\langle \omega - u, v(\omega) \rangle,
\]

which is the desired inequality.

Let us now discuss the case in which \( f \) is Lipschitz in \( \Omega \). It suffices to prove the assertion for any \( M > M_0 \). Indeed, if the latter is true then for any fixed closed ball \( B \) around the origin the compact convex sets \( B \cap (M_1 \pm f)(\Omega) \) converge to \( B \cap (M_1 \pm f)(\Omega) \) as \( M \downarrow M_0 \) in the sense of the Hausdorff distance (see [10] p.183). But the class of compact convex sets is closed with respect to the Hausdorff distance (see [10] p.184). It follows that \( B \cap (M_1 \pm f)(\Omega) \) is a closed convex set if \( B \cap (M_1 \pm f)(\Omega) \) is closed and convex for any \( M > M_0 \).

We also recall that by Kirszbraun’s theorem (see FEDERER [10]) we may extend \( f \) as a Lipschitz function defined in all \( \mathbb{R}^n \) having the same Lipschitz constant \( M_0/2 \). Thus, for \( M > M_0 \), \( M_1 + f \) is a bi-Lipschitz homeomorphism from \( \mathbb{R}^n \) and, for \( V_M = (M_1 + f)(\Omega) \), we have

\[
V_M = \bigcap_{j=1}^{N} \{ u : \tilde{G}_j(u) \leq 0 \},
\]

(7)
with \( \tilde{G}_j = G_j \circ (M_1 + f)^{-1} \), \( j = 1, \ldots, N \). Now, from the part of the theorem already proved, \( V_M \) is locally convex at all points of \( (\partial V_M)_s \equiv (M_1 + f)(\partial \Omega_s) \). But, by (7), \( V_M \) cannot fail to be locally convex only at points of \( \partial V_M - (\partial V_M)_s \).

Indeed, let \( \omega' \in \partial V_M - (\partial V_M)_s \) be such a point and \( B_{\omega'} \) be a closed ball around it such that \( B_{\omega'} \cap V_M \) is not convex. Let us choose \( B_{\omega'} \) so small that \( B_{\omega'} \cap V_M \) is connected. This is possible since \( V_M \) is homeomorphic with \( \overline{\Omega} \). Clearly, any plane \( \Pi \) passing through \( \omega' \), possessing non-empty intersection with the interior of \( B_{\omega'} \cap V_M \), can be approximated by planes of this kind whose intersection with \( B_{\omega'} \cap \partial V_M \) is entirely in \( \omega' \cup [B_{\omega'} \cap (\partial V_M)_s] \), in terms of the Hausdorff distance of their intersections with \( B_{\omega'} \), for instance. But, if \( \Pi \) satisfies this generic intersection property, then \( \Pi \cap B_{\omega'} \cap \partial V_M \) can be identified with a closed subset \( A \subseteq \mathbb{R}^2 \) which is the intersection of regions of the type \( \{u \in \mathbb{R}^2 : \phi(u) \leq 0\} \), where \( \phi \) is a Lipschitz function, \( \phi^{-1}(0) \) is a Lipschitz curve and \( \phi^{-1}(0) \cap \partial A \) is smooth. However, such plane regions cannot fail to be locally convex only at corner points, as may be easily seen. Now, if \( p \) and \( q \) are any two points in \( B_{\omega'} \cap V_M \) such that the line segment \([p, q]\) connecting them fails to be entirely contained in \( B_{\omega'} \cap V_M \), then \( \omega', p, q \) belong to a plane that can be arbitrarily approximated by planes with the generic intersection property described above. So, \( p, q \) are the limits of a sequence of points \( p_a, q_a \) such that the planes containing \( \omega', p_a, q_a \) possess the property referred to. Hence, any point in \([p, q]\) is also a limit of points in the segments \([p_a, q_a]\) which in turn belong to \( B_{\omega'} \cap V_M \), by the convexity of the intersection of those planes with this set. Since \( B_{\omega'} \cap V_M \) is closed it must contain the entire segment \([p, q]\), giving rise to a contradiction. Hence, \( V_M \) must be locally convex at all its points, and so, by Theorem 2 it must be convex.

Finally, in this case, inequality (4) holds for all \( u \in \overline{\Omega}, \omega \in \partial \Omega \) and \( v(\omega) \) in the outer normal cone of \( \overline{\Omega} \) at \( \omega \) because of the convexity of \( (M_1 \pm f)(\overline{\Omega}) \) and the fact that the outer normals of \( \partial \Omega_s \) at \( \omega \in \partial \Omega_s \) are also outer normals of \( (M_1 \pm f)(\partial \Omega_s) \) at \( (M_1 \pm f)(\overline{\Omega}) \). On the other hand, any \( v(\omega) \) in the outer normal cone at \( \omega \in \partial \Omega \) is a positive linear combination of limits of outer normals \( v(\omega_a) \), at \( \omega_a \in \partial \Omega_s \), with \( \omega_a \to \omega \), for which the inequality holds at the corresponding points with \( u \) fixed.

As a first corollary of the above theorem we have the following result, which is of some interest in itself.

**Corollary 1.** Let \( \Omega \) and \( \partial \Omega \) be as in (1)–(3). Assume that \( f : \mathbb{R}^n \to \mathbb{R}^n \) is Lipschitz continuous and \( f|\partial \Omega \) is smooth. Let \( M_0 = 2 \text{Lip}(f) \) and \( g(u) = u + M^{-1}f(u) \), for some \( M \geq M_0 \). Suppose that

\[
(f(\omega), v(\omega)) \leq 0 \quad (8)
\]

for all \( \omega \in \partial \Omega \) and \( v(\omega) \) in the outer normal cone of \( \Omega \) at \( \omega \). Then (C1) and (C2) imply that \( g(\overline{\Omega}) \subseteq \overline{\Omega} \). In particular, when equality holds in (8) the equality \( g(\overline{\Omega}) = \overline{\Omega} \) holds for \( M \geq M_0 \).

**Proof.** The proof follows from the fact that if \( u \in \overline{\Omega}, \omega \in \partial \Omega \) and \( v(\omega) \) is in the outer normal cone of \( \Omega \) at \( \omega \), then we have

\[
(g(u) - \omega, v(\omega)) = (g(u) - g(\omega), v(\omega)) + M^{-1}(f(\omega), v(\omega)) \leq 0,
\]
which in turn implies that \( g(u) \in \Omega \) for any \( u \in \Omega \). Finally, in the case where the equality holds in (8), using the first part for both \( f \) and \(-f\) we conclude that, for any \( \omega \in \partial \Omega \), both \( \omega + M^{-1}f(\omega) \) and \( \omega - M^{-1}f(\omega) \) belong to \( \Omega \). But, since \( \omega \in \partial \Omega \) is in the line segment joining these two points, convexity of \( \Omega \) implies that they both should also belong to \( \partial \Omega \). Hence, for \( M \geq M_0 \), \( g \) is obviously injective, \( g(\Omega) \subseteq \Omega \) and \( g(\partial \Omega) \subseteq \partial \Omega \). It is also easy to see that \( |g(u)| \to \infty \) as \( |u| \to \infty \), and so \( g \) extends to an injective map of \( \mathbb{S}^n \), the \( n \)-dimensional unity sphere, into itself, via stereographic projection. We conclude, using Brower’s degree theory, that \( g(\Omega) = \Omega \).

**Remark 1.** The above corollary can be directly applied, for instance, to prove invariance of \( \Omega \) under Euler and certain Runge-Kutta type schemes applied to the system of ordinary differential equations \( u = f(u) \), for \( \Omega \) and \( f \) satisfying its hypotheses. Indeed, we recall that the Euler scheme is given by

\[
u_{k+1} = \nu_k + hf(\nu_k),
\]

where \( h = \Delta t \), while the Runge-Kutta type schemes we consider are given by

\[
u_{k+1} = \frac{1}{6} (\nu_k + hf(\nu_k)) + \frac{1}{6} (z_{k1} + a_1hf(z_{k1})) + \frac{1}{6} (z_{k2} + a_2hf(z_{k2})) + \frac{1}{6} (z_{k3} + a_3hf(z_{k3})),
\]

where

\[
z_{k1} = \nu_k + (1 - a_1)hf(\nu_k),
\]

\[
z_{k2} = z_{k1} + (a_1 - a_2)hf(z_{k1}),
\]

\[
z_{k3} = z_{k2} + (a_2 - a_3)hf(z_{k2}),
\]

with \( 1 > a_1 > a_2 > a_3 > 0 \). So, the invariance of \( \Omega \) under the Euler scheme follows immediately from the corollary if we choose \( h \leq (2M_0)^{-1} \). The invariance under the Runge-Kutta type schemes above follows by observing that, by the corollary, the expressions inside the parentheses belong to \( \Omega \), for \( h \leq (2M_0)^{-1} \). So, \( \nu_{k+1} \) is a convex combination of points in \( \Omega \), hence, it is a point in \( \Omega \).

We observe that in the above results the only restrictions imposed on \( \nabla f \) were its local or uniform boundedness, the smoothness of \( f \) over \( \partial \Omega \), and \((C2)\). We are going to impose some additional restrictions on \( \nabla f \) in order to get information about the image of \( f \) itself.

We first recall that a \( n \times n \) matrix \( A \) is said to be symmetrizable if there exists a symmetric positive definite matrix \( P \) such that \( PA \) is symmetric. It is easy to see that a matrix \( A \) is symmetrizable if and only if it is diagonalizable, that is, there exists a non-singular matrix \( S \) such that \( SAS^{-1} \) is a diagonal matrix with only real entries.

We now consider the case in which, over some suitable set \( S \subseteq \mathbb{R}^n \), \( f \) satisfies the following two conditions:

- \( \nabla f(u) \) is continuous and symmetrizable over \( S \) by a continuous symmetric matrix \( P(u) > 0 \); \( \nabla f(u) \) is continuous and symmetrizable over \( S \) by a continuous symmetric matrix \( P(u) > 0 \); \( \nabla f(u) \) is continuous and symmetrizable over \( S \) by a continuous symmetric matrix \( P(u) > 0 \); \( \nabla f(u) \) is continuous and symmetrizable over \( S \) by a continuous symmetric matrix \( P(u) > 0 \); \( \nabla f(u) \) is continuous and symmetrizable over \( S \) by a continuous symmetric matrix \( P(u) > 0 \);
- all eigenvalues of \( \nabla f(u) \) are positive for all \( u \in S \).
Theorem 4. Let $\Omega$ and $\partial \Omega$ be as in (1)–(3) and $U \subseteq \mathbb{R}^n$ be an open connected set containing $\Omega \cup \partial \Omega$. Assume that $f : U \rightarrow \mathbb{R}^n$ is locally Lipschitz continuous, smooth over $\partial \Omega$, and $\nabla f(u)$ satisfies (9) and (10) with $S = \partial \Omega$. Suppose that $\Omega$ and $f$ satisfy conditions (C1) and (C2) of Theorem 3. Then, given $\omega_0 \in \partial \Omega \cap U$, there exists an open ball $B(\omega_0)$ around $\omega_0$ such that $f(\partial \Omega \cap B(\omega_0))$ is locally convex at all points of $f(\partial \Omega \cap B(\omega_0))$. Moreover, for any $u \in \Omega \cap B(\omega_0)$ and $\omega \in \partial \Omega \cap B(\omega_0)$,

$$ (f(u) - f(\omega), v(\omega)) \leq 0, $$

where $v(\omega)$ is the unit outer normal of $\partial \Omega$ at $\omega$.

Further, if $\nabla f(u)$ satisfies (9), (10) with $S = U$ and $\Omega \subseteq U$, then $f(\Omega)$ is convex and the inequality (11) holds for all $u \in \Omega$, $\omega \in \partial \Omega$ and $v(\omega)$ in the outer normal cone of $\Omega$ at $\omega$.

Proof. We begin by proving the first part of the statement. Let $\omega_0$ be any fixed point in $\partial \Omega$, and $r_0 > 0$ such that $f$ is a bi-Lipschitz homeomorphism from $B_0$ onto $f(B_0)$ and $B_0 \cap \partial \Omega = B_0 \cap \partial \Omega$, where $B_0 = B(\omega_0, r_0) \subseteq U$. This is clearly possible by (10) since $f$ is smooth at $\omega_0$. We will prove that $f(B_0 \cap \Omega)$ is locally convex at all points of $f(B_0 \cap \partial \Omega)$. So, assume, for contradiction, that $f(B_0 \cap \Omega)$ fails to be locally convex at a certain $v_0 \in \partial \Omega \cap f(B_0 \cap \partial \Omega)$ and let $u_0 \in B_0 \cap \partial \Omega$, be given by $f(u_0) = v_0$. Let $\varepsilon > 0$ be smaller than the minimum of the eigenvalues of $\nabla f(\omega)$, for $\omega \in B_0 \cap \partial \Omega$. Define

$$ g(u) = u_0 + \frac{1}{\varepsilon}(f(u) - f(u_0)). $$

For this part of the proof, the only point which requires a different reasoning with respect to the corresponding part of the proof of Theorem 3 is the existence of a cone. Let $\alpha : [-\delta, \delta] \rightarrow \partial V_r$, with $\alpha(0) = u_0$ and $p = \alpha(-\delta)$, $q = \alpha(\delta)$, for some $0 < \delta < \delta_0$, as in the proof of Theorem 3. Given $\xi, \eta \in \mathbb{R}^n$, we define

$$ (\xi, \eta)_u := (P(u)\xi, \eta), \quad ||\xi||_u := (P(u)\xi, \xi)^{1/2}. $$

We have

$$ (\nabla g^{-1}(u_0)\alpha'(0), \alpha'(0))_{u_0} > \varepsilon M_0^{-1}|\alpha'(0)|_{u_0}, $$

with $M_0 > \varepsilon$ equal to the maximum of the eigenvalues of $\nabla f(u_0)$. A similar inequality holds also, obviously, for $-\alpha'(0)$. Now, we clearly have

$$ (\nabla g^{-1}(u_0)\alpha'(0), \alpha'(0))_{u_0} \leq |\alpha'(0)|_{u_0}, $$

which, from (12), gives

$$ (\nabla g^{-1}(u_0)\alpha'(0), \alpha'(0))_{u_0} > \varepsilon M_0^{-1}|\nabla g^{-1}(u_0)\alpha'(0)|_{u_0}||\alpha'(0)||_{u_0}. $$

Now, inequality (13) says that $\nabla g^{-1}(u_0)\alpha'(0)$ lies in a convex cone around $\alpha'(0)$ like the one in Fig. 1. It follows that for $\delta > 0$ sufficiently small, $g^{-1}(q)$ and $q$ are in the interior of the same sheet of this cone, and $g^{-1}(p)$ and $p$ are in the interior of the opposite one. The remainder of the proof of this part of the statement follows exactly
as in the proof of Theorem 3. So, for the inequality (11) we assume that $B(\omega_0)$ is so small that $f(B(\omega_0) \cap \overline{\Omega})$ is contained in a convex subset $C \subseteq f(B) \cap \overline{\Omega}$, which is possible from the local convexity of the latter at $f(\omega_0)$. Again, for $\omega \in \partial \Omega_p \cap B(\omega_0)$, if $v(\omega)$ is the outer unit normal of $\partial \Omega_p$ at $\omega$, then $v(\omega)$ is also the outer unit normal of $f(\partial \Omega_p \cap B(\omega_0))$ at $f(\omega)$. Hence, inequality (11) follows from the convexity of $C$.

Let us now prove the last part of the statement. We assume that (9), (10) hold with $S = \mathcal{U}$ and $\overline{\Omega} \subseteq \mathcal{U}$. It suffices to prove that $f_\epsilon(\overline{\Omega})$ is convex for any $\epsilon > 0$, where $f_\epsilon = sI + f$. We claim that $f_\epsilon$ is a locally bi-Lipschitz homeomorphism from $\mathcal{U}$ onto $f_\epsilon(\mathcal{U})$. We first prove that $f_\epsilon$ is a one to one map of $\mathcal{U}$ onto $f_\epsilon(\mathcal{U})$. Clearly, $f_\epsilon$ is a local homeomorphism. In particular, given any two points $z, w \in f_\epsilon(\mathcal{U})$, there is a smooth path contained in $f_\epsilon(\mathcal{U})$ joining $z$ to $w$, which is the image of a smooth path $\alpha : [0, 1] \to \partial \Omega$, with $f_\epsilon(\alpha(0)) = z, f_\epsilon(\alpha(1)) = w$. Therefore, we may define the function $d : f_\epsilon(\mathcal{U}) \times f_\epsilon(\mathcal{U}) \to \mathbb{R}_+$ by

$$d(z, w) = \inf \int_0^1 \left| \frac{d f_\epsilon(\alpha(t))}{dt} \right|_{\alpha(t)} dt,$$

where the infimum is to be taken among all smooth paths $\alpha : [0, 1] \to \mathcal{U}$, with $f_\epsilon(\alpha(0)) = z, f_\epsilon(\alpha(1)) = w$. It is not difficult to verify that $d$ defines a metric in $f_\epsilon(\mathcal{U})$. We also notice that $\nabla f_\epsilon(u)$ is symmetric and positive definite with respect to the inner product $(\cdot, \cdot)_{\alpha}$ for all $u \in \mathcal{U}$. Hence, given $u, v \in \mathcal{U}$ and a smooth path $\alpha : [0, 1] \to \mathcal{U}$, with $\alpha(0) = u, \alpha(1) = v$, we have

$$\int_0^1 \left| \frac{d f_\epsilon(\alpha(t))}{dt} \right|_{\alpha(t)} dt \geq \epsilon \int_0^1 |\alpha'(t)|_{\alpha(t)} dt \geq \epsilon d_P(u, v),$$

where $d_P$ is the Riemannian metric induce by $P$ in $\mathcal{U}$. Hence,

$$d(f_\epsilon(u), f_\epsilon(v)) \geq \epsilon d_P(u, v), \quad (14)$$

and so $f_\epsilon$ is a one to one map of $\mathcal{U}$ onto $f_\epsilon(\mathcal{U})$. In particular, $f_\epsilon(\partial \Omega) = \partial f_\epsilon(\Omega)$.

Since $f_\epsilon$ is clearly an open map, we conclude that $f_\epsilon$ is a homeomorphism between $\mathcal{U}$ and $f_\epsilon(\mathcal{U})$ as asserted.

Now, from the part already proved, we get that $f_\epsilon(\overline{\Omega})$ is locally convex at all points of $f_\epsilon(\partial \Omega_p)$. It follows as in the proof of Theorem 3 that $f_\epsilon(\overline{\Omega})$ is convex. Finally, in this case, inequality (11) is obtained in the limit from the corresponding ones for $f_\epsilon$, as $\epsilon \downarrow 0$. The latter hold for all $u \in \overline{\Omega}, \omega \in \partial \Omega$ and $v(\omega)$ in the outer normal cone of $f_\epsilon(\overline{\Omega})$ at $\omega$ because of the convexity of $f_\epsilon(\overline{\Omega})$ and the fact that the outer normals of $\partial \Omega_p$ at $\omega \in \partial \Omega_p$ are also outer normals of $f(\partial \Omega_p)$ at $f(\omega)$.

**Remark 2.** We observe that, when $f$ is Lipschitz continuous, $\nabla f$ is symmetrizable in $\mathcal{U} \supseteq \overline{\Omega}$, and (C1), (C2) of Theorem 3 hold, the above theorem implies that $(M \pm f)(\overline{\Omega})$ is convex if $M > \rho(\nabla f)$, where $\rho(\nabla f)$ denotes the supremum in $\overline{\Omega}$ of the spectral radius of $\nabla f$, that is,

$$\rho(\nabla f) = \sup_{u \in \overline{\Omega}} \max_{1 \leq j \leq n} |\lambda_j(u)|.$$

This statement provides an extension of Lemma 2.1 of [34].
We now have a corollary of Theorem 4 which is the analogue of Corollary 1 and is of interest in fixed point theory.

**Corollary 2.** Let \( \Omega \) and \( \partial \Omega \) be as in (1)–(3) and \( \mathcal{U} \subseteq \mathbb{R}^n \) be an open connected set. Assume that \( f : \mathcal{U} \to \mathbb{R}^n \) is locally Lipschitz continuous, smooth over \( \partial \Omega \), and \( \nabla f(u) \) satisfies (9), (10) with \( S = \mathcal{U} \) and \( \Omega \subseteq \mathcal{U} \). Suppose that
\[
(f(\omega) - \omega, \nu(\omega)) \leq 0
\]
for all \( \omega \in \partial \Omega \) and \( \nu(\omega) \) in the outer normal cone of \( \Omega \) at \( \omega \). Then (C1) and (C2) imply that \( f(\Omega) \subseteq \Omega \).

**Proof.** As in the proof of Corollary 1 the assertion follows from the fact that if \( u \in \Omega \), \( \omega \in \partial \Omega \) and \( \nu(\omega) \) is in the outer normal cone of \( \Omega \) at \( \omega \), we have
\[
(f(u) - \omega, \nu(\omega)) = (f(u) - f(\omega), \nu(\omega)) + (f(\omega) - \omega, \nu(\omega)) \leq 0,
\]
which in turn implies that \( f(u) \in \Omega \) for any \( u \in \Omega \).

We end this section with some remarks about possible extensions of the results given here to the context of infinite-dimensional spaces. One such immediate extension is as follows. Let us assume that \( f \) is a Lipschitz continuous compact map from a separable Hilbert space \( \mathcal{H} \) into itself. Let \( \Omega \) be as in (1)–(3) where now the \( G_j \) are defined in \( \mathcal{H} \). We again suppose that \( f \) is smooth over \( \partial \Omega \). We now add the assumption that \( \nabla^2 G_j(\omega)|_{T_\omega(\partial \Omega)} \) is compact, viewed as a linear symmetric transformation of \( T_\omega(\partial \Omega) \subseteq \mathcal{H} \) for all \( \omega \in S_j \), \( j = 1, \ldots, N \). Then Theorems 3, 4 and the corresponding corollaries hold with \( \mathbb{R}^n \) replaced by \( \mathcal{H} \). The proofs of such results are identical to the proofs of the corresponding finite-dimensional ones with the only modifications being that instead of the spectral theorem for symmetric operators in \( \mathbb{R}^n \) we use the spectral theorem for symmetric compact operators in separable Hilbert spaces, and Brower’s degree theory is replaced by Leray-Schauder’s. We omit further details. General results about maps of convex sets in Banach spaces together with applications will be given elsewhere.

### 3. Finite difference schemes for systems of conservation laws

In [11] we studied invariant regions under Lax-Friedrichs schemes for nonlinear systems of conservation laws in several space variables. For simplicity of notation only, let us henceforth restrict our discussion to the two-dimensional case. So, these systems have the form
\[
\partial_t u + \partial_x f(u) + \partial_y g(u) = 0, \quad t > 0, \quad (x, y) \in \mathbb{R}^2,
\]
where \( u \) takes its values in \( \mathbb{R}^n \), \( f(u) \) and \( g(u) \) are vector fields in some domain \( \mathcal{U} \subseteq \mathbb{R}^n \).

In this section we strengthen the result in [11] and obtain other more general results concerning a wider class of finite-difference schemes for (16), with initial data
\[
u(x, y, 0) = u_0(x, y).
\]
We consider the Lax-Friedrichs scheme given by
\[
\begin{align*}
\mathbf{u}^{k+1}_{ij} & = \frac{1}{4}(\mathbf{u}^{k}_{i-1,j} + \mathbf{u}^{k}_{i+1,j} + \mathbf{u}^{k}_{i,j-1} + \mathbf{u}^{k}_{i,j+1}) \\
& - \frac{\lambda x}{2}(f(\mathbf{u}^{k+1}_{i+1,j}) - f(\mathbf{u}^{k}_{i-1,j})) - \frac{\lambda y}{2}(g(\mathbf{u}^{k+1}_{i,j+1}) - g(\mathbf{u}^{k}_{i,j-1})),
\end{align*}
\]  
(18)
and the so-called flux-splitting schemes. The latter are based on a decomposition of the flux functions \( f, g \) of the form
\[
\begin{align*}
f = f^+ + f^-, \quad g = g^+ + g^-,
\end{align*}
\]  
(19)
such that the Jacobian matrices \( \nabla f^\pm, \nabla g^\pm \) are diagonalizable, being the eigenvalues of those with \(^+\) superscript positive and the eigenvalues of those with \(^-\) superscript negative. Given such decompositions, let
\[
\begin{align*}
F(u, v) = f^+(u) + f^-(v), \quad G(u, v) = g^+(u) + g^-(v).
\end{align*}
\]  
The flux-splitting schemes are given by
\[
\begin{align*}
\mathbf{u}^{k+1}_{ij} & = \mathbf{u}^{k}_{i,j} - \lambda x(F(\mathbf{u}^{k}_{i,j}, \mathbf{u}^{k}_{i+1,j}) - F(\mathbf{u}^{k}_{i,j}, \mathbf{u}^{k}_{i,j})) \\
& - \lambda y(G(\mathbf{u}^{k}_{i,j}, \mathbf{u}^{k}_{i,j+1}) - G(\mathbf{u}^{k}_{i,j-1}, \mathbf{u}^{k}_{i,j})).
\end{align*}
\]  
(20)
Here, as usual, \( i, j \in \mathbb{Z}, k = 0, 1, 2, \ldots \), where \( \lambda x = \Delta t/\Delta x, \lambda y = \Delta t/\Delta y \) and we start by setting \( \mathbf{u}^0_{ij} = \mathbf{u}_0(i \Delta x, j \Delta y) \).

Since the early 80’s, flux-splitting schemes have received a great deal of attention, as can be seen from the vast number of articles published on this subject in the last two decades (e.g., see [40, 26, 27, 4], and the references therein). The main point about the flux-splitting in connection with finite-difference schemes is that upstream schemes such as Godunov’s and numerous of its modifications (see [21, 15]) usually give better resolution of the discontinuities. On the other hand, they are restricted to hyperbolic systems and have difficult implementation when the eigenvalues do not possess one and the same fixed sign throughout the whole region of the state space where the solutions take their values. A flux-splitting induces a decomposition of the original scheme into a sum of two trivial upwind schemes of much easier implementation. Nevertheless, the simplest splittings share similar problems of resolution of discontinuities with Lax-Friedrichs, which, by the way, in the case of hyperbolic systems can also be written as a flux-splitting scheme (cf. [35]). We, nevertheless, treat Lax-Friedrichs as an independent scheme since it has a simple direct definition which does not require any hyperbolicity restriction over the flux functions to have a meaning. In this connection, we would like to remark that systems of mixed type can be given hyperbolic splittings [35]. In this case, of course, the decomposing pairs cannot be simultaneously symmetrized in non-hyperbolic regions.

**Definition 1.** We say that \( \Omega \) is invariant under the schemes (18) or (20) if there exists \( \lambda_0 > 0 \) such that for \( \max(\lambda^x, \lambda^y) \leq \lambda_0, \) if \( \mathbf{u}^0_{ij} \in \Omega \) for all \( i, j \in \mathbb{Z} \), then \( \mathbf{u}^k_{ij} \in \Omega \) for all \( i, j \in \mathbb{Z} \) and \( k \in \mathbb{Z}_+ \).
We also consider a weaker notion of invariance, of a local character, which is designed for the study of systems whose flux functions are not globally Lipschitz, such as the Euler equations for non-isentropic gas dynamics which will be discussed in Section 4.

**Definition 2.** Let \( f, g, f^\pm, g^\pm \) be locally Lipschitz continuous maps from an open convex set \( \mathcal{U} \subseteq \mathbb{R}^n \) to \( \mathbb{R}^n \). We say that \( \Omega \subseteq \mathbb{R}^n \) is locally invariant under the schemes (18) or (20) applied to (16) at \( \overline{u} \in \mathcal{U} \cap \Omega \) if there exists \( r > 0 \) such that

\[
B(\overline{u}, r) = \{ u \in \mathbb{R}^n : |u - \overline{u}| \leq r \} \subseteq \mathcal{U}
\]

and \( \lambda_0(\overline{u}, r) > 0 \), with the property that whenever, for some \( i, j \in \mathbb{Z} \) and \( k \geq 0 \), we have \( u^k_{ij}, u^k_{ij \pm 1} \in B(\overline{u}, r) \cap \overline{\Omega} \) and \( \max\{\lambda^x, \lambda^y\} < \lambda_0(\overline{u}, r) \), then \( u^{k+1}_{ij} \in \overline{\Omega} \). This property is non-trivial only when \( \overline{u} \in \partial\Omega \).

**Theorem 3.** allows us to give a final characterization for the invariant regions under the Lax-Friedrichs schemes.

**Theorem 5.** Let \( \Omega \) and \( \partial\Omega \), be as in (1)-(3) and \( \mathcal{U} \subseteq \mathbb{R}^n \) be an open set containing \( \Omega \cup \partial\Omega \). Assume that \( f, g : \mathcal{U} \to \mathbb{R}^n \) are locally Lipschitz continuous and smooth over \( \partial\Omega \). If (C1) holds and (C2) is verified for \( f \) and \( g \), then \( \overline{\Omega} \) is locally invariant under the Lax-Friedrichs scheme (18) at all \( \overline{u} \in \partial\Omega \). In this case, \( \lambda_0(\overline{u}, r) \) may be taken as \( \left( 4 \max\{\text{Lip}(f, \overline{u}, r), \text{Lip}(g, \overline{u}, r)\} \right)^{-1} \), where \( \text{Lip}(f, \overline{u}, r) = \text{Lip}(f|B(\overline{u}, r)), \overline{u} \in \partial\Omega \).

Moreover, in the case where \( f, g \) are defined and (globally) Lipschitz over \( \overline{\Omega} \), \( \overline{\Omega} \) is invariant under (18) if and only if those conditions hold. For \( \lambda_0 \), in the latter case, we may take \( \left( 4 \max\{\text{Lip}(f), \text{Lip}(g)\} \right)^{-1} \).

**Proof.** We first prove the second part of the statement where \( f, g \) are assumed to be defined and Lipschitz continuous over \( \overline{\Omega} \). The proof of this part of Theorem 5 follows as in [11]. Here, besides the improvement of dropping condition (C3), defined therein, as a consequence of Theorem 3, we also establish a more precise estimate for \( \lambda_0 \). The main idea for the proof of the sufficiency part is to prove that for any \( \omega \in \partial\Omega \) we must have

\[
\langle u^{k+1}_{ij}, \omega, \nu(\omega) \rangle \leq 0
\]

if \( \nu(\omega) \) is the outer normal cone of \( \Omega \) at \( \omega \), provided that \( u^k_{ij} \in \overline{\Omega} \) for all \( i, j \in \mathbb{Z} \). The validity of inequality (21) for all \( \omega \in \partial\Omega \) and (C1) immediately imply \( u^{k+1}_{ij} \in \overline{\Omega} \). So, for \( M \geq M_0 = 2 \max\{\text{Lip}(f), \text{Lip}(g)\} \), let us define \( f_M = M1 + f \) and \( g_M = M1 + g \). From (C1), (C2) and Theorem 3 we have

\[
\langle u - \omega, \nu(\omega) \rangle \leq 0,
\]

\[
\langle f_M(u) - f_M(\omega), \nu(\omega) \rangle \leq 0,
\]

\[
\langle g_M(u) - g_M(\omega), \nu(\omega) \rangle \leq 0,
\]
for \( u \in \overline{\Omega} \) and \( \nu(\omega) \) in the outer normal cone of \( \Omega \) at \( \omega \in \partial \Omega \). Then, from (18) we get
\[
\langle u_{j+1}^{k} - \omega, \nu(\omega) \rangle 
\leq \left( \frac{u_{j+1}^{k} + u_{j+1}^{k-1}}{4} - \frac{\omega}{2}, \nu(\omega) \right) 
- \frac{\lambda_x}{2} \left( f(u_{j+1}^{k}) - f(\omega), \nu(\omega) \right) 
- \frac{\lambda_y}{2} \left( g(u_{j+1}^{k}) - g(\omega), \nu(\omega) \right) 
+ \left( \frac{1}{4} - \frac{M_0 \lambda_x}{2} \right) \langle u_{j-1}^{k} - \omega, \nu(\omega) \rangle 
+ \left( \frac{1}{4} - \frac{M_0 \lambda_y}{2} \right) \langle u_{j-1}^{k} - \omega, \nu(\omega) \rangle 
\leq - \frac{1}{4} \left( u_{j+1}^{k} - \omega, \nu(\omega) \right) \left( -1 + 2 \lambda_x \left( f(u_{j+1}^{k}) - f(\omega), \nu(\omega) \right) \right) 
- \frac{1}{4} \left( u_{j+1}^{k} - \omega, \nu(\omega) \right) \left( -1 + 2 \lambda_y \left( g(u_{j+1}^{k}) - g(\omega), \nu(\omega) \right) \right) 
+ \left( \frac{1}{4} - \frac{M_0 \lambda_x}{2} \right) \langle u_{j-1}^{k} - \omega, \nu(\omega) \rangle 
+ \left( \frac{1}{4} - \frac{M_0 \lambda_y}{2} \right) \langle u_{j-1}^{k} - \omega, \nu(\omega) \rangle,
\]
where we used (23), (24) in the first inequality. Now, Theorem 3 and (22) imply that the right-hand side of the last inequality above is non-positive for \( \lambda^x, \lambda^y \leq (2M_0)^{-1} \) and so (21) follows. This concludes the sufficiency part of the statement.

For the converse, let us first assume that (C1) is not true. We will prove that \( \overline{\Omega} \) cannot be invariant under (18). It suffices to consider initial data not depending on the second subscript. If (C1) does not hold, we can find \( u_0^{i-1} = u_i^{k} \) and \( u_1^{i+1} \) in \( \overline{\Omega} \) such that \( (3u_0^{i-1} + u_0^{i+1})/4 \) is out of \( \overline{\Omega} \). This follows from the fact that, given any \( 0 < \theta_0 < 1 \), a closed set is convex if and only if with any two of its points \( u, v \) it also contains \( \theta_0 u + (1 - \theta_0) v \), which is an easy exercise in convexity. Then, for \( \lambda^x = 0 \) in (18) with \( k = 0 \), we find that \( u_1^{i} \) is out of \( \overline{\Omega} \), which will also hold for \( \lambda^x > 0 \) sufficiently small, and so \( \overline{\Omega} \) is not invariant under (18). Now, let us assume for contradiction that (C2) does not hold, say, for \( f \) and that \( \overline{\Omega} \) is invariant under (18). Since we have already proved that (C1) is necessary we may assume that \( \overline{\Omega} \) is convex. Hence we can find \( \omega \in \partial \Omega \), so that
\[
\langle \nabla f(\omega) \tau, \nu(\omega) \rangle < 0
\]
for some \( \tau \in T_\omega(\partial \Omega) \), where \( \nu(\omega) \) is the unit outer normal vector at \( \omega \). Let \( \omega(t), t \in (-\delta, \delta) \) be a smooth curve in \( \partial \Omega \), with \( \omega(0) = \omega, \omega'(0) = \tau \). Then, for \( t > 0 \) sufficiently small, we have
\[
\frac{1}{4} \left( \omega(t) - \omega, \nu(t) \right) - \frac{\lambda^x}{2} \left( f(\omega(t)) - f(\omega), \nu(t) \right) > 0.
\]
Therefore, setting \( u_k^0 = u_{k-1}^0 = \omega, u_{k+1}^0 = \omega(t) \) in (18) with \( k = 0 \), it follows from the above inequality that \( (u_1^0 - \omega, \nu(\omega)) > 0 \) and so \( u_1^0 \notin \Omega \), contradicting our assumption.

We now consider the first part of the statement concerning the local invariance of \( \Omega \) at all points of \( \partial \Omega \). By Theorem 3, given \( u \in \partial \Omega \), it is possible to obtain an open ball \( B_0 \subseteq B_1 \subseteq \mathcal{U} \) around it such that \( f_M(B_0 \cap \Omega) \) is contained in a convex subset of \( f_M(B_1 \cap \Omega) \), the latter being homeomorphic through \( f_M \) to \( B_1 \cap \Omega \), with \( M > M_0 = 2\text{Lip}(f(B_1)) \). Let \( B \subseteq B_0 \) be another ball around \( u \) with radius \( r \) equal to, say, 1/2 the radius of \( B_0 \), and let \( \lambda_0(u, r) \) be chosen so that if \( u_{i,j}^{k+1} \), given by (18), must belong to \( B_0 \) provided that \( \max(\lambda^+, \lambda^-) < \lambda_0 \). Then, the same reasoning as above leads to the validity of (21) for all \( \omega \in B_0 \cap \partial \Omega \) and \( \nu(\omega) \), the corresponding unit outer normal. On the other hand, by the choice of \( \lambda_0 \) we already know that (21) is true for \( \omega \in \partial B_0 \cap \Omega \) and \( \nu(\omega) \), the corresponding outer normal of \( \partial B_0 \) provided that \( \max(\lambda^+, \lambda^-) \leq \lambda_0 \). Hence, by the convexity of \( B_0 \cap \Omega \) we conclude that \( u_{i,j}^{k+1} \in B_0 \cap \Omega \) as desired, completing the proof.

Concerning the flux-splitting schemes given by (20) we have the following result.

**Theorem 6.** Let \( \Omega \) and \( \partial \Omega \) be as in (1)–(3) and \( \mathcal{U} \subseteq \mathbb{R}^n \) be an open set containing \( \Omega \cup \partial \Omega \). Assume that \( f^\pm, g^\pm : \mathcal{U} \to \mathbb{R}^n \) are locally Lipschitz continuous in \( \mathcal{U} \) and smooth over \( \partial \Omega \). Suppose that \( \pm \nabla f^\pm, \pm \nabla g^\pm \) satisfy (9), (10), with \( S = \partial \Omega \). If (C1) holds and (C2) is verified for \( f^\pm \) and \( g^\pm \), then \( \Omega \) is locally invariant under the flux-splitting scheme (20) at all \( \overline{\Omega} \in \partial \Omega \). In this case, \( \lambda_0(u, r) \) may be taken as

\[
(2 \max \left\{ \rho(\nabla f^\pm, \overline{u}, r), \rho(\nabla g^\pm, \overline{u}, r) \right\})^{-1},
\]

where \( \rho(\nabla f^\pm, \overline{u}, r) \) denotes the supremum of the spectral radius of \( \nabla f^\pm \) in \( B(\overline{u}, r) \).

Moreover, if \( f^\pm, g^\pm \) are Lipschitz continuous over \( \mathcal{U} \), \( \pm \nabla f^\pm, \pm \nabla g^\pm \) satisfy (9), (10), with \( S = \mathcal{U} \) and \( \overline{\Omega} \subseteq \mathcal{U} \), then \( \overline{\Omega} \) is invariant under the flux-splitting scheme (20).

In the latter case, we may take \( \lambda_0 \) as

\[
(2 \max \left\{ \rho(\nabla f^\pm), \rho(\nabla g^\pm) \right\})^{-1}.
\]

Conversely, if \( \overline{\Omega} \) is convex and invariant under the flux-splitting scheme (20), then (C2) must be verified for \( f^\pm \) and \( g^\pm \).

**Proof.** It will suffice to prove the second part of the statement in which \( f^\pm, g^\pm \) satisfy (9), (10), with \( S = \mathcal{U} \) and \( \overline{\Omega} \subseteq \mathcal{U} \). The first part, concerning local invariance, will follow from the same reasoning used in the proof of the first part, as in the case of Theorem 5, and it will become clear. So let us assume \( f^\pm, g^\pm \) to be defined and Lipschitz in \( \mathbb{R}^n \). Again the sufficiency part reduces to showing (21). Now from
(C1), (C2) and Theorem 4 we have
\[
\langle u - \omega, \nu(\omega) \rangle \leq 0, \quad (25)
\]
\[
\langle f^+(u) - f^+(\omega), \nu(\omega) \rangle \leq 0, \quad (26)
\]
\[
\langle g^+(u) - g^+(\omega), \nu(\omega) \rangle \leq 0, \quad (27)
\]
\[
\langle f^-(u) - f^-(\omega), \nu(\omega) \rangle \geq 0, \quad (28)
\]
\[
\langle g^-(u) - g^-(\omega), \nu(\omega) \rangle \geq 0, \quad (29)
\]
keeping the notation in the proof of Theorem 5. Then, from (20), we get
\[
\langle u^{k+1}_{ij} - \omega, \nu(\omega) \rangle \leq \langle u^k_{ij} - \omega, \nu(\omega) \rangle
\]
\[
- \frac{\lambda x}{2} (f^+(u^k_{ij}) - f^+(\omega), \nu(\omega)) - \frac{\lambda y}{2} (g^+(u^k_{ij}) - g^+(\omega), \nu(\omega))
\]
\[
+ \frac{\lambda x}{2} (f^-(u^k_{ij}) - f^-(\omega), \nu(\omega)) + \frac{\lambda y}{2} (g^-(u^k_{ij}) - g^-(\omega), \nu(\omega))
\]
\[
\leq \langle u^k_{ij} - \omega, \nu(\omega) \rangle - (1 - M_0(\lambda x + \lambda y)) \langle u^k_{ij} - \omega, \nu(\omega) \rangle,
\]
where in the first inequality we used Theorem 4, (26)–(29), and in the last one (25) and again Theorem 4 (Remark 2) with
\[
M_0 = \max \{\rho(\nabla f^\pm), \rho(\nabla g^\pm)\}.
\]
Hence, for $\lambda x, \lambda y \leq (2M_0)^{-1}$, (21) follows. This proves the sufficiency part. The last part of the statement follows by an argument entirely similar to that used in the proof of the last part of the converse of Theorem 5.

**Remark 3.** The symmetrizations to which Theorem 6 refers can be made, for each of the functions $\nabla f^\pm(u), \nabla g^\pm(u)$, by different positive definite symmetric matrices $P(u)$. In particular, flux-splittingss of systems of mixed type are not excluded.

**Remark 4.** We cannot expect the invariance of $\Omega$ under (20) applied to a general system (16) admitting a flux-splitting as above to imply both (C1) and (C2). To see this it suffices to consider the trivial case in which $f \equiv g \equiv 0$. Hence, some additional condition excluding this trivial case must be imposed in order to get a partial converse. Besides the partial converse given by Theorem 6, stating that invariance plus (C1) implies (C2) we have the following facts. In the case of systems (16) of two equations, $n = 2$, if $\Omega$ is invariant under the flux-splitting scheme (20), and if, for each $j = 1, \ldots, N$, (C2) holds in $\partial \Omega \cap \{G_j = 0\}$ for at least one among
Then (C1) holds, that is, \(\Omega\) must be convex. More generally, for systems (16) with any number of equations, if \(\Omega\) is invariant under the flux-splitting scheme (20), and if, for each \(j = 1, \ldots, N\), (C2) holds in \(\partial \Omega \cap \{G_j = 0\}\) for at least one among \(f^\pm, g^\pm\), say, \(f^+\), and besides, \(\nabla^2 G_j(\omega)|_{\partial \Omega}\) and \(\nabla f^+(\omega)|_{\partial \Omega}\) possess a common basis of eigenvectors, then (C1) must hold. The proof of these facts relies on the observation that the one-dimensional upstream scheme with flux-vector given by any one of \(f^\pm, g^\pm\), say \(f^+\), may be expressed as an average of the values of the solution of a Riemann problem with left and right states given by, say, \(u_{k-1}^i\) and \(u_k^i\), when these two states are sufficiently close. Hence, if \(\Omega\) were not convex it would be possible, by the assumptions, to find a rarefaction curve \(u(s), s \in (-\delta_0, \delta_0)\), contained in \(\partial \Omega\), upward convex with respect to the outer normal at \(u(0)\). In this case, the average of the values of the Riemann solution given by the rarefaction wave connecting \(u(-\delta)\) to \(u(\delta)\), for \(0 < \delta < \delta_0\) sufficiently small, would fall outside \(\Omega\) and the latter could not be invariant. We omit further details.

**Remark 5.** The one-dimensional case of Theorem 5 has a rather wide range of direct applications to systems of conservation laws for which the corresponding theorem in [5], concerning the artificial viscosity approximation for (16), is applicable. These include \(p\)-systems, nonlinear elasticity equations, Euler equations of isentropic gas dynamics, plane waves in nonlinear electromagnetism, etc. For such applications we refer to, e.g., [33, 11]. Examples of applications to systems which change type (hyperbolic-elliptic) can be found in [11, 13]. Theorem 5 also directly applies to the multi-dimensional systems in Temple’s class considered in [12]. We will discuss below an application of this theorem to the multi-D system of Euler equations in compressible fluid dynamics.

**Remark 6.** The one-dimensional case of Theorem 6 is directly applicable to the entropy flux-splittings for \(p\)-systems and Euler equations of isentropic gas dynamics which are presented and discussed in [4]. We also discuss below an application of this theorem to the multi-D system of Euler equations in compressible fluid dynamics.

### 4. Application to the compressible Euler equations

In this section we apply the results of Section 3 to the system of Euler equations for compressible fluids, which in two space variables reads

\[
\begin{bmatrix}
\rho \\
\rho u \\
\rho v \\
\rho E \\
\end{bmatrix}_t + \begin{bmatrix}
\rho u \\
\rho u^2 + p \\
\rho uv \\
u(\rho E + p) \\
\end{bmatrix}_x + \begin{bmatrix}
\rho v \\
\rho uv \\
\rho v^2 + p \\
v(\rho E + p) \\
\end{bmatrix}_y = 0.
\]

Here, as usual, \(\rho\) is the density, \(v = (u, v)\) is the vector velocity, \(p\) is the pressure and \(E = \frac{1}{2}|v|^2 + e\) is the total specific energy while \(e\) is the internal specific energy. Other important variables connected to (30) are the absolute temperature \(T\) and
the entropy $S$. The thermodynamic variables $\rho$, $p$, $e$, $T$ are always non-negative, with $p \to 0$ as $\rho \to 0$ and $e \to 0$ as $T \to 0$. Nernst normalization assuming that $S \to 0$ as $T \to 0$ ensures that $S$ is also always non-negative. The limit cases in which $\rho = 0$ (vacuum) or $T = 0$ (absolute zero) will not be considered here, and so we assume all the thermodynamic variables $\rho$, $p$, $e$, $T$, $S$ to be positive:

$$\rho, \ p, \ e, \ T, \ S > 0.$$  

The second law of thermodynamics relates the variables $\rho$, $p$, $e$, $S$, $T$ by the differential formula

$$de = T \, dS + \frac{p}{\rho^2} \, d\rho.$$  \hspace{1cm} (31)

To (30) and (31) we add three equations of state giving three among the four variables $p$, $e$, $S$, $T$ as functions of $\rho$ and the other remaining variable, say,

$$p = p(\rho, T), \quad e = e(\rho, T), \quad S = S(\rho, T).$$

These equations of state are not independent since they must be compatible with (31), which implies that

$$\rho^2 \frac{\partial e}{\partial \rho} = -T^2 \frac{\partial}{\partial T} \left( \frac{p}{T} \right),$$

$$\frac{\partial e}{\partial T} = T \frac{\partial S}{\partial T},$$

$$\rho^2 \frac{\partial S}{\partial \rho} = \frac{1}{T} \left( \rho^2 \frac{\partial e}{\partial \rho} - p \right).$$  \hspace{1cm} (32)

Another frequently used form of the equations of state is

$$p = p(\rho, S), \quad e = e(\rho, S), \quad T = T(\rho, S),$$

whose compatibility equations dictated by (31) are

$$\rho^2 \frac{\partial e(\rho, S)}{\partial \rho} = p(\rho, S),$$

$$\frac{\partial e(\rho, S)}{\partial S} = T(\rho, S),$$

$$\rho^2 \frac{\partial T(\rho, S)}{\partial \rho} = \frac{\partial p(\rho, S)}{\partial S}.\hspace{1cm} (33)$$

Clapeyron’s law for ideal gases reads

$$p = R\rho T,$$  \hspace{1cm} (34)

where $R > 0$ is the gas constant. The ideal gas is said to be polytropic if

$$e = c_v T,$$  \hspace{1cm} (35)
for a positive constant $c_v$, known as the specific heat at constant volume. For an ideal polytropic gas, the equation of state for the pressure as a function of the variables $\rho, S$ becomes

$$p = \kappa \epsilon^{S/c_v} \rho^\gamma,$$

(36)

where $\kappa$ is a positive constant, $\epsilon$ denotes the base of the Napierian logarithms, and $\gamma = (R + c_v)/c_v$. The equations of state for $e$ and $T$ follow directly from (34) and (35).

In the conservative variables $u = (\rho, \rho u, \rho v, \rho E)$ the system (30) has the form (16) with $f(u) = (\rho u, \rho u^2 + p, \rho uv, \rho uE + pu)$ and $g(u) = (\rho v, \rho uv, \rho v^2 + p, \rho vE + pv)$. To obtain $p$ as a function of the conservative variables we first obtain it as a function of $(\rho, e)$, which is always possible, and then replace $e$ by $E - \frac{1}{2}|v|^2$.

For smooth solutions, the system (30) can be rewritten in the form

$$\rho t + (\rho u)x + (\rho v)y = 0,$$

(37)

$$u_t + uu_x + \frac{p_x}{\rho} + uu_y = 0,$$

(38)

$$v_t + uv_x + \frac{p_y}{\rho} + vv_y = 0,$$

(39)

$$S_t + uS_x + vS_y = 0,$$

(40)

with the help of (31). In particular, (40) shows that $S$ is a Riemann invariant for (30), or, more specifically, that

$$\nabla S \nabla f(u) = u \nabla S \quad \text{and} \quad \nabla S \nabla g(u) = v \nabla S,$$

(41)

where the $\nabla$ is taken with respect to the conservative variables. Also from equations (37)–(40) we easily see that $\nabla f(u)$ and $\nabla g(u)$ are diagonalizable provided that

$$\frac{\partial p(\rho, S)}{\partial \rho} > 0,$$

(42)

in which case we write $\frac{\partial p(\rho, S)}{\partial \rho} = c^2$ where $c$ is known as the local sound speed. In this case, the eigenvalues of $\nabla f(u)$ are $u \pm c$ and $u$ with multiplicity 2 and those of $\nabla g(u)$ are $v \pm c$ and $v$ with multiplicity 2.

Equation (40), valid for smooth solutions, can also be written in conservative form as

$$(\rho S)_t + (\rho uS)_x + (\rho vS)_y = 0.$$

(43)

Physical non-smooth solutions, which correspond to irreversible processes, should satisfy, instead of (43), the Clausius inequality:

$$(\rho S)_t + (\rho uS)_x + (\rho vS)_y \geq 0.$$

(44)

In terms of the theory of conservation laws (see, e.g., [7,33]) (43), (44) together with the hyperbolicity condition (42) express the fact that $\eta(u) = -\rho S$ is a convex
mathematical entropy for the system (16) with entropy flux components \( q_1(u) = -\rho \, u \, S \) and \( q_2(u) = -\rho \, v \, S \), that is,

\[
\nabla q_1(u) = \nabla q(u) \nabla f(u), \quad \nabla q_2(u) = \nabla q(u) \nabla g(u),
\]

and

\[
\nabla^2 q(u) \geq 0.
\]

Equations (45) imply, in particular, that \( \nabla^2 q(u) \) is a symmetrizer for both \( \nabla f(u) \) and \( \nabla g(u) \) whenever the strict inequality in (46) holds.

Now let us assume that the equations of state are set in the form

\[
p = p(\rho, e), \quad S = S(\rho, e), \quad T = T(\rho, e),
\]

whose compatibility relations imposed by (31) imply

\[
\rho^2 S_\rho + p S_e = 0, \quad S_e = \frac{1}{T}.
\]

In particular, the conditions \( T > 0 \) and (42) transform into

\[
S_e > 0, \quad \rho^2 p(\rho, e) + pp_e(\rho, e) > 0.
\]

In [33], it is proved that conditions (47) imply that \( -S \), as a function of the conservative variables, is quasi-convex. Due to the smoothness of \( -S \), this is equivalent to saying that the domains

\[
\Omega_r = \{ u \in U \subseteq \mathbb{R}^4 : S \geq r \}
\]

are convex sets for any \( r > 0 \), where

\[
U = \{ u = (u_1, \ldots, u_4) \in \mathbb{R}^4 : u_1 > 0, \ u_1 u_4 > \frac{1}{2} (u_2^2 + u_3^2) \}.
\]

In the one-dimensional case, this can also be obtained from the convexity of \(-\rho \, S\), which in turn is a consequence of the hyperbolicity condition (42) and the Clausius inequality (44) together with (43), as an application of a general result of Lax [20].

We observe that, since \( S \to 0 \) as \( e \to 0 \) and \( e = 0 \) on \( \partial U \), the sets \( \Omega_r, r > 0 \), above are entirely contained in \( U \), except for the \( u_4 \)-axis where \( u_1, u_2, u_3 = 0 \). This can be better visualized in the special case of ideal polytropic gases.

The convexity of the domains \( \Omega_r, r > 0 \), and the fact, expressed by (41), that \( S \) is a Riemann invariant for both \( f \) and \( g \) allow the application of Theorem 5 to conclude that, for any \( r > 0 \), \( \Omega_r \) is locally invariant under the Lax-Friedrichs scheme applied to (30), according to Definition 2.

**Theorem 7.** For any \( r > 0 \), the domain \( \Omega_r \) given by (48) is locally invariant under the Lax-Friedrichs scheme (18) applied to the system of compressible Euler equations (30), at all points of \( \partial \Omega_r \cap U \).
Let us now focus on flux-splitting schemes for (30). We will restrict the discussion to ideal polytropic gases. The first flux-splitting schemes for these were proposed in [31] and [36]. Since then a vast literature has been dedicated to the construction of flux-splitting schemes for (30), most of which are specifically for the polytropic gases (see [40,4] and the references therein). Among these a special role is played by those splittings obtained through a kinetic formulation of the Euler equations which seem to have been first proposed in [18] and [28] (see also [29]).

Following previous work by Deshpande [8], Perthame in [26, 27] analyzes such schemes which belong to the so-called class of Boltzmann-type schemes whose general theory was formulated in [15]. The important feature of the analysis in [8,26,27] is related to the verification of the entropy property for the corresponding flux-splitting schemes, that is, the validity of (44) for the approximate solution. In particular, this amounts to the verification that $-\rho S$ is a mathematical convex entropy, in the sense of (45), shared by all fluxes $f^\pm, g^\pm$, which implies that $\nabla f^\pm, \nabla g^\pm$ are simultaneously symmetrizable by $\nabla^2 (-\rho S)$. It follows as above, using the conservation equation for $\rho$, that the thermodynamic entropy $S$ is a Riemann invariant also shared by all fluxes $f^\pm, g^\pm$, that is, $\nabla S$ is a left-eigenvector of $\nabla f^\pm, \nabla g^\pm$. Hence, the flux-splitting schemes based on kinetic formulation constructed in [26,27] for the Euler equations (30), for polytropic gases, satisfy all the conditions of Theorem 6. Again, from the convexity of the domains $\Omega_r$, $r > 0$, defined in (48), and the fact that $S$ is a Riemann invariant common to all fluxes $f^\pm, g^\pm$ it follows that these regions are locally invariant for these schemes. More generally, we have the following result.

**Theorem 8.** Let (19) be a flux-splitting for the compressible Euler equations (30) such that $-\rho S$ is a mathematical entropy shared by all fluxes $f^\pm, g^\pm$ (e.g., [26, 27]). For any $r > 0$, the domain $\Omega_r$ given by (48) is locally invariant under the corresponding flux-splitting scheme (20) at all points of $\partial \Omega_r \cap U$.

Acknowledgements. This research was partially supported by CNPq, under the grant 352871/96-1 and FAPERJ, grant E-26/151.890/2000. H. F. would like to thank his colleagues at IMPA, Alcides Lins Neto, Benar Svaiter and Carlos Isnard, for many stimulating conversations during the preparation of this paper.

**References**


Instituto Nacional
de Matemática Pura e Aplicada – IMPA
Estrada Dona Castorina, 110
22460-320 Rio de Janeiro RJ, Brasil
e-mail: hermano@impa.br

*(Accepted May 25, 2001)*

*Published online October 1, 2001 – © Springer-Verlag (2001)*