Qualitative properties of trajectories of control systems: a survey

F. H. Clarke[*,†] Yu. S. Ledyaev[‡,†]
R. J. Stern[§,†] and P. R. Wolenski[¶]

CRM-2203

August 1994

[*] Centre de recherches mathématiques, Université de Montréal, C.P. 6128, Succ. centre-ville, Montréal, Québec, H3C 3J7, Canada
[†] Research supported by the Natural Sciences and Engineering Research Council of Canada, and by le fonds FCAR du Québec
[‡] Steklov Mathematics Institute, Moscow 117966, Russia
[§] Department of Mathematics & Statistics, Concordia University, Montréal, Québec, H4B 1R6, Canada
[¶] Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803, USA

Abstract

We present a unified approach to a complex of related issues in control theory, one based to a great extent on the methods of nonsmooth analysis. The issues include invariance, stability, equilibria, monotonicity, the Hamilton-Jacobi equation, feedback synthesis, and necessary condtions.

Résumé

On présente une approche unifiée à un ensemble de questions portant sur le comportement des trajectories d'un système contrôlé. Les sujets traités comprennent l'invariance (faible et forte), la stabilité, les équilibres, la monotonie, les solutions généralisées de l'équation Hamilton-Jacobi (incluant solutions viscosité), les feedbacks, et l'optimalité. Les techniques principales sont celles de l'analyse non lisse.

1 Introduction

This article considers certain basic properties of a finite-dimensional deterministic control system. The focus is on the following issues:

- weak invariance (*some* trajectory remains in a given set)
- strong invariance(*every* trajectory remains in a given set)
- equilibria (the system has rest points)
- local attainability of a set (the system can be steered in finite time to a given set, or to its interior)
- Lyapounov stability of invariant sets (functional criteria for stability)
- monotonicity along trajectories (functional counterparts of invariance)
- verification functions (sufficient conditions in optimal control)
- the Hamilton-Jacobi equation (generalized solution concepts)
- optimal and suboptimal feedback (control synthesis)
- Hamiltonian inclusions (necessary conditions and controllability in optimal control)

Of course, a complete overview of these topics is quite impractical here. What we present is a special point of view based upon a unifying approach that turns out to be quite natural, in our opinion, and to yield a number of satisfying answers to some basic questions. It is noteworthy that the approach is largely geometric in its inspiration, and gives rise to relatively simple, frequently algorithmic methods. Many of the results presented here are very recent or forthcoming, but some are older. This survey is the first to unify this body of work conceptually and technically, and it is our hope that it will be helpful in increasing contact among the several schools of thought involved in these issues.

The instruments employed are principally those of nonsmooth analysis and geometry, and the recurrent themes include invariance, feedback, the (true) Hamiltonian, and value functions. The dual interpretation of phenomena in geometric or analytical terms is frequent.

The control system studied here will be modeled as a differential inclusion:

$$\dot{x}(t) \in F\big(x(t)\big). \tag{1.1}$$

The essential equivalence of (1.1) to the explicitly parametrized system

$$\dot{x}(t) = f\big(x(t), u(t)\big), \quad u(t) \in U \tag{1.2}$$

is now well-understood, and we shall not dwell upon it. The model (1.1) is especially convenient when it is state trajectories (rather than controls) that are being focused upon. A *trajectory* (of F on a given interval $[0,T]$) is as usual an absolutely continuous function $x\colon [0,T] \to \mathbb{R}^n$ such that (1.1) holds a.e. on $[0,T]$.

F is a mapping from $\mathbb{R}^n$ to the subsets of $\mathbb{R}^n$. The following Standing Hypothesis will be in force throughout:

(SH). (a) For every $x \in \mathbb{R}^n$, $F(x)$ is a nonempty compact convex set;

(b) $x \mapsto F(x)$ is upper semicontinuous;

(c) For certain constants γ and c, and for all $x \in \mathbb{R}^n$,

$$v \in F(x) \Longrightarrow |v| \le \gamma |x| + c.$$

It is the case for each of the results contained in this article that, at least in part, the Standing Hypothesis could be weakened and the conclusion would remain valid. We comment occasionally on this issue, but by and large we sacrifice stating minimal hypotheses to the goal of emphasizing the main ideas.

An additional hypothesis on F that intervenes later is that it be locally Lipschitz: to every bounded set S in $\mathbb{R}^n$ there corresponds a constant K such that

$$F(x_2) \subseteq F(x_1) + K|x_2 - x_1| B \quad \forall x_1, x_2 \in S,$$

where B denotes the closed unit ball in $\mathbb{R}^n$.

The lower and upper Hamiltonians corresponding to F are functions defined respectively by

$$h(x,p) := \min\bigl\{\langle p,v\rangle : v \in F(x)\bigr\}$$
$$H(x,p) := \max\bigl\{\langle p,v\rangle : v \in F(x)\bigr\}.$$

We proceed to recall some elements of nonsmooth analysis (see for example [25]). Let S be a closed subset of $\mathbb{R}^n$, and let $s \in S$. If $x \notin S$, and if one of the closest points to x in S is s, the vector $x - s$ (and any nonnegative scalar multiple of it) is said to be a *proximal normal* to S at s. The set of such vectors is the proximal normal cone to S at s, denoted $N_S^P(s)$ (we set $N_S^P(s) = \{0\}$ if no x as described above exists).

Now let $f\colon \mathbb{R}^n \to (-\infty,\infty]$ be a lower semicontinuous, extended-valued function, and let x be a point such that $f(x)$ is finite. An element ζ of $\mathbb{R}^n$ is a *proximal subgradient* of f at x provided that for some $\sigma \geq 0$, and for all y in some neighborhood of x, we have

$$f(y) - f(x) + \sigma|y-x|^2 \geq \langle \zeta, y-x\rangle.$$

The (possibly empty) set of proximal subgradients of f at x is denoted $\partial_P f(x)$.

The link between normals and subgradients is via the *epigraph* of f, the set

$$\operatorname{epi}(f) := \bigl\{(x,r) \in \mathbb{R}^n \times \mathbb{R} \colon f(x) \leq r\bigr\}.$$

Then

$$\zeta \in \partial_P f(x) \iff (\zeta,-1) \in N^P_{\operatorname{epi}(f)}\bigl(x, f(x)\bigr).$$

Another tool of nonsmooth analysis is the classical tangent cone to the set S at $x \in S$. The *directional* or *Dini tangent cone* (also called the contingent or Bouligand cone) is the set of all limits of the form

$$\lim_{i\to\infty} \frac{s_i - x}{t_i}$$

where $s_i \in S$, $s_i \to s$, and where $t_i \downarrow 0$. We denote it $T_S^D(x)$. If $d_S(\cdot)$ signifies the distance function corresponding to S, then $T_S^D(x)$ consists of those v for which the Dini derivate $Dd_S(x;v)$ is zero, where in generic terms Df is defined by

$$Df(x;v) := \liminf_{\substack{v'\to v\\ t\downarrow 0}} \frac{f(x+tv') - f(x)}{t}.$$

Another very useful tangent cone is obtained by using a different directional derivate of d_S. The C-tangent cone, $T_S^C(x)$, consists of those v for which

$$d_S^0(x;v) := \limsup_{\substack{y\to x\\ t\downarrow 0}} \frac{d_S(y+tv) - d_S(y)}{t} \leq 0.$$

Finally, we recall the multidirectional mean value inequality [28]: given f and x as above and a compact convex subset Y of $\mathbb{R}^n$, for any $\varepsilon > 0$, there exists $z \in [x,Y] + \varepsilon B$ and $\zeta \in \partial_P f(z)$ such that

$$\min_Y f - f(x) < \langle \zeta, y - x\rangle + \varepsilon \quad \forall y \in Y,$$
$$f(z) < \min_{[x,Y]} f + \Bigl|\min_Y f - f(x)\Bigr| + \varepsilon,$$

where $[x,Y]$ is the "interval" $\operatorname{co}\bigl(\{x\} \cup Y\bigr)$.

The plan of the paper is described by the list of issues given at the beginning of the introduction, which coincides with the titles of sections 2 through 11.

2 Weak invariance

In studying the properties of the trajectories of a multifunction F in connection with a set S, we find it convenient to speak of the *system* (S,F), as for example in the case of the following important property:

Definition 2.1. The system (S,F) is said to be weakly invariant provided that for any $x_0 \in S$, there exists a trajectory $x(\cdot)$ for F satisfying $x(0) = x_0$ and $x(t) \in S$, $t \geq 0$.

We now give a proximal criterion for weak invariance, one whose proof is important since it introduces two recurrent themes: proximal aiming and feedback. Recall the Standing Hypothesis (SH) on F, and assume (as we always will henceforth) that S is closed; the lower Hamiltonian h corresponding to F was defined in §1. We remark that the proximal aiming technique used in the proof has a predecessor in differential games, namely the design of extremal strategies proposed by Krasovskii and Subbotin [61].

Theorem 2.1. *The system (S, F) is weakly invariant iff*

$$h(x,\zeta) \leq 0 \quad \forall \zeta \in N_S^P(x), \quad \forall x \in S. \tag{2.1}$$

Proof. Let us assume that (2.1) holds; we proceed to show that (S, F) is weakly invariant. To this end, let $x_0 \in S$ be given. We shall explicitly construct a trajectory $x(\cdot)$ with $x(0) = x_0$ which remains in S.

We begin by associating to each x in $\mathbb{R}^n$ a point $s = s(x)$ nearest to x in S. When S is not convex, there are in general many such nearest points s; $s(x)$ is merely a choice among them. The vector $x - s(x)$ lies in $N_S^P\big(s(x)\big)$ by construction, so by (2.1) there exists $v \in F\big(s(x)\big)$ such that $\big\langle v, x - s(x)\big\rangle \leq 0$. We denote an arbitrary choice of such v by $f(x)$. The remainder of the proof consists of showing that we can solve (in a certain sense) the Cauchy problem

$$\dot{x}(t) = f\big(x(t)\big), \quad x(0) = x_0 \tag{2.2}$$

and that the resulting solution $x(\cdot)$ lies in S. Then $\dot{x} \in F\big(s(x)\big) = F(x)$ and $x(\cdot)$ is a trajectory as required. The underlying technique here is new in this context, quite robust in applying to other settings, and we label it "proximal aiming", since it corresponds to choosing a velocity v ("aiming") in a direction determined by (actually opposite to) a proximal $x - s(x)$.

The difficulty of solving (2.2) is that f is not generally continuous, that $f(x)$ is an arbitrary element of $F(x)$ when $x \in S$, and that in fact we did not even insist on its being measurable. We take a pragmatic approach therefore, based upon the well-known numerical solution procedure which constructs piecewise-linear "approximate solutions" (Euler polygons) for partitions of the time interval. Later we pass to the limit. This procedure leads to what we shall call an *Euler solution* of (2.2), a solution concept familiar in differential games [61].

Let $\pi = \{t_0, t_1, \ldots, t_N\}$ denote a partition of the interval $[0, T]$, where T is fixed, and where we shall always take $t_0 = 0$, $t_N = T$. Once T is fixed, along with x_0, it is routine to reduce to the case in which S is compact, which we assume henceforth to simplify the proof. On $[t_0, t_1]$, we consider the differential equation

$$\dot{x}(t) = f(x_0), \quad x(t_0) = x_0$$

which has constant right side. We denote by x_1 the resulting value $x(t_1)$. On the next subinterval $[t_1, t_2]$ we take

$$\dot{x}(t) = f(x_1), \quad x(t_1) = x_1,$$

and so on, generating a piecewise-linear function $x_\pi(\cdot)$ on $[0, T]$ whose "nodes" are (t_i, x_i) $(i = 0, 1, \ldots, N)$. Now let x be any uniform limit of a sequence $x_{\pi_i}(\cdot)$, where the partitions π_i have diameter decreasing to 0. (Such a limit does exist, since f, and hence $\dot{x}_\pi$, is uniformly bounded.) Then x is termed an Euler solution of (2.1), and we claim that $x(t) \in S \quad \forall t \geq 0$.

To see this, note first that

$$d_S(x_1) \leq M(t_1 - t_0),$$

where M is a uniform bound on F in S, and since $x_0 \in S$. Then

$$\begin{aligned}
d_S^2(x_2) &\leq \big|x_2 - s(x_1)\big|^2 \qquad \text{(since } s(x_1) \in S\text{)} \\
&= |x_2 - x_1|^2 + \big|x_1 - s(x_1)\big|^2 + 2\big\langle x_2 - x_1, x_1 - s(x_1)\big\rangle \\
&\leq M^2(t_2 - t_1)^2 + d_S^2(x_1) + 2\int_{t_1}^{t_2} \big\langle \dot{x}_\pi(t), x_1 - s(x_1)\big\rangle\, dt \\
&\leq M^2\big\{(t_2 - t_1)^2 + (t_1 - t_0)^2\big\} + 2\int_{t_1}^{t_2} \big\langle f(x_1), x_1 - s(x_1)\big\rangle\, dt \\
&\leq M^2\big\{(t_2 - t_1)^2 + (t_1 - t_0)^2\big\},
\end{aligned}$$

since the integrand is nonpositive by construction of f. In general we have

$$\begin{aligned}
d_S^2(x_k) &\leq M^2 \sum_{i=0}^{k-1} (t_{i+1} - t_i)^2 \\
&\leq M^2 T \operatorname{diam}(\pi).
\end{aligned}$$

It follows that as $\text{diam}(\pi) \to 0$, the nodes x_k of the approximants converge to S, so that their uniform limit $x(\cdot)$ lies in S. Finally, note that each piecewise-linear approximation $x_\pi(\cdot)$ is an "approximate trajectory": for $t \in (t_i, t_{i+1})$ we have

$$\dot{x}_\pi(t) = f(x_i) \in F\big(s(x_i)\big),$$

where $s(x_i)$ is close to $x_\pi(t)$ as $\text{diam}(\pi) \to 0$. It follows readily from the sequential compactness of approximate trajectories (see [24]) that $x(\cdot)$ is a trajectory. This completes the proof that (2.1) implies weak invariance. The opposite implication is simpler, and stems from Theorem 2.2 below. □

Remark 2.1. (a) We remark again that the proof of the theorem constructed (via "proximal aiming") a function f on $\mathbb{R}^n$ with the property that, for all $x_0 \in S$, and for any Euler solution $x(\cdot)$ of $\dot{x} = f(x)$, $x(0) = x_0$, we have $x(t) \in S \quad \forall t \geq 0$. However, $f(x)$ is not necessarily continuous, and is not designed to achieve invariance except "in the limit". The question of constructing bona fide (continuous or Lipschitz) invariant *feedback laws* is addressed in §4 and §10.

(b) Of course, implicit in Theorem 2.1 is a result asserting the existence of a trajectory. One can derive local existence of the initial-value problem $\dot{x} \in F(x)$, $x(0) = x_0 \in \mathbb{R}^n$ in the absence of any prescribed set S by the following device: given x_0, and F satisfying (SH) *near* x_0, set $S = x_0 + \delta\overline{B}$, where $\overline{B}$ is the closed unit ball in $\mathbb{R}^n$ and redefine F on $bdry(S)$ in such a way as to ensure that F is upper semicontinuous on S and that $x \in bdry(S) \Longrightarrow 0 \in F(x)$. Then Theorem 2.1 applies to yield (locally) a trajectory for F beginning at x_0. See [2] [47] for a thorough discussion of existence theory, which is an issue of some depth.

(c) When S is compact, the global growth condition (c) of (SH) can be dropped. Note also that F need be defined (i.e., nonempty) only on S itself for purposes of the proof.

Remark 2.2. The type of invariance we have been discussing (which is also known as "viability" [1]) is "positive" or "forward" invariance; i.e., relative to *future* time t. A different concept is (weak) "backward" invariance or "preinvariance" of (S, F): the existence of a trajectory $x(\cdot)$ on $(-\infty, 0]$ satisfying $x(0) = x_0$ and whose values $x(t)$ lie in S for all t *prior* to 0. We can effect a time reversal of this type by considering $-F$ instead of F. The resulting characterization of weak backward invariance is the following:

$$H(x, \zeta) \geq 0 \quad \forall \zeta \in N_S^P(x), \quad \forall x \in S \tag{2.3}$$

where H is the upper Hamiltonian defined in §1.

Remark 2.3. For ease of exposition, we shall generally limit ourselves to the *autonomous* case $F(x)$ instead of that in which F has explicit t-dependence: $F(t, x)$. However, all our results extend rather easily to nonautonomous situations, via two different routes. The first and easier of the two is the familiar bookkeeping device of state augmentation: we consider $\tilde{x} = (t, x)$ to be the state, and we replace F by the (autonomous) multifunction $\widetilde{F}(\tilde{x}) = \{1\} \times F(t, x)$. To proceed, we require of course that $\widetilde{F}$ satisfy the relevant hypotheses. For the purposes of Theorem 2.1, for example, this amounts to $F(t, x)$ being upper semicontinuous in (t, x). Under this condition, the theorem carries over in evident fashion. It is also possible to allow the set S to depend on t, a line we shall not pursue here.

The second (nonreductionist) approach to considering nonautonomous problems allows time t to be treated differently from the state, in fact with less regular behavior. A familiar assumption in such a setting would be that $F(t, x)$ is $\mathcal{L} \times \mathcal{B}$ measurable in (t, x) (see [24]), which is certainly true if F is measurable in t and continuous in x. It is easy to adapt the proof of Theorem 2.1 to this setting (sequential compactness of approximate trajectories still holds), where the feedback $f(t, x)$ now depends on t too.

Let us just state a sample result in this nonautonomous setting. The system (S, F) is now called weakly invariant (on $(0, \infty)$, say) if for every $(t_0, x_0) \in (0, \infty) \times S$, there is a trajectory $x(\cdot)$ of F on $[t_0, \infty)$ such that $x(t_0) = x_0$ and such that $x(t) \in S \quad \forall t \geq t_0$. Then (S, F) is weakly invariant if

$$h(t, x, \zeta) \leq 0 \quad \forall \zeta \in N_S^P(x), \quad \forall t > 0, \quad \forall x \in S,$$

where $h(t, x, \zeta) := \min\big\{\langle v, \zeta\rangle : v \in F(t, x)\big\}$.

Definition 2.2. The *attainable set* $\mathcal{A}(\S_\prime; \tau)$ from x_0 at time τ is the set of all points of the form $x(\tau)$, where $x(\cdot)$ is a trajectory for F on $[0, \tau]$ satisfying $x(0) = x_0$.

We are now in position to give a result which subsumes for the first time the other types of criteria known to imply weak invariance. In its essentials, the proof is a simple consequence of Theorem 2.1. ("co" signifies "convex hull") The cone T_S^D of D-tangents to S was defined in §1.

Theorem 2.2. *The following are equivalent*:

1. $F(x) \cap T_S^D(x) \neq \varnothing \quad \forall x \in S$;
2. $F(x) \cap \operatorname{co} T_S^D(x) \neq \varnothing \quad \forall x \in S$;
3. $h(x,\zeta) \leq 0 \quad \forall \zeta \in N_S^P(x), \quad \forall x \in S$;
4. (S,F) *is weakly invariant*;
5. $\forall x \in S$, $\forall \varepsilon > 0$, $\exists \delta \in (0,\varepsilon)$ *such that* $\mathcal{A}(\S;\delta) \cap \mathcal{S} \neq \varnothing$.

Proof. That (1) implies (2) is a tautology, and (2) yields (3) because any element v of $\operatorname{co} T_S^D(x)$ lies in the polar of $N_S^P(x)$ (i.e., $\langle v, \zeta \rangle \leq 0 \quad \forall \zeta \in N_S^P(x)$), as is easily seen. The implication (3) $\Rightarrow$ (4) was proven in Theorem 2.1, and (4) $\Rightarrow$ (5) is a tautology. There remains to show that (5) $\Rightarrow$ (1). Let $x_i(\cdot)$ be a trajectory on $[0,\delta_i]$ with $x_i(0) = x$, $x_i(\delta_i) \in S$, where $\delta_i \downarrow 0$. For any $\varepsilon > 0$, for all i sufficiently large, we have

$$\frac{x_i(\delta_i) - x}{\delta_i} = \frac{1}{\delta_i} \int_0^{\delta_i} \dot{x}_i(t)\, dt \in F(x) + \varepsilon B,$$

since F is upper semicontinuous. The left side (at least along a subsequence) converges to an element of $T_S^D(x)$ by definition. Since ε is arbitrary, (1) follows. ☐

Remark 2.4. The criterion (1) has a long history of discovery and rediscovery beginning in the case of a function F with Nagumo, and continuing with Brézis, Crandall, Hartman, Martin, Yorke, and others (consult for example [1] [16] [54] [74], as well as [27] and its references.); it was stated first in the present context by Haddad [55]. The interesting refinement (2) is due to Ushakov [53]; we remark that a direct proof of the equivalence of (1) and (2) seems quite nontrivial. The criterion (3) introduced here has the advantage of requiring to be checked only at points x for which $N_S^P(x)$ is nontrivial, while still subsuming the others. It has the further property of giving rise to feedback laws via proximal aiming, as shown in the proof of Theorem 2.1.

Having explored the issue of being able to stay in a given set, can we characterize the property of having no choice but to remain?

3 Strong invariance

The system (S,F) is said to be *strongly invariant* provided that *all* trajectories $x(\cdot)$ whose initial value $x(0)$ lies in S remain in S thereafter: $x(t) \in S \quad \forall t \geq 0$. In contrast to weak invariance, note that consideration of this type of invariance presupposes the possibility of trajectories that exit from S, and hence logically requires that F be defined on a set strictly containing S.

In view of the previous section, it is perhaps natural to guess that strong invariance would be characterized by a condition such as $F(x) \subseteq T_S^D(x) \quad \forall x \in S$. This turns out to be false without additional hypotheses on F. A simple example in $\mathbb{R}$ is provided by $S = \{0\}$ and $F(x) = \{x^{1/3}\}$; $F(0)$ is contained in $T_S^D(0) = \{0\}$, yet *some* solutions of $\dot{x} = x^{1/3}$ beginning at 0 leave S.

In the following, T_S^C refers to the C-tangent cone (see §1). As ever, the Standing Hypothesis is in force.

Theorem 3.1. *Let F be locally Lipschitz. The following are equivalent*:

1. $F(x) \subseteq T_S^C(x) \quad \forall x \in S$;
2. $F(x) \subseteq T_S^D(x) \quad \forall x \in S$;
3. $F(x) \subseteq \operatorname{co} T_S^D(x) \quad \forall x \in S$;
4. $H(x,\zeta) \leq 0 \quad \forall \zeta \in N_S^P(x), \quad \forall x \in S$;
5. (S,F) *is strongly invariant*;
6. $\forall x \in S$, $\forall \delta > 0$ *sufficiently small*, $\mathcal{A}(\S;\delta) \subseteq \mathcal{S}$.

Proof. The implications (1) $\Rightarrow$ (2) $\Rightarrow$ (3) are tautologies, while (3) $\Rightarrow$ (4) is immediate from the fact that $\operatorname{co} T_S^D(x)$ lies in the polar of $N_S^P(x)$ (see the proof of Theorem 2.2). The equivalence of (5) and (6) is easy. That (5) implies (1) follows from nonsmooth calculus; the detailed argument dates from [16]. The missing link is therefore the implication (4) $\Rightarrow$ (5), which we proceed to prove. Let $y(\cdot)$ be a trajectory with $y(0) =: x_0 \in S$. It suffices to prove that $y(t)$ lies in S for $t \leq T$, for some $T > 0$. Consider the multifunction

$$\widetilde{F}(t,x) := \Big\{ v \in F(x) \colon \big|v - \dot{y}(t)\big| \leq K\big|x - y(t)\big| \Big\},$$

where K is the Lipschitz rank of F.

Then, in view of the Lipschitz hypothesis, $\widetilde{F}$ is nonempty valued on S at least for $t \leq T$, if $T > 0$ is small enough, as well as having convex compact values. It is easy to verify as well that $\widetilde{F}$ is $\mathcal{L}\times\mathcal{B}$ measurable, and upper semicontinuous in x, and that the system $(S, \widetilde{F})$ inherits condition (2.1) from (4), since the lower Hamiltonian $\tilde{h}$ for $\widetilde{F}$ satisfies $\tilde{h} \leq H$. Thus the nonautonomous version of Theorem 2.1 (see Remark 2.3) yields the weak invariance of $(S, \widetilde{F})$. Therefore $\widetilde{F}$ admits a trajectory x on $[0, T]$ with $x(0) = x_0$ remaining in S. But x satisfies

$$\big|\dot{x}(t) - \dot{y}(t)\big| \leq k\big|x(t) - y(t)\big|, \quad x(0) = y(0),$$

so that by the Gronwall inequality $x(\cdot)$ and $y(\cdot)$ coincide. □

Remark 3.1. We pause to comment on the role of T_S^C.

(a) It is false that one can add to the list of equivalences in the weak invariance Theorem 2.2 the condition "$T_S^C(x) \cap F(x) \neq \varnothing \quad \forall x \in S$", even when F is Lipschitz. A counterexample is given in [37]. However, the property "$\operatorname{int} T_S^C(x) \cap F(x) \neq \varnothing \quad \forall x \in S$" turns out to be relevant to the issue of feedback construction within S (see Remark 2.1(a)). Since the multifunction $T_S^C(\cdot)$ is lower semicontinuous on S when $\operatorname{int} T_S^C(x)$ is nonempty on S, the Michael selection theorem in this case yields the existence of a continuous selection f of F on S for which (S, f) is weakly invariant. Furthermore, (see §4) we can go on to show that when F is Lipschitz there is a *Lipschitz* selection f of F such that (S, f) is (strongly) invariant. This conclusion is false if $F(x)$ intersects merely $T_S^C(x)$; we do require the interior of $T_S^C(x)$.

(b) A closed subset Ω of $\mathbb{R}^n$ is called *weakly avoidable* if for all $x_0 \notin \operatorname{int}\Omega$, there is a trajectory x of F with $x(0) = x_0$ such that $x(t) \notin \operatorname{int}\Omega \quad \forall t \geq 0$. Let Ω satisfy the following property, for all $x \in \Omega$,

$$\operatorname{int} T_\Omega^C(x) \neq 0 \;\text{ and }\; T_\Omega^C(x) = T_\Omega^D(x)$$

(the latter property asserts that Ω is "regular" in the sense of nonsmooth analysis). Then a necessary and sufficient condition for the weak avoidability of Ω is

$$\big\{-F(x)\big\} \cap T_X^C(x) \neq \varnothing \quad \forall x \in \Omega.$$

This follows from our earlier weak invariance results (applied to $S = \operatorname{cl}(\operatorname{comp}(\Omega))$) upon noting the formula [37] [38] $T_S^C(x) = -T_\Omega^C(x)$.

(c) A further important role for T_S^C occurs in the consideration of equilibria of F in weakly invariant sets; see the next section.

A (weak or strong) invariant set S resembles an equilibrium point of a differential equation, in the sense that it corresponds to states at which the flow can remain ... is there some more precise relationship between equilibria and invariance?

4 Equilibria

There is a classical connection between flow invariance and the existence of equilibria, which we now recall. Let $f \colon \mathbb{R}^n \to \mathbb{R}^n$ be a locally Lipschitz function, and consider the Cauchy problem

$$\dot{x}(t) = f\big(x(t)\big), \quad x(0) = x_0. \tag{4.1}$$

Suppose that S is flow invariant; that is, the (unique) solution $x(\cdot)$ of (4.1) has values $x(t) \in S \quad \forall t \geq 0$ whenever $x_0 \in S$. Does it follow that S contains an equilibrium? (i.e., a point $x^* \in S$ such that $f(x^*) = 0$.)

The answer is affirmative if S has certain topological properties, for example when S is homeomorphic to a compact convex set K in $\mathbb{R}^m$. This means that for some continuous function $h\colon K \to S$ having continuous inverse, one has $h(K) = S$.

The proof of this well-known result is an application of the Brouwer fixed point theorem (see for example [10]). Indeed, it is equivalent to it, and we shall sketch the derivation of Brouwer's Theorem from the "invariance implies equilibrium" result to motivate later results.

It is sufficient to prove Brouwer's Theorem for the case of a Lipschitz mapping g from the closed ball B to itself; we seek a fixed point of g. Consider the differential equation $\dot{x} = f(x)$, $x(0) = x_0$, where $f(x) := g(x) - x$. Suppose that f fails to admit an equilibrium in B. We claim then that B is flow invariant.

For let $x(\cdot)$ solve the Cauchy problem. If and when $|x(t)| = 1$, we have

$$\frac{d}{dt}\left|x(t)\right|^2 = \left\langle 2x, g(x) - x \right\rangle < 0,$$

so that $x(t)$ cannot leave B. This establishes invariance, and it follows that f has an equilibrium in B. But this corresponds to a fixed point for g, as required.

Let us now turn to the case of multifunctions, where the well-known extension of Brouwer's Theorem is due to Kakutani: let S be compact convex, and let $G\colon S \rightrightarrows S$ have compact convex nonempty values and be upper semicontinuous; then G has a fixed point in S (i.e., a point x^* such that $x^* \in G(x^*)$).

It turns out that a dynamic version of this result (for equilibria) exists, though it may not have been viewed in this light. We are referring to a theorem of Browder [11] which also applies to an upper semicontinuous multifunction F defined on a compact convex set S. The theorem asserts that provided the following tangency condition holds

$$F(x) \cap T_S(x) \neq \varnothing \quad \forall x \in S, \tag{4.2}$$

then F admits an equilibrium in S (i.e., a point $x^* \in S$ such that $0 \in F(x^*)$). Here, $T_S(x)$ refers to the tangent cone of a convex set; we have $T_S(x) = T_S^D(x) = T_S^C(x)$, in the notation of earlier sections.

The reason we refer to Browder's Theorem as the dynamic version of Kakutani's Theorem is that we recognize (via Theorem 2.2) the condition (4.2) as a criterion giving weak invariance of the system (S, F). Nonetheless, no dynamic considerations are involved in Browder's proof. Incidentally, deriving Kakutani's Theorem from Browder's is simple: just note

$$F(x) := G(x) - x \subseteq S - x \subseteq T_S(x),$$

and apply Browder's Theorem to F.

Let us now emphasize one distinction between the single-valued case (Brouwer and its dynamic analogue) and the multi-valued one (Kakutani, Browder): the latter results are confined to *convex* sets S, in contrast to the former, where S need only be homeomorphic to a compact convex set (say). Now, a very natural conjecture to make is that when S has this property, then the weak invariance (viability) of (S, F) implies that the system (S, F) admits an equilibrium (i.e., a point $x \in S$ such that $0 \in F(x)$). This turns out to be false, even when F is Lipschitz; a first example is given in [37].

One way to view this state of affairs is to conclude that the tangency criterion

$$F(x) \cap T_S^D(x) \neq \varnothing \quad \forall x \in S,$$

while precisely right to characterize weak invariance, is not strong enough to force F to have an equilibrium. It turns out that the tangent cone T_S^C is highly useful here, however, as illustrated by the following new result from [37]:

Theorem 4.1. *Let S satisfy*

(i) S is homeomorphic to a compact convex set;

(ii) $\operatorname{int} T_S^C(x) \neq \varnothing \quad \forall x \in S$.

Suppose that F satisfies the Standing Hypothesis together with

(iii) $F(x) \cap T_S^C(x) \neq \varnothing \quad \forall x \in S$.

Then the system (S, F) admits an equilibrium.

As shown in [37], none of (i), (ii) or (iii) can be deleted for the conclusion to hold; nor, as stated above, can T_S^C be replaced by T_S^D. We remark that Theorem 4.1 subsumes the theorem of Browder cited earlier (the convex case).

There is a close link between condition (iii) of the theorem and the existence of "invariant feedback" for (S, F), one that can be used to prove the theorem. We are referring to the following fact:

Lemma 4.1. *Let F be locally Lipschitz, and suppose*

$$F(x) \cap \operatorname{int} T_S^C(x) \neq \varnothing \quad \forall x \in S, \tag{4.3}$$

where S is compact. Then there exists a Lipschitz function f on S such that

$$f(x) \in F(x) \cap \operatorname{int} T_S^C(x) \quad \forall x \in S.$$

We sketch a proof. Let us admit for $F(x)$ a representation of the form $\varphi(x, U)$, where φ is continuous, Lipschitz in x, and U is a compact set. For each $x \in S$ there exists by hypothesis a point u_x in U such that

$$\varphi(x, u_x) \in \operatorname{int} T_S^C(x).$$

When $\operatorname{int} T_S^C(x) \neq \varnothing$, as is the case here, it is known that $T_S^C(\cdot)$ is lower semicontinuous at x. We deduce the existence of a neighborhood $N(x)$ such that

$$\varphi(y, u_x) \in \operatorname{int} T_S^C(y) \quad \forall y \in N(x).$$

Let $N(x_i)$ $(i = 1, 2, \ldots, N)$ be a finite subcovering of S, and let $\rho_i(\cdot)$ be a locally Lipschitz partition of unity associated with it. Set

$$f(y) := \sum_{i=1}^{N} \rho_i(y) \varphi(y, u_{x_i}).$$

Then f is Lipschitz on S by construction. Furthermore, $f(y) \in F(y)$ by the convexity of the set $F(y)$. Finally, if $\rho_i(y) \neq 0$ then $y \in N(x_i)$, whence

$$\varphi(y, u_{x_i}) \in \operatorname{int} T_S^C(y).$$

This implies that $f(y) \in \operatorname{int} T_S^C(y)$ by the convexity of T_S^C and completes the proof of the Lemma.

Given this, the proof of Theorem 4.1 is at hand. First, reduce the theorem by standard upper appoximation to the case in which F is Lipschitz and satisfies (4.3). Next, apply the Lemma to deduce the existence of f. Finally, observe that

$$f(x) \in \operatorname{int} T_S^C(x) \Rightarrow f(x) \in T_S^D(x) \Rightarrow (S, f) \text{ invariant},$$

so that f has a zero in S by the classical result cited earlier. But then (S, F) has an equilibrium, as required.

Remark 4.1. The set S of Theorem 4.1 is compact and, as a result of hypothesis (ii), has nonempty interior. This indicates an intrinsically finite-dimensional nature to the approach. However, a different method [37] involving more nonsmooth analysis yields the following:

Theorem 4.2. *Let S lie in a Banach space and be lipeomorphic to a compact convex set. If F satisfies* (iii), *then (S, F) admits an equilibrium.*

We refer to [37] for the proof, and for further developments. Related results in the finite dimensional case were obtained in [44] by topological methods.

Having characterized the possibility of remaining in a given set, it is natural to ask: how can we recognize that nearby points lying outside the set can be steered to it?

5 Local attainability of a set

The principal purpose of this section is to establish conditions which guarantee that all points in a neighborhood of a given set S can be steered via a trajectory to S in finite time. That is, the question is: given $x_0 \notin S$, when is there $\tau > 0$ such that $\mathcal{A}(\S_\prime; \tau) \cap \mathcal{S} \neq \varnothing$?

For simplicity, we assume for the remainder of this article that S is compact. Let us now postulate the following strengthened proximal normality criterion (cf. (2.1)): for some $\delta \geq 0$,

$$h(x, \zeta) \leq -\delta |\zeta| \quad \forall \zeta \in N_S^P(x), \quad \forall x \in S. \tag{5.1}$$

The case $\delta = 0$ reduces to (2.1) considered earlier.

The following "Lemma of the Alternative" lies at the heart of an algorithmic approach to attainability which seems to yield the best convergence rate. Its proof relies on a new Mean Value Inequality [28] and on a characterization of proximal subgradients of distance functions [40].

Lemma 5.1. *Let U be a set containing S, and let F be Lipschitz with constant K on U, and let $|v| \leq M$ for all $v \in F(x)$, $\forall x \in U$. Then, for any $t > 0$, and for any $u \in U$, for any $\varepsilon > 0$, there exists $v \in F(u)$ such that one of the two following conclusions holds:*

(a) $d_S(u + tv) \leq (Kt + 1)d_S(u) + 2KMt^2 - \delta t$; or

(b) for some $\tau \in [0, t]$, $u + \tau v \in S + \varepsilon B$.

Further, if $\delta = 0$, then (a) always holds

Proof. The Mean Value Inequality [28] (see also §1) applied to the function d_S, to the point u, and to the set $u + tF(u)$ gives the existence of a point z in $\big[u, u + tF(u)\big] + \varepsilon B$ and a vector $\zeta \in \partial_P d_S(z)$ such that the following inequality holds for all $w \in F(u)$:

$$\min_{v \in F(u)} d_S(u + tv) - d_S(u) < \langle \zeta, tw \rangle + \varepsilon, \tag{5.2}$$

and such that

$$d_S(z) \leq d_S(u) + tM + \varepsilon. \tag{5.3}$$

Suppose first that z belongs to S. Since we know that z is of the form $u + \tau v$ for some $v \in F(u)$ and $\tau \in [0, t]$, we immediately deduce alternative (b) of the Lemma. So now let us suppose that $z \notin S$, and see how to obtain (a).

When $z \notin S$, it follows from results in [40] that ζ must have the form $(z - y)/d_S(z)$, where y is the closest point in S to z. Consequently $|\zeta| = 1$ and $\zeta \in N_S^P(y)$. Now condition (5.1) applies to give the existence of some $\overline{v}$ in $F(y)$ such that

$$\langle \zeta, \overline{v} \rangle \leq -\delta. \tag{5.4}$$

Invoking the Lipschitz property for F, we derive the existence of some $w \in F(u)$ so that

$$|w - \overline{v}| \leq K|y - u|. \tag{5.5}$$

Returning to (5.2) armed with precisely this w, we get

$$\begin{aligned}
\min_{v \in F(u)} d_S(u + tv) - d_S(u) &< \langle \zeta, tw \rangle + \varepsilon \\
&= t\langle \zeta, w - \overline{w} \rangle + t\langle \zeta, \overline{v} \rangle + \varepsilon \\
&\leq t|\zeta|\,|w - \overline{v}| - \delta t + \varepsilon \\
&\leq Kt|y - u| - \delta t + \varepsilon \qquad \text{(by (5.5))} \\
&\leq Kt\big\{|y - z| + |z - u|\big\} - \delta t + \varepsilon \\
&\leq Kt\big\{d_S(z) + tM + \varepsilon\big\} - \delta t + \varepsilon \\
&\leq Kt\big\{d_S(u) + 2tM + 2\varepsilon\big\} - \delta t + \varepsilon \qquad \text{(by (5.2))}.
\end{aligned}$$

Since ε is arbitrary, this implies alternative (a). □

The point of alternative (a) of the lemma is that for $\delta > 0$, it will give an iterative step in which d_S is decreased; on the other hand, alternative (b) will correspond to termination: we are almost in S. We now illustrate the use of Lemma 5.1 in designing algorithms that produce trajectories attaining S.

Theorem 5.1. *Let the hypotheses of Lemma 5.1 hold, with $\delta > 0$. Then for all x_0 in a neighborhood of S, there is a trajectory $x(\cdot)$ of F such that $x(0) = x_0$ and such that for some $s \geq 0$ one has $x(s) \in S$.*

Proof. Let us choose $\delta_0 > 0$ such that the M and K of Lemma 5.1 are valid on $U := S + \delta_0 B$, and sufficiently small so that

$$\min\{e^{Kt}\delta_0 - \delta t \colon t \geq 0\} < 0.$$

We proceed to take any $t^* \geq 0$ such that $e^{Kt^*}\delta_0 - \delta t^* < 0$. We shall now establish the assertion of the theorem for any x_0 such that $d_S(x_0) < \gamma\delta_0 e^{-Kt^*}$, where $\gamma \in (0, 1)$, where the time s in the statement of the theorem satisfies $s \leq t^*$.

Let N be a positive integer and set $t = t^*/N$. We further choose N large enough to ensure

$$\lambda := \delta t - 2KMt^2 > 0, \qquad \text{as well as}$$

$$\gamma\delta_0 + \frac{KMt^*}{N} < \delta_0, \qquad \text{and} \tag{5.6}$$

$$\delta_0 - \delta t^* + \frac{2KM(t^*)^2}{N} < 0. \tag{5.7}$$

We proceed to define a piecewise-linear function x^N such that $x^N(0) = x_0$ and $x^N(t_N) \in S$ for some $t_N \in [0, t^*]$. The values x_i of $x^N(it)$ are determined recursively as follows:

We apply Lemma 5.1 to obtain $v_0 \in F(x_0)$ such that either

$$\begin{aligned} d_S(x_0 + tv_0) &\leq (Kt+1)d_S(x_0) + 2KMt^2 - \delta t \\ &= (Kt+1)d_S(x_0) - \lambda \end{aligned} \tag{5.8}$$

or

$$x_0 + \tau v_0 \in S + \frac{1}{N}B \text{ for some } \tau \in [0, t]. \tag{5.9}$$

In the latter case (called "termination"), we set $t_N = \tau$ and $x_1 = x_0 + \tau v_0$. Otherwise, we set $x_1 = x_0 + tv_0$. Observe that by (5.8),

$$d_S(x_1) \leq e^{Kt} d_S(x_0) = e^{Kt^*/N} d_S(x_0) < \gamma\delta_0,$$

so that x_1 lies in U. This permits us to iterate the procedure; in general terms we find $v_i \in F(x_i)$ and set $x_{i+1} := x_i + tv_i$ as long as the first alternative continues to hold. From

$$d_S(x_{i+1}) \leq (Kt+1)d_S(x_i) - \lambda$$

we derive

$$d_S(x_i) \leq (Kt+1)^i d_S(x_0) - i\lambda \qquad (i = 1, 2, \ldots). \tag{5.10}$$

This last inequality implies $d_S(x_i) < \gamma\delta_0$ as long as $i \leq N$, and hence all such x_i belong to U for $i \leq N$, justifying the recursion to that point. The process terminates as soon as the second alternative holds, and the first index i for which this occurs is labeled J. We claim that $J < N$. For if (5.8) holds for $i = N$, then

$$\begin{aligned} d_S(x_i) &\leq \left(\frac{Kt^*}{N} + 1\right)^N d_S(x_0) - i\lambda \\ &< e^{Kt^*} d_S(x_0) - N\{\delta t - 2KMt^2\} \\ &< \delta_0 - \delta t^* + \frac{2KM(t^*)^2}{N}, \end{aligned}$$

but this is negative by (5.7), a contradiction since $d_S(x_i) \geq 0$. This confirms that $J < N$; i.e., that termination of the process must occur. Thus $x_J = x_{J-1} + \tau v_{J-1} \in S + 1/NB$ for some $\tau \in (0, t]$ and $v_{J-1} \in F(x_{J-1})$. The resulting function $x^N(s)$ is defined for time $0 \leq s \leq (J-1)t + \tau =: t_N^*$ by setting

$$\begin{aligned} x^N(it) &= x_i \qquad \text{for } i = 0, 1, \ldots, J-1 \\ x^N(t_N^*) &= x_J, \end{aligned}$$

and by requiring that x^N be linear between the given values. Note that $t_N^* = (J-1)t + \tau < (N-1)t + t = Nt = t^*$, and that $x^N(0) = x_0$, $x^N(t_N) \in S + 1/NB$, and also that each x^N is Lipschitz of rank M.

The sequence $\{x^N\}$ of functions admits a subsequence converging uniformly. Also, the x^N are "approximate trajectories" of F. For example, when s lies in $[it, (i+1)t]$ for some $i = 0, 1, \ldots, J-2$, then

$$\begin{aligned} d_S\big(x^N(s)\big) &\leq d_S(x_i) + KtM \\ &< \gamma\delta_0 + \frac{KMt^*}{N} < \delta_0 \qquad \text{by (5.6)}. \end{aligned}$$

Thus $x^N(s) \in U$, and we deduce

$$\begin{aligned} \dot{x}^N(s) = v_i &\in F(x_i) \\ &\subseteq F\big(x^N(s)\big) + K\big|x^N(s) - x_i\big| \\ &\subseteq F\big(x^N(s)\big) + \frac{KMt^*}{N}. \end{aligned}$$

A straightforward application of the sequential compactness theorem for differential inclusions [24] yields that an appropriate subsequence of $\{x^N\}$ converges to a trajectory $x(\cdot)$ for F which begins at x_0 and whose value $x(s)$ at some $s \le t^*$ lies in S. $\square$

In keeping with the results of prior sections, we now show that the proximal normal hypothesis of the theorem is implied by a tangential requirement in terms of T_S^C.

Proposition 5.1. *Let F be locally Lipschitz, and suppose*

$$F(x) \cap \operatorname{int}\big[T_S^C(x)\big] \neq \varnothing \quad \forall x \in S.$$

Then there exists $\delta > 0$ for which (5.1) holds. (Thus the conclusion of Theorem 5.1 holds.)

Proof. It suffices to produce $\delta > 0$ such that for any given $x \in S$, there exists $v \in F(x)$ such that $v + \delta B \subseteq T_S^C(x)$. It is readily seen that then for every $x \in S$, there exists $v \in F(x)$ for which

$$\langle \zeta, v\rangle \le -\delta|\zeta| \quad \forall \zeta \in N_S^P(x)$$

which is (5.1).

By assumption, we can associate with each $x \in S$ a number $\varepsilon(x) > 0$ and a vector $v(x) \in F(x)$ such that $v(x) + \varepsilon(x)B \subseteq T_S^C(x)$. Due to the lower semicontinuity property of the tangent cone $T_S^C(x)$, one has

$$T_S^C(x) \subseteq T_S^C(x') + \frac{\varepsilon(x)}{2}$$

for all $x' \in S$ sufficiently near x. Hence there exists an open ball $U(x)$ around x such that

$$x' \in U(x) \Longrightarrow v(x) + \frac{\varepsilon(x)}{2}B \subseteq T_S^C(x').$$

We further insist that the radius of $U(x)$ be less than $\varepsilon(x)/4K$, where K is the Lipschitz rank of F on U. Now, the family of balls $\big\{U(x)\big\}_{x\in S}$ constitutes an open cover of S, and by the compactness of S, there exists a finite subcover $\big\{U(x_1), U(x_2), \ldots, U(x_k)\big\}$. Define

$$\delta = \min\left\{\frac{\varepsilon(x_i)}{4} : i = 1, 2, \ldots, k\right\}.$$

We claim that this δ has the required property. In order to see why, let $x \in S$. Then $x \in U(x_i)$ for some i, and so

$$\|x - x_i\| < \frac{\varepsilon(x_i)}{4K}.$$

Furthermore, there exists $v \in F(x)$ such that

$$\big\|v - v(x_i)\big\| \le K\|x - x_i\| < \frac{\varepsilon(x_i)}{4}.$$

By our construction, we also have

$$v(x_i) + \frac{\varepsilon(x_i)}{2}B \subseteq T_S^C(x).$$

Hence $v + \delta B \subseteq T_S^C(x)$, which completes the proof. $\square$

Remark 5.1. (a) A recurrent theme in the article involves obtaining strong versions of invariance-type results by replacing h by H, as in Theorems 2.1 and 3.1; it arises here as well. For example, if (5.1) is replaced by

$$H(x,\zeta) \le -\delta|\zeta| \quad \forall \zeta \in N_S^P(x), \quad \forall x \in S, \tag{5.11}$$

where $\delta > 0$, then it can be shown [36] that *all* trajectories beginning sufficiently near S enter S in finite time. Likewise, a sufficient condition for this property is

$$F(x) \subseteq \operatorname{int}\bigl[T_S^C(x)\bigr] \quad \forall x \in S$$

(cf. Prop. 5.1).

(b) A careful analysis of the proof algorithm allows us to be specific about the size of the neighborhood whose existence is asserted in Theorem 5.1, and also to give a "rate of convergence" to S. Specifically, let the bound M and the Lipschitz constant K pertain to the neighborhood $S + \delta_0 B$, where δ_0 is small enough to ensure

$$\frac{\delta}{2} - 2K\delta_0 =: c > 0.$$

Then [36] we have the following estimate for the attaining time t^*, for any x_0 in $S + \delta_0 B$:

$$t^* \le \frac{d_S(x_0)}{c}.$$

Further, for the attaining trajectory $x(\cdot)$, we have

$$d_S\bigl(x(t)\bigr) \le d_S\bigl(x(s)\bigr) - (t-s)c, \quad 0 \le s < t \le t^*,$$

so that $x(\cdot)$ approaches S at a prescribed rate c, as measured by the distance function. As shown in [86], estimates of this type can be used to show that the minimal time function is Lipschitz.

(c) As shown in [36], the developments of this section can be extended to a continuous (non-Lipschitz) multifunction; even the merely upper semicontinuous case can be treated, modulo a uniform type of hypothesis *near* (rather than on) the boundary of S.

(d) Consider now the issue of reaching the *interior* of the set S along some trajectory, which appears to be a new kind of consideration. Of course, in order to be able to do so, this interior must first of all exist (i.e., be nonempty). A natural conjecture would be the following: if the hypotheses of Theorem 5.1 are satisfied, and if $\operatorname{int} S \neq \varnothing$, with $S = \operatorname{cl}(\operatorname{int} S)$, then for all points x_0 near S, there is a trajectory from x_0 which enters the interior of S. This turns out to be false, but it is possible to give extra conditions on S which rescue the statement. One of these, familiar by now, is that $\operatorname{int} T_S^C(x) \neq \varnothing$ for $x \in S$; see [36] for details.

Local attainability of a set can be viewed as a type of stability property. Is there a connection with Lyapounov's celebrated approach to the stability of equilibria?

6 Lyapounov stability of invariant sets

The concept of weakly invariant set for a differential inclusion can be thought of as an extension of that of equilibrium point for a differential equation: each corresponds to states that can be maintained. This analogy can be made precise in the context of stability considerations when treated by the classical Lyapounov approach. First, a definition:

Definition 6.1. The system (S,F) is locally weakly stable if for all $\varepsilon > 0$ there exists $\delta > 0$ such that, for any point x_0 within distance δ of S (i.e., $d_S(x_0) < \delta$), there exists a trajectory $x(\cdot)$ with $x(0) = x_0$ and $d_S\bigl(x(t)\bigr) \le \varepsilon \quad \forall t \ge 0$.

When S is compact and F satisfies the Standing Hypothesis, it follows readily (from sequential compactness of trajectories) that the local weak stability of (S,F) implies the weak invariance of (S,F); thus only weakly invariant systems are candidates for stability. The following gives a criterion for this unorthodox set stability in otherwise very familiar terms: the function φ below is a *Lyapounov function*.

Theorem 6.1. *In order that S be locally weakly stable, it is necessary and sufficient that there exist a lower semicontinuous function $\varphi\colon \mathbb{R}^n \to [0,\infty]$ satisfying:*

(A1) $\forall \varepsilon > 0 \quad \exists \delta > 0$ *such that* $S + \delta B \subseteq \bigl\{x\colon \varphi(x) \le \varepsilon\bigr\}$;

(A2) $\forall \varepsilon > 0 \quad \exists \delta > 0$ *such that* $\{x \colon \varphi(x) \leq \delta\} \subseteq S + \varepsilon B$*;*

(B) For some $\delta_0 > 0$*, for all* $x_0 \in S + \delta_0 B$*, there is at least one trajectory* x *satisfying* $x(0) = x_0$ *and*

$$\varphi\big(x(t)\big) \leq \varphi(x_0), \quad t \geq 0.$$

Remark 6.1. (A1) and (A2) together imply that $S = \{x \colon \varphi(x) = 0\}$, but say more than this in requiring that the level sets of φ converge to S appropriately.

Proof of the theorem. Suppose first that such a function φ exists; let us confirm the stability of (S, F). Let any $\varepsilon > 0$ be given, and let δ correspond to it as in (A2). Now invoke (A1) to obtain $\delta' < \delta_0$ such that $d_S(x_0) < \delta'$ implies $\varphi(x_0) \leq \delta$. Finally, we note that by (B), for any such x_0, there is a trajectory $x(\cdot)$ beginning at x_0 such that $\varphi\big(x(t)\big) \leq \varphi(x_0) \leq \delta$. But then $x(t)$ lies in $S + \varepsilon B$ for $t \geq 0$, as required.

Now let (S, F) be locally weakly stable; we proceed to exhibit a Lyapounov function φ. There exist ε_1, $\delta_1 > 0$ such that given x_0 in $S + \delta_1 B$, some trajectory $x(\cdot)$ from x_0 satisfies $d_S\big(x(t)\big) \leq \varepsilon_1$ $(t \geq 0)$. Let us define

$$\varphi(\alpha) := \min\left\{\delta_1, \inf_{x(0)=\alpha}\left\{\sup_{t\geq 0} d_S\big(x(t)\big)\right\}\right\},$$

where the inf is taken over all trajectories x satisfying $x(0) = \alpha$. An argument involving sequential compactness shows that if $\varphi(\alpha) < \delta_1$, then there is a trajectory x at which the infimum above is attained. Armed with this, one shows that φ is lower semicontinuous.

The property (A1) is an immediate consequence of the definition φ, given that (S, F) is stable. We turn now to (A2). Let $\varepsilon > 0$ be given, with $\varepsilon < \delta_1$. If α satisfies $\varphi(\alpha) \leq \varepsilon$, then there is a trajectory $x(\cdot)$ with $x(0) = \alpha$ such that $d_S\big(x(t)\big) \leq \varepsilon$, $t \geq 0$. Therefore $d_S\big(x(0)\big) = d_S(\alpha) \leq \varepsilon$, which shows

$$\{\alpha \colon \varphi(\alpha) \leq \varepsilon\} \subseteq S + \varepsilon B,$$

confirming (A2). □

Remark 6.2. (a) It is possible to extend other kinds of Lyapounov stability from equilibrium points to invariant sets as well, notably the asymptotic variety. We refer to the discussion in Deimling [47] for the equilibrium point context, which inspired the results of this section. See also related results in [49] [75].

(b) It is possible in Theorem 6.1 to modify (B) by requiring that the trajectory $x(\cdot)$ be such that $t \to \varphi\big(x(t)\big)$ be nonincreasing, a somewhat more familiar form of the Lyapounov criterion stemming from "energy function" considerations. This requires no essential modifications in the proof. To handle the issue of "strong local stability", in which *all* trajectories starting near S must remain near S, we remark that it suffices to modify (B) by requiring that all trajectories from x_0 satisfy the stated inequality.

The property that a function φ be decreasing along some (or all) trajectories is central to Lyapounov stability; is there a differential characterization of this property when φ is merely lower-semicontinuous?

7 Monotonicity along trajectories

We saw in the preceding section, in connection with Lyapounov stability, the relevance of having functions φ which are monotone along (certain) trajectories. It turns out that this property is central to other issues as well, so much so that we shall name it, and require subdifferential criteria to recognize it. Let φ be a function mapping $\mathbb{R}^n$ to $(-\infty, \infty]$.

Definition 7.1. The system (φ, F) is weakly decreasing provided that for any α in $\mathbb{R}^n$, there exists a trajectory $x(\cdot)$ with $x(0) = \alpha$ such that $\varphi\big(x(t)\big) \leq \varphi(\alpha) \quad \forall t \geq 0$.

We recall the Standing Hypothesis (SH) on F, and proceed to:

Theorem 7.1. *Let* φ *be lower semicontinuous. Then* (φ, F) *is weakly decreasing iff*

$$h\big(x, \partial_P\varphi(x)\big) \leq 0 \quad \forall x \in \mathbb{R}^n. \tag{7.1}$$

(The condition (7.1) in the statement of the theorem requires $h(x, \zeta) \leq 0 \quad \forall \zeta \in \partial_P\varphi(x)$, $\forall x \in \mathbb{R}^n$. Note that the inequality is automatically true when $\partial_P\varphi(x)$ is empty: there is nothing to check at such points.)

Proof. Let

$$\begin{aligned} S &:= \operatorname{epi}(\varphi) \\ &= \big\{(x,r) \in X \times \mathbb{R} \colon \varphi(x) \le r\big\}, \end{aligned}$$

which is closed because φ is lower semicontinuous, and let $\widetilde{F}(x,r)$ be defined as $F(x) \times \{0\}$. Then it follows from the definitions that (φ, F) is weakly decreasing iff $(S, \widetilde{F})$ is weakly invariant. According to Theorem 2.2, the system $(S, \widetilde{F})$ is weakly invariant iff for every (x,r) in S, for every $(\zeta, \sigma) \in N_S^P(x,r)$, one has

$$h_{\widetilde{F}}(x,r,\zeta,\sigma) = h(x,\zeta) \le 0. \tag{7.2}$$

We need only to prove therefore that this last condition is equivalent to (7.1).

First, let (7.2) hold for all $(\zeta,\sigma) \in N_S^P(x,r)$. Consider $\zeta \in \partial_P \varphi(x)$. Then $(\zeta, -1) \in N_S^P\big(x, \varphi(x)\big)$, so that, by (7.2), we have $h(x,\zeta) \le 0$; i.e. (7.1) follows.

Now let (7.1) hold, and consider any (ζ,σ) in $N_S^P(x,r)$. Then (ζ,σ) belongs to $N_S^P\big(x,\varphi(x)\big)$, as is easily shown, and $\sigma \le 0$ necessarily. If $\sigma \ne 0$ then $-\zeta/\sigma$ belongs to $\partial_P\varphi(x)$, whence $h(x, -\zeta/\sigma) \le 0$. Since h is positively homogeneous in its last variable, we get $h(x,\zeta) \le 0$, i.e. (7.2). If σ is zero, it is known [38] that $(\zeta, 0)$ can be approximated to an arbitrary tolerance by (ζ', σ') where $\sigma' < 0$ and (ζ', σ') belongs to $N_S^P\big(x', \varphi(x')\big)$, for x' arbitrarily near x. Then, as above, we have $h(x', \zeta') \le 0$. Since h is lower semicontinuous, we deduce $h(x,\zeta) \le 0$; that is, (7.2) holds. ☐

Remark 7.1. (a) When φ is continuous, (φ, F) is weakly decreasing iff there is a trajectory $x(\cdot)$ for each α as in Definition 7.1 such that the function $t \to \varphi\big(x(t)\big)$ is decreasing (not necessarily strictly), which helps justify our terminology. In the lower semicontinuous case, this basic question appears to be open.

(b) When φ is the indicator of a set S, then (φ, F) is weakly decreasing iff (S, F) is weakly invariant.

We now examine a strong version of the monotonicity property.

Definition 7.2. The system (φ, F) is strongly decreasing provided that for any trajectory $x(\cdot)$ on $[0,\infty)$, we have $\varphi\big(x(t)\big) \le \varphi\big(x(0)\big)$ for $t \ge 0$; equivalently, provided that $t \to \varphi\big(x(t)\big)$ is decreasing whenever $x(\cdot)$ is a trajectory.

Theorem 7.2. *Let φ be lower semicontinuous, and let F be locally Lipschitz. Then (φ, F) is strongly decreasing iff*

$$H\big(x, \partial_P\varphi(x)\big) \le 0 \quad \forall x. \tag{7.3}$$

We omit the proof, since it follows very closely that of Theorem 7.1, with an appeal to Theorem 3.1 instead of Theorem 2.1. Note that the Lipschitz property of F is essential here, as in §3.

The following question now arises: what property (if any) corresponds to the other two Hamiltonian inequalities that can be paired with (7.1) and (7.3)? For example, consider

$$H\big(x, \partial_P\varphi(x)\big) \ge 0 \quad \forall x. \tag{7.4}$$

By carrying out a time-reversal (replacing F by $-F$), it can be deduced as a corollary of Theorem 7.1 that (7.4) characterizes the following property: for all x_0, there exists a trajectory $x(\cdot)$ on $(-\infty, 0]$ with $x(0) = x_0$ such that

$$\varphi\big(x(t)\big) \le \varphi(x_0), \quad t < 0.$$

(When φ is continuous, it is equivalent to require that $t \to \varphi\big(x(t)\big)$ be increasing on $(-\infty, 0]$.) We call the system *weakly preincreasing* in this case, the prefix "pre" serving to indicate prior time: φ increases to value $\varphi(x_0)$ on a previous part of some trajectory through x_0. Of course, when φ is the indicator function of a set S, the preceding is entirely consistent with what we called "weak preinvariance" in Remark 2.2.

The fourth and last Hamiltonian inequality to be considered is

$$h\big(x, \partial_P\varphi(x)\big) \ge 0 \quad \forall x. \tag{7.5}$$

This turns out (when F is Lipschitz) to characterize (φ, F) being *strongly increasing*: $t \to \varphi\big(x(t)\big)$ is increasing on $[0,\infty)$ for any trajectory x. Again, time-reversal will serve to prove this fact, together with the observation that "strongly increasing" and "strongly preincreasing" (to use an undefined but transparent term) are identical properties. We caution the reader that the properties are actually distinct when "weakly" replaces "strongly", as can be shown by simple examples [38].

Having now understood the four Hamiltonian inequalities that can be written for proximal subgradients, we become aware of some missing characterizations. For example, the property "weakly increasing" is not covered above, and it remains an open question whether this property of a system (φ, F) can be characterized via $\partial_P\varphi$. (If so, more must be involved than the type of Hamiltonian inequality considered thus far.) This difficulty is somewhat surprising, since $\partial_P\varphi$ is equally applicable to characterizing increase or decrease when trajectories are not involved [42].

When φ is continuous (or upper semicontinuous), there is an obvious recourse: observe that (φ, F) is weakly increasing iff $(-\varphi, F)$ is weakly decreasing, and invoke Theorem 7.1. We derive:

Theorem 7.3. *Let φ be upper semicontinuous. Then (φ, F) is weakly increasing iff*

$$H\big(x, \partial^P\varphi(x)\big) \geq 0 \quad \forall x. \tag{7.6}$$

Here, $\partial^P\varphi(x)$ denotes the *proximal superdifferential* of φ at x:

$$\partial^P\varphi(x) := -\partial_P(-\varphi)(x).$$

Let us now gather our conclusions together for the case of continuous φ. There are $2^3 = 8$ possible proximal inequalities, but only six distinct monotonicity properties. We present a detailed analysis of the issue of monotonicity along trajectories both because of the importance it turns out to have and because clarity seems to require it.

Theorem 7.4. *Let φ be continuous and F be locally Lipschitz. The following monotonicity properties of the system (φ, F) are all distinct, and each is characterized by its corresponding proximal inequality $(\forall x)$:*

(a) *weakly decreasing* $h\big(x, \partial_P\varphi(x)\big) \leq 0$;
(b) *weakly increasing* $H\big(x, \partial^P\varphi(x)\big) \geq 0$;
(c) *strongly decreasing* $H\big(x, \partial_P\varphi(x)\big) \leq 0$ *or* $H\big(x, \partial^P\varphi(x)\big) \leq 0$;
(d) *strongly increasing* $h\big(x, \partial_P\varphi(x)\big) \geq 0$ *or* $h\big(x, \partial^P\varphi(x)\big) \geq 0$;
(e) *weakly predecreasing* $h\big(x, \partial^P\varphi(x)\big) \leq 0$;
(f) *weakly preincreasing* $H\big(x, \partial_P\varphi(x)\big) \geq 0$.

We saw in §6 that in applying Lyapounov's method for stability, the need to consider weakly decreasing systems (φ, F) arose; we know now that this corresponds to the inequality $h(\partial_P\varphi) \leq 0$. Does the opposite inequality have any historical antecedents?

8 Verification functions

We now briefly describe how a venerable idea originating with Legendre leads to inequalities of the type $h(\partial_P\varphi) \geq 0$. The context is optimization, and we modify the variational setting in which the idea came to light to a control one consistent with our focus on trajectories.

Consider the following optimal control problem (P):

$$\text{Minimize}\, \ell\big(x(T)\big) \tag{8.1}$$

over all trajectories $x(\cdot)$, that are solutions of

$$\dot{x}(t) \in F\big(x(t)\big) \tag{8.2}$$

which satisfy a given initial condition

$$x(0) = x_0. \tag{8.3}$$

Here, x_0 and the "horizon" T are prescribed, as well as F of course, and $\ell\colon \mathbb{R}^n \to (-\infty, \infty]$ is a given extended-valued lower semicontinuous function.

As long as there is at least one admissible arc x (i.e., one satisfying (8.2) (8.3)) for which $\ell\big(x(T)\big) < \infty$, it follows that the problem admits a solution and has finite *value* V (the value of the minimum in (8.1) when x is an optimal solution). We remark that the device of allowing ℓ to be extended-valued implies that explicit endpoint constraints of the form $x(T) \in D$ can be implicitly incorporated in our formulation by taking $\ell\big(x(T)\big) = +\infty$ whenever $x(T)$ fails to lie in D.

Now suppose that we have a *feasible arc* $\bar{x}(\cdot)$ that we suspect of being optimal for our problem. Thus $\bar{x}$ is admissible and $\ell\big(\bar{x}(T)\big) < \infty$. How can we confirm that $\bar{x}$ is a solution? Here's one way: produce a smooth (C^1)

function $\varphi(t,x)$ such that

$$\varphi_t(t,x) + \langle \varphi_x(t,x), v \rangle \geq 0 \quad \forall (t,x,v) \tag{8.4}$$

$$\varphi(T,\cdot) = \ell(\cdot) \tag{8.5}$$

$$\varphi(0,x_0) = \ell\big(\bar{x}(T)\big) \tag{8.6}$$

Let us see how the existence of φ proves that $\bar{x}$ is optimal. Let x be any other feasible arc. Then (in view of (8.4))

$$\frac{d}{dt}\varphi\big(t,x(t)\big) = \varphi_t\big(t,x(t)\big) + \big\langle \varphi_x\big(t,x(t)\big), \dot{x}(t) \big\rangle \geq 0 \text{a.e.}$$

Integrating this on $[0,T]$ and invoking (8.5) yields

$$\ell\big(x(T)\big) \geq \varphi(0,x_0) = \ell\big(\bar{x}(T)\big). \quad \text{(by (8.6))}.$$

So $\bar{x}$ gives the least possible value V of $\ell\big(x(T)\big)$, as required to show. It also follows that $\varphi(0,x_0) = V$.

In this simple argument, the "Hamilton-Jacobi inequality" (8.4) was really only used to deduce that the map $t \to \varphi\big(t,x(t)\big)$ is increasing whenever x is a trajectory. In the terminology of the previous section, we simply want (φ,F) to be strongly increasing; we know (Theorem 7.4) that this can be characterized in proximal terms when φ depends only on x. It is no surprise that the characterization can be extended to the case of t-dependence; let us proceed to note the result.

For convenience of notation, we define an extended Hamiltonian $h_e \colon \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ as follows:

$$h_e \colon (x,\theta,\zeta) \mapsto \theta + h(x,\zeta). \tag{8.7}$$

Now let $\varphi(t,x)$ be lower semicontinuous. Essentially as before, the system (φ,F) is termed strongly increasing if $t \to \varphi\big(t,x(t)\big)$ is increasing whenever x is a trajectory. By considering t as an extra state variable subject to $\dot{t} = 1$, and by taking $F_e(x) := \{1\} \times F(x)$, we derive as a trivial consequence of Theorem 7.4 all the corresponding proximal characterizations of monotonicity when φ has t-dependence. Thus, for example, (φ,F) is strongly increasing iff $h_e\big(\partial_P\varphi(t,x)\big) \geq 0 \quad \forall (t,x)$. This means that for any (t,x), for any element (θ,ζ) of $\partial_P\varphi(t,x)$, we have

$$h_e(x,\theta,\zeta) = \theta + h(x,\zeta) \geq 0.$$

If we wish to restrict attention to only those arguments t lying in a given interval, $(-\infty,T]$ say, then some care must be exercised at the boundary. Here is an extension to this setting of the verification argument whose proof is an easy variation of the classical one, given Theorem 7.4:

Proposition 8.1. *Let $\bar{x}$ be feasible for (P), and suppose there exists a lower semicontinuous $\varphi(t,x)$ satisfying*

$$h_e\big(x,\partial_P\varphi(t,x)\big) \geq 0 \quad \forall (t,x) \in (-\infty,T) \times \mathbb{R}^n \tag{8.8}$$

$$\limsup_{t \uparrow T, y \to x} \varphi(t,y) \leq \ell(x) \quad \forall x \tag{8.9}$$

$$\varphi(0,x_0) = \ell\big(\bar{x}(T)\big) \tag{8.10}$$

Then $\bar{x}$ solves (P), and

$$V = \varphi(0,x_0). \tag{8.11}$$

This is an extension to nonsmooth φ of Legendre's approach to sufficient conditions, which in the calculus of variations has also been called the "royal road of Caratheodory". A lower semicontinuous function φ satisfying (8.8) – (8.10) is called a *verification function* (for $\bar{x}$).

The obvious question to ask at this point is how applicable the method turns out to be, or to rephrase this: can we be sure that a verification function φ for $\bar{x}$ exists when $\bar{x}$ is optimal?

Considerable insight into this question arises from applying the technique of invariant imbedding to the equality (8.11). Suppose that instead of the problem (P) considered above, we consider a family of problems $P(\tau,\alpha)$ parametrized by the initial data (τ,α); i.e., the initial condition is

$$x(\tau) = \alpha$$

rather than $x(0) = x_0$. Let $V(\tau,\alpha)$ denote the value of $P(\tau,\alpha)$; then we observe that the verification argument actually gives not only $V(0,x_0) = \varphi(0,x_0)$ as before (where φ is still as in Prop. 8.1), but also

$$V(\tau,\alpha) \geq \varphi(\tau,\alpha) \quad \forall (\tau,\alpha).$$

Further, by its very definition V satisfies (8.10). We are quite naturally led to consider whether we could take V itself as the function φ in Proposition 8.1. We remark that V is lower semicontinuous, as a consequence (once again!) of compactness of trajectories.

In taking $\varphi = V$, the issue of restricted domain arises: $V(\tau, \alpha)$ is only defined for $\tau \le T$. Thus proximal subgradients of V only make sense for $\tau < T$. But in this range, we can easily see that V indeed satisfies (8.8)! The reason: $V\big(t, x(t)\big)$ is always increasing when x is a trajectory; the minimum of $\ell\big(x(T)\big)$ can only be greater or equal starting from an intermediate point $\big(t', x(t')\big)$ than it was from an earlier (less committed) point $\big(\tau, x(\tau)\big)$. (This is an instance of the logic known as the *principle of optimality*.)

It is not hard to see that V satisfies

$$\liminf_{t \uparrow T, y \to x} V(t, y) = V(T, x) = \ell(x). \tag{8.12}$$

(That the left side is no greater than the right follows from lower semicontinuity, and noting that V is constant along optimal arcs then gives (8.12) whenever $\ell(x) < \infty$; the case $\ell(x) = +\infty$ is easily handled separately.) Alas, V fails in general to satisfy (8.9), although (8.12) holds. However, in those cases in which V is continuous, (8.9) evidently follows from (8.12) and all the requisite properties of a verification function are satisfied by V. Now, it can be shown from well-known properties of trajectories that V is continuous if ℓ is continuous. We obtain therefore:

Proposition 8.2. *When ℓ is continuous, a feasible arc $\bar{x}$ is optimal iff there exists a continuous verification function for $\bar{x}$; the value function V is one such verification function for any optimal arc.*

This is a satisfying justification of the verification method when ℓ is continuous, but the latter hypothesis is restrictive in that it rules out endpoint constraints. A deeper result applicable to endpoint-constrained problems is the following, which is the culmination of a long-term effort to completely justify the method [19] [57] [70] [31] [25]:

Theorem 8.1. *When (P) is normal, a feasible arc $\bar{x}$ is optimal iff there exists a Lipschitz continuous verification function φ for $\bar{x}$.*

The concept of normality pertains to necessary conditions for (P); see §11 for the definition. The proof of the theorem is too involved to be given here; we remark however that the verification function φ that it produces is not (and cannot be) the value function when endpoint constraints are active. The generalized Hamilton-Jacobi inequality (actually, equality) in [19] [57] [70] (see also [25]) is one of the early such generalizations; it is stated in terms of the generalized gradient of φ (see §10), but this is equivalent to the proximal form (8.8), as follows readily from nonsmooth analysis and the concavity of $h_e(x, \cdot)$. We refer to [25, Chapter 2] for a fairly determined effort to explain clearly the various aspects of the verification method, including such points as its relationship to dynamic programming and illustrations of its use. See also [48] [62] [12] [75] for more on sufficient conditions.

There are in general many possible verification functions for a given optimal $\bar{x}$. We have seen in this section, however, how naturally the value function V arises in connection with the verification method, and hence the associated Hamilton-Jacobi inequality. Might it be possible to establish an even closer relationship, perhaps even a characterization of V in Hamilton-Jacobi terms?

9 The Hamilton-Jacobi equation

The following theorem shows that the value function is the unique lower semicontinuous solution of a suitable generalization of the classical Hamilton-Jacobi equation

$$\varphi_t + H(x, \varphi_x) = 0.$$

Recall that $h_e(x, \theta, \zeta)$ is defined as $\theta + h(x, \zeta)$, and that the local Lipschitz property for F has joined the standing hypotheses.

Theorem 9.1. *There is a unique lower semicontinuous function $\varphi \colon (-\infty, T] \times \mathbb{R}^n \to (-\infty, \infty]$ satisfying*

$$h_e\big(x, \partial_P \varphi(t, x)\big) = 0 \quad \forall (t, x) \in (-\infty, T) \times \mathbb{R}^n \tag{9.1}$$

$$\liminf_{t \uparrow T, y \to x} \varphi(t, y) = \ell(x) = \varphi(T, x) \quad \forall x \in \mathbb{R}^n. \tag{9.2}$$

That unique function is the value function V.

Remark 9.1. In essence, (9.2) *defines* $\varphi(T,\cdot)$ via a limit inferior and requires that the result coincide with ℓ. This is essential for the uniqueness; to see why, just consider adding a positive constant to φ for $t < T$: this preserves lower semicontinuity as well as (9.1) ...

Proof. That V satisfies (9.2) was noted earlier (see (8.12)), as well as "half" of (9.1); i.e., the inequality (8.8). There remains to show

$$h_e\big(x, \partial_P V(t,x)\big) \le 0 \quad \forall (t,x) \in (-\infty, T) \times \mathbb{R}^n. \tag{9.3}$$

But whenever $V(\tau,\alpha)$ is finite, there is an optimal arc $\bar{x}$ for the problem $P(\tau,\alpha)$, and along $\bar{x}$, V is constant (i.e., $t \to V\big(t, \bar{x}(t)\big)$ is constant on $[\tau, T]$). Thus the system (V,F) is weakly decreasing relative to $t \in (-\infty, T)$, so that (9.3) holds by Theorem 7.4. We have shown that V satisfies (9.1) (9.2).

Now let φ be any other function as described in the theorem. Let us show first that $V \le \varphi$. To this end, let (τ,α) be any point (with $\tau < T$) at which φ is finite. Since φ satisfies $h_e(x, \partial_P \varphi) \le 0$, the system (φ, F) is weakly decreasing relative to $t < T$, so that for any $T' < T$ there is a trajectory x on $[\tau, T']$ such that

$$\varphi\big(T', x(T')\big) \le \varphi(\tau,\alpha).$$

Taking a sequence $T' \uparrow T$, semicontinuity and compactness of trajectories leads to a trajectory x such that $\ell\big(x(T)\big) = \varphi\big(T, x(T)\big) \le \varphi(\tau,\alpha)$. This implies $V(\tau,\alpha) \le \varphi(\tau,\alpha)$.

We now proceed to show $V \ge \varphi$. Let (τ,α) be any point at which V is finite. Then there exists a trajectory $\bar{x}$ optimal for $P(\tau,\alpha)$. By (9.2), there is a sequence (t_i, y_i), with $t_i < T$, converging to $\big(T, \bar{x}(T)\big)$, and such that $\varphi(t_i, y_i) \to \ell\big(\bar{x}(T)\big) = V(\tau,\alpha)$. By elementary properties of differential inclusions regarding the dependence on initial conditions (applied in reverse time), there exists (for i large) a trajectory x_i from a point (τ_i, α_i) to (t_i, y_i), where $(\tau_i,\alpha_i) \to (\tau,\alpha)$. Since φ satisfies $h_e(x, \partial_P \varphi) \ge 0$, the system (φ, F) is strongly increasing for $t \in [\tau_i, t_i]$, whence

$$\varphi(\tau_i, \alpha_i) \le \varphi(t_i, y_i).$$

This implies $\varphi(\tau,\alpha) \le V(\tau,\alpha)$ as required. □

The proof actually establishes two "comparison theorems" that we proceed to note formally:

Corollary 1. *Let $\varphi\colon (-\infty, T] \times \mathbb{R}^n \to (-\infty, \infty]$ be lower semicontinuous and satisfy*

(a) $h_e\big(x, \partial_P \varphi(t,x)\big) \le 0 \quad \forall (t,x) \in (-\infty, T) \times \mathbb{R}^n$,

(b) $\ell(x) \le \varphi(T,x) \quad \forall x \in \mathbb{R}^n$.

Then $\varphi \ge V$.

Corollary 2. *Let $\varphi\colon (-\infty, T] \times \mathbb{R}^n \to (-\infty, \infty]$ be lower semicontinuous and satisfy*

(a) $h_e\big(x, \partial_P \varphi(t,x)\big) \ge 0 \quad \forall (t,x) \in (-\infty, T) \times \mathbb{R}^n$,

(b) $\liminf_{t \uparrow T, y \to x} \varphi(t,y) \le \ell(x) \quad \forall x \in \mathbb{R}^n$.

Then $\varphi \le V$.

Remark 9.2. Corollary 1 is valid without the Lipschitz hypothesis on F, in contrast to its companion.

When ℓ (and hence V) is continuous, we can work with continuous functions, and the boundary condition simplifies. We can also invoke the proximal supergradient as well as the subgradient via Theorem 7.4(d) to characterize strong increase, and the very same proof gives:

Corollary 3. *Let ℓ be continuous. V is then the unique continuous function satisfying*

(a) $h_e\big(x, \partial_P \varphi(t,x)\big) \le 0 \quad \forall (t,x) \in (-\infty, T) \times \mathbb{R}^n$,

(b) $h_e\big(x, \partial^P \varphi(t,x)\big) \ge 0 \quad \forall (t,x) \in (-\infty, T) \times \mathbb{R}^n$,

(c) $\varphi(T,\cdot) = \ell(\cdot)$.

We shall remain in the continuous case for the remainder of this section, in order to facilitate comparison with two well-known generalizations of the concept of solution to the Hamilton-Jacobi equation, namely Subbotin's *minimax solutions* and the *viscosity solutions* of Crandall-Lions. These solution concepts are expressed not in proximal terms, but via other constructs of nonsmooth analysis which we need to review as a preliminary.

The *lower Dini derivate* of f at x in the direction v, denoted $Df(x;v)$, is defined by

$$Df(x;v) := \liminf_{\substack{v' \to v \\ \lambda \downarrow 0}} \frac{f(x + \lambda v') - f(x)}{\lambda}.$$

The vector ζ is called a *Dini subgradient* of f at x provided one has

$$Df(x;v) \geq \langle \zeta, v \rangle \quad \forall v \in \mathbb{R}^n$$

(alternate terminology: Fréchet subgradient).

The set of Dini subgradients of f at x is denoted $\partial_D f(x)$. In much of the viscosity solution literature, this is called simply the subdifferential, and it is defined equivalently as those ζ such that for some C^1 function g with $g'(x) = \zeta$, the function $f - g$ has a minimum at x. Minimax solutions are defined directionally via Df, and viscosity solutions via the complementary subdifferential ∂_D.

The following consequence of Theorem 9.1 asserts that V is the unique minimax solution of the Hamilton-Jacobi equation:

Corollary 4. *V is the unique continuous function $\varphi \colon (-\infty, T] \times \mathbb{R}^n \to \mathbb{R}$ satisfying*

- *(a)* $\inf_{v \in F(x)} D\varphi(t, x; 1, v) \leq 0 \quad \forall (t, x) \in (-\infty, T) \times \mathbb{R}^n,$
- *(b)* $\sup_{v \in F(x)} D\varphi(t, x; -1, -v) \leq 0 \quad \forall (t, x) \in (-\infty, T) \times \mathbb{R}^n,$
- *(c)* $\varphi(T, \cdot) = \ell(\cdot)$.

Proof. It suffices to prove that the two conditions (a) of Corollaries 3 and 4 are equivalent, as well as (b). First, let (a) of Corollary 4 hold; now, let (θ, ζ) belong to $\partial_P \varphi(t, x)$. We have, by definition, for (t', x') near (t, x),

$$\varphi(t', x') - \varphi(t, x) + \sigma\{|t' - t|^2 + |x' - x|^2\} \geq \theta(t' - t) + \langle \zeta, x' - x \rangle. \tag{9.4}$$

Let $\varepsilon > 0$ be given. Put $(t', x') = (t + \lambda, x + \lambda v')$ in (9.4) for $\lambda > 0$ small, where $v' \to v$, and where $v \in F(x)$ satisfies

$$D\varphi(t, x; 1, v) < \varepsilon.$$

Then (9.4) gives

$$\varepsilon > D\varphi(t, x; 1, v) \geq (\theta, \zeta) \cdot (1, v) \geq h_e(x, \theta, \zeta).$$

We deduce (a) of Corollary 3, since ε is arbitrary.

Now let (a) of Corollary 3 hold, and suppose that for some $\varepsilon > 0$, we have

$$D\varphi(t, x; 1, v) > \varepsilon \quad \forall v \in F(x).$$

Then for some $\delta > 0$ we have

$$D\varphi(t, x; 1, v) > \frac{\varepsilon}{2} \quad \forall v \in F(x) + \delta B. \tag{9.5}$$

It can be shown (see [29], which gives the appropriate extension of Subbotin's initial result [78]) that (9.5) implies the existence, for any $r > 0$, of $(\theta, \zeta) \in \partial_P \varphi(t', x')$, where (t', x') is of distance less than r from (t, x), such that

$$(\theta, \zeta) \cdot (1, v) > \frac{\varepsilon}{3} \quad \forall v \in F(x) + \delta B$$

(this is a nontrivial fact from nonsmooth analysis).

Thus, as soon as $F(x')$ lies within $F(x) + \delta B$ we deduce

$$h_e(x', \theta, \zeta) \geq \frac{\varepsilon}{3} \text{ for some } (\theta, \zeta) \in \partial_P \varphi(t', x'),$$

contradicting (a) of Corollary 3 as required.

The proof that the (b) parts are equivalent is similar to the above; we omit it. ◻

We turn now to the viscosity solutions, which are usually defined via ∂_D. It is easy to prove directly from the definitions that $\partial_P f(x) \subseteq \partial_D f(x)$. A somewhat deeper result [29] is that the graph of $\partial_P f$ is dense in that of $\partial_D f$; i.e. given $\zeta \in \partial_D f(x)$ and $\varepsilon > 0$, there exist x' and ζ' within ε of x and ζ respectively such that $\zeta' \in \partial_P f(x')$ (actually, a bit more is true: one can also arrange to have $f(x')$ within ε of $f(x)$, even when f is just lower semicontinuous). Of course when f is continuous, there is a Dini superdifferential $\partial^D f$ for which approximation by the proximal supergradient $\partial^D f$ holds. This fact immediately implies the coincidence of proximal solutions and viscosity solutions:

Corollary 5. *The statement of Corollary 3 remains true if in* (a) *and* (b) *we replace $\partial_P\varphi$ by $\partial_D\varphi$ and $\partial^P\varphi$ by $\partial^D\varphi$.*

Remark 9.3. To avoid possible confusion, we should point out that requiring (a) (b) of the Corollary (with $\partial_D\varphi$ and $\partial^D\varphi$) corresponds to a viscosity solution of the equation $-h_e = 0$, i.e. of

$$-\varphi_t - h(x, \varphi_x) = 0,$$

and *not* of $h_e = 0$ (there is a distinction). Equivalently, V is the viscosity solution of $-\varphi_t + H(x, -\varphi_x) = 0$. Perturbing terminal rather than final endpoints would give a value function satisfying (in the viscosity sense) a Hamilton-Jacobi equation without the two minus signs.

We turn now to arbitrary first-order partial differential equations, to which each of the three solution concepts given above can be extended.

Let $F\colon \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ be a continuous function, and let Ω be an open subset of $\mathbb{R}^n$. Our purpose is to discuss generalized solutions of the partial differential equation

$$F\big(x, u(x), \nabla u(x)\big) = 0, \quad x \in \Omega. \tag{9.6}$$

Definition 9.1. A lower semicontinuous function $u\colon \mathbb{R}^n \to (-\infty, \infty]$ [an upper semicontinuous function $v\colon \mathbb{R}^n \to [-\infty, \infty)$] is said to be a viscosity supersolution [subsolution] of (9.6) if it satisfies the following condition: whenever there is a point $x \in \Omega$ at which the function is finite and a function φ Fréchet differentiable at x such that $u - \varphi$ attains a local minimum [$v - \varphi$ attains a local maximum] at x, then relation (9.7) [(9.8)] below holds:

$$F\big(x, u(x), \nabla\varphi(x)\big) \geq 0, \tag{9.7}$$

$$F\big(x, v(x), \nabla\varphi(x)\big) \leq 0. \tag{9.8}$$

(This varies slightly from the terminology in [45] by allowing semisolutions to be extended-valued.) A *viscosity solution* of (9.6) is a continuous function which is simultaneously a viscosity supersolution and subsolution.

Definition 9.2. A lower semicontinuous function $u\colon \mathbb{R}^n \to (-\infty, \infty]$ [an upper semicontinuous function $v\colon \mathbb{R}^n \to [-\infty, \infty)$] is said to be a proximal supersolution [subsolution] of (9.6) if relation (9.9) [(9.10)] holds:

$$F\big(x, u(x), \zeta\big) \geq 0 \quad \forall \zeta \in \partial_P u(x), \quad \forall x \in \Omega, \tag{9.9}$$

$$F\big(x, v(x), \zeta\big) \leq 0 \quad \forall \zeta \in \partial^P v(x), \quad \forall x \in \Omega, \tag{9.10}$$

Again, a *proximal solution* of (9.6) is a continuous function which is both a proximal subsolution and supersolution.

Definition 9.3. A lower semicontinuous function $u\colon \mathbb{R}^n \to (-\infty, \infty]$ [an upper semicontinuous function $v\colon \mathbb{R}^n \to [-\infty, \infty)$] is said to be a minimax supersolution [subsolution] of (9.6) if relation (9.11) [(9.12)] holds for any $x \in \Omega$ at which the function is finite:

$$\sup_p \inf_v \Big\{ \underline{D}u(x; v) - \langle p, v\rangle - F\big(x, u(x), p\big) \Big\} \leq 0 \tag{9.11}$$

$$\inf_p \sup_v \Big\{ \overline{D}u(x; v) - \langle p, v\rangle - F\big(x, u(x), p\big) \Big\} \geq 0. \tag{9.12}$$

Here, $\underline{D}u$ signifies the lower Dini derivate Du already used above, whereas $\overline{D}u$ is the upper Dini derivate in which "lim sup" replaces "lim inf" in the definition. A *minimax solution* of (9.6) is a continuous function which is both a minimax subsolution and supersolution.

Theorem 9.2. *Let F be continuous, and let $w\colon \Omega \to \mathbb{R}$ be continuous. Then the following are equivalent*:

(a) w is a minimax solution;

(b) w is a viscosity solution;

(c) w is a proximal solution.

In fact, each of the semisolution concepts are equivalent for merely semicontinuous functions.

The proof, which is a nontrivial exercise in nonsmooth analysis, can be extended to Hilbert space. We refer to [29] for details. Note that a priori, it is proximal solutions that are easiest to confirm: w need be checked only at those points for which $\partial_P w \neq \varnothing$ (for the supersolution part, say), a proper subset of those points for which $\partial_D w \neq \varnothing$. On the other hand, once w is known to be a solution, it inherits, as a minimax solution, properties that hold at *all* points.

Remark 9.4. (a) Using L. C. Young's framework of generalized flows, R. B. Vinter [85] has used linear/convex duality to prove a significant result complementary to Corollary 2 above: V is revealed as the supremum of all C^1 functions φ satisfying (8.4) together with $\varphi(T,\cdot) \leq \ell(\cdot)$.

(b) There are of course many Hamiltonians other than h for which the study of the Hamilton-Jacobi equation arises; the crucial property of $h(t,x,p)$ making possible the type of approach presented here is its concavity in p. In particular, the one-sided mode of definition of discontinuous solution is problematic without this property; see Barles [4]. Other types of underlying control problems (differential games, stopping times, stochastic control, evolution equations, higher order ...), give rise to multiple variants, many featuring interesting and difficult technical issues; see [3] [5] [6] [13] [15] [46] [45] [50] [51] [52] [56] [76] [77] [80] [81] [84] and the many references therein.

(c) Several schools have participated over the past two decades in developing the nonsmooth theory of the Hamilton-Jacobi equation, not always with acknowledgment of intellectual debt. The state of the art circa 1976 is described [7] where the "almost everywhere" type of (Lipschitz) solution dominates, and where Fleming's "artificial viscosity" approximation method is outlined. To our knowledge, the first truly pointwise (subdifferential) definition of generalized solution (again for Lipschitz functions) apeared in 1977 [57], in connection with the verification method (see §8). (This turns out to be a certain viscosity semisolution.) Subbotin [76] inaugurated the two-sided approach to defining a nonsmooth solution, with Dini derivates and in the context of differential games. Subsequently, Crandall and Lions [42] developed further Fleming's approach; the (eventual) two-sided, subdifferential approach to defining viscosity solutions, carried out for continuous functions and with attendant uniqueness theorem, consituted a breakthrough that vindicated the nonsmooth analysis approach to the issue.

(d) It appears to be Barron and Jensen [6] who first demonstrated the possibility of giving a single subdifferential characterization in certain cases, and for merely lower semicontinuous solutions. The results of this section owe an intellectual debt to them, and also to Subbotin, who first stressed the relevance of invariance.

We have seen that the value function is the unique solution of a suitably generalized Hamilton-Jacobi equation. To calculate it, we could presumably solve the family of underlying optimal control problems. (There is some irony in this, for we were led to the Hamilton-Jacobi equation in the first place as a tool for solving a single one of those problems!) There has also been considerable work done on computing V directly, for example by multigrid methods. Suppose then that we did succeed in finding V; how can we use this information to calculate the actual solutions of the optimal control problem?

10 Optimal and suboptimal feedback

The Hamilton-Jacobi inequality

$$h_e\big(x, \partial_P \varphi(t,x)\big) \leq 0, \quad t < T \tag{10.1}$$

together with the boundary condition

$$\varphi(T,\cdot) \geq \ell(\cdot) \tag{10.2}$$

are useful in producing upper bounds for the value V of the optimal control problem introduced in §8; as we saw in Corollary 1 to Theorem 9.1, these conditions imply $V \leq \varphi$. Thus for each $(\tau,\alpha) \in (-\infty, T] \times \mathbb{R}^n$ there is a trajectory $\bar{x}$ with $\ell\big(\bar{x}(T)\big) \leq \varphi(\tau,\alpha)$. We address now the issue of actually constructing such a trajectory. In the special case in which $\varphi = V$, this becomes the issue of finding optimal trajectories.

It is interesting to recall the classical approach to this issue. Assuming that φ is smooth, this would direct us to select for each (t,x) a point $\overline{v}(t,x)$ in $F(x)$ at which the minimum defining $h_e\big(x,\nabla\varphi(t,x)\big)$ is attained; i.e. such that

$$\varphi_t(t,x) + \big\langle \varphi_x(t,x), \overline{v}\big\rangle = h_e\big(x,\nabla\varphi(t,x)\big).$$

Then, we proceed to define a trajectory $\bar{x}$ via

$$\dot{\bar{x}}(t) = \overline{v}\big(t,\bar{x}(t)\big), \quad \bar{x}(\tau) = \alpha.$$

If all this is possible, we derive

$$\ell\big(\bar{x}(T)\big) \le \varphi(\tau,\alpha), \tag{10.3}$$

as follows:

$$\begin{aligned}
\ell\big(\bar{x}(T)\big) - \varphi(\tau,\alpha) &\le \varphi\big(T,\bar{x}(T)\big) - \varphi(\tau,\alpha) \\
&= \int_0^T \frac{d}{dt}\varphi\big(t,\bar{x}(t)\big)\,dt \\
&= \int_0^T \Big\{\varphi_t\big(t,\bar{x}(t)\big) + \Big\langle \varphi_x\big(t,\bar{x}(t)\big), \dot{\bar{x}}\Big\rangle\Big\}\,dt \\
&= \int_0^T h_e\Big(\bar{x}(t),\nabla\varphi\big(t,\bar{x}(t)\big)\Big)\,dt \le 0.
\end{aligned}$$

The difficulties with this dynamic programming approach are intrinsic (smoothness of φ, regularity of $\overline{v}$, existence of $\bar{x}$), but it is of note that it attempts to construct a *feedback* giving rise to the required trajectory. Our approach allows us to rescue the approach in essentially these terms. The integration step above is still not possible, but proximal methods produce the required system monotonicity. We remark that the concept of Euler solution mentioned below was defined in §2.

Theorem 10.1. *Let the lower semicontinuous function $\varphi\colon (-\infty,T]\times\mathbb{R}^n \to (-\infty,\infty]$ satisfy* (10.1) (10.2)*, and let $(\tau,\alpha)\in(-\infty,\infty)\times\mathbb{R}^n$ be given. Then there exists a selection $\overline{v}$ of F* (i.e., *a function $\overline{v}(t,x)$ with values in $F(x)\quad\forall(t,x)$*) *with the property that every Euler solution $\bar{x}$ of the initial-value problem*

$$\dot{x} = \overline{v}(t,x), \quad x(\tau) = \alpha$$

satisfies $\ell\big(\bar{x}(T)\big) \le \varphi(\tau,\alpha)$.

Proof. As we shall see, the theorem is not just an abstract existence theorem: we shall construct $\overline{v}$ quite explicitly. To do so, let us consider the set

$$S := \big\{(t,x)\in[-\infty,T]\times\mathbb{R}^n\colon \varphi(t,x)\le\varphi(\tau,\alpha)\big\} \cup \big\{(t,x)\colon t\ge T, x\in\mathbb{R}^n\big\}.$$

Let us note that S is closed, since φ is lower semicontinuous, but also that the system (S,F_e) is weakly invariant, where (as in §8) $F_e(x) := \{1\}\times F(x)$. This is essentially a restatement of the fact that (φ,F) is weakly decreasing, which follows from (10.1).

Theorem 2.2 says that weak invariance is equivalent to a certain proximal normal condition, precisely the one used in Theorem 2.1 to construct (via "proximal aiming") a function $f(t,x)$ generating invariant trajectories (see Remark 2.1). In evident notation that treats t as the first component of an augmented state variable (t,x), we had defined $f(t,x)$ by the following procedure:

Let (s,y) be any point in S closest to (t,x), and let $f(t,x)$ be a point in $F(y)$ minimizing $v\mapsto\langle v,x-y\rangle$. It follows then from the proximal normal condition that

$$t - s + \big\langle f(t,x), x-y\big\rangle \le 0. \tag{10.4}$$

The proof of Theorem 2.1 uses (10.4) to show that any Euler solution of

$$(\dot{t},\dot{x}) = \big(1, f(t,x)\big)$$

beginning in S remains in S thereafter.

The function f is not quite a feedback, so we define $\overline{v}(t,x)$ as follows: $\overline{v}(t,x)$ is the point in $F(x)$ closest to $f(t,x)$ (which itself lies in $F(y)$). The Lipschitz condition on F gives:

$$\left|\overline{v}(t,x) - f(t,x)\right| \le K|x-y|.$$

Together with (10.4), this implies

$$\begin{aligned} t - s + \big\langle \overline{v}(t,x), x-y \big\rangle \\ &\le t - s + \big\langle f(t,x), x-y \big\rangle + K|x-y|^2 \\ &\le K d_S^2(t,x). \end{aligned} \tag{10.5}$$

The next step in the proof is to note that in the construction used to prove Theorem 2.1, it suffices to have (10.5) instead of (10.4): the square distance term necessitates only routine modifications. Thus any Euler solution of $\dot{x} = \overline{v}(t,x)$ for which $\big(\tau, x(\tau)\big) = (\tau,\alpha)$ lies in S is such that $\big(t, x(t)\big) \in S$, $t \ge \tau$. By definition then,

$$\varphi\big(t, x(t)\big) \le \varphi(\tau,\alpha) \text{ for } \tau < t < T.$$

This easily implies $\ell\big(x(T)\big) \le \varphi(\tau,\alpha)$ as required. $\square$

Remark 10.1. Let us summarize in alternate terms the definition of a suitable proximal feedback $\overline{v}(t,x)$: pick any (s,y) in S closest to (t,x), and let $\overline{v}(t,x)$ in $F(x)$ be a point at which the minimum defining $h(x, x-y)$ is attained. It is possible to construct $\overline{v}$ explicitly in simple cases; we observe however that in general $\overline{v}$ cannot be constructed so as to be continuous. Indeed, it is well known that optimality may require discontinuous feedback laws. See [9] [73] for related results.

For practical purposes of control design, it is sometimes desirable to have a more regular feedback; the Lipschitz feedback laws are especially agreeable since they generate a unique trajectory once the initial condition is specified. Can semisolutions of the Hamilton-Jacobi equation be used to generate Lipschitz feedbacks? We now give one such result, in terms of the *generalized gradient* $\partial\varphi$ of a locally Lipschitz map φ. For such functions, in finite dimensions, we may define $\partial\varphi$, as follows [16] [25]:

$$\partial\varphi(t,x) := \operatorname{cl}\operatorname{co}\big\{\lim \nabla\varphi(t_i,x_i)\big\},$$

where the set of limits is generated by all sequences (t_i,x_i) converging to (t,x) along which $\nabla\varphi(t_i,x_i)$ exists. Below, an "ε-feedback" $\overline{v}$ is a function satisfying $\overline{v}(t,x) \in F(x) + \varepsilon B$ for all (t,x).

Theorem 10.2. *Let φ be locally Lipschitz and satisfy*

$$h_e\big(x, \partial\varphi(t,x)\big) < \Delta \tag{10.6}$$

together with the boundary condition (10.2). Then for any $\varepsilon > 0$, there is an ε-feedback $\overline{v}$ Lipschitz in x such that for any $(\tau,\alpha) \in (-\infty, T) \times \mathbb{R}^n$, the solution x (in the classical sense) to

$$\dot{x}(t) = \overline{v}\big(t, x(t)\big), \quad x(\tau) = \alpha$$

satisfies

$$\ell\big(x(T)\big) < \varphi(\tau,\alpha) + \Delta(T-\tau). \tag{10.7}$$

Remark 10.2. (a) We omit the proof of the theorem, which is based upon the calculus of $\partial\varphi$. Note that the value function V satisfies $h_e(x, \partial_P V) = 0$ by Theorem 9.1, but ∂V contains $\partial_P V$ and in general we will have $h_e\big(x, \partial V(t,x)\big)$ greater than 0 for certain (t,x) (we remark that V will be Lipschitz when ℓ is). Taking

$$\Delta > \sup_{(t,x)} h_e\big(x, \partial V(t,x)\big)$$

allows us to apply the theorem to get (for any $\varepsilon > 0$) what might be called a "Δ-optimal ε-feedback". We can think of Δ as the cost paid (relative to optimality) for having a Lipschitz feedback law.

(b) Besides the regularity of φ, an important distinction between the feedbacks produced by Theorems 10.1 and 10.2 concerns the initial values (τ,α) for which they are valid (i.e. for which the estimates (10.3) or (10.7) hold). That of Theorem 10.1 is constructed especially for the given point (τ,α) used in defining S; it also works for any other (τ',α') satisfying $\varphi(\tau',\alpha') \le \varphi(\tau,\alpha)$. On the other hand, the feedback of Theorem 10.2 yields the estimate (10.7) for *any* initial data.

(c) The design of *universal* feedbacks in the context of Theorem 2.1 (i.e., those yielding estimate (10.3) for *any* initial (τ, α)), is an important problem, especially when the control system is subject to uncertainty, or in differential games [8] [59] [60] [61]. The methods of this article extend to such situations: universal feedbacks are constructed in [39], which is the first part of forthcoming work on systems with uncertainty. We remark that when uncertainty is present, a type of Lavrentiev phenomenon can hold: the minimum cost ℓ obtainable by continuous feedback can be strictly greater than that obtainable by discontinuous ones.

Given the central role played by the value function, it is natural to ask for necessary conditions for the underlying optimal control problem that could help us identify optimal trajectories. One might expect the Hamiltonian to figure once again in such results. Are there such necessary conditions?

11 Hamiltonian inclusions

In this section we will describe briefly a Hamiltonian theory of necessary conditions for optimal control. The theory and its applications constitute by now a considerable body of work, so we shall limit ourselves to a presentation of the basic ideas most related to previous sections, together with a thumbnail guide to the literature.

We consider now the same problem (P) considered before, and we suppose that a given trajectory $x(\cdot)$ solves (P). For ease of exposition, we suppose initially that ℓ has the form

$$\ell(x) = g(x) + \psi_S(x),$$

where g is locally Lipschitz and where ψ_S is the *indicator function* of a closed set S (hence $\psi_S(x) = 0$ if $x \in S$, and $+\infty$ otherwise). The *generalized normal cone* $N_S(x)$ to S at x is a now-familiar object (see [24]) needed for the statement below.

Theorem 11.1. *If x solves (P), then there exist an arc $p(\cdot)$ and a scalar λ equal to 0 or 1 with $\|p\| + \lambda \neq 0$ such that*

$$\big(-\dot p(t), \dot x(t)\big) \in \partial H\big(x(t), p(t)\big) \quad \text{a.e. in } [0,T] \tag{11.1}$$

$$-p(T) \in \lambda \partial g\big(x(T)\big) + N_S\big(x(T)\big). \tag{11.2}$$

For some scalar h, we also have

$$H\big(x(t), p(t)\big) = h \quad \forall t \in [0,T]. \tag{11.3}$$

Remark 11.1. (a) The generalized gradient ∂H was introduced in the previous section. We refer to (11.1) as a *Hamiltonian inclusion* and (11.2) as a *transversality condition*. The Hamiltonian inclusion was introduced in 1973 by Clarke. There is quite a bit to be said about such topics as its relationship to the celebrated Pontryagin maximum principle (from which it is distinct), its many variants, including versions for free time or state-constrained problems, the case in which F is unbounded, and analogous results for the more complex generalized problem of Bolza; we refer to [18] [21] [23] [26] [64,65,66,67] [72] and to the many references therein.

(b) The case of a general ℓ gives rise to a transversality condition of the form

$$-\big(p(T), \lambda\big) \in N_{\operatorname{epi}\ell}\big(x(T), \ell\big(x(T)\big)\big).$$

Regarding transversality, Mordukhovic [68] has observed that with no actual modifications in the proofs, the term ∂g in (11.2), for example, can be replaced by a potentially smaller object $\partial_L g$, the set of limiting proximal subgradients [25] [58] [69] [71]; this is not true of ∂H in (11.1), however.

(c) The stipulation $\|p\| + \lambda \neq 0$ is present to avoid triviality of the necessary conditions, since $\lambda = 0$ and $p \equiv 0$ always satisfy (11.1) (11.2). When there fails to be any nontrivial arc $p(\cdot)$ satisfying (11.2) for $\lambda = 0$, then x is called *normal*. (P) itself is called normal when every solution x to (P) is normal. When $S = \mathbb{R}^n$ (i.e., when there is no active endpoint constraint on $x(T)$), then (P) is automatically normal. This follows from the fact that any p satisfying (11.1) is either everywhere nonzero or else identically zero.

Proof of Theorem 11.1. There are two known routes to the Hamiltonian inclusion: the original exact penalization approach which employs a nonsmooth penalty Lagrangian and a generalized Euler equation (see for example [24]), and a more recent one [26] which uses "proximal decoupling". We shall sketch the latter, since it reprises two of our main themes in this article: proximals and value functions.

For simplicity, let us take the case $T = 1$, $\ell(x) = \langle w, x\rangle$ for some given vector w. Consider the value function V defined on L^2 as follows: $V(\alpha)$ is the minimum of $\ell\big(x(1)\big)$ over all arcs x satisfying $x(0) = x_0$ and the (perturbed) differential inclusion

$$\dot{x}(t) \in F\big(x(t) + \alpha(t)\big) \text{a.e.}, \quad t \in [0, 1].$$

Suppose that $\partial_P V(0)$ contains an element ζ. Then locally, and ignoring the quadratic terms, which complicate matters significantly but can be dealt with in the ensuing arguments, we have

$$V(\alpha) - V(0) \geq \langle \zeta, \alpha\rangle.$$

Now let (u, v) be any pair in L^2 near $(x, \dot{x})$, satisfying $v(t) \in F\big(u(t)\big)$ a.e. Let y be the arc such that $\dot{y} = v$ a.e., $y(0) = x_0$. Then

$$\dot{y}(t) \in F\big(y(t) + \alpha(t)\big),$$

provided we take $\alpha := u - y$, which is near 0 if (u, v) is near $(x, \dot{x})$. But then $V(\alpha) \leq \ell\big(y(1)\big)$ by definition of V; also, $V(0) = \ell\big(x(1)\big)$ since x solves (P). Substituting into the inequality above yields

$$\left\langle w, \int_0^1 v(t)\, dt \right\rangle - \int_0^1 \big\langle \zeta(t), u(t) - y(t)\big\rangle\, dt \geq \big\langle w, x(1)\big\rangle.$$

Recall that $\dot{y} = v$; using integration by parts on the left side, and defining an arc p by $\dot{p} = \zeta$, $p(1) = -w$ produces (when the dust settles, and still ignoring some quadratic terms):

$$\int_0^1 \big\{\langle \dot{p}, u\rangle + \langle p, v\rangle\big\}\, dt \leq \int_0^1 \big\{\langle \dot{p}, x\rangle + \langle p, \dot{x}\rangle\big\}\, dt \tag{11.4}$$

whenever (u, v) in L^2 lies near $(x, \dot{x})$ and satisfies

$$\big(u(t), v(t)\big) \in \mathrm{gr}(F) \text{a.e.} \tag{11.5}$$

Notice the effect to this point: the constraint $\dot{u} = v$ has been removed (u and v are "decoupled"), and certain linear terms have been added to the quantity being extremized. The p defining these terms is essentially the proximal subgradient ζ, and will turn out to be the "adjoint variable" satisfying the Hamiltonian inclusion with x. This is just one case of a general principle linking subgradients of a value function to the "Lagrange multipliers" occurring in the necessary conditions, a theme we shall see again below.

We see from (11.4) that $(x, \dot{x})$ solves a certain linear problem relative to the decoupled constraint (11.5). The structure of this new problem is simple enough to make its direct analysis relatively straightforward. We can deduce

$$(\dot{p}, p) \in N^P_{\mathrm{gr}(F)}(x, \dot{x}) \text{a.e.}$$

The final step is the derivation of the Hamiltonian inclusion (11.1) from this relation; see [26]. We have omitted here not only this step, but the complications stemming from the fact that $\partial_P V(0)$ may be empty; in general we must have recourse to $\zeta_i \in \partial_P V(\alpha_i)$, where $\alpha_i \to 0$ in L^2, and carry out a limiting procedure. □

Value function analysis

There is an intimate relationship between the arcs p occurring in the Hamiltonian inclusion and the differential properties of certain value functions, one that has been exploited for a variety of issues ranging from controllability to sensitivity to periodic solutions. Let us motivate this relationship by means of a somewhat heuristic discussion involving the same value function V that played an important role in sections 8 and 9.

Let (θ, ζ) belong to $\partial_P V(\overline{\tau}, \overline{\alpha})$. Then ignoring quadratic terms which would not contribute in any case to first order necessary conditions, we have (locally)

$$V(\tau, \alpha) - V(\overline{\tau}, \overline{\alpha}) \geq \theta(\tau - \overline{\tau}) + \langle \zeta, \alpha - \overline{\alpha}\rangle. \tag{11.6}$$

Now let $\overline{x}$ solve $P(\overline{\tau}, \overline{\alpha})$, so that

$$\overline{x}(\overline{\tau}) = \overline{\alpha}, \quad V(\overline{\tau}, \overline{\alpha}) = \ell\big(\overline{x}(T)\big).$$

Let x be any arc feasible for $P(\tau, \alpha)$; then $V(\tau, \alpha) \leq \ell\big(x(T)\big)$, and substituting into (11.6) we conclude that the quantity

$$\ell\big(x(T)\big) - \theta\tau - \langle \zeta, \alpha\rangle = \ell\big(x(T)\big) - \theta\tau - \big\langle \zeta, x(\tau)\big\rangle$$

is minimized over (local) choices of both τ and of trajectories x on $[\tau, T]$ by taking $\tau = \overline{\tau}$ and $x = \overline{x}$. This is a free-time optimal control problem to which applies an appropriate version of the necessary conditions of Theorem 11.1. These give the existence of p such that the Hamiltonian inclusion (11.1) holds, as well as

$$-p(T) \in \partial \ell\big(\overline{x}(T)\big), \quad -p(\tau) = \zeta$$
$$H\big(\overline{x}(t), p(t)\big) = h = \theta.$$

We see therefore that $\big(h, -p(\tau)\big)$ belongs to $\partial_P V(\tau, \alpha)$, establishing another instance of the general principle that relates the subdifferential of V to the "multipliers" p that figure in the optimality conditions. Well known heuristically, this case of the principle was first justified rigorously in a control context by Clarke and Vinter [32] [83] in terms of ∂V and all along the trajectory. We also deduce from the preceding

$$\theta + h\big(\overline{x}(\tau), \zeta\big) = \theta - H\big(\overline{x}(\tau), -\zeta\big) = h - H\big(\overline{x}(\tau), p(\tau)\big) = 0,$$

another way to obtain the proximal Hamilton-Jacobi equation studied earlier. As a final consideration, consider the possibilities as we take sequences (τ_i, α_i) of points (τ, α) such as above. If the corresponding $(\theta_i, \zeta_i) \in \partial_P V(\tau_i, \alpha_i)$ become unbounded, it can be shown that an abnormal multiplier results. Since Lipschitz behavior of V is equivalent to local boundedness of its proximal subgradients, we are led to the conclusion that "normal problems give rise to Lipschitz value functions", a general principle that can be explicitly implemented in many circumstances.

What is the interest in knowing that a value function is Lipschitz? To answer this question, consider an example: let $x(\cdot)$ be a trajectory on $[0, T]$ such that $x(0) \in S_0$, $x(T) \in S_1$. A natural question in control is whether for all α, β small, there is still a trajectory y satisfying $y(0) \in S_0 + \alpha$, $y(T) \in S_1 + \beta$, a controllability issue. Further, stability is of interest: can we choose y to satisfy, for example,

$$\int_0^T \big|\dot{x}(t) - \dot{y}(t)\big|\, dt \le K\big|(\alpha, \beta)\big|,$$

for some K? It is easy to see that the answers to these questions are affirmative if a certain value function W is Lipschitz, namely the one for which $W(\alpha, \beta)$ is defined as the minimum of

$$\int_0^T \big|\dot{x}(t) - \dot{y}(t)\big|\, dt$$

over all trajectories y beginning in $S_0 + \alpha$ and ending in $S_1 + \beta$. And then the stability rank K is the Lipschitz constant of W, a quantity that we could seek to estimate via multipliers p associated to x, in keeping with the discussion above. The details of this example are given in [24], where W is shown to be Lipschitz if x is normal.

Value function analysis of this type is made possible by the good properties of the Hamiltonian inclusion, among them the critical one of being "preserved in the limit". A number of challenging solvability and sensitivity problems have been addressed via this methodology, including discontinuous free time problems [41], time delayed control problems [34], functional differential equations [35], stochastic control [63], multiprocesses [33], periodic Hamiltonian trajectories [17] [20], infinite-dimensional perturbations [22] [64], time optimal and generally parametrized problems [30]. Sufficiency via ∂V was broached in [82]; see also [14]. Chapter 4 of [25] surveys the state of value function analysis *circa* 1989, but the subject has been developing rapidly.

References

[1] J.-P. Aubin, *Viability Theory*. Birkhauser, Boston, 1991.

[2] J.-P. Aubin and A. Cellina, *Differential Inclusions*. Springer-Verlag, New York, 1984.

[3] M. Bardi and P. Soravia, *Hamilton-Jacobi equations with singular boundary conditions on a free boundary and applications to differential games*. Trans. Amer. Math. Soc. 325 (1991), 205-229.

[4] G. Barles, *Discontinuous viscosity solutions of first-order Hamilton-Jacobi equations: a guided visit*. Nonlinear Analysis **20** (1993), 1123–1134,

[5] E. N. Barron and R. Jensen, *Optimal control and semicontinuous viscosity solutions*. Proc. Amer. Math. Soc. **113** (1991), 397–402.

[6] ———, *Semicontinuous viscosity solutions for Hamilton-Jacobi equations with convex Hamiltonians*. Comm. Partial Diff. Equations **15** (1990), 1713–1742.

[7] S. H. Benton Jr., *The Hamilton-Jacobi Equation: a Global Approach*, Academic Press, New York, 1977.

[8] L. D. Berkovitz, *Characterizations of the values of differential games*. Appl. Math. Optim. **27** (1988), 177-183.

[9] ———, *Optimal feedback controls*. SIAM J. Control Optim. **27** (1989), 991–1007.

[10] N. P. Bhatia and G. P. Szego, *Stability Theory of Dynamical Systems*. Springer-Verlag, Berlin, 1970.

[11] F. Browder, *The fixed point theory of multivalued mappings in topological vector spaces*. Math. Ann. **177** (1968), 283–301.

[12] R. Bulirsh, H.J. Pesh, *The maximum principle, Bellman's equation, and Carathéodory's work*, J. Optim. Theory Appl. **80** (1994) 199–225.

[13] P. Cannarsa and G. Da Prato, *Second-order Hamilton-Jacobi equations in infinite dimensions*. SIAM J. Control. Optim. **29** (1991), 474–492.

[14] P. Cannarsa and H. Frankowska, *Value function and optimality conditions for semilinear control problems*. Appl. Math. Optim. **26** (1992), 139–169.

[15] P. Cannarsa and M. Soner, *Generalized one-sided estimates for solutions of Hamilton-Jacobi equations and applications*. Nonlinear Analysis **13** (1989), 305–323.

[16] F. H. Clarke, *Generalized gradients and applications*. Trans. Amer. Math. Soc. 205 (1975), 247–262.

[17] ———, *Solutions périodiques des équations hamiltoniennes*. Comptes Rendus Acad. Sci. Paris **287** (1978), 951–952.

[18] ———, *Optimal control and the true Hamiltonian*, SIAM Review **21** (1979), 157–166.

[19] ———, *The applicability of the Hamilton-Jacobi verification technique*. In Proceeding of the IFIP Conference (New York 1981), pp. 88–94, System Modeling and Optimization, Springer-Verlag, Berlin, New York, 1982.

[20] ———, *On Hamiltonian flows and symplectic transformations*. SIAM J. Control Optim. 20 (1982), 355–359.

[21] ———, *Hamiltonian trajectories and local minima of the dual action*. Trans. Amer. Math. Soc. **287** (1985), 239–251.

[22] ———, *Perturbed optimal control problems*. IEEE Trans. Automatic Control **31** (1986), 535–542.

[23] ———, *Hamiltonian analysis of the generalized problem of Bolza*. Trans. Amer. Math. Soc. **301** (1987), 385–400.

[24] ———, *Optimization and Nonsmooth Analysis*. Classics in Applied Mathematics, Vol. 5, Society for Industrial and Applied Mathematics, Philadelphia, 1990. (Originally published by Wiley Interscience, New York, 1983.)

[25] F. H. Clarke, *Methods of Dynamic and Nonsmooth Optimization*. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 57, Society for Industrial and Applied Mathematics, Philadelphia, 1989.

[26] ———, *A decoupling principle in the calculus of variations*. J. Math. Analysis Appl. **172** (1993), 92–105.

[27] F. H. Clarke and J.-P. Aubin, *Monotone invariant solutions to differential inclusions*. J. London Math. Soc. (2) **16** (1977), 357-366.

[28] F. H. Clarke and Yu. S. Ledyaev, *Mean value inequalities in Hilbert space*. Trans. Amer. Math. Soc. **344** (1994), 307-324.

[29] ———, *Mean value inequalities*. Proc. Amer. Math. Soc, in press.

[30] F. H. Clarke and P.D. Loewen, *The value function in optimal control: Sensitivity, controllability, and time-optimality*. SIAM J. Control Optim. **24** (1986), 243–263.

[31] F. H. Clarke and R. B. Vinter, *Local optimality conditions and Lipschitzian solutions to the Hamilton-Jacobi equation*, SIAM J. Control Optim. **21** (1983) 856–870.

[32] ———, *The relationship between the maximum principle and dynamic programming*. SIAM J. Control Optim. **25** (1987) 1291–1311.

[33] ______, *Optimal multiprocesses.* SIAM J. Control Optim. **27** (1989), 1072–1091.

[34] F. H. Clarke and P.R. Wolenski, *The sensitivity of optimal control problems to time delay.* SIAM J. Control Optim. **29** (1991), 1176–1214.

[35] ______, *Necessary conditions for functional differential inclusions.* To appear.

[36] ______, *Control of systems to sets and their interiors*, J. Optim. Theory Appl., to appear.

[37] F. H. Clarke, Yu. S. Ledyaev and R. J. Stern, *Fixed points and equilibria in nonconvex sets.* Nonlinear Analysis, to appear.

[38] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern and P. R. Wolenski, *Subdifferential and Nonlinear Analysis.* Monograph in preparation.

[39] F. H. Clarke, Yu. S. Ledyaev and A. I. Subbotin, *Universal feedback controls.* To appear.

[40] F. H. Clarke, Yu. S. Ledyaev and P. R. Wolenski, *Proximal analysis and minimization principles.* To appear.

[41] F. H. Clarke, P. D. Loewen and R. B. Vinter, *Differential inclusions with free time.* Ann. Inst. H. Poincaré, Analyse Nonlinéaire **5** (1988), 573–593.

[42] F. H. Clarke, R. J. Stern and P. R. Wolenski, *Subgradient criteria for monotonicity, the Lipschitz condition, and convexity.* Canadian J. Math. **45** (1993), 1167–1183.

[43] ______, *Proximal smoothness and the lower $-C^2$ property.* To appear.

[44] B. Cornet, *Euler characteristics and fixed-point theorems*, to appear.

[45] M. G. Crandall, H. Ishii and P.-L. Lions, *User's guide to viscosity solutions of second order partial differential equations.* Bull. Amer. Math. Soc. **27** (1992), 1–67.

[46] M. G. Crandall and P.-L. Lions, *Viscosity solutions of Hamilton-Jacobi equations.* Trans. Amer. Math. Soc. **277** (1983), 1–42.

[47] K. Deimling, *Multivalued Differential Equations.* D. Gruyter, Berlin, 1992.

[48] M. A. H. Dempster and J. J. Ye, *, Necessary and sufficient optimality conditions for control of piecewise deterministic Markov processes.* Stochastics and Stochastics Reports **40** (1992), 125–145.

[49] A. F. Filippov, *Differential Equations with Discontinuous Righthand Sides.* Kluwer, 1988.

[50] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions.* Springer-Verlag, New York, 1993.

[51] H. Frankowska, *Hamilton-Jacobi equations: Viscosity solutions and generalized gradients.* J. Math. Anal. Appl. **141** (1989) 21–26.

[52] ______, *Lower semicontinuous solutions of the Hamilton-Jacobi equation.* SIAM J. Control Optim. **31** (1993) 257–272.

[53] H. G. Guseinov, A. I. Subbotin and V. N. Ushakov, *Derivatives for multivalued mappings with applications to game-theoretical problems of control.* Problems of Control Information Theory **14** (1985), 155–167.

[54] Kh. Guseinov and V. N. Ushakov, *Strongly and weakly invariant sets with respect to differential inclusions*, Differencial'nye Uravneniya **26** (1990) 1888–1894.

[55] G. Haddad, *Monotone trajectories of differential inclusions and functional differential inclusions with memory.* Israel J. Math. **39** (1981), 83–100.

[56] U. G. Haussman, *Generalized solutions of the Hamilton-Jacobi equation of stochastic control.* SIAM J. Control Optim. **32** (1994), 728–743.

[57] D. Havelock, *A Generalization of the Hamilton-Jacobi Equation*, Master's thesis (supervisor F. H. Clarke), University of British Columbia, 1977.

[58] A. D Ioffe, *Proximal analysis and approximate subdifferentials.* J. London Math. Soc. **41** (1990), 175–192.

[59] N. N. Krasovskii, *Differential games, approximation and formal models*, Math. USSR Sbornik **35** (1979) 795–822.

[60] ______, *The Control of a Dynamic Systems*, Nauka, Moscow, 1986 (in Russian).

[61] N. N. Krasovskii and A. I. Subbotin, *Positional Differential Games*, Nauka, Moscow, 1974. (in Russian: French translation: *Jeux Différentiels*, Editions MIR, Moscou, 1979; revised English translation: *Game-theoretical Control Problems*, Springer-Verlag, New York, 1988.)

[62] V. F. Krotov, *Global methods in optimal control theory*, in "Advances in Nonlinear Dynamics and Control: a report from Russia", pp. 76–121, vol. 17, Birkhauser, Boston, 1993.

[63] P. D. Loewen, *Parameter sensitivity in stochastic optimal control.* Stochastics **22** (1987), 1–40.

[64] ______, *Optimal Control via Nonsmooth Analysis.* CRM Proceedings and Lecture Notes, Vol. 2, American Mathematical Society, Providence, 1993.

[65] P. D. Loewen and R. T. Rockafellar, *The adjoint arc in nonsmooth optimization.* Trans. Amer. Math. Soc. **325** (1991), 39–72.

[66] ______, *Optimal control of unbounded differential inclusions.* SIAM J. Control Optim. **32** (1994), 442–470.

[67] J. Mawhin and M. Willem, *Critical Point Theory and Hamiltonian Systems.* Springer-Verlag, New York, 1989.

[68] B. Mordukhovich, *Metric approximations and necessary conditions for general classes of nonsmooth extremal problems.* Dokl. Akad. Nauk SSSR **254** (1980), 1072–1076.

[69] ______, *Lipschitzian stability of constraint systems and generalized equations.* Nonlinear Analysis **22** (1994), 173–206.

[70] D. Offin, *A Hamilton-Jacobi Approach to the Differential Inclusion Problem*, Master's thesis (supervisor F. H. Clarke), University of British Columbia, 1979.

[71] R. T. Rockafellar, *Extensions of subgradient calculus with applications to optimization.* Nonlinear Analysis **9** (1985), 665–698.

[72] ______, *Dualization of subgradient conditions for optimality.* Nonlinear Analysis **20** (1993), 627–646.

[73] J. D. L. Rowland and R. B. Vinter, *Construction of optimal feedback controls.* Systems and Control Letters **16** (1991) 357–367.

[74] O. Roxin, *Stability in general control systems*, J. Diff. Equations **1** (1965) 115–150.

[75] D. Shevitz and B. Paden, *Lyapounov stability theory of nonsmooth systems*, IEEE Trans. Automatic Control **39** (1994) 1910–1913.

[76] A. I. Subbotin, *A generalization of the basic equation of the theory of differential games.* Soviet Math. Dokl. **22** (1980), 358–362.

[77] ______, *Generalization of the main equation of differential game theory.* J. Optim. Theory Appl. **43** (1984), 103–133.

[78] ______, *On a property of the subdifferential*, Mat. Sbornik **182** (1991) 1315–1330. (English translation: Math. USSR Sbornik **74** (1993) 63–78.)

[79] ______, *Minimax Inequalities and Hamilton-Jacobi Equations.* Nauka, Moscow, 1991. (English translation to appear)

[80] ______, *Continuous and discontinuous solutions of boundary-value problems for first-order partial differential equations.* Doklady Ross. A.N. **323** (1992) (in Russian).

[81] ______, *Viable characteristics of Hamilton-Jacobi equations.* To appear.

[82] N. N. Subbotina, *The maximum principle and the superdifferential of the value function.* Problems of Control and Information Theory **18** (1989), 151-160.

[83] R. B. Vinter, *New results on the relationship between dynamic programming and the maximum principle*. Math. Control Signals Systems **1** (1988), 97–105.

[84] ______, *Uniqueness of solutions to the Hamilton-Jacobi equation: a system theoretic proof*. System Control Lett., to appear.

[85] ______, *Convex duality and nonlinear optimal control*. SIAM J. Control Optim. **31** (1993), 518–538.

[86] R. Yue, *On the properties of Bellman's function in time optimal control problems*, to appear.

[87] V. M. Zeidan, *Sufficiency conditions with minimal regularity assumptions*. Appl. Math. Optim. **20** (1989) 19–31.