International Institute for
Applied Systems Analysis
IIASA www.iiasa.ac.at

Some Characterizations of Optimal Trajectories in Control Theory

Cannarsa, P. & Frankowska, H.

IIASA Working Paper

WP-89-083

November 1989

WORKING PAPER

SOME CHARACTERIZATIONS OF OPTIMAL TRAJECTORIES IN CONTROL THEORY

Piermarco Cannarsa
Halina Frankowska

November 1989
WP-89-083

IIASA
International Institute
for Applied Systems Analysis

SOME CHARACTERIZATIONS OF OPTIMAL TRAJECTORIES IN CONTROL THEORY

Piermarco Cannarsa [1)]
Halina Frankowska [2)]

November 1989
WP-89-083

1) Dipartimento di Matematica, Universitá di Pisa, via F. Buonarroti 2, Pisa, Italy

2) Ceremade, Université de Paris-Dauphine, Paris, France & International Institute for Applied Systems Analysis

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

The authors provide several characterizations of optimal trajectories for the classical Meyer problem arising in optimal control. For this purpose they study the regularity of directional derivatives of the value function: for instance it is shown that for smooth control systems the value function V is continuously differentiable along an optimal trajectory $x : [t_0, 1] \to \mathbf{R}^n$ provided V is differentiable at the initial point $(t_0, x(t_0))$. Then they deduce the upper semicontinuity of the optimal feedback map and address the problem of optimal design, obtaining sufficient conditions for optimality. Finally it is shown that the optimal control problem may be reduced to a viability problem.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

Contents

Some characterizations of optimal trajectories in control theory

Piermarco Cannarsa
&
Halina Frankowska

Introduction

Consider the optimal control problem

$$\text{minimize } g(x(1))$$

over all solutions of the control system

$$x' = f(t, x, u(t)), \ \ u(t) \in U \tag{1}$$

satisfying the initial condition

$$x(0) = \xi_0. \tag{2}$$

We recall that by a simple change of variables the classical Bolza problem in control theory

$$\text{minimize} \left\{ \varphi(x(1)) + \int_0^1 L(t, x(t), u(t))dt \right\}$$

over the trajectory-control pairs (x, u) of (1), (2) may be reduced to the one under consideration.

The goal of the optimal control theory is to find necessary and sufficient conditions for optimality and to construct optimal trajectories. While several results establishing necessary conditions are available in the form of maximum principle, it is difficult to complete these conditions to sufficient ones. In this paper we show that this additional information may be obtained from some properties of the value function defined by

$$V(t_0, x_0) = \inf\{ g(x(1)) \mid x \text{ is a solution of (1) on } [t_0, 1], \ x(t_0) = x_0 \}$$

When the data of the problem are Lipschitz, then the value function is locally Lipschitz. When it is differentiable it satisfies the Hamilton-Jacobi equation:

$$-\frac{\partial V}{\partial t}(t,x) + H\left(t,x,-\frac{\partial V}{\partial x}(t,x)\right) = 0, \quad V(1,\cdot) = g(\cdot) \tag{3}$$

where the Hamiltonian H is defined by

$$H(t,x,p) = \sup_{u\in U} \langle p, f(t,x,u)\rangle$$

In general, even in very regular situations, the value function is not differentiable. But still it solves the Hamilton-Jacobi equation (3) in the viscosity sense (see [12], [13]).

Furthermore V can be characterized as the unique viscosity solution of (3). So it inherits many qualitative properties of this class of solutions, such as stability and comparison theorems and also enjoys some numerical advantages (see for instance [9]).

Although, as we have just recalled, the value function is not necessarily differentiable, we prove in this paper that the differentiability of V is preserved along optimal trajectories. More precisely, we show that if V is differentiable at some point (t_0, x_0) and $\overline{x}$ denotes any optimal solution starting from x_0 at time t_0, then for every $t \in [t_0, 1]$, V is differentiable at $(t, \overline{x}(t))$ (see Corollary 4.3). Actually, the derivative $-V_x'(t,\overline{x}(t))$ is equal to the co-state of the Pontriagin maximum principle, which we recall in Section 3.

The value function is also a good tool to characterize optimal trajectories. It is well known that V is nondecreasing along trajectories of (1) and is constant along optimal trajectories.

When the Hamiltonian H is smooth enough and the value function is differentiable at $(0,\xi_0)$, then the following necessary and sufficient condition for optimality holds true (Lemma 3.5):

Let $x(\cdot)$, $p(\cdot)$ solve the Hamiltonian system

$$\begin{cases} x'(t) = \frac{\partial H}{\partial p}(t,x(t),p(t)) \\ p'(t) = -\frac{\partial H}{\partial x}(t,x(t),p(t)), \quad t \in [0,1] \end{cases} \tag{4}$$

Then x is optimal if and only if $x(0) = \xi_0$, $p(0) = -V_x'(0,\xi_0)$ (see Theorem 3.4 for a more general statement.)

Even when the Hamiltonian is not smooth, the value function may still be used to construct the optimal feedback map:

$$G(t,x) = \left\{ v \in f(t,x,U) \mid \frac{\partial V}{\partial(1,v)}(t,x) = 0 \right\} \tag{5}$$

Namely the following property holds true: a trajectory $\overline{x}$ of (1) is optimal for our optimization problem if and only if it is a solution of the differential inclusion

$$(6) \qquad x' \in G(t,x), \; x(0) = \xi_0$$

We refer to [16], [4] for some developments in this direction.

To investigate regularity properties of the set-valued map G we prove the existence of the directional derivatives of V. For this aim we show that under very general assumptions on the control system the value function is semiconcave (see Theorem 4.1).

As a consequence of the semiconcavity of V, we obtain that the feedback map G is upper semicontinuous and has nonempty compact images (see Theorem 5.1).

In particular whenever the feedback map G is single-valued, it is continuous. From the above it follows that in this particular case optimal trajectories are continuously differentiable.

Moreover if the data are convex, then G has convex values and the inclusion (6) fits the well investigated framework of upper semicontinuous convex valued maps. In particular solutions of (6) can be obtained as limits of Euler curves.

When the map G does not have convex images the above characterization of optimal trajectories is not easy to apply. To overcome this difficulty we provide an alternative approach based on viability theory.

Namely we observe that solving the optimal control problem is equivalent to solving a control system with state constraints:

$$\begin{cases} i) & t' = 1 \\ ii) & x' = f(t,x,u), \; u \in U \\ iii) & z' = 0 \\ iv) & (t, x(t), z(t)) \in \text{Graph}\,(V) \\ v) & t(0) = 0, \; x(0) = \xi_0, \; z(0) = V(0,\xi_0) \end{cases}$$

The last problem is a viability one and may be approached using many results of viability theory (see [19], [2], [1] and bibliographies contained therein). In particular solutions of such system can be constructed using Euler curves. We underline that in this case dynamics $i) - iii)$ remain regular, but we have to keep trajectories in the set Graph(V) according to the relation $iv)$.

Finally, we treat the case involving the end point constraints $(x(1) \in K_1)$ via penalization techniques. We show that the value function of such problem may be approximated by the value function of problems with free end points (see Theorem 7.1). A result of the same nature holds true for optimal trajectories.

The plan of the paper is as follows. Section 1 contains basic material on the value function. In Section 2 we recall some definitions of set-valued gradients and investigate properties of semiconcave functions. Necessary and sufficient conditions for optimality are described in Section 3, while Section 4 is devoted to the semiconcavity of the value function. The optimal feedback map is studied in Section 5 and viability theory is applied to optimal trajectories in Section 6. In Section 7 we address the problem with end point constraints.

1 Value function in optimal control

Consider a complete separable metric space U and a continuous function

$$f : [0,1] \times \mathbf{R}^n \times U \rightarrow \mathbf{R}^n$$

We associate with it the control system

$$(7) \qquad x'(t) = f(t, x(t), u(t)), \quad u(t) \in U \text{ almost everywhere}$$

An absolutely continuous function $x : [t_0, t_1] \rightarrow \mathbf{R}^n$ is called a trajectory of (7) if there exists a measurable function $u : [t_0, t_1] \rightarrow U$ such that $x'(t) = f(t, x(t), u(t))$ almost everywhere in $[t_0, t_1]$.

Throughout the whole paper we impose the following assumptions on f

$$(8) \begin{cases} i) & \exists\, k \in L^1(0,1;\mathbf{R}_+), \ \forall\, (t,u) \in [0,1] \times U, \ f(t,\cdot,u) \text{ is } k(t) - \text{Lipschitz} \\ ii) & \exists\, \gamma > 0 \text{ such that } \forall\, (t,u) \in [0,1] \times U, \ \|f(t,x,u)\| \leq \gamma(\|x\| + 1) \end{cases}$$

Let $g : \mathbf{R}^n \rightarrow \mathbf{R}$ be a locally Lipschitz function and $\xi_0 \in \mathbf{R}^n$ be given.

We investigate the minimization problem

$$(9) \qquad \text{minimize } \{g(x(1)) \mid x \text{ is a solution of (7) on } [0,1], \ x(0) = \xi_0\}$$

The dynamic programming approach associates with this problem the value function defined by

$$(10) \quad V(t_0, x_0) = \inf \{g(x(1)) \mid x \text{ is a solution of (7) on } [t_0, 1], \ x(t_0) = x_0\}$$

Our assumptions allow to apply the relaxation theorem from [2] to show that V is actually equal to the value function of the relaxed problem:

Consider the convexified differential inclusion

$$(11) \qquad x'(t) \in \overline{co} f(t, x(t), U) \text{ almost everywhere}$$

We recall that an absolutely continuous function $x : [t_0, t_1] \mapsto \mathbf{R}^n$ is called a trajectory of (11) if for almost every $t \in [t_0, t_1]$, $x'(t) \in \overline{co} f(t, x(t), U)$. We associate with (11) the following minimization problem

$$\text{(12)} \qquad \text{minimize } \{g(x(1)) \mid x \text{ is a solution of (11) on } [0,1],\ x(0) = \xi_0\}$$

The corresponding value function is given by

$$V^{co}(t_0, x_0) \ = \ \inf \{g(x(1)) \mid x \text{ is a solution of (11) on } [t_0, 1],\ x(t_0) = x_0\}$$

Theorem 1.1 *For all* $(t_0, x_0) \in [0,1] \times \mathbf{R}^n$ *we have*

$$V(t_0, x_0) = V^{co}(t_0, x_0) = \min\{g(x(1)) \mid x \text{ is a solution of (11) on } [t_0, 1],\ x(t_0) = x_0\}$$

Proof — From the relaxation theorem (see [2]) and the parametrization theorem [2] we know that the closure in the metric of uniform convergence of trajectories of (7) defined on the time interval $[t_0, 1]$ is equal to the set of trajectories of (11) defined on $[t_0, 1]$. This ends the proof. □

It is well known that the value function is nondecreasing along trajectories of (7) and therefore a trajectory $x : [t_0, 1] \to \mathbf{R}^n$ satisfies $V(t_0, x(t_0)) = g(x(1))$ if and only if $V(t, x(t)) \equiv g(x(1))$. This leads to a verification technique in optimal control:

A trajectory $x : [0,1] \to \mathbf{R}^n$ of the control system (7) is optimal for the problem (9) if and only if $x(0) = \xi_0$ and $V(t, x(t)) \equiv const$ (in this case $V(t, x(t)) \equiv g(x(1))$)

Hence instead of looking for an optimal trajectory for the problem (9) one can search a trajectory of (7) satisfying the initial condition and such that the value function is constant along it.

We recall that the directional derivative of a function $\varphi : \mathbf{R}^n \to \mathbf{R}$ at $x_0 \in X$ in the direction $\Theta \in X$ (when it exists) is defined by

$$\frac{\partial \varphi}{\partial \Theta}(x_0) \ = \ \lim_{h \to 0+} \frac{\varphi(x_0 + h\Theta) - \varphi(x_0)}{h}$$

Proposition 1.2 *The value function V is locally Lipschitz. Furthermore for every trajectory x of (7) on $[0,1]$ and for almost every $t \in [0,1]$ there exists the directional derivative*

$$\frac{\partial V}{\partial (1, x'(t))}(t, x(t))$$

Proof — Local Lipschitz continuity of V is a well known result. It can be checked by the arguments similar to [15, Theorem 4.2, p.85] (see also [16]).

Fix a trajectory $x(\cdot)$. Then the function $t \to \varphi(t) := V(t, x(t))$ is absolutely continuous. Fix t such that φ and x are differentiable at t. Then

$$\lim_{h \to 0+} \frac{V(t+h, x(t) + hx'(t)) - V(t, x(t))}{h} = \lim_{h \to 0+} \frac{V(t+h, x(t+h)) - V(t, x(t))}{h}$$

and the proof follows. □

When the value function is differentiable it has many properties related to dynamics of system.

For instance

Proposition 1.3 *If for some* $(t_0, x_0) \in [0, 1[\times \mathbf{R}^n$ *and* $v \in \overline{co} f(t_0, x_0, U)$, V *has the directional derivative at* (t_0, x_0) *in the direction* $(1, v)$ *then this directional derivative is nonnegative.*

Proof — Consider a solution $x(\cdot)$ of the differential inclusion (11) satisfying $x(t_0) = x_0$, $x'(t_0) = v$ (by [2] such solution does exist). Since V is locally Lipschitz at (t_0, x_0) and nondecreasing along trajectories of (11), thanks to Theorem 1.1, we obtain

$$\lim_{h \to 0+} \frac{V(t_0+h, x_0+hv) - V(t_0, x_0)}{h} = \lim_{h \to 0+} \frac{V(t_0+h, x(t_0+h)) - V(t_0, x_0)}{h} \geq 0 \quad \square$$

Unfortunately in the great majority of cases the value function is not differentiable and many attempts to overcome this difficulty recently appeared in the literature (see [12], [13], [4], [16] and bibliographies contained therein). In the Section 4 we provide sufficient conditions for the value function V to have directional derivatives in all directions.

To characterize optimal trajectories we introduce two following feedback maps $G : [0, 1] \times \mathbf{R}^n \rightsquigarrow \mathbf{R}^n$ and $G^{co} : [0, 1] \times \mathbf{R}^n \rightsquigarrow \mathbf{R}^n$ defined respectively by

$$G(t, x) = \left\{ v \in f(t, x, U) \mid \frac{\partial V}{\partial(1, v)}(t, x) = 0 \right\}$$

and

$$G^{co}(t, x) = \left\{ v \in \overline{co} f(t, x, U) \mid \frac{\partial V}{\partial(1, v)}(t, x) = 0 \right\}$$

Then we have the following characterizations of optimal trajectories:

Theorem 1.4 *The following two statements are equivalent:*

i) x *is a trajectory of the differential inclusion*

$$x' \in G(t, x) \tag{13}$$

defined on the time-interval $[t_0, 1]$.

ii) x is a trajectory of the control system (7) defined on the time-interval $[t_0, 1]$ *and for every* $t \in [t_0, 1]$, $V(t, x(t)) = g(x(1))$.

For the relaxed system (11) the following two statements are equivalent:

iii) x is a trajectory of the differential inclusion

$$x' \in G^{co}(t, x) \tag{14}$$

defined on the time-interval $[t_0, 1]$.

iv) x is a trajectory of the differential inclusion (11) defined on the time-interval $[t_0, 1]$ *and for every* $t \in [t_0, 1]$, $V(t, x(t)) = g(x(1))$.

Proof — Fix a trajectory x of (7) defined on time interval $[t_0, 1]$ and set $\varphi(t) = V(t, x(t))$ for every $t \in [t_0, 1]$. From Proposition 1.2 for almost all $t \in [t_0, 1]$

$$\varphi'(t) = \frac{\partial V}{\partial(1, x'(t))}(t, x(t))$$

Assume that *i*) holds true. Thus $\varphi'(t) = 0$ almost everywhere in $[t_0, 1]$. Consequently φ is constant equal to $V(1, x(1)) = g(x(1))$. Assume next that *ii*) holds true. Then, differentiating the map $t \to \varphi(t)$, we obtain that for every $t \in [t_0, 1[$, $\varphi'(t) = 0$. Thus

$$\frac{\partial V}{\partial(1, x'(t))}(t, x(t)) = 0$$

almost everywhere and therefore for almost all $t \in [t_0, 1]$, $x'(t) \in G(t, x(t))$. The proof of the second statement is analogous and is omitted. □

Corollary 1.5 *A trajectory $x : [0, 1] \to \mathbf{R}^n$ is an optimal solution of the optimal control problem (9) if and only if it is a solution of the differential inclusion (13) and $x(0) = \xi_0$. An analogous statement holds true for the relaxed problem (12) and the differential inclusion (14).*

Proof — Since V is nondecreasing along trajectories of the control system (7) we deduce that $\overline{x}(\cdot)$ is optimal for the control problem (9) if and only if V is constant along $\overline{x}$. Theorem 1.4 ends the proof. □

Theorem 1.6 *For every $t_0 \in [0, 1]$, $x_0 \in \mathbf{R}^n$ inclusion (14) has at least one solution satisfying $x(t_0) = x_0$.*

Proof — Consider the optimal control problem

$$\text{minimize } g(x(1))$$

over the solutions of the differential inclusion

$$x'(t) \in \overline{co} f(t, x(t), U) \text{ a.e. in } [t_0, 1], \quad x(t_0) = x_0$$

By Theorem 1.1 it has at least one optimal solution $\overline{x}$. Furthermore $V(t, \overline{x}(t)) \equiv g(\overline{x}(1))$. Theorem 1.4 ends the proof. □

The map G introduced above, in general, does not enjoy any regularity properties and this is why it is difficult to obtain solutions of the differential inclusion (13). In Sections 4 and 5 we provide some sufficient conditions for upper semicontinuity of G and in section 6 we reduce the problem to a problem with state constraints. The advantage of this approach lies in the possibility to exploit results of viability theory and, in particular, to get solutions of (14) as limits of Euler curves.

2 Some preliminaries on nonsmooth functions

Consider an open set $\Omega \subset \mathbf{R}^n$ and a function $\varphi : \Omega \to \mathbf{R}$. When it is not differentiable it is possible to define its gradient taking weaker limits of differential quotients.

Definition 2.1 *Let $x_0 \in \Omega$. The* superdifferential *of φ at x_0 is the closed convex set defined as follows:*

$$D^+\varphi(x_0) = \left\{ p \in \mathbf{R}^n \mid \limsup_{x \to x_0} \frac{\varphi(x) - \varphi(x_0) - \langle p, x - x_0 \rangle}{\|x - x_0\|} \leq 0 \right\}$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product.

The subdifferential *is defined in a similar way:*

$$D^-\varphi(x_0) = \left\{ p \in \mathbf{R}^n \mid \liminf_{x \to x_0} \frac{\varphi(x) - \varphi(x_0) - \langle p, x - x_0 \rangle}{\|x - x_0\|} \geq 0 \right\}$$

It is not difficult to show that φ is Fréchet differentiable at x_0 if and only if both super and subdifferentials are not empty at x_0. In this case

$$D^+\varphi(x_0) = D^-\varphi(x_0) = \{ \varphi'(x_0) \}$$

We always have $D^+\varphi(x_0) = -D^-(-\varphi)(x_0)$.

The super and subdifferential may also be characterized using *the Dini* directional derivatives, which are defined in the following way:

Definition 2.2 *The* lower Dini derivative *of φ at x_0 in the direction Θ is given by*

$$\partial^-\varphi(x_0)(\Theta) \;=\; \liminf_{h\to 0+,\ \Theta'\to\Theta} \frac{\varphi(x_0+h\Theta')-\varphi(x_0)}{h}$$

and the upper Dini derivative *of φ at x_0 in the direction Θ is defined by*

$$\partial^+\varphi(x_0)(\Theta) \;=\; \limsup_{h\to 0+,\ \Theta'\to\Theta} \frac{\varphi(x_0+h\Theta')-\varphi(x_0)}{h} \tag{15}$$

Clearly

$$\partial^-\varphi(x_0) = -\partial^+(-\varphi)(x_0) \tag{16}$$

When φ is Lipschitz at x_0 then the definition may be simplified as follows

$$\partial^-\varphi(x_0)(\Theta) \;=\; \liminf_{h\to 0+} \frac{\varphi(x_0+h\Theta)-\varphi(x_0)}{h}$$

and

$$\partial^+\varphi(x_0)(\Theta) \;=\; \limsup_{h\to 0+} \frac{\varphi(x_0+h\Theta)-\varphi(x_0)}{h}$$

From [16, Lemma 2.7] we know that

$$D^-\varphi(x_0) \;=\; \{\, p\in\mathbf{R}^n \mid\ \forall\,\Theta\in\mathbf{R}^n,\ \partial^-\varphi(x_0)(\Theta) \;\geq\; <p,\Theta>\,\}$$

and

$$D^+\varphi(x_0) \;=\; \{\, p\in\mathbf{R}^n \mid\ \forall\,\Theta\ \in\ \mathbf{R}^n,\ \partial^+\varphi(x_0)(\Theta) \;\leq\; <p,\Theta>\,\} \tag{17}$$

Definition 2.3 *Assume that φ is Lipschitz at $x_0\in\Omega$. The* regularized lower derivative *of φ at x_0 in the direction $\Theta\in\mathbf{R}^n$ is defined by*

$$\varphi^o_-(x_0,\Theta) \;=\; \liminf_{h\to 0+,\ x\to x_0} \frac{\varphi(x+h\Theta)-\varphi(x)}{h}$$

This notion is a "lower version" of Clarke's definition of directional derivative. Indeed it can be easily checked that

$$\varphi^o_-(x_0,\Theta) = -\varphi^o(x_0,-\Theta) = -(-\varphi)^o(x_0,\Theta) \tag{18}$$

where $\varphi^o(x_0,\Theta)$ denotes the directional derivative from [10].

Proposition 2.4 *Let $\varphi:\mathbf{R}^n\to\mathbf{R}$ be Lipschitz at $x_0\in\mathbf{R}^n$. Then the function $\Theta\to\varphi^o_-(x_0,\Theta)$ is concave.*

This result may be deduced from [10, Proposition 2.1.1].

We investigate next the closedness of the level sets of the regularized lower derivative.

Proposition 2.5 *Let $\varphi : \mathbf{R}^n \to \mathbf{R}$ be a locally Lipschitz function and define the set-valued map $Q : \mathbf{R}^n \leadsto \mathbf{R}^n$ by*

$$Q(x) \;=\; \{\, \Theta \in \mathbf{R}^n \mid \varphi^o_-(x,\Theta) \;\le\; 0 \,\}$$

Then Q has nonempty closed images and the graph of the map Q is closed.

Proof — Clearly for every x, $0 \in Q(x)$. It remains to show that for every sequence $(x_n, \Theta_n) \in \mathbf{R}^n \times \mathbf{R}^n$ converging to some (x, Θ) and satisfying $\Theta_n \in Q(x_n)$ we have $\Theta \in Q(x)$. Fix such a sequence and let $\varepsilon_n \to 0$. By the definition of $\varphi^o_-(x_n, \Theta_n)$ there exist $h_n \to 0+$, $x'_n \to x$ be such that for every n

$$\frac{\varphi(x'_n + h_n\Theta_n) - \varphi(x'_n)}{h_n} \;\le\; \varepsilon_n$$

Consequently

$$\varphi^o_-(x,\Theta) \;\le\; \liminf_{n\to\infty} \frac{\varphi(x'_n + h_n\Theta_n) - \varphi(x'_n)}{h_n} \;\le\; 0$$

This ends the proof. □

Definition 2.6 *Assume that φ is Lipschitz at $x_0 \in \Omega$. The* generalized gradient *of φ at x_0 is defined by*

$$(19) \qquad \partial\varphi(x_0) \;=\; \{\, p \in \mathbf{R}^n \mid \; \forall\, \Theta \in \mathbf{R}^n, \; \varphi^o_-(x_0, \Theta) \;\le\; < p, \Theta > \,\}$$

We denote by $D^\star\varphi(x_0)$ the set of all cluster points of gradients $\varphi'(x_n)$ when x_n converge to x_0:

$$D^\star\varphi(x_0) \;=\; \{\, \lim_{n\to\infty} \varphi'(x_n) \mid x_n \to x_0 \;\;\&\;\; \varphi'(x_n) \text{ does exist and is converging} \,\}$$

In view of (18) the above definition of the generalized gradient is equivalent to the one given by Clarke.

It is clear that $D^\star\varphi(x_0)$ is compact. From [10, Theorem 2.5.1] follows that

$$(20) \qquad \partial\varphi(x_0) \;=\; co(D^\star\varphi(x_0))$$

where co denotes the convex hull.

Let us denote by B the closed unit ball in $\mathbf{R}^n$.

Definition 2.7 *Consider a convex subset K of $\mathbf{R}^n$ and a function $\varphi : K \to \mathbf{R}$. It is called semiconcave if there exists a function $\omega : \mathbf{R}_+ \times \mathbf{R}_+ \to \mathbf{R}_+$ such that*

$$(21) \qquad \forall\, r \le R,\ t \le T,\ \ \omega(r,t) \ \le\ \omega(R,T)\ \ \&\ \ \forall\, R > 0\ \lim_{t\to 0+} \omega(R,t) \ =\ 0$$

and for every $R > 0$, $\lambda \in [0,1]$ and any points x, $y \in K \cap RB$

$$\lambda\varphi(x) \ +\ (1-\lambda)\varphi(y) \ \le\ \varphi(\lambda x \ +\ (1-\lambda)y) \ +\ \lambda(1-\lambda)\|x-y\|\omega(R,\|x-y\|)$$

We say that φ is semiconcave at x_0 if there exists a neighborhood of x_0 such that the restriction of φ to it is semiconcave.

We call the above function ω a modulus of semiconcavity of φ.

Usually in the definition of semiconcavity $\omega(r,t) = ct$ for a nonnegative constant c (see [20], [21]), or $\omega(r,t) = ct^\alpha$ for $c \ge 0$ and $\alpha \in]0,1]$ ([7]). We observe that every concave function $\varphi : K \to \mathbf{R}$ is semiconcave (with ω equal to zero). Furthermore

Proposition 2.8 *Let K be a convex subset of $\mathbf{R}^n$ and $\varphi : \mathbf{R}^n \to \mathbf{R}$ be continuously differentiable on a neighborhood of K. Then φ is semi-concave.*

This is a well known result, we provide its proof for the seek of completeness.

Proof — Fix $R > 0$, x, $y \in K \cap RB$ and $\lambda \in [0,1]$. From the mean value theorem there exist t, $t_1 \in [0,1]$ such that

$$\varphi(\lambda x \ +\ (1-\lambda)y) \ =\ \varphi(x) \ +\ \varphi'(x + t(1-\lambda)(y-x))(1-\lambda)(y-x)$$

and

$$\varphi(\lambda x \ +\ (1-\lambda)y) \ =\ \varphi(y) \ +\ \varphi'(y + t_1\lambda(x-y))\lambda(x-y)$$

Multiplying the above equalities by λ and $(1-\lambda)$ respectively and adding them yields

$$\begin{aligned}&\lambda\varphi(x) \ +\ (1-\lambda)\varphi(y) \ \le\ \varphi(\lambda x \ +\ (1-\lambda)y)\\ &+\ (\varphi'(x + t(1-\lambda)(y-x)) - \varphi'(y + t_1\lambda(x-y)))\,\lambda(1-\lambda)(x-y)\end{aligned}$$

Then taking $\omega(R,\cdot)$ equal to the modulus of continuity of φ' over $K \cap RB$ we end the proof. □

Example 1. Consider a subset K of $\mathbf{R}^n$ and let $dist(x,K)$ denote the distance from a point $x \in \mathbf{R}^n$ to K. Define the function $\varphi : \mathbf{R}^n \to \mathbf{R}_+$ by $\varphi(x) = dist(x,K)^2$. We claim that φ is semiconcave.

Indeed fix $x, y \in \mathbf{R}^n$, $\lambda \in [0,1]$ and set $x_\lambda = \lambda x + (1-\lambda)y$. Let $a \in \overline{K}$ (the closure of K) be such that $\|x_\lambda - a\| = dist(x_\lambda, K)$. Then

$$(22) \qquad \begin{cases} \varphi(x_\lambda) = \|\lambda(x-a) + (1-\lambda)(y-a)\|^2 \\ = \lambda^2 \|x-a\|^2 + (1-\lambda)^2 \|y-a\|^2 + 2\lambda(1-\lambda) < x-a, y-a > \end{cases}$$

On the other hand

$$\|x-y\|^2 = \|x-a\|^2 + \|y-a\|^2 - 2 < x-a, y-a >$$

Hence

$$2\lambda(1-\lambda) < x-a, y-a > = \lambda(1-\lambda)\left(\|x-a\|^2 + \|y-a\|^2 - \|x-y\|^2\right)$$

This and (22) imply

$$\begin{aligned} \varphi(x_\lambda) &= \lambda \|x-a\|^2 + (1-\lambda)\|y-a\|^2 - \lambda(1-\lambda)\|x-y\|^2 \\ &\geq \lambda\varphi(x) + (1-\lambda)\varphi(y) - \lambda(1-\lambda)\|x-y\|^2 \end{aligned}$$

Consequently φ is semiconcave. □

In general a Lipschitz function does not have directional derivatives. Our next aim is to show that for a semi-concave at x_0 function φ the directional derivatives exist and coincide with regularized lower derivatives. This result was proved in [7], [8]. We provide a different proof of this fact for the seek of completeness.

Theorem 2.9 *Let $x_0 \in \mathbf{R}^n$ and $\varphi : \mathbf{R}^n \to \mathbf{R}$ be Lipschitz and semiconcave at x_0. Then for every $\Theta \in \mathbf{R}^n$ the directional derivative $\frac{\partial \varphi}{\partial \Theta}(x_0)$ exists and is equal to the regularized lower derivative $\varphi^o_-(x_0, \Theta)$:*

$$(23) \qquad \forall\, \Theta \in \mathbf{R}^n, \ \frac{\partial \varphi}{\partial \Theta}(x_0) = \varphi^o_-(x_0, \Theta)$$

Consequently $D^+\varphi(x_0) \neq \emptyset$ and

$$(24) \qquad D^+\varphi(x_0) = \partial\varphi(x_0) = co(D^\star\varphi(x_0))$$

Proof — It is enough to consider the case $\|\Theta\| \leq 1$. Let $\delta > 0$ be such that φ is semiconcave on $B_{2\delta}(x_0)$ with semiconcavity modulus $\omega(\cdot) := \omega(2\delta, \cdot)$. Fix $x \in B_\delta(x_0)$, $\Theta \in B$ and observe that for all $0 < h_1 \leq h_2 \leq \delta$ we have

$$\begin{cases} \varphi(x + h_1\Theta) - \varphi(x) = \varphi\left(\frac{h_1}{h_2}(x + h_2\Theta) + \left(1 - \frac{h_1}{h_2}\right)x\right) - \varphi(x) \\ \geq \frac{h_1}{h_2}\varphi(x + h_2\Theta) + \left(1 - \frac{h_1}{h_2}\right)\varphi(x) - \varphi(x) - \frac{h_1}{h_2}\left(1 - \frac{h_1}{h_2}\right)h_2\|\Theta\|\,\omega(h_2\|\Theta\|) \\ = \frac{h_1}{h_2}\varphi(x + h_2\Theta) - \frac{h_1}{h_2}\varphi(x) - h_1\left(1 - \frac{h_1}{h_2}\right)\|\Theta\|\,\omega(h_2\|\Theta\|) \end{cases}$$

Consequently for all $0 < h_1 \leq h_2 \leq \delta$

$$\frac{\varphi(x+h_1\Theta)-\varphi(x)}{h_1} \geq \frac{\varphi(x+h_2\Theta)-\varphi(x)}{h_2} - \left(1-\frac{h_1}{h_2}\right)\omega(h_2\,\|\Theta\|)$$

and we proved that for every $x \in B_\delta(x_0)$

$$(25)\ \ \forall\, 0 < h' \leq h \leq \delta,\ \ \frac{\varphi(x+h'\Theta)-\varphi(x)}{h'} \geq \frac{\varphi(x+h\Theta)-\varphi(x)}{h} - \omega(h\,\|\Theta\|)$$

Thus for every $0 < h \leq \delta$

$$\liminf_{h'\to 0+} \frac{\varphi(x_0+h'\Theta)-\varphi(x_0)}{h'} \geq \frac{\varphi(x_0+h\Theta)-\varphi(x_0)}{h} - \omega(h\,\|\Theta\|)$$

Taking $\limsup_{h\to 0+}$ in the right-hand side of the above inequality yields that the directional derivative $\frac{\partial\varphi}{\partial\Theta}(x_0)$ does exist. Clearly $\frac{\partial\varphi}{\partial\Theta}(x_0) \geq \varphi^o_-(x_0,\Theta)$. To prove the opposite fix $\varepsilon > 0$ and $0 < \lambda < \delta$. From the continuity of φ it follows that there exists $0 < \alpha < \delta$ such that for all $x \in B_\alpha(x_0)$

$$\frac{\varphi(x_0+\lambda\Theta)-\varphi(x_0)}{\lambda} \leq \frac{\varphi(x+\lambda\Theta)-\varphi(x)}{\lambda} + \varepsilon$$

Thus, using (25), we obtain that

$$\frac{\varphi(x_0+\lambda\Theta)-\varphi(x_0)}{\lambda} \leq \inf_{x\in B_\alpha(x_0),\ h\in]0,\lambda]} \frac{\varphi(x+h\Theta)-\varphi(x)}{h} + \omega(\lambda\,\|\Theta\|) + \varepsilon$$

Letting ε, α and λ converge to zero we end the proof of the first statement. The second one results from (23) recalling (17), (19) and (20). □

Proposition 2.10 *Let $\varphi : \mathbf{R}^n \to \mathbf{R}$ be Lipschitz and semiconcave at x_0. If $D^+\varphi(x_0)$ is a singleton then φ is differentiable at x_0 and*

$$D^*\varphi(x_0) = \{\ \varphi'(x_0)\ \}$$

In particular, if $D^+\varphi(x)$ is a singleton for all x near x_0, then φ is continuously differentiable at x_0.

The proof follows by exactly the same arguments as the ones in [7, Corollaries 4.11, 4.12]

Definition 2.11 *Let $K \subset \mathbf{R}^n$ be convex and $\varphi : K \to \mathbf{R}$ be given. It is called semiconvex (respectively semiconvex at x_0) whenever $-\varphi$ is semiconcave (respectively semiconcave at x_0).*

Proposition 2.12 *Let $\varphi : \mathbf{R}^n \to \mathbf{R}$, $x_0 \in \mathbf{R}^n$. If φ is Lipschitz at x_0 and both semiconvex and semiconcave at x_0, then φ is continuously differentiable on a neighborhood of x_0.*

Proof — Since φ and $-\varphi$ are semiconcave at x_0, by Theorem 2.9, there exists a neighborhood $\mathcal{N}$ of x_0 such that for all $x \in \mathcal{N}$

$$D^+\varphi(x) = \partial\varphi(x), \quad D^+(-\varphi)(x) = \partial(-\varphi)(x)$$

Furthermore

$$D^-\varphi(x) = -D^+(-\varphi)(x) = -\partial(-\varphi)(x) = \partial\varphi(x)$$

the last equality being a straightforward consequence of (20). Hence both $D^+\varphi(x)$ and $D^-\varphi(x)$ are nonempty and therefore φ is differentiable on $\mathcal{N}$. The conclusion follows from Proposition 2.10. □

3 Necessary and sufficient conditions for optimality

We provide next a sufficient condition for optimality which involves the superdifferential defined in the previous section.

Theorem 3.1 *Assume that (8) hold true and let $\overline{x} : [0, 1] \to \mathbf{R}^n$ be a solution of the control system (7), $\overline{x}(0) = \xi_0$ and $\overline{u}$ be a corresponding control. If for almost every $t \in [0, 1]$ there exists $p(t) \in \mathbf{R}^n$ such that*

$$(26) \qquad (< p(t), \overline{x}'(t) >, \ -p(t)) \in D^+V(t, \overline{x}(t))$$

then $\overline{x}$ is optimal for the problem (9).

Proof — Consider the absolutely continuous function $\psi(t) = V(t, \overline{x}(t))$ and let $t \in [0, 1]$ be such that the derivatives $\psi'(t)$ and $\overline{x}'(t)$ do exist. We first observe that (16) and (17) imply that

$$(27) \qquad \begin{cases} 0 = \langle (< p(t), \overline{x}'(t) >, -p(t)), (1, \overline{x}'(t)) \rangle \geq \partial^+V(t, \overline{x}(t))(1, \overline{x}'(t)) \\ = \limsup_{h \to 0+} \frac{V(t+h, \overline{x}(t)+h\overline{x}'(t)) - V(t, \overline{x}(t))}{h} \\ = \limsup_{h \to 0+} \frac{V(t+h, \overline{x}(t+h)) - V(t, \overline{x}(t))}{h} \\ = \psi'(t) \end{cases}$$

This yields that ψ is nonincreasing. Since the value function is also nondecreasing along trajectories of the control system (7) (see Section 1), we deduce that the map $t \to V(t, \overline{x}(t))$ is constant. So $\overline{x}$ is optimal. □

The above map p may be constructed using the co-state variable of the Maximum Principle which is stated below.

We associate with the control system (7) the Hamiltonian $H : [0,1] \times \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ defined by

$$H(t,x,p) = \sup_{u \in U} < p, f(t,x,u) >$$

Under the assumptions of Section 1 it is continuous, locally Lipschitz with respect to (x,p) and convex with respect to the third variable.

Theorem 3.2 *Assume that (8) hold true and that f is differentiable with respect to x and g is differentiable. A trajectory-control pair $(\overline{x}, \overline{u})$ of control system (7) with $\overline{x}(0) = \xi_0$ is optimal for the problem (9) if and only if the solution $p : [0,1] \to \mathbf{R}^n$ of the adjoint equation*

$$(28) \qquad - p'(t) = \left(\frac{\partial f}{\partial x}(t, \overline{x}(t), \overline{u}(t))\right)^{\star} p(t), \ \ p(1) = -g'(\overline{x}(1))$$

satisfies the maximum principle

$$(29) \qquad < p(t), f(t, \overline{x}(t), \overline{u}(t)) > = \max_{u \in U} < p(t), f(t, \overline{x}(t), u) > \ \text{ a.e. in } [0,1]$$

and the transversality conditions

$$(30) \qquad (H(t, \overline{x}(t), p(t)), -p(t)) \in D^+V(t, \overline{x}(t)) \ \text{ a.e. in } [0,1]$$

$$(31) \qquad - p(t) \in D_x^+ V(t, \overline{x}(t)) \ \text{ for every } t \in [0,1]$$

where $D_x^+V(t, \overline{x}(t))$ denotes the superdifferential of $V(t,\cdot)$ at $\overline{x}(t)$.

Furthermore if V is semiconcave, then (30) holds true everywhere in $[0,1]$.

Remark — The above condition is a joined form of the maximum principle and the co-state inclusions (30), (31). The necessary condition of the above type was proved in ([16]) under somewhat different assumptions. An inclusion on co-state p similar to (31) in non-smooth case was derived in [11]. □

Proof — Sufficiency is a straightforward consequence of Theorem 3.1 and (29), (30). The fact that (28) and (29) are necessary is the well known Pontriagin's maximum principle.

To prove the necessity of (30) fix $t \in [0,1[$ such that $\overline{x}'(t) = f(t, \overline{x}(t), \overline{u}(t))$ and the equality (29) holds true and let $\Theta \in \mathbf{R}^n$. Consider the solution $w(\cdot)$ of the linearized along $(\overline{x}, \overline{u})$ system

$$(32) \qquad \begin{cases} w'(s) &= \frac{\partial f}{\partial x}(s, \overline{x}(s), \overline{u}(s))w(s), \quad s \in [0,1] \\ w(t) &= \Theta \end{cases}$$

For every $h > 0$, let x_h be the solution to the differential equation

$$(33) \qquad \begin{cases} x'(s) &= f(s, x(s), \overline{u}(s)), \quad s \in [0,1] \\ x(t) &= \overline{x}(t) + h\Theta \end{cases}$$

From the variational equation we know that the quotients

$$\frac{x_h - \overline{x}}{h}$$

converge uniformly to w. Fix $\alpha \in \mathbf{R}$. Hence from (28) and (29), using that V is nondecreasing along trajectories of (7) and constant along $\overline{x}$, we deduce that

$$\partial^+ V(t, \overline{x}(t))(\alpha, \alpha\overline{x}'(t) + \Theta)$$

$$= \limsup_{h\to 0+} \left(V(t + \alpha h, \overline{x}(t) + h(\alpha\overline{x}'(t) + w(t))) - V(t, \overline{x}(t))\right)/h$$

$$= \limsup_{h\to 0+} \left(V(t + \alpha h, \overline{x}(t + \alpha h) + hw(t + \alpha h)) - V(t, \overline{x}(t))\right)/h$$

$$= \limsup_{h\to 0+} \left(V(t + \alpha h, x_h(t + \alpha h)) - V(t, \overline{x}(t))\right)/h$$

$$\leq \limsup_{h\to 0+} \left(g(x_h(1)) - g(\overline{x}(1))\right)/h = g'(\overline{x}(1))w(1) = \langle -p(t), w(t)\rangle$$

$$= \langle -p(t), -\alpha\overline{x}'(t)\rangle + \langle -p(t), \alpha\overline{x}'(t) + \Theta\rangle$$

$$= \alpha H(t, \overline{x}(t), p(t)) + \langle -p(t), \alpha\overline{x}'(t) + \Theta\rangle$$

Hence we deduce that for every $\Theta_1 \in \mathbf{R}^n$

$$\partial^+ V(t, \overline{x}(t))(\alpha, \Theta_1) \leq \alpha H(t, \overline{x}(t), p(t)) + \langle -p(t), \Theta_1\rangle$$

Consequently, $(H(t, \overline{x}(t), p(t)), -p(t)) \in D^+V(t, \overline{x}(t))$ and the proof of (30) follows from (17). To prove (31) observe that for every $t \in [0,1]$, $\Theta \in \mathbf{R}^n$ and the solution w of (32)

$$\langle -p(t), \Theta\rangle = g'(\overline{x}(1))w(1) \geq \limsup_{h\to 0+}(V(t, \overline{x}(t) + h\Theta) - V(t, \overline{x}(t)))/h$$

$$= \partial_x^+ V(t, \overline{x}(t))(\Theta)$$

This and (17) imply (31). When V is semiconcave, then the last statement follows from (30), continuity of $H(\cdot)$, $p(\cdot)$, $\bar{x}(\cdot)$ and (24). □

Remark — When the Hamiltonian H is differentiable with respect to (x,p), then from arguments similar to [18, Remark 4.10] it follows that $\bar{x}$ and the co-state p of the last theorem satisfy the Hamiltonian system

$$\begin{cases} \bar{x}'(t) &= \frac{\partial H}{\partial p}(t, \bar{x}(t), p(t)) \\ p'(t) &= -\frac{\partial H}{\partial x}(t, \bar{x}(t), p(t)) \end{cases} \square$$

It is well known that for every $(t,x) \in [0,1] \times \mathbf{R}^n$ at which V is differentiable we have

$$(34) \qquad -\frac{\partial V}{\partial t}(t,x) + H\left(t, x, -\frac{\partial V}{\partial x}(t,x)\right) = 0$$

When V is not differentiable at (t,x) the above equation has to be understood in the viscosity sense (see [12], [13]).

Since the Hamiltonian is continuous we immediately deduce from (34) that

$$(35)\ \forall\, (t,x) \in [0,1] \times \mathbf{R}^n,\ \forall\, (p_t, p_x) \in D^{\star}V(t,x),\ \ -p_t + H(t,x,-p_x) = 0$$

We show next that in Theorem 3.2 whenever $p(0) = -V_x'(0,\xi_0)$, we have the equality in the inclusion (31).

Theorem 3.3 *Assume that (8) hold true and that f is differentiable with respect to x and g is differentiable. Suppose further that the derivative $V_x'(t_0, x_0)$ does exist and let $\bar{x}$ be an optimal solution for the problem (10). Consider the co-state $p : [t_0, 1] \to \mathbf{R}^n$ corresponding to $\bar{x}$ and given by Theorem 3.2, where the interval $[0,1]$ is replaced by $[t_0, 1]$ and ξ_0 by x_0. Then*

$$-p(t) = D_x^+ V(t, \bar{x}(t)) \text{ for all } t \in [t_0, 1]$$

In the next Section we show that under some additional regularity assumptions on f for all t, $p(t)$ is equal to the derivative of the value function $V_x'(t, \bar{x}(t))$ whenever $V_x'(t_0, x_0)$ does exists.

Proof — We already know from Theorem 3.2 that

$$-p(t) \in D_x^+ V(t, \bar{x}(t)) \text{ for all } t \in [t_0, 1]$$

Thus $p(t_0) = -V_x'(t_0, x_0)$.

Let $\overline{u}$ be an optimal control corresponding to $\overline{x}$. Fix Θ and let w, x_h have the same meaning as in the proof of Theorem 3.2 with t replaced by t_0. Then, since V is nondecreasing along trajectories of the control system (7) and constant along $\overline{x}$, for all $t \in [t_0, 1]$

$$\langle p(t_0), \Theta\rangle \;=\; \langle -V'_x(t_0, x_0), \Theta\rangle \;=\; -\lim_{h\to 0+} \frac{V(t_0, x_0 + h\Theta) - V(t_0, x_0)}{h}$$

$$\geq\; -\limsup_{h\to 0+} \frac{V(t, x_h(t)) - V(t, \overline{x}(t))}{h}$$

$$=\; -\limsup_{h\to 0+} \frac{V(t, \overline{x}(t) + hw(t)) - V(t, \overline{x}(t))}{h} \;=\; -\partial^+_x V(t, \overline{x}(t))(w(t))$$

where $\partial^+_x V(t, \overline{x}(t))(w(t))$ denotes the upper Dini derivative of $V(t, \cdot)$ at $\overline{x}(t)$ in the direction $w(t)$.

Using (17) we deduce that for every $q \in D^+_x V(t, \overline{x}(t))$ we have

$$\langle p(t_0), \Theta\rangle \;\geq\; \langle -q, w(t)\rangle \;=\; \langle -q, X(t)\Theta\rangle \;=\; \langle -X(t)^\star q, \Theta\rangle$$

where X denotes the fundamental solution of

$$\begin{cases} X'(t) &= \frac{\partial f}{\partial x}(t, \overline{x}(t), \overline{u}(t))X(t), \;\; t \in [t_0, 1] \\ X(t_0) &= Id \end{cases}$$

Since $\Theta \in \mathbf{R}^n$ is arbitrary, we have $p(t_0) = -X(t)^\star q$. On the hand, $p(\cdot)$ being a solution of (28), we know that $p(t_0) = X(t)^\star p(t)$. Since for every $t \in [t_0, 1]$, the matrix $X(t)$ is nonsingular we proved that $-p(t) = q$. This yields that $D^+_x V(t, \overline{x}(t))$ is single valued and ends the proof. □

Whenever H happens to be more regular we can prove the following theorem concerning optimal design.

For every (t_0, x_0) we define

$$D^\star_x V(t_0, x_0) \;=\; D^\star W(x_0) \tag{36}$$

where W is given by $W(x) = V(t_0, x)$.

Theorem 3.4 *Assume that (8) holds true, that f is differentiable with respect to x, g is differentiable and for every $R > 0$ there exists a nonnegative integrable function $l_R \in L^1(0, 1; \mathbf{R}_+)$ such that for all $x, y, p, q \in RB$*

$$\begin{cases} \left\|\frac{\partial H}{\partial x}(t, x, p) - \frac{\partial H}{\partial x}(t, y, q)\right\| + \left\|\frac{\partial H}{\partial p}(t, x, p) - \frac{\partial H}{\partial p}(t, y, q)\right\| \\ \leq\; l_R(t)(\|x - y\| + \|p - q\|) \end{cases} \tag{37}$$

Let $(t_0, x_0) \in [0,1] \times \mathbf{R}^n$ *and* $p_0 \in \mathbf{R}^n$ *be such that*

$$-p_0 \in D_x^+ V(t_0, x_0) \tag{38}$$

If $x(\cdot)$, $p(\cdot)$ *solves the differential equation*

$$\begin{cases} x'(t) = \frac{\partial H}{\partial p}(t, x(t), p(t)) \\ p'(t) = -\frac{\partial H}{\partial x}(t, x(t), p(t)), \quad t \in [t_0, 1] \end{cases} \tag{39}$$

and

$$\begin{cases} x(t_0) = x_0 \\ p(t_0) = p_0 \end{cases} \tag{40}$$

and if the sets $f(t, x, U)$ *are convex and compact, then* $x(\cdot)$ *is an optimal solution of problem (10).*

Remark — The above theorem extends a result of [8] which concerned a problem in Calculus of Variations. For such problems condition (37) is natural. It is much more restrictive for nonlinear control systems. We observe that (37) is satisfied whenever the variables x and u are "separated":

$$f(t, x, u) = \varphi(t, x) + \psi(t, u)$$

where $\varphi(t, \cdot)$ has $k_R(t)$–Lipschitz gradient and the boundary of $\psi(t, U)$ is sufficiently smooth. □

Lemma 3.5 *Under all assumptions of Theorem 3.4 suppose that the derivative* $\frac{\partial V}{\partial x}(t_0, x_0)$ *does exist. Then* $x(\cdot)$ *is optimal for the problem (10) if and only if there exists an absolutely continuous* $p : [t_0, 1] \to \mathbf{R}^n$ *such that* $(x, p)(\cdot)$ *solves (39) and*

$$-p(t_0) = \frac{\partial V}{\partial x}(t_0, x_0)$$

Proof — Assume that $x(\cdot)$ is an optimal solution of (10). By Theorem 3.2 applied with the interval $[0, 1]$ replaced by $[t_0, 1]$ and ξ_0 by x_0 and by the remark following it, there exists an absolutely continuous $p : [t_0, 1] \to \mathbf{R}^n$ such that $(x, p)(\cdot)$ is a solution of (39) and $-p(t_0) \in D_x^+ V(t_0, x_0)$. Since V is differentiable with respect to x at (t_0, x_0) we deduce that $-p(t_0) = \frac{\partial V}{\partial x}(t_0, x_0)$. Conversely, let (x, p) solve the Hamiltonian system (39) and

$$x(t_0) = x_0, \quad p(t_0) = -\frac{\partial V}{\partial x}(t_0, x_0)$$

Let $\overline{x}$ be an optimal solution of (10) and $\overline{p}$ be the corresponding co-state given by Theorem 3.2. Then for the same reasons as before

$$\overline{p}(t_0) = -\frac{\partial V}{\partial x}(t_0, x_0)$$

So, $(x, p) = (\overline{x}, \overline{p})$ by uniqueness. □

Proof of Theorem 3.4 — By the very definition of $D_x^*V(t_0, x_0)$ it follows that there exists a sequence x_k converging to x_0 such that $V(t_0, \cdot)$ is differentiable at x_k and

$$-p_0 = \lim_{k\to\infty} \frac{\partial V}{\partial x}(t_0, x_k)$$

Let $\overline{x}_k$ be an optimal trajectory for the problem (10) with x_0 replaced by x_k. Then from Lemma 3.5 there exists $\overline{p}_k$ such that $(\overline{x}_k, \overline{p}_k)$ solves (39) and $-\overline{p}_k(t_0) = \frac{\partial V}{\partial x}(t_0, x_k)$. By the continuous dependence of solutions on the initial conditions we obtain that the sequence $(\overline{x}_k, \overline{p}_k)$ converges uniformly to a solution $(\overline{x}, \overline{p})$ of (39) satisfying

$$\overline{x}(t_0) = x_0, \;\; \overline{p}(t_0) = p_0$$

So $x = \overline{x}$ by uniqueness and

$$V(t_0, x_0) = \lim_{k\to\infty} V(t_0, \overline{x}_k(t_0)) = \lim_{k\to\infty} g(\overline{x}_k(1)) = g(\overline{x}(1)) = g(x(1))$$

and therefore x is optimal. □

Remark —

i) By minor modifications of the above arguments it is easy to show that condition (38) may be replaced by the following one

$$(41) \qquad (H(t_0, x_0, p_0), -p_0) \in D^*V(t_0, x_0)$$

In general (38) and (41) are not comparable. If V is semiconcave, however, then (38) is more restrictive than (41) in view of Proposition 4.2 below.

ii) In general we do not know if either (38) or (41) is necessary for $x(\cdot)$ to be optimal. This is the case for (41) in Calculus of Variations (see [8]), since, then, for any optimal trajectory $\overline{x}(\cdot)$, V is differentiable at every point $(t, \overline{x}(t))$ with $t_0 < t < 1$.

Other examples of problems for which (38) is necessary, are given by optimal control problems having unique optimal trajectory for the initial state (t_0, x_0). □

4 Semiconcavity properties of the value function

We provide a sufficient condition for semi-concavity of the value function $V : [0,1] \times \mathbf{R}^n \to \mathbf{R}$ introduced in the first section. Throughout the whole section we suppose for simplicity that f does not depend on time. Moreover we assume

$$
(42) \begin{cases}
i) & f : \mathbf{R}^n \times U \to \mathbf{R}^n \text{ is continuous} \\
ii) & \exists\, M > 0 \text{ such that } \forall\, (x,u) \in \mathbf{R}^n \times U,\ \|f(x,u)\| \le M(\|x\| + 1) \\
iii) & \exists\, L > 0, \forall\, x_1, x_2 \in \mathbf{R}^n,\ u \in U,\ \|f(x_1,u) - f(x_2,u)\| \le L\,\|x_1 - x_2\| \\
iv) & \exists\, \omega : \mathbf{R}_+ \times \mathbf{R}_+ \to \mathbf{R}_+ \text{ such that (21) holds true and} \\
 & \forall\, \lambda \in [0,1],\ \forall\, u \in U,\ \forall\, R > 0,\ \forall\, x_0,\ x_1 \in RB \\
 & \|\lambda f(x_0,u) + (1-\lambda) f(x_1,u) - f(\lambda x_0 + (1-\lambda)x_1, u)\| \\
 & \le \lambda(1-\lambda)\,\|x_1 - x_0\|\,\omega(R, \|x_1 - x_0\|) \\
v) & g : \mathbf{R}^n \to \mathbf{R} \text{ is locally Lipschitz and semiconcave}
\end{cases}
$$

Remark —

1) Assumption $iv)$ holds true in particular when f is continuously differentiable with respect to x uniformly in u:

There exists a function $\omega : \mathbf{R}_+ \times \mathbf{R}_+ \to \mathbf{R}_+$ satisfying (21) such that

$$
\forall\, u \in U,\ \forall\, x_1, x_2 \in RB,\quad \left\| \frac{\partial f}{\partial x}(x_1,u) - \frac{\partial f}{\partial x}(x_2,u) \right\| \le \omega(R, \|x_1 - x_2\|)
$$

It can be proved in a way similar to Proposition 2.8.

2) Vice versa, Proposition 2.12 implies that if f satisfies $iv)$, then f is continuously differentiable with respect to x. □

The main result of this section is the following:

Theorem 4.1 *If (42) hold true, then the value function is semi-concave on* $[0,1] \times \mathbf{R}^n$.

Proof — For every $t \in [0,1]$ and measurable function $u : [t,1] \to U$, we denote by $y(\cdot; t, x, u)$ the solution of the system

$$
\begin{cases}
y'(s) = f(y(s), u(s)) \\
y(t) = x
\end{cases}
$$

The Gronwall lemma implies that

$$
(43) \qquad \forall\, x \in RB,\ \forall\, s \in [t,1],\ \|y(s)\| \le C_R := (R + M)e^M
$$

moreover for all $t \in [0,1]$, $s \in [t,1]$, x_0, $x_1 \in \mathbf{R}^n$ and all measurable functions $u : [t,1] \to U$ we have

$$\|y(s;t,x_1,u) - y(s;t,x_0,u)\| \le e^{L(s-t)}\|x_1 - x_0\| \tag{44}$$

Step 1. We claim that there exists $\omega_1 : \mathbf{R}_+ \times \mathbf{R}_+ \to \mathbf{R}_+$ satisfying (21) such that for all $0 \le t \le s \le 1$, $R > 0$, x_0, $x_1 \in RB$, $\lambda \in [0,1]$ and a measurable function $u : [t,1] \to U$ we have

$$\begin{aligned}&\|\lambda y(s;t,x_1,u) + (1-\lambda)y(s;t,x_0,u) - y(s;t,\lambda x_0 + (1-\lambda)x_1,u)\|\\ &\quad\le \lambda(1-\lambda)\|x_1 - x_0\|\,\omega_1(R,\|x_1 - x_0\|)\end{aligned}$$

Indeed set $x_\lambda = \lambda x_0 + (1-\lambda)x_1$ and define

$$y_\lambda(\tau) = \lambda y(\tau;t,x_1,u) + (1-\lambda)y(\tau;t,x_0,u) - y(\tau;t,x_\lambda,u)$$

Then

$$\begin{cases} y_\lambda'(\tau) = \\ \lambda f(y(\tau;t,x_1,u),u(\tau)) + (1-\lambda)f(y(\tau;t,x_0,u),u(\tau)) - f(y(\tau;t,x_\lambda,u),u(\tau)) \\ y_\lambda(t) = 0 \end{cases}$$

Thus by assumptions (42) *iii*) and *iv*) and (43)

$$\begin{aligned}&\|y_\lambda'(\tau)\| \le \\ &\lambda(1-\lambda)\|y(\tau;t,x_1,u) - y(\tau;t,x_0,u)\|\,\omega(C_R,\|y(\tau;t,x_1,u) - y(\tau;t,x_0,u)\|) + L\|y_\lambda(\tau)\|\end{aligned}$$

and our claim follows from (44) and the Gronwall lemma.

Step 2. We claim that there exists $\omega_2 : \mathbf{R}_+ \times \mathbf{R}_+ \to \mathbf{R}_+$ satisfying (21) such that for all $t \in [0,1]$, $\lambda \in [0,1]$, $R > 0$ and x_0, $x_1 \in RB$ the following inequality holds true

$$\begin{aligned}&\lambda V(t,x_1) + (1-\lambda)V(t,x_0) - V(t,\lambda x_1 + (1-\lambda)x_0)\\ &\le \lambda(1-\lambda)\|x_1 - x_0\|\,\omega_2(R,\|x_1 - x_0\|)\end{aligned}$$

Indeed define x_λ as above, fix $\varepsilon > 0$ and a control u_ε such that

$$V(t,x_\lambda) > g(y(1;t,x_\lambda,u_\varepsilon)) - \varepsilon$$

Let ω_g denotes a modulus of semiconcavity of g and L_R a Lipschitz constant of g on the ball of radius C_R. Then from (44) and Step 1 we get

$$\begin{aligned}&\lambda V(t,x_1) + (1-\lambda)V(t,x_0) - V(t,x_\lambda)\\ &< \lambda g(y(1;t,x_1,u_\varepsilon)) + (1-\lambda)g(y(1;t,x_0,u_\varepsilon)) - g(y(1;t,x_\lambda,u_\varepsilon)) + \varepsilon\\ &\le \lambda(1-\lambda)\|y(1;t,x_1,u_\varepsilon) - y(1;t,x_0,u_\varepsilon)\|\,\omega_g(C_R,\|y(1;t,x_1,u_\varepsilon) - y(1;t,x_0,u_\varepsilon)\|)\\ &+ L_R\|\lambda y(1;t,x_1,u_\varepsilon) + (1-\lambda)y(1;t,x_0,u_\varepsilon) - y(1;t,x_\lambda,u_\varepsilon)\| + \varepsilon\\ &\le L_R'\lambda(1-\lambda)\|x_1 - x_0\|\left(\omega_g(C_R,e^L\|x_1 - x_0\|) + \omega_1(R,\|x_1 - x_0\|)\right) + \varepsilon\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary our claim follows.

Thus we proved the semiconcavity of $V(t,\cdot)$.

Step 3. Consider next $0 \le t_1 < t_0 \le 1$, $R > 0$ and let $x_0,\ x_1 \in RB,\ \lambda \in [0,1]$. Define

$$x_\lambda = \lambda x_1 + (1-\lambda)x_0, \quad t_\lambda = \lambda t_1 + (1-\lambda)t_0$$

Pick any $\varepsilon > 0$ and let u_ε be such that

$$V(t_0, y(t_0; t_\lambda, x_\lambda, u_\varepsilon)) < V(t_\lambda, x_\lambda) + \varepsilon$$

Define

$$\tau(s) = \begin{cases} \lambda s + (1-\lambda)t_0, & \text{if } t_1 \le s \le t_0 \\ s & \text{otherwise} \end{cases} \tag{45}$$

Since the value function is nondecreasing along trajectories of our control system we have

$$\begin{cases} \lambda V(t_1, x_1) + (1-\lambda)V(t_0, x_0) - V(t_\lambda, x_\lambda) \le \\ \lambda V(t_0, y(t_0; t_1, x_1, u_\varepsilon \circ \tau)) + (1-\lambda)V(t_0, x_0) - V(t_0, y(t_0; t_\lambda, x_\lambda, u_\varepsilon)) + \varepsilon \end{cases} \tag{46}$$

Set $y_1(s) = y(s; t_1, x_1, u_\varepsilon \circ \tau)$, $y_\lambda(s) = y(s; t_\lambda, x_\lambda, u_\varepsilon)$. Let K_R denote the Lipschitz constant of V on $[0,1] \times C_R B$. By (46) and Step 2 we obtain

$$\begin{cases} \lambda V(t_1, x_1) + (1-\lambda)V(t_0, x_0) - V(t_\lambda, x_\lambda) \\ \le \lambda(1-\lambda) \|y_1(t_0) - x_0\| \omega_2(C_R, \|y_1(t_0) - x_0\|) \\ + K_R \|\lambda y_1(t_0) + (1-\lambda)x_0 - y_\lambda(t_0)\| \end{cases} \tag{47}$$

On the other hand from assumption (42) *ii*) follows that

$$\forall\, s \in [t_1, t_0], \ \|y_1(s) - x_0\| \le \|x_1 - x_0\| + M_R(t_0 - t_1) \tag{48}$$

where $M_R = M(1 + C_R)$. Set

$$z(s) = \lambda y_1(\tau^{-1}(s)) + (1-\lambda)x_0 - y_\lambda(s)$$

and notice that $z(t_\lambda) = 0$, $z(t_0) = \lambda y_1(t_0) + (1-\lambda)x_0 - y_\lambda(t_0)$. Furthermore, using (42) *iii*), we obtain the following estimates

$$\begin{cases} \|z'(s)\| = \|f(y_1 \circ \tau^{-1}(s), u_\varepsilon(s)) - f(y_\lambda(s), u_\varepsilon(s))\| \\ \le L\|y_1 \circ \tau^{-1}(s) - y_\lambda(s)\| \le L\|z(s)\| + L(1-\lambda)\|y_1 \circ \tau^{-1}(s) - x_0\| \end{cases}$$

Therefore from the Gronwall inequality and (48) we deduce that

$$\begin{cases} \|z(t_0)\| \le L\int_{t_\lambda}^{t_0} (1-\lambda)\|y_1 \circ \tau^{-1}(s) - x_0\| e^{L(s-t_0)} ds \\ \le Le^L \lambda(1-\lambda)(t_0 - t_1)(\|x_1 - x_0\| + M_R(t_0 - t_1)) \end{cases} \tag{49}$$

Inequalities (47), (48) imply the conclusion. □

Proposition 4.2 *Assume that the value function is semiconcave at a point* $(t_0, x_0) \in [0,1] \times \mathbf{R}^n$. *If* $D_x^+V(t_0, x_0)$ *is a singleton, then* V *is differentiable at* (t_0, x_0) *and* $D^*V(t_0, x_0) = \{ V'(t_0, x_0) \}$.

Here at boundary points ($t_0 \in \{ 0, 1 \}$) the above differentiability of course has to be understood in one sided sense.

Proof — Let $\pi_x : \mathbf{R} \times \mathbf{R}^n \to \mathbf{R}^n$ denote the projection on $\mathbf{R}^n$. Since

$$\pi_x D^+V(t_0, x_0) \subset D_x^+V(t_0, x_0) =: \{ p_0 \},$$

by (35) and (24) we conclude that

$$(p_t, p_x) \in D^*V(t_0, x_0) \implies p_x = p_0, \; p_t = H(x_0, -p_0)$$

Hence $D^+V(t_0, x_0)$ is a singleton. The conclusion follows from Proposition 2.10. □

Corollary 4.3 *Assume (42), that* g *is differentiable and that the derivative* $V_x'(t_0, x_0)$ *does exist, and let* $\bar{x}$ *be an optimal solution of problem (10). Then for all* $t \in [t_0, 1]$, V *is differentiable at* $(t, \bar{x}(t))$ *and*

$$D^*V(t, \bar{x}(t)) = \{ V'(t, \bar{x}(t)) \}$$

Conversely assume that $x : [t_0, 1] \to \mathbf{R}^n$ *is a solution of (7) and that for every* $t \in [t_0, 1]$, V *is differentiable at* $(t, x(t))$. *If the sets* $f(x, U)$ *are convex and compact and*

$$-\frac{\partial V}{\partial x}(t, x(t))x'(t) = H\left(x(t), -\frac{\partial V}{\partial x}(t, x(t))\right) \text{ a.e. in } [t_0, 1] \qquad (50)$$

then x *is optimal for problem (10).*

Proof — The first statement follows immediately from Proposition 4.2 and Theorem 3.3. To prove the second one fix $\bar{t} \in [t_0, 1]$ and let $\bar{x} : [\bar{t}, 1] \to \mathbf{R}^n$ be an optimal solution of problem (10) with (t_0, x_0) replaced by $(\bar{t}, x(\bar{t}))$.

We already know that V is semiconcave. By Theorem 3.2 there exists $p(\bar{t}) \in \mathbf{R}^n$ such that

$$(H(x(\bar{t}), p(\bar{t})), -p(\bar{t})) = V'(\bar{t}, x(\bar{t}))$$

Since $\bar{t} \in [t_0, 1]$ is arbitrary, assumption (50) and Theorem 3.1 end the proof. □

Usually the value function is not everywhere differentiable. However that is always the case for "convex" problems and continuously differentiable cost, as we prove below (see also [5], [6], [7]).

Proposition 4.4 *Assume that (42) holds true, g is convex and*

$$\text{Graph}(f(\cdot, U)) \text{ is closed and convex} \tag{51}$$

Then V is continuously differentiable on $[0,1] \times \mathbf{R}^n$ *and convex with respect to the second variable.*

Proof — By Theorem 1.1, assumption (51) yields that for every $(t_0, x_0) \in [0,1] \times \mathbf{R}^n$ there exists a solution $\overline{x}$ of the control system

$$x' = f(x(t), u(t)), \; u(t) \in U, \;\; x(t_0) = x_0$$

satisfying $V(t_0, x_0) = g(\overline{x}(1))$.

Fix $t_0 \in [0,1]$, x_0, $x_1 \in \mathbf{R}^n$, $\lambda \in [0,1]$ and consider trajectories $x : [t_0, 1] \to \mathbf{R}^n$ and $y : [t_0, 1] \to \mathbf{R}^n$ such that $V(t_0, x_0) = g(x(1))$, $V(t_0, x_1) = g(y(1))$. Define the trajectory $z : [t_0, 1] \to \mathbf{R}^n$ by $z(t) = \lambda x(t) + (1-\lambda)y(t)$. Then, using (51), we obtain that z is a solution of the control system (7). Thus, by convexity of g,

$$V(t_0, \lambda x_0 + (1-\lambda)x_1) \;\le\; g(z(1)) \;\le\; \lambda V(t_0, x_0) + (1-\lambda)V(t_1, x_1)$$

and therefore $V(t_0, \cdot)$ is convex.

Next, as $V(t, \cdot)$ is both convex and semiconcave for all $t \in [0,1]$, Proposition 2.12 yields that $V(t, \cdot)$ is continuously differentiable on $\mathbf{R}^n$. The conclusion now follows from Proposition 4.2. □

5 Optimal feedback

One of the major issues of optimal control theory is to find an "equation" for optimal trajectories. Theorem 1.4 provides an inclusion formulation. However in general the set-valued map G is not regular enough to make us able to solve the inclusion (13). The situation is comparable to having an ordinary differential equation with nonsmooth right hand side: it may have solutions, but this solution can not be obtained as say limits of Euler curves.

That is why we have to investigate regularity properties of G. In this section we show that under the assumptions of Theorem 4.1, the feedback map G^{co} is upper semicontinuous and that so is G if we assume in addition that the sets $f(x, U)$ are closed.

In this section we assume again that the control system (7) is atonomous, i.e., f does not depend on time.

Results of Sections 2 and 4 imply that under assumptions (42) the feedback maps $G : [0,1] \times \mathbf{R}^n \mapsto \mathbf{R}^n$ and $G^{co} : [0,1] \times \mathbf{R}^n \mapsto \mathbf{R}^n$ defined in Section 1 are respectively equal to

$$G(t,x) \;=\; \{\, v \,\in\, f(x,U) \mid V^o_-(t,x)(1,v) = 0 \,\}$$

and

$$G^{co}(t,x) \;=\; \{\, v \,\in\, \overline{co} f(x,U) \mid V^o_-(t,x)(1,v) = 0 \,\}$$

Theorem 5.1 *Let us assume that (42) holds true. Then G^{co} has compact nonempty images and is upper semicontinuous. The same holds true for the map G if we assume in addition that the sets $f(x,U)$ are closed.*

Proof — From Theorems 4.1 and 2.9 we know that for every $(t,x) \in [0,1[\times \mathbf{R}^n$ and every $\Theta \in \mathbf{R}^n$ the directional derivative $\frac{\partial V}{\partial(1,\Theta)}(t,x)$ exists and is equal to the regularized lower derivative $V^o_-((t,x),(1,\Theta))$. Define the set-valued map

$$\hat{Q} : [0,1] \times \mathbf{R}^n \rightsquigarrow \mathbf{R}^n$$

by

$$\hat{Q}(t,x) \;=\; \{\, \Theta \,\in\, \mathbf{R}^n \mid V^o_-(t,x)(1,\Theta) \leq 0 \,\}$$

From Proposition 2.5 we know that the set $\mathrm{Graph}(\hat{Q})$ is closed. On the other hand Proposition 1.3 implies that for every $v \in \overline{co} f(x,U)$, $\frac{\partial V}{\partial(1,v)}(t,x) \geq 0$. Thus

$$G(t,x) \;=\; \hat{Q}(t,x) \cap f(x,U), \;\; G^{co}(t,x) \;=\; \hat{Q}(t,x) \cap \overline{co} f(x,U)$$

This and the assumptions on f imply that graphs of the set-valued maps G, G^{co} are closed. Furthermore G^{co} takes its values in a compact set. From [2, p.42] follows that G and G^{co} are upper semicontinuous. □

Corollary 5.2 *Let us assume that (42) holds true and that the sets $f(x,U)$ are closed. If the map G is single-valued, then the function $(t,x) \to G(t,x)$ is continuous.*

A typical example of a nonlinear control system with closed convex images is the affine system:

$$x' \;=\; f(x) \,+\, \sum_{i=1}^{k} u_i g_i(x), \;\; u_i \in [a,b]$$

where f and g_i are continuous functions from $\mathbf{R}^n$ to itself.

The feedback map G defined above, in general, does not have convex images because the map of directional derivatives is concave.

For this reason, in general, the feedback inclusion (13) is very difficult to investigate. When V happens to be differentiable and the sets $f(x,U)$ are closed and convex, then for obvious reasons the map G has convex compact images. Proposition 4.4 provides a sufficient condition for continuous differentiability of V.

Theorem 5.3 ***Assume that (42), (51) hold true and that g is convex. Then G has convex compact images and is upper semicontinuous. Furthermore if for every x the set $f(x,U)$ is strictly convex, then G is single valued and continuous.***

Proof — By Proposition 4.4 we know that V is continuously differentiable. This yields that for all $(t,x) \in [0,1[\times \mathbf{R}^n$ the set

$$G(t,x) = f(x,U) \cap \{\Theta \in \mathbf{R}^n \mid V'(t,x)(1,\Theta) = 0\}$$

is convex. Theorem 5.1 ends the proof of the first statement. From Proposition 1.3 it follows that for all $(t,x) \in [0,1[\times \mathbf{R}^n$

$$v \in G(t,x) \iff v \in f(x,U) \ \& \ \sup_{u\in U} \left\langle -\frac{\partial V}{\partial x}(t,x),\ f(x,u) \right\rangle = \left\langle -\frac{\partial V}{\partial x}(t,x),\ v \right\rangle$$

This and strict convexity of $f(x,U)$ imply that G is single valued. Corollary 5.2 completes the proof. □

Let us assume that G is upper semicontinuous and has convex compact images. We have already mentioned that solutions of (13) may be constructed as limits of Euler curves.

An alternative approach comes from Cellina's approximate selection theorem (see [2, Theorem 1.12.1, p.84]). Namely this theorem states that for every $\varepsilon > 0$ and $R > 0$ there exists a locally Lipschitz map $g_{\varepsilon R} : [0,1] \times RB \to \mathbf{R}^n$ satisfying

$$\text{(52)} \qquad \mathrm{Graph}(g_{\varepsilon R}) \subset \mathrm{Graph}(G) + \varepsilon B$$

With every $\varepsilon > 0$, $R > 0$ we associate the solution $x_{\varepsilon R}$ of the differential equation

$$\begin{cases} x' &= g_{\varepsilon R}(x) \\ x(0) &= \xi_0 \end{cases}$$

Then from assumptions (8) follows that for some $R > 0$, $x_{\varepsilon R}$ are defined on the whole interval $[0,1]$ and the sequence $\{ x_{\varepsilon R} \}_{\varepsilon \in]0,1]}$ is bounded in $C(0,1)$. Hence also $\{ g \circ x_{\varepsilon R} \}_{\varepsilon \in]0,1]}$ is bounded and therefore the functions $x_{\varepsilon R}$ are equicontinuous. This and the Ascoli-Arzela theorem imply that for some sequence $\varepsilon_n \to 0+$, the subsequence $\{ x_{\varepsilon_n R} \}_{n \geq 1}$ converges to an absolutely continuous function $x : [0,1] \to \mathbf{R}^n$. From (52) we deduce that x is a solution of the feedback inclusion (13) and thereby it is optimal.

6 Viability approach to optimal control

In this section we provide an alternative approach to optimal trajectories based on viability techniques.

We first observe the following characterization of optimal trajectories:

Theorem 6.1 *Assume that f satisfies (8). Then a solution $\overline{x}$ of the control system (7) defined on the time interval $[0,1]$ is optimal if and only if the function $t \to (t, \overline{x}(t), V(0,\xi_0))$ a solution of the viability problem*

$$(53) \qquad \begin{cases} t' = 1 \\ x'(t) = f(t,x(t),u(t)), \; u(t) \in U \text{ is measurable} \\ z'(t) = 0 \\ (t,x(t),z(t)) \in \text{Graph}(V) \text{ for all } t \in [0,1] \\ t(0) = 0, \; x(0) = \xi_0, \; z(0) = V(0,\xi_0) \end{cases}$$

Proof — We already observed that $\overline{x}(\cdot)$ is optimal if and only if the map $t \to V(t,\overline{x}(t))$ is constant on the time interval $[0,1]$. On the other hand $t \to (t,\overline{x}(t),z(t))$ is a solution of (53) if and only if $z(t) = V(t,\overline{x}(t)) \equiv const$ and $\overline{x}(\cdot)$ is a solution of (7) satisfying $\overline{x}(0) = \xi_0$. □

Inclusion (53) is a viability problem which may be approached using many results of viability theory. Actually viability technique may be applied not only to the value function V but also to any continuous function W satisfying some inequalities from [16]. To state results in this direction we need the following definition.

Definition 6.2 *Consider a function $W : [0,1] \times \mathbf{R}^n \to \mathbf{R}$ and let $(t,x) \in [0,1] \times \mathbf{R}^n$. The contingent derivative of W at (t,x) in the direction $(w,v) \in \mathbf{R} \times \mathbf{R}^n$ is a subset of $\mathbf{R}$ defined by*

$$DW(t,x)(w,v) := \left\{ u \in \mathbf{R} \mid \liminf_{h\to 0+,(w',v')\to(w,v)} dist\left(u, \frac{W(t+hw', x+hv') - W(t,x)}{h}\right) = 0 \right\}$$

Theorem 6.3 *Consider a continuous function $W : [0,1] \times \mathbf{R}^n \to \mathbf{R}$ and assume that f satisfies (8). If for every $(t,x) \in [0,1] \times \mathbf{R}^n$*

$$0 \in \{DW(t,x)(1,v) \mid v \in \overline{co} f(t,x,U)\}$$

then for all (t_0,x_0) there exists a solution $\overline{x}$ of the differential inclusion

$$x'(t) \in \overline{co} f(t,x(t),U) \text{ a.e. in } [t_0,1], \; x(t_0) = x_0$$

such that $W(t,\overline{x}(t)) \equiv W(1,\overline{x}(1))$.

Proof — It is not restrictive to assume that $t_0 = 0$. We extend W on $\mathbf{R}_+ \times \mathbf{R}^n$ by setting for all $t > 1$, $W(t,x) = W(1,x)$. Define the closed set K =Graph(W) and the map $F_1(t,x) = \{1\} \times \overline{co} f(t,x,U) \times \{0\}$. Set

$$\hat{F}(t,x) = \begin{cases} F_1(t,x) & \text{if } t \in [0,1[\\ \overline{co}(\{0\} \cup F_1(1,x)) & \text{if } t \geq 1 \end{cases}$$

Then for every $(t,x) \in \mathbf{R}_+ \times \mathbf{R}^n$, the contingent cone $T_K(t,x,W(t,x))$ to K at $(t,x,W(t,x))$ is equal to Graph$(DW(t,x))$. Hence, by our assumption, for every $(t,x) \in [0,1[\times\mathbf{R}^n$ there exists $u \in F_1(t,x)$ such that $(1,u,0) \in T_K(t,x,W(t,x))$. Furthermore, for every $t \geq 1$ and $x \in \mathbf{R}^n$, we have $0 \in \hat{F}(t,x)$. This proves that

$$\forall\, (t,x) \in \mathbf{R}_+ \times \mathbf{R}^n, \;\; \hat{F}(t,x) \cap T_K(t,x,W(t,x)) \neq \emptyset$$

By the assumptions $\hat{F}$ is continuous and has closed convex images. Consequently, by the Haddad viability theorem [19], the constrained system

$$\begin{cases} y'(t) \in \hat{F}(t,y(t)) & \text{almost everywhere} \\ y(t) \in K & \text{for all } t \\ y(0) = (0, x_0, W(0,x_0)) \end{cases}$$

has a solution $\overline{y} = (z_0, x, z) : [t_0, t_1] \to \mathbf{R} \times \mathbf{R}^n \times \mathbf{R}$ for some $t_1 > t_0$. Using the assumptions on f and Haddad's theorem, we extend this solution on the time interval $[t_0, 1]$. Then, from definition of K and $\hat{F}$, $z_0(t) = t$, $z(t) = W(t,x(t))$. On the other hand $z'(t) = 0$ almost everywhere in $[t_0,1]$ and therefore $z \equiv const$. This ends the proof. □

Theorem 6.4 *Consider a continuous function $W : [0,1] \times \mathbf{R}^n \to \mathbf{R}$ and assume that f does not depend on t and satisfies (8). If $W(1,\cdot) = g(\cdot)$ and*

$$\sup_{v \in \overline{co} f(x,U)} \inf D(-W)(t,x)(1,v) \leq 0$$

then for every solution $y(\cdot) = (t,x,z)(\cdot)$ of

$$(54) \qquad \begin{cases} t' = 1 \\ x'(t) = f(t,x(t),u(t)), \;\; u(t) \in U \text{ is measurable} \\ z'(t) = 0 \\ (t,x(t),z(t)) \in \text{Graph}\,(W) \;\; \text{for all } t \in [0,1] \\ t(0) = 0, \; x(0) = \xi_0, \; z(0) = W(0,\xi_0) \end{cases}$$

defined on the time interval $[0,1]$, the trajectory $x(\cdot)$ is optimal for the problem (9).

Proof — From [16] we deduce that W is nondecreasing along trajectories of (7). On the other hand if $y(\cdot) = (t,x,z)(\cdot)$ is a solution of (54) defined on the time interval $[0,1]$ then $W(t,x(t)) \equiv const$. □

7 Problem with end point constraints

In this section we investigate the case when the additional end point constraint is present:

$$x(1) \in K_1$$

where K_1 is a given closed subset of $\mathbf{R}^n$. The corresponding value function is defined by

$$V(t_0, x_0) = inf\{g(x(1)) \mid x \text{ is a solution of (7) on } [t_0, 1],\ x(t_0) = x_0,\ x(1) \in K_1\}$$

We observe that $V(t_0, x_0) = +\infty$ whenever no trajectory starting at x_0 at time t_0 hits K_1 at time one.

In this more general case the value function may be discontinuous and one has either to develop a verification technique for a larger class of functions (some results in this direction were obtained in [16] or to try to reduce the problem to a new one, where the data fits the Lipschitzian framework. We shall follow this second strategy and apply the penalization technique.

We provide only a convergence result showing that the problem with end point constraints may be approximated by free end point ones. Further developments are left to future work.

We impose on the functions f and g the same assumptions as in Section 1 and we consider the family of penalized problems: with every $\varepsilon > 0$ we associate the minimization problem

$$(P_\varepsilon) \quad \text{minimize } \left\{ g(x(1)) + \frac{1}{\varepsilon} dist(x, K_1)^2 \mid x(\cdot) \text{ is a solution of (7)},\ x(0) = \xi_0 \right\}$$

Define functions g_ε from $\mathbf{R}^n$ to $\mathbf{R}$ by

$$\forall\, x \in \mathbf{R}^n,\ \ g_\varepsilon(x) \ = \ g(x) \ + \ \frac{1}{\varepsilon}\, dist(x, K_1)^2$$

The value function V_ε corresponding to the problem (P_ε) is defined by (10) with g replaced by g_ε.

Since g_ε is locally Lipschitz we deduce that V_ε is also locally Lipschitz continuous with the Lipschitz constant depending on ε. Hence the results obtained in previous sections may be applied to V_ε.

Furthermore if g is semiconcave, then, using Example 1 from Section 2, we show that also the functions g_ε are semiconcave. This and Theorem 4.1 yield that under assumptions (42) for every $\varepsilon > 0$ the value function V_ε is semiconcave on $[0,1] \times \mathbf{R}^n$. Consequently, results concerning regularity of optimal feedback may be applied to penalized problems.

The aim of this section is to prove the convergence of V_ε to V.

Theorem 7.1 *Assume that f satisfies (8). If the sets $f(t,x,U)$ are closed and convex, then for every $(t,x) \in [0,1] \times \mathbf{R}^n$ the function $\mathbf{R}_+ \ni \varepsilon \to V_\varepsilon(t,x)$ is nonincreasing. Furthermore for every $\varepsilon > 0$, $V_\varepsilon(t,x) \le V(t,x)$ and*

$$\lim_{\varepsilon \to 0+} V_\varepsilon(t,x) = V(t,x)$$

Proof — The first two statements are obvious. Fix $(t,x) \in [0,1] \times \mathbf{R}^n$ and set $W(t,x) = \lim_{\varepsilon \to 0+} V_\varepsilon(t,x)$. Clearly $W(t,x) \le V(t,x)$. To show the opposite it is enough to consider the case $W(t,x) < +\infty$. Consider trajectory control pairs $(y^\varepsilon, u^\varepsilon)$ of control system (7) satisfying

$$V^\varepsilon(t,x) = g^\varepsilon(y^\varepsilon(1))$$

(they exist by Theorem 1.1). Then, by the relaxation theorem [2], there exists a sequence $\varepsilon_n \to 0+$ and a trajectory $y(\cdot)$ of (7) defined on $[t,1]$ such that $y^{\varepsilon_n} \to y$ uniformly on $[t,1]$. On the other hand

$$0 \le dist(y^\varepsilon(1), K_1)^2 \le \varepsilon\,(W(t,x) - g(y^\varepsilon(1))$$

and therefore, taking the limit in the above inequality, we obtain $y(1) \in K_1$. Furthermore from the inequality

$$V^\varepsilon(t,x) \ge g(y^\varepsilon(1))$$

we deduce that $W(t,x) \ge g(y(1)) \ge V(t,x)$. □

Corollary 7.2 *Under all assumptions of Theorem 7.1 consider a sequence $\varepsilon_n \to 0+$ and let $x^{\varepsilon_n}(\cdot)$ be an optimal solution to the problem (P_{ε_n}). If problem (P) has at least one solution, then every cluster point $x(\cdot)$ of $\{x^{\varepsilon_n}(\cdot)\}$ in the metric of uniform convergence is an optimal solution of (P).*

Proof — Indeed, since (P) has a solution, by Theorem 7.1, for all $n > 0$,

$$V^{\varepsilon_n}(t, x^{\varepsilon_n}(t)) \equiv const \le V(0,\xi_0) < +\infty$$

and taking limit we deduce that $V(t,x(t)) \equiv const < +\infty$. Thus $x(1) \in K_1$ and x is optimal. □

The above results imply that to find an optimal solution of the problem with end point constraints we can take any cluster point of solutions of penalized problems when $\varepsilon \to 0+$. On the other hand penalized problems may be addressed using theorems of previous sections.

References

[1] AUBIN J.-P. (to appear) VIABILITY THEORY.

[2] AUBIN J.-P. & CELLINA A. (1984) DIFFERENTIAL INCLUSIONS. Springer-Verlag (Grundlehren der Math. Wissenschaften), Vol.264

[3] AUBIN J.-P. & FRANKOWSKA H. (to appear) SET-VALUED ANALYSIS. Birkhäuser, Systems and Control: Foundations and Applications

[4] BERKOVITZ L. (to appear) *On optimal trajectories of a control problem.* SIAM J. on Control and Opt.

[5] BARBU V. & DA PRATO G. (1982) HAMILTON-JACOBI EQUATIONS IN HILBERT SPACES. Pitman

[6] BARDI M. & EVANS C.L. (1984) *On Hopf's formula for solutions of Hamilton-Jacobi Equations.* Nonlinear Anal., TMA, 8, pp. 1373-1381

[7] CANNARSA P. & SONER H.M. (1987) *On the singularities of the viscosity solutions to Hamilton-Jacobi-Bellman equations.* Indiana University Math. J. 46, pp. 501-524

[8] CANNARSA P. & SONER H.M. (1989) *Generalized one-sided estimates for solutions of Hamilton-Jacobi Equations and applications.* Nonlinear Anal., TMA, 13, pp. 305-323

[9] CAPUZZO DOLCETTA I. & ISHII H. (1984) *Approximate solutions of the Bellman equation of deterministic control theory.* Applied Math. and Optim. 11, pp. 161-181

[10] CLARKE F.H. (1983) OPTIMIZATION AND NONSMOOTH ANALYSIS. Wiley interscience

[11] CLARKE F.H. & VINTER R.B. (1987) *The relationship between the maximum principle and dynamic programming.* SIAM J. on Control and Opt. 25, pp. 1291-1311

[12] CRANDALL M.G. & LIONS P.L. (1983) *Viscosity solutions of Hamilton-Jacobi equations.* Trans.Am.Math.Soc., 277, pp. 1-42

[13] CRANDALL M.G., EVANS L.C. & LIONS P.L. (1984) *Some properties of viscosity solutions of Hamilton-Jacobi equations.* Trans.Am.Math.Soc., 282, pp. 487-502

[14] CRANDALL M.G. & LIONS P.L. (1986) *On existence and uniqueness of solutions of Hamilton-Jacobi equations.* Nonlinear Anal., TMA, 10, pp. 353-370

[15] FLEMING W.H., & RISHEL W.H., (1975) DETERMINISTIC AND STOCHASTIC OPTIMAL CONTROL. Springer-Verlag, New York

[16] FRANKOWSKA H. (1989) *Optimal trajectories associated to a solution of contingent Hamilton-Jacobi equations.* Applied Math. and Optim., 19, pp. 291-311

[17] FRANKOWSKA H. (1989) *Hamilton-Jacobi equation: viscosity solutions and generalized gradients.* J. of Math. Analysis and Appl., 141, pp. 21-26

[18] FRANKOWSKA H. (1989) *Contingent cones to reachable sets of control systems.* SIAM J. on Control and Opt. 27, pp. 170-198

[19] HADDAD G. (1981) *Monotone trajectories of differential inclusions with memory.* Israel J. Maths. 39, pp. 38-100

[20] ISHII H. (1984) *Uniqueness of unbounded viscosity solutions of Hamilton-Jacobi equations.* Indiana University Math. J. 43, pp. 721-748

[21] ISHII H. & LIONS P.L. (1988) *Viscosity solutions of fully nonlinear second order elliptic partial differential equation.* Preprint CEREMADE, 8820

[22] LIONS P.L. (1982) GENERALIZED SOLUTIONS OF HAMILTON-JACOBI EQUATIONS. Pitman, London

[23] LIONS P.L. & SOUGANIDIS P.E. (1985) *Differential games, optimal control and directional derivatives of viscosity solutions of Bellman's and Isaaks' equations.* SIAM J. on Control, 23, pp. 566-583