Sensitivity relations for the Mayer problem with differential inclusions

Piermarco Cannarsa, Hélène Frankowska, Teresa Scarinci

To cite this version:

Piermarco Cannarsa, Hélène Frankowska, Teresa Scarinci. Sensitivity relations for the Mayer problem with differential inclusions. ESAIM: Control, Optimisation and Calculus of Variations, EDP Sciences, 2015, 21 (3), pp.789 - 814. <10.1051/cocv/2014050 >. <hal-00991204v3>

HAL Id: hal-00991204

https://hal.inria.fr/hal-00991204v3

Submitted on 26 Sep 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sensitivity relations for the Mayer problem with differential inclusions

P. Cannarsa[1], **H. Frankowska**[2], **T. Scarinci**[1,2]

[1] *Dipartimento di Matematica, Università di Roma Tor Vergata, Via della Ricerca Scientifica 1, 00133 Roma, Italy.*

[2] *CNRS, IMJ-PRG, UMR 7586, Sorbonne Universités, UPMC Univ Paris 06, Univ Paris Diderot, Sorbonne Paris Cité, Case 247, 4 Place Jussieu, 75252 Paris, France.*

Abstract

In optimal control, sensitivity relations are usually understood as inclusions that identify the pair formed by the dual arc and the Hamiltonian as a suitable generalized gradient of the value function, evaluated along a given minimizing trajectory. In this paper, sensitivity relations are obtained for the Mayer problem associated with the differential inclusion $\dot{x} \in F(x)$ and applied to express optimality conditions. The first application of our results concerns the maximum principle and consists in showing that a dual arc can be constructed for *every* element of the superdifferential of the final cost as a solution of an adjoint system. The second and last application we discuss in this paper concerns optimal design. We show that one can associate a family of optimal trajectories, starting at some point (t, x), with every nonzero reachable gradient of the value function at (t, x), in such a way that families corresponding to distinct reachable gradients have *empty* intersection.

Keywords: Mayer problem, differential inclusions, optimality conditions, sensitivity relations.
MSC Subject classifications: 34A60, 49J53.

1 Introduction

Given a complete separable metric space U and a vector field $f : \mathbb{R}^n \times U \to \mathbb{R}^n$, smooth with respect to x, for any point $(t_0, x_0) \in (-\infty, T] \times \mathbb{R}^n$ and Lebesgue measurable map $u : [t_0, T] \to U$ let us denote by $x(\cdot; t_0, x_0, u)$ the solution of the Cauchy problem

$$\begin{cases} \dot{x}(t) = f(x(t), u(t)) & \text{a.e. } t \in [t_0, T], \\ x(t_0) = x_0, \end{cases} \tag{1.1}$$

that we suppose to exist on the whole interval $[t_0, T]$. Then, given a smooth function $\phi : \mathbb{R}^n \to \mathbb{R}$, we are interested in minimizing the final cost $\phi(x(T; t_0, x_0, u))$ over all controls u.

In the Dynamic Programming approach to such a problem, one seeks to characterize the value function V, that is,

$$V(t_0, x_0) = \inf_{u(\cdot)} \phi(x(T; t_0, x_0, u(\cdot))) \qquad (t_0, x_0) \in (-\infty, T] \times \mathbb{R}^n, \tag{1.2}$$

as the unique solution, in a suitable sense, of the Hamilton-Jacobi equation

$$\begin{cases} -\partial_t v(t, x) + H(x, -v_x(t, x)) = 0 & \text{in } (-\infty, T) \times \mathbb{R}^n \\ v(T, x) = \phi(x) & x \in \mathbb{R}^n, \end{cases}$$

where the Hamiltonian H is defined as

$$H(x,p) = \sup_{u \in U} \langle p, f(x,u) \rangle \qquad (x,p) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Now, the classical method of characteristics ensures that, given $t_0 \in \mathbb{R}^n$ and as long as V is smooth, along any solution of the system of ODEs

$$\begin{cases} \dot{x}(t) = \nabla_p H(x(t), p(t)), & x(T) = z \\ -\dot{p}(t) = \nabla_x H(x(t), p(t)), & p(T) = -\nabla \phi(z) \end{cases} \qquad t \in [t_0, T], \tag{1.3}$$

the gradient of V satisfies

$$(H(x(t), p(t)), -p(t)) = \nabla V(t, x(t)), \quad \forall\, t \in [t_0, T]. \tag{1.4}$$

It is well known that the characteristic system (1.3) is also a set of necessary optimality conditions for any optimal solution $x(\cdot)$ of the Mayer problem (1.2). Observe that $\nabla V(t,x)$ allows to "measure" sensitivity of the optimal cost with respect to (t,x). For this reason, (1.4) is called a *sensitivity relation* for problem (1.2). Obviously, the above calculation is just formal because, in general, V cannot be expected to be smooth. On the other hand, relation (1.4) is important for deriving sufficient optimality conditions, as we recall in Section 2. This fact motivates interest in generalized sensitivity relations for nonsmooth value functions.

To the best of our knowledge, the first "nonsmooth result" in the above direction was obtained by Clarke and Vinter in [10] for the Bolza problem, where, given an optimal trajectory $\overline{x}(\cdot)$, an associated arc $p(\cdot)$ is constructed to satisfy the partial sensitivity relation

$$-p(t) \in \partial_x V(t, \overline{x}(t)) \;\; a.e. \;\; t \in [t_0, T]. \tag{1.5}$$

Here, $\partial_x V$ denotes Clarke's generalized gradient of V in the second variable. Subsequently, for the same problem, Vinter [17] proved the existence of an arc satisfying the full sensitivity relation

$$(H(\overline{x}(t), p(t)), -p(t)) \in \partial V(t, \overline{x}(t)) \text{ for all } t \in (t_0, T), \tag{1.6}$$

with ∂V equal to Clarke's generalized gradient in (t,x).

Full sensitivity relations were recognized as necessary and sufficient conditions for optimality in [4], where the first two authors of this paper studied the Mayer problem for the parameterized control system (1.1), with f depending also on time. More precisely, replacing the Clarke generalized gradient with the Fréchet superdifferential $\partial^+ V$, the full sensitivity relation

$$(H(t, \overline{x}(t), p(t)), -p(t)) \in \partial^+ V(t, \overline{x}(t)) \qquad \text{a.e. } t \in [t_0, T], \tag{1.7}$$

together with the maximum principle

$$\langle p(t), \dot{\overline{x}}(t) \rangle = H(t, \overline{x}(t), p(t)) \qquad \text{a.e. } t \in [t_0, T], \tag{1.8}$$

for some $p(t) \in \mathbb{R}^n$, was shown to actually characterize optimal trajectories. Earlier, in [15], a similar result had been proved under stronger regularity assumptions with $p(\cdot)$ equal to the dual arc.

Following the above papers, the analysis has been extended in several directions. For instance, in [5], sensitivity relations were adapted to the minimum time problem for the parameterized control system

$$\dot{x}(t) = f(x(t), u(t)) \quad t \geq 0, \tag{1.9}$$

taking the form of the inclusion

$$-p(t) \in \partial^+ T(\overline{x}(t)) \text{ for all } t \in [0, T(x_0)), \tag{1.10}$$

where $T(\cdot)$ denotes the minimum time function for a target K, $\partial^+ T$ is the Fréchet superdifferential of T, and $\overline{x}(\cdot)$ is an optimal trajectory starting from x_0 which attains K at time $T(x_0)$. In [6], the above result has been extended to nonparameterized systems described by differential inclusions. For optimal control problems with state constraints, sensitivity relations were derived in [3] and [12] using a suitable relaxation of the limiting subdifferential of the value function.

Obtaining sensitivity relations in terms of the Fréchet and/or proximal superdifferential of the value function for the differential inclusion

$$\dot{x}(s) \in F(x(s)) \qquad \text{a.e. } s \in [t_0, T], \tag{1.11}$$

with the initial condition

$$x(t_0) = x_0, \tag{1.12}$$

is far from straightforward, when F cannot be parameterized as

$$F(x) = \{f(x, u) \ : \ u \in U\}$$

with f smooth in x. The main goal of the present work is to prove both partial and full sensitivity relations for the Mayer problem

$$\inf \phi(x(T)), \tag{1.13}$$

the infimum being taken over all absolutely continuous functions $x : [t_0, T] \to \mathbb{R}^n$ that satisfy (1.11)-(1.12). The main assumptions we impose on the data, expressed in terms of the Hamiltonian

$$H(x, p) = \sup_{v \in F(x)} \langle v, p \rangle \qquad (x, p) \in \mathbb{R}^n \times \mathbb{R}^n, \tag{1.14}$$

require $H(\cdot, p)$ to be semiconvex, $H(x, \cdot)$ to be differentiable for $p \neq 0$, and $\nabla_p H(\cdot, p)$ locally Lipschitz continuous. We refer the reader to [8], where this set of hypotheses was used to obtain the semiconcavity of the value function, for a detailed discussion of their role in lack of a smooth parameterization of F.

For the Mayer problem (1.11)-(1.13), we shall derive sensitivity relations like (1.5) and (1.6) for both the proximal and Fréchet superdifferentials of the value function. More precisely, let $\overline{x} : [t_0, T] \to \mathbb{R}^n$ be an optimal trajectory of problem (1.13) and let $\overline{p} : [t_0, T] \to \mathbb{R}^n$ be an absolutely continuous function (that we call *dual arc*) such that the pair $(\overline{x}, \overline{p})$ solves the *Hamiltonian inclusion*

$$\begin{cases} -\dot{\overline{p}}(t) \in \partial_x^- H(\overline{x}(t), \overline{p}(t)) \\ \ \ \dot{\overline{x}}(t) \in \partial_p^- H(\overline{x}(t), \overline{p}(t)) \end{cases} \qquad \text{a.e. in } [t_0, T], \tag{1.15}$$

and

$$-\overline{p}(T) \in \partial \phi(\overline{x}(T)), \tag{1.16}$$

where $\partial_x^- H$ and $\partial_p^- H$ denote the Fréchet subdifferentials of H with respect to x and p, respectively [1]. Then we show that *all* such arcs $\overline{p}(\cdot)$ satisfy the proximal partial sensitivity relation

$$-\overline{p}(t) \in \partial_x^{+,pr} V(t, \overline{x}(t)) \text{ for all } t \in [t_0, T], \tag{1.17}$$

[1]We will see that $\overline{p}(\cdot)$ is in the set of differentiability of the map $H(x, \cdot)$ when $p(T) \neq 0$. In that case, the second inclusion in (1.15) becomes $\dot{\overline{x}}(t) = \nabla_p H(\overline{x}(t), \overline{p}(t))$ for all $t \in [t_0, T]$.

whenever

$$-\overline{p}(T) \in \partial^{+,pr}\phi(\overline{x}(T)), \tag{1.18}$$

where $\partial^{+,pr}$ denotes the proximal superdifferential. Moreover, replacing $\partial^{+,pr}\phi(\overline{x}(T))$ by the Fréchet superdifferential $\partial^{+}\phi(\overline{x}(T))$ in (1.18), we derive the full sensitivity relation

$$(H(\overline{x}(t),\overline{p}(t)),-\overline{p}(t)) \in \partial^{+}V(t,\overline{x}(t)) \text{ for all } t \in (t_0,T). \tag{1.19}$$

Thanks to (1.19) we can recover, under suitable assumptions, the same set of necessary and sufficient conditions for optimality that appears in the context of smooth parameterized systems.

From a technical viewpoint, we note that the proof of (1.17) and (1.19) is entirely different from the one for parameterized control systems. Indeed, in the latter case, the conclusion is obtained by appealing to the variational equation of (1.1). In the present context, such a strategy is impossible to follow because F admits no smooth parameterization, in general. As in [8], the role of the variational equation is here played by the maximum principle.

After obtaining sensitivity relations, we discuss two applications of (1.17) to the Mayer problem. Our first application concerns optimality conditions. Under our assumptions on H, the maximum principle in its available forms associates, with any optimal trajectory $\overline{x}:[t_0,T]\to\mathbb{R}^n$ of problem (1.13), a dual arc $\overline{p}:[t_0,T]\to\mathbb{R}^n$ such that $(\overline{x},\overline{p})$ satisfies (1.15) and the transversality condition

$$-\overline{p}(T) \in \partial\phi(\overline{x}(T)), \tag{1.20}$$

see, for instance, [9]. Here, for F locally strongly convex (see Section 2.6 for the definition of locally strongly convex multifunction), we construct multiple dual arcs $p(\cdot)$ satisfying the maximum principle

$$H(\overline{x}(t),p(t)) = \langle p(t),\dot{\overline{x}}(t)\rangle \quad \text{ a.e. in } [t_0,T], \tag{1.21}$$

by solving, for any $q \in \partial^{+,pr}\phi(\overline{x}(T))$, the *adjoint system*

$$\begin{cases} -\dot{p}(s) \in \partial_x^{-}H(\overline{x}(s),p(s)) & \text{a.e. in } [t_0,T]\,,\\ -p(T) = q. \end{cases}$$

Moreover, these dual arcs satisfy the full sensitivity relation (1.19).
Our second application aims to clarify the connection between the set of all reachable gradients of V at some point (t,x), $\partial^{*}V(t,x)$, and the optimal trajectories at (t,x). When the control system is parameterized as in (1.1), such a connection is fairly well understood: one can show that any nonzero reachable gradient of V at (t,x) can be associated with an optimal trajectory starting from (t,x), and the map from $\partial^{*}V(t,x)\setminus\{0\}$ into the family of optimal trajectories is one-to-one (see [7, Theorem 7.3.10]). In this paper, we use a suitable version of (1.17) to prove an analogue of the above result (see Theorem 5.3 for more details) which takes into account the lack of uniqueness for the initial value problem (1.15)-(1.18).

This paper is organized as follows. In Section 2, we set our notation, introduce the main assumptions, and recall preliminary results from nonsmooth analysis and control theory. In Section 3, sensitivity relations are derived in terms of the proximal and Fréchet superdifferentials. Finally, an application to the maximum principle is obtained in Section 4, and a result connecting reachable gradients of V with optimal trajectories in Section 5.

2 Preliminaries

2.1 Notation

Let us start by listing various basic notations and quickly reviewing some general facts for future use. Standard references are [7, 9].

We denote by $\mathbb{R}^+$ the set of strictly positive real numbers, by $|\cdot|$ the Euclidean norm in $\mathbb{R}^n$, and by $\langle \cdot, \cdot \rangle$ the inner product. $B(x, \epsilon)$ is the closed ball of radius $\epsilon > 0$ and center x. S^{n-1} is the unit sphere in $\mathbb{R}^n$. ∂E and $int(E)$ are the boundary and the interior of a subset E of $\mathbb{R}^n$, respectively. Given a continuous function $x : [t_0, T] \to \mathbb{R}^n$, the compact tubular neighborhood of $x([t_0, T])$ is defined by, for $r \geq 0$,

$$D_r(x([t_0, T])) := \{x \in \mathbb{R}^n : \ |x - x(t)| \leq r \text{ for some } t \in [t_0, T]\}.$$

For any continuous function $f : [t_0, t_1] \to \mathbb{R}^n$, let $\|f\|_\infty = \max_{t \in [t_0, t_1]} |f(t)|$. When f is Lebesgue integrable, let $\|f\|_{\mathcal{L}^1([t_0, t_1])} = \int_{t_0}^{t_1} |f(t)| \, dt$. $W^{1,1}([t_0, T]; \mathbb{R}^n)$ is the set of all absolutely continuous functions $x : [t_0, T] \to \mathbb{R}^n$, which we usually refer to as *arcs*.
Consider now a real-valued function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, where Ω is an open set, and suppose that f is locally Lipschitz. We denote by $\nabla f(\cdot)$ its gradient, which exists a.e. in Ω. A vector ζ is a *reachable gradient* of f at $x \in \Omega$ if there exists a sequence $\{x_i\} \subset \Omega$ such that f is differentiable at x_i for all $i \in \mathbb{N}$ and

$$x = \lim_{i \to \infty} x_i, \quad \zeta = \lim_{i \to \infty} \nabla f(x_i).$$

We denote by $\partial^* f(x)$ the set of all such vectors. Furthermore, the *(Clarke's) generalized gradient* of f at $x \in \Omega$, denoted by $\partial f(x)$, is the set of all the vectors ζ such that

$$\langle \zeta, v \rangle \leq \limsup_{\substack{y \to x, \\ h \to 0^+}} \frac{f(y + hv) - f(y)}{h}, \quad \forall v \in \mathbb{R}^n. \tag{2.1}$$

It is known that $co(\partial^* f(x)) = \partial f(x)$, where $co(A)$ denotes the convex hull of a subset A of $\mathbb{R}^n$.
Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ be any real-valued function defined on a open set $\Omega \subset \mathbb{R}^n$. For any $x \in \Omega$, the sets

$$\partial^- f(x) = \left\{ p \in \mathbb{R}^n : \liminf_{y \to x} \frac{f(y) - f(x) - \langle p, y - x \rangle}{| \ y - x \ |} \geq 0 \right\},$$
$$\partial^+ f(x) = \left\{ p \in \mathbb{R}^n : \limsup_{y \to x} \frac{f(y) - f(x) - \langle p, y - x \rangle}{| \ y - x \ |} \leq 0 \right\}$$

are called the *(Fréchet) subdifferential* and *superdifferential* of f at x, respectively. A vector $p \in \mathbb{R}^n$ is a *proximal supergradient* of f at $x \in \Omega$ if there exist two constants $c, \rho \geq 0$ such that

$$f(y) - f(x) - \langle p, y - x \rangle \leq c|y - x|^2, \ \forall y \in B(x, \rho).$$

The set of all proximal supergradients of f at x is called the *proximal superdifferential* of f at x, and is denoted by $\partial^{+,pr} f(x)$. Note that $\partial^{+,pr} f(x)$ is a subset of the Fréchet superdifferential of f at x.
For a mapping $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, associating to each $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ a real number, $\nabla_x f$, $\nabla_y f$ are its partial derivatives (when they do exist). The partial generalized gradient or partial Fréchet/proximal sub/superdifferential will be denoted in a similar way.

Let Ω be an open subset of $\mathbb{R}^n$. $C^1(\Omega)$ and $C^{1,1}(\Omega)$ are the spaces of all the functions with continuous and Lipschitz continuous first order derivatives on Ω, respectively.
Let $K \subset \mathbb{R}^n$ be a convex set. For $\overline{v} \in K$, recall that the *normal cone to K at $\overline{v}$* (in the sense of convex analysis) is the set

$$N_K(\overline{v}) = \{p \in \mathbb{R}^n : \langle p, v - \overline{v} \rangle \leq 0, \forall v \in K\}.$$

A well-known separation theorem implies that the normal cone at any $v \in \partial K$ contains a half line. Moreover, if K is not a singleton and has a C^1 boundary when $n > 1$, then all normal cones at the boundary points of K are half lines.
Finally, recall that a set-valued map $F : X \rightrightarrows Y$ is *strongly injective* if $F(x) \cap F(y) = \emptyset$ for any two distinct points $x, y \in X$.

2.2 Locally semiconcave functions

Here, we recall the notion of semiconcave function in $\mathbb{R}^n$ and list some results useful in this paper. Further details may be found, for instance, in [7].

We write $[x, y]$ to denote the segment with endpoints x, y for any $x, y \in \mathbb{R}^n$.

Definition 2.1. Let $A \subset \mathbb{R}^n$ be an open set. We say that a function $u : A \to \mathbb{R}$ is (linearly) *semiconcave* if it is continuous in A and there exists a constant c such that

$$u(x + h) + u(x - h) - 2u(x) \leq c|h|^2,$$

for all $x, h \in \mathbb{R}^n$ such that $[x - h, x + h] \subset A$. The constant c above is called a semiconcavity constant for u in A. We denote by $SC(A)$ the set of functions which are semiconcave in A. We say that a function u is semiconvex on A if and only if $-u$ is semiconcave on A.

Finally, recall that u is locally semiconcave in A if for each $x \in A$ there exists an open neighborhood of x where u is semiconcave.

In the literature, semiconcave functions are sometimes defined in a more general way. However, in the sequel we will mainly use the previous definition and properties recalled in the following proposition.

Proposition 2.2. *Let $A \subset \mathbb{R}^n$ be an open set, $u : A \to \mathbb{R}$ be a semiconcave function with a constant of semiconcavity c, and $x \in A$. Then,*

1. *a vector $p \in \mathbb{R}^n$ belongs to $\partial^+ u(x)$ if and only if*
$$u(y) - u(x) - \langle p, y - x \rangle \leq c|y - x|^2 \tag{2.2}$$
for any point $y \in A$ such that $[y, x] \subset A$. Consequently, $\partial^+ u(x) = \partial^{+,pr} u(x)$.

2. *$\partial u(x) = \partial^+ u(x) = co(\partial^* u(x))$.*

3. *If $\partial^+ u(x)$ is a singleton, then u is differentiable at x.*

If u is semiconvex, then (2.2) holds reversing the inequality and the sign of the quadratic term and the other two statements are true with the Fréchet/proximal subdifferential instead of the Frećhet/proximal superdifferential.
In proving our main results we shall require the semiconvexity of the map $x \mapsto H(x, p)$. Let us recall its consequence which will be used later on.

Lemma 2.3 ([8, Corollary 1]). *Let H be as in (1.14). If H is locally Lipschitz and the map $x \mapsto H(x,p)$ is locally semiconvex, uniformly for p in all bounded subsets of $\mathbb{R}^n$, then*

$$\partial H(x,p) \subset \partial_x^- H(x,p) \times \partial_p^- H(x,p), \quad \forall (x,p) \in \mathbb{R}^n \times \mathbb{R}^n.$$

2.3 Differential inclusions and standing assumptions

We recall that the Hausdorff distance between two compact sets $A_i \subset \mathbb{R}^n,\ i = 1,2$, is

$$dist_{\mathcal{H}}(A_1,A_2) = \max\{dist^+_{\mathcal{H}}(A_1,A_2), dist^+_{\mathcal{H}}(A_2,A_1)\},$$

where $dist^+_{\mathcal{H}}(A_1,A_2) = \inf\{\epsilon : A_1 \subset A_2 + B(0,\epsilon)\}$ is the semidistance. We say that a multifunction $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ with nonempty and compact values is locally Lipschitz if for each $x \in \mathbb{R}^n$ there exists a neighborhood K of x and a constant $c > 0$ depending on K so that $dist_{\mathcal{H}}(F(z), F(y)) \leq c \mid z - y \mid$ for all $z, y \in K$.
Throughout this paper, we assume that the multifunction F satisfies a collection of classical conditions of the theory of differential inclusions, the so-called *Standing Hypotheses*:

$$(SH) \begin{cases} (i) & F(x) \text{ is nonempty, convex, compact for each } x \in \mathbb{R}^n, \\ (ii) & F \text{ is locally Lipschitz with respect to the Hausdorff metric,} \\ (iii) & \exists \gamma > 0 \text{ so that } \max\{|v| : v \in F(x)\} \leq \gamma(1 + |x|)\, \forall x \in \mathbb{R}^n. \end{cases}$$

Assumptions (SH)(i)-(ii) guarantee the existence of local solutions of (1.11)-(1.12) and (SH)(iii) guarantees that solutions are defined on $[t_0, T]$. Basic notions concerning differential inclusions can be found, for instance, in the monograph [1].

For the sake of brevity, we usually refer to the Mayer problem (1.11)-(1.13) as $\mathcal{P}(t_0, x_0)$. Assuming (SH) and lower semicontinuity of ϕ implies that $\mathcal{P}(t_0, x_0)$ has at least one *optimal solution*, that is, a solution $\overline{x}(\cdot) \in W^{1,1}([t_0, T]; \mathbb{R}^n)$ of (1.11) and (1.12) such that

$$\phi(\overline{x}(T)) \leq \phi(x(T)),$$

for any trajectory $x(\cdot) \in W^{1,1}([t_0, T]; \mathbb{R}^n)$ of (1.11) satisfying (1.12). Actually, the Standing Hypotheses were first introduced assuming only the upper semicontinuity of F instead of $(SH)(ii)$. Although upper semicontinuity suffices to deduce the existence of optimal trajectories, in this paper we prefer to formulate (ii) as above because we will often take advantage of Lipschitz continuity.

Under assumption (SH) one can show that it is possible to associate with each optimal trajectory $x(\cdot)$ for $\mathcal{P}(t_0, x_0)$ an arc $p(\cdot)$ such that the pair $(x(\cdot), p(\cdot))$ satisfies an Hamiltonian inclusion.

Theorem 2.4 ([9, Theorem 3.2.6]). *Assume (SH) and that $\phi : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz. If $\overline{x}(\cdot)$ is an optimal solution for $\mathcal{P}(t_0, x_0)$, then there exists an arc $\overline{p} : [t_0, T] \to \mathbb{R}^n$ which, together with $\overline{x}(\cdot)$, solves the differential inclusion*

$$(-\dot{p}(s), \dot{x}(s)) \in \partial H(x(s), p(s)), \ \textit{a.e.} \quad s \in [t_0, T], \tag{2.3}$$

$$-\, p(T) \in \partial \phi(x(T)). \tag{2.4}$$

If (q, v) belongs to $\partial H(x,p)$, then $v \in F(x)$ and $\langle p, v \rangle = H(x,p)$. Thus, system (2.3) encodes the equality

$$H(\overline{x}(t), \overline{p}(t)) = \langle \overline{p}(t), \dot{\overline{x}}(t) \rangle \text{ for a.e. } t \in [t_0, T]. \tag{2.5}$$

This equality shows that the scalar product $\langle v, \overline{p}(t)\rangle$ is maximized over $F(\overline{x}(t))$ by $v = \dot{\overline{x}}(t)$. For this reason, the previous result is known as the *maximum principle* (in Hamiltonian form).
Recall now that the *value function* $V : (-\infty, T] \times \mathbb{R}^n \to \mathbb{R}$ associated with the Mayer problem is defined by: for all $(t_0, x_0) \in (-\infty, T] \times \mathbb{R}^n$,

$$V(t_0, x_0) = \inf \left\{ \phi(x(T)) : x \in W^{1,1}\left([t_0, T]; \mathbb{R}^n\right) \text{ satisfies (1.11) and (1.12)} \right\}. \tag{2.6}$$

Under assumptions (SH), V is locally Lipschitz and solves in the viscosity sense the Hamilton-Jacobi equation

$$\begin{cases} -\partial_t u(t,x) + H(x, -u_x(t,x)) = 0 & \text{in } (-\infty, T) \times \mathbb{R}^n, \\ u(T,x) = \phi(x), \ x \in \mathbb{R}^n, \end{cases} \tag{2.7}$$

where H is the Hamiltonian associated to F. Indeed, if the multifunction F satisfies assumption (SH), then it always admits a parameterization as a locally Lipschitz function (see, e.g., [2, Theorem 7.9.2]) and the result is well-known for the Lipschitz-parametric case (see, e.g., [7]).

Proposition 2.5. *Assume (SH) and that $\phi : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz. Then the value function of the Mayer problem is the unique viscosity solution of the problem* (2.7)*, where the Hamiltonian H is given by* (1.14)*.*

We conclude this part recalling that V satisfies the *dynamic programming principle*. Hence, if $y(\cdot)$ is any trajectory of the system (1.11)-(1.12), then the function $s \to V(s, y(s))$ is nondecreasing, and it is constant if and only if $y(\cdot)$ is optimal.

2.4 Sufficient conditions for optimality

In the control literature, it is well known that the full sensitivity relation involving the Fréchet superdifferential of V, coupled with the maximum principle, is a sufficient condition for optimality. For reader's convenience, we adapt this result to our context omitting the proof which is similar to the one of [4, Theorem 4.1]. Recall that (see, e.g., [2])

$$\begin{aligned} \partial^+ V(t,y) = \big\{ &(p', p'') \in \mathbb{R} \times \mathbb{R}^n : \\ &\forall (\theta', \theta'') \in \mathbb{R} \times \mathbb{R}^n, D_\downarrow V(t,y)(\theta', \theta'') \le p'\theta' + \langle p'', \theta'' \rangle \big\}, \end{aligned} \tag{2.8}$$

where the *upper Dini derivative* of V at (t,y) in the direction (θ', θ'') is given by

$$D_\downarrow V(t,y)(\theta', \theta'') := \limsup_{\tau \to 0^+} \frac{V(t + \tau\theta', y + \tau\theta'') - V(t,y)}{\tau}. \tag{2.9}$$

Theorem 2.6. *Assume (SH), suppose ϕ is locally Lipschitz, and let $x : [t_0, T] \to \mathbb{R}^n$ be a solution of system* (1.11)-(1.12)*. If, for almost every $t \in [t_0, T]$, there exists $p(t) \in \mathbb{R}^n$ such that*

$$\begin{gathered} \langle p(t), \dot{x}(t) \rangle = H(x(t), p(t)), \\ (H(x(t), p(t)), -p(t)) \in \partial^+ V(t, x(t)), \end{gathered} \tag{2.10}$$

then x is optimal for problem $\mathcal{P}(t_0, x_0)$.

2.5 Main assumptions

We impose further conditions on the Hamiltonian associated with F:

$$(H1)\begin{cases} \text{For each nonempty, convex and compact subset } K \subseteq \mathbb{R}^n, \\ (i)\ \exists\ c \geq 0 \text{ so that } , \forall p \in \mathbb{R}^n, x \mapsto H(x,p) \text{ is semiconvex on } K \text{ with constant } c|p|, \\ (ii)\ \nabla_p H(x,p) \text{ exists and is Lipschitz continuous in } x \text{ on } K, \text{ uniformly for } p \in \mathbb{R}^n \setminus \{0\}. \end{cases}$$

Remark 2.7. Note that $H(x,\cdot)$ is positively homogeneous of degree one, and $\nabla_p H(x,\cdot)$ is positively homogeneous of degree zero. Then, assuming $(H1)$ is equivalent to requiring that

$$(H1^*)\begin{cases} \text{for each non empty, convex and compact subset } K \subseteq \mathbb{R}^n, \\ (i)\ \exists\ c \geq 0 \text{ so that } , \forall p \in S^{n-1}, x \mapsto H(x,p) \text{ is semiconvex on } K \text{ with constant } c, \\ (ii)\ \nabla_p H(x,p) \text{ exists and is Lipschitz continuous in } x \text{ on } K, \text{ uniformly for } p \in S^{n-1}. \end{cases}$$

Some examples of multifunctions satisfying (SH) and $(H1)$ are given in [8].

Let us start by analyzing the meaning of the above assumptions beginning with $(H1)(i)$ which is equivalent to the *mid-point property* of the multifunction F on K, that is,

$$dist^+_{\mathcal{H}}\left(2F(x), F(x+z)+F(x-z)\right) \leq c \mid z \mid^2,$$

for all x, z so that $x, x \pm z \in K$. Another consequence of $(H1)(i)$ is the fact that the generalized gradient of H splits into two components, as described in Lemma 2.3. This implies that every solution of (2.3) is also a solution of the Hamiltonian inclusion (2.11) below.

Theorem 2.8 ([8, Corollary 2])**.** *Assume that (SH) and $(H1)(i)$ hold and suppose $\phi : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz. If $\overline{x}(\cdot)$ is an optimal solution for $\mathcal{P}(t_0, x_0)$, then there exists an arc $\overline{p} : [t_0, T] \to \mathbb{R}^n$ which, together with $\overline{x}(\cdot)$, satisfies the system*

$$\begin{cases} -\dot{p}(s) & \in & \partial_x^- H(x(s), p(s)), \\ \dot{x}(s) & \in & \partial_p^- H(x(s), p(s)), \end{cases} \quad a.e. \quad s \in [t_0, T] \tag{2.11}$$

and the transversality condition

$$-p(T) \in \partial\phi(x(T)). \tag{2.12}$$

Given an optimal trajectory $\overline{x}(\cdot)$, any arc $\overline{p}(\cdot)$ satisfying (2.11) and the *tranversality condition* (2.12) is called a *dual arc* associated with $\overline{x}(\cdot)$. Recall that any solution to (2.3) solves also the system (2.11).

Remark 2.9. Under the additional hypothesis that $\partial^+\phi(\overline{x}(T)) \neq \emptyset$, for every $q \in \partial^+\phi(\overline{x}(T))$ there exists an arc $\overline{p}$ such that the pair $(\overline{x}, \overline{p})$ solves (2.11) and the condition $-p(T) = q$. Indeed, since $q \in \partial^+\phi(\overline{x}(T))$, there exists a function $g \in C^1(\mathbb{R}^n; \mathbb{R})$ such that $g \geq \phi$, $g(\overline{x}(T)) = \phi(\overline{x}(T))$, and $\nabla g(\overline{x}(T)) = q$ (see, for instance, [7, Proposition 3.1.7]). Note that $\overline{x}$ is still optimal for the Mayer problem (1.11)-(1.13) with ϕ replaced by g. Thus, by Theorem 2.8 there exists an arc $\overline{p}$ such that the pair $(\overline{x}, \overline{p})$ solves (2.11) and satisfies the terminal condition $-p(T) = q$.

Remark 2.10. Let $x : [t_0, T] \to \mathbb{R}^n$ be continuous and p be a solution of $-\dot{p}(s) \in \partial_x^- H(x(s), p(s))$ a.e. in $[t_0, T]$. Observe that, if p vanishes at some time $t \in [t_0, T]$, then it must vanish at every time. Indeed, let $K \subset \mathbb{R}^n$ be a compact set containing $x([t_0, T])$. Denoting by c_K a Lipschitz constant of F on K, we have that $c_K|p|$ is a Lipschitz constant for $H(\cdot, p)$ on the same set. Indeed, let $x, y \in K$ and v_x be such that $H(x,p) = \langle v_x, p\rangle$. By (SH), there exists $v_y \in F(y)$ such that

$$H(x,p) - H(y,p) \leq \langle v_x - v_y, p\rangle \leq c_K|p||x-y|. \tag{2.13}$$

Recalling (2.1), it follows that

$$|\zeta| \le c_K|p|, \;\forall \zeta \in \partial_x H(x,p), \forall x \in K, \forall p \in \mathbb{R}^n. \tag{2.14}$$

Hence, in view of the differential inclusion verified by p,

$$|\dot{p}(s)| \le c_K|p(s)|, \text{ for a.e. } s \in [t_0, T]. \tag{2.15}$$

By Gronwall's Lemma, we obtain that either $p(s) \neq 0$ for every $s \in [t_0, T]$, or $p(s) = 0$ for every $s \in [t_0, T]$. Consequently, under assumptions $(H1)$ a dual arc associated with an optimal trajectory vanishes either for all times or never.

Concerning the assumption $(H1)(ii)$, the existence of the gradient of H with respect to p is equivalent to the fact that the argmax set of $\langle v, p\rangle$ over $v \in F(x)$ is the singleton $\{\nabla_p H(x,p)\}$, for each $p \neq 0$. Thus, the following relation holds:

$$H(x,p) = \langle \nabla_p H(x,p), p\rangle, \;\forall p \neq 0. \tag{2.16}$$

Moreover, it is easy to see that, for every x, the boundary of the sets $F(x)$ contains no line segment. The main consequence of the local Lipschitzianity of the map $x \mapsto \nabla_p H(x,p)$ is the following result, the proof of which is straightforward.

Lemma 2.11 ([8, Proposition 3]). *Assume (SH) and $(H1)$. Let $p : [t,T] \to \mathbb{R}^n$ be continuous and nonvanishing in $[t,T]$. Then, for each $x \in \mathbb{R}^n$, the Cauchy problem*

$$\begin{cases} \dot{y}(s) = \nabla_p H(y(s), p(s)) & \text{for all } s \in [t,T]\,, \\ y(t) = x, \end{cases} \tag{2.17}$$

has a unique solution $y(\cdot; t, x)$. Moreover, for every $r > 0$ there exists a constant k such that

$$|y(s;t,x) - y(s;t,z)| \le e^{k(T-t)}|z - x|, \;\forall z, x \in B(0,r), \forall s \in [t,T]. \tag{2.18}$$

Remark 2.12. Note that the map $p \mapsto \nabla_p H(x,p)$ is continuous for $p \neq 0$. Thus, the local Lipschitzianity of the map $x \mapsto \nabla_p H(x,p)$ implies that $(x,p) \mapsto \nabla_p H(x,p)$ is a continuous map, for $p \neq 0$. This is the reason why the ODE (2.17) is verified everywhere on $[t,T]$, not just almost everywhere. Suppose now $x(\cdot)$ is optimal for $\mathcal{P}(t_0, x_0)$ and $p(\cdot)$ is any nonvanishing dual arc associated with $x(\cdot)$—if it does exist. Then, Lemma 2.11 implies that $x(\cdot)$ is the unique solution of the Cauchy problem (2.17) with $t = t_0$, $x(t_0) = x_0$, and $p(\cdot)$ equals such a dual arc. Furthermore, in this case, $x(\cdot)$ is of class C^1 and the maximum principle (2.5) holds true for all $t \in [t_0, T]$, interpreting $\dot{x}$ as the left and right derivatives[2] of x at T at t_0, respectively.

2.6 R-convex sets

Let A be a compact and convex subset of $\mathbb{R}^n$ and $R > 0$.

Definition 2.13. The set A is $R-$*convex* if, for each $z, y \in \partial A$ and any vectors $n \in N_A(z), m \in N_A(y)$ with $\mid n \mid = \mid m \mid = 1$, the following inequality holds true

$$\mid z - y \mid \le R \mid n - m \mid . \tag{2.19}$$

[2]This standard convention will be used throughout the paper.

The concept of $R-$convex set is not new. It is a special case of *hyperconvex sets* (with respect to the ball of radius R and center zero) introduced by Mayer in [13]. A study of hyperconvexity appears also in [16, 14]. The notion of R-convexity was considered, among others, by Levitin, Poljak, Frankowska, Olech, Pliś, Lojasiewicz, and Vian (they called these sets R-regular, R-convex, as well strongly convex). We first recall some interesting characterizations of $R-$convex sets.

Proposition 2.14 ([11, Proposition 3.1])**.** *Let A be a compact and convex subset of $\mathbb{R}^n$. Then the following conditions are equivalent*

1. *A is $R-$convex,*
2. *A is the intersection of a family of closed balls of radius R,*
3. *for any two points $x, y \in \partial A$ such that $|x-y| \leq 2R$, each arc of a circle of radius R which joins x and y and whose length is not greater that πR is contained in A,*
4. *for each $z \in \partial A$ and any $n \in N_A(z), \mid n \mid = 1$, the ball of center $z - Rn$ and radius R contains A, that is $\mid z - Rn - x \mid \leq R$ for each $x \in A$,*
5. *for each $z \in \partial A$ and any vector $n \in N_A(z)$ with $|n| = 1$, we have the inequality*

$$|z-x| \leq \sqrt{2R}\langle z-x, n\rangle^{\frac{1}{2}}, \ \forall x \in A. \tag{2.20}$$

R-convex sets are obviously convex. Moreover, the boundary of an R-convex set A satisfies a generalized lower bound for the curvature, even though ∂A may be a nonsmooth set. Indeed, for every point $x \in \partial A$ there exists a closed ball B_x of radius R such that $x \in \partial B_x$ and $A \subset B_x$. This fact suggests that, in some sense, the curvature of ∂A is bounded below by $1/R$.

Definition 2.15. A multifunction $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *locally strongly convex* if for each compact set $K \subset \mathbb{R}^n$ there exists $R > 0$ such that $F(x)$ is R-convex for every $x \in K$.

We can reformulate the above property of F in an equivalent Hamiltonian form. Here, we denote by $F_p(x)$ the argmax set of $\langle v, p\rangle$ over $v \in F(x)$. The existence of $\nabla_p H(x,p)$ is equivalent to the fact that the set $F_p(x)$ is the singleton $\{\nabla_p H(x,p)\}$, whenever $p \neq 0$.

$$(H2) \begin{cases} \text{For every compact } K \subset \mathbb{R}^n, \text{ there exists a constant } c' = c'(K) > 0 \text{ such that for all} \\ x \in K,\ p \in \mathbb{R}^n, \text{ we have: } v_p \in F_p(x) \Rightarrow \langle v - v_p, p\rangle \leq -c'|p||v-v_p|^2,\ \forall v \in F(x). \end{cases}$$

In the next lemma we show that the local strong convexity of F is equivalent to assumption $(H2)$ for the associated Hamiltonian, giving also a result connecting $(H2)$ with the regularity of H.

Lemma 2.16. *Suppose that $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a multifunction satisfying (SH). Let K be any convex and compact subset of $\mathbb{R}^n$. Then*

1. *$(H2)$ holds with a constant c' on K if and only if $F(x)$ satisfies the R-convexity property for all $x \in K$ with radius $R = (2c')^{-1}$.*
2. *If $(H2)$ holds, then $\nabla_p H(x,p)$ exists for all $x \in K$ and $p \in \mathbb{R}^n \setminus \{0\}$ and is Hölder continuous in x on K with exponent $1/2$, uniformly for $p \in \mathbb{R}^n \setminus \{0\}$.*

Proof. For all $x \in K$ and $v \in \partial F(x)$, we have $v \in F_{y_v}(x)$ for all $y_v \in N_{F(x)}(v)$. Therefore, $(H2)$ holds with constant c' on K if and only if for any $y_v \in N_{F(x)}(v)$ with $\mid y_v \mid = 1$ we have

$$\langle v - \overline{v}, y_v \rangle \geq c' |v - \overline{v}|^2, \ \forall \overline{v} \in F(x),$$

or equivalently,

$$\mid v - \overline{v} \mid \leq \sqrt{2 \frac{1}{2c'}} \langle v - \overline{v}, y_v \rangle^{\frac{1}{2}}, \ \forall \overline{v} \in F(x),$$

for all $y_v \in N_{F(x)}(v)$ with $\mid y_v \mid = 1$. By Proposition 2.14, this is equivalent to the $(2c')^{-1}-$convexity of $F(x)$ for each $x \in K$. For the proof of the second statement we refer to [8, Proposition 4]. □

The second statement of the above lemma is not an equivalence, in general, as is shown in the example below. Moreover, assumption $(H1)$ does not follow from $(H2)$.

Example 2.17. Let us denote by $M \subset \mathbb{R}^2$ the intersection of the epigraph of the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^4$, and the closed ball $B(0, R)$, $R > 0$. Let us consider the multifunction $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ that associates set M with any $x \in \mathbb{R}^2$. Observe that M fails to be strongly convex, since the curvature at $x = 0$ is equal to zero. Moreover, since M is a closed convex set and its boundary contains no line, the argmax of $\langle v, p \rangle$ over $v \in M$ is a singleton for each $p \neq 0$. So, the Hamiltonian $H(p) = \sup_{v \in M} \langle v, p \rangle$ is differentiable for each $p \neq 0$. Note that the gradient $\nabla_p H$ is constant with respect to the x variable. Consequently, the Hamiltonian satisfies $(H1)$ but F does not satisfy $(H2)$.

3 Sensitivity relations

In this section, we discuss partial and full sensitivity relations for both the Fréchet and proximal superdifferential of the value function.

Theorem 3.1. *Assume (SH), $(H1)$ and let $\phi : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz. Let $\overline{x} : [t_0, T] \to \mathbb{R}^n$ be an optimal solution for $\mathcal{P}(t_0, x_0)$ and let $\overline{p} : [t_0, T] \to \mathbb{R}^n$ be an arc such that $(\overline{x}, \overline{p})$ solves the system*

$$\begin{cases} -\dot{p}(t) & \in & \partial_x^- H(\overline{x}(t), p(t)) \\ \dot{\overline{x}}(t) & \in & \partial_p^- H(\overline{x}(t), p(t)) \end{cases} \qquad a.e. \ in \ [t_0, T], \tag{3.1}$$

and satisfies the transversality condition

$$-\overline{p}(T) \in \partial^{+,pr} \phi(\overline{x}(T)). \tag{3.2}$$

Then, there exist constants $c, R > 0$ such that, for all $t \in [t_0, T]$ and all $h \in B(0, R)$,

$$V(t, \overline{x}(t) + h) - V(t, \overline{x}(t)) \leq \ \langle -\overline{p}(t), h \rangle + c \mid h \mid^2 . \tag{3.3}$$

Consequently, $\overline{p}(\cdot)$ satisfies the proximal partial sensitivity relation

$$-\overline{p}(t) \in \partial_x^{+,pr} V(t, \overline{x}(t)) \ for \ all \ t \in [t_0, T]. \tag{3.4}$$

To prove the above theorem, we need the following lemma.

Lemma 3.2. *Assume $\overline{p}(T) \neq 0$ and fix $t \in [t_0, T)$. For $h \in \mathbb{R}^n$, let $x_h : [t,T] \to \mathbb{R}^n$ be the solution of the problem*

$$\begin{cases} \dot{x}(s) = \nabla_p H(x(s), \overline{p}(s)), & s \in [t, T], \\ x(t) = \overline{x}(t) + h. \end{cases} \tag{3.5}$$

Let us consider a compact tubular neighborhood of $\overline{x}([t_0,T])$, $D_r(\overline{x}([t_0,T]))$, for some $r > 0$. Then, there exist constants k and c_1, depending only on r, such that for all $h \in B\left(0, re^{-k(T-t_0)}\right)$ it holds that

$$|x_h(s) - \overline{x}(s)| \leq e^{k(T-t_0)} \mid h \mid, \text{ for all } s \in [t,T], \tag{3.6}$$

and

$$\langle \overline{p}(t), h \rangle + \langle -\overline{p}(T), x_h(T) - \overline{x}(T) \rangle \leq c_1 \mid h \mid^2 . \tag{3.7}$$

Proof. First, recall that, thanks to Remark 2.10, $\overline{p}(t) \neq 0$ for all $t \in [t_0, T]$. Hence, $\overline{x}(\cdot)$ is the unique solution of the Cauchy problem

$$\begin{cases} \dot{x}(s) = \nabla_p H(x(s), \overline{p}(s)) \text{ for all } s \in [t, T], \\ x(t) = \overline{x}(t). \end{cases} \tag{3.8}$$

Let $k > 0$ be a Lipschitz constant for $\nabla_p H(\cdot, p)$ on $D_r(\bar{x}([t_0,T]))$ for $p \neq 0$. We claim that x_h takes values in $D_r(\bar{x}([t_0,T]))$ for all $h \in B\left(0, re^{-k(T-t_0)}\right)$. By contradiction, suppose that there exists $\overline{s}$, $t < \overline{s} < T$, such that $|x_h(\overline{s}) - \overline{x}(\overline{s})| = r$ and $|x_h(s) - \overline{x}(s)| < r$ for all $s \in [t, \overline{s})$. Then, by standard arguments based on Gronwall's lemma we conclude that

$$|\overline{x}(\overline{s}) - x_h(\overline{s})| \leq e^{k(\overline{s}-t)}|h|. \tag{3.9}$$

Since $\overline{s} < T$, for all $h \in B\left(0, re^{-k(T-t_0)}\right)$ it holds that

$$|\overline{x}(\overline{s}) - x_h(\overline{s})| < r. \tag{3.10}$$

This contradicts the fact that $|x_h(\overline{s}) - \overline{x}(\overline{s})| = r$. Thus, $x_h([t,T]) \subset D_r(\overline{x}([t_0,T]))$ for all $h \in B\left(0, re^{-k(T-t_0)}\right)$, and using again the Gronwall's lemma it is easy to deduce that (3.6) holds true for all $h \in B\left(0, re^{-k(T-t_0)}\right)$.

In order to prove (3.7), note that

$$\langle \overline{p}(t), h \rangle + \langle -\overline{p}(T), x_h(T) - \overline{x}(T) \rangle = \int_t^T \frac{d}{ds} \langle -\overline{p}(s), x_h(s) - \overline{x}(s) \rangle ds$$

$$= \int_t^T \langle -\dot{\overline{p}}(s), x_h(s) - \overline{x}(s) \rangle ds \; + \int_t^T \langle -\overline{p}(s), \dot{x}_h(s) - \dot{\overline{x}}(s) \rangle \; ds := (I) + (II).$$

Let $c' > 0$ be such that $c'|p|$ is the semiconvexity constant of $H(\cdot, p)$ on $D_r(\overline{x}([t_0,T]))$. Since $-\dot{\overline{p}}(s) \in \partial_x^- H(\overline{x}(s), \overline{p}(s))$ a.e. in $[t_0, T]$, we obtain that

$$(I) \leq \int_t^T \left(c' \mid \overline{p}(s) \mid \cdot \mid x_h(s) - \overline{x}(s) \mid^2 + H(x_h(s), \overline{p}(s)) - H(\overline{x}(s), \overline{p}(s)) \right) \; ds.$$

Then, from (3.6),

$$(I) \leq \int_t^T \left(c' \mid \overline{p}(s) \mid \cdot \mid h \mid^2 e^{2k(T-t_0)} + H(x_h(s), \overline{p}(s)) - H(\overline{x}(s), \overline{p}(s)) \right) \; ds$$

$$\leq (T-t_0)c' \parallel \overline{p} \parallel_\infty e^{2k(T-t_0)}|h|^2 + \int_t^T H\left(x_h(s), \overline{p}(s)\right) - H(\overline{x}(s), \overline{p}(s)))\ ds.$$

Now recalling (2.16), (3.5) and (3.8), we get

$$\begin{aligned}(II) &= \int_t^T \langle -\overline{p}(s), \nabla_p H(x_h(s), \overline{p}(s)) - \nabla_p H(\overline{x}(s), \overline{p}(s))\rangle ds \\ &= \int_t^T \left(-H(x_h(s), \overline{p}(s)) + H(\overline{x}(s), \overline{p}(s))\right) ds.\end{aligned}$$

Adding up the previous relations, it follows that (3.7) holds true with $c_1 := (T-t_0)c' \parallel \overline{p} \parallel_\infty e^{2k(T-t_0)}$. The lemma is proved. □

Proof of Theorem 3.1. First note that estimate (3.3) is immediate for $t = T$. Let us observe that, in view of Remark 2.10, it holds that:

(i) either $\overline{p}(t) \neq 0$ for all $t \in [t_0, T]$,

(ii) or $\overline{p}(t) = 0$ for all $t \in [t_0, T]$.

We shall analyze each of the above cases separately. Suppose, first, that $\overline{p}(t) \neq 0$ for all $t \in [t_0, T]$ and fix $t \in [t_0, T)$. Then $\overline{x}(\cdot)$ is the unique solution of the Cauchy problem

$$\begin{cases} \dot{x}(s) = \nabla_p H(x(s), \overline{p}(s)) \quad \text{for all } s \in [t, T]\,, \\ x(t) = \overline{x}(t). \end{cases} \tag{3.11}$$

First note that, owing to (3.2), there exist constants $c, r > 0$ so that, for all $z \in B(0, r)$,

$$\phi(\overline{x}(T) + z) - \phi(\overline{x}(T)) \leq \langle -\overline{p}(T), z\rangle + c|z|^2. \tag{3.12}$$

Let $k > 0$ be a Lipschitz constant for $\nabla_p H(\cdot, p)$ on $D_r(\overline{x}([t_0, T]))$, where $D_r(\overline{x}([t_0, T]))$ is a compact tubular neighborhood of $\overline{x}([t_0, T])$ as in Lemma 3.2. For each $h \in B\left(0, re^{-k(T-t_0)}\right)$, let $x_h(\cdot)$ be the solution of problem (3.5). By the optimality of $\overline{x}(\cdot)$, the very definition of the value function, and the dynamic programming principle, we have that, for all $h \in B\left(0, re^{-k(T-t_0)}\right)$,

$$V(t, \overline{x}(t) + h) - V(t, \overline{x}(t)) + \langle \overline{p}(t), h\rangle \leq \phi(x_h(T)) - \phi(\overline{x}(T)) + \langle \overline{p}(t), h\rangle. \tag{3.13}$$

Since $|x_h(T) - \overline{x}(T)| \leq r$, from (3.12) and (3.13) we conclude that for each $h \in B\left(0, re^{-k(T-t_0)}\right)$,

$$\begin{aligned} &V(t, \overline{x}(t) + h) - V(t, \overline{x}(t)) + \langle \overline{p}(t), h\rangle \\ &\quad \leq \langle \overline{p}(t), h\rangle + \langle -\overline{p}(T), x_h(T) - \overline{x}(T)\rangle + c \mid x_h(T) - \overline{x}(T) \mid^2 . \end{aligned} \tag{3.14}$$

Therefore, in view of (3.14) and Lemma 3.2, we have that for all $h \in B\left(0, re^{-k(T-t_0)}\right)$,

$$V(t, \overline{x}(t) + h) - V(t, \overline{x}(t)) + \langle \overline{p}(t), h\rangle \leq c_2 \mid h \mid^2, \tag{3.15}$$

where $c_2 := c_1 + ce^{2k(T-t_0)}$. The proof of (3.3) in case (i) is thus complete.
Next, suppose we are in case (ii), that is $p(t) = 0$ for all $t \in [t_0, T]$. Let $t \in [t_0, T)$ be fixed. Then, by Filippov's Theorem (see, e.g., Theorem 10.4.1 in [2]), there exists a constant k_1, independent of $t \in [t_0, T]$, such that, for any $h \in \mathbb{R}^n$ with $\mid h \mid\leq 1$, the initial value problem

$$\begin{cases} \dot{x}(s) \in F(x(s)) \quad a.e. \ in\ [t, T], \\ x(t) = \overline{x}(t) + h. \end{cases} \tag{3.16}$$

has a solution, $x_h(\cdot)$, that satisfies the inequality

$$\| x_h - \overline{x} \|_\infty \leq e^{k_1(T-t_0)} \mid h \mid . \tag{3.17}$$

By the optimality of $\overline{x}(\cdot)$, the very definition of the value function, and the dynamic programming principle it follows that

$$V(t, \overline{x}(t) + h) - V(t, \overline{x}(t))\rangle \leq \phi(x_h(T)) - \phi(\overline{x}(T)). \tag{3.18}$$

Moreover, owing to (3.2) an recalling that $\overline{p}(\cdot)$ is equal to zero at each point, there exist constants $c, r > 0$ so that, for all $z \in B(0,r)$,

$$\phi(\overline{x}(T) + z) - \phi(\overline{x}(T)) \leq c|z|^2. \tag{3.19}$$

Let $r_1 := \min\{1, re^{-k_1(T-t_0)}\}$. In view of (3.17)-(3.19), we obtain that, for all $t \in [t_0, T]$ and $h \in B(0, r_1)$,

$$V(t, \overline{x}(t) + h) - V(t, \overline{x}(t)) \leq ce^{2k_1(T-t_0)} \mid h \mid^2 . \tag{3.20}$$

The proof is complete also in case (ii). $\square$

Remark 3.3. The above proof shows also that: if $\overline{x} : [t_0, T] \to \mathbb{R}^n$ is optimal for $\mathcal{P}(t_0, x_0)$ and $\overline{p} : [t_0, T] \to \mathbb{R}^n$ is an absolutely continuous function such that $(\overline{x}, \overline{p})$ solves the system

$$\left\{ \begin{array}{rcl} -\dot{p}(t) & \in & \partial_x^- H(\overline{x}(t), p(t)) \\ \dot{\overline{x}}(t) & \in & \partial_p^- H(\overline{x}(t), p(t)), \end{array} \right. \quad -\overline{p}(T) \in \partial^{+,pr}\phi(\overline{x}(T)) \quad \text{a.e. in } \left[\overline{t}, T\right], \tag{3.21}$$

for some $t_0 \leq \overline{t} < T$, then (3.3) holds true for all $t \in [\overline{t}, T]$, with uniform constants R, c on $[\overline{t}, T]$.

Remark 3.4. One can easily adapt the previous proof to show that the above inclusion holds true with the Fréchet superdifferential as well, that is, if $-\overline{p}(T) \in \partial^+\phi(\overline{x}(T))$, then

$$-\overline{p}(t) \in \partial_x^+ V(t, \overline{x}(t)) \text{ for all } t \in [t_0, T].$$

In this case, the term $c \mid h \mid^2$ in (3.3) is replaced by $o(\mid h \mid)$.

Theorem 3.5. *Assume (SH), (H1) and let $\phi : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz. Let $\overline{x} : [t_0, T] \to \mathbb{R}^n$ be an optimal solution for problem $\mathcal{P}(t_0, x_0)$ and let $\overline{p} : [t_0, T] \to \mathbb{R}^n$ be such that $(\overline{x}, \overline{p})$ solves the system*

$$\left\{ \begin{array}{rcl} -\dot{p}(t) & \in & \partial_x^- H(\overline{x}(t), p(t)) \\ \dot{\overline{x}}(t) & \in & \partial_p^- H(\overline{x}(t), p(t)) \end{array} \right. \quad a.e. \text{ in } [t_0, T], \tag{3.22}$$

and satisfies the transversality condition

$$-\overline{p}(T) \in \partial^+\phi(\overline{x}(T)). \tag{3.23}$$

Then, $\overline{p}(\cdot)$ satisfies the full sensitivity relation

$$(H(\overline{x}(t), \overline{p}(t)), -\overline{p}(t)) \in \partial^+ V(t, \overline{x}(t)) \text{ for all } t \in (t_0, T). \tag{3.24}$$

Proof. In view of Remark 2.10, it holds that

(i) either $\overline{p}(t) \neq 0$ for all $t \in [t_0, T]$,

(ii) or $\overline{p}(t) = 0$ for all $t \in [t_0, T]$.

Suppose to be in case (i), that is $\overline{p}(t) \neq 0$ for all $t \in [t_0, T]$. Let $t \in (t_0, T)$ be fixed. Hence, $\overline{x}(\cdot)$ is the unique solution of the Cauchy problem

$$\begin{cases} \dot{x}(s) = \nabla_p H(x(s), \overline{p}(s)) \text{ for all } s \in [t, T], \\ x(t) = \overline{x}(t). \end{cases} \tag{3.25}$$

Consider now any $(\alpha, \theta) \in \mathbb{R} \times \mathbb{R}^n$ and, for every $\tau > 0$, let x_τ be the solution of the differential equation

$$\begin{cases} \dot{x}(s) = \nabla_p H(x(s), \overline{p}(s)) \text{ for all } s \in [t, T], \\ x(t) = \overline{x}(t) + \tau\theta. \end{cases} \tag{3.26}$$

By (2.9) and (3.26), we have that

$$D_\downarrow V(t, \overline{x}(t))(\alpha, \alpha\dot{\overline{x}}(t) + \theta) = \limsup_{\tau \to 0^+} \frac{V(t + \alpha\tau, x_\tau(t) + \tau\alpha\dot{\overline{x}}(t)) - V(t, \overline{x}(t))}{\tau}. \tag{3.27}$$

Moreover, from (3.25) and (3.26),

$$\begin{aligned} | \, x_\tau(t + \alpha\tau) - x_\tau(t) - \tau\alpha\dot{\overline{x}}(t) \, | &\leq \left| \int_t^{t+\alpha\tau} |\nabla_p H(x_\tau(s), \overline{p}(s)) - \nabla_p H(\overline{x}(t), \overline{p}(t))| \, ds \right| \\ \leq \left| \int_t^{t+\alpha\tau} |\nabla_p H(x_\tau(s), \overline{p}(s)) - \nabla_p H(\overline{x}(s), \overline{p}(s))| \; ds \right| & \\ + \left| \int_t^{t+\alpha\tau} |\nabla_p H(\overline{x}(s), \overline{p}(s)) - \nabla_p H(\overline{x}(t), \overline{p}(t))| \; ds \right|. & \end{aligned} \tag{3.28}$$

By (3.6), (3.28), using also that the map $x \mapsto \nabla_p H(x, p)$ is locally Lipschitz for $p \neq 0$ and the map $s \mapsto \nabla_p H(\overline{x}(s), \overline{p}(s))$ is continuous, we conclude that

$$| \, x_\tau(t + \alpha\tau) - x_\tau(t) - \tau\alpha\dot{\overline{x}}(t) \, | = o(\tau). \tag{3.29}$$

Hence, from (3.27), (3.29), using that V is locally Lipschitz, the dynamic programming principle, and the transversality condition (3.23) we deduce that

$$\begin{aligned} D_\downarrow V(t, \overline{x}(t))(\alpha, \alpha\dot{\overline{x}}(t) + \theta) &\leq \limsup_{\tau \to 0^+} \frac{V(t + \alpha\tau, x_\tau(t + \alpha\tau)) - V(t, \overline{x}(t))}{\tau} \\ \leq \limsup_{\tau \to 0^+} \frac{\phi(x_\tau(T)) - \phi(\overline{x}(T))}{\tau} &\leq \limsup_{\tau \to 0^+} \frac{\langle -\overline{p}(T), x_\tau(T) - \overline{x}(T) \rangle}{\tau}. \end{aligned} \tag{3.30}$$

In view of (3.7), the above upper limit does not exceed $\langle -\overline{p}(t), \theta \rangle$. Recalling Remark 2.12, we have that $H(\overline{x}(t), \overline{p}(t)) = \langle \dot{\overline{x}}(t), \overline{p}(t) \rangle$ for all $t \in [t_0, T]$. Thus, we finally obtain

$$D_\downarrow V(t, \overline{x}(t))(\alpha, \alpha\dot{\overline{x}}(t) + \theta) \leq \alpha H(\overline{x}(t), \overline{p}(t)) + \langle -\overline{p}(t), \alpha\dot{\overline{x}}(t) + \theta \rangle. \tag{3.31}$$

Hence, for all $\alpha \in \mathbb{R}$ and $\theta_1 \in \mathbb{R}^n$,

$$D_\downarrow V(t, \overline{x}(t))(\alpha, \theta_1) \leq \alpha H(\overline{x}(t), \overline{p}(t)) + \langle -\overline{p}(t), \theta_1 \rangle. \tag{3.32}$$

The conclusion (3.24), in case (i), follows from (2.8), (2.9), and (3.32).

Now, suppose to be in case (ii), that is $p(s) = 0$ for all $s \in [t_0, T]$. Thanks to (2.8), (2.9), and the fact that $H(x, 0) = 0$ the inclusion (3.24) holds true if and only if, for all $(\alpha, \theta) \in \mathbb{R} \times \mathbb{R}^n$,

$$D_\downarrow V(t, \overline{x}(t))(\alpha, \theta) \leq 0 \quad \text{for all } t \in (t_0, T). \tag{3.33}$$

Let $t \in (t_0, T)$ and $(\alpha, \theta) \in \mathbb{R} \times \mathbb{R}^n$ be fixed. By Filippov's Theorem (see, e.g., [2, Theorem 10.4.1]), there exists a constant k such that, for any $0 < \tau < 1$, the initial value problem

$$\begin{cases} \dot{x}(s) \in F(x(s)) \quad a.e. \ \ in \ [t + \alpha\tau, T], \\ x(t + \alpha\tau) = \overline{x}(t) + \tau\theta, \end{cases} \tag{3.34}$$

has a solution, $x_\tau(\cdot)$, that satisfies the inequality

$$\| x_\tau - \overline{x} \|_\infty \leq k\tau. \tag{3.35}$$

Hence, by the dynamic programming principle, (3.23), (2.9) and (3.35) we deduce that

$$\begin{aligned} D_\downarrow V(t, \overline{x}(t))(\alpha, \theta) &= \limsup_{\tau \to 0^+} \frac{V(t + \alpha\tau, x_\tau(t + \alpha\tau)) - V(t, \overline{x}(t))}{\tau} \\ &\leq \limsup_{\tau \to 0^+} \frac{\phi(x_\tau(T)) - \phi(\overline{x}(T))}{\tau} \leq 0. \end{aligned} \tag{3.36}$$

Then, the conclusion holds true also in case (ii). □

Remark 3.6.

- The above reasoning can be also applied to the end-point $t = T$ to prove that

$$\limsup_{t \nearrow T, \ y \to \overline{x}(T)} \frac{V(t, y) - V(T, \overline{x}(T)) + \langle \overline{p}(T), y - \overline{x}(T) \rangle}{|y - \overline{x}(T)| + |t - T|} \leq 0. \tag{3.37}$$

  Similarly, in the case of $t = t_0$, one can show that

$$\limsup_{t \searrow t_0, \ y \to \overline{x}(t_0)} \frac{V(t, y) - V(t_0, \overline{x}(t_0)) + \langle \overline{p}(t_0), y - \overline{x}(t_0) \rangle}{|y - \overline{x}(t_0)| + |t - t_0|} \leq 0. \tag{3.38}$$

- If, in addition, $\nabla_p H(\cdot, \cdot)$ is locally Lipschitz in $\mathbb{R}^n \times (\mathbb{R}^n \setminus \{0\})$, then one can show that the proximal full sensitivity relation

$$(H(\overline{x}(t), \overline{p}(t)), -\overline{p}(t)) \in \partial^{+,pr} V(t, \overline{x}(t)) \text{ for all } t \in (t_0, T) \tag{3.39}$$

  holds true when $-\overline{p}(T) \in \partial^{+,pr} \phi(\overline{x}(T))$. Note that the full sensitivity relation (3.39) implies the partial version (3.4) on (t_0, T). However, in Theorem 3.1 we have proved (3.4) without assuming the local Lipschitzianity of $\nabla_p H(\cdot, \cdot)$.

4 Necessary and sufficient conditions for optimality

The first result of this section can be seen as a strengthening of the maximum principle. Roughly speaking, we want to prove that *every* solution $\overline{p}$ of system (4.1) below, where q is any proximal supergradient of the final cost, provides a dual arc associated with the optimal trajectory $\overline{x}$, i.e. $(\overline{x}, \overline{p})$ satisfies the maximum principle (2.5). In the proof of a similar result for parameterized control systems (see, e.g., [7, Theorem in 7.3.1.]), a crucial role is played by the analysis of the response of the system to a variation of a fixed control. This approach is unavailable, in general, for differential inclusions. Here, the role of the variational equation will be played by the partial sensitivity relations of the previous section under a further assumption on the multifunction F.

Theorem 4.1. *Assume (SH), (H1), and suppose $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is locally strongly convex and $\phi : \mathbb{R}^n \to \mathbb{R}$ locally Lipschitz. Let $-\infty < t_0 < T$ and let $\overline{x} : [t_0, T] \to \mathbb{R}^n$ be an optimal solution of $\mathcal{P}(t_0, x_0)$ and let $q \in \partial^{+,pr}\phi(\overline{x}(T))$. Then,*

- *there exists a solution $\overline{p} : [t_0, T] \to \mathbb{R}^n$ of the differential inclusion*

$$\begin{cases} -\dot{p}(s) \in \partial_x^- H(\overline{x}(s), p(s)) & a.e. \text{ in } [t_0, T]\,, \\ -p(T) = q. \end{cases} \tag{4.1}$$

- *any solution $\overline{p}(\cdot)$ to (4.1) satisfies the maximum principle*

$$H(\overline{x}(t), \overline{p}(t)) = \langle \dot{\overline{x}}(t), \overline{p}(t) \rangle \quad a.e. \text{ in } [t_0, T]. \tag{4.2}$$

Moreover, if $q \neq 0$, then (4.2) holds true everywhere in $[t_0, T]$.

Proof. Let $\overline{x}(\cdot)$ be an optimal solution of $\mathcal{P}(t_0, x_0)$. Define the multifunction $G : [t_0, T] \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ by $G(s, p) = \partial_x^- H(\overline{x}(s), p)$. Observe that:

- for each $(t, p) \in [t_0, T] \times \mathbb{R}^n$, $G(t, p)$ is nonempty compact and convex set;
- by a known property of the generalized gradient, there exists a constant $k > 0$ such that, $\forall (s, p) \in [t_0, T] \times \mathbb{R}^n$ and $\forall v \in G(s, p)$, it holds that $|v| \leq k|p|$;
- G is upper semicontinuous.

In order to verify the last property, let us prove that G has a closed graph in $[t_0, T] \times \mathbb{R}^n \times \mathbb{R}^n$. The conclusion follows because a multifunction taking values in a compact set and having a closed graph is upper semicontinuous (see e.g. Corollary 1 p. 41 in [1]). The graph of G is

$$Graph(G) = \{((t, p), q)\,, (t, p) \in [t_0, T] \times \mathbb{R}^n : q \in \partial_x^- H(\overline{x}(t), p)\}.$$

Let $((t_i, p_i), q_i)$ be a sequence in $Graph(G)$ which converges to some $((t, p), q)$. Since $q_i \in \partial_x^- H(\overline{x}(t_i), p_i)$, there exists an open set A containing $\overline{x}([t_0, T])$ and a constant $c = c(A)$ such that

$$H(y, p_i) - H(\overline{x}(t_i), p_i) - \langle q_i, y - \overline{x}(t_i) \rangle \geq -c|p_i||y - \overline{x}(t_i)|^2 \tag{4.3}$$

for any point $y \in A$ and i large enough. Passing to the limit in (4.3), we see that $q \in \partial_x^- H(\overline{x}(t), p)$. This proves that the graph of G is closed.

From the above three properties we deduce the existence of at least one solution, $\overline{p}(\cdot)$, of (4.1) on $[t_0, T]$. Let us now show that every solution $\overline{p}(\cdot)$ to (4.1) satisfies (4.2). Hence, in the sequel, let q and $\overline{p}$ be fixed as above. If $q = 0$, then, thanks to Remark 2.10, we conclude that the solution of (4.1) vanishes on $[t_0, T]$. In this case, equality (4.2) is obvious. Consider, next, the case of $q \neq 0$. Then, Remark 2.10 ensures that $\overline{p}(\cdot)$ never vanishes on $[t_0, T]$. So, let us derive (4.2) in the stronger form

$$\dot{\overline{x}}(t) = \nabla_p H(\overline{x}(t), \overline{p}(t)) \tag{4.4}$$

for all $t \in [t_0, T]$. The proof is divided into two steps.

First step. In this step, we prove (4.4) on a suitable interval $(T-\overline{\tau},T]$.
Observe that, since $q\in\partial^{+,pr}\phi(\overline{x}(T))$, there exist constants $r_1>0$ and $c_1>0$ such that

$$\phi(y)-\phi(\overline{x}(T))+\langle\overline{p}(T),y-\overline{x}(T)\rangle\le c_1|y-\overline{x}(T)|^2, \tag{4.5}$$

whenever $y\in B(\overline{x}(T),r_1)$. Thus, let us consider the tubular compact neighborhood of $\overline{x}([t_0,T])$ defined by

$$D_{r_1}(\overline{x}([t_0,T]))=\{x\in\mathbb{R}^n:\ \exists t\in[t_0,T]\text{ such that }|x-\overline{x}(t)|\le r_1\}.$$

The following constants will be used in the rest of the proof:

- k is the Lipschitz constant of $\nabla_pH(\cdot,p)$, for $p\neq 0$, on $D_{r_1}(\overline{x}([t_0,T]))$,
- c' is such that $c'|p|$ is the semiconvexity constant of $H(\cdot,p)$ on $D_{r_1}(\overline{x}([t_0,T]))$,
- R is the radius of strong convexity of the multifunction F on $D_{r_1}(\overline{x}([t_0,T]))$.

We start with a brief observation. Consider the Cauchy problem: for all $\tau_1\in(0,T-t_0]$,

$$\begin{cases}\dot{x}(t)=\nabla_pH(x(t),\overline{p}(t))\ \text{ for all } t\ge T-\tau_1,\\ x(T-\tau_1)=\overline{x}(T-\tau_1).\end{cases} \tag{4.6}$$

If $\tau_1>0$ is such that

$$\tau_1e^{k\tau_1}=\frac{r_1}{\|\,\dot{\overline{x}}-\nabla_pH(\overline{x},\overline{p})\,\|_\infty+1},$$

then the solution x to (4.6) is well-defined on the interval $[T-\tau_1,T]$ and $|x(t)-\overline{x}(t)|\le r_1$ for all $t\in[T-\tau_1,T]$. Indeed, otherwise, there exists $\overline{t}$, $T-\tau_1<\overline{t}<T$, such that $|\overline{x}(\overline{t})-x(\overline{t})|=r_1$ and $|x(t)-\overline{x}(t)|<r_1$ for all $t\in[T-\tau_1,\overline{t})$. On the other hand, note that:

$$\begin{aligned}|\overline{x}(t)-x(t)|&\le\int_{T-\tau_1}^{t}|\dot{\overline{x}}(s)-\nabla_pH(x(s),\overline{p}(s))|\,ds\\
&\le\int_{T-\tau_1}^{t}\big(|\nabla_pH(\overline{x}(s),\overline{p}(s))-\nabla_pH(x(s),\overline{p}(s))|+|\dot{\overline{x}}(s)-\nabla_pH(\overline{x}(s),\overline{p}(s))|\big)ds\\
&\le k\int_{T-\tau_1}^{t}|\overline{x}(s)-x(s)|ds+\int_{T-\tau_1}^{t}|\dot{\overline{x}}(s)-\nabla_pH(\overline{x}(s),\overline{p}(s))|ds\\
&\le k\int_{T-\tau_1}^{t}|\overline{x}(s)-x(s)|ds+(t-T+\tau_1)\big(\|\,\dot{\overline{x}}-\nabla_pH(\overline{x},\overline{p})\,\|_\infty+1\big).\end{aligned} \tag{4.7}$$

By the Gronwall inequality, we have that:

$$|\overline{x}(\overline{t})-x(\overline{t})|\le(\overline{t}-T+\tau_1)e^{k(\overline{t}-T+\tau_1)}\big(\|\,\dot{\overline{x}}-\nabla_pH(\overline{x},\overline{p})\,\|_\infty+1\big). \tag{4.8}$$

Then, since $T-\tau_1<\overline{t}<T$,

$$|\overline{x}(\overline{t})-x(\overline{t})|<r_1. \tag{4.9}$$

This contradicts the choice of $\overline{t}$.

Now, set $\overline{\tau}:=\min\{\tau_1,\tau_2\}$, where τ_1 is as above and $\tau_2>0$ is such that

$$e^{k\tau_2}\tau_2^{\frac{1}{2}}\sqrt{\frac{2R(c_1+c'\,\|\,\overline{p}\,\|_\infty\,\tau_1)}{\min_{[t_0,T]}|\overline{p}|}}=\frac{1}{2}.$$

Let us prove the equality in (4.4) on $(T-\overline{\tau},T]$. Define the admissible trajectory $x_1:[t_0,T]\to\mathbb{R}^n$ in the following way:

- on the interval $[t_0, T-\overline{\tau})$, $x_1(\cdot)$ coincides with the optimal trajectory $\overline{x}(\cdot)$,
- on the interval $[T-\overline{\tau}, T]$, $x_1(\cdot)$ is the solution of the Cauchy problem:

$$\begin{cases} \dot{x}_1(t) = \nabla_p H(x_1(t), \overline{p}(t)) \ \text{ for all } t \in [T-\overline{\tau}, T], \\ x_1(T-\overline{\tau}) = \overline{x}(T-\overline{\tau}). \end{cases} \tag{4.10}$$

Recall that $D_{r_1}(\overline{x}([t_0, T]))$. We are going to give a first estimate for $\|\overline{x} - x_1\|_\infty$. Proceeding as in (4.7), we have that for all $t \in [T-\overline{\tau}, T]$,

$$|\overline{x}(t) - x_1(t)| \ \leq k \int_{T-\overline{\tau}}^{t} |\overline{x}(s) - x_1(s)| ds + \int_{T-\overline{\tau}}^{t} |\dot{\overline{x}}(s) - \nabla_p H(\overline{x}(s), \overline{p}(s))| ds. \tag{4.11}$$

By the Gronwall inequality, the above estimate yields that, for every $t \in [T-\overline{\tau}, T]$,

$$|\overline{x}(t) - x_1(t)| \leq e^{k\overline{\tau}} \int_{T-\overline{\tau}}^{T} |\dot{\overline{x}}(s) - \nabla_p H(\overline{x}(s), \overline{p}(s))| ds. \tag{4.12}$$

The next point is to bound the right-hand side of (4.12). Since $x_1(T) \in B(\overline{x}(T), r_1)$, by (4.5) we have that

$$\phi(x_1(T)) - \phi(\overline{x}(T)) + \langle \overline{p}(T), x_1(T) - \overline{x}(T) \rangle \leq c_1 |x_1(T) - \overline{x}(T)|^2. \tag{4.13}$$

Observe that, since $\phi(x_1(T)) - \phi(\overline{x}(T)) \geq 0$,

$$\langle \overline{p}(T), x_1(T) - \overline{x}(T) \rangle \leq c_1 |x_1(T) - \overline{x}(T)|^2. \tag{4.14}$$

Furthermore, since $x(T-\overline{\tau}) = \overline{x}(T-\overline{\tau})$, we have that

$$\begin{aligned} \langle \overline{p}(T), x_1(T) - \overline{x}(T) \rangle &= \int_{T-\overline{\tau}}^{T} \langle \dot{\overline{p}}(s), x_1(s) - \overline{x}(s) \rangle \ ds + \int_{T-\overline{\tau}}^{T} \langle \overline{p}(s), \dot{x}_1(s) - \dot{\overline{x}}(s) \rangle \ ds \\ &= (I) + (II). \end{aligned} \tag{4.15}$$

The semiconvexity of $H(\cdot, p)$ yields

$$(I) \geq \int_{T-\overline{\tau}}^{T} \big(H(\overline{x}(s), \overline{p}(s)) - H(x_1(s), \overline{p}(s)) - c' |\overline{p}(s)| |x_1(s) - \overline{x}(s)|^2 \big) \, ds. \tag{4.16}$$

As for (II), recalling (2.16) we have that

$$(II) = \int_{T-\overline{\tau}}^{T} \big(H(x_1(s), \overline{p}(s)) - H(\overline{x}(s), \overline{p}(s)) + \langle \overline{p}(s), \nabla_p H(\overline{x}(s), \overline{p}(s)) - \dot{\overline{x}}(s) \rangle \big) \, .$$

Now, since $F(x)$ is R-convex for all $x \in D_{r_1}(\overline{x}([t_0, T]))$, we can invoke Lemma 2.16 and $(H2)$ to bound the above scalar product. We thus obtain

$$(II) \geq \int_{T-\overline{\tau}}^{T} \bigg(H(x_1(s), \overline{p}(s)) - H(\overline{x}(s), \overline{p}(s)) + \frac{1}{2R} |\overline{p}(s)| |\nabla_p H(\overline{x}(s), \overline{p}(s)) - \dot{\overline{x}}(s)|^2 \bigg) \, ds. \tag{4.17}$$

Combining (4.15)-(4.17) we conclude that

$$\begin{aligned} &\langle \overline{p}(T), x_1(T) - \overline{x}(T) \rangle \\ &\qquad \geq \int_{T-\overline{\tau}}^{T} \Big(- c' |\overline{p}(s)| |x_1(s) - \overline{x}(s)|^2 + \frac{1}{2R} |\overline{p}(s)| |\nabla_p H(\overline{x}(s), \overline{p}(s)) - \dot{\overline{x}}(s)|^2 \Big) ds. \end{aligned} \tag{4.18}$$

Inequalities (4.14) and (4.18) imply that

$$\int_{T-\overline{\tau}}^{T} |\dot{\overline{x}}(s) - \nabla_p H(\overline{x}(s), \overline{p}(s))|^2 ds \leq \frac{2R}{\min_{[t_0,T]} |\overline{p}|}(c_1 + c' \parallel \overline{p} \parallel_\infty \overline{\tau}) \parallel x_1 - \overline{x} \parallel_\infty^2 . \qquad (4.19)$$

So, we can go back to (4.12) and estimate the integral on the right-hand side using (4.19) and the Hölder inequality. We obtain that:

$$\parallel \overline{x} - x_1 \parallel_\infty \leq e^{k\overline{\tau}} \overline{\tau}^{\frac{1}{2}} \sqrt{\frac{2R(c_1 + c' \parallel \overline{p} \parallel_\infty \overline{\tau})}{\min_{[t_0,T]} |\overline{p}|}} \parallel x_1 - \overline{x} \parallel_\infty . \qquad (4.20)$$

By the choice of $\overline{\tau}$, it follows that $\parallel \overline{x} - x_1 \parallel_\infty = 0$. So, by (4.19), we finally conclude that

$$\dot{\overline{x}}(t) = \nabla_p H(\overline{x}(t), \overline{p}(t)) \text{ a.e. in } [T - \overline{\tau}, T]. \qquad (4.21)$$

Since arc $\overline{p}(\cdot)$ never vanishes, the pair $(\overline{x}(\cdot), \overline{p}(\cdot))$ stays in the set where H is differentiable with respect to p. Thus, (4.21) is true for all $t \in (T - \overline{\tau}, T]$.

Second step. In this step we prove (4.4) in the whole interval $[t_0, T]$.
Set

$$\Gamma := \{\tau \in (0, T - t_0] : \text{ (4.4) holds true in } (T - \tau, T]\}.$$

In the first step we have constructed an element of such a set, proving that Γ is nonempty. So, it is enough to show that the supremum of Γ, τ_m, is equal to $T - t_0$. Arguing by contradiction, suppose $\tau_m < T - t_0$. By Remark 3.3, there exist constants $c_2, r_2 > 0$ such that

$$V(t, y) - V(t, \overline{x}(t)) + \langle \overline{p}(t), y - \overline{x}(t) \rangle \leq c_2 \mid y - \overline{x}(t) \mid^2, \qquad (4.22)$$

for all $t \in (T - \tau_m, T]$ and for all $y \in B(\overline{x}(t), r_2)$. Moreover, since r_2 and c_2 are uniform with respect to time, (4.22) still holds true when $t = T - \tau_m$. So,

$$-\overline{p}(T - \tau_m) \in \partial_x^{+,pr} V(T - \tau_m, \overline{x}(T - \tau_m)). \qquad (4.23)$$

Finally, consider the Mayer problem with final cost $V(T - \tau_m, \cdot)$ and time horizon $T - \tau_m$, that is,

$$\inf\{V(T - \tau_m, y(T - \tau_m)) : \dot{y} \in F(y) \; a.e. \text{ in } [t_0, T - \tau_m], \; y(t_0) = x_0\}.$$

Observe that $\overline{x}$, restricted to $[t_0, T - \tau_m]$, is an optimal trajectory of the above problem. Thus, recalling again that

$$\begin{cases} -\dot{\overline{p}}(s) \in \partial_x^- H(\overline{x}(s), \overline{p}(s)) \quad \text{ a.e. in } [t_0, T - \tau_m], \\ -\overline{p}(T - \tau_m) \in \partial_x^{+,pr} V(T - \tau_m, \overline{x}(T - \tau_m)), \end{cases} \qquad (4.24)$$

by our first step one can deduce the existence of a constant $\overline{\overline{\tau}} > 0$ such that the equality in (4.4) holds true everywhere in $(T - \tau_m - \overline{\overline{\tau}}, T - \tau_m]$. But this statement contradicts the fact that τ_m is the supremum of Γ. We conclude that $\tau_m = T - t_0$, and the equality in (4.4) holds true everywhere in $[t_0, T]$. This concludes the proof.

□

Remark 4.2. The above result and Theorem 3.5 together imply that, for any $q \in \partial^{+,pr}\phi(\overline{x}(T))$ and any solution $\overline{p}$ of (4.1), both the maximum principle and the full sensitivity relation (3.24) hold true. This is a stronger conclusion than the one of Theorem 2.8, which only affirms that the maximum principle (2.5) holds true for some $q \in \partial\phi(\overline{x}(T))$ and some solution $\overline{p}(\cdot)$. For this reason, using the proximal superdifferential of ϕ instead of the generalized gradient seems more appropriate whenever $(H1)$ is satisfied and F is locally strongly convex.

Now, we are ready to give a set of necessary and sufficient conditions for optimality.

Theorem 4.3. *Assume $(SH),(H1)$. Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be locally strongly convex and let $\phi : \mathbb{R}^n \to \mathbb{R}$ be locally semiconcave. A solution $\overline{x} : [t_0, T] \to \mathbb{R}^n$ of system* (1.11)-(1.12) *is optimal for $\mathcal{P}(t_0, x_0)$ if and only if, for every $q \in \partial^+\phi(\overline{x}(T))$, any solution $\overline{p} : [t_0, T] \to \mathbb{R}^n$ of the differential inclusion*

$$-\dot{p}(t) \in \partial_x^- H(\overline{x}(t), p(t)) \quad a.e.\ in\ [t_0, T] \tag{4.25}$$

with the transversality condition

$$-p(T) = q, \tag{4.26}$$

satisfies the full sensitivity relation

$$(H(\overline{x}(t), \overline{p}(t)), -\overline{p}(t)) \in \partial^+ V(t, \overline{x}(t)) \quad for\ all\ t \in (t_0, T), \tag{4.27}$$

and the maximum principle

$$H(\overline{x}(t), \overline{p}(t)) = \langle \overline{p}(t), \dot{\overline{x}}(t) \rangle \quad a.e.\ in\ [t_0, T]. \tag{4.28}$$

Proof. Sufficiency follows from Theorem 2.6. The fact that the existence of an arc $p(\cdot)$ satisfying (4.26) and (4.28) is a necessary condition for optimality is guaranteed by Theorem 4.1 and the semiconcavity of ϕ. Finally, the full sensitivity relation (4.27) is a consequence of Theorem 3.5. □

5 Relations between reachable gradients of the value function and optimal trajectories

In the calculus of variations, the existence of a one-to-one correspondence between the set of minimizers starting from a point (t, x) and the set of all reachable gradients of the value function V at (t, x) is a well-known fact. This allows, among other things, to identify the singular set of the value function—that is, the set of points at which V is not differentiable—as the set of points (t, x) for which the given funtional admits more than one minimizer. The aim of this last section is to investigate the above property for Mayer functionals with differential inclusions. Here, the difficulty consists in the nonsmoothness of the Hamiltonian at $p = 0$ that will force us to study separately the case of $0 \in \partial^* V(t, x)$. When the Hamiltonian and the terminalt cost are of class $C^{1,1}_{loc}(\mathbb{R}^n \times (\mathbb{R}^n \setminus \{0\}))$ and $C^1(\mathbb{R}^n)$, respectively, there is an injective map from $\partial^* V(t, x) \setminus \{0\}$ to the set of all optimal trajectories starting from (t, x) (see, for instance, [7, Theorem 7.3.10] where the authors study parameterized optimal control problems with smooth data). However, we assume below neither the existence of a smooth parameterization, nor such a regularity of the Hamiltonian. Our assumptions for the terminal cost function are also milder than in [7]. It is precisely the lack of regularity of H that represents the main difficulty, since it does not guarantee the uniqueness of solutions to system (5.1) below. We shall prove that, in our case, there exists an injective set-valued map from $\partial^* V(t, x) \setminus \{0\}$ into the set of optimal trajectories starting from (t, x).

Lemma 5.1. *Assume* (SH) *and* $(H1)$*, and suppose* ϕ *is locally Lipschitz and such that* $\partial^+\phi(z) \neq \emptyset$ *for all* $z \in \mathbb{R}^n$*. Then, given a point* $(t,x) \in (-\infty, T) \times \mathbb{R}^n$ *and a vector* $\overline{p} = (\overline{p}_t, \overline{p}_x) \in \partial^* V(t,x) \setminus \{0\}$*, there exists at least one pair* $(y(\cdot), p(\cdot))$ *that satisfies the system*

$$\begin{cases} \dot{y}(s) &= \nabla_p H(y(s), p(s)) \quad \text{for all } s \in [t,T], \\ -\dot{p}(s) &\in \partial_x^- H(y(s), p(s)) \quad \text{a.e. in } [t,T], \end{cases} \tag{5.1}$$

together with the initial conditions

$$\begin{cases} y(t) = x, \\ p(t) = -\overline{p}_x, \end{cases} \tag{5.2}$$

such that $y(\cdot)$ *is optimal for* $\mathcal{P}(t,x)$*.*

Proof. Observe, first, that if $(\overline{p}_t, \overline{p}_x) \in \partial^* V(t,x) \setminus \{0\}$, then $\overline{p}_x \neq 0$. Indeed, V satisfies $-V_t + H(x, -V_x(t,x)) = 0$ at every point of differentiability. Therefore, taking limits, we have that $-p_t + H(x, -p_x) = 0$ for every $(\overline{p}_t, \overline{p}_x) \in \partial^* V(t,x)$. Since $H(x,0) = 0$, we conclude that if $(p_t, p_x) \neq 0$, then $p_x \neq 0$.
Since $(\overline{p}_t, \overline{p}_x) \in \partial^* V(t,x)$, we can find a sequence $\{(t_k, x_k)\}$ such that V is differentiable at (t_k, x_k) and

$$\lim_{k \to \infty} (t_k, x_k) = (t,x), \quad \lim_{k \to \infty} \nabla_x V(t_k, x_k) = \overline{p}_x.$$

Let $y_k(\cdot)$ be an optimal trajectory for $\mathcal{P}(t_k, x_k)$. Since $\overline{p}_x \neq 0$, there exists $\overline{k} > 0$ such that $\nabla_x V(t_k, x_k) \neq 0$ for all $k > \overline{k}$. By Remark 2.9, there exists an arc p_k such that (y_k, p_k) satisfies

$$\begin{cases} \dot{y}_k(s) \in \partial_p^- H(y_k(s), p_k(s)) \quad \text{a.e. in } [t_k, T], & y_k(t_k) = x_k, \\ -\dot{p}_k(s) \in \partial_x^- H(y_k(s), p_k(s)) \quad \text{a.e. in } [t_k, T], & -p_k(T) \in \partial^+\phi(y_k(T)). \end{cases} \tag{5.3}$$

Moreover, by Theorem 3.1, we have that $-p_k(t_k) \in \partial_x^+ V(t_k, x_k)$. Since V is differentiable at (t_k, x_k), it follows that $-p_k(t_k) = \nabla_x V(t_k, x_k)$. Thus, for $k > \overline{k}$ we have that $p_k(s) \neq 0$ for all $s \in [t_k, T]$ (see Remark 2.10), and the first inclusion in (5.3) becomes

$$\dot{y}_k(s) = \nabla_p H(y_k(s), p_k(s)) \quad \text{for all } s \in [t_k, T]. \tag{5.4}$$

Now, we extend y_k and p_k on $[t-1, t_k)$ by setting $y(s) = y(t_k)$, $p(s) = p(t_k)$ for all $s \in [t-1, t_k)$. The last point of the reasoning consists of proving that the sequence $(y_k(\cdot), p_k(\cdot))$, after possibly passing to a subsequence, converges to a pair $(y(\cdot), p(\cdot))$ that verifies the conclusions of the lemma. It is easy to prove that the sequences of functions $\{p_k\}_k$ and $\{y_k\}_k$ are uniformly bounded and uniformly Lipschitz continuous in $[t-1, T]$, using Gronwall's inequality together with estimates (2.15) and $(SH)(iii)$, respectively. Hence their derivatives are essentially bounded. Therefore, after possibly passing to subsequences, we can assume that the sequence $(y_k(\cdot), p_k(\cdot))$ converges uniformly in $[t-1, T]$ to some pair of Lipschitz functions $(y(\cdot), p(\cdot))$. Moreover, $\dot{p}_k(\cdot)$ converges weakly to $\dot{p}(\cdot)$ in $L^1([t-1, T]; \mathbb{R}^n)$. Now, observe that

$$((y_k(s), p_k(s)), -\dot{p}_k(s)) \in Graph(M) \quad \text{a.e. in } [t_k, T],$$

where M is the multifunction defined by $M(x,p) := \partial_x^- H(x,p)$. The multifunction M is upper semicontinuous; this can be easily derived as it was done for G in the proof of Theorem 4.1. Hence, from Theorem 7.2.2. in [2] it follows that $-\dot{p}(s) \in \partial_x^- H(y(s), p(s))$ for a.e. $s \in [t,T]$. Moreover, we have that

$$p(t) = \lim_{k \to \infty} p_k(t_k) = -\lim_{k \to \infty} \nabla_x V(t_k, x_k) = -\overline{p}_x \neq 0.$$

So, we appeal again to Remark 2.10 to deduce that $p(\cdot)$ never vanishes on $[t,T]$. Since the map $(x,p) \mapsto \nabla_p H(x,p)$ is continuous for $p \neq 0$, we can pass to the limit in (5.4) and deduce that $\dot{y}(s) = \nabla_p H(y(s), p(s))$ for all $s \in [t,T]$. In conclusion, $(y(\cdot), p(\cdot))$ is a solution of the system in (5.1) with initial conditions $y(t) = x,\ p(t) = -\overline{p}_x$. This implies of course that $\dot{y}(s) \in F(y(s))$ for all $s \in [t,T]$. Moreover, since V is continuous and $y_k(\cdot)$ is optimal, we have

$$\phi(y(T)) = \lim_{k\to\infty} \phi(y_k(T)) = \lim_{k\to\infty} V(t_k, y_k(t_k)) = V(t, y(t)),$$

which means that $y(\cdot)$ is optimal for $\mathcal{P}(t,x)$. $\square$

The *nondegeneracy condition* $\partial^+\phi(z) \neq \emptyset$ for all $z \in \mathbb{R}^n$ that we have assumed above is verified, for instance, when ϕ is differentiable or semiconcave.

Remark 5.2. From the above proof it follows that, if the cost is supposed to be of class $C^1(\mathbb{R}^n)$, then the pair $(y(\cdot), p(\cdot))$ in Lemma 5.1 satisfies $\nabla\phi(y(T)) = -p(T)$. Moreover, p is a dual arc associated to y.

Now, fix $(t,x) \in (-\infty, T) \times \mathbb{R}^n$. For any $\overline{p} = (\overline{p}_t, \overline{p}_x) \in \partial^* V(t,x) \setminus \{0\}$, we denote by $\mathcal{R}(\overline{p})$ the set of all trajectories $y(\cdot)$ that are solutions of (5.1)-(5.2), and optimal for $\mathcal{P}(t,x)$. The above lemma guarantees that the set-valued map $\mathcal{R}$ that associates with any $p \in \partial^* V(t,x) \setminus \{0\}$ the set $\mathcal{R}(p)$ has nonempty values. Now let us prove that $\mathcal{R}$ is strongly injective. We will use the "difference set":

$$\partial_x^- H(x,p) - \partial_x^- H(x,p) := \{a - b :\ a, b \in \partial_x^- H(x,p)\}.$$

Theorem 5.3. *Assume (SH), (H1), and suppose that ϕ is of class $C^1(\mathbb{R}^n)$ and*

(H3) for each $z \in \mathbb{R}^n$, $F(z)$ is not a singleton and has a C^1 boundary (for $n > 1$),

(H4) $\mathbb{R}^+ p \cap (\partial_x^- H(z,p) - \partial_x^- H(z,p)) = \emptyset \quad \forall p \neq 0$ for all $z \in \mathbb{R}^n$.

Then for any $\overline{p}_1, \overline{p}_2 \in \partial^* V(t,x) \setminus \{0\}$ with $\overline{p}_1 \neq \overline{p}_2$, we have that $\mathcal{R}(\overline{p}_1) \cap \mathcal{R}(\overline{p}_2) = \emptyset$.

Proof. Suppose to have two elements $\overline{p}_i = (\overline{p}_{i,t}, \overline{p}_{i,x}) \in \partial^* V(t,x) \setminus \{0\},\ i = 1,2$, with $\overline{p}_1 \neq \overline{p}_2$. Note that the Hamilton-Jacobi equation implies that $\overline{p}_{1,x} \neq \overline{p}_{2,x}$ if and only $\overline{p}_1 \neq \overline{p}_2$. Furthermore, suppose that there exist two pairs $(y(\cdot), p_i(\cdot)),\ i = 1,2$ that are solutions of system (5.1) with $p_i(t) = \overline{p}_{i,x}$ and $y(\cdot)$ is optimal for $\mathcal{P}(t,x)$. Then, we get

$$\dot{y}(s) = \nabla_p H(y(s), p_1(s)) = \nabla_p H(y(s), p_2(s)) \quad \text{for all } s \in [t,T].$$

This implies that

$$p_i(s) \in N_{F(y(s))}(\dot{y}(s)),\ i = 1,2 \ \text{ for all } s \in [t,T].$$

By $(H3)$, the normal cone $N_{F(y(s))}(\dot{y}(s))$ is a half-line. Recalling also that $p_i\,(i = 1,2)$ never vanishes, it follows that there exists $\lambda(s) > 0$ such that $p_2(s) = \lambda(s) p_1(s)$, for every $s \in [t,T]$. The function $\lambda(\cdot)$ is differentiable a.e. on $[t,T]$ because

$$\lambda(s) = \frac{|\ p_2(s)\ |}{|\ p_1(s)\ |}.$$

By (5.1), since $\beta \partial_x^- H(x,p) = \partial_x^- H(x, \beta p)$ for each $\beta > 0$, it follows that

$$-\dot{p}_2(s) = -\lambda(s)\dot{p}_1(s) - \dot{\lambda}(s) p_1(s) \in \lambda(s) \partial_x^- H(y(s), p_1(s)) \quad \text{for a.e. } s \in [t,T].$$

Dividing by $\lambda(s)$,

$$-\frac{\dot{\lambda}(s)}{\lambda(s)}p_1(s) \in \partial_x^- H(y(s), p_1(s)) - \partial_x^- H(y(s), p_1(s)) \quad \text{for a.e. } s \in [t, T]. \tag{5.5}$$

Since the "difference set" in $(H4)$ is symmetric, $(H4)$ is equivalent to the condition

$$(\mathbb{R} \setminus \{0\})\, p \cap \big(\partial_x^- H(x,p) - \partial_x^- H(x,p)\big) = \emptyset \quad \forall p \neq 0. \tag{5.6}$$

From (5.5) and (5.6) we obtain that $\dot{\lambda}(s) = 0$ a.e. in $[t, T]$. This gives that λ is constant. Moreover, since $\phi \in C^1(\mathbb{R}^n)$, we have that $\nabla\phi(y(T)) = -p_i(T)$, $i = 1, 2$, which implies that $\lambda = 1$. But this yields $\overline{p}_{1,x} = \overline{p}_{2,x}$, which contradicts the inequality $\overline{p}_1 \neq \overline{p}_2$. Hence, we can assert that the functions $y_i(\cdot)$, $i = 1, 2$, are different. □

Remark 5.4. Assumption $(H4)$ is verified, for instance, when $x \mapsto H(x, p)$ is differentiable, without any Lipschitz regularity of the map $x \mapsto \nabla_p H(x, p)$.

Example 5.5. The above theorem is false when $\phi \notin C^1(\mathbb{R}^n)$. Indeed, without such an assumption there could exist infinitely many reachable gradients at a point at which the optimal trajectory is unique. Let us consider the state equation in the one-dimensional space given by $\dot{x} = u \in [-1, 1]$. Let the cost function be defined by

$$\phi(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) + 3x & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

This function is differentiable, with derivative

$$\phi'(x) = \begin{cases} -\cos(\frac{1}{x}) + 2x\sin(\frac{1}{x}) + 3 & \text{if } x \neq 0, \\ 3 & \text{if } x = 0 \end{cases}$$

which is discontinuous at zero because $\cos(\frac{1}{x})$ oscillates as $x \to 0$. Therefore, this function is differentiable but not of class $C^1(\mathbb{R})$. Moreover, $\partial^*\phi(0) = [2, 4]$.
Note that ϕ is strictly increasing. Thus, for any $(t, x) \in [0, T] \times \mathbb{R}$, there exists a unique optimal control that is the constant one $u \equiv -1$, and so the unique optimal trajectory is $x(s) = x - s + t$. The value function is $V(t, x) = \phi(x - T + t)$, and so it has the same regularity as the final cost ϕ. Summarizing, $\partial^* V(t, x) = \{(a, a) : \ a \in [2, 4]\}$ at any point (t, x) such that $x = T - t$, but there exists a unique optimal trajectory starting from such points.

Now let us consider the case when $\overline{p} = 0 \in \partial^* V(t, x)$.

Theorem 5.6. *Assume (SH), $(H1)$ and suppose that $\phi \in C^1(\mathbb{R}^n)$. Let $(t, x) \in (-\infty, T) \times \mathbb{R}^n$ be such that $0 \in \partial^* V(t, x)$. Then there exists an optimal trajectory $y : [t, T] \to \mathbb{R}^n$ for $\mathcal{P}(t, x)$ such that $\nabla\phi(y(T)) = 0$. Consequently, the corresponding dual arc is equal to zero.*

Proof. Since $0 \in \partial^* V(t, x)$, we can find a sequence $\{(t_k, x_k)\}$ such that V is differentiable at (t_k, x_k) and

$$\lim_{k\to\infty} (t_k, x_k) = (t, x), \ \lim_{k\to\infty} \nabla V(t_k, x_k) = 0.$$

Let $y_k(\cdot)$ be an optimal trajectory for $\mathcal{P}(t_k, x_k)$ and $p_k(\cdot)$ be a dual arc. We extend y_k and p_k on $[t - 1, T]$ as in the proof of Lemma 5.1. By Theorem 7.2.2 in [2], we can assume, after possibly passing to a subsequence, that $y_k(\cdot)$ converges uniformly to $y(\cdot)$ which is a trajectory of our system on $[t,T]$. Since

$$\phi(y(T)) = \lim_{k\to\infty} \phi(y_k(T)) = \lim_{k\to\infty} V(t_k, x_k) = V(t,x),$$

it follows that y is optimal for $\mathcal{P}(t,x)$. Furthermore, the sensitivity relation in Remark 3.4 holds true and so, recalling that V is differentiable at (t_k, x_k), we have

$$-\,p_k(t_k) = \nabla_x V(t_k, x_k) \to 0. \tag{5.7}$$

By (5.7), (2.15) and the Gronwall's inequality, we get that $p_k(T) \to 0$ when $k\to\infty$. We conclude that

$$\nabla\phi(y(T)) = \lim_{k\to\infty} \nabla\phi(y_k(T)) = \lim_{k\to\infty} -p_k(T) = 0.$$

□

As applications of the above results we obtain the following corollaries.

Corollary 5.7. *Under the same assumptions of Theorem 5.3, for every $(t,x)\in(-\infty,T]\times\mathbb{R}^n$ there exist at least as many optimal solutions of $\mathcal{P}(t,x)$ as elements of $\partial^* V(t,x)$.*

Proof. Fix $(t,x)\in(-\infty,T]\times\mathbb{R}^n$. By Lemma 5.1 and Remark 5.2, to every $p=(p_t,p_x)\in\partial^*V(t,x)\smallsetminus\{0\}$ corresponds a pair of arcs (y,p) satisfying (5.1) and such that $-p(t)=p_x$, y is optimal for $\mathcal{P}(t,x)$ and $-p(T)=\nabla\phi(y(T))$. Moreover, since the arc p never vanishes, we have that $\nabla\phi(y(T))\neq 0$. Theorem 5.3 implies that optimal solutions corresponding to distinct elements of $\partial^*V(t,x)\smallsetminus\{0\}$ are distinct. On the other hand, if $0\in\partial^*V(t,x)$, then Thereom 5.6 yields the existence of an optimal trajectory y for $\mathcal{P}(t,x)$ such that $0=\nabla\phi(y(T))$. These facts give our claim. □

Corollary 5.8. *Assume (SH), $(H1)$, $\phi\in C^1(\mathbb{R}^n)$ and suppose also that $\nabla\phi(x)\neq 0$ for all $x\in\mathbb{R}^n$. Then $0\notin\partial^*V(t,x)$ for all $(t,x)\in(-\infty,T)\times\mathbb{R}^n$.*

Corollary 5.9. *Assume (SH), $(H1)$, $(H3)$, $(H4)$ and let $\phi\in C^1(\mathbb{R}^n)\cap SC(\mathbb{R}^n)$. If V fails to be differentiable at a point $(t,x)\in(-\infty,T)\times\mathbb{R}^n$, then there exist two or more optimal trajectories starting from (t,x).*

Proof. Since V is semiconcave (see [8]), if it is not differentiable at a point $(t,x)\in(-\infty,T)\times\mathbb{R}^n$, then we can find two distinct elements $p_1,p_2\in\partial^*V(t,x)$. If p_1,p_2 are both nonzero, we can apply Theorem 5.3 to find two distinct optimal trajectories. If one of the two vectors is zero, for instance p_1, then there exists at least an associated optimal trajectory $y_1(\cdot)$ such that $\nabla\phi(y_1(T))=0$ by Theorem 5.6, but for any optimal trajectory $y_2(\cdot)$ associated to p_2 it holds that $\nabla\phi(y_2(T))\neq 0$ by Lemma 5.1 and Remark 2.10. □

It might happen that two or more optimal trajectories actually start from a point (t,x) at which V is differentiable. However, if $H\in C^{1,1}_{loc}(\mathbb{R}^n\times(\mathbb{R}^n\smallsetminus\{0\}))$, then it is well-know that such a behaviour can only occur when the gradient of V at (t,x) vanishes (see e.g. Theorem 7.3.14 and Example 7.2.10(iii) in [7]). More can be said in one space dimension as we explain below.

Example 5.10. When $n=1$ it is easy to show that, if V is differentiable at some point (t_0,x_0) with $V_x(t_0,x_0)\neq 0$, then there exists a unique optimal trajectory starting from (t_0,x_0). Indeed, in this case, $F(x)=[f(x),g(x)]$ for suitable functions $f,g:\mathbb{R}\to\mathbb{R}$, with $f\le g$, such that $-f$ and g are locally semiconvex. So,

$$H(x,p)=\begin{cases} f(x)p, & p<0,\\ g(x)p, & p\ge 0.\end{cases} \tag{5.8}$$

If $\overline{x}(\cdot)$ is an optimal trajectory at (t_0, x_0) and $\overline{p}(\cdot)$ is a dual arc associated with $\overline{x}(\cdot)$, then by Remark 3.4 we have that $0 \neq V_x(t_0, x_0) = -\overline{p}(t_0)$. Therefore, $0 \neq \overline{p}(t)$ for all $t \in [t_0, T]$ by Remark 2.10. Thus, (5.8) and the maximum principle yield

$$\dot{\overline{x}}(t) = \begin{cases} f(\overline{x}(t)), & \text{if } V_x(t_0, x_0) > 0, \\ g(\overline{x}(t)), & \text{if } V_x(t_0, x_0) < 0. \end{cases} \tag{5.9}$$

Since f and g are both locally Lipschitz, $\overline{x}(\cdot)$ is the unique solution of (5.9) satisfying $x(t_0) = x_0$.

Acknowledgements

Partial support of this research by the European Commission (FP7-PEOPLE-2010-ITN, Grant Agreement no. 264735-SADCO), and by the INdAM National Group GNAMPA is gratefully acknowledged.

References

[1] Jean-Pierre Aubin and Arrigo Cellina. *Differential inclusions*. Vol. 264. Gründlehren der Mathematischen Wissenschaften. Berlin: Springer-Verlag, 1984.

[2] Jean-Pierre Aubin and Hélène Frankowska. *Set-valued analysis*. Vol. 2. Systems & Control: Foundations & Applications. Birkhäuser Boston Inc., 1990.

[3] Piernicola Bettiol, Hélène Frankowska, and Richard Vinter. “Improved sensitivity relations in state constrained optimal control”. *Applied Mathematics and Optimization* (2014).

[4] Piermarco Cannarsa and Halina Frankowska. “Some characterizations of optimal trajectories in control theory”. *SIAM J. Control Optim.* 29 (1991), pp. 1322–1347.

[5] Piermarco Cannarsa, Hélène Frankowska, and Carlo Sinestrari. “Optimality conditions and synthesis for the minimum time problem”. *Set-Valued Anal.* 8.1-2 (2000), pp. 127–148.

[6] Piermarco Cannarsa, Francesco Marino, and Peter Wolenski. “The dual arc inclusion with differential inclusions”. *Nonlinear Anal.* 79 (2013), pp. 176–189.

[7] Piermarco Cannarsa and Carlo Sinestrari. *Semiconcave functions, Hamilton-Jacobi equations, and optimal control*. Birkhäuser Boston Inc., 2004.

[8] Piermarco Cannarsa and Peter R. Wolenski. “Semiconcavity of the value function for a class of differential inclusions”. *Discrete Contin. Dyn. Syst.* 29.2 (2011), pp. 453–466.

[9] Frank H. Clarke. *Optimization and nonsmooth analysis*. New York: John Wiley & Sons Inc., 1983.

[10] Frank H. Clarke and Richard B. Vinter. “The relationship between the maximum principle and dynamic programming”. *SIAM J. Control Optim.* 25 (1987), pp. 1291–1311.

[11] Halina Frankowska and Czesław Olech. “R-convexity of the integral of set-valued functions”. Johns Hopkins Univ. Press, 1981, pp. 117–129.

[12] Hélène Frankowska and Marco Mazzola. “On relations of the adjoint state to the value function for optimal control problems with state constraints”. *Nonlinear Differential Equations and Applications* 20 (2013), pp. 361–383.

[13] Anton E. Mayer. “Eine Überkonvexität”. *Math. Z.* 39.1 (1935), pp. 511–531.

[14] L. Pasqualini. “Superconvexité”. *Bull. de Cl. (5) XXV* (1939), pp. 18–24.

[15] Nina N. Subbotina. “The maximum principle and the superdifferential of the value function”. *Problems Control Inform. Theory/Problemy Upravlen. Teor. Inform.* 18.3 (1989), pp. 151–160.

[16] Paul Vincensini. “Sur les figures superconvexes planes”. *Bull. Soc. Math. France* 64 (1936), pp. 197–208.

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