REGULARITY ALONG OPTIMAL TRAJECTORIES OF THE VALUE FUNCTION OF A MAYER PROBLEM

Carlo Sinestrari[1]

Abstract. We consider an optimal control problem of Mayer type and prove that, under suitable conditions on the system, the value function is differentiable along optimal trajectories, except possibly at the endpoints. We provide counterexamples to show that this property may fail to hold if some of our conditions are violated. We then apply our regularity result to derive optimality conditions for the trajectories of the system.

1. Introduction

Let us consider the optimal control problem of Mayer type:

$$(\mathbf{M}) \qquad \text{minimize } g(y(T))$$

over all trajectories $y(\cdot)$ of the control system

$$y'(t) = f(y(t), u(t)), \qquad t \in [t_0, T] \tag{1.1}$$

satisfying the initial condition

$$y(t_0) = x_0. \tag{1.2}$$

Here the control $u(\cdot)$ takes values in some given set $U \subset \mathbb{R}^m$. We want to analyze some relations between the structure of the minimizers of problem $(\mathbf{M})$ and the differentiability properties of the associated value function, defined as

$$V(t_0, x_0) = \inf\{g(y(T)) \ : \ y \text{ is a solution of (1.1)}, \ y(t_0) = x_0\},$$

for all $(t_0, x_0) \in [0, T] \times \mathbb{R}^n$. It is well known that V is not differentiable in general, even if the data of the problem are smooth, and that it is a viscosity solution of the Hamilton-Jacobi-Bellman equation associated with $(\mathbf{M})$, which has the form

$$\begin{cases} -\partial_t V(t,x) + H(x, \nabla V(t,x)) = 0, & (t,x) \in]0, T[\times \mathbb{R}^n \\ V(T,x) = g(x) & x \in \mathbb{R}^n, \end{cases} \tag{1.3}$$

where the Hamiltonian H is defined as

$$H(x,p) = \sup_{u \in U} -f(x,u) \cdot p.$$

Observe that, up to a sign, $H(x, \cdot)$ is the support function of the convex hull of the set

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

Therefore, its behaviour depends on the properties of $f(x,U)$. However, for every choice of f, U, the function H is homogeneous of degree one with respect to p and is not differentiable at $p = 0$ (if we exclude the trivial case where $f(x,U)$ is a singleton). Because of these properties the analysis of problem **(M)** presents some additional difficulties, compared to the cases (*e.g.* calculus of variations [8, 11]) where the Hamiltonian is smooth and strictly convex in p.

The main aim of this paper is to find conditions on the system ensuring that V is differentiable along optimal trajectories, except possibly at the endpoints; that is, if $y : [t_0, T] \to \mathbb{R}^n$ achieves the minimum in problem **(M)**, then V is differentiable at all points of the form $(s, y(s))$, for $s \in]t_0, T[$. Roughly speaking, this shows that the set where V is differentiable is large enough to contain the interior points of all optimal trajectories. Such a property is well known for problems in calculus of variations (see *e.g.* [11]); it has also been proved for some classes of optimal control problems with exit time [4, 6]. The argument used in these papers, however, relies on the strict convexity of the Hamiltonian and so it cannot be directly extended to the Mayer problem considered here. Moreover, one can give examples where the value function of **(M)** is not differentiable along optimal trajectories (see Sect. 3). Therefore, suitable restrictions have to be imposed on the system if one wants to prove differentiability. Until now, the only result in this direction has been obtained in [9]; under the assumption that $f(x,U)$ is convex with a smooth boundary, it is proved that the superdifferential of V at the interior points of an optimal trajectory has dimension at most one. Such a result does not imply differentiability, but shows that the value function cannot have an arbitrary singularity at these points.

In this paper we provide a set of conditions which ensures the differentiability of V along an optimal trajectory $y(\cdot)$. Our hypotheses are stated in Theorem 3.5 and can be summarized as follows. We require the final cost g to be semiconcave (see Sect. 2 for the definition) and to be differentiable at $y(T)$. Moreover, we assume that the set $f(x,U)$ is convex with a smooth boundary and that H is differentiable for $p \neq 0$. Finally, we require a nondegeneracy condition at the initial point of the trajectory, namely that 0 does not belong to the set of reachable gradients $D^*V(t_0, x_0)$. This set is defined as the collection of all p such that $p = \lim_{k \to \infty} DV(t_k, x_k)$, where (t_k, x_k) are points of differentiability for V and $(t_k, x_k) \to (t_0, x_0)$. We also prove that, in the case where $f = f(u)$, our regularity result holds even without the smoothness assumption on the boundary of $f(U)$ (see Th. 3.8).

We then give counterexamples showing that some of the above requirements cannot be removed, *e.g.* the differentiability of g at $y(T)$ or the nondegeneracy condition at the initial point. Let us remark that the smoothness of $f(x,U)$ is related to the properties of the Hamiltonian; it ensures, in fact, that $H(x,p)$ is strictly convex with respect to p along any nonradial direction. Roughly speaking, the strict convexity of H provides strong information on the structure of the generalized differentials of V along an optimal trajectory (see Lem. 3.3); using this, we can show that, if V is not differentiable at some point of the form $(s, y(s))$ with $t_0 < s < T$, then g is not differentiable at $y(T)$, in contradiction with our assumptions.

In Section 4 we apply our regularity result to the analysis of optimality conditions for problem **(M)**. It is known that to every nonzero vector $p \in D^*V(t_0, x_0)$ corresponds an optimal trajectory for **(M)** obtained by solving the associated Hamiltonian system (see [3]). Here we prove that, under the assumptions of Section 3, such a correspondence is one-to-one provided $0 \notin D^*V(t_0, x_0)$. Thus, we obtain a complete characterization of the optimal trajectories starting from such points. This generalizes to the Mayer problem the results proved in [8] for problems in calculus of variations and in [4, 6] for optimal control problems with exit time.

2. Preliminaries

Let a compact set $U \subset \mathbb{R}^m$ and a continuous function $f : \mathbb{R}^n \times U \to \mathbb{R}^n$ be given. For a fixed $T > 0$, we consider the control system

$$\begin{cases} y'(t) = f(y(t), u(t)), & t \in [t_0, T] \text{ a.e.} \\ y(t_0) = x_0, \end{cases} \tag{2.1}$$

where $t_0 \in [0, T]$, $x_0 \in \mathbb{R}^n$ and $u \in L^\infty([t_0, T], U)$. The assumptions made on f in the sequel ensure the existence and uniqueness of a solution to (2.1) in the whole interval $[t_0, T]$; we denote this solution by $y(\cdot; t_0, x_0, u)$. We then assume that a function $g \in C(\mathbb{R}^n)$, called *final cost*, is given, and we consider, for any $(t_0, x_0) \in [0, T] \times \mathbb{R}^n$, the problem

(M) minimize $g(y(T; t_0, x_0, u))$ over all controls $u : [t_0, T] \to U$.

A control $u : [t_0, T] \to U$ such that the infimum in **(M)** is attained is called *optimal* for problem **(M)** with initial point (t_0, x_0), and the corresponding solution $y(\cdot) = y(\cdot; t_0, x_0, u)$ of (2.1) is called an *optimal trajectory*.

Throughout the paper we make the following assumptions on f, g:

(H1) (i) There exists $K > 0$ such that $|f(x_2, u) - f(x_1, u)| \le K|x_2 - x_1|$, for all $x_1, x_2 \in \mathbb{R}^n$, $u \in U$;
   (ii) f_x exists and is continuous; in addition, for any $R > 0$ there exists $K(R) > 0$ such that $\|f_x(x_2, u) - f_x(x_1, u)\| \le K(R)|x_2 - x_1|$, for all $x_1, x_2 \in B(0, R)$, $u \in U$;
   (iii) $g \in \text{Lip}_{\text{loc}}(\mathbb{R}^n)$.

Let us remark that all our results can be extended to the case where the function f in (2.1) depends also on t; we restrict ourselves to the autonomous case to simplify the exposition.

Definition 2.1. For any $(t, x) \in [0, T] \times \mathbb{R}^n$, we define

$$V(t, x) = \inf\{g(y(T; t, x, u)) \ : \ u : [t, T] \to U \text{ measurable }\}\cdot$$

The function V is called the *value function* of the control problem **(M)**.

We now collect some basic results about problem **(M)** and about nonsmooth functions which are needed in our analysis. For more details about these topics the reader may consult, in addition to the references quoted in the following, the forthcoming monograph [7]. We begin by recalling the definition of semiconcavity.

Definition 2.2. A function $v : A \to \mathbb{R}$, with $A \subset \mathbb{R}^n$ convex, is called *semiconcave* if there exists a nondecreasing function $\omega : [0, \infty[\to [0, \infty[$ such that $\omega(\rho) \to 0$ as $\rho \to 0$ and

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

for all $x, y \in A$ and $\lambda \in [0, 1]$. We denote by $SC_{loc}(A)$ the class of the functions which are semiconcave on all convex compact subsets of A.

There is a more restrictive notion of semiconcavity, often used in the literature (*e.g.* [2, 13]), where the function ω in the above definition has the form $\omega(h) = Ch$ for some $C \ge 0$. The definition we have chosen here has the consequence that any function of class C^1 is locally semiconcave. In fact, it is easily checked that, if $v \in C^1(A)$, then v is semiconcave in any compact convex set $S \subset A$, taking $\omega(\cdot)$ equal to the modulus of continuity of Dv in S.

We recall that, if $v : \Omega \to \mathbb{R}$ with $\Omega \subset \mathbb{R}^n$, the *superdifferential* of v at a point $x \in \Omega$ is the set

$$D^+v(x) = \left\{ p \in \mathbb{R}^n : \limsup_{y \to x} \frac{v(y) - v(x) - p \cdot (y - x)}{|y - x|} \le 0 \right\}\cdot$$

We observe that, if v is a locally Lipschitz function, then $D^+v(x)$ is a compact convex set (possibly empty) at every point. If v is differentiable at x, then $D^+v(x)$ coincides with the singleton $\{Dv(x)\}$, where Dv denotes

the usual gradient of v. For a locally Lipschitz function $v : \Omega \to \mathbb{R}$, we also define the set of *reachable gradients*

$$D^*v(x) = \{\, p \;:\; \exists\{x_k\} \subset \Omega \text{ such that } x_k \to x, v \text{ is differentiable at } x_k, p = \lim_{k\to\infty} Dv(x_k)\}.$$

This set is nonempty as a corollary of Rademacher's theorem. In the case of a semiconcave function we have the following properties (see [3, 8]).

Theorem 2.3. *Let $v \in SC_{loc}(\Omega)$, with $\Omega \subset \mathbb{R}^n$ open. Then*

  (i) *v is locally Lipschitz in Ω;*
  (ii) *if $x_k \to x$, $p_k \in D^+v(x_k)$ and $p_k \to p$, then $p \in D^+v(x)$;*
  (iii) *for every $x \in \Omega$, $D^+v(x)$ is nonempty and coincides with the convex hull of $D^*v(x)$;*
  (iv) *v is differentiable at x if and only if $D^+v(x)$ is a singleton.*

In the following we consider functions $v = v(t, x)$ where $(t, x) \in \mathbb{R} \times \mathbb{R}^n$. For such functions, we use the symbol Dv (resp. D^+v, D^*v) to denote the gradient (resp. the generalized gradients) with respect to all variables, while we denote with $\nabla v, \nabla^+ v, \nabla^* v$ the corresponding gradients taken with respect to the x variable only, *i.e.* the gradients of the function $x \to v(t, x)$ for fixed t. We denote by Π_x the x-projection of a vector, that is, if $q = (q_t, q_x) \in \mathbb{R} \times \mathbb{R}^n$, then $\Pi_x(q) := q_x$. We have $\Pi_x Dv(t, x) = \nabla v(t, x)$ at every point (t, x) where v is differentiable, while the same relation does not necessarily hold for the generalized differentials. However, if v is semiconcave we have (see[1], Lem. 5)

$$v \in SC_{loc} \quad \Longrightarrow \quad \nabla^+ v(t, x) = \Pi_x(D^+v(t, x)). \tag{2.2}$$

The interest of semiconcavity in control problems comes from the fact that in many cases the value function is semiconcave provided the data are semiconcave. On the contrary, it is well known that other regularity properties, *e.g.* differentiability, are usually not inherited by the value function. For the Mayer problem considered here we have the following result (see [3], Th. 5.1).

Theorem 2.4. *If assumption (H1) holds and if $g \in SC_{loc}(\mathbb{R}^n)$, then $V \in SC_{loc}([0, T] \times \mathbb{R}^n)$.*

As recalled in the introduction, the *Hamiltonian function* associated with system (2.1) is

$$H(x, p) = \max_{u \in U} -p \cdot f(x, u). \tag{2.3}$$

It is well known (see *e.g.* [2, 13]) that the value function V is a viscosity solution of the Hamilton-Jacobi-Bellman equation (1.3) in the sense of Crandall-Lions. From this property it is easy to deduce that V satisfies

$$-p_t + H(x, p_x) = 0, \qquad \forall\, (p_t, p_x) \in D^*V(t, x), \quad (t, x) \in\,]0, T[\, \times \mathbb{R}^n. \tag{2.4}$$

This equality can be extended to all elements of $D^+V(t, x)$ provided x lies along an optimal trajectory and is not an endpoint, as shown by the following result (see [3], Prop. 4.4).

Theorem 2.5. *Let assumption (H1) hold, and let $y : [t, T] \to \mathbb{R}^n$ be an optimal trajectory for problem* **(M)** *with initial point (t, x). Then, for any $\tau \in\,]t, T[$, we have*

$$-p_t + H(y(\tau), p_x) = 0, \qquad \forall\, (p_t, p_x) \in D^+V(\tau, y(\tau)). \tag{2.5}$$

Let $u : [t, T] \to U$ be an optimal control for a point (t, x) and let $y = y(\cdot; t, x, u)$ be the corresponding optimal trajectory. An arc $p : [t, T] \to \mathbb{R}^n$ is called a *dual arc* associated with (u, y) if it solves the adjoint linearized equation

$$p'(s) = -f_x^T(y(s), u(s))\, p(s), \qquad s \in [t, T] \text{ a.e.} \tag{2.6}$$

(here f_x^T denotes the transpose of the matrix f_x) and satisfies, for a.e. $s \in [t, T]$, the property

$$-f(y(s), u(s)) \cdot p(s) \geq -f(y(s), v) \cdot p(s), \qquad \forall v \in U. \tag{2.7}$$

The well known *Pontryagin maximum principle* ensures the existence of a dual arc, satisfying a suitable condition at the terminal time called *transversality condition*. One possible statement of this result, which is sufficient for our purposes, is the following.

Theorem 2.6. *Let assumption (H1) hold, let u be an optimal control for problem* **(M)** *with initial point (t, x), and let $y(\cdot) = y(\cdot; t, x, u)$ be the corresponding optimal trajectory. Suppose that $D^+ g(y(T))$ is nonempty. Then, for any $q \in D^+ g(y(T))$, the arc $p : [t, T] \to \mathbb{R}^n$ obtained solving equation (2.6) with the terminal condition $p(T) = q$ is a dual arc associated with (u, y). In addition, p satisfies*

$$p(s) \in \nabla^+ V(s, y(s)), \qquad \forall s \in [t, T].$$

Proof. See [3], Theorem 4.2. See also *e.g.* [10, 15] for related results.   □

If the Hamiltonian H is differentiable, the maximum principle can be restated as an ordinary differential system. The following statement takes into account the fact that, for the problem we are considering, we cannot expect H to be differentiable at $p = 0$.

Theorem 2.7. *Suppose that hypothesis (H1) is satisfied and that the Hamiltonian H defined in (2.3) belongs to $C_{loc}^{1,1}(\mathbb{R}^n \times (\mathbb{R}^n \setminus \{0\}))$. Let $y : [t, T] \to \mathbb{R}^n$ be an optimal trajectory for the point (t, x) and let $p : [t, T] \to \mathbb{R}^n$ be a dual arc associated with y such that $p(\bar{s}) \neq 0$ for some $\bar{s} \in [t, T]$. Then $p(s) \neq 0$ for all $s \in [t, T]$, the arcs y, p are of class C^1 and solve the system*

$$\begin{cases} y'(s) = -H_p(y(s), p(s)) \\ p'(s) = H_x(y(s), p(s)) \end{cases} \qquad s \in [t, T]. \tag{2.8}$$

Proof. See *e.g.* [3]. The property that $p(s) \neq 0$ for all $s \in [t, T]$ follows from the fact that equation (2.6) is linear.   □

Let us recall some properties of convex sets which are needed for the analysis of the Hamiltonian H. Given $V \subset \mathbb{R}^n$ compact and convex, the *support function* of V is defined by

$$\sigma_V(p) = \max_{v \in V} v \cdot p.$$

A vector $q \in \mathbb{R}^n$ is called *normal* (in the sense of convex analysis) to V at a point $v \in V$ if it satisfies

$$\langle q, v - w \rangle \geq 0, \qquad \forall w \in V,$$

or, equivalently, if $\sigma_V(q) = q \cdot v$. The set of all normal vectors to V at v is called *normal cone* to V at v and is denoted by $N_V(v)$. If V possesses an outer normal ν at $v \in \partial V$ in the usual smooth sense, then it is easily seen that $N_V(v) = \{\lambda \nu : \lambda \geq 0\}$. A well known result in convex analysis (see [14]) is the following:

Proposition 2.8. *Given a compact convex set $V \subset \mathbb{R}^n$ and $p \in \mathbb{R}^n$, we have that σ_V is differentiable at p if and only if there exists a unique $v_p \in V$ such that $\sigma_V(p) = p \cdot v_p$, and in that case we have $D\sigma_V(p) = v_p$.*

The proposition shows that σ_V is not differentiable at $p = 0$ unless V is a singleton, and that $\sigma_V \in C^1(\mathbb{R}^n \setminus \{0\})$ if and only if V is strictly convex. Let us prove also the following technical lemma.

Lemma 2.9. *Let V be a compact convex set containing more than one element. Suppose that $\sigma_V \in C^1(\mathbb{R}^n \setminus \{0\})$. Then, for any $v \in V$, 0 is an extreme point of $N_V(v)$ (i.e. it cannot be written as a proper convex combination of elements of $N_V(v)$).*

Proof. Suppose that there exists $\bar{v} \in V$ such that 0 is not an extreme point of $N_V(\bar{v})$. This easily implies that $N_V(\bar{v})$ contains a line passing through 0; in particular, we can find $q_1 \neq 0$ such that q_1 and $-q_1$ both belong to $N_V(\bar{v})$. Recalling the definition of $N_V(\bar{v})$, we deduce that $q_1 \cdot (\bar{v} - v) = 0$ for all $v \in V$. Therefore, the set $\{v \in V \; : \; v \cdot q_1 = \sigma_V(q_1)\}$ coincides with V. Since V contains more than one element, Proposition 2.8 shows that σ_V is not differentiable at q_1, in contradiction with our hypotheses. $\qquad\square$

Finally, let us recall the following result, which characterizes the regions where the support function σ_V has a linear behaviour (see [5], Lem. 5.6).

Lemma 2.10. *Let $V, M \subset \mathbb{R}^n$ be convex sets, with V compact. The following two properties are equivalent:*

   (i) $\exists \bar{v} \in V \; : \; M \subset N_V(\bar{v})$;
   (ii) $\sigma_V(tp_0 + (1-t)p_1) = t\sigma_V(p_0) + (1-t)\sigma_V(p_1), \;\; \forall p_0, p_1 \in M, \;\; \forall t \in [0,1]$.

In particular, if V has a smooth boundary, the normal cone $N_V(\bar{v})$ is a half-line starting from the origin for any $\bar{v} \in \partial V$, and therefore the support function is strictly convex along all directions except the radial ones.

3. REGULARITY ALONG OPTIMAL TRAJECTORIES

We now study the differentiability of the value function along the optimal trajectories of our control problem. We need to introduce, in addition to (H1), the following assumption.

  (H2) (i) The set $f(x, U)$ is convex for all $x \in \mathbb{R}^n$;
        (ii) H belongs to $C_{loc}^{1,1}(\mathbb{R}^n \times (\mathbb{R}^n \setminus \{0\}))$;
        (iii) $g \in SC_{loc}(\mathbb{R}^n)$.

Remark 3.1. From Proposition 2.8 we see that property (ii) of (H2) is related to the strict convexity of the set $f(x, U)$. In fact, if (ii) holds then the boundary of $f(x, U)$ cannot have flat parts; however, it may have corners.

For our analysis we need a result about optimal trajectories obtained in [3], Theorem 4.6. In the original statement of this result g is assumed differentiable, but the proof applies also to the case where g is semiconcave, as already remarked in [3].

Theorem 3.2. *Let assumptions $(H1), (H2)$ hold. Given $(t, x) \in [0, T[\times \mathbb{R}^n$ and $\bar{p} = (\bar{p}_t, \bar{p}_x) \in D^*V(t, x)$ such that $\bar{p} \neq 0$, let us consider system (2.8) with initial conditions $y(t) = x$, $p(t) = \bar{p}_x$. Then the corresponding solution (y, p) is defined in $[t, T]$ and satisfies $p(s) \neq 0$ for $s \in [t, T]$; in addition, y is an optimal trajectory for (t, x) and p is a dual arc associated with y such that $p(t) \in \nabla^+V(s, y(s))$ for all $s \in]t, T]$.*

Before giving our main result, we prove some preliminary lemmas.

Lemma 3.3. *Assume property $(H1)$ and let $y : [t, T] \to \mathbb{R}^n$ be an optimal trajectory for problem* **(M)**. *Then, for all $s \in]t, T[$, the set $\Pi_x(D^+V(s, y(s)))$ is contained in some normal cone to the convex hull of $-f(y(s), U)$.*

Proof. Let us fix any $\bar{s} \in]t, T[$, and let us set for simplicity $\bar{x} = y(\bar{s})$. We deduce from Theorem 2.5 that every element of $D^+V(\bar{s}, \bar{x})$ has the form $(H(\bar{x}, q), q)$ for some $q \in \mathbb{R}^n$. Since D^+V is a convex set, this implies that, for all $q_0, q_1 \in \Pi_x(D^+V(\bar{s}, \bar{x}))$ and $\lambda \in [0, 1]$, we have

$$(\; \lambda H(\bar{x}, q_0) + (1-\lambda)H(\bar{x}, q_1) \; , \;\; \lambda q_0 + (1-\lambda)q_1 \;) \in D^+V(\bar{s}, \bar{x}).$$

Again using Theorem 2.5, we find

$$\lambda H(\bar{x}, q_0) + (1-\lambda)H(\bar{x}, q_1) = H(\; \bar{x} \; , \; \lambda q_0 + (1-\lambda)q_1 \;).$$

Since $H(\bar{x}, \cdot)$ is the support function of the convex hull of $-f(\bar{x}, U)$, we can apply Lemma 2.10 and conclude that $\Pi_x(D^+V(\bar{s}, \bar{x}))$ is contained in some normal cone to $\mathrm{co}\,(-f(\bar{x}, U))$. $\qquad\square$

Before proving the next lemma, we remark a simple property which will be often used in the following. Recalling (2.4), and using the fact that $H(x,0) = 0$ for all x, we deduce that

$$q = 0 \iff \Pi_x(q) = 0 \qquad \forall q \in D^*V(t,x), \ \forall (t,x) \in]0,T[\times \mathbb{R}^n. \tag{3.1}$$

Similarly, if $y : [t,T] \to \mathbb{R}^n$ is an optimal trajectory, then Theorem 2.5 implies that

$$q = 0 \iff \Pi_x(q) = 0 \qquad \forall q \in D^+V(s,y(s)), \ \forall s \in]t,T[. \tag{3.2}$$

Observe that in the previous formulas we use the same symbol 0 to denote the zero vector in $\mathbb{R} \times \mathbb{R}^n$ and in $\mathbb{R}^n$.

Lemma 3.4. *Let assumptions (H1), (H2) hold, and suppose that $f(x,U)$ contains more than one element for all $x \in \mathbb{R}^n$. Let $(t,x) \in [0,T[\times \mathbb{R}^n$ be a point such that $0 \notin D^*V(t,x)$. Then, if $y : [t,T] \to \mathbb{R}^n$ is any optimal trajectory for (t,x), we have that $0 \notin D^+V(s,y(s))$ for all $s \in]t,T]$.*

Proof. Let (t,x) be such that $0 \notin D^*V(t,x)$, and let y be an optimal trajectory for (t,x). We observe that we can find $\varepsilon > 0$ such that $0 \notin D^*V(s,y(s))$ for $s \in [t,t+\varepsilon]$; this is an easy consequence of the definition of D^*V. We take any $\bar{s} \in]t, t+\varepsilon]$ and set for simplicity $\bar{x} = y(\bar{s})$. We argue by contradiction and suppose that $0 \in D^+V(\bar{s},\bar{x})$. Then, by Theorem 2.3, the zero vector (in $\mathbb{R} \times \mathbb{R}^n$) is a convex combination of elements of $D^*V(\bar{s},\bar{x})$. Looking at the space components, we have that the zero vector (in $\mathbb{R}^n$) is a combination of elements $q_1, \ldots, q_n \in \Pi_x(D^*V(\bar{s},\bar{x}))$. Since $\Pi_x(D^*V(\bar{s},\bar{x})) \subset \Pi_x(D^+V(\bar{s},\bar{x}))$, we can use Lemmas 2.9 and 3.3 to conclude that $q_1, \ldots, q_n$ are all zero. Thus, $0 \in \Pi_x(D^*V(\bar{s},\bar{x}))$, which also implies, by (3.1), that $0 \in D^*V(\bar{s},\bar{x})$, contradicting our choice of $\bar{s}$. Hence, we have shown that

$$0 \notin D^+V(s,y(s)) \qquad \forall s \in]t, t+\varepsilon]. \tag{3.3}$$

To conclude our proof, we have to take into account the case $s \in]t+\varepsilon, T]$. We argue again by contradiction and suppose that there is $\bar{s} \in]t+\varepsilon, T]$ such that $0 \in D^+V(\bar{s},y(\bar{s}))$. We can consider the Mayer problem with final time $\bar{s}$ and final cost $V(\bar{s}, \cdot)$; the trajectory y is optimal for (t,x) also in this case. We apply Theorem 2.6 and deduce that there exists $p : [t,\bar{s}] \to \mathbb{R}^n$ which is a dual arc associated with y and satisfies $p(\bar{s}) = 0$, $p(s) \in \nabla^+V(s,y(s))$ for all $s \in [t,\bar{s}]$. Since the equation satisfied by p is linear, we have that $p \equiv 0$, and therefore $0 \in \nabla^+V(s,y(s))$ for all $s \in [t,\bar{s}]$. By (2.2) and by (3.2), this implies that $0 \in D^+V(s,y(s))$ for all $s \in]t,\bar{s}]$, in contradiction with (3.3). $\qquad \square$

We are now ready to give our result about the differentiability of V along optimal trajectories. Let us introduce the following additional hypothesis:

(H3) The set $f(x,U)$ has a C^1 boundary for all $x \in \mathbb{R}^n$, if $n > 1$; it is a nondegenerate segment for all $x \in \mathbb{R}$, if $n = 1$.

Theorem 3.5. *Let assumptions (H1), (H2), (H3) hold, and let $y : [t,T] \to \mathbb{R}^n$ be an optimal trajectory for problem **(M)** with initial point $(t,x) \in [0,T[\times \mathbb{R}^n$. Suppose that $0 \notin D^*V(t,x)$ and that g is differentiable at $y(T)$. Then V is differentiable at all points of the form $(s,y(s))$, with $t < s < T$.*

Proof. Let us consider an optimal trajectory y starting from a point (t,x) satisfying our hypotheses. We fix any $\bar{s} \in]t,T[$ and we set $\bar{x} = y(\bar{s})$. Let us consider the dual arc p associated with y such that $p(T) = Dg(y(T))$. By Theorem 2.6 and by (2.2), it satisfies $p(\bar{s}) \in \nabla^+V(\bar{s},\bar{x}) = \Pi_x(D^+V(\bar{s},\bar{x}))$.

Let us now take any $\bar{q} \in \Pi_x(D^*V(\bar{s},\bar{x}))$. Since $D^*V \subset D^+V$, we have also $\bar{q} \in \Pi_x(D^+V(\bar{s},\bar{x}))$. By Lemma 3.3, $\Pi_x(D^+V(\bar{s},\bar{x}))$ is contained in some normal cone to $-f(\bar{x},U)$. By assumption (H3), any normal cone to $-f(\bar{x},U)$ is a half-line starting from the origin. In addition, from Lemma 3.4 and (3.2), we know that $0 \notin \Pi_x(D^+V(\bar{s},\bar{x}))$. This implies that $\bar{q} = \mu p(\bar{s})$ for some $\mu > 0$.

Now let us set $p_1(s) = \mu p(s)$ for $s \in [\bar{s},T]$. By the homogeneity properties of H, the pair (y,p_1) solves system (2.8) with conditions $y(\bar{s}) = \bar{x}$, $p_1(\bar{s}) = \bar{q}$. By Theorem 3.2, p_1 is a dual arc associated with y in $[\bar{s},T]$

and satisfies $p_1(T) = Dg(y(T))$. But then the pair (y, p_1) coincides with (y, p), since they satisfy the same terminal conditions. In particular, $\bar{q} = p_1(\bar{s}) = p(\bar{s})$. Thus we have proved that $\Pi_x(D^*V(\bar{s}, \bar{x})) = \{p(\bar{s})\}$.

By (2.4), $D^*V(\bar{s}, \bar{x})$ reduces to the singleton $\{ (H(p(\bar{s})) , p(\bar{s})) \}$. Proposition 2.3 then implies that $D^+V(\bar{s}, \bar{x})$ is also a singleton and that V is differentiable at $(\bar{s}, \bar{x})$. $\qquad\square$

Example 3.6. It is easy to construct examples where the hypotheses of the previous theorem are satisfied. For instance, one can take $U \subset \mathbb{R}^n$ to be a smooth uniformly convex set (*e.g.* an ellipsoid) and $f(x, u) = A(x)u + b(x)$, where $A(x)$ is a nonsingular $n \times n$-matrix, $b(x)$ is a vector valued function, $A(x), b(x)$ are of class $C^{1,1}$. Then, if $g \in C^1(\mathbb{R}^n)$, the previous theorem can be applied to all points (t, x) such that $0 \notin D^*V(t, x)$. Let us remark that in the next section (Cor. 4.4) we will prove that if $Dg(x) \neq 0$ for all x then $0 \notin D^*V(t, x)$ for all (t, x), and so in this case the previous theorem applies to all optimal trajectories.

Example 3.7. Let us now give some counterexamples showing that the value function may fail to be differentiable along optimal trajectories if some of the hypotheses in Theorem 3.8 are violated. We consider the one-dimensional system governed by the equation $y' = u$, where $u \in U = [-1, 1]$. Then it is easy to see that, for any given $g \in C(\mathbb{R})$, the value function of problem **(M)** is

$$V(t, x) = \min\{g(z) \ : \ |z - x| \leq T - t\}.$$

(i) Consider first a final cost $g : \mathbb{R} \to \mathbb{R}$ which is strictly increasing. Then, for any $(t, x) \in [0, T] \times \mathbb{R}$, the unique optimal control is the constant one $u \equiv -1$, and the corresponding trajectory is $y(s) = x + t - s$. The value function is $V(t, x) = g(x + t - T)$. Notice that, if $y(T) = x + t - T$ is a point of nondifferentiability for g, then V is not differentiable at $(s, y(s))$, for any $s \in [t, T]$. In this case, the optimal trajectory y consists entirely of points of nondifferentiability for V. Such a behaviour is excluded in Theorem 3.5 by the assumption that g is differentiable at $y(T)$.

(ii) Let us consider the same system, taking now as a final cost any function g with the following properties:

$$\begin{aligned}
&\text{(a) } g(x) = 0 \text{ for all } x \leq -2, \\
&\text{(b) } g(x) = -x \text{ for all } x \geq 0, \\
&\text{(c) } g(x) \geq 0 \text{ for } x \in [-2, 0].
\end{aligned}$$

We assume that the values of g in $[-2, 0]$ are such that $g \in C^1(\mathbb{R})$. We consider problem **(M)** with $T = 2$. Although we have not completely specified g, our requirements suffice to characterize the optimal trajectories starting from points (t, x) with $t \in [0, 1]$. In fact, the points reachable at time $T = 2$ by an admissible trajectory starting from (t, x) are those of the interval $[x - 2 + t, x + 2 - t]$. If $t \in [0, 1]$ and $x \leq t - 2$ we have $x - 2 + t \leq 2t - 4 \leq -2$ and $x + 2 - t \leq 0$, and so it is easily seen that the minimum of g on $[x - 2 + t, x + 2 - t]$ is 0. On the other hand, if $x \geq t - 2$, we find that the optimal trajectory is the right-most one ending at $x + 2 - t$. It follows that, for $t \in [0, 1]$,

$$V(t, x) = \begin{cases} 0 & x \leq t - 2 \\ t - 2 - x & x \geq t - 2. \end{cases}$$

Observe that the segment $y(s) = s - 2$ is an optimal trajectory for the point $(0, -2)$. However, the points of this trajectory are singular for all $s \in [0, 1]$, since $\nabla^+V(s, y(s)) = [-1, 0]$ for all $s \in [0, 1]$. This shows that the hypothesis that $0 \notin D^*V(t, x)$ in Theorem 3.5 is essential.

In the case where the function f governing system (2.1) is independent of x, we can prove differentiability of V along optimal trajectories without requiring the smoothness of $f(U)$ as in assumption (H3).

Theorem 3.8. *Let assumptions* (H1), (H2) *hold, and suppose that* $f = f(u)$. *Given* $(t, x) \in [0, T[\times \mathbb{R}^n$ *such that* $0 \notin D^*V(t, x)$, *let* $y : [t, T] \to \mathbb{R}^n$ *be an optimal trajectory for* (t, x) *and suppose that* g *is differentiable at* $y(T)$. *Then* V *is differentiable at all points of the form* $(s, y(s))$, *with* $t < s < T$.

Proof. We consider only the case where $f(U)$ is not a singleton, since otherwise the result is trivial. Let us take a point $(t, x) \in [0, T[\times \mathbb{R}^n$ such that $0 \notin D^*V(t, x)$. By Theorem 2.6 and Lemma 3.4, any optimal trajectory y starting from (t, x) has a nonzero dual arc p. Since $f = f(u)$, we have that $H = H(p)$; solving (2.8) we find that $p(s) \equiv \bar{p}$ and $y(s) = x - (s - t)DH(\bar{p})$ for some constant $\bar{p} \neq 0$. Thus, any optimal trajectory starting from (t, x) is a straight line.

Now, given any optimal trajectory y starting from (t, x), let us fix $\bar{s} \in]t, T[$ and let us set $\bar{x} = y(\bar{s})$. By Theorem 2.3, to prove differentiability of V at $(\bar{s}, \bar{x})$, it suffices to show that $D^*V(\bar{s}, \bar{x})$ is a singleton. Let us take any $\bar{q} \in \Pi_x(D^*V(\bar{s}, \bar{x}))$. From Lemma 3.4 and from (3.1) we deduce that $\bar{q} \neq 0$. By Theorem 3.2, if we set

$$\tilde{y}(s) = \bar{x} - (s - \bar{s})DH(\bar{q}), \qquad s \in [\bar{s}, T]$$

we have that $\tilde{y}$ is optimal for $(\bar{s}, \bar{x})$ and that $\bar{q} \in D^+g(\tilde{y}(T))$. Let us define the trajectory

$$\hat{y}(s) = \begin{cases} y(s) & s \in [t, \bar{s}] \\ \tilde{y}(s) & s \in [\bar{s}, T]. \end{cases}$$

It is easily checked that $\hat{y}$ is optimal for (t, x). Therefore, $\hat{y}$ is a straight line and, since it coincides with y in $[t, \bar{s}]$, it must coincide with y on $[\bar{s}, T]$ also. Now, recalling that $\bar{q} \in D^+g(\tilde{y}(T))$, we deduce that $\bar{q} = Dg(y(T))$. This shows that the space component of any element of $D^*V(\bar{s}, \bar{x})$ coincides with $Dg(y(T))$, which implies, by (2.4), that $D^*V(\bar{s}, \bar{x})$ is a singleton. $\qquad \square$

4. OPTIMALITY CONDITIONS

We now show how the results of the previous section can be applied to derive further optimality conditions in addition to those of Theorem 3.2. We assume throughout that assumptions (H1), (H2) are satisfied. Let us fix some notation. For a given $(t, x) \in [0, T[\times \mathbb{R}^n$, we denote by $\mathcal{F}_{t,x}$ the set of optimal trajectories for (t, x) and we call $\mathcal{M}_{t,x}$ the map from $D^*V(t, x) \backslash \{0\}$ to $\mathcal{F}_{t,x}$ obtained by solving system (2.8) as described in Theorem 3.2. It is natural to investigate whether this map is one-to-one, that is, injective and surjective; if this holds, then one has a complete description of the optimal trajectories of **(M)**. A result of this kind has been proved for other classes of optimal control problems [4,6,8]; however, it is easy to see that it cannot hold in full generality for the problem considered here. For instance, in Example 3.7 (i) the optimal trajectory is unique for all initial points, even for the points where V is not differentiable and thus D^*V is not a singleton. However, such a behaviour is related to the nondifferentiability of the final cost and can be ruled out assuming that $g \in C^1(\mathbb{R}^n)$. More substantial difficulties arise when $0 \in D^*V(t, x)$; in this case there can be optimal trajectories starting from (t, x) which are not solutions of class C^1 of system (2.8). Therefore, we have to deal with this case separately in the following analysis.

Theorem 4.1. *If $g \in C^1(\mathbb{R}^n)$ the map $\mathcal{M}_{t,x}$ is injective for all (t, x). Moreover, if either assumption (H3) holds or $f = f(u)$, then $\mathcal{M}_{t,x}$ is one-to-one at all points (t, x) such that $0 \notin D^*V(t, x)$.*

Proof. Recall that different elements of $D^*V(t, x)$ have different space components, by (2.4). Consider two elements $\bar{p}, \bar{q} \in D^*V(t, x) \backslash \{0\}$ and suppose that the corresponding solutions of (2.8), which we denote by $(y, p), (\hat{y}, q)$, are such that $y \equiv \hat{y}$. Since g is differentiable, we have $p(T) = Dg(y(T)) = Dg(\hat{y}(T)) = q(T)$. But then the pairs (y, p) and $(\hat{y}, q)$ satisfy system (2.8) with the same terminal conditions, and therefore coincide. In particular $\bar{p} = \bar{q}$, which proves that our map is injective.

Suppose now that either assumption (H3) holds or $f = f(u)$, let y be any optimal trajectory for (t, x) and let p be the dual arc associated with y such that $p(T) = Dg(y(T))$. Then, by Theorems 2.6, 3.5 and 3.8, V is differentiable at $(s, y(s))$ and satisfies $\nabla V(s, y(s)) = p(s)$ for all $s \in]t, T[$. It follows that $p(t) = \lim_{s \to t} p(s)$ is the space component of an element of $D^*V(t, x)$. Since the pair (y, p) solves (2.8) by Theorem 2.7, this proves our assertion. $\qquad \square$

Let us turn to the case where $0 \in D^*V(t,x)$.

Theorem 4.2. *Let $(t,x) \in [0,T[\times\mathbb{R}^n$ be such that $0 \in D^*V(t,x)$. Then there exists $y : [t,T] \to \mathbb{R}^n$ optimal trajectory for (t,x) such that $0 \in D^+g(y(T))$.*

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

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

Let y_k be an optimal trajectory for (t_k,x_k) and let p_k be an associated dual arc satisfying $p_k(T) \in D^+g(y_k(T))$. By standard compactness properties of trajectories we can assume, after possibly passing to a subsequence, that $y_k \to y$ uniformly, where y is an trajectory of (2.1) starting from (t,x). Since

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

we see that y is optimal. In addition, we have $p_k(t_k) = \nabla V(t_k,x_k) \to 0$. Since the equation satisfied by p_k is linear, with a coefficient which is uniformly bounded by assumption (H1), this implies that $p_k(T) \to 0$ as well. By Proposition 2.3 (ii), we conclude that $0 \in D^+g(y(T))$. □

Observe that the above result is weaker than the one in Theorem 3.2 because it only gives the existence of the optimal trajectory y without saying that y is a solution of class C^1 of system (2.8) with suitable initial conditions. However, such a statement cannot be improved; at the points where $0 \in D^*V(t,x)$ we cannot expect to characterize the optimal trajectories by means of the map $\mathcal{M}_{t,x}$, as the following example shows.

Example 4.3. Consider again the one-dimensional system governed by $y' = u \in [-1,1]$. Suppose that the final cost g has a global minimum at some point x_0. Then, for any $t < T$, we have $V(t,x) \equiv g(x_0)$ for all x such that $|x - x_0| < T - t$; any trajectory y going from (t,x) to (T,x_0) with $|y'| \le 1$ a.e. is optimal. Therefore, V is differentiable at all such points (t,x) and has zero gradient. The optimal trajectories are infinitely many and not necessarily differentiable; in addition, their structure is different at each of these points, although the gradient of V is the same. Therefore, at these points we cannot recover the optimal trajectories from the knowledge of $D^*V(t,x)$ alone, as it happens in Theorem 4.1.

We can derive some easy consequences of Theorem 4.2.

Corollary 4.4. *If $0 \notin D^+g(x)$ for all $x \in \mathbb{R}^n$, then $0 \notin D^*V(t,x)$ for all $(t,x) \in [0,T] \times \mathbb{R}^n$.*

Corollary 4.5. *Suppose that $g \in C^1(\mathbb{R}^n)$ with $Dg(x) \neq 0$ for all $x \in \mathbb{R}^n$, and that either assumption (H3) holds or $f = f(u)$. Then V is differentiable at a point $(t,x) \in [0,T] \times \mathbb{R}^n$ if and only if the optimal trajectory starting at (t,x) is unique.*

Proof. By Proposition 2.3, V is differentiable at (t,x) if and only if $D^*V(t,x)$ is a singleton. Thus, the statement follows from the previous corollary and from Theorem 4.1. □

The equivalence stated in Corollary 4.5 is well known for problems in calculus of variations [11, 12] and for optimal control problems with exit time [4, 6]. If the hypothesis $Dg \neq 0$ is not satisfied we can prove the following weaker version of this property (see also [3], Th. 4.7).

Theorem 4.6. *Suppose that $g \in C^1(\mathbb{R}^n)$ and that either assumption (H3) holds or $f = f(u)$. Then, for any $(t,x) \in [0,T[\times\mathbb{R}^n$ the following holds: if V is not differentiable at (t,x) then there exists more than one optimal trajectory for problem* **(M)** *with initial point (t,x). On the other hand, if V is differentiable at (t,x) with nonzero gradient, the optimal trajectory is unique.*

Proof. Suppose that V is not differentiable at a point $(t,x) \in [0,T[\times\mathbb{R}^n$. By Proposition 2.3, we can find two distinct elements $\bar{p}, \tilde{p} \in D^*V(t,x)$. Then we can apply Theorems 3.2 or 4.2 to find two optimal trajectories

$\bar{y}, \tilde{y}$ associated with $\bar{p}, \tilde{p}$. If $\bar{p}, \tilde{p}$ are both nonzero, the two trajectories are distinct by Theorem 4.1. They are distinct also if one of the two vectors, say for instance $\tilde{p}$, is zero, since in that case we have $Dg(\bar{y}(T)) \neq 0$ and $Dg(\tilde{y}(T)) = 0$. This proves the first part of the theorem. The second one follows from Theorem 4.1.   □

The previous theorem leaves open the possibility that the optimal trajectory is not unique at a point where V is differentiable with zero gradient. Such a behaviour can actually occur, as we have seen in Example 4.3.

References

[1] P. Albano and P. Cannarsa, Propagation of singularities for solutions of nonlinear first order partial differential equations. *Arch. Ration. Mech. Anal.* **162** (2002) 1-23.

[2] M. Bardi and I. Capuzzo Dolcetta, *Optimal control and viscosity solutions of Hamilton-Jacobi equations.* Birkhäuser, Boston (1997).

[3] P. Cannarsa and H. Frankowska, Some characterizations of optimal trajectories in control theory. *SIAM J. Control Optim.* **29** (1991) 1322-1347.

[4] P. Cannarsa, C. Pignotti and C. Sinestrari, Semiconcavity for optimal control problems with exit time. *Discrete Contin. Dyn. Syst.* **6** (2000) 975-997.

[5] P. Cannarsa and C. Sinestrari, Convexity properties of the minimum time function. *Calc. Var.* **3** (1995) 273-298.

[6] P. Cannarsa and C. Sinestrari, On a class of nonlinear time optimal control problems. *Discrete Contin. Dyn. Syst.* **1** (1995) 285-300.

[7] P. Cannarsa and C. Sinestrari, *Semiconcave functions, Hamilton-Jacobi equations and optimal control.* Birkhäuser, Boston (2004).

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

[9] P. Cannarsa and M. E. Tessitore, On the behaviour of the value function of a Mayer optimal control problem along optimal trajectories, in *Control and estimation of distributed parameter systems* (Vorau, 1996). *Internat. Ser. Numer. Math.* **126** 81-88 (1998).

[10] F.H. Clarke and R.B. Vinter, The relationship between the maximum principle and dynamic programming. *SIAM J. Control Optim.* **25** (1987) 1291-1311.

[11] W.H. Fleming, The Cauchy problem for a nonlinear first order partial differential equation. *J. Diff. Eq.* **5** (1969) 515-530.

[12] N.N. Kuznetzov and A.A. Šiškin, On a many dimensional problem in the theory of quasilinear equations. *Z. Vyčisl. Mat. i Mat. Fiz.* **4** (1964) 192-205.

[13] P.L. Lions, *Generalized solutions of Hamilton-Jacobi equations.* Pitman, Boston (1982).

[14] R.T. Rockafellar, *Convex Analysis.* Princeton University Press, Princeton (1970).

[15] X.Y. Zhou, Maximum principle, dynamic programming and their connection in deterministic control. *J. Optim. Theory Appl.* **65** (1990) 363-373.