SEMICONCAVITY RESULTS FOR OPTIMAL CONTROL PROBLEMS ADMITTING NO SINGULAR MINIMIZING CONTROLS

P. CANNARSA AND L. RIFFORD

ABSTRACT. Semiconcavity results have generally been obtained for optimal control problems in absence of state constraints. In this paper, we prove the semiconcavity of the value function of an optimal control problem with end-point constraints for which all minimizing controls are supposed to be nonsingular.

1. INTRODUCTION

The mathematical literature is rich of results that describe the regularity of the value function of optimal control problems without state constraints, much less so if constraints are present.

For instance, given $(t, x) \in [0, \infty) \times \mathbb{R}^n$, consider the optimal control problem which consists of minimizing, with respect to $u(\cdot)$, the Bolza type functional

$$J(t, x; u(\cdot)) = \int_0^t L(y_u(s; t, x), u(s))ds + \ell(y_u(0; t, x))$$

where $y_u(\cdot; t, x)$ is the solution of the state equation

$$\dot{y}(s) = f(y(s), u(s)) \quad \text{a.e. in } (0, t), \qquad y(t) = x\,.$$

If f is sufficiently smooth, then the value function

$$v(t, x) := \inf \left\{ J(t, x; u(\cdot)) \mid u(\cdot) \in L^1 \right\} \qquad (t, x) \in [0, \infty) \times \mathbb{R}^n$$

can be shown to be, roughly speaking, as regular as the problem data L and ℓ, where the term *regular* stands for continuous, Lipschitz continous, or semiconcave, see, *e.g.*, [4] and [8].

We recall that a function $g : \Omega \to \mathbb{R}$ defined on an open set $\Omega \subset \mathbb{R}^N$ is said to be *locally semiconcave* if for each compact convex set $K \subset \Omega$, there is a positive constant C_K such that

$$\mu g(x) + (1 - \mu)g(y) - g(\mu x + (1 - \mu)y) \leq \mu(1 - \mu)C_K|x - y|^2$$

for any $\mu \in [0, 1]$, any $x, y \in K$. The importance of semiconcavity in control theory is widely acknowledged. Initially used as a tool for uniqueness in dynamic programmimg, it is nowadays mainly regarded as a property ensuring

---

Dipartimento di Matematica, Università di Roma Tor Vergata, Via della Ricerca Scientifica 1, 00133 Roma, Italy (e-mail: cannarsa@mat.uniroma2.it).

Université de Nice-Sophia Antipolis, Labo. J.A. Dieudonné, UMR 6621, Parc Valrose, 06108 Nice, France (e-mail: Ludovic.Rifford@math.cnrs.fr).

This research was completed, in part, while the first author was visiting the Université de Lyon 1 and, in part, while th e second was visiting the Università di Roma Tor Vergata on an INdAM grant. The authors are grateful to the above institutions for their support.

better regularity than a.e. differentiability: indeed, the Hausdorff dimension of the singular set of a semiconcave function can be sharply estimated, and the way how singularities propagate is fairly well understood, see [8]. Moreover, semiconcavity has been successfully applied to Lyapunov stability and feedback stabilization for nonlinear control systems, see for example [18], [19], [23], and [20] for further references.

In the presence of state constraints, however, it turns out that the only semiconcavity results that are available are restricted to optimal exit time problems, see [5], [6], [7], and [8]. In particular, for the above problems, no constraints can be active on the interior of trajectories and terminal time must be *free*.

In the present paper, we are interested in obtaining the semiconcavity of the value function of a *fixed* terminal time Bolza problem, with initial cost ℓ replaced by an end-point constraint. More precisely, given $x_0 \in \mathbb{R}^n$, for any control $u(\cdot) \in \mathcal{U} := L^1([0,\infty);\mathbb{R}^m)$, let us denote by $x_u(\cdot)$ the solution of the Cauchy problem

$$\dot{x}(s) = f(x(s), u(s)), \quad s > 0 \text{ a.e.}, \qquad x(0) = x_0, \tag{1}$$

on the interval $[0,\infty)$[1]. The value function $V : (0,\infty) \times \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is then defined as

$$V(t,x) := \inf \left\{ \int_0^t L(x_u(s), u(s))ds \mid u(\cdot) \in \mathcal{U} \text{ s.t. } x_u(t) = x \right\}, \tag{2}$$

with the convention that $V(t,x) = \infty$ if there is no control $u(\cdot) \in \mathcal{U}$ such that $x_u(t) = x$. This problem is much more complicated that the one with an initial cost: to begin with, V may well be equal to ∞ on a large part of $(0,\infty) \times \mathbb{R}^n$. Also, in this case there may be abnormal extremals, which can be associated, roughly speaking, to non-Lipschitz regularity points of the corresponding value function. To cope with such difficulties we will use the approach of geometric control, assuming that our problem admits no singular optimal controls (see section 2 for definitions). Moreover, since our method is based on the Pontryagin Maximum Principle, we will restrict the class of control system to affine systems of the form

$$\dot{x} = f(x,u) := f_0(x) + \sum_{i=1}^m u_i f_i(x), \tag{3}$$

where $f_0, f_1, \cdots, f_m$ are m vector fields on $\mathbb{R}^n$, and where $u = (u_1, \cdots, u_m)$ belongs to $\mathbb{R}^m$. We will suppose that:

(A1) the family $\{f_0, f_1, \cdots, f_m\}$ consists of vector fields of class $C^{1,1}_{loc}$ on $\mathbb{R}^n$ with sublinear growth, *i.e.*, such that

$$|f_i(x)| \leq M(|x|+1), \quad \forall x \in \mathbb{R}^n, \quad \forall i = 0, 1, \cdots, m$$

for some constant $M > 0$;

(A2) the Lagrangian L satisfies the following conditions:

(i) for any $x \in \mathbb{R}^n$, the function $u \mapsto L(x,u)$ is of class C^2, and $(x,u) \mapsto D^2_u L(x,u)$ is continuous on $\mathbb{R}^n \times \mathbb{R}^m$ with positive definite values;

[1] Here, we assume for sake of simplicity that any solution $x_u(\cdot)$ is defined on $[0,\infty)$.

(ii) there exist $c_0 \geq 0$ and $\theta : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\theta(q)/q \to +\infty$ as $q \to +\infty$, and

$$L(x,u) \geq \theta(|u|_m) - c_0, \quad \forall x \in \mathbb{R}^n, \forall u \in \mathbb{R}^m \,;$$

(iii) for all $r > 0$ there exists $K(r) > 0$ such that

$$|\zeta| \leq K(r)\theta(|u|_m),$$

for all $x \in B_r, u \in \mathbb{R}^m$ and $\zeta \in \partial_x L(x,u)$;

(iv) L is locally semiconcave in the x-variable uniformly for u in all compact sets of $\mathbb{R}^m$, that is, for each compact convex set $K \subset \mathbb{R}^n$ and each compact set $U \subset \mathbb{R}^m$, there is a constant $C_{K,U} > 0$ such that

$$\mu L(x,u) + (1-\mu)L(y,u) - L(\mu x + (1-\mu)y, u) \\ \leq \mu(1-\mu)C_{K,U}|x-y|^2,$$

for an y $\mu \in [0,1]$, any $x, y \in K$, and any $u \in U$.

In order to prove the semiconcavity of the value function of optimal control problems without of state constraints one commonly applies PDE techniques based on comparison arguments, or else direct methods which use ad hoc perturbations of optimal trajectories, see, e.g., [8]. In the present case, our technique is completely different: invoking a nonsmooth version of Pontryagin's Maximum Principle, we manage to represent optimal trajectories as a family of arcs parametrized by the elements of a suitable compact set. Then, the smooth dependence of such a family on parameters yields the required regularity.

As a corollary of our main result, we derive the semiconcavity of the distance function associated with a sub-Riemannian structure. We note that the regularity of such a function has so far been investigated only in a subanalytic set-up, see [1], [13], and [24].

The *outline* of the paper is the following. In section 2, we introduce the end-point mapping and the notion of singular control. In section 3, we derive regularity properties of the value function, and in section 4 we prove optimality conditions based on the regularity of V. Section 5 is devoted to a special class of problems associated with the so-called *fat distributions*, while section 6 studies the distance function in the general sub-Riemannian case.

Notation

Throughout this paper, we denote by $\langle \cdot, \cdot \rangle$ and $|\cdot|$, respectively, the Euclidean scalar product and norm in the state space $\mathbb{R}^n$. For any $x \in \mathbb{R}^n$ and any $r > 0$, we set $B(x,r) := \{y \in \mathbb{R}^n : |y-x| < r\}$, and we use the abbreviations $B_r := B(0,r)$, $B := B_1$.

We denote by $\langle \cdot, \cdot \rangle_m$ and $| \cdot |_m$, respectively, the Euclidean scalar product and norm in the control space $\mathbb{R}^m$.

For any matrix M, we denote by M^* the transpose of M, and by $\|M\|$ its norm (with respect to Euclidean norm).

For any control $u(\cdot) \in \mathcal{U} := L^1([0,\infty);\mathbb{R}^m)$, we denote by $\|u(\cdot)\|_1$ the L^1 norm of $u(\cdot)$.

2. The End-Point mapping

Let a point $x_0 \in \mathbb{R}^n$ and some time $t>0$ be fixed. The *end-point mapping* associated with system (3) (with initial state x_0 at time t) is the function defined by

$$E^{x_0,t} : \mathcal{U} \longrightarrow \mathbb{R}^n, \qquad u(\cdot) \longmapsto x_u(t).$$

Recall that $\mathcal{U}$ is a Banach space with the L^1-norm. The differential of the end-point mapping is described by the following well-known result.

Proposition 2.1. *Under assumption* (A1)*, $E^{x_0,t}$ is of class C^1 on $\mathcal{U}$, and its differential at some control $u(\cdot)$ is given by the linear operator*

$$dE^{x_0,t}(u(\cdot)) : \mathcal{U} \longrightarrow \mathbb{R}^n, \quad v(\cdot) \longmapsto \zeta(t),$$

where $\zeta(\cdot)$ is the unique solution of the Cauchy problem

$$\dot{\zeta}(s) = A(s)\zeta(s) + B(s)v(s), \quad s \in [0,t] \quad a.e., \quad \zeta(0)=0.$$

Here, matrices $A(s)$ and $B(s)$ are defined by

$$A(s) := \frac{\partial f}{\partial x}(x_u(s),u(s)) \qquad B(s) := \frac{\partial f}{\partial u}(x_u(s),u(s)),$$

for a.e. $s \in [0,t]$.

The proof of Proposition 2.1 is straightforward, see [15] or [24].

Remark 2.2. Under assumption (A1), the function

$$\mathcal{E}^{x_0} : (0,\infty) \times \mathcal{U} \longrightarrow \mathbb{R}^n, \quad (t,u(\cdot)) \longmapsto x_u(t),$$

is indeed of class $C^{1,1}_{loc}$ on $(0,\infty)\times\mathcal{U}$. This fact will be useful in the sequel.

Notice that, by definition,

$$A(s) = df_0(x_u(s)) + \sum_{i=1}^m u_i(s) df_i(x_u(s)),$$

and

$$B(s) = (f_1(x_u(s)), \cdots, f_m(x_u(s)))$$

for a.e. $s \in [0,t]$. So, by Proposition 2.1, the differential of $E^{x_0,t}$ at $u(\cdot)$ corresponds to the end-point mapping associated with the system obtained linearizing (3) along $(x_u(\cdot),u(\cdot))$, with initial condition 0 at time $t=0$. Therefore we can represent $dE^{x_0,t}(u(\cdot))$ by

$$dE^{x_0,t}(u(\cdot)) : \mathcal{U} \longrightarrow \mathbb{R}^n, \quad v(\cdot) \longmapsto S(t)\int_0^t S(s)^{-1}B(s)v(s)ds, \tag{4}$$

where $S(\cdot)$ is the solution of the Cauchy problem

$$\dot{S}(s) = A(s)S(s), \quad S(0)=I_n.$$

We now introduce a notion which is crucial for our approach.

Definition 2.3. *A control $u(\cdot) \in \mathcal{U}$ is said to be* singular *for $E^{x_0,t}$ if $dE^{x_0,t}(u(\cdot))$ is not surjective. Otherwise, $u(\cdot)$ is said to be* nonsingular *or* regular.

Let us define the pre-Hamiltonian $H_0 : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ by

$$\begin{aligned} H_0(x,p,u) &:= \langle p, f(x,u)\rangle \\ &= \langle p, f_0(x)\rangle + \sum_{i=1}^m u_i \langle p, f_i(x)\rangle, \end{aligned} \tag{5}$$

for any triple $(x,p,u) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$. Notice that H_0 is of class $C^{1,1}_{loc}$ in the x variable, and of class C^∞ in p,u. Adopting Hamiltonian formalism, we have the following well-known characterization of singular controls.

Proposition 2.4. *A control $u(\cdot) \in \mathcal{U}$ is singular for $E^{x_0,t}$ if and only if there exists an absolutely continuous arc $p(\cdot) : [0,t] \to \mathbb{R}^n \setminus \{0\}$ such that*

$$\left\{ \begin{array}{rcl} \dot{x}_u(s) &=& \nabla_p H_0(x_u(s), p(s), u(s)) \\ -\dot{p}(s) &=& \nabla_x H_0(x_u(s), p(s), u(s)) \end{array} \right. \tag{6}$$

and

$$\nabla_u H_0(x_u(s), p(s), u(s)) = 0, \tag{7}$$

for a.e. $s \in [0,t]$.

In particular, given a control $u(\cdot) \in \mathcal{U}$, along the associated trajectory $x_u(\cdot) : [0,t] \to \mathbb{R}^n$ we have

$$\left\{ \begin{array}{rcl} \nabla_x H_0(x_u(s), p, u(s)) &=& A(s)^* p \\ \nabla_p H_0(x_u(s), p, u(s)) &=& f_0(x_u(s)) + B(s)u(s) \\ \nabla_u H_0(x_u(s), p, u(s)) &=& B(s)^* p, \end{array} \right.$$

for any $s \in [0,t]$ and any $p \in \mathbb{R}^n$. Consequently, a control $u(\cdot) \in \mathcal{U}$ is singular for $E^{x_0,t}$ if and only if there exists an absolutely continuous arc $p(\cdot) : [0,t] \to \mathbb{R}^n \setminus \{0\}$ such that

- (6) is satisfied a.e. on $[0,t]$
- $p(\cdot)$ is orthogonal to each vector $f_1(x_u(\cdot)), \cdots, f_m(x_u(\cdot))$ on $[0,t]$.

Example 2.5. By Proposition 2.4, it can be easily seen that the control system, known as the "nonholonomic integrator",

$$\dot{x} = u_1 f_1(x) + u_2 f_2(x) = u_1 \begin{pmatrix} 1 \\ 0 \\ x_2 \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ 1 \\ -x_1 \end{pmatrix}$$

does not admit nontrivial singular controls. In other terms, for each $t > 0$ and each $u(\cdot) \in L^1([0,t];\mathbb{R}^m) \setminus \{0\}$, the mapping $dE^{x_0,t}(u(\cdot))$ is surjective. Actually, this property is satisfied by a general class of control systems which will be studied later on in this paper (see Section 5).

3. Properties of the value function V

3.1. **Existence of optimal controls.**

Recall that the value function $V : (0,\infty) \times \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is defined for each pair $(t,x) \in (0,\infty) \times \mathbb{R}^n$ as the infimum of the cost functional

$$C_t(u(\cdot)) := \int_0^t L(x_u(s), u(s))ds,$$

over all control $u(\cdot) \in \mathcal{U}$ steering x_0 to x in time t. If no such control exist, then we set $V(t,x) = \infty$.

Proposition 3.1. *Assume* (A1)-(A2) *and let* $(t,x) \in (0,\infty) \times \mathbb{R}^n$. *If there exists a control steering* x_0 *to* x *in time* t*, then there also exists a control* $u(\cdot) \in \mathcal{U}$*, steering* x_0 *to* x *in time* t*, which minimizes* $C_t(\cdot)$.

The proof of the above result is based on the following lemma which will be very useful in the sequel.

Lemma 1. *Let* (A1)-(A2) *be satisfied and let* $(u_k(\cdot))_k$ *be a sequence of controls in* $\mathcal{U}$ *such that* $\{C_t(u_k(\cdot)\}_k$ *is bounded. Then, there exists a control* $u_\infty(\cdot) \in \mathcal{U}$ *such that* $x_{u_k}(\cdot)$ *converges uniformly to* $x_{u_\infty}(\cdot)$ *on* $[0,t]$*, and* $u_k(\cdot)$ *converges to* $u_\infty(\cdot)$ *in the weak-*L^1 *topology.*

The proofs of Proposition 3.1 and Lemma 1 being very classical, they are left to the reader.

3.2. **Continuity of the value function.** Let Ω be an open subset of $(0,\infty) \times \mathbb{R}^n$. We state the following new assumption on our optimal control problem:

(A3) for every $(t,x) \in \Omega$, we have $V(t,x) < \infty$, and for any control $u(\cdot) \in \mathcal{U}$ steering x_0 to x in time t which minimizes $C_t(\cdot)$, the linear operator $dE^{x_0,t}(u(\cdot))$ is surjective.

Proposition 3.2. *Under assumptions* (A1)-(A3)*,* V *is continuous on* Ω.

Proof. As a first step, let us show that V is lower semicontinuous on Ω. Consider a sequence of points $\{(t_k,x_k)\}_k$ in Ω which converges to $(t,x) \in \Omega$ and such that $V(t_k,x_k) \to \lambda$ as $k \to \infty$. We have to prove that $V(t,x) \leq \lambda$. By Proposition 3.1, for each k there exists a control $u_k(\cdot) \in \mathcal{U}$ such that $V(t_k,x_k) = C_{t_k}(u_k(\cdot))$. Hence, $\{C_{t_k}(u_k(\cdot))\}_k$ is bounded (since it converges to λ) and such that $x_k(t_k;x_0,u_k(\cdot)) = x_k \to x$ as $k \to \infty$. We note that, without loss of generality, we can assume that $\{C_{t+1}(u_k(\cdot))\}_k$ is also bounded. This can be easily seen by possibly modifying $u_k(\cdot)$ on $(t_k,t+1)$ for large enough k, that is, taking $u_k(s) := 0$ on $(t_k,t+1)$. Then, by (A1), $|x_{u_k}(s)| \leq (|x_k| + M(t+1-t_k))e^{M(t+1-t_k)}$ for any $s \in (t_k,t+1)$. This implies that $\int_{t_k}^{t+1} L(x_{u_k}(s),u_k(s))ds$ is uniformly bounded in k.

Next, by Lemma 1, we deduce that up to a subsequence, there exists a control $u_\infty(\cdot) \in \mathcal{U}$ such that $x_{u_k}(\cdot)$ converges uniformly to the absolutely continuous function $x_\infty(\ cdot) := x_{u_\infty}$ on $[0,t+1]$, and $u_k(\cdot)$ converges to $u_\infty(\cdot)$ in the weak-L^1 topology. By uniform convergence, $x_\infty(0) = x_0$. Furthermore,

$$|x_\infty(t) - x| \leq |x_\infty(t) - x_k(t)| + |x_k(t) - x_k(t_k)| + |x_k - x|.$$

The first and last terms above clearly tend to zero as $k \to \infty$.. The second one is bounded by $\int_{t_k}^{t} |\dot{x}_k(s)|ds$ which tends to zero since $(\dot{x}_k(\cdot))_k$ is equiabsolutely integrable. Hence, $x_\infty(t) = x$. By the same argument as in the end of the proof of Proposition 3.1, one can show that $\lim_{k\to\infty} C_{t_k}(u_k(\cdot)) = C_t(u(\cdot))$. So, $V(t,x) \leq \lambda$. This proves that V is lower semicontinuous in Ω.

Let us now prove that V is continuous in Ω. Let $(\bar{t},\bar{x}) \in \Omega$. Since $V(\bar{t},\bar{x}) < \infty$, there exists $u(\cdot) \in \mathcal{U}$ such that

$$E^{x_0,\bar{t}}(u(\cdot)) = \bar{x} \quad \text{and} \quad V(\bar{t},\bar{x}) = \int_0^{\bar{t}} L(x_u(s),u(s))ds.$$

Moreover, by assumption (A3), $dE^{x_0,\bar{t}}(u(\cdot))$ is surjective. Hence, there exists n controls $v_1(\cdot), \cdots, v_n(\cdot) \in \mathcal{U}$ such that the linear mapping

$$dE^{x_0,\bar{t}}(u(\cdot)) : \text{span}\{v_1(\cdot), \cdots, v_n(\cdot)\} \longrightarrow \mathbb{R}^n$$

is an isomorphism. In particular, the n vectors

$$dE^{x_0,\bar{t}}(u(\cdot))\,(v_1(\cdot))\,, \cdots, dE^{x_0,\bar{t}}(u(\cdot))\,(v_n(\cdot))$$

are linearly independent. Define $F : (0, \infty) \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$F(t, x, \lambda) := E^{x_0,t}\left(u(\cdot) + \sum_{i=1}^{m} \lambda_i v_i(\cdot)\right) - x,$$

for any triple $(t, x, \lambda) \in (0, \infty) \times \mathbb{R}^n \times \mathbb{R}^n$. Then F is of class C^1 (see Remark 2.2); moreover, $F(\bar{t}, \bar{x}, 0) = 0$ and the differential

$$D_\lambda F(\bar{t}, \bar{x}, 0) = \left(dE^{x_0,\bar{t}}(u(\cdot))\,(v_1(\cdot))\,|\cdots|dE^{x_0,\bar{t}}(u(\cdot))\,(v_n(\cdot))\right)$$

is an isomorphism. Hence, by the Implicit Function Theorem, for some neighborhoods $\mathcal{V}$ of $(\bar{t}, \bar{x})$ in $(0, \infty) \times \mathbb{R}^n$ and $\mathcal{V}'$ of 0_n in $\mathbb{R}^n$, there exists a unique function $g : \mathcal{V} \to \mathbb{R}^n$ of class C^1, with $g(0, 0) = 0$, such that for any $(t, x) \in \mathcal{V}$ and $\lambda \in \mathcal{V}'$,

$$F(t, x, \lambda) = 0 \Longleftrightarrow \lambda = g(t, x).$$

Therefore, for every $(t, x) \in \mathcal{V}$, $u(\cdot) + \sum_{i=1}^{m} g(t, x)_i v_i(\cdot)$ steers x_0 to x in time t. Thus, for any pair $(t, x) \in \mathcal{V}$,

$$V(t, x) \leq C_t\left(u(\cdot) + \sum_{i=1}^{m} g(t, x)_i v_i(\cdot)\right).$$

Letting $(t, x) \to (\bar{t}, \bar{x})$, the last inequality yields that V is upper semicontinuous on Ω. This completes the proof. □

3.3. **Semiconcavity of the value function.** In this section we prove two semiconcavity results for the value function V of problem defined in (2). First we will study system (3) with no drift term, since no additional assumptions will be needed in this case.

Theorem 1. *If assumptions* (A1)-(A3) *hold and* $f_0 \equiv 0$*, then* V *is locally semiconcave on* Ω.

For the proof of Theorem 1 we consider the pseudo-Hamiltonian $\widetilde{H}$: $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ which is defined as follows:

$$\widetilde{H}(x, p, u) := \langle p, f(x, u)\rangle - L(x, u), \quad \forall (x, p, u) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m. \tag{8}$$

We notice that $\widetilde{H}$ is locally Lipschitz in the x variable, of class C^∞ in the p variable, and of class C^2 in the u variable. For each triple $(x, p, u) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$, we denote by $\nabla_p \widetilde{H}(x, p, u)$ and $\nabla_u \widetilde{H}(x, p, u)$ its classical gradients in the p and u variables, and by $\partial_x \widetilde{H}(x, p, u)$ its partial generalized gradient in the x variable. We refer the reader to the books [10, 12] for calculus rules with generalized gradients. We have, for any (x, p, u),

$$\begin{cases} \partial_x \widetilde{H}(x,p,u) & = & \sum_{i=1}^m u_i df_i(x)^* p - \partial_x L(x,u) \\ \nabla_p \widetilde{H}_0(x,p,u) & = & \sum_{i=1}^m u_i f_i(x) \\ \nabla_{u_i} \widetilde{H}_0(x,p,u) & = & \langle p, f_i(x)\rangle - \nabla_{u_i} L(x,u). \end{cases}$$

which implies that for each pair $(x,p) \in \mathbb{R}^n \times \mathbb{R}^n$,

$$\nabla_u \widetilde{H}(x,p,u) = 0 \Longleftrightarrow \nabla_u L(x,u) = (\langle p, f_1(x)\rangle, \cdots, \langle p, f_m(x)\rangle)^*.$$

On the other hand, on account of assumption (A2), there exists a locally Lipschitz map $\Phi : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}^m$ such that, for every $x \in \mathbb{R}^n$, $\Phi(x,\cdot)$ is a diffeomorphism of class C^1 from $\mathbb{R}^m$ into itself, and, for any $v \in \mathbb{R}^m$,

$$\nabla_u L(x,u) = v \quad \Longleftrightarrow \quad u = \Phi(x,v).$$

Let us set, for any pair $(x,p) \in \mathbb{R}^n \times \mathbb{R}^m$,

$$(X(x,p))_i = \langle p, f_i(x)\rangle, \quad \forall i = 1, \cdots, m.$$

Then, for any $(x,p,u) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$,

$$\nabla_u \widetilde{H}(x,p,u) = 0 \quad \Longleftrightarrow \quad u = \Phi(x, X(x,p)). \tag{9}$$

Hence, the Hamiltonian $H(x,p) = \max_{u \in \mathbb{R}^m} \left\{ \widetilde{H}(x,p,u) \right\}$ takes the form

$$\begin{aligned} H(x,p) &= \widetilde{H}(x,p,\Phi(x,X(x,p)) \qquad (10) \\ &= \langle p, f(x,\Phi(x,X(x,p))\rangle - L(x,\Phi(x,X(x,p))) \\ &= \langle \Phi(x,X(x,p)), X(x,p)\rangle_m - L(x,\Phi(x,X(x,p))), \end{aligned}$$

for any pair $(x,p) \in \mathbb{R}^n \times \mathbb{R}^n$. By construction, H is locally Lipschitz in the x variable and of class C^1 in the p variable. In addition, by (9)-(10),

$$\begin{cases} \partial_x H(x,p) & = & \partial_x \widetilde{H}(x,p,\Phi(x,X(x,p))) \\ \nabla_p H(x,p) & = & \nabla_p \widetilde{H}(x,p,\Phi(x,X(x,p))). \end{cases}$$

The version of Pontryagin's Maximum Principle we give below is adapted from a recent fundamental result by Clarke [11] to the problem of interest to this paper.

Proposition 3.3. *Under assumption* (A3)*, if $\bar{u}(\cdot) \in \mathcal{U}$ is a minimizing control steering x_0 to x in time t, then there exists an absolutely continuous arc $p(\cdot) : [0,t] \to \mathbb{R}^n$ such that the pair $(\bar{x}(\cdot) := x_{\bar{u}}(\cdot), p(\cdot))$ is a solution of the Hamiltonian differential inclusion*

$$\begin{cases} \dot{\bar{x}}(s) & = & \nabla_p H(\bar{x}(s), p(s)) \\ -\dot{p}(s) & \in & \partial_x H(\bar{x}(s), p(s)) \end{cases} \tag{11}$$

for almost every $s \in [0,t]$, and such that the function

$$s \longmapsto H(\bar{x}(s), p(s)) \text{ is constant on } [0,t]. \tag{12}$$

In particular, $\bar{x}(\cdot)$ is of class $C^{1,1}$, while $p(\cdot)$ and $\bar{u}(\cdot)$ are Lipschitz on $[0,t]$.

Proof. We will recast our problem in Mayer's form introducing, as usual, an extra state variable. Given a control $u(\cdot) \in \mathcal{U}$, let $y_u(\cdot)$ be the solution of the Cauchy problem

$$\dot{y}(s) = L(x_u(s), u(s)), \quad s \in [0, \infty) \quad \text{a.e.}, \quad y(0) = 0.$$

If we set, for every $(x, y, u) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m$,

$$X(x, y, u) := \begin{pmatrix} f(x, u) \\ L(x, u) \end{pmatrix},$$

then $\bar{u}(\cdot)$ minimizes the terminal cost

$$g(x_u(t), y_u(t)) := y_u(t)$$

over all controls $u(\cdot) \in \mathcal{U}$ and all absolutely continuous arcs $(x_u(\cdot), y_u(\cdot)) : [0, t] \to \mathbb{R}^n \times \mathbb{R}$ satisfying

$$\begin{pmatrix} \dot{x}_u(s) \\ \dot{y}_u(s) \end{pmatrix} = X(x_u(s), y_u(s), u(s)), \quad s \in [0, t] \quad \text{a.e.}$$

and

$$x_u(0) = x_0, \quad y_u(0) = 0, \quad x_u(t) = x.$$

Let us write the above Mayer problem as an optimization problem for a differential inclusion with closed graph. Set, for every $(x, y, z) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m$,

$$F(x, y, z) := \left\{ \begin{pmatrix} \sum_{i=1}^m u_i f_i(x) \\ L(x, u) + \delta \\ u \end{pmatrix} \;\Big|\; u \in \mathbb{R}^m, \delta \geq 0 \right\}.$$

By construction, the multifunction F has closed graph in $(\mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m)^2$, denote it by G. Besides, the trajectory

$$\bar{Z}(\cdot) := \left(\bar{x}(\cdot), \bar{y}(\cdot) := y_{\bar{u}}(\cdot), \bar{z}(\cdot) := \int_0^{\cdot} u(s) ds \right) : [0, t] \longrightarrow \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m$$

minimizes the terminal cost

$$\ell(x(t), y(t), z(t)) := y(t)$$

over all trajectories of the differential inclusion

$$(\dot{x}(s), \dot{y}(s), \dot{z}(s)) \in F\left(x(s), y(s), z(s)\right), \quad s \in [0, t] \quad \text{a.e.} \tag{13}$$

satisfying the constraints

$$x(0) = x_0, \quad y(0) = 0, \quad z(0) = 0, \quad x(t) = x. \tag{14}$$

Our aim is now to apply Theorem 3.4.1 of [11]. Denoting by $|\cdot|_*$ the Euclidean norm in $\mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m$, we claim that, for every $R > 0$, there exists a summable function $k_R : [0, t] \to \mathbb{R}$, bounded below by a positive constant, such that for almost all $s \in [0, t]$, and every $(Z, V) \in G$ satisfying

$$\left| Z - \bar{Z}(s) \right|_* < R \quad \text{and} \quad \left| V - \dot{\bar{Z}}(s) \right|_* < R, \tag{15}$$

one has

$$(\alpha, \beta) \in N_G^P(Z, V) \quad \Longrightarrow \quad |\alpha|_* \leq k_R(s) |\beta|_*$$

(we refer the reader to [11], [12] for the definition of the proximal normal cone $N_G^P(Z, V)$.) For let $r_R > R$ be such that $\bar{x}(s) \in B_{r_R - R}$ for every $s \in [0, t]$, and denote by $\widetilde{K}_R$ a constant $\geq K(r_R)$ such that $|df_i(x)| \leq \widetilde{K}_R$ for all $x \in B_{r_R}$. By (A2)(ii), there exists C_1 such that $q \leq \theta(q)$ for all $r \geq C_1$. Let $s \in [0,t]$ be such that $\dot{\bar{Z}}(s)$ exists, $(Z,V) := ((x,y,z),(v,w,u)) \in G$ satisfies (15), and fix a vector

$$(\alpha,\beta) = (\alpha_1,\alpha_2,\alpha_3,\beta_1,\beta_2,\beta_3) \in N_G^P(Z,V).$$

Note that, necessarily,

$$|x - \bar{x}(s)| < R \quad \text{and} \quad L(x,u) - L(\bar{x}(s),\bar{u}(s)) < R. \tag{16}$$

We need the following result whose proof is given in the Appendix.

Lemma 2. *For every $(Z,V) \in (\mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m) \times (\mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m)$ and every $(\alpha = (\alpha_1,\alpha_2,\alpha_3), \beta = (\beta_1,\beta_2,\beta_3)) \in N_G^L(Z,V)$, we have $\alpha_2 = \alpha_3 = 0$, $\beta_2 \geq 0$. Moreover*

$$\beta_2 < 0 \quad \Longrightarrow \quad \frac{1}{(-\beta_2)}\left(\alpha_1 + \left(\sum_{i=1}^m u_i df_i(x)\right)^* \beta_1\right) \in \partial_x L(x,u), \tag{17}$$

$$\beta_2 < 0 \quad \Longrightarrow \frac{1}{(-\beta_2)}\left(\beta_3 + (\langle \beta_1, f_1(x)\rangle, \cdots, \langle \beta_1, f_m(x)\rangle)^*\right) = \nabla_u L(x,u), \tag{18}$$

and

$$\beta_2 = 0 \quad \Longrightarrow \quad \alpha_1 + \left(\sum_{i=1}^m u_i df_i(x)\right)^* \beta_1 = 0, \tag{19}$$

$$\beta_2 = 0 \quad \Longrightarrow \quad \beta_3 + (\langle \beta_1, f_1(x)\rangle, \cdots, \langle \beta_1, f_m(x)\rangle)^* = 0. \tag{20}$$

Now, suppose $\beta_2 < 0$. Then (16), (17) and (A2)(iii) yield

$$\begin{aligned}
|\alpha|_* &= |\alpha_1| \\
&\leq \left|\alpha_1 + \left(\sum_{i=1}^m u_i df_i(x)\right)^* \beta_1\right| + \left|\left(\sum_{i=1}^m u_i df_i(x)\right)^* \beta_1\right| \\
&\leq K(r_R)|\beta_2|\theta\left(|u|_m\right) + \sqrt{m}\widetilde{K}_R|u|_m|\beta_1| \\
&\leq \widetilde{K}_R\theta\left(|u|_m\right)|\beta_2| + \sqrt{m}\widetilde{K}_R \max\left\{C_1, \theta((|u|_m)\right\}|\beta_1| \\
&\leq \sqrt{m}\widetilde{K}_R \max\left\{C_1, \theta((|u|_m)\right\}\left(|\beta_1| + |\beta_2|\right) \\
&\leq \sqrt{m}\widetilde{K}_R \max\left\{C_1, L(x,u) + c_0\right\}\left(|\beta_1| + |\beta_2| + |\beta_3|_m\right) \\
&\leq \sqrt{m}\widetilde{K}_R \max\left\{C_1, L(\bar{x}(s),\bar{u}(s)) + R + c_0\right\}\left(|\beta_1| + |\beta_2| + |\beta_3|_m\right) \\
&\leq k_R(s)|\beta|_*,
\end{aligned}$$

where

$$k_R(s) := \sqrt{3}\sqrt{m}\widetilde{K}_R \max\left\{C_1, L(\bar{x}(s),\bar{u}(s)) + R + c_0\right\}.$$

On the other hand, if $\beta_2 = 0$, then (16), (19) and (A2)(ii) imply that

$$\begin{aligned} |\alpha|_* &= |\alpha_1| \\ &\le \left| \left(\sum_{i=1}^m u_i df_i(x) \right)^* \beta_1 \right| \\ &\le \sqrt{m}\widetilde{K}_R |u|_m |\beta_1| \\ &\le \sqrt{m}\widetilde{K}_R \max\{C_1, \theta((|u|_m)\} \mid beta_1| \\ &\le k_R(s)|\beta|_*. \end{aligned}$$

Consequently, since $\int_0^t L(\bar{x}(s), \bar{u}(s))ds < \infty$, we have proved that, for every $R > 0$, there exists a summable function $k_R : [0,t] \to \mathbb{R}$, bounded below by a positive constant, such that for almost all $s \in [0,t]$ and for every $(Z,V) \in G$ satisfying (15),

$$(\alpha, \beta) \in N_G^P(Z,V) \quad \Longrightarrow \quad |\alpha|_* \le k_R(s)|\beta|_*.$$

This proves our claim. From the proof of Corollary 3.5.3 in [11], we deduce that the necessary conditions of Theorem 3.4.1 in [11] hold. Therefore, there exist a number $\lambda_0 \in \{0,1\}$ and an absolutely continuous arc $P(\cdot) = (p_1(\cdot), p_2(\cdot), p_3(\cdot)) : [0,t] \to \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m$ such that:

(i) for every $s \in [0,t], (\lambda_0, P(s)) \neq (0,0)$;

(ii) $-p_2(t) = \lambda_0$, $p_3(t) = 0$;

(iii) for almost every $s \in [0,t]$, we have

$$\dot{P}(s) \in \operatorname{co}\left\{ w \mid (w, P(s)) \in N_G^L\left(\bar{Z}(s), \dot{\bar{Z}}(s)\right) \right\};$$

(iv) for almost every $s \in [0,t]$, we have

$$\langle P(s), V \rangle_* \le \left\langle P(s), \dot{\bar{Z}}(s) \right\rangle_*, \quad \forall V \in F\left(\bar{Z}(s)\right)$$

where $\langle \cdot, \cdot \rangle_*$ denotes the Euclidean scalar product in $\mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m$;

(v) there exists a constant h such that

$$\left\langle P(s), \dot{\bar{Z}}(s) \right\rangle_* = h, \quad s \in [0,t] \quad \text{a.e..}$$

Owing to Lemma 2, assertion (iii) can be written as

$$-\dot{p}_1(s) \in \sum_{i=1}^m \bar{u}_i(s) df_i(\bar{x}(s))^* p_1(s) + p_2(s)\partial_x L(\bar{x}(s), \bar{u}(s)), \tag{21}$$

and

$$\dot{p}_2(s) = \dot{p}_3(s) = 0, \tag{22}$$

for almost every $s \in [0,t]$. Hence, (iv) implies that

$$\left\langle p_1(s), \sum_{i=1}^m u_i f_i(x) \right\rangle - \lambda_0 L(\bar{x}(s), u) \le \left\langle p_1(s), \sum_{i=1}^m \bar{u}_i(s) f_i(\bar{x}(s)) \right\rangle - \lambda_0 L(\bar{x}(s), \bar{u}(s)), \tag{23}$$

for almost every $s \in [0,t]$. Notice that, if $\lambda_0 = 0$, then by (ii) and (23), we obtain that $p_2(s) = p_3(s) = 0$ for any $s \in [0,t]$ and

$$\left\langle p_1(s), \sum_{i=1}^m u_i f_i(x) \right\rangle \leq \left\langle p_1(s), \sum_{i=1}^m \bar{u}_i(s) f_i(\bar{x}(s)) \right\rangle,$$

for almost every $s \in [0,t]$ and all $u \in \mathbb{R}^m$. Thus, $H_0(\bar{x}(s), p_1(s)) = 0$ for every $s \in [0,t]$, and, by (21),

$$-\dot{p}_1(s) = \nabla_x H_0(\bar{x}(s), p_1(s), \bar{u}(s)), \quad s \in [0,t] \quad \text{a.e..}$$

Since $P(s) \neq 0$ for every $s \in [0,t]$, this contradicts assumption (A3) in view of Proposition 2.4. Therefore, $\lambda = 1$. This implies that $p_2(s) = -1$ for every $s \in [0,t]$, which yields, in turn, (11) and (12). □

We are now ready to prove Theorem 1.

Proof of Theorem 1. Let $(\bar{t}, \bar{x}) \in \Omega$ and let $\delta > 0$ be such that

$$G := [\bar{t} - \delta, \bar{t} + \delta] \times \overline{B}(\bar{x}, \delta) \subset \Omega.$$

Let $\mathcal{K} \subset [\bar{t} - \delta, \bar{t} + \delta] \times \mathcal{U}$ be the set of all pairs $(t, u(\cdot))$ for which there exists a pair of absolutely continuous arcs $(x_u(\cdot), p_u(\cdot)) : [0,t] \longrightarrow \mathbb{R}^n \times \mathbb{R}^n$ satisfying the following properties:

(i) $x_u(0) = x_0$ and $x_u(t) \in \overline{B}(\bar{x}, \delta)$;
(ii) $(x_u(\cdot), p_u(\cdot))$ is a solution of the Hamiltonian inclusion (11) on $[0,t]$;
(iii) $s \mapsto H(x_u(s), p_u(s))$ is constant on $[0,t]$;
(iv) $u(s) := \Phi(x_u(s), X(x_u(s), p_u(s)))$ for any $s \in [0,t]$;
(v) $u(s) = 0$ fo r all $s \in (t, \infty)$;
(vi) $V(t, x_u(t)) = C_t(u(\cdot))$.

Proposition 3.3 ensures that, for any $(t,x) \in G$, there exists $u(\cdot) \in \mathcal{U}$ such that $(t, u(\cdot)) \in \mathcal{K}$ and $x_u(t) = x$. Moreover, $\mathcal{K}$ has useful compactness properties, as our next result shows.

Lemma 3. *There is a constant $K > 0$ such that, for every $(t, u(\cdot)) \in \mathcal{K}$,*

$$|u(s)|_m < K, \quad \forall s \in [0,t], \tag{24}$$

and

$$|u(s) - u(s')|_m < K|s - s'|, \quad \forall s, s' \in [0,t]. \tag{25}$$

Proof. First of all, since V is continuous on Ω, V is bounded on all compact subsets of Ω. Hence, by (vi), there is $C > 0$ such that $C_t(u(\cdot)) \leq C$ for every $(t, u(\cdot)) \in \mathcal{K}$. Also by assumption (A2), there exists $C_1 > 0$ such that

$r \le \theta(r)$ for all $r \ge C_1$. There fore, for every $(t, u(\cdot)) \in \mathcal{K}$, we have

$$\begin{aligned}
&\|u(\cdot)\|_1 \\
= &\int_{[0,t]\cap\{|u|_m \ge C_1\}} |u(s)|_m ds + \int_{[0,t]\cap\{|u|_m < C_1\}} |u(s)|_m ds \\
\le &\int_{[0,t]\cap\{|u|_m \ge C_1\}} \theta(|u(s)|_m) ds + tC_1 \\
\le &\int_0^t [L(x_u(s), u(s)) + c_0] ds + tC_1 \\
\le &\; C + (c_0 + C_1)t \\
\le &\; C + (c_0 + C_1)(\bar{t} + \delta) =: \tilde{C}. \qquad (26)
\end{aligned}$$

Consequently, recalling assumption (A1) and applying Gronwall's Lemma, we conclude that all trajectories $x_u(\cdot)$ associated with elements $(t, u(\cdot)) \in \mathcal{K}$ are uniformly bounded, that is, there is a compact set $\mathcal{C} \subset \mathbb{R}^n$ such that, for every $(t, u(\cdot)) \in \mathcal{K}$,

$$x_u(s) \in \mathcal{C} \qquad \forall s \in [0, t]. \qquad (27)$$

On the other hand, inequality (26) also says that for every $(t, u(\cdot)) \in \mathcal{K}$, there exists $s_u \in [0, t]$ such that

$$|u(s_u)|_m \le \frac{2\tilde{C}}{\bar{t} - \delta}. \qquad (28)$$

Let $\tilde{M}$ be a positive constant such that

$$|\langle u, \nabla_u L(x, u)\rangle_m) - L(x, u)| \le \tilde{M}, \qquad (29)$$

for any $x \in \mathcal{C}$ and any $u \in \mathbb{R}^m$ satisfying $|u|_m \le 2\tilde{C}/(\bar{t} - \delta)$. By (28)-(29), (iv), and the fact that $f_0 \equiv 0$, we deduce that, for every $(t, u(\cdot)) \in \mathcal{K}$,

$$\begin{aligned}
|H(x_u(s_u)), p_u(s_u))| &= \langle u(s_u), \nabla_u L(x_u(s_u), u(s_u))\rangle - L(x_u(s_u), u(s_u)) \\
&\le \tilde{M}. \qquad (30)
\end{aligned}$$

Let now $\widehat{M}$ be another positive constant such that

$$|L(x, u)| \le \widehat{M},$$

for any $x \in \mathcal{C}$ and any $u \in \mathbb{R}^m$ satisfying $|u|_m \le 1$. We need the following lemma.

Lemma 4. *If we define* $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *by*

$$\forall (x, u) \in \mathbb{R}^n \times \mathbb{R}^m, \quad h(x, u) := \langle u, \nabla_u L(x, u)\rangle_m - L(x, u), \qquad (31)$$

then we have that

$$h(x, u) \ge \frac{\theta(|u|_m)}{|u|_m} - \frac{c_0 + \widehat{M}}{|u|_m} - \widehat{M}, \quad \forall (x, u) \in \mathcal{C} \times \mathbb{R}^m \setminus \{0\}. \qquad (32)$$

Proof. Fix $x \in \mathcal{C}, u \in \mathbb{R}^m \setminus \{0\}$, and set $v := u/|u|_m \in \overline{B_m(0, 1)}$. Define the convex function of class C^2, $\widetilde{L} : [0, \infty) \to \mathbb{R}$, by

$$\forall \alpha \ge 0, \quad \widetilde{L}(\alpha) := L(x, \alpha v),$$

and define $\widetilde{h}:[0,\infty)\to\mathbb{R}$ by

$$\forall \alpha\geq 0,\quad \widetilde{h}(\alpha):=\alpha\widetilde{L}'(\alpha)-\widetilde{L}(\alpha).$$

Then, for every $\alpha\geq 1$,

$$\begin{aligned}
\widetilde{h}(\alpha) &= \widetilde{h}(1)+\int_1^\alpha \widetilde{h}'(r)dr\\
&= \widetilde{h}(1)+\int_1^\alpha r\widetilde{L}\text{"}(r)dr\\
&\geq \widetilde{h}(1)+\int_1^\alpha \widetilde{L}\text{"}(r)dr\\
&= \widetilde{h}(1)+\widetilde{L}'(\alpha)-\widetilde{L}'(1)\\
&\geq \widetilde{h}(1)+\frac{\widetilde{L}(\alpha)-\widetilde{L}(0)}{\alpha}-\widetilde{L}'(1) \quad \text{(by convexity of } \widetilde{L})\\
&= -\widetilde{L}(1)+\frac{\widetilde{L}(\alpha)-\widetilde{L}(0)}{\alpha}\\
&\geq \frac{\theta(\alpha|v|_m)}{\alpha}-\frac{c_0}{\alpha}-\frac{\widehat{M}}{\alpha}-\widehat{M},
\end{aligned}$$

in view of assumption (A2) and the definition of $\widetilde{L}$ and $\widetilde{h}$. Taking $\alpha=|u|_m$, we conclude easily. □

We now return to the proof of Lemma 3. Since

$$\lim_{q\to\infty}\frac{\theta(q)}{q}-\frac{c_0+\widehat{M}}{q}-\widehat{M}=+\infty,$$

there exists $C_2>0$ such that

$$h(x,u)>2\widetilde{M},$$

for any $x\in\mathcal{C}$ and any $u\in\mathbb{R}_m$ satisfying $|u|_m>C_2$. Thus, by (iii) and (30) we deduce that, for every $(t,u(\cdot))\in\mathcal{K}$,

$$|u(s)|_m\leq C_2 \qquad \forall s\in[0,t], \tag{33}$$

which in turn gives (24). Furthermore, we know that for every $(t,u(\cdot))\in\mathcal{K}$,

$$-\dot{p}_u(s)\in\sum_{i=1}^m (u(s))_i df_i(x_u(s))^* p_k(s)-\partial_x L(x_u(s),u_(s)),\quad s\in[0,t] \text{ a.e.} \tag{34}$$

Hence by (27), (33) and Gronwall's Lemma, there exists a constant $M'>0$ such that for every $(t,u(\cdot))\in\mathcal{K}$,

$$|p_u(s)|\leq M'|p_u(0)|,\quad \forall s\in[0,t]. \tag{35}$$

Next, we claim that, for some constant $P>0$,

$$\forall (t,u(\cdot))\in\mathcal{K},\quad |p_u(0)|\leq P. \tag{36}$$

For suppose there exists a sequence $\{(t_k,u_k(\cdot))\}_k\in\mathcal{K}$ such that $\{p_0^k:=p_{u_k}(0)\}_k$ satisfies $|p_0^k|\to\infty$ as $k\to\infty$. Define, for any k,

$$\widehat{p}_k(s):=\frac{p_{u_k}(s)}{|p_0^k|}\qquad \forall s\in[0,t_k].$$

By (34), we have that

$$-\dot{\widehat{p}}_k(s) \in \sum_{i=1}^{m}(u_k(s))_i df_i(x_{u_k}(s))^*\widehat{p}_k(s) - \frac{1}{|p_0^k|}\partial_x L(x_{u_k}(s), u_k(s))$$

for almost every $s \in [0, t_k]$. Since, by (33), $\{u_k(\cdot)\}_k$ is uniformly bounded in L^∞ and, by (27), $x_k(\cdot)$ are all included in the compact set $\mathcal{C}$, the sequence $\{\widehat{p}_k\}_k$ is uniformly bounded and equicontinuous. By the Ascoli-Arzèla Theorem, we deduce that, up to a subsequence, the pair $(x_{u_k}(\cdot), \widehat{p}_k(\cdot))$ converges uniformly to some pair $(x^\infty(\cdot), p^\infty(\cdot))$, and $u_k(\cdot)$ converges to some $u_\infty(\cdot)$ in the weak-L^1 topology. Moreover, $t_k \to t \in [\bar{t} - \delta, \bar{t} + \delta]$ and $C_t(u_\infty(\cdot)) = V(t, x)$ (by the same argument as in the end of the proof of Proposition 3.1). Furthermore, recalling the linear dependence of H_0 with respect to u,

$$\begin{cases} \dot{x}_\infty(s) &= \nabla_p H_0(x_\infty(s), p_\infty(s), u_\infty(s)) \\ -\dot{p}_\infty(s) &= \nabla_x H_0(x_\infty(s), p_\infty(s), u_\infty(s)), \end{cases}$$

for almost every $s \in [0, t]$. Also, by (iii) and (30), we know that, for every k and every $u \in \mathbb{R}^m$,

$$\begin{aligned} \forall s \in [0, t_k], \quad \tilde{H}(x_{u_k}(s), p_{u_k}(s), u) &\leq H(x_{u_k}(s), p_{u_k}(s)) \\ &= H(x_{u_k}(s_{u_k}), p_{u_k}(s_{u_k})) \\ &\leq \tilde{M}. \end{aligned}$$

Hence, for any k, any $u \in \mathbb{R}^m$ such that $|u|_m \leq 1$, and any $s \in [0, t_k]$,

$$\sum_{i=1}^{m} u_i \left\langle \frac{p_k(s)}{|p_0^k|}, f_i(x_k(s)) \right\rangle \leq \frac{\widetilde{M} + \widehat{M}}{|p_0^k|}.$$

Passing to the limit in the above inequality, we obtain

$$\sum_{i=1}^{m} u_i \langle p_\infty(s), f_i(x_\infty(s)) \rangle \leq 0$$

for any $u \in \mathbb{R}^m$ such that $|u|_m \leq 1$. This implies that, for any $s \in [0, t]$, $p_\infty(s)$ is orthogonal to each vector $f_1(x_\infty(s)), \cdots, f_m(x_\infty(s))$. So, invoking Proposition 2.4, we conclude that $u_\infty(\cdot)$ is a singular control for $E^{x_0,t}$, in contrast with assumption (A3). This proves our claim.

Summing up, we have proved that, for every $(t, u(\cdot)) \in \mathcal{K}$,

$$x_u(s) \in \mathcal{C}\,, \quad |u(s)|_m \leq C_2\,, \quad \text{and} \quad |p_u(s)| \leq M'P$$

for every $s \in [0, t]$. By (11), we deduce that, for every $(t, u(\cdot)) \in \mathcal{K}$, the derivatives $\dot{x}_u(\cdot)$ and $\dot{p}_u(\cdot)$ are uniformly bounded on $[0, t]$. Since, by (iv),

$$u(s) = \Phi(x_u(s), X(x_u(s), p_u(s))), \quad \forall s \in [0, t],$$

the uniform Lipschitz estimate (25) easily follows. □

We can now complete the proof of Theorem 1. We denote by $\mathcal{U}_\infty$ the set of $u(\cdot) \in \mathcal{U}$ which satisfy (24) and (25) on $[0, \bar{t} + \delta]$. We shall regard any control $u(\cdot)$, such that $(t, u(\cdot)) \in \mathcal{K}$, as defined on $[0, \bar{t} + \delta]$ which is always the case possibly extending its domain of definition to $[0, \bar{t} + \delta]$ by taking

$u(s) = u(t)$ for every $s \in [t, \bar{t} + \delta]$. We shall equip $\mathcal{U}_\infty$ with the uniform norm $\|\cdot\|_\infty$ on $[0, \bar{t} + \delta]$.

Lemma 5. *There exist $r, R > 0$ such that, for every $(t, u(\cdot)) \in \mathcal{K}$, there exists a mapping of class $C^{1,1}$,*

$$\mathcal{F}_{t,u} : (t - r, t + r) \times B(x := x_u(t), r) \to \mathcal{U}_\infty ,$$

with $d\mathcal{F}_{t,u}$ R-Lipschitz, which satisfies $\mathcal{F}_{t,u}(t, x) = (t, u(\cdot))$ and

$$\forall s \in (t - r, t + r), \forall y \in B(x, r), \quad E^{x_0,s}\left(\mathcal{F}_{t,u}(s, y)\right) = y.$$

Proof. Fix $(\hat{t}, \hat{u}(\cdot)) \in \mathcal{K}$. By assumption (A3), there are n controls

$$v_1^{\hat{u}}(\cdot), \cdots, v_n^{\hat{u}}(\cdot) \in \mathcal{U}$$

such that the linear operator given by

$$\begin{array}{rcl} dE^{x_0,\hat{t}}(\hat{u}(\cdot)) : \operatorname{span}\{v_1^{\hat{u}}(\cdot), \cdots, v_n^{\hat{u}}(\cdot)\} & \longrightarrow & \mathbb{R}^n \\ v(\cdot) & \longmapsto & dE^{x_0,\hat{t}}(\hat{u}(\cdot))(v(\cdot)), \end{array}$$

i s a linear isomorphism. Since the mapping $(t, u(\cdot)) \mapsto E^{x_0,t}(u(\cdot))$ is of class C^1 on $[0, \bar{t} + \delta] \times \mathcal{U}_\infty$, there exists a constant $\rho_{\hat{t},\hat{u}} > 0$ such that

$$\begin{array}{rcl} dE^{x_0,t}(u(\cdot)) : \operatorname{span}\{v_1^{\hat{u}}(\cdot), \cdots, v_n^{\hat{u}}(\cdot)\} & \longrightarrow & \mathbb{R}^n \\ v(\cdot) & \longmapsto & dE^{x_0,t}(u(\cdot))(v(\cdot)), \end{array}$$

is an isomorphism for every $(t, u(\cdot)) \in [0, \bar{t} + \delta] \times \mathcal{U}_\infty$ satisfying

$$|t - \hat{t}| < \rho_{\hat{t},\hat{u}} \quad \text{and} \quad \|u(\cdot) - \hat{u}(\cdot)\|_\infty < \rho_{\hat{t},\hat{u}}. \tag{37}$$

Define

$$\begin{array}{rcl} \mathcal{E}^u_{\hat{t},\hat{u}} : (0, \infty) \times \mathbb{R}^n & \longrightarrow & (0, \infty) \times \mathbb{R}^n \\ (t, \lambda := (\lambda_1, \cdots, \lambda_n)) & \longmapsto & \left(t, E^{x_0,t}\left(u(\cdot) + \sum_{i=1}^n \lambda_i v_i^{\hat{u}}(\cdot)\right)\right), \end{array}$$

Then, for some constant $\mu_{\hat{t},\hat{u}} > 0$,

$$\left|\det\left(d\left(\mathcal{E}^u_{\hat{t},\hat{u}}\right)(t, 0_n)\right)\right| \geq \mu_{\hat{t},\hat{u}}$$

for every $(t, u(\cdot))$ satisfying (37). By the compactness of $\mathcal{K}$, there exist a finite set J and J pairs $(t_j, u_j) \in \mathcal{K} (j = 1, \cdots, J)$ such that

$$\mathcal{K} \subset \bigcup_{j \in J} \left\{(t_j - \rho_{t_j,u_j}, t_j + \rho_{t_j,u_j}, t_j) \times B_X\left(u_j(\cdot), \rho_{t_j,u_j}\right)\right\}.$$

Set

$$\mu := \min_{j \in J}\{\mu_{t_j,u_j}\}.$$

Therefore, by construction, we have for every $(t, u(\cdot)) \in \mathcal{K}$,

$$\left|\det\left(d\left(\mathcal{E}^u_{t_j,u_j}\right)(t, 0_n)\right)\right| \geq \mu,$$

for some $j \in J$. In other terms, for for every $(t, u(\cdot)) \in \mathcal{K}$, there is $j \in J$ such that the linear mapping

$$\begin{array}{rcl} d\left(\mathcal{E}^u_{t_j,u_j}\right)(t, 0_n) : \mathbb{R} \times \mathbb{R}^n & \longrightarrow & \mathbb{R} \times \mathbb{R}^n \\ (s, \lambda) & \longmapsto & \left(s, \sum_{i=1}^n \lambda_i dE^{x_0,t}(u(\cdot))\left(v_i^{u_j}(\cdot)\right)\right), \end{array}$$

is an isomorphism satisfying

$$\left|\det\left(d\left(\mathcal{E}^{u}_{t_j,u_j}\right)(t,0_n)\right)\right| \geq \mu.$$

Then, we apply the Inverse Mapping Theorem to the mapping $\mathcal{E} := \mathcal{E}^{u}_{t_j,u_j}$ obtaining an inverse of the form $\mathcal{E}^{-1}(s,y) = (s,\lambda^{t,u}(s,y))$. Observe that the radius $r>0$ of the cylinder $(t-r,t+r)\times B(x=x_u(t),r)$ which is contained in the image of a neighborhood of $(t,u(\cdot))$ and the Lipschitz constant of $d\mathcal{E}^{-1}$ depend uniquely on the norms of $\mathcal{E}, d\mathcal{E}$ in a neighborhood of $(t,u(\cdot))$ and of $(d\mathcal{E})^{-1}$ in a neighborhood of $\mathcal{E}(t,u(\cdot))$. Since $\mathcal{K}$ is compact and the absolute values of the determinants of $d\mathcal{E}(t,u(\cdot))$ are uniformly bounded below by a positive constant, we conclude taking

$$\mathcal{F}_{t,u}(s,y) = u(\cdot) + \sum_{i=1}^{n} \lambda_i^{t,u}(s,y)v_i^{u_j}(\cdot)$$

for every $(s,y)\in(t-r,t+r)\times B(x=x_u(t),r)$. □

Let us return to the proof of Theorem 1 and consider $(t,x)\in G$ and $u(\cdot)\in\mathcal{U}$ such that $(t,u(\cdot))\in\mathcal{K}$ and $x_u(t)=x$. By Lemma 5, there exists a diffeomorphism of class $C^{1,1}$, $\mathcal{F}_{t,u}:(t-r,t+r)\times B_r(x)\to(0,\infty)\times\mathcal{U}_\infty$, which satisfies $\mathcal{F}_{t,u}(t,x)=(t,u(\cdot))$ and such that $D\mathcal{F}_{t,u}$ is R-Lipschitz. We have, by the definition of V,

$$V(t,x) = C_t\left(\mathcal{F}_{t,u}(t,x)\right), \tag{38}$$

and

$$V(t',y) \leq C_{t'}\left(\mathcal{F}_{t,u}(t',y)\right), \quad \forall (t',y)\in(t-r,t+r)\times B_r(x). \tag{39}$$

Moreover, we have the following lemma that we shall prove in the Appendix.

Lemma 6. *The function $(t,u(\cdot))\in[0,\bar{t}+\delta]\times\mathcal{U}_\infty \mapsto C_t(u(\cdot))$ is locally semiconcave, that is, for each pair $(t,u(\cdot))\in[0,\bar{t}+\delta]\times\mathcal{U}_\infty$ there are constants $\rho, C>0$ such that*

$$\left|C_{t_1}(u^1(\cdot)) - C_{t_2}(u^2(\cdot))\right| \leq C\left(|t_1-t_2| + \|u^1(\cdot)-u^2(\cdot)\|_\infty\right) \tag{40}$$

and

$$\begin{aligned}\mu C_{t_1}(u^1(\cdot)) + (1-\mu)C_{t_2}(u^2(\cdot)) - C_{\mu t_1+(1-\mu)t_2}(\mu u^1(\cdot)+(1-\mu)u^2(\cdot))\\ \leq \mu(1-\mu)C(|t_1-t_2|^2 + \|u^1(\cdot)-u^2(\cdot)\|_\infty^2)\end{aligned} \tag{41}$$

$\forall \mu \imath n[0,1]$, $\forall t_1,t_2\in[0,\bar{t}+\delta]$ with $|t_1-t_2|\leq\rho$, and $\forall u^1(\cdot),u^2(\cdot)\in\mathcal{U}_\infty$ satisfying $\|u^1(\cdot)-u^2(\cdot)\|_\infty\leq\rho$.

To complete the proof of Theorem 1, let $(t_1,x_1),(t_0,x_0)\in G$ be such that $|t_1-t_0|<r, |x_1-x_0|<r$. For $\mu\in[0,1]$, set $t_\mu := \mu t_1+(1-\mu)t_0$ and $x_\mu := \mu x_1+(1-\mu)x_0$. Since $(t_\mu,x_\mu)\in G$, there exists $u_\mu\in\mathcal{U}$ such that $(t_\mu,u_\mu)\in\mathcal{K}$ and $x_\mu = x_{u_\mu}(t_\mu)$. Also, since $(t_1,x_1),(t_0,x_0)\in(t_\mu-r,t_\mu+r)\times B_r(x_\mu)$, in

view of (38)-(39) and Lemma 6, we obtain

$$
\begin{aligned}
&\mu V(t_1,x_1)+(1-\mu)V(t_0,x_0)-V(t_\mu,x_\mu)\\
&\le \mu C_{t_1}\left(\mathcal{F}_{t_\mu,u_\mu}(t_1,x_1)\right)+(1-\mu)C_{t_0}\left(\mathcal{F}_{t_\mu,u_\mu}(t_0,x_0)\right)-C_{t_\mu}\left(\mathcal{F}_{t_\mu,u_\mu}(t_\mu,x_\mu)\right)\\
&\qquad=\mu C_{t_1}\left(\mathcal{F}_{t_\mu,u_\mu}(t_1,x_1)\right)+(1-\mu)C_{t_0}\left(\mathcal{F}_{t_\mu,u_\mu}(t_0,x_0)\right)\\
&\qquad\quad-C_{t_\mu}\Big(\mu\mathcal{F}_{t_\mu,u_\mu}(t_1,x_1)+(1-\mu)\mathcal{F}_{t_\mu,u_\mu}(t_0,x_0)\Big)\\
&\quad+C_{t_\mu}\Big(\mu\mathcal{F}_{t_\mu,u_\mu}(t_1,x_1)+(1-\mu)\mathcal{F}_{t_\mu,u_\mu}(t_0,x_0)\Big)-C_{t_\mu}\Big(\mathcal{F}_{t_\mu,u_\mu}(t_\mu,x_\mu)\Big)\\
&\qquad\le C\mu(1-\mu)\left\|\mathcal{F}_{t_\mu,u_\mu}(t_1,x_1)-\mathcal{F}_{t_\mu,u_\mu}(t_0,x_0)\right\|^2_{(0,\infty)\times\mathcal{U}_\infty}\\
&\quad+C\left\|\mu\mathcal{F}_{t_\mu,u_\mu}(t_1,x_1)+(1-\mu)\mathcal{F}_{t_\mu,u_\mu}(t_0,x_0)-\mathcal{F}_{t_\mu,u_\mu}(t_\mu,x_\mu)\right\|_{(0,\infty)\times\mathcal{U}_\infty}.
\end{aligned}
$$

Since $(t,x)\mapsto\mathcal{F}_{t_\mu,u_\mu}(t,x)$ is $C^{1,1}$, the conclusion follows. □

In order to allow for a drift in (3), we impose the additional assumptions below.

(A3)' (i) For all $r>0$ there exists $K_1(r)$ such that

$$L(x,u)\ge K_1(r)\left(|u|_m^2-1\right)$$

for all $x\in B_r, u\in\mathbb{R}^m$.

(ii) For all $r>0$ there exists $K_2(r)>0$ such that

$$L(x,u)\le K_2(r)\left(|u|_m^2+1\right)$$

for all $x\in B_r, u\in\mathbb{R}^m$.

(iii) For all $r>0$ there exists $K_3(r)>0$ such that

$$|\nabla_u L(x,u)|_m\le K_3(r)\left(|u|_m+1\right)$$

for all $x\in B_r, u\in\mathbb{R}^m$.

(iv) For all $r>0$ there exists $K_4(r)>0$ such that

$$x\in B_r, u\in\mathbb{R}^m, \zeta\in\partial_x L(x,u)\Longrightarrow|\zeta|\le K_4(r)\left(|u|_m^2+1\right).$$

Then, we have the following result.

Theorem 2. *If assumptions* (A1)-(A3) *and* (A3)' *hold, then* V *is locally semiconcave in* Ω*.*

Proof of Theorem 2. As in the proof of Theorem 1, for a fixed $(\bar t,\bar x)\in\Omega$ let $\delta>0$ be such that $G:=[\bar t-\delta,\bar t+\delta]\times\overline{B}(\bar x,\delta)\subset\Omega$. Let $\mathcal{K}\subset[\bar t-\delta,\bar t+\delta]\times\mathcal{U}$ be the the (nonempty) set of all pairs $(t,u(\cdot))$ for which there exists a pair of absolutely continuous arcs $(x(\cdot),p(\cdot)):[0,t]\longrightarrow\mathbb{R}^n\times\mathbb{R}^n$ which satisfies properties (i)-(vi). As in the proof of Theorem 1, we note that, for some $\tilde C>0$ and every $(t,u(\cdot))\in\mathcal{K}$,

$$\|u(\cdot)\|_1=\int_0^t|u(s)|_m\,ds\le\tilde C.$$

Thus, by assumption (A1) and Gronwall's Lemma, there is $r>0$ such that for every $(t,u(\cdot))\in\mathcal{K}$, $|x_u(s)|\le r$ for all $s\in[0,t]$. Hence, recalling (A3)' (i) and repeating the reasoning used to obtain (26), we conclude that, for some constant $\hat{C} > 0$ and every $(t, u(\cdot)) \in \mathcal{K}$,

$$\|u(\cdot)\|_2 = \sqrt{\int_0^t |u(s)|_m^2 ds} \leq \hat{C}. \qquad (42)$$

Let us also observe that, by (A1) and (A3)' (ii)-(iv), for every $t \in (0, \infty)$, the map $C_t : u(\cdot) \in L^2([0,t];\mathbb{R}^m) \longrightarrow C_t(u(\cdot)) \in \mathbb{R}$, is locally Lipschitz in L^2-norm. Therefore, for every $(t, u(\cdot)) \in \mathcal{K}$, the nonsmooth Lagrange multiplier theorem (see [10, 12]) ensures that there exists $\bar{p}_u \in \mathbb{R}^n$ satisfying $\bar{p}_u^* dE^{x_0,t}(u(\cdot)) \in \partial C_t(u(\cdot))$. So,

$$\langle \bar{p}_u, dE^{x_0,t}(u(\cdot))(v(\cdot))\rangle \in \langle \partial C_t(u(\cdot)), v(\cdot)\rangle_{L^2}, \quad \forall v(\cdot) \in L^2([0,t];\mathbb{R}^m). \quad (43)$$

Actually, $\bar{p}_u$ can be related to the adjoint arc $p(\cdot)$ of Proposition 3.3 as follows.

Lemma 7. *There exists an absolutely continuous arc $p(\cdot) : [0,t] \to \mathbb{R}^n$ for which* (11) *is satisfied for almost every $s \in [0,t]$ and such that $p(t) = \bar{p}_u$.*

Proof. We note that, for every $\xi \in \partial C_t(u(\cdot))$, there exists $\eta \in L^\infty([0,t];\mathbb{R}^n)$ satisfying $\eta(s) \in \partial_x L(x_u(s), u(s))$ for every $s \in [0,t]$, such that

$$\langle \xi, v(\cdot)\rangle_{L^2} = \int_0^t \langle \eta(s), dE^{x_0,s}(u(\cdot))(v(\cdot))\rangle + \langle \nabla_u L(x_u(s), u(s)), v(s)\rangle_m ds,$$

for every $v(\cdot) \in L^2([0,t];\mathbb{R}^m)$. Moreover, the first term of the above right-hand side can be written as

$$\begin{aligned}
&\int_0^t \langle \eta(s), dE^{x_0,s}(u(\cdot))(v(\cdot))\rangle ds \\
&= \int_0^t \langle \eta(s), \int_0^s S(s)S(r)^{-1}B(r)v(r)dr\rangle ds \\
&= \iint_0^t \int_0^s \langle \eta(s), S(s)S(r)^{-1}B(r)v(r)\rangle drds \\
&= \int_0^t \int_r^t \langle \eta(s), S(s)S(r)^{-1}B(r)v(r)\rangle dsdr \\
&= \int_0^t \langle \int_s^t \left(S(r)S(s)^{-1}B(s)\right)^* \eta(r)dr, v(s)\rangle_m ds \\
&= \int_0^t \langle B(s)^* \left(S(s)^{-1}\right)^* \int_s^t S(r)^*\eta(r)dr, v(s)\rangle_m ds.
\end{aligned}$$

Furthermore, for every $v(\cdot) \in L^2([0,t];\mathbb{R}^m)$,

$$\begin{aligned}
\langle \bar{p}_u, dE^{x_0,t}(u(\cdot))(v(\cdot))\rangle_{L^2} &= \langle \bar{p}_u, \int_0^t S(t)S(s)^{-1}B(s)v(s)ds\rangle \\
&= \int_0^t \langle \bar{p}_u, \int_0^t S(t)S(s)^{-1}B(s)v(s)\rangle ds \\
&= \int_0^t \langle B(s)^* \left(S(s)^{-1}\right)^* S(t)^*\bar{p}_u, v(s)\rangle_m ds.
\end{aligned}$$

Let us set

$$p(s) := \left(S(s)^{-1}\right)^* S(t)^* \bar{p}_u - \left(S(s)^{-1}\right)^* \int_s^t S(r)^* \eta(r) dr, \quad \forall s \in [0,t].$$

Then,

$$B(s)^* p(s) = \nabla_u L(x_u(s), u(s)), \quad \forall s \in [0,t]\,.$$

This proves that $u(s) = \Phi(x_u(s), X(x_u(s), p(s)))$ for every $s \in [0,t]$. Also,

$$-\dot{p}(s) = A(s)^* p(s) - \eta(s) \in \partial_x H(x_u(s), p(s)), \text{ a.e. } s \in [0,t].$$

This concludes the proof of the lemma. □

We now proceed to show the following result.

Lemma 8. *There exists a constant $K > 0$ such that, for every $(t, u(\cdot)) \in \mathcal{K}$,*

$$|u(s)|_m < K, \quad \forall s \in [0,t],$$

and

$$|u(s) - u(s')|_m < K|s - s'|, \quad \forall s, s' \in [0,t].$$

Proof. To begin, recall that, for every $(t, u(\cdot)) \in \mathcal{K}$ and some constant $R > 0$,

$$\|u(\cdot)\|_1, \|u(\cdot)\|_2, \|x(\cdot)\|_\infty \leq R. \tag{44}$$

Let us now prove that there is a constant $P > 0$ such that, for every $(t, u(\cdot)) \in \mathcal{K}$ and every $\bar{p}_u \in \mathbb{R}^n$ satisfying (43), we have,

$$|\bar{p}_u| \leq P. \tag{45}$$

We argue by contradiction: suppose there exist sequences $(t_k, u_k(\cdot))_k \in \mathcal{K}$ and $(\bar{p}_k)_k \in \mathbb{R}^n$ such that (43) holds for every k, and $|\bar{p}_k| \to \infty$ as $k \to \infty$. Thus, for every k and every $v(\cdot) \in L^2([0,t_k]; \mathbb{R}^m)$,

$$\left\langle \frac{\bar{p}_k}{|\bar{p}_k|}, dE^{x_0,t_k}(u_k(\cdot))(v(\cdot)) \right\rangle \in \frac{1}{|\bar{p}_k|} \langle \partial C_{t_k}(u_k(\cdot)), v(\cdot) \rangle_{L^2}, .$$

Recalling that the map $(t, u(\cdot)) \mapsto dE^{x_0,t}(u(\cdot))$ is continuous in the weak L^1-topology (see [24]), in the limit as $k \to \infty$ we obtain that there exist $\bar{p} \in \mathbb{R}^n$ with $|\bar{p}| = 1$ and a pair $(t, u(\cdot)) \in \mathcal{K}$, with $u \in L^2([0,t]; \mathbb{R}^m)$, such that $\langle \bar{p}, dE^{x_0,t}(u(\cdot))(v(\cdot)) \rangle = 0$ for every $v(\cdot) \in L^2([0,t]; \mathbb{R}^m)$. Since $u(\cdot)$ minimizes $C_t(\cdot)$, we have obtained a contradiction. Consequently, there is $P > 0$ such that (45) holds for every $(t, u(\cdot)) \in \mathcal{K}$ and every $\bar{p}_u \in \mathbb{R}^n$ satisfying (43). Let M be a positive constant such that $\|df_i(x)^*\| \leq M$ for every $x \in B_r$ and every $i = 0, \cdots, m$. Recalling Lemma 7 we have that, for every $(t, u(\cdot)) \in \mathcal{K}$ and $\bar{p}_u \in \mathbb{R}^n$ satisfying (43), there is an absolutely continuous arc $p_u(\cdot) : [0,t] \to \mathbb{R}^n$ satisfying $p_u(t) = \bar{p}_u$ and

$$-\dot{p}_u(s) \in \sum_{i=1}^m (u(s))_i df_i(x_u(s))^* p_u(s) - \partial_x L(x_u(s), u(s)), \quad \text{a.e. } s \in [0,t].$$

By (44) and assumption (A3)' (iv), we deduce that, for every $(t, u(\cdot)) \in \mathcal{K}$,

$$|\dot{p}_u(s)| \leq R\sqrt{m}|u(s)|_m |p_u(s)| + K_4(R)\left(|u(s)|_m + 1\right),\ quad \forall s \in [0,t].$$

Owing to (45) and Lemma 7 the above estimate implies that, for all $(t, u(\cdot)) \in \mathcal{K}$ and $s \in [0, t]$,

$$\begin{aligned} & |p_u(s)| \\ \leq\ & |p_u(t)| + \int_s^t K_4(R)\left(|u(r)|_m + 1\right) dr + \int_s^t R\sqrt{m}|u(r)|_m |p_u(r)| dr \\ \leq\ & P + K_4(R)\left(C + (\bar{t} + \delta)\right) + \int_s^t R\sqrt{m}|u(r)|_m |p_u(r)| dr \\ :=\ & Q + \int_s^t R\sqrt{m}|u(r)|_m |p_u(r)| dr. \end{aligned}$$

By Gronwall's Lemma, we deduce that, for every $s \in [0, t]$,

$$|p_u(s)| \leq Q \exp\left(\int_s^t R\sqrt{m}|u(r)|_m dr\right) \leq Q e^{R\sqrt{m}C}.$$

Then, we conclude the reasoning arguing as in the proof of Lemma 3. $\square$

The proof of Theorem 2 is now complete. $\square$

4. Properties of optimal trajectories

In this section, we will derive a few results relating the differentiability of V at a given point with the uniqueness of the optimal control at the same point. A similar analysis is known in the literature for finite horizon problems with an initial cost and for exit time problems, see [8]. Here, we are interested in problems with finite horizon and fixed initial condition. Although our results will be analogous to those obtained in the aforementioned situations, the technique of proof is—in the present context—totally different, yet still based on the semiconcavity of V. Throught this section, we assume that $f_0 \equiv 0$.

To begin, let us recall that the Hamiltonian $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is defined by

$$H(x, p) := \max_{u \in \mathbb{R}^m} \left\{ \widetilde{H}(x, p, u) \right\} = \max_{u \in \mathbb{R}^m} \left\{ \langle p, f(x, u) \rangle - L(x, u) \right\}$$

for any $(x, p) \in \mathbb{R}^\times \mathbb{R}^n$. Let us also recall that, as well-known, V is a viscosity solution of the Hamilton-Jacobi equation

$$\frac{\partial V}{\partial t} + H(x, D_x V) = 0 \quad \text{in} \quad \Omega. \tag{46}$$

We will need the following assumptions.

(A4) The Lagrangian L is of class $C^{1,1}_{loc}$ in the x variable.

(A5) For every $(t, x) \in \Omega$ and for any control $u(\cdot)$ steering x_0 to x in time t which minimizes $C_t(\cdot)$, the trajectory $x_u(\cdot)$ remains in Ω for $s \in (0, t]$. Moreover, for every (t_1, x_1) and (t_2, x_2) in Ω, and for any control $u(\cdot)$ steering x_0 to x_2 in time t_2 which minimizes the cost functional with $x_u(t_1) = x_1$, we have that $dE^{x_1, t_2 - t_1}(u(t_1 + \cdot))$ is surjective.

Let $(t,x) \in \Omega$ and $u(\cdot) \in \mathcal{U}$ be a control steering x_0 to x in time t which minimizes $C_t(\cdot)$. Under assumptions (A1)-(A4), Proposition 3.3 implies that there exists an absolutely continuous arc $p(\cdot) : [0,t] \to \mathbb{R}^n$ such that the pair $(x(\cdot) := x_u(\cdot), p(\cdot))$ is a solution of the Hamiltonian differential system

$$\left\{ \begin{array}{rcl} \dot{x}(s) & = & \nabla_p H(x(s), p(s)) \\ -\dot{p}(s) & = & \nabla_x H(x(s), p(s)), \end{array} \right. \tag{47}$$

for almost every $s \in [0,t]$. Note that, in view of assumption (A4), $(x(\cdot), p(\cdot))$ is the solution of a locally Lipschitz differential equation. Hence, it is of class $C^{1,1}$. Also, observe that the above Hamiltonian system can be rewritten as

$$\left\{ \begin{array}{rcl} \dot{x}(s) & = & \nabla_p \widetilde{H}(x(s), p(s), u(s)) \\ -\dot{p}(s) & = & \nabla_x \widetilde{H}(x(s), p(s), u(s)), \end{array} \right. \tag{48}$$

for almost every $s \in [0,t]$, where

$$u(s) = \Phi(x(s), X(x(s), p(s))) \, . \tag{49}$$

Lemma 9. *Under assumptions* (A1)-(A5)*, for every $(t,x) \in \Omega$ and every control $u(\cdot) \in \mathcal{U}$ steering x_0 to x in time t such that $V(t,x) = C_t(u(\cdot))$, we have that, for every $s \in (0,t)$, $u(\cdot)$ is the unique control in $L^1([0,s];\mathbb{R}^m)$ steering x_0 to $x_u(s)$ and such that $V(s, x_u(s)) = C_s(u(\cdot))$.*

Proof. We argue by contradiction. Let $\bar{s} \in (0,t)$ and and let $u'(\cdot)$ be a control steering x_0 to $x_u(\bar{s})$ in time $\bar{s}$ which is minimizing and such that $u'(\cdot) \neq u(\cdot)$ on $[0,\bar{s}]$. Then, we have two different controls which minimize the cost between x_0 and x in time t. By assumption (A4), there are two absolutely continuous arcs $p(\cdot), p'(\cdot) : [0,t] \to \mathbb{R}^n$ such that the triples $(x(\cdot), u(\cdot), p(\cdot))$ and $(x'(\cdot) := x_{u'}(\cdot), u'(\cdot), p'(\cdot))$ satisfy the Hamiltonian system (48). In addition, since $u(\cdot)$ and $u'(\cdot)$ do not coincide on $[0,t]$, then $p'(s) \neq p(s)$ for any $s \in [\bar{s},t]$. On the other hand, since both trajectories $x(\cdot)$ and $x'(\cdot)$ and both controls $u(\cdot)$ and $u'(\cdot)$ coincide on $[\bar{s},t]$, we have that, for almost every $s \in [\bar{s},t]$,

$$\left\{ \begin{array}{rcl} \dot{x}(s) & = & \sum_{i=1}^m u_i(s) f_i(x(s)) \\ -\left(\dot{p'}(s) - \dot{p}(s)\right) & = & \sum_{i=1}^m u_i(s) df_i(x(s))^* \left(p'(s) - p(s)\right) . \end{array} \right.$$

Furthermore, by (49), we have that, for any $s \in [\bar{s},t]$,

$$\langle p'(s) - p(s), f(x(s), u) \rangle = 0, \quad \forall u \in \mathbb{R}^m.$$

In other terms, $u(\cdot)$ is a singular control for $E^{x(\bar{s}), t-\bar{s}}$. This contradicts assumption (A5) and proves the result. □

We need now the following lemma.

Lemma 10. *For every $(t,x) \in \Omega$ and every $(\zeta_t, \zeta_x) \in D^*V(t,x)$, there exists a solution $(x(\cdot), p(\cdot))$ of the Hamiltonian system* (47) *such that the corresponding control given by* (49) *satisfies $p(t) = \zeta_x$ and minimizes $C_t(\cdot)$.*

Proof. Let us first prove that for every $(t,x) \in \Omega$ at which V is differentiable, there exists a solution $(x(\cdot), p(\cdot))$ of the hamiltonian system (47) such that the corresponding control given by (49) is minimizing and $p(t) = \nabla_x V(t,x)$. Let $u(\cdot) \in \mathcal{U}$ be a control steering x_0 to x such that $V(t,x) = C_t(u(\cdot))$. Since V is differentiable at (t,x), there exists a function $\phi : \mathbb{R}^n \to \mathbb{R}$ of class C^1

with $\nabla_x V(t,x) = \nabla\phi(x)$ and such that $y \mapsto V(t,y) - \phi(y)$ attains a local minimum at x. Thus, there exists a neighbourhood $\mathcal{V}$ of $u(\cdot)$, contained in $\mathcal{U}$, such that $V(t,x) \le V(t,x_v(t)) - \phi(x_v(t)) + \phi(x)$ for every control $v(\cdot) \in \mathcal{V}$. Moreover, the very definition of V yields $V(t,x_v(t)) \le C_t(v(\cdot))$. Therefore, $V(t,x) \le C_t(v(\cdot)) - \phi(x_v(t)) + \phi(x)$, for every control $v(\cdot) \in \mathcal{V}$. In particular, $u(\cdot)$ is a solution of the minimization problem

$$\min_{v(\cdot)\in\mathcal{V}} \left\{ \int_0^t L(x_v(s), v(s))ds - \phi(x_v(t)) + \phi(x) \right\}.$$

By the Pontryagin Maximum Principle, we deduce the existence of an absolutely continuous arc $p(\cdot)$ satisfying (47) such that $p(t) = \nabla_x V(t,x)$. Now, let (t,x) be any point in Ω and $(\zeta_t, \zeta_x) \in D^*V(t,x)$. By definition, there exists a sequence $\{(t_k, x_k)\}_k$ in Ω such that V is differentiable at (t_k, x_k) and $\lim \nabla V(t_k, x_k) = (\zeta_t, \zeta_x)$. For each k, we denote by $u_k(\cdot)$ a minimizing control joining x_0 to x_k. On account of the first part of this proof, we know that, for each k, there exists an adjoint arc $p_k(\cdot) : [0, t_k] \to \mathbb{R}^n$ satisfying $p_k(t_k) = \zeta_{x_k}$. Passing to the limit as $k \to \infty$ gives the result. □

Theorem 3. *Under assumptions* (A1)-(A5)*, for every* $(t,x) \in \Omega$ *and every* $u(\cdot) \in \mathcal{U}$ *steering* x_0 *to* x *in time* t *such that* $V(t,x) = C_t(u(\cdot))$*, the function* V *is differentiable at* $(s, x_u(s))$ *for every* $s \in (0,t)$*. Moreover,* $u(\cdot)$ *is the unique control in* $L^1([0,s];\mathbb{R}^m)$ *steering* x_0 *to* $x_u(s)$ *and such that* $V(s, x_u(s)) = C_s(u(\cdot))$*. Furthermore, if* $p(\cdot) : [0,s] \to \mathbb{R}^n \setminus \{0\}$ *satisfies* (47) *on* $[0,s]$*, then* $p(s) = \nabla_x V(s, x_u(s))$*.*

Proof. We argue by contradiction. If V is not differentiable at $(s, x(s))$ for some $s \in (0,t)$, then, by semiconcavity, V possesses at least two distinct limiting subgradients $(\zeta_s^1, \zeta_{x(s)}^1), (\zeta_s^1, \zeta_{x(s)}^1)$ at the point $(s, x(s))$. Since V is a viscosity solution of (46), we have that

$$\zeta_s^i + H\left(x(s), \zeta_{x(s)}^i\right) = 0, \quad \forall i = 1, 2.$$

Thus, necessarily, $\zeta_{x(s)}^1 \neq \zeta_{x(s)}^2$, and the above lemma yields the first statement. The second assertion follows from Lemma 9. Lemma 10 ensures the existence of $p(\cdot) : [0,s] \to \mathbb{R}^n$ such that (47) is satisfied on $[0,s]$, and $p(s) = \nabla_x V(s, x_u(s))$. If there is another adjoint arc for which (47) is satisfied on $[0,s]$, then, as in the proof of Lemma 9, we deduce that $u(\cdot)$ is singular for $E^{x_0,s}$, which contradicts assumption (A3). □

Theorem 4. *Under assumptions* (A1)-(A5)*, for every* $(t,x) \in \Omega$*,* V *is differentiable at* (t,x) *if and only if there is a unique* $u(\cdot) \in L^1([0,t];\mathbb{R}^m)$ *steering* x_0 *to* x *in time* t *such that* $V(t,x) = C_t(u(\cdot))$*.*

Proof. Assume that V is differentiable at $(t,x) \in \Omega$, and suppose there are two distinct controls $u_1(\cdot) \neq u_2(\cdot)$ steering x_0 to x in time t such that $V(t,x) = C_t(u_i(\cdot))$ for $i = 1,2$. Then there exist two arcs $p_1(\cdot), p_2(\cdot) : [0,t] \to \mathbb{R}^n \setminus \{0\}$ satisfying (47) on $[0,t]$. Both vectors $p_1(t)$ and $p_2(t)$ being equal to $\nabla_x V(t,x)$, we have that $p_1(t) = p_2(t)$. This implies that $u_1(\cdot) = u_2(\cdot)$.

Assume now that there is a unique $u(\cdot) \in L^1([0,t];\mathbb{R}^m)$ steering x_0 to x in time t such that $V(t,x) = C_t(u(\cdot))$. If V is not differentiable at (t,x), then it possesses at least two distinct elements in $D_x^*V(t,x)$. The conclusion easily follows. □

In view of the above results, one concludes that, for every $(t,x) \in \Omega$, we have a one-to-one mapping between the set of minimizing controls steering x_0 to x in time t, and the set of limiting gradients $D^*V(t,x)$.

5. Strongly bracket generating control systems

Throught this section, we assume that $f_0 \equiv 0$ and that $f_1, \cdots, f_m$ are smooth vector fields satisfying (A1). We recall that if X, Y are two smooth vector fields on $\mathbb{R}^n$, then the Lie bracket $[X,Y]$ at $x \in \mathbb{R}^n$ is defined by

$$[X,Y](x) := dX(x)(Y(x)) - dY(x)(X(x)).$$

The control system is said to be *strongly bracket generating* on $\mathbb{R}^n$ if the following assumption is satisfied:

(*A6*) For every $x \in \mathbb{R}^n$ and every $(v_1, \cdots, v_m) \in \mathbb{R}^m \setminus \{0_m\}$,

$$\text{span}\big\{f_j(x) \bigm| 1 \le j \le m\big\} + \text{span}\left\{\left[\sum_{i=1}^m v_i f_i, f_j\right](x) \biggm| 1 \le j \le m\right\} = \mathbb{R}^n.$$

Whenever $f_1(x), \cdots, f_m(x)$ are linearly independent for any $x \in \mathbb{R}^n$, the family $f_1, \cdots, f_m$ defines a so-called nonsingular distribution Δ on $\mathbb{R}^n$, i.e.,

$$\Delta(x) := \text{span}\left\{f_1(x), \cdots, f_m(x)\right\}, \quad \forall x \in \mathbb{R}^n.$$

If vector fields $f_1, \cdots, f_m$ satisfy assumption (A6), then the corresponding distribution Δ is said to be fat (cf. [13], [15], [25]). It can be proved that, given a pair (m,n), there may be no fat distributions of rank m in $\mathbb{R}^n$, see [15], [17]. We have the following well-known result.

Proposition 5.1. *If assumption* (A6) *hold, then, for any $x \in \mathbb{R}^n$ and any $t > 0$, any control $u(\cdot) \in \mathcal{U}$ which is not identically zero on the interval* $[0,t]$ *is nonsingular for $E^{x,t}$.*

Proof. We use the characterization of singular controls given by Proposition 2.4. Let us argue by contradiction and assume that there exists a solution $(x(\cdot), p(\cdot), u(\cdot))$ of (6)-(7) on $[0,t]$. Then, for every $i = 1, \ldots, m$,

$$\langle p(s), f_i(x(s))\rangle = 0, \quad \forall s \in [0,t].$$

Differentiating the above equality and using (6) yields, for a.e. $s \in [0,t]$,

$$\begin{aligned}
0 = \frac{d}{ds}\left(\langle p(s), f_i(x(s))\right) &= \langle \dot p(s), f_i(x(s))\rangle + \langle p(s), df_i(x(s))(\dot x(s))\rangle \\
= -\langle \sum_{j=1}^m u_j(s) df_j(x(s))^* p(s), f_i(x(s))\rangle &+ \langle p(s), \sum_{j=1}^m u_j(s) df_i(x(s))(f_j(x(s)))\rangle \\
= \sum_{j=1}^m u_j(s)\left(-\langle p(s), df_j(x(s))(f_i(x(s))\rangle + \langle p(s), df_i(x(s))(f_j(x(s)))\rangle\right) \\
= \sum_{j=1}^m u_j(s)\langle p(s), [f_i, f_j](x(s))\rangle &= \left\langle p(s), \left[f_i, \sum_{j=1}^m f_j\right](x(s))\right\rangle.
\end{aligned}$$

Fix $\bar{s} \in [0,t]$ such that $p(\cdot)$ and $x(\cdot)$ are differentiable at $\bar{s}$ and such that $u(\bar{s}) \neq 0$, and set $v := u(\bar{s}) \in \mathbb{R}^m$. We obtain

$$\left\langle p(\bar{s}), \left[\sum_{j=1}^{m} f_j, f_i\right](x(\bar{s}))\right\rangle = 0,$$

for every $i = 1, \ldots, m$, which contradicts assumption (A6). □

Furthermore, the Chow-Rashevsky Theorem (see [3], [9], [16]) asserts that if system (3) has no drift, then under assumption (A6), for any pair $(x, y) \in R^n$ and any $t > 0$ there exists some control $u(\cdot) \in \mathcal{U}$ such that $x(t; x, u(\cdot)) = y$. Hence, we obtain the following result as a corollary.

Corollary 5.2. *If assumptions* (A1)-(A2) *and* (A6) *hold and if* $f_0 \equiv 0$, *then the value function* V *is continuous on* $(0, \infty) \times \mathbb{R}^n$ *and semiconcave on* $(0, \infty) \times (\mathbb{R}^n \setminus \{x_0\})$.

Example 5.3. The nonholonomic integrator of Example 2.5 is strongly bracket generating. Indeed, it is easy to check that

$$[f_1, f_2](x) = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}, \qquad \forall x \in \mathbb{R}^3$$

So, the three vectors $f_1(x), f_2(x), [f_1, f_2](x)$ form a basis of $\mathbb{R}^3$.

6. The sub-Riemannian distance

Throught this section, we assume that $f_0 \equiv 0$, that $f_1, \ldots, f_m$ are smooth vector fields satisfying (A1), and that the following assumption is satisfied:

(*A7*) for every $x \in \mathbb{R}^n$, $f_1(x), \ldots, f_m(x)$ are linearly independent.

For any family $\mathcal{F}$ of smooth vector fields (*i.e.*, $\mathcal{F} \subset C^\infty(\mathbb{R}^n, \mathbb{R}^n)$), we denote by $\mathrm{Lie}(\mathcal{F})$ the Lie algebra of all vector fields generated by $\mathcal{F}$, that is, the smallest vector subspace S of $C^\infty(\mathbb{R}^n, \mathbb{R}^n)$ satisfying

$$[f, g] \in S, \quad \forall f \in \mathcal{F}, \quad \forall g \in S.$$

For any point $x \in \mathbb{R}^n$, $\mathrm{Lie}(\mathcal{F})(x)$ denotes the set of all vectors $f(x) \in \mathbb{R}^n$ with $f \in \mathcal{F}$. We say that $\{f_1, \ldots, f_m\}$ satisfy *Hörmander's bracket generating condition* on $\mathbb{R}^n$ if the following assumption is satisfied:

(*A8*) for every $x \in \mathbb{R}^n$, $\quad \mathrm{Lie}\,\{f_1, \ldots, f_m\}(x) = \mathbb{R}^n$.

Hereafter, we assume that (A7)-(A8) are satisfied and set

$$\Delta(x) := \mathrm{span}\,\{f_1(x), \cdots, f_m(x)\}, \quad \forall x \in \mathbb{R}^n.$$

According to the classical Chow-Rashevsky theorem (see [3], [9], [16]), the control system (3) is small time locally controllable at any point of $\mathbb{R}^n$.

Let $g(\cdot, \cdot)$ be a Riemannian metric on $\mathbb{R}^n$, associated with a smooth positive definite symmetric matrix $Q(x)$, that is

$$g_x(v, w) = \langle Q(x)v, w\rangle, \quad \forall x \in \mathbb{R}^n, \quad \forall v, w \in R^n.$$

The pair (Δ, g) defines what we call a *sub-Riemannian distribution* of rank m on $\mathbb{R}^n$, that is a smooth distribution of constant rank m which satisfies Hörmander's bracket generating condition on $\mathbb{R}^n$. We refer the reader to [14], [15] for an extensive study of sub-Riemannian distributions.

An absolutely continuous arc $\gamma : [0,1] \to \mathbb{R}^n$ is said to be *horizontal* if $\dot{\gamma}(t) \in \Delta(x)$ for a.e. $t \in [0,1]$. For any given $x_0 \in \mathbb{R}^n$, we denote by $\mathcal{H}_{x_0}$ the set of horizontal arcs satisfying $\gamma(0) = x_0$. For any $x \in \mathbb{R}^n$, the *sub-Riemannian distance* between x_0 and x, denoted by $d_{SR}(x_0, \cdot)$, is defined as the minimal length of an horizontal arc joining x_0 to x, that is,

$$d_{SR}(x_0, x) := \inf \left\{ \text{length}(\gamma) \;\middle|\; \gamma \in \mathcal{H}_{x_0}\,,\ \gamma(1) = x \right\}$$

where

$$\text{length}(\gamma) := \int_0^1 \sqrt{g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))}dt\,.$$

We note that, since system (3) is small time locally controllable, it is easy to prove that the map $x \mapsto d_{SR}(x_0, x)$ is well defined and continuous in $\mathbb{R}^n$. Furthermore, denoting by $F(x)$ the set $\{v \in \Delta(x) \mid g_x(v,v) \le 1\}$ for any $x \in \mathbb{R}^n$, it is straightforward to show that $d_{SR}(x_0, \cdot)$ coincides with the minimum time $T_{x_0}(x)$ needed to steer x to x_0 along a trajectory of the differential inclusion

$$\dot{x}(t) \in F(x(t)). \tag{50}$$

For any $x \in \mathbb{R}^n$, the *sub-Riemannian energy* between x_0 and x, denoted by $e_{SR}(x_0, x)$, is defined by

$$e_{SR}(x_0, x) := \inf \left\{ \int_0^1 g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))dt \;\middle|\; \gamma \in \mathcal{H}_{x_0} \text{ s.t. } \gamma(1) = x \right\}.$$

The following lemma is fundamental. For sake of completeness, we provide its easy proof.

Lemma 11. *For every $x \in \mathbb{R}^n$, $d_{SR}(x_0, x) = \sqrt{e_{SR}(x_0, x)}$.*

Proof. First, we observe that, for every horizontal arc $\gamma(\cdot)$ satisfying $\gamma(0) = x_0$ and $\gamma(1) = x$, the Cauchy-Schwarz inequality yields

$$\left(\int_0^1 \sqrt{g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))}dt \right)^2 \le \int_0^1 g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))dt\,.$$

Taking the infimum over γ, the above inequality implies $d_{SR}(x_0, x)^2 \le e_{x_0}(x)$ for every $x \in \mathbb{R}^n$. On the other hand, for all $x \in \mathbb{R}^n$ and every $\epsilon > 0$, there exists a horizontal curve $\gamma \in \mathcal{H}_{x_0}$, with $\gamma(1) = x$, such that

$$\text{length}(\gamma) = \int_0^1 \sqrt{g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))}dt \le d_{SR}(x_0, x) + \epsilon.$$

Define $\phi : [0,1] \to [0, \text{length}(\gamma)]$ by

$$\phi(s) := \int_0^s \sqrt{g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))}dt, \quad \forall s \in [0,1].$$

Note that ϕ is strictly increasing, hence one-to-one from $[0,1]$ to $[0, \text{length}(\gamma)]$. Set $\psi := \phi^{-1} : [0, \text{length}(\gamma)] \to [0,1]$, and define $\xi : [0,1] \to \mathbb{R}^n$ by $\xi(t) := \gamma(\psi(\text{length}(\gamma)t))$ for any $t \in [0,1]$. It clear that $\xi \in \mathcal{H}_{x_0}$ and $\xi(1) = x$. Moreover, one can easily check that

$$g_{\xi(t)}\left(\dot{\xi}(t), \dot{\xi}(t)\right) = \text{length}(\gamma)^2 \quad \text{for a.e. } t \in [0,1]\,.$$

Consequently,

$$e_{SR}(x_0,x) \le \int_0^1 g_{\xi(t)}\left(\dot{\xi}(t),\dot{\xi}(t)\right)dt = \text{length}(\gamma)^2 = \left(d_{SR}(x_0,x)+\epsilon\right)^2.$$

Letting ϵ tend to 0 completes the proof of the lemma. □

Since $f_1,\cdots,f_m$ satisfy (A7), for every $\gamma \in \mathcal{H}_{x_0}$, there is a unique control $u(\cdot) \in L^1([0,1];\mathbb{R}^m)$ such that

$$\dot{\gamma}(t) = \sum_{i=1}^m u_i(t) f_i(\gamma(t)) \quad \text{for a.e. } t \in [0,1].$$

Hence,

$$g_{\gamma(t)}(\dot{\gamma}(t),\dot{\gamma}(t)) = \sum_{i=1}^m u_i(t)^2 g_{\gamma(t)}(f_i(\gamma(t)), f_i(\gamma(t))) + \sum_{i,j=1, i\neq j}^m u_i(t)u_j(t) g_{\gamma(t)}(f_i(\gamma(t)), f_i(\gamma(t))).$$

Define the Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ by

$$L(x,u) := \sum_{i=1}^m u_i^2 g_x(f_i(x), f_i(x)) + \sum_{i,j=1,i\neq j}^m u_i u_j g_x(f_i(x), f_j(x)).$$

Under the above assumptions, it is easy to show that L satisfies (A2). Moreover, by Lemma 11,

$$d_{SR}(x_0,x)^2 = \inf\left\{\int_0^1 L(x_u(s),u(s))ds \mid u(\cdot) \in \mathcal{U} \text{ s.t. } x_u(1) = x\right\},$$

for every $x \in \mathbb{R}^n$. Furthermore, by (A8) together with Chow-Rashevsky's Theorem, $d_{SR}(x_0,x)$ is finite for every $x \in \mathbb{R}^n$. Therefore, Proposition 3.1 implies that, for every $x \in \mathbb{R}^n$, there is a minimizing control $u(\cdot) \in L^1([0,1];\mathbb{R}^m)$ steering x_0 to x such that

$$d_{SR}(x_0,x)^2 = \int_0^1 L(x_u(s),u(s))ds.$$

We now need the following assumption.

($A9$) Every minimizing control steering x_0 to $x \neq x_0$ is regular.

Note that, under assumption (A9), Pontryagin's Maximum Principle ensures that all minimizing controls are smooth. Moreover, applying Theorem 1 and Chow-Rashevsky's Theorem, we obtain the following result.

Theorem 5. *Let $x_0 \in \mathbb{R}^n$. If $f_0 \equiv 0$ and assumptions* (A1), (A7)-(A9) *hold, then the function $d_{SR}(x_0,\cdot) = T_{x_0}$ is continuous on $\mathbb{R}^n$ and locally semiconcave on $\mathbb{R}^n \setminus \{x_0\}$.*

Hence, in the special case of fat distributions, Corollary 5.2 yields the result below.

Corollary 6.1. *Let $x_0 \in \mathbb{R}^n$ and (Δ, g) be a sub-Riemannian distribution on $\mathbb{R}^n$ such that Δ is fat on $\mathbb{R}^n$. Then the function $d_{SR}(x_0,\cdot) = T_{x_0}$ is continuous on $\mathbb{R}^n$ and locally semiconcave on $\mathbb{R}^n \setminus \{x_0\}$.*

Theorem 5 provides useful information on the regularity of the sub-Riemannian distance function in the smooth case. For example, using classical results on the structure of the singular sets of locally semiconcave functions (see [2], [8]), we can deduce that the Hausdorff dimension of the set of points at which $d_{SR}(x_0,\cdot)$ fails to be differentiable does not exceed $n-1$. Furthermore, the semiconcavity of the sub-Riemannian distance is fundamental to study the stabilization problem for nonholonomic distributions, see [23]. Finally, we observe that, since the Lagrangian associated to (Δ, g) is smooth in the x variable, one can show that the sub-Riemannian distance function can be written locally as an infimum of "uniformly" smooth functions. Such an approach leads to further regularity results, see [21] and [22].

Appendix

6.1. **Proof of Lemma 2.** Let $(Z,V)=((x,y,z),(v,w,u))\in(\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}^m)^2$ and $(\alpha,\beta)=((\alpha_1,\alpha_2,\alpha_3),(\beta_1,\beta_2,\beta_3))\in N_G^P(Z,V)$. By the definition of F, we have that $v=\sum_{i=1}^m u_i f_i(x)$ and there exists $\delta\geq 0$ such that $w=L(x,u)+\delta$. Since F is independent of y and z, $\alpha_2=\alpha_3=0$. Moreover, by the definition of $N_G^P(Z,V)$, ther e exists $\sigma>0$ such that

$$\begin{aligned}\langle\alpha_1,x'-x\rangle+\langle\beta_1,v'-v\rangle+\beta_2(w'-w)+\langle\beta_3,u'-u\rangle_m\\ \leq\sigma|x'-x|^2+\sigma(y'-y)^2+\sigma|z'-z|_m^2\\ +\sigma|v'-v|^2+\sigma(w'-w)^2+\sigma|u'-u|_m^2,\end{aligned}\tag{51}$$

for every $((x',y',z'),(v',w',u'))\in(\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}^m)\times(\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}^m)$. Hence, it is easy to see that $\beta_2\leq 0$, and that $\beta_2=0$ whenever $\delta>0$. Assume, first, $\beta_2<0$ (so that $\delta=0$). Applying (51) at point $((x',y,z),(f(x',u),L(x',u),u))\in G$, for every $x'\in\mathbb{R}^n$, gives

$$\begin{aligned}\langle\alpha_1,x'-x\rangle+\langle\beta_1,f(x',u)-f(x,u)\rangle+\beta_2\left(L(x',u)-L(x,u)\right)\\ \leq\sigma|x'-x|^2+\sigma\left|f(x',u)-f(x,u)\right|^2+\sigma\left(L(x',u)-L(x,u)\right)^2,\end{aligned}\tag{52}$$

which can be written as

$$\begin{aligned}L(x',u)-L(x,u)\\ +\sigma|x'-x|^2+\sigma\left|f(x',u)-f(x,u)\right|^2+\sigma\left(L(x',u)-L(x,u)\right)^2\\ \geq\left\langle\frac{\alpha_1}{(-\beta_2)},x'-x\right\rangle+\left\langle\frac{\beta_1}{(-\beta_2)},f(x',u)-f(x,u)\right\rangle\end{aligned}$$

for every $x' \in \mathbb{R}^n$. Now, for every $x' \in \mathbb{R}^n$ satisfying $|x'-x| \leq 1$,

$$\big|f(x',u) - f(x,u)\big| \leq \big|f_0(x') - f_0(x)\big| + \sum_{i=1}^m |u_i| \, \big|f_i(x') - f_i(x)\big|$$

$$= \Big| \int_0^1 \frac{d}{d\lambda} \big\{ f_0(x + \lambda(x'-x)) \big\} \, d\lambda \Big| + \sum_{i=1}^m |u_i| \Big| \int_0^1 \frac{d}{d\lambda} \big\{ f_i(x + \lambda(x'-x)) \big\} \, d\lambda \Big|$$

$$\leq \int_0^1 \big| df_0(x + \lambda(x'-x)) \big| \, |x'-x| d\lambda + \sum_{i=1}^m |u_i| \int_0^1 \big| df_i(x + \lambda(x'-x)) \big| \, |x'-x| d\lambda$$

$$\leq K|x'-x| + \sqrt{m} K |u|_m |x'-x|,$$

where K is a positive constant such that $|df_i(x')| \leq K$ for all $i = 0, \dots, m$ and $x' \in x + \overline{B}$. We also obtain, by (A2) (ii),

$$L(x',u) - L(x,u) \leq K(|x|+1)\theta\left(|u|_m\right)|x'-x|$$

for every $x' \in x + \overline{B}$. Since, for every $x' \in \mathbb{R}^n$,

$$f(x',u) - f(x,u) = \sum_{i=1}^m u_i df_i(x)(x'-x) + o(x'-x),$$

we obtain that, for every $x' \in x + \overline{B}$,

$$\begin{aligned} & L(x',u) - L(x,u) \\ & \qquad + \sigma\left(1 + K\left(1 + \sqrt{m}|u|_m\right) + K(|x|+1)\theta\left(|u|_m\right)\right)|x'-x|^2 \\ & \qquad \geq \left\langle \frac{\alpha_1}{(-\beta_2)}, x'-x \right\rangle + \left\langle \frac{\beta_1}{(-\beta_2)}, \sum_{i=1}^m u_i df_i(x)(x'-x) \right\rangle + o(x'-x). \end{aligned}$$

Therefore,

$$\frac{1}{(-\beta_2)}\left(\alpha_1 + A^*\beta_1\right) \in \partial_x L(x,u) \quad \text{where} \quad A := \sum_{i=1}^m u_i df_i(x)\,.$$

Now, apply (51) to $((x,y,z),(f(x,u'),L(x,u'),u')) \in G$, for any $u' \in \mathbb{R}^m$, to obtain

$$\begin{aligned} & \langle \beta_1, f(x,u') - f(x,u) \rangle + \beta_2\left(L(x,u') - L(x,u)\right) + \langle \beta_3, u'-u \rangle_m \\ & \qquad \leq \sigma\left|f(x,u') - f(x,u)\right|^2 + \sigma\left(L(x,u') - L(x,u)\right)^2 + \sigma|u'-u|_m^2, \end{aligned} \tag{53}$$

which can be written as

$$\begin{aligned} & L(x,u') - L(x,u) + \sigma\left|f(x,u') - f(x,u)\right|^2 + \sigma\left(L(x,u') - L(x,u)\right)^2 + \sigma|u'-u|_m^2 \\ & \qquad \geq \left\langle \frac{\beta_1}{(-\beta_2)}, f(x',u) - f(x,u) \right\rangle + \left\langle \frac{\beta_3}{(-\beta_2)}, u'-u \right\rangle_m \end{aligned}$$

for every $u' \in \mathbb{R}^m$. Also, for every $u' \in \mathbb{R}^m$,

$$\left|f(x,u') - f(x,u)\right| \leq \sqrt{m} \max\left\{|f_i(x)| \mid i = 1, \dots, m\right\}|u'-u|_m\,.$$

Thus, denoting by K a Lipschitz constant for $u' \mapsto L(x,u)$ on the ball centered at u with radius 1, we have $L(x,u') - L(x,u) \leq K|u'-u|_m$ for every $u' \in \mathbb{R}^m$ with $|u'-u|_m \le 1$. As above, this implies that

$$\frac{1}{(-\beta_2)}\left(\widehat{\beta}_1+\beta_3\right)=\nabla_u L(x,u) \quad \text{where} \quad \widehat{\beta}_1 := (\langle \beta_1, f_1(x)\rangle, \ldots, \langle \beta_1, f_m(x)\rangle)^*$$

Suppose now $\beta_2 = 0$. By (52) and the above estimates, we easily deduce that

$$\alpha_1 + A^*\beta_1 = 0, \quad \text{where} \quad A := \sum_{i=1}^m u_i df_i(x).$$

On the other hand, (53) yields

$$\frac{1}{(-\beta_2)}\left(\widehat{\beta}_1+\beta_3\right)=0 \quad \text{where} \quad \widehat{\beta}_1 := (\langle \beta_1, f_1(x)\rangle, \cdots, \langle \beta_1, f_m(x)\rangle)^*.$$

The fact that the same properties are satisfied whenever $(\alpha,\beta) \in N_G^P(Z,V)$ is easy to prove.

6.2. **Proof of Lemma 6.** Let $t_1 \le t_2 \in I := [0, \bar{t}+\delta]$ and let $u^1(\cdot), u^2(\cdot) \in \mathcal{U}_\infty$ be such that $\|u^1(\cdot)\|_\infty, \|u^2(\cdot)\|_\infty \le K$. Define

$$x_1(\cdot) := x_{u^1}(\cdot), x_2(\cdot) := x_{u^2}(\cdot).$$

Observ e that, by assumptions (A1)-(A2), there is a constant $C > 0$ such that

$$\|x_1(\cdot)\|_\infty, \|\dot{x}_1(\cdot)\|_\infty, \|x_2(\cdot)\|_\infty, \|\dot{x}_2(\cdot)\|_\infty \le C.$$

Moreover, by regularity of the lagrangian in both variables, we have that

$$|L(x,u) - L(x',u')| \le C\left(|x-x'| + |u-u'|_m\right)$$

for every $x, x' \in \mathbb{R}^n$ and $u, u' \in \mathbb{R}^m$ satisfying $|x|, |x'| \le C$ and $|u|_m, |u'|_m \le K$. In addition, by Gronwall's Lemma we conclude that

$$|x_2(s) - x_1(s)| \le C\left\|u^2(\cdot) - u^1(\cdot)\right\|_\infty, \quad \forall s \in [0, t_1].$$

Therefore, estimate (40) can be derived as follows

$$\begin{aligned}
&|C_{t_1}(u^1(\cdot)) - C_{t_2}(u^2(\cdot))| \\
&\le \left|\int_{t_1}^{t_2} L(x_2(s), u^2(s))ds\right| + \int_0^{t_1} \left|L(x_1(s), u^1(s)) - L(x_2(s), u^2(s))\right|ds \\
&\qquad \le C|t_1 - t_2| + C\int_0^{t_1}\Big(|x_1(s) - x_2(s)| + |u^1(s) - u^2(s)|\Big)ds \\
&\qquad\qquad \le C\Big(|t_1 - t_2| + \|u^2(\cdot) - u^1(\cdot)\|_\infty\Big)
\end{aligned}$$

Now, in order to prove (41), let $\mu \in [0,1]$ and define

$$\begin{aligned}
t_\mu &= \mu t_1 + (1-\mu)t_2 \\
u^\mu(\cdot) &= \mu u^1(\cdot) + (1-\mu)u^2(\cdot) \\
x_\mu(\cdot) &= x_{u^\mu}(\cdot)
\end{aligned}$$

We note that

$$\begin{aligned}\mu C_{t_1}(u_1(\cdot)) + (1-\mu)C_{t_2}(u_2(\cdot)) - C_{t_\mu}(u^\mu(\cdot)) \\ = \mu C_{t_1}(u_1(\cdot)) + (1-\mu)C_{t_1}(u_2(\cdot)) - C_{t_1}(u^\mu(\cdot)) \\ + (1-\mu)\int_{t_1}^{t_2} L(x_2(s),u^2(s))ds - \int_{t_1}^{t_\mu} L(x_\mu(s),u^\mu(s))ds \quad (54)\end{aligned}$$

The change of variables $s := \mu t_1 + (1-\mu)t$ gives

$$\int_{t_1}^{t_\mu} L(x_\mu(s),u^\mu(s))ds = (1-\mu)\int_{t_1}^{t_2} L\Big(x_\mu\big(\mu t_1+(1-\mu)t\big), u^\mu\big(\mu t_1+(1-\mu)t\big)\Big)dt\,.$$

Hence,

$$\begin{aligned}(1-\mu)\int_{t_1}^{t_2} L(x_2(s),u^2(s))ds - \int_{t_1}^{t_\mu} L(x_\mu(s),u^\mu(s))ds \\ = (1-\mu)\int_{t_1}^{t_2}\Big[L(x_2(t),u^2(t)) - L\Big(x_\mu\big(\mu t_1+(1-\mu)t\big), u^\mu\big(\mu t_1+(1-\mu)t\big)\Big)\Big]\,dt \\ = (1-\mu)\Big[I_1+I_2+I_3+I_4\Big]\end{aligned}$$

where

$$\begin{aligned}I_1 &:= \int_{t_1}^{t_2}\Big[L(x_2(t),u^2(t)) - L\Big(x_2\big(\mu t_1+(1-\mu)t\big),u^2(t)\Big)\Big]\,dt\\ I_2 &:= \int_{t_1}^{t_2}\Big[L\Big(x_2\big(\mu t_1+(1-\mu)t\big),u^2(t)\Big) - L\Big(x_2\big(\mu t_1+(1-\mu)t\big),u^\mu(t)\Big)\Big]\,dt\\ I_3 &:= \int_{t_1}^{t_2}\Big[L\Big(x_2\big(\mu t_1+(1-\mu)t\big),u^\mu(t)\Big) - L\Big(x_\mu\big(\mu t_1+(1-\mu)t\big),u^\mu(t)\Big)\Big]\,dt\end{aligned}$$

and

$$\begin{aligned}I_4 := \int_{t_1}^{t_2}\Big[L\Big(x_\mu\big(\mu t_1+(1-\mu)t\big),u^\mu(t)\Big)\\ -L\Big(x_\mu\big(\mu t_1+(1-\mu)t\big),u^\mu\big(\ mut_1+(1-\mu)t\big)\Big)\Big]\,dt\end{aligned}$$

In view of our preliminary considerations, we have that

$$\begin{aligned}I_1 \le C\int_{t_1}^{t_2}\big|x_2(t)-x_2\big(\mu t_1+(1-\mu)t\big)\big|\,ds\\ \le C\mu\int_{t_1}^{t_2}(t-t_1)dt = \frac{C}{2}\,\mu(t_2-t_1)^2\end{aligned}$$

Moreover,

$$I_2 \le C\int_{t_1}^{t_2}\big|u^2(t)-u^\mu(t)\big|_m\,dt \le C\mu|t_2-t_1|\ \big\|u^2(\cdot)-u^1(\cdot)\big\|_\infty$$

Also, observe that, again by Gronwall's Lemma,

$$|x_2(s)-x_\mu(s)| \le C\,\big\|u^2(\cdot)-u^\mu(\cdot)\big\|_\infty \le C\mu\,\big\|u^2(\cdot)-u^1(\cdot)\big\|_\infty$$

for all $s \in [t_1, t_\mu]$. Therefore,

$$I_3 \leq C \int_{t_1}^{t_2} \Big| x_2\big(\mu t_1 + (1-\mu)t\big) - x_\mu\big(\mu t_1 + (1-\mu)t\big) \Big| dt$$
$$\leq C\mu(t_2 - t_1) \left\| u^2(\cdot) - u^1(\cdot) \right\|_\infty .$$

Finally, recalling (25), we obtain

$$I_4 \leq C \int_{t_1}^{t_2} \Big| u^\mu(t) - u^\mu\big(\mu t_1 + (1-\mu)t\big) \Big| dt \leq \frac{KC}{2} \mu (t_2 - t_1)^2$$

Summing up, we conclude that

$$(1-\mu) \int_{t_1}^{t_2} L(x_2(s), u^2(s)) ds - \int_{t_1}^{t_\mu} L(x_\mu(s), u^\mu(s)) ds$$
$$\leq C\mu(1-\mu)(t_2 - t_1) \left\| u^2(\cdot) - u^1(\cdot) \right\|_\infty + C\mu(1-\mu)(t_2 - t_1)^2 \quad (55)$$

The quantity $\mu C_{t_1}(u^1(\cdot)) + (1-\mu) C_{t_1}(u^2(\cdot)) - C_{t_1}(u^\mu(\cdot))$ can be written as

$$\begin{aligned} & \mu C_{t_1}(u^1(\cdot)) + (1-\mu) C_{t_1}(u^2(\cdot)) - C_{t_1}(u^\mu(\cdot)) \\ = & \int_0^{t_1} \Big[\mu L(x_1(s), u^1(s)) + (1-\mu) L(x_2(s), u^2(s)) - L(x_\mu(s), u^\mu(s)) \Big] ds \\ = & \int_0^{t_1} \Big[\mu L(x_1(s), u^\mu(s)) + (1-\mu) L(x_2(s), u^\mu(s)) - L(x_\mu(s), u^\mu(s)) \Big] ds \\ & +\mu \int_0^{t_1} \Big[L(x_1(s), u^1(s)) - L(x_1(s), u^\mu(s)) \Big] ds \\ & +(1-\mu) \int_0^{t_1} \Big[L(x_2(s), u^2(s)) - L(x_2(s), u^\mu(s)) \Big] ds \end{aligned}$$

From assumption (A2)(ii), we know that, for some constant $C' > 0$,

$$\mu L(x,u) + (1-\mu) L(x',u) - L\left(\mu x + (1-\mu)x', u\right) \leq C' \mu (1-\mu) |x - x'|^2$$

for every every $\mu \in [0,1]$, every $x, x' \in \mathbb{R}^n$ satisfying $|x|, |x'| \leq C$ and every $u, u' \in \mathbb{R}^m$ satisfying $|u|_m, |u'|_m \leq K$. Moreover, since the end-point mapping E^{x_0,t_1} is of class $C^{1,1}$, we can also assume that,

$$|\mu x_1(s) + (1-\mu) x_2(s) - x_\mu(s)| \leq C\mu(1-\mu) \left\| u^1(\cdot) - u^2(\cdot) \right\|_\infty^2, \quad \forall s \in [0, t_1].$$

Hence we obtain,

$$\begin{aligned}
&\int_0^{t_1} \Big[\mu L(x_1(s),u^\mu(s)) + (1-\mu)L(x_2(s),u^\mu(s)) - L(x_\mu(s),u^\mu(s))\Big]ds \\
=\ &\int_0^{t_1} [\mu L(x_1(s),u^\mu(s)) + (1-\mu)L(x_2(s),u^\mu(s)) \\
&\qquad\qquad - L\big(\mu x_1(s)+(1-\mu)x_2(s),u^\mu(s)\big)\big]\,ds \\
&\qquad + \int_0^{t_1} \Big[L\big(\mu x_1(s)+(1-\mu)x_2(s),u^\mu(s)\big) - L(x_\mu(s),u^\mu(s))\Big]ds \\
\le\ & C'\mu(1-\mu)\int_0^{t_1} |x_1(s)-x_2(s)|^2\,ds \\
&\qquad + C\int_0^{t_1} |\mu x_1(s)+(1-\mu)x_2(s)-x_\mu(s)|\,ds \\
\le\ & C\mu(1-\mu)\left\|u^1(\cdot)-u^2(\cdot)\right\|_\infty^2 .
\end{aligned}$$

On the other hand, denoting by $h(s)$ the difference $u^1(s)-u^2(s)$ for all $s\in[0,t_1]$, we have that

$$\begin{aligned}
&\int_0^{t_1} \Big[L(x_1(s),u^1(s)) - L(x_1(s),u^\mu(s))\Big]ds \\
=\ &(1-\mu)\int_0^{t_1}\int_0^1 \langle \nabla_u L\left(x_1(s),u^\mu(s)+\lambda(1-\mu)h(s)\right), h(s)\rangle\, d\lambda ds,
\end{aligned}$$

and

$$\begin{aligned}
&\int_0^{t_1} \Big[L(x_2(s),u^2(s)) - L(x_2(s),u^\mu(s))\Big]ds \\
=\ &-\mu\int_0^{t_1}\int_0^1 \langle \nabla_u L\left(x_2(s),u^\mu(s)-\lambda\mu h(s)\right), h(s)\rangle\, d\lambda ds
\end{aligned}$$

Thus,

$$\begin{aligned}
&\mu\int_0^{t_1} \Big[L(x_1(s),u^1(s)) - L(x_1(s),u^\mu(s))\Big]ds \\
&\qquad + (1-\mu)\int_0^{t_1} \Big[L(x_2(s),u^2(s)) - L(x_2(s),u^\mu(s))\Big]ds \\
&\qquad\qquad = \mu(1-\mu)\int_0^{t_1}\int_0^1 \langle \Lambda(\lambda,s), h(s)\rangle\, d\lambda ds,
\end{aligned}$$

where

$$\begin{aligned}
&\Lambda(\lambda,s) \\
&\quad = \nabla_u L\left(x_1(s),u^\mu(s)+\lambda(1-\mu)h(s)\right) - \nabla_u L\left(x_2(s),u^\mu(s)-\lambda\mu h(s)\right)
\end{aligned}$$

for every $\lambda\in[0,1]$ and every $s\in[0,t_1]$. Since

$$\left|\nabla_u L(x,u) - \nabla_u L(x',u')\right| \le C\big(|x-x'| + |u-u'|_m\big)$$

for every $x, x' \in \mathbb{R}^n$ satisfying $|x|, |x'| \leq C$ and every $u, u' \in \mathbb{R}^m$ satisfying $|u|_m, |u'|_m \leq K$, we obtain that, for every $\lambda \in [0,1]$ and every $s \in [0, t_1]$,

$$\begin{aligned} & \Lambda(\lambda, s) \leq C\,|h(s)|_m + C\,|x_1(s) - x_2(s)| \\ \leq \quad & C\,\left\|u^2(\cdot) - u^1(\cdot)\right\|_\infty . \end{aligned}$$

Therefore,

$$\begin{aligned} \mu \int_0^{t_1} & \Big[L(x_1(s), u^1(s)) - L(x_1(s), u^\mu(s))\Big] ds \\ & + (1-\mu) \int_0^{t_1} \Big[L(x_2(s), u^2(s)) - L(x_2(s), u^\mu(s))\Big] ds \\ & \leq \mu(1-\mu)\left\|u^2(\cdot) - u^1(\cdot)\right\|_\infty^2 \quad (56) \end{aligned}$$

The conclusion follows from (54), (55), (56).

References

[1] A. Agrachev. Compactness for Sub-Riemannian length-minimizers and subanalyticity. Control theory and its applications (Grado, 1998), *Rend. Sem. Mat. Univ. Politec. Torino*, 56(4):1-12, 2001.

[2] G. Alberti, L. Ambrosio and P. Cannarsa. On the singularities of convex functions. *Manuscripta Math.*, 76(3-4):421–435, 1992.

[3] A. Bellaiche. The tangent space in sub-Riemannian geometry. In *Sub-Riemannian Geometry, Birkhäuser*, 1–78, 1996.

[4] P. Cannarsa and H. Frankowska. Some characterizations of optimal trajectories in control theory. *SIAM J. Control Optim.*, 29(6):1322-1347, 1991.

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

[6] P. Cannarsa and C. Sinestrari. On a class of nonlinear time optimal control problems. *Discrete Contin. Dynam. Systems*, 1(2):285-300, 1995.

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

[8] P. Cannarsa and C. Sinestrari. *Semiconcave functions, Hamilton-Jacobi equations, and optimal control.* Progress in Nonlinear Differential Equations and their Applications, 58. Birkhäuser Boston Inc., Boston, MA, 2004.

[9] C.-L. Chow. Über systeme von linearen partiellen differentialgleichungen ester ordnung. *Math. Ann.*, 117: 98-105, 1939.

[10] F.H. Clarke. *Optimization and no nsmooth analysis* Canadian Mathematical Society Series of Monographs and Advanced Texts, John Wiley & Sons Inc., New York, 1983.

[11] F. Clarke. Necessary conditions in dynamic optimization. Mem. Amer. Math. Soc., vol. 173, number 816, 2005.

[12] F.H. Clarke, Yu.S. Ledyaev, R.J. Stern, and P.R. Wolenski. *Nonsmooth Analysis and Control Theory.* Graduate Texts in Mathematics, vol. 178. Springer-Verlag, New York, 1998.

[13] Z. Ge. Horizontal path spaces and Carnot-Carathéodory metrics. *Pacific J. Math.*, 161(2):255-286, 1993.

[14] I. Kupka. Géométrie sous-riemannienne. *Astérisque*, (241):Exp. No. 817, 5, 351-380, 1997. Séminaire Bourbaki, Vol. 1995/96.

[15] R. Montgomery. *A tour of subriemannian geometries, their geodesics and applications*, volume 91 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2002.

[16] P.K. Rashevsky. About connecting two points of a completely nonholonomic space by admissible curve. *Uch. Zapiski Ped. Inst. Libknechta*, 2:83-94, 1938.

[17] C. B. Rayer. The exponential map for the Lagrange problem on differentiable manifold. *Phil. Trans. Roy. Soc. London, A, Math. and Phys.*, 1127, 299-344, 1967.

[18] L. Rifford. Existence of Lipschitz and semiconcave control-Lyapunov functions. *SIAM J. Control Optim.*, 39(4):1043–1064, 2000.

[19] L. Rifford. Semiconcave control-Lyapunov functions and stabilizing feedbacks. *SIAM J. Control Optim.*, 41(3):659–681, 2002.

[20] L. Rifford. The Stabilization Problem: AGAS and SRS Feedbacks. In *Optimal Control, Stabilization, and Nonsmooth Analysis*, Lectures Notes in Control and Information Sciences, 301, Springer-Verlag, Heidelberg (2004), 173–184.

[21] L. Rifford. A Morse-Sard theorem for the distance function on Riemannian manifolds. *Manuscripta Math.*, 113:251–265, 2004.

[22] L. Rifford. À propos des sphères sous-riemanniennes. *Bull. Belg. Math. Soc. Simon Stevin*, 13(3):521–526, 2006.

[23] L. Rifford and E. Trélat. On the stabilization problem for nonholonomic distributions. *J. Eur. Math. Soc.*, to appear.

[24] E. Trélat. Some properties of the value function and its level sets for affine control systems with quadratic cost. *J. of Dynamical and Control Systems*, 6(4):511-541, 2000.

[25] A. Weinstein. Fat bundles and symplectic manifolds. *Adv. in Math.*, 37, 239-250, 1980.