Semiconcave Functions, Hamilton–Jacobi Equations, and Optimal Control

Piermarco Cannarsa & Carlo Sinestrari

Contents

Chapter 1

A model problem

The purpose of this chapter is to give an outline of the dynamic programming method in control theory and to show how the class of semiconcave functions naturally appears in this context. We do this by considering a classical problem in calculus of variations.

In Section 1.1, after defining our problem, we introduce the dynamic programming principle and the notion of value function, which are the starting point of the so–called Bellman's *dynamic programming method* in control theory. Then, in Section 1.2, we restrict our attention to the case where the integrand of our problem has no explicit (t, x)–dependence; in this case the value function admits a simple representation formula due to Hopf.

In Section 1.3 we observe that the value function is solution of a partial differential equation called *Hamilton–Jacobi* (or sometimes Hamilton–Jacobi–Bellman) equation. However, this property holds only almost everywhere, because our value function is Lipschitz continuous but not differentiable in general. This fact constitutes a major problem in the use of the dynamic programming approach. In fact, the Cauchy problem for a Hamilton–Jacobi equation has, in general, infinitely many Lipschitz continuous solutions satisfying the equation almost everywhere. So, such a property is not enough to characterize the value function.

Before seeing how to overcome this difficulty, we give in Section 1.4 an account of the classical method of characteristics for Hamilton–Jacobi equations. This technique gives in an elementary way a local existence result for smooth solutions, and at the same time shows that no global smooth solution exists in general. Although the method is completely independent of the control–theoretic interpretation of the equation, there is an interesting connection between the solutions of the characteristic system and optimal trajectories of the corresponding problem in control or calculus of variations.

In Section 1.5 we introduce the notion of semiconcavity and we show how this property provides a possible answer to the questions outlined before. On one hand, one can prove that the value function is semiconcave. On the other hand, a uniqueness result holds for Lipschitz continuous solutions of Hamilton–Jacobi equations which are also semiconcave. Thus, the value function can be characterized as the unique semiconcave solution of the equation.

Let us point out that the application of semiconcavity to uniqueness results has nowadays mainly historical interest, since it has been later incorporated in the more general theory of viscosity solutions. Stronger motivations for studying semiconcavity for the analysis of optimal control problems are now others: semiconcavity is a finer regularity property than Lipschitz continuity and in the problems where the value function is semiconcave the approach of dynamic programming becomes simpler and more powerful.

We conclude the chapter by describing, in Section 1.6, the connection between Hamilton-Jacobi equations and hyperbolic conservation laws, both for classical and for generalized solutions. In particular, we show that semiconcavity corresponds to a one–sided Lipschitz estimate satisfied by the admissible solutions of conservation laws.

1.1 A problem in calculus of variations

Given $0 < T \leq +\infty$, we set $Q_T =]0,T[\, \times \mathbb{R}^n$. We suppose that two continuous functions

$$L : \overline{Q}_T \times \mathbb{R}^n \to \mathbb{R}, \qquad u_0 : \mathbb{R}^n \to \mathbb{R}$$

are given. The function L will be called *running cost*, or *lagrangian*, while u_0 is called *initial cost*. We assume that both functions are bounded from below.

For fixed $(t,x) \in \overline{Q}_T$, we set

$$\mathcal{A}(t,x) = \{y \in W^{1,1}([0,t];\mathbb{R}^n) \;:\; y(t) = x\}$$

and we consider the problem of minimizing the *cost functional*

$$J_t[y] = \int_0^t L(s, y(s), \dot{y}(s))\, ds + u_0(y(0))$$

over all arcs $y \in \mathcal{A}(t,x)$.

Definition 1.1.1 *The function $u : \overline{Q}_T \to \mathbb{R}$ defined as*

$$u(t,x) = \inf_{y \in \mathcal{A}(t,x)} J_t[y] \tag{1.1.1}$$

is called the value function *of our minimization problem.*

By our assumptions u is finite everywhere. In addition we have

$$u(0,x) = u_0(x). \tag{1.1.2}$$

It turns out that the knowledge of the value function u simplifies substantially the resolution of our minimization problem. It is important therefore to find an alternative characterization of this function which allows to compute it without referring directly to the definition.

The following result is called *Bellman's optimality principle* or *dynamic programming principle* and is the starting point for the study of the properties of u.

Theorem 1.1.2 *Let $(t,x) \in \overline{Q}_T$ and $y \in \mathcal{A}(t,x)$. Then, for all $t' \in [0,t]$,*

$$u(t,x) \le \int_{t'}^{t} L(s, y(s), \dot{y}(s))\, ds + u(t', y(t')). \tag{1.1.3}$$

In addition, $u(t,x) = J_t[y]$ if and only if equality holds in (1.1.3) *for all $t' \in [0,t]$.*

Proof — For fixed $t' \in [0,t]$, let z be an arc in $W^{1,1}([0,t'];\mathbb{R}^n)$ such that $z(t') = y(t')$. If we set

$$\xi(s) = \begin{cases} z(s), & s \in [0,t'], \\ y(s), & s \in [t',t], \end{cases}$$

we have that $\xi \in \mathcal{A}(t,x)$ and therefore

$$u(t,x) \le J_t[\xi] = \int_{t'}^{t} L(s,y,\dot{y})\, ds + \int_0^{t'} L(s,z,\dot{z})\, ds + u_0(z(0)).$$

Taking the infimum over all $z \in \mathcal{A}(t', y(t'))$ we obtain (1.1.3).

If (1.1.3) holds as an equality for all $t' \in [0,t]$, then choosing $t' = 0$ yields that y is a minimizer for J_t. Conversely, if y is a minimizer we find, by the definition of u and by inequality (1.1.3),

$$\int_0^t L(s,y,\dot{y})\, ds + u_0(y(0)) = u(t,x) \le \int_{t'}^{t} L(s,y,\dot{y})\, ds + u(t', y(t')) \tag{1.1.4}$$

for any given $t' \in [0,t]$. This implies that $J_{t'}[y] \le u(t', y(t'))$. Since by definition $J_{t'}[y] \ge u(t', y(t'))$, we must have equality in (1.1.4), and therefore also in (1.1.3). ■

Theorem 1.1.3 *Let $(t,x) \in \overline{Q}_T$. Then, for all $t' \in [0,t]$,*

$$u(t,x) = \inf_{y \in \mathcal{A}(t,x)} \left\{ \int_{t'}^{t} L(s, y(s), \dot{y}(s))\, ds + u(t', y(t')) \right\}. \tag{1.1.5}$$

Proof — Given $\varepsilon > 0$, let $y \in \mathcal{A}(t,x)$ be such that

$$u(t,x) + \varepsilon \geq \int_0^t L(s,y,\dot{y})\,ds + u_0(y(0)).$$

Then

$$\begin{aligned} u(t,x) &\geq \int_{t'}^t L(s,y,\dot{y})\,ds + \int_0^{t'} L(s,y,\dot{y})\,ds + u_0(y(0)) - \varepsilon \\ &\geq \int_{t'}^t L(s,y,\dot{y})\,ds + u(t',y(t')) - \varepsilon. \end{aligned}$$

By the arbitrariness of ε we deduce that $u(t,x)$ is greater than or equal to the right–hand side of (1.1.5). The converse inequality follows from Theorem 1.1.2. ■

1.2 The Hopf formula

Let us consider the special case when $L(t,x,q) = L(q)$ and let us assume

$$\begin{cases} \text{(i) } L \text{ is convex and } \lim\limits_{|q|\to\infty} \dfrac{L(q)}{|q|} = +\infty \\ \text{(ii) } u_0 \in \mathrm{Lip}\,(\mathbb{R}^n). \end{cases} \tag{1.2.6}$$

Under these hypotheses the value function of our problem admits a simple representation formula, called *Hopf's formula.*

Theorem 1.2.1 *The value function u satisfies*

$$u(t,x) = \min_{z\in\mathbb{R}^n} \left[tL\left(\frac{x-z}{t}\right) + u_0(z)\right] \tag{1.2.7}$$

for all $(t,x) \in \overline{Q}_T$.

Proof — Observe that the minimum in (1.2.7) exists thanks to hypotheses (1.2.6). Let us denote by $v(t,x)$ the left–hand side of (1.2.7).

For fixed $(t,x) \in \overline{Q}_T$ and $z \in \mathbb{R}^n$, let us set

$$y(s) = z + \frac{s}{t}(x-z), \qquad 0 \leq s \leq t.$$

Then $y \in \mathcal{A}(t,x)$ and therefore

$$u(t,x) \leq J_t[\xi] = tL\left(\frac{x-z}{t}\right) + u_0(z).$$

Taking the infimum over z we conclude that $u \le v$ on $\overline{Q}_T$.

To prove the opposite inequality, let us take $\zeta \in \mathcal{A}(t,x)$. From Jensen's inequality it follows that

$$L\left(\frac{x-\zeta(0)}{t}\right) = L\left(\frac{1}{t}\int_0^t \dot{\zeta}(s)\,ds\right) \le \frac{1}{t}\int_0^t L(\dot{\zeta}(s))\,ds$$

and therefore

$$v(t,x) \le u_0(\zeta(0)) + tL\left(\frac{x-\zeta(0)}{t}\right) \le J_t[\zeta].$$

Taking the infimum over $\zeta \in \mathcal{A}(t,x)$ we obtain that $v \le u$. ∎

Using Hopf's formula we can prove a first regularity property of u.

Theorem 1.2.2 *Under the assumptions* (1.2.6) *the value function u is Lipschitz continuous in $\overline{Q}_T$. More precisely, we have*

$$|u(t',x') - u(t,x)| \le L_0|x-x'| + L_1|t-t'|, \tag{1.2.8}$$

where $L_0 = \mathrm{Lip}\,(u_0)$ and $L_1 \ge 0$ is a suitable constant.

Proof — Let us first observe that (1.2.7) implies

$$\frac{u(t,x)-u_0(x)}{t} \le L(0) \tag{1.2.9}$$

for all (t,x) with $t>0$. Let us now take $(t,x),(t',x') \in Q_T$ and let $y \in \mathbb{R}^n$ be such that

$$u(t,x) = tL\left(\frac{x-y}{t}\right) + u_0(y).$$

Then we find, using (1.2.9),

$$\begin{aligned} L\left(\frac{x-y}{t}\right) &= \frac{u(t,x)-u_0(x)}{t} + \frac{u_0(x)-u_0(y)}{t} \\ &\le L(0) + L_0\frac{|x-y|}{t}. \end{aligned}$$

Since L is superlinear, we can find a constant $C_1 > 0$ depending on $L_0, L(0)$ such that $L(q) \ge L(0) + L_0|q|$ if $|q| > C_1$. Then the previous inequality implies

$$\left|\frac{x-y}{t}\right| \le C_1. \tag{1.2.10}$$

Let us now set

$$y' = x' - t'\frac{x-y}{t}.$$

Then we have

$$\frac{x-y}{t} = \frac{x'-y'}{t'}, \qquad y'-y = x'-x+(t-t')\frac{x-y}{t},$$

and so Hopf's formula (1.2.7), together with (1.2.10), yields

$$\begin{aligned} u(t',x') - u(t,x) &\le (t'-t)L\left(\frac{x-y}{t}\right) + u_0(y') - u_0(y) \\ &\le |t-t'| \max_{|q|\le C_1} L(q) + L_0|y'-y| \\ &\le \left(C_1 + \max_{|q|\le C_1} L(q)\right)|t-t'| + L_0|x'-x|. \end{aligned}$$

By interchanging the role of (x,t) and (x',t'), we obtain the reverse inequality, and so the conclusion. ∎

A well-known theorem due to Rademacher (see e.g. [42]) asserts that a Lipschitz continuous function is differentiable almost everywhere, and so the previous result immediately implies

Corollary 1.2.3 *Under hypotheses* (1.2.6)*, the value function u is differentiable a.e. in Q_T.*

On the other hand, it is easy to see that u is in general not differentiable everywhere, even if u_0 and L are differentiable.

Example 1.2.4 Let us consider problem (1.1.1) with $n = 1$, $L(q) = q^2/2$ and the initial cost given by

$$u_0(z) = \begin{cases} -z^2 & \text{if } |z| < 1 \\ 1-2|z| & \text{if } |z| \ge 1. \end{cases}$$

Observe that u_0 is of class C^1 but not C^2. This is not essential for the behavior we are going to describe; one can build examples with C^∞ data exhibiting similar properties (see Example 6.1.5).

Let us compute the value function using Hopf's formula. We find easily that the minimum in (1.2.7) is assumed for

$$\begin{cases} z = \dfrac{x}{1-2t} & \text{if } t < 1/2 \text{ and } |x| < 1-2t \\ z = x + \mathrm{sgn}(x)2t & \text{if } |x| \ge 1-2t \ge 0 \text{ or if } t \ge 1/2, \end{cases}$$

yielding

$$u(t,x) = \begin{cases} -\dfrac{x^2}{1-2t} & \text{if } t < 1/2 \text{ and } |x| < 1-2t \\ 1-2(|x|+t) & \text{if } |x| \ge 1-2t \ge 0 \text{ or if } t \ge 1/2. \end{cases}$$

Therefore $u(t,x)$ is not differentiable at the points of the form $(t,0)$ with $t \geq 1/2$.

An important issue of the approach we are describing is to obtain precise results about the regularity of the value function u. Roughly speaking, we want to know how "badly" u can behave. We have seen in Theorem 1.2.2 that u is at least Lipschitz continuous. More refined information about the regularity of u will be obtained later in this chapter (see Section 1.5).

1.3 Hamilton–Jacobi equations

In this section we show that the value function u solves a suitable partial differential equation. We assume throughout that hypotheses (1.2.6) are satisfied. We use the notation

$$u_t = \frac{\partial u}{\partial t}, \qquad \nabla u = \left(\frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_n}\right).$$

Theorem 1.3.1 *Let u be differentiable at a point $(t,x) \in Q_T$. Then*

$$u_t(t,x) + H(\nabla u(t,x)) = 0, \tag{1.3.11}$$

where

$$H(p) = \sup_{q \in \mathbb{R}^n} [p \cdot q - L(q)]. \tag{1.3.12}$$

Equation (1.3.11) is called *Hamilton–Jacobi equation*. The function H is called the *hamiltonian* associated to our problem of calculus of variations. In general, a function defined as in (1.3.12) is called the *Legendre–Fenchel transform* of L (see Appendix A.1).

Proof — Let (t,x) be a point where u is differentiable. Given $q \in \mathbb{R}^n$, $s > 0$, let us set $y(\tau) = x + q(\tau - t)$. Then we obtain from (1.1.3)

$$u(t+s, x+sq) \leq \int_t^{t+s} L(\dot{y}(\tau))d\tau + u(t,x),$$

which implies

$$\frac{u(t+s, x+sq) - u(t,x)}{s} \leq L(q).$$

Letting $s \downarrow 0$ we obtain

$$u_t(t,x) + q \cdot \nabla u(t,x) - L(q) \leq 0,$$

and we deduce, by the arbitrariness of q,

$$u_t(t,x) + H(\nabla u(t,x)) \leq 0.$$

To prove the converse inequality, let us take $\hat{x} \in \mathbb{R}^n$ such that

$$u(t,x) = tL\left(\frac{x-\hat{x}}{t}\right) + u_0(\hat{x}).$$

Such an $\hat{x}$ exists by Hopf's formula. We then set, for $s \in [0,t[$,

$$x_s = \frac{s}{t}x + \left(1 - \frac{s}{t}\right)\hat{x}.$$

Since $(x-\hat{x})/t = (x_s - \hat{x})/s$, we deduce from Hopf's formula

$$\begin{aligned} u(t,x) - u(s,x_s) &\geq tL\left(\frac{x-\hat{x}}{t}\right) + u_0(\hat{x}) - sL\left(\frac{x_s-\hat{x}}{s}\right) + u_0(\hat{x}) \\ &= (t-s)L\left(\frac{x-\hat{x}}{t}\right). \end{aligned}$$

Dividing by $t-s$ and letting $s \uparrow t$ we obtain

$$u_t(t,x) + \frac{x-\hat{x}}{t} \cdot \nabla u(t,x) - L\left(\frac{x-\hat{x}}{t}\right) \geq 0,$$

and we conclude that $u_t(t,x) + H(\nabla u(t,x)) \geq 0$. ∎

From Corollary 1.2.3 and Theorem 1.3.1 it follows that the value function u given from (1.2.7) satisfies

$$\begin{cases} u_t + H(\nabla u) = 0 & (t,x) \in Q_T \text{ a.e.} \\ u(0,x) = u_0(x) & x \in \mathbb{R}^n. \end{cases} \tag{1.3.13}$$

The involutive character of the Legendre transform implies that the correspondence between calculus of variations and Hamilton–Jacobi equations is valid in both directions. More precisely, let H, u_0 be given such that

$$\begin{cases} \text{(i) } H \text{ is convex and } \lim\limits_{|p|\to\infty} \dfrac{H(p)}{|p|} = +\infty \\ \text{(ii) } u_0 \in \mathrm{Lip}\,(\mathbb{R}^n) \end{cases} \tag{1.3.14}$$

and suppose that we want to solve problem (1.3.13). To this purpose, we can define L to be the Legendre transform of H, i.e.

$$L(q) = \max_{p \in \mathbb{R}^n}[p \cdot q - H(p)].$$

Then, L satisfies property (1.2.6)-(i) and H is the Legendre transform of L, i.e. (1.3.12) holds. Therefore, Hopf's formula (1.2.7) yields a Lipschitz function u that solves (1.3.13).

The property of solving problem (1.3.13) almost everywhere, however, is not enough to characterize the value function u. Indeed, such a problem can have more than one solution in the class $\mathrm{Lip}(\overline{Q}_T)$, as the next example shows.

Example 1.3.2 The problem

$$\begin{cases} u_t + u_x^2 = 0 & (t,x) \in \mathbb{R}_+ \times \mathbb{R} \text{ a.e.} \\ u(0,x) = 0 & x \in \mathbb{R}. \end{cases} \tag{1.3.15}$$

admits the solution $u \equiv 0$. However, for any $a > 0$, the function u_a defined as

$$u_a(t,x) = \begin{cases} 0 & \text{if } |x| \geq at \\ a|x| - a^2 t & \text{if } |x| < at \end{cases}$$

is a Lipschitz function satisfying the equation almost everywhere together with its initial condition.

Simple examples like the above one show that the property of solving the equation almost everywhere is too weak and does not suffice to provide a satisfactory notion of generalized solution. It is therefore desirable to find some additional condition which ensures uniqueness and characterizes the value function among the Lipschitz continuous solutions of the equation. A possible way of doing this relies on the semiconcavity property of the value function and will be pursued in Section 1.5.

1.4 The method of characteristics

For the study of problem (1.3.13) it is interesting to consider the method of characteristics, which is a classical approach to find regular solutions to first order partial differential equations. As we show below, the method of characteristics provides a local smooth solution of the equation, if the data are smooth, but at the same time shows that in general no global regular solution exists.

Suppose that H, u_0 are in $C^2(\mathbb{R}^n)$, and let u be a C^2 solution of problem (1.3.13) in some strip Q_T. For fixed $z \in \mathbb{R}$, let us denote by $X(t;z)$ the solution of the ordinary differential equation (here the dot denotes differentiation with respect to t)

$$\dot{X} = DH(\nabla u(t,X)), \qquad X(0) = z. \tag{1.4.16}$$

Such a solution is defined in some maximal interval $[0, T_z[$ (although it will later turn out that $T_z = T$ for all z). The curve $t \to X(t; z)$ is called the *characteristic curve* associated with u and starting from the point $(0, z)$. Let us now set

$$U(t; z) = u(t, X(t; z)), \qquad P(t; z) = \nabla u(t, X(t; z)). \tag{1.4.17}$$

Then, using the fact that u solves problem (1.3.13) we find that

$$\dot{U} = u_t(t, X) + \nabla u(t, X) \cdot \dot{X} = -H(P) + DH(P) \cdot P,$$

$$\dot{P} = \nabla u_t(t, X) + \nabla^2 u(t, X)\dot{X} = \nabla(u_t + H(\nabla u))(t, X) = 0.$$

Therefore P is constant, and so the right–hand side of (1.4.16) is also constant. Thus, X is defined in $[0, T[$ and we can compute explicitly X, U, P obtaining

$$\begin{cases} P(t; z) = \nabla u_0(z) \\ X(t; z) = z + tDH(\nabla u_0(z)) \\ U(t; z) = u_0(z) + t[DH(\nabla u_0(z)) \cdot \nabla u_0(z) - H(\nabla u_0(z))]. \end{cases} \tag{1.4.18}$$

Observe that the right–hand side of (1.4.18) is no longer defined in terms of the solution u, but only depends on the initial value u_0. This suggests that, even without assuming in advance the existence of a solution, one can use these formulas to define one. This approach is called the method of characteristics. As we are now going to show, such a construction can be in general carried out only locally in time.

We need the following classical result about the global invertibility of maps (see e.g. [7, Th. 3.1.8]).

Theorem 1.4.1 *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be of class C^1 and proper (that is, $f^{-1}(K)$ is compact whenever K is compact). If f is locally invertible, then f is a global C^1 diffeomorphism of $\mathbb{R}^n$ into itself.*

We can now give the local existence result for classical solutions based on the method of characteristics.

Theorem 1.4.2 *Let $u_0, H \in C^2(\mathbb{R}^n)$ be given, and suppose that ∇u_0 and $\nabla^2 u_0$ are bounded. Let us set*

$$T^* = \sup\{t > 0 \ : \ I + tD^2 H(\nabla u_0(z))\nabla^2 u_0(z) \text{ is invertible for all } z \in \mathbb{R}^n\}.$$

Then problem (1.3.13) *has a unique solution $u \in C^2(\mathbb{R} \times [0, T^*[\,)$.*

Proof — Let X, U, P be defined as in (1.4.18) and let us fix $t \in [0, T^*[$. By our assumptions, the jacobian of the map $z \to X(t; z)$ is uniformly bounded and nondegenerate for every z; this implies that $z \to X(t; z)$ is proper and locally invertible. By Theorem 1.4.1, such a map has a C^1 inverse, which we denote by $x \to Z(t; x)$. After setting

$$u(t, x) = U(t, Z(t; x)), \qquad (t, x) \in [0, T^*[\times \mathbb{R}^n,$$

let us check that u is a C^2 solution of problem (1.3.13).

From the definition it is clear that u is of class at least C^1. Let us compute its derivatives. By (1.4.18) we have

$$\begin{aligned} U_z(t, z) &= \nabla u_0(z) + t \nabla u_0(z) D^2 H(\nabla u_0(z)) \nabla^2 u_0(z) \\ &= P(t, z) X_z(t, z) \end{aligned}$$

where the subscript denotes partial differentiation. This implies, by the definition of Z,

$$\nabla u(t, x) = U_z(t; Z(t; x)) Z_x(t; x) = P(t; Z(t, x)). \tag{1.4.19}$$

We also find

$$\begin{aligned} u_t(t, x) &= U_t(t; Z(t, x)) + U_z(t; Z(t, x)) Z_t(t; x) \\ &= DH(P) \cdot P - H(P) + P X_z Z_t. \end{aligned}$$

where we have written for simplicity P, Z, etc. instead of $P(t; Z(t, x))$, $Z = Z(t; x)$ respectively. Keeping into account that $X(t; Z(t; x)) \equiv x$ we obtain

$$X_t(t; Z) + X_z(t; Z) Z_t = 0.$$

Therefore, since $X_t = DH(P)$ by (1.4.18),

$$u_t(t, x) = X_t \cdot P - H(P) + P X_z Z_t = -H(P).$$

This equality, together with (1.4.19), implies that $u \in C^2$ and satisfies problem (1.3.12) (the fact that the initial condition is satisfied is a direct consequence of the definition). The uniqueness follows from the remarks we have done before the statement of theorem. ■

Corollary 1.4.3 *Let u_0 and H be as in the previous theorem. Set*

$$M_0 = \sup_{z \in \mathbb{R}^n} |\nabla u_0(z)|, \quad M_1 = \sup_{z \in \mathbb{R}^n} ||\nabla^2 u_0(z)||, \quad M_2 = \sup_{|p| \leq M_0} ||D^2 H(p)||.$$

Then problem (1.3.13) *has a C^2 solution at least up to time $T^* = M_1 M_2^{-1}$. If H and u_0 are both convex (or both concave) then the problem has a C^2 solution for all times.*

The time T^* given by Theorem 1.4.2 is optimal, as shown by the next result.

Theorem 1.4.4 *Let H, u_0 and T^* be as in Theorem 1.4.2. If $T' > T^*$, then no C^2 solution of problem* (1.3.13) *exists in* $[0, T'[\, \times \mathbb{R}^n$.

We give two different proofs of this statement; the argument of the second one is more intuitive, but can be applied only in the case $n = 1$.

First proof — We argue by contradiction and suppose that there exists a solution $u \in C^2([0, T'[\, \times \mathbb{R}^n)$, with $T' > T^*$. By definition of T^*, there exist $t^* \in]T^*, T'[\,$, $z^* \in \mathbb{R}^n$ and $\theta \in \mathbb{R}^n \setminus \{0\}$ such that

$$X_z(t^*; z^*)\theta = \theta + t^* D^2 H(\nabla u_0(z^*))\nabla^2 u_0(z^*)\theta = 0.$$

This implies in particular that

$$P_z(t^*; z^*)\theta = \nabla^2 u_0(z^*)\theta \neq 0.$$

Differentiating (1.4.17) with respect to z we obtain

$$\nabla^2 u(t^*, X(t^*, z^*))X_z(t^*, z^*) = P_z(t^*, z^*).$$

Taking product with θ in both sides of this equality yields a contradiction. ∎

Second proof — Here we restrict ourselves to the case $n = 1$. As in the first proof, we argue by contradiction and suppose that there exists a solution u as described in the statement. Then we take $t^* < T'$, $z^* \in \mathbb{R}$ such that $1 + t^* H''(u_0'(z^*))u_0''(z^*) = 0$. If we fix any $\hat{t} \in]t^*, T'[\,$, we have $1 + \hat{t} H''(u_0'(z^*))u_0''(z^*) < 0$. Then the function $z \to X(\hat{t}; z)$ is decreasing in some neighborhood of z^*, and so there exist $z_1 < z_2$ such that $X(\hat{t}; z_1) > X(\hat{t}; z_2)$. Then there also exists $s < \hat{t}$ such that $X(s; z_1) = X(s; z_2)$. It follows that

$$u_0'(z_i) = P(s; z_i) = u_x(s, X(s; z_i)), \qquad i = 1, 2,$$

which implies $u_0'(z_1) = u_0'(z_2)$. But then we deduce from (1.4.18) that $X(s; z_2) - X(s; z_1) = z_2 - z_1 \neq 0$, in contradiction with our choice of s. ∎

As it is clear from the second proof, the critical time T is the time at which the characteristic lines start to cross. Intuitively speaking, since ∇u is constant along any of these lines, the crossing of characteristics corresponds to the formation of discontinuities in the gradient of u.

Let us now focus on the case of a convex hamiltonian, which is the relevant one for the problem of calculus of variations considered in this chapter. We have seen that in this case the value function given by Hopf's formula is a Lipschitz function solving the Hamilton–Jacobi equation almost everywhere. We now show some important connections between the method of characteristics and the problem of calculus of variations. We consider L, u_0 satisfying assumptions (1.2.6). In addition, we assume that L is strictly convex and that u_0 is differentiable. Let us denote by H the Legendre transform of L; by Theorem A.1.4, H is differentiable. Observe that such hypotheses are weaker than the ones in Theorem 1.4.2, but they are enough to ensure that the functions P, X, U in (1.4.18) are well defined.

Our first remark is that minimizers for the problem in calculus of variations are also characteristic curves for the Hamilton–Jacobi equation.

Proposition 1.4.5 *Let $L, u_0 : \mathbb{R}^n \to \mathbb{R}$ satisfy assumptions* (1.2.6)*; suppose in addition that L is strictly convex and that u_0 is of class C^1. Given $(t,x) \in [0,\infty[\times\mathbb{R}^n$, let $y(\cdot)$ be a minimizer for problem* (1.1.1). *Then $y(\cdot) = X(\cdot, z)$ for some $z \in \mathbb{R}^n$.*

Proof — Let $y(\cdot)$ be a minimizer for (t,x) and let us set $z = y(0)$. In the proof of Hopf's formula (Theorem 1.2.1) we have seen that y is linear and so $y(s) = z + s(x-z)/t$. In addition, z is a minimizer for the right–hand side of (1.2.7), which implies that $DL((x-z)/t) = \nabla u_0(z)$. By property (A.1.5) of the Legendre transform, this implies that $H(\nabla u_0(z)) = (x-z)/t$, and so $y(s) = X(s; z)$. ■

Let us set for $(t,x) \in [0,+\infty[\times\mathbb{R}^n$,

$$\mathcal{Z}(t;x) = \{z \in \mathbb{R}^n \ : \ X(t;z) = x\}. \tag{1.4.20}$$

If the hypotheses of Theorem 1.4.2 are satisfied then such a set is a singleton for small times; in general it is a compact set, by our assumption that u_0 is Lipschitz continuous.

Proposition 1.4.6 *Given L, u_0 as in the previous proposition, let $u(t,x)$ be the value function of problem* (1.1.1). *Then*

$$u(t,x) = \min_{z \in \mathcal{Z}(t;x)} U(t;z). \tag{1.4.21}$$

Proof — As we have seen in the proof of the previous proposition, when one takes the minimum in Hopf's formula (1.2.7) it is enough to consider the values of z such that $x = X(t;z)$. In other words,

$$u(t,x) = \min_{z \in \mathcal{Z}(t;x)} \left[u_0(z) + tL\left(\frac{x-z}{t}\right)\right].$$

The condition $z \in \mathcal{Z}(t;x)$ is equivalent to $DH(\nabla u_0(z)) = (x-z)/t$. Therefore, recalling identity (A.1.4), we have, for $z \in \mathcal{Z}(t;x)$,

$$U(t;z) = u_0(z) + tL\left(\frac{x-z}{t}\right)$$

and so the two sides of (1.4.21) coincide. ∎

In other words, the solution given by Hopf's formula is a suitable selection of the multivalued function obtained by the method of characteristics. Let us point out that such a behavior is peculiar of Hamilton–Jacobi equations with convex hamiltonian: if H is nonconvex there is in general no way of finding a weak solution by taking such a selection.

Corollary 1.4.7 *Let $L, u_0 \in C^2(\mathbb{R}^n)$ satisfy assumptions* (1.2.6). *Suppose in addition that D^2L is positive definite everywhere and that $\nabla^2 u_0$ is uniformly bounded. Then there exists $T^* > 0$ such that the value function of problem* (1.1.1) *is of class C^2 in $[0,T^*[\,\times\mathbb{R}^n$. If u_0 is convex, then $T^* = +\infty$ and so the value function is C^2 everywhere. In addition, given (t,x) with $t < T^*$, there exists a unique minimizer for (t,x) and is given by*

$$y(s) = x + (s-t)DH(\nabla u(t,x)). \tag{1.4.22}$$

Proof — By Theorem A.1.5 the Legendre transform of L is of class C^2. Hence, Theorem 1.4.2 ensures the existence of $T^* > 0$ such that $\mathcal{Z}(t,x)$ is a singleton for $t < T^*$ and such that the method of characteristics yields a smooth solution of the Hamilton–Jacobi equation in $[0,T^*[\,\times\mathbb{R}^n$. By the previous theorem, this smooth solution coincides with the value function. As observed in Corollary 1.4.3, $T^* = +\infty$ if u_0 is convex. Finally, (1.4.22) follows from (1.4.19) and (1.4.18). ∎

The previous theorem shows how the dynamic programming approach can provide a solution to the problem of calculus of variations we are considering. Such a result, however, relies on the smoothness of the value function and we have seen that such a property does not hold in general for all times. When one tries to extend this approach to the cases when the value function is only Lipschitz, several difficulties arise. We have already mentioned the non–uniqueness of Lipschitz continuous solutions of the Hamilton–Jacobi equation. It is also not obvious how to restate equation (1.4.22) at the points where u is not differentiable. These issues will motivate many of the topics treated in the remainder of the book.

1.5 Semiconcavity of Hopf's solution

In this section we consider a further regularity property of the solution to (1.3.13) given by Hopf's formula, known as *semiconcavity*. We first give a definition.

Definition 1.5.1 *Let S be a subset of $\mathbb{R}^N$. We say that a function $u : S \to \mathbb{R}$ is* semiconcave with linear modulus *if there exists $C > 0$ such that*

$$\lambda u(y_1) + (1-\lambda)u(y_2) - u(\lambda y_1 + (1-\lambda)y_2) \leq \lambda(1-\lambda)C|y_1 - y_2|^2 \quad (1.5.23)$$

for any pair $y_1, y_2 \in S$, such that the segment $[y_1, y_2]$ is contained in S and for any $\lambda \in [0,1]$.

Often in the literature the above functions are called semiconcave, without further specifications. We will use instead the term semiconcave for a broader class of functions (see Definition 2.1.1), defined replacing the quantity $|y_1 - y_2|^2$ in the right–hand side of (1.5.23) by $|y_1 - y_2|\omega(|y_1 - y_2|)$, with $\omega(\rho)$ any function such that $\omega(\rho) \to 0$ as $\rho \to 0$. The function ω will be called modulus of semiconcavity, and this explains our choice of terminology in the previous definition. We denote by $\mathrm{SCL}(S)$ the class of functions which are semiconcave with a linear modulus in every compact subset of S.

The interest of the above definition is made clearer by the following characterization.

Proposition 1.5.2 *Given $u : A \to \mathbb{R}$, with $A \subset \mathbb{R}^N$ open set, and given $C \geq 0$, the following properties are equivalent:*

(a) *inequality* (1.5.23) *is satisfied;*

(b) *the function $y \to u(y) - C|y|^2$ is concave in every convex subset of A;*

(c) *$u \in C(A)$ and satisfies $u(y+h) + u(y-h) - 2u(y) \leq 2C|h|^2$ for any y, h such that $[y-h, y+h] \subset A$;*

(d) *for any $\nu \in \mathbb{R}^N$ such that $|\nu| = 1$ we have $\partial^2_{\nu\nu} u \leq 2C$ in A in the sense of distributions.*

Proof — We notice that, for any $y_1, y_2 \in \mathbb{R}^N$,

$$\begin{aligned} &\lambda|y_1|^2 + (1-\lambda)|y_2|^2 - |\lambda y_1 + (1-\lambda)y_2|^2 \qquad (1.5.24)\\ &= \lambda(1-\lambda)(|y_1|^2 + |y_2|^2 - 2\langle y_1, y_2\rangle) = \lambda(1-\lambda)|y_1 - y_2|^2. \end{aligned}$$

This shows that inequality (1.5.23) holds if and only if the function $y \to u(y) - C|y|^2$ is concave. Therefore (a) and (b) are equivalent.

A similar computation shows that, if (c) holds, then the function $v(y) = u(y) - C|y|^2$ is continuous and satisfies $v(y+h)+v(y-h)-2v(y) \leq 0$ for any y, h such that $[y-h, y+h] \subset A$. It is well known that this property implies the concavity of v in the convex subsets of A. Therefore (c) implies (b). On the other hand, (a) obviously implies (c) and so the three properties are equivalent.

To conclude the proof, let us observe that $\partial^2_{\nu\nu}|y|^2 \equiv 2$ for any unit vector ν. Thus u satisfies property (d) if and only if the function $v(y) = u(y) - C|y|^2$ satisfies $\partial^2_{\nu\nu} v \leq 0$ in the sense of distributions for any unit vector ν. Since this property characterizes concave functions, we deduce the equivalence of (b) and (d). ∎

We now prove the semiconcavity of the function u defined by Hopf's formula (1.2.7). We assume throughout that assumptions (1.2.6) hold.

Theorem 1.5.3 *Suppose that*

$$\begin{aligned} &(i)\ L \in C^2(\mathbb{R}^n),\ D^2L(q) \leq \frac{2}{\alpha} I \quad \forall\, q \in \mathbb{R}^n \\ &(ii)\ u_0(x+h) + u_0(x-h) - 2u_0(x) \leq C_0|h|^2, \quad \forall\, x, h \in \mathbb{R}^n \end{aligned} \tag{1.5.25}$$

for suitable constants $\alpha > 0$, $C_0 \geq 0$. Then there exists a constant $C_1 \geq 0$ such that

$$\begin{aligned} & u(t+s, x+h) + u(t-s, x-h) - 2u(t,x) \\ \leq \ & \frac{2tC_0}{2t + \alpha(t^2 - s^2)C_0}(|h| + C_1|s|)^2 \end{aligned} \tag{1.5.26}$$

for all $t > 0$, $s \in\,]-t, t[\,$, $x, h \in \mathbb{R}^n$.

Proof — For fixed t, s, x, h as in the statement of the theorem, let us choose $\hat{x} \in \mathbb{R}^n$ such that

$$u(t,x) = tL\left(\frac{x - \hat{x}}{t}\right) + u_0(\hat{x}).$$

We set, for $\lambda \geq 0$,

$$x^+_\lambda = \hat{x} + \lambda\left(h - s\frac{x - \hat{x}}{t}\right), \qquad x^-_\lambda = \hat{x} - \lambda\left(h - s\frac{x - \hat{x}}{t}\right).$$

Recalling also estimate (1.2.10) from the proof of Theorem 1.2.2 we have, for some $C_1 \geq 0$,

$$\begin{aligned} &(i) && \frac{x^+_\lambda + x^-_\lambda}{2} = \hat{x} \\ &(ii) && x^+_\lambda - x^-_\lambda = 2\lambda\left(h - s\frac{x - \hat{x}}{t}\right) \\ &(iii) && |x^+_\lambda - x^-_\lambda| \leq 2\lambda(|h| + C_1|s|). \end{aligned} \tag{1.5.27}$$

Using Hopf's formula (1.2.7) we can estimate

$$\begin{aligned} &u(t+s,x+h)+u(t-s,x-h)-2u(t,x)\\ &\leq\ 2t\left[\frac{t+s}{2t}L\left(\frac{x+h-x_\lambda^+}{t+s}\right)+\frac{t-s}{2t}L\left(\frac{x-h-x_\lambda^-}{t-s}\right)-L\left(\frac{x-\hat{x}}{t}\right)\right]\\ &\quad +u_0(x_\lambda^+)+u_0(x_\lambda^-)-2u_0(\hat{x}). \end{aligned} \tag{1.5.28}$$

From (1.5.27)(iii) and (1.5.25)(ii) it follows that

$$u_0(x_\lambda^+)+u_0(x_\lambda^-)-2u_0(\hat{x})\leq C_0\lambda^2(|h|+C_1|s|)^2. \tag{1.5.29}$$

We now take $q_0,q_1\in\mathbb{R}^n$, $\theta\in[0,1]$ and set $q_\theta=\theta q_1+(1-\theta)q_0$. Using assumption (1.5.25)(i) and the equivalence between (a) and (d) in Proposition 1.5.2, we have

$$\theta L(q_1)+(1-\theta)L(q_0)-L(q_\theta)\leq\frac{\theta(1-\theta)}{\alpha}|q_1-q_0|^2. \tag{1.5.30}$$

Therefore, observing that

$$\frac{t+s}{2t}\frac{x+h-x_\lambda^+}{t+s}+\frac{t-s}{2t}\frac{x-h-x_\lambda^-}{t-s}=\frac{x-\hat{x}}{t},$$

we obtain from (1.5.30)

$$\begin{aligned} &\frac{t+s}{2t}L\left(\frac{x+h-x_\lambda^+}{t+s}\right)+\frac{t-s}{2t}L\left(\frac{x-h-x_\lambda^-}{t-s}\right)-L\left(\frac{x-\hat{x}}{t}\right)\\ &\qquad\leq\ \frac{t^2-s^2}{\alpha(2t)^2}\left|\frac{x+h-x_\lambda^+}{t+s}-\frac{x-h-x_\lambda^-}{t-s}\right|^2. \end{aligned} \tag{1.5.31}$$

In addition

$$\begin{aligned} &\frac{x+h-x_\lambda^+}{t+s}-\frac{x-h-x_\lambda^-}{t-s}\\ &=\ \frac{(t-s)(x+h)-(t+s)(x-h)}{t^2-s^2}-\frac{(t-s)x_\lambda^+-(t+s)x_\lambda^-}{t^2-s^2}\\ &=\ 2\frac{th-sx+s\hat{x}}{t^2-s^2}-\frac{2\lambda t}{t^2-s^2}\left(h-s\frac{x-\hat{x}}{t}\right)\\ &=\ \frac{2(1-\lambda)t}{t^2-s^2}\left(h-s\frac{x-\hat{x}}{t}\right), \end{aligned}$$

which implies, recalling again (1.2.10)

$$\left|\frac{x+h-x_\lambda^+}{t+s}-\frac{x-h-x_\lambda^-}{t-s}\right|\leq\frac{2(1-\lambda)t}{t^2-s^2}(|h|+C_1|s|). \tag{1.5.32}$$

From (1.5.28), (1.5.29), (1.5.31) and (1.5.32) we obtain, for all $\lambda \geq 0$,

$$\begin{aligned} u(t+s,x+h) + u(t-s,x-h) - 2u(t,x) \\ \leq \ C(\lambda)[|h| + C_1|s|]^2, \end{aligned}$$

where

$$C(\lambda) = \frac{2t}{\alpha(t^2-s^2)}(1-\lambda)^2 + C_0\lambda^2.$$

It is easily checked that

$$\min_{\lambda \geq 0} C(\lambda) = \frac{2tC_0}{2t + \alpha C_0(t^2-s^2)}, \tag{1.5.33}$$

and this proves (1.5.26). ∎

Corollary 1.5.4 *Let assumptions* (1.2.6) *be satisfied and let u be the function defined by Hopf's formula* (1.2.7).

(i) *If L satisfies property* (1.5.25)(i) *for some $\alpha > 0$, then there exists C_1 such that*

$$u(t+s,x+h) + u(t-s,x-h) - 2u(t,x) \leq \frac{2t}{\alpha(t^2-s^2)}(|h| + C_1|s|)^2.$$

(ii) *If u_0 satisfies* (1.5.25)(ii) *for some $C_0 > 0$, then there exists C_1 such that*

$$u(t+s,x+h) + u(t-s,x-h) - 2u(t,x) \leq C_0(|h| + C_1|s|)^2.$$

(iii) *If u_0 is concave, then so is u.*

Proof. Formally, statements (i) and (ii) are obtained by letting respectively $C_0 \to \infty$ and $\alpha \downarrow 0$ in estimate (1.5.26). A precise motivation can be obtained by a suitable adaptation of the proof of the previous theorem. To prove (i), we apply inequality (1.5.28) with $\lambda = 0$. With this choice of λ we have $x_\lambda^+ = x_\lambda^- = \hat{x}$; thus, the terms with u_0 drop out while the terms involving L can be estimated as before. To prove (ii) we choose instead $\lambda = 1$; then it is easily checked that the total contribution of the terms with L in (1.5.28) is zero, while the remaining terms can be estimated as before. Finally, statement (iii) follows from (ii) taking $C_0 = 0$. ∎

We now show that the semiconcavity property singles out Hopf's solution among all Lipschitz continuous solutions of the Hamilton–Jacobi equation.

Theorem 1.5.5 *Let $H \in C^2(\mathbb{R}^n)$ be convex and let that $u_1, u_2 \in \mathrm{Lip}\,(\overline{Q}_T)$ be solutions of* (1.3.13) *such that, for any $t > 0$ and $i = 1, 2$,*

$$u_i(t, x+h) + u_i(t, x-h) - 2u_i(t, x) \leq C\left(1 + \frac{1}{t}\right)|h|^2, \qquad x, h \in \mathbb{R}^n \tag{1.5.34}$$

for some $C > 0$. Then $u_1 = u_2$ everywhere in $\overline{Q}_T$.

We first recall an elementary algebraic property.

Lemma 1.5.6 *Let A, B be two symmetric $n \times n$ matrices. Suppose that $0 \leq A \leq \Lambda I$ and $B \leq kI$ for some $\Lambda, k > 0$. Then*

$$\mathrm{trace}(AB) \leq nk\Lambda.$$

Proof — Let $\{e_1, \ldots, e_n\}$ be on orthonormal basis of eigenvectors of B and let $\{\mu_1, \ldots, \mu_n\}$ be the corresponding eigenvalues. Then our assumptions imply

$$\mu_i \leq k, \qquad 0 \leq \langle e_i, Ae_i\rangle \leq \Lambda, \qquad \forall\, i = 1, \ldots, n.$$

It follows

$$\begin{aligned} \mathrm{trace}(AB) &= \sum_{i=1}^n \langle e_i, ABe_i\rangle = \sum_{i=1}^n \mu_i \langle e_i, Ae_i\rangle \\ &\leq k\sum_{i=1}^n \langle e_i, Ae_i\rangle \leq nk\Lambda. \; \blacksquare \end{aligned}$$

Proof of Theorem 1.5.5 — Setting $\bar{u}(t, x) = u_1(t, x) - u_2(t, x)$ we have, for $(t, x) \in Q_T$ a.e.,

$$\bar{u}_t(t, x) = H(\nabla u_2(t, x)) - H(\nabla u_1(t, x)) = -b(t, x)\nabla\bar{u}(t, x),$$

where

$$b(t, x) = \int_0^1 DH(r\nabla u_2(t, x) + (1-r)\nabla u_1(t, x))dr.$$

Let $\phi : \mathbb{R} \to [0, \infty[$ be a C^1 function to be fixed later. Setting $v = \phi(\bar{u})$ we obtain, for $(t, x) \in Q_T$ a.e.,

$$v_t(t, x) + b(t, x) \cdot \nabla v(t, x) = 0. \tag{1.5.35}$$

Let $k \in C^\infty(\mathbb{R}^n)$ be a nonnegative function whose support is contained in the unit sphere and whose integral is 1. We define, for $i = 1, 2$, $\varepsilon > 0$ and $(t, x) \in Q_T$,

$$u_i^\varepsilon(t, x) = \frac{1}{\varepsilon^n}\int_{\mathbb{R}^n} u_i(t, y)k\left(\frac{x-y}{\varepsilon}\right)dy.$$

By well-known properties of convolutions the functions u_i^ε are of class C^∞ with respect to x and satisfy

$$\begin{aligned} &(i)\ |\nabla u_i^\varepsilon| \leq \mathrm{Lip}\,(u_i) \\ &(ii) \text{ for all } t>0,\ \nabla u_i^\varepsilon(t,\cdot) \to \nabla u_i(t,\cdot) \text{ a.e. as } \varepsilon \to 0. \end{aligned} \tag{1.5.36}$$

In addition, u_i^ε satisfies the semiconcavity estimate (1.5.34) for all $\varepsilon > 0$. Therefore, using property (d) in Proposition 1.5.2,

$$\nabla^2 u_i^\varepsilon(t,x) \leq C\left(1+\frac{1}{t}\right) I, \qquad (t,x) \in Q_T. \tag{1.5.37}$$

Setting

$$b_\varepsilon(t,x) = \int_0^1 DH(r\nabla u_2^\varepsilon(t,x) + (1-r)\nabla u_1^\varepsilon(t,x))dr,$$

we can rewrite equation (1.5.35) in the form

$$v_t + \mathrm{div}\,(vb_\varepsilon) = (\mathrm{div}\, b_\varepsilon)v + (b_\varepsilon - b)\cdot \nabla v.$$

Let us set

$$R = \max\{|DH(p)| \ :\ |p| \leq \max(\mathrm{Lip}\,(u_1), \mathrm{Lip}\,(u_2))\},$$

$$\Lambda = \max\{||D^2 H(p)|| \ :\ |p| \leq \max(\mathrm{Lip}\,(u_1), \mathrm{Lip}\,(u_2))\}.$$

By (1.5.37) and Lemma 1.5.6 it follows

$$\begin{aligned} \mathrm{div}\, b_\varepsilon &= \\ &= \int_0^1 \sum_{k,l=1}^n \frac{\partial^2 H}{\partial p_k \partial p_l}(r\nabla u_2^\varepsilon + (1-r)\nabla u_1^\varepsilon)\left(r\frac{\partial^2 u_2^\varepsilon}{\partial x_k \partial x_l} + (1-r)\frac{\partial^2 u_1^\varepsilon}{\partial x_k \partial x_l}\right) dr \\ &\leq n\Lambda C(1+1/t). \end{aligned} \tag{1.5.38}$$

This is the estimate where the semiconcavity assumption on the u_i's is used.

Given $(t_0, x_0) \in Q_T$, let us introduce the function

$$E(t) = \int_{B(x_0, R(t_0-t))} v(t,x)dx, \qquad 0 \leq t \leq t_0$$

and the cone

$$\mathcal{C} = \{(t,x) \ :\ 0 \leq t \leq t_0,\ x \in B(x_0, R(t_0-t))\},$$

where we use the notation

$$B(x_0, R(t_0-t)) = \{x \in \mathbb{R}^n \ :\ |x-x_0| < R(t_0-t)\}.$$

Then E is Lipschitz continuous and satisfies, for a.e. $t > 0$,

$$\begin{aligned}
E'(t) &= \int_{B(x_0,R(t_0-t))} v_t dx - R\int_{\partial B(x_0,R(t_0-t))} v dS \\
&= \int_{B(x_0,R(t_0-t))} \{-\mathrm{div}\,(vb_\varepsilon) + (\mathrm{div}\,b_\varepsilon)v + (b_\varepsilon - b)\cdot\nabla v\}dx \\
&\quad -R\int_{\partial B(x_0,R(t_0-t))} v dS \\
&= -\int_{\partial B(x_0,R(t_0-t))} v(b_\varepsilon\cdot\nu + R)dS \\
&\quad +\int_{B(x_0,R(t_0-t))} \{(\mathrm{div}\,b_\varepsilon)v + (b_\varepsilon - b)\cdot\nabla v\}dx \\
&\leq \int_{B(x_0,R(t_0-t))} \{(\mathrm{div}\,b_\varepsilon)v + (b_\varepsilon - b)\cdot\nabla v\}dx \\
&\leq n\Lambda C\left(1+\frac{1}{t}\right)E(t) + \int_{B(x_0,R(t_0-t))} (b_\varepsilon - b)\cdot\nabla v\, dx.
\end{aligned}$$

Letting $\varepsilon \to 0$ we obtain, by (1.5.36)(ii),

$$E'(t) \leq n\Lambda C\left(1+\frac{1}{t}\right)E(t), \qquad t\in\,]0,t_0[\text{ a.e.} \tag{1.5.39}$$

We now choose the function ϕ in the definition of v. We fix $\eta > 0$ and define ϕ such that $\phi(z) = 0$ if $|z| \leq \eta[\mathrm{Lip}\,(u_1) + \mathrm{Lip}\,(u_2)]$ and $\phi(z) > 0$ otherwise. Then the assumption that $u_1 = u_2$ for $t = 0$ implies that $v(t,x) = 0$ if $t \leq \eta$. Thus we obtain, from (1.5.39) and the Gronwall's inequality,

$$0 \leq E(t) \leq E(\eta)\exp\left(\int_\eta^t n\Lambda C\left(1+\frac{1}{s}\right)ds\right) = 0$$

for any $t \in [0,t_0]$. Hence

$$|u_2 - u_1| \leq \eta\,[\,\mathrm{Lip}\,(u_1) + \mathrm{Lip}\,(u_2)]\ \text{on } \mathcal{C}.$$

By the arbitrariness of $\eta > 0$, we deduce that u_1 and u_2 coincide in $\mathcal{C}$, and, in particular, $u_1(t_0,x_0) = u_2(t_0,x_0)$. ■

We can summarize the results we have obtained about problem (1.3.13) as follows.

Corollary 1.5.7 *Let H, u_0 satisfy hypotheses* (1.3.14) *. Suppose in addition that $H \in C^2(\mathbb{R}^n)$ and that either H is uniformly convex or u_0 is semiconcave with a linear modulus. Then there exists a unique $u \in \mathrm{Lip}\,([0,\infty[\,\times\mathbb{R}^n)$ which solves problem* (1.3.13) *almost everywhere and which satisfies*

$$u(t,x+h) + u(t,x-h) - 2u(t,x) \leq C\left(1+\frac{1}{t}\right)|h|^2, \quad x,h\in\mathbb{R}^n, t>0, \tag{1.5.40}$$

for a suitable $C > 0$. In addition, u is given by Hopf's formula (1.2.7), *taking as L the Legendre transform of H.*

Proof — We recall that, if H is uniformly convex and L is the Legendre transform of H, then L satisfies property (1.5.25)-(i) (see (A.1.6) in the appendix). The result then follows from Theorem 1.3.1, Corollary 1.5.4 and Theorem 1.5.5. ∎

1.6 Semiconcavity and entropy solutions

In this section we analyze the connection between Hamilton–Jacobi equations and another important class of partial differential equations, the so–called *hyperbolic conservation laws.*

Given $H \in C^2(\mathbb{R})$ and $T \in]0, \infty]$, let $u \in C^2([0, T[\times \mathbb{R})$ be a solution of

$$u_t + H(u_x) = 0. \tag{1.6.41}$$

Then, if we set $v(t, x) := u_x(t, x)$ and differentiate (1.6.41) with respect to x, we see that v solves

$$v_t + H(v)_x = 0. \tag{1.6.42}$$

Conversely, if v is a C^1 solution of (1.6.42) and we set, for a given $x_0 \in \mathbb{R}$,

$$u(t, x) := \int_{x_0}^{x} v(t, y)\, dy - \int_0^t H(v(s, x_0))\, ds \tag{1.6.43}$$

then u is a solution of (1.6.41).

Equation (1.6.42) is called a hyperbolic conservation law. We observe that the above transformation can no longer be done when one considers equations in several space dimensions or systems of equations.

Just like Hamilton–Jacobi equations, conservation laws do not possess in general global smooth solutions. It is interesting therefore to investigate the relation between the two equations when dealing with generalized solutions. Conservation laws are in divergence form and it is natural to consider solutions in the sense of distributions. More precisely, let us consider equation (1.6.42) with initial data

$$v(0, x) = v_0(x), \qquad x \in \mathbb{R} \text{ a.e.}, \tag{1.6.44}$$

where $v_0 \in L^\infty(\mathbb{R})$. Then we say that $v \in L^\infty([0, T[\times \mathbb{R})$ is a *weak solution* of problem (1.6.42)–(1.6.44) if it satisfies

$$\begin{aligned} &\int_0^T \int_{\mathbb{R}} [v(t, x)\phi_t(t, x) + H(v(t, x))\phi_x(t, x)]\, dxdt \\ &\qquad = -\int_{\mathbb{R}} v_0(x)\phi(0, x)\, dx, \qquad \forall \phi \in C_0^\infty([0, T[\times \mathbb{R}), \end{aligned} \tag{1.6.45}$$

Here C_0^∞ denotes the class of C^∞ functions with compact support.

Proposition 1.6.1 *Let* $u \in \mathrm{Lip}\,([0,T[\,\times \mathbb{R})$ *be an almost everywhere solution of* (1.6.41) *and let* $v(t,x) = u_x(t,x)$. *Then* v *is a weak solution of* (1.6.42)–(1.6.44) *with* $v_0(x) = u_x(0,x)$.

Conversely, let $v \in L^\infty([0,T[\times \mathbb{R})$ *be a weak solution of* (1.6.42)–(1.6.44) *for some* $v_0 \in L^\infty(\mathbb{R})$. *Then, there exists* $u \in \mathrm{Lip}\,([0,T[\,\times \mathbb{R})$ *which satisfies* (1.6.41) *almost everywhere and is such that* $u_x(t,x) = v(t,x)$ *for a.e.*$(t,x) \in [0,T[\,\times \mathbb{R}$ *and* $u_x(0,x) = v_0(x)$ *for a.e.* $x \in \mathbb{R}$.

Proof — Let us first assume that we have a Lipschitz continuous function u which solves (1.6.41) almost everywhere. Then, if $\phi \in C_0^\infty([0,T[\,\times \mathbb{R})$, we have

$$\begin{aligned} \int_0^T \int_{\mathbb{R}} u_x(t,x)\phi_t(t,x) &= -\int_0^T \int_{\mathbb{R}} u(t,x)\phi_{xt}(t,x) \\ &= \int_0^T \int_{\mathbb{R}} u_t(t,x)\phi_x(t,x) + \int_{\mathbb{R}} u(0,x)\phi_x(0,x) \end{aligned}$$

Therefore, if we set $v := u_x$, we obtain

$$\begin{aligned} &\int_0^T \int_{\mathbb{R}} [v(t,x)\phi_t(t,x) + H(v(t,x))\phi_x(t,x)]\,dxdt \\ &= \int_0^T \int_{\mathbb{R}} [u_t(t,x) + H(u_x(t,x))]\phi_x(t,x)\,dxdt + \int_{\mathbb{R}} u(0,x)\phi_x(0,x)\,dx \\ &= -\int_{\mathbb{R}} u_x(0,x)\phi(0,x)\,dx. \end{aligned}$$

Thus, v is a weak solution of (1.6.42) with initial value $u_x(0,x)$.

Conversely, let us suppose that v is a weak solution of (1.6.42)–(1.6.44). Let us introduce the sets

$$\mathcal{A} = \{x \in \mathbb{R} : x \text{ is a Lebesgue point for } v(t,\cdot) \text{ for } t \in]0,T[\text{ a.e.}\}$$

$$\mathcal{B} = \{t \in]0,T[: t \text{ is a Lebesgue point for } v(\cdot,x) \text{ for } x \in \mathbb{R} \text{ a.e.}\}.$$

By well–known results of measure theory, both sets have a complement of measure zero. For given $x_0, x_1 \in \mathcal{A}$ and $t_0, t_1 \in \mathcal{B}$, let us take a sequence $\phi_n \in C_0^\infty(\,]0,T[\,\times \mathbb{R})$ converging to the characteristic function of the rectangle $[t_0,t_1] \times [x_0,x_1]$. Passing to the limit in (1.6.45) we obtain

$$\int_{x_0}^{x_1} [v(t_1,x) - v(t_0,x)]dx + \int_{t_0}^{t_1} [H(v(t,x_1)) - H(v(t,x_0))]dt = 0. \quad (1.6.46)$$

Let us now fix any $x_0 \in \mathcal{A}$ and define u as in (1.6.43). Then, if $x_1 \in \mathcal{A}$, we have $u_x(t,x_1) = v(t,x_1)$ for a.e. t. Therefore we obtain from (1.6.43) and (1.6.46)

$$u(t_1,x_1) - u(t_0,x_1) + \int_{t_0}^{t_1} H(u_x(t,x_1))dt = 0$$

for all $x_1 \in \mathcal{A}$ and $t_0, t_1 \in \mathcal{B}$. This shows that u solves (1.6.41) a.e.; the other properties are direct consequences of the definition. ∎

This result suggests that, in the case that H is convex, we can extend to conservation laws the existence and uniqueness result for Hamilton–Jacobi equations of the previous chapter, provided we require that the solutions satisfy a "differentiated" form of the semiconcavity inequality.

Corollary 1.6.2 *Given $H \in C^2(\mathbb{R})$ such that $H''(p) \geq c > 0$ and given $v_0 \in L^\infty(\mathbb{R})$ there exists a unique $v \in L^\infty([0,\infty[\times\mathbb{R})$ which is a weak solution of problem* (1.6.42)–(1.6.44) *and satisfies a.e.*

$$v(y,t) - v(x,t) \leq \frac{y-x}{ct}, \qquad x < y, \ t > 0. \tag{1.6.47}$$

Proof — We observe that inequality (1.6.47) is equivalent to the property that for all $t > 0$, $x \to v(x,t) - x/(ct)$ is nonincreasing. Therefore, if $u \in \mathrm{Lip}\,([0,\infty[\,\times\mathbb{R})$ is such that $u_x = v$ a.e., then v satisfies (1.6.47) if and only if, for all $t > 0$, $x \to u(x,t) - x^2/(2ct)$ is concave. The result then follows from Corollary 1.5.4, Theorem 1.5.5 and Proposition 1.6.1, taking into account also the characterizations of semiconcavity given in Proposition 1.5.2. ∎

Condition (1.6.47) is called *Oleinik one–sided Lipschitz inequality*. A weak solution of (1.6.41)–(1.6.44) satisfying the Oleinik inequality is called an *entropy solution*. There are other equivalent definitions of entropy solutions of scalar conservation laws. The most powerful one was given by Kruzhkov [61] replacing equality (1.6.45) by a family of integral inequalities. Kruzhkov's definition covers also the cases where H is not convex and where the space dimension is greater than one.

Bibliographical notes. The dynamic programming approach for the analysis of control problems was started by Bellman [21] in the 50's and it gradually became a standard tool in control theory (see e.g. the books [51, 68, 47]). However, in the special case of calculus of variations considered in this chapter, many of the ideas we have described were already present in much older works. For instance, the connection between variational problems and Hamilton–Jacobi equations is well-known in classical mechanics. In this context, an approach based on the introduction of the value function was already used by Carathéodory [30].

The representation formula (1.2.7) was given by Conway and Hopf in [34, 53]. An analogous formula had been previously found for solution of conservation laws by Lax [65] and by Oleinik [70].

The first uniqueness results for semiconcave solutions of Hamilton–Jacobi equations were obtained by Kruzhkov and Douglis (see e.g. [59, 40, 60]). Our presentation of Theorem 1.5.5 follows [42].

The semiconcavity of solutions of evolution Hamilton–Jacobi equations was initially studied only with respect to the space variables, since such a property is strong enough for obtaining a uniqueness result. More recently, the work of Cannarsa and Soner [27, 28] underlined the interest of the semiconcavity with respect the whole set of variables for the analysis of the singular set.

The connection between hyperbolic conservation laws and Hamilton–Jacobi equations has been known since the beginning of the study of the two theories. However, a precise statement of the equivalence of the problems like in Proposition 1.6.1 was first given in [34]. The uniqueness statement of Corollary 1.6.2 was proved by Oleinik in [70]. In our exposition we have reversed the historical order, since Oleinik's uniqueness theorem was proved before the corresponding statement for Hamilton–Jacobi equations. Oleinik's paper inspired the aforementioned studies by Kruzhkov and Douglis and in particular suggested that the semiconcavity property could provide a uniqueness criterion for Hamilton–Jacobi equations.

Chapter 2

Definition and examples of semiconcave functions

This chapter is devoted to the basic properties of the class of semiconcave functions. Our analysis will be focused in particular on equivalent characterizations of semiconcavity, on the differential properties of semiconcave functions, and on some significant examples.

In Section 2.1 we give the definition of semiconcavity in a more general way than we did in the previous chapter; we derive some basic consequences of this property, like the local Lipschitz continuity. We show that the infimum of a family of smooth functions satisfying some suitable uniform bound is semiconcave; such a property is an easy consequence of the definition, but explains why semiconcave functions appear in a natural way in the analysis of minimization problems.

Then we consider some examples in Section 2.2, like the distance function or the solutions to certain partial differential equations. We analyze the role of semiconcavity in the so–called vanishing viscosity method for Hamilton–Jacobi equations, a method which has been extensively used in the past to obtain existence results of generalized solutions.

In Section 2.3 we recall some properties which are peculiar of semiconcave functions with a linear modulus, like Alexandroff's theorem or Jensen's lemma.

In Section 2.4 we investigate the relation between viscous Hamilton-Jacobi equations and the heat equation induced by the Cole-Hopf transformation, showing that semiconcavity corresponds to the Li-Yau differential Harnack inequality for the heat equation. Finally, in Section 2.5 we analyze the relation between semiconcavity and a generalized one–sided estimate, a property which will be applied later in the book to prove semiconcavity of viscosity solutions.

2.1 Definition and basic properties

Throughout the chapter S will be a subset of $\mathbb{R}^n$.

Definition 2.1.1 *We say that a function* $u : S \to \mathbb{R}$ *is* semiconcave *if there exists a nondecreasing upper semicontinuous function* $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ *such that* $\lim_{\rho \to 0^+} \omega(\rho) = 0$ *and*

$$\lambda u(x) + (1-\lambda)u(y) - u(\lambda x + (1-\lambda)y) \leq \lambda(1-\lambda)|x-y|\omega(|x-y|) \quad (2.1.1)$$

for any pair $x, y \in S$*, such that the segment* $[x, y]$ *is contained in* S *and for any* $\lambda \in [0, 1]$.

A function ω satisfying the properties of the previous definition is called a *modulus of semiconcavity* for u in S. We call *locally semiconcave* a function which is semiconcave on every compact subset of its domain of definition. We denote by $\mathrm{SC}\,(S)$ the space of all locally semiconcave functions in S. As in the previous chapter, we use the notation $\mathrm{SCL}\,(S)$ for the functions which are locally semiconcave with a linear modulus. Finally, a function v is called *semiconvex* if $-v$ is semiconcave.

Inequality (2.1.1) is a weakened form of the concavity inequality: we do not require the right–hand side to be nonnegative, as for concave functions, but only to be less than a term of order higher than $|x - y|$. In Chapter 1 (see Definition 1.5.1) we have already seen the definition of semiconcavity corresponding to the case of ω linear, which is the most common in the literature (see e.g. [67], [37], [17]). Semiconcave functions with a linear modulus enjoy some special properties, as we will see in the following. However, many significant results about semiconcave functions hold for an arbitrary modulus, and for this reason it is convenient for us to work with the more general definition given above.

Obviously, a concave function is also semiconcave. Another immediate example of semiconcave functions is given by the smooth ones.

Proposition 2.1.2 *Let* $u \in C^1(A)$*, with* A *open. Then both* u *and* $-u$ *are locally semiconcave in* A *with modulus equal to the modulus of continuity of* Du.

Proof — Let ω_K denote the modulus of continuity of Du in a compact set $K \subset A$ and let $x, y \in K$ be such that the segment $[x, y] \subset K$. We have

$$\begin{aligned}
&\lambda u(x) + (1-\lambda)u(y) - u(\lambda x + (1-\lambda)y) \\
&\quad = \lambda[u(x) - u(x + (1-\lambda)(y-x)] + (1-\lambda)[u(y) - u(y + \lambda(x-y))] \\
&\quad = -\lambda\langle Du(\xi_1), (1-\lambda)(y-x)\rangle - (1-\lambda)\langle Du(\xi_2), \lambda(x-y)\rangle,
\end{aligned}$$

for suitable points ξ_1 and ξ_2 belonging to $[x, y]$. Since $|\xi_1 - \xi_2| \le |x - y|$ we obtain

$$\begin{aligned} & |\lambda u(x) + (1-\lambda)u(y) - u(\lambda x + (1-\lambda)y)| \\ = & \ \lambda(1-\lambda)|\langle Du(\xi_1) - Du(\xi_2), x - y\rangle| \\ \le & \ \lambda(1-\lambda)|x-y|\omega_K(|x-y|). \ \blacksquare \end{aligned}$$

The converse result is also true, that is, a function which is both semiconcave and semiconvex is continuously differentiable.

Proposition 2.1.3 *If $u : A \to \mathbb{R}$, with A open, is both semiconcave and semiconvex in A, then $u \in C^1(A)$. In addition, if the moduli of semiconcavity and semiconvexity of u have both the form $\omega(\rho) = C\rho^\alpha$ for some $\alpha \in]0, 1]$, then $u \in C^{1,\alpha}_{loc}(A)$.*

Proof — Consider first the case $\alpha = 1$ where the modulus is linear. Then, by Proposition 1.5.2, all directional second derivatives of u are distributions locally bounded both from above and from below, and so they belong to $L^\infty_{loc}(A)$. Now any mixed second derivative is the linear combination of directional ones, as follows:

$$\partial^2_{\xi\eta} u = \frac{1}{4}\left(\partial^2_{(\xi+\eta)(\xi+\eta)} u - \partial^2_{(\xi-\eta)(\xi-\eta)} u\right), \qquad \forall \xi, \eta \in \mathbb{R}^n.$$

We deduce that $D^2 u \in L^\infty_{loc}(A)$ and that $Du \in \mathrm{Lip}_{loc}(A)$.

In the case $\alpha < 1$, the assertion follows from [78, Prop. V.8–V.9] (see [67, p. 259] for an independent proof). The proof of the C^1 regularity of u for a general modulus will be given later in this chapter (see Corollary 3.3.7). $\blacksquare$

The previous result can be used to show that a given function is not semiconcave.

Example 2.1.4 The function $u(x) = |x|$ is not semiconcave in any neighborhood of 0. In fact, u is convex; were u also semiconcave, it would be differentiable by the previous proposition.

Let us now observe an immediate corollary of the definition of semiconcavity.

Proposition 2.1.5 *Let $\{u_\alpha\}_{\{\alpha \in \mathcal{A}\}}$ be a family of functions defined in S and semiconcave with the same modulus ω. Then the function $u := \inf_{\alpha \in \mathcal{A}} u_\alpha$ is also semiconcave in S with modulus ω provided $u > -\infty$.*

Combining Propositions 2.1.2 and 2.1.5 we obtain the following result, which provides interesting examples of semiconcave functions.

Corollary 2.1.6 *The infimum of a family of C^1 functions which are uniformly bounded from below and have equicontinuous gradients is a semiconcave function. If the functions of the family are of class C^2 and have second derivatives uniformly bounded from above, then the infimum is semiconcave with linear modulus.*

Functions defined as infimum of smooth functions occur in the study of optimization problems (see the "lower–C^1" and "lower–C^2" functions in [73] or the "marginal functions" in [17]). We will analyze later in more detail the relation between semiconcave functions and infima of smooth functions (see Section 3.4).

We shall see later (see Theorem 3.4.2) that a converse to this result is also true; that is, any semiconcave function can be represented as the infimum of a suitable family of regular functions.

We have seen in Proposition 1.5.2 that semiconcave functions with a linear modulus can be written as the sum of a concave function and a smooth one. We show now that this property fails when we consider semiconcave functions with a general modulus.

Proposition 2.1.7 *Given $\alpha \in]0,1[$, there exists $u : [0,M] \to \mathbb{R}$ which is semiconcave with modulus $\omega(r) = Cr^\alpha$ for some $M, C > 0$ and that cannot be written in the form $u = u_1 + u_2$ with u_1 concave and $u_2 \in C^1([0,M])$.*

Proof — Let us choose $\beta > 1$ such that $\alpha\beta < 1$. Let us set

$$s_0 = 0, \qquad s_n = \sum_{k=1}^{n} k^{-\beta} \quad (n \geq 1), \qquad L = \lim_{n\to\infty} s_n.$$

We then define a function $v : [0, L+1] \to \mathbb{R}$ by setting

$$v(x) = (x - s_n)^\alpha, \qquad \text{if } x \in [s_n, s_{n+1}[$$

and $v(x) = 0$ if $x \in [L, L+1]$. The function v is discontinuous at s_n for all $n \geq 1$ and continuous elsewhere. Let us observe that

$$\sum_{n=1}^{\infty} [v(s_n-) - v(s_n+)] = \sum_{n=1}^{\infty} n^{-\alpha\beta} = +\infty. \tag{2.1.2}$$

It is also easy to see that v satisfies the one–sided Hölder estimate

$$v(y) - v(x) \leq (y-x)^\alpha, \qquad \forall\, x, y \quad 0 \leq x \leq y \leq L+1. \tag{2.1.3}$$

Let us now set

$$u(x) = \int_0^x v(y)\,dy, \qquad x \in [0, L+1].$$

Using (2.1.3) we obtain, for any x, y with $0 \le x \le y \le L+1$,

$$\begin{aligned}
&\lambda u(x) + (1-\lambda)u(y) - u(\lambda x + (1-\lambda)y) \\
&= \lambda(1-\lambda)(y-x)\int_0^1 [v(y - \theta\lambda(y-x)) - v(x + \theta(1-\lambda)(y-x))]\,d\theta \\
&\le \lambda(1-\lambda)(y-x)\int_0^1 (1-\theta)(y-x)^\alpha d\theta = \lambda(1-\lambda)(y-x)^{1+\alpha}/2.
\end{aligned}$$

Thus, u is semiconcave with modulus $\omega(r) = r^\alpha/2$. Let us now assume that there exist two functions $u_1, u_2 : [0, L+1] \to \mathbb{R}$, u_1 concave and $u_2 \in C^1$ such that $u = u_1 + u_2$. Then setting $v_i = u_i'$ for $i = 1, 2$, we have that $v_2 \in C[0, L+1]$ while v_1 is defined for every $x \neq s_n$ and is decreasing. In addition, $v_1 + v_2 = v$. Since v_2 is continuous, we have for every n

$$v(s_n-) - v(s_n+) = v_1(s_n-) - v_1(s_n+).$$

Since v_1 is decreasing, we deduce from (2.1.2) that $v_1(x) \to -\infty$ as $x \to L^-$. This is a contradiction, since v_1 is defined and decreasing in $[0, L+1]$. Therefore there exist no functions u_1, u_2 as above. This proves the proposition. ■

Like concavity, the semiconcavity property implies the local Lipschitz continuity, as shown in the next theorem.

Theorem 2.1.8 *A semiconcave function $u : S \to \mathbb{R}$ is locally Lipschitz continuous in the interior of S.*

Proof — We divide the proof in different steps. It is not restrictive to assume that S is open.

STEP 1: u is locally bounded from below.

Given $x_0 \in S$, we take a closed cube Q centered at x_0 and contained in S. We denote by L the diameter of Q and by $x_1, x_2, \ldots, x_{2^n}$ the vertices of Q. Let us set

$$m_0 = \min\{u(x_i) \ : \ i = 1, 2, \ldots, 2^n\}.$$

Let x_i and x_j be two consecutive vertices of Q. Using the semiconcavity inequality (2.1.1) we obtain, for any $\lambda \in [0,1]$,

$$\begin{aligned}
&u(\lambda x_i + (1-\lambda)x_j) \\
&\ge \lambda u(x_i) + (1-\lambda)u(x_j) - \lambda(1-\lambda)|x_i - x_j|\omega(|x_i - x_j|) \\
&\ge m_0 - \frac{L\omega(L)}{4}.
\end{aligned}$$

This shows that u is bounded from below on the 1–dimensional faces of Q. We can repeat the procedure by taking any convex combination of two points lying on different 1–dimensional faces, and we obtain that $u \geq m_0 - L\omega(L)/2$ on the 2–dimensional faces. Iterating the procedure n times, we obtain that u bounded from below on the whole cube Q.

STEP 2: u is locally bounded from above.

Given $x_0 \in S$, we consider now $R > 0$ such that $\overline{B(x_0, R)} \subset S$. By the first step, u is greater than some constant m on $\overline{B(x_0, R)}$. Using (2.1.1) we find, for any $x \in \overline{B(x_0, R)}$,

$$\begin{aligned} & \frac{|x-x_0|}{R+|x-x_0|}\, u\left(x_0 - R\frac{x-x_0}{|x-x_0|}\right) + \frac{R}{R+|x-x_0|}\, u(x) - u(x_0) \\ \leq\ & \frac{R|x-x_0|}{R+|x-x_0|}\, \omega(R+|x-x_0|). \end{aligned}$$

Therefore

$$\begin{aligned} u(x) \ &\leq\ -\frac{|x-x_0|}{R}\, m + \frac{R+|x-x_0|}{R}\, u(x_0) + |R|\omega(R+|x-x_0|) \\ &\leq\ |m| + 2|u(x_0)| + R\omega(2R), \end{aligned}$$

which shows that u is bounded from above in $\overline{B(x_0, R)}$.

STEP 3: u is locally Lipschitz continuous.

We first observe that the semiconcavity inequality (2.1.1) implies the following: if z belongs to the segment $[x, y] \subset S$, then

$$\frac{u(x)-u(z)}{|x-z|} - \frac{u(z)-u(y)}{|z-y|} \leq \omega(|x-y|).$$

Given $x_0 \in S$, let us fix $R > 0$ such that $\overline{B(x_0, R)} \subset S$. By the first two steps, there exist constants m, M such that $m \leq u \leq M$ in $\overline{B(x_0, R)}$. Given any $x, y \in B(x_0, R/2)$, consider the straight line through x, y and call x', y' the points of this line at distance R from x_0, ordered in such a way that $x \in [x', y]$ and $y \in [x, y']$. Then we have

$$\frac{u(x')-u(x)}{|x'-x|} - \omega(|x'-y|) \leq \frac{u(x)-u(y)}{|x-y|} \leq \frac{u(y)-u(y')}{|y-y'|} + \omega(|y'-x|),$$

which implies

$$\frac{|u(x)-u(y)|}{|x-y|} \leq \frac{2(M-m)}{R} + \omega(2R). \ \blacksquare$$

Remark 2.1.9 It follows from the last step of the above proof that the Lipschitz constant of u in a given set $B \subset\subset S$ depends only on B, on $\sup_B |u|$ and on the modulus of semiconcavity of u.

Observe that the previous theorem is immediate for semiconcave function with linear modulus. In fact, it is well–known that concave functions are locally Lipschitz continuous and, using Proposition 1.5.2(b), this property can be immediately extended to semiconcave function with linear modulus. However, as we have seen in the above proof, Lipschitz continuity can be proved for arbitrary semiconcave functions by an easy adaptation of the argument used for concave functions. The same remark applies to most of the properties considered in the rest of the chapter.

The next result shows that the linear growth near 0 is a threshold for a modulus of semiconcavity, because the only semiconcave functions whose modulus tends to zero faster than linear are the concave functions.

Proposition 2.1.10 *Let $u : A \to \mathbb{R}$ be a semiconcave function, with A open, and with a modulus ω such that*

$$\lim_{\rho \to 0^+} \frac{\omega(\rho)}{\rho} = 0.$$

Then u is concave on all convex subsets of A.

Proof — We first observe that it suffices to prove the result in the case when u is smooth. In fact, the general case can be obtained by regularizing u, since the convolution of a semiconcave function with a mollifier is semiconcave with the same modulus.

We may therefore assume that $u \in C^2(A)$. Then, for any $x \in A$ and $\nu \in S^{n-1}$,

$$\begin{aligned} \partial^2_{\nu\nu} u(x) &= \lim_{h \to 0^+} \frac{u(x+h\nu) + u(x-h\nu) - 2u(x)}{h^2} \\ &\leq \lim_{h \to 0^+} \frac{h\omega(2h)}{h^2} = 0. \end{aligned}$$

This shows that u is concave in A. ∎

Checking inequality (2.1.1) for any $\lambda \in [0,1]$ may often increase the technical difficulty of proving that a given function is semiconcave. We have seen in Proposition 1.5.2 that, when the modulus is linear, it suffices to prove the inequality for the midpoint of any segment. The same property holds for a large class of moduli, as shown by the next result.

Theorem 2.1.11 *Let $u : S \to \mathbb{R}$ be a continuous function. Suppose that there exists a nondecreasing upper semicontinuous function $\tilde{\omega} : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{\rho\to 0^+} \tilde{\omega}(\rho) = 0$ and*

$$u(x) + u(y) - 2u\left(\frac{x+y}{2}\right) \leq \frac{|x-y|}{2}\tilde{\omega}(|x-y|) \tag{2.1.4}$$

for any x, y such that the segment $[x, y]$ is contained in S. Then u is semiconcave in S with modulus

$$\omega(\rho) = \sum_{h=0}^{\infty} \tilde{\omega}\left(\frac{\rho}{2^h}\right)$$

provided the right–hand side is finite. If $\tilde{\omega}$ is linear then we can take $\omega = \tilde{\omega}$.

Proof — Let x_0, x_1 be such that $[x_0, x_1] \subset S$. We set, for $\lambda \in]0,1[$,

$$x_\lambda = \lambda x_1 + (1-\lambda)x_0.$$

We also define, for $k \geq 1$,

$$D_k := \{x_\lambda \ : \ \lambda = \frac{j}{2^k} \text{ for some } j = 0, 1, \ldots, 2^k\}.$$

We claim that

$$\lambda u(x_1) + (1-\lambda)u(x_0) - u(x_\lambda) \leq \lambda(1-\lambda)L\sum_{h=0}^{k-1} \tilde{\omega}\left(\frac{L}{2^h}\right), \ \forall x_\lambda \in D_k, \tag{2.1.5}$$

where we have set $L := |x_1 - x_0|$. We proceed by induction on k. In the case $k = 1$ inequality (2.1.5) is easily checked; in fact it holds by hypothesis if $\lambda = 1/2$, while it is trivial if $\lambda = 0, 1$. We now assume that (2.1.5) is satisfied for a certain $k \geq 1$ and take $x_\lambda \in D_{k+1} \backslash D_k$. If we set $\mu = \lambda - 2^{-k-1}$, $\nu = \lambda + 2^{-k-1}$ then $x_\mu, x_\nu \in D_k$. Therefore, we deduce from (2.1.5)

$$\begin{aligned} &(\mu+\nu)u(x_1) + (2-\mu-\nu)u(x_0) - u(x_\mu) - u(x_\nu) \\ &\leq \ (\mu(1-\mu) + \nu(1-\nu))L\sum_{h=0}^{k-1} \tilde{\omega}\left(\frac{L}{2^h}\right), \end{aligned}$$

which implies, by the definition of μ, ν,

$$\begin{aligned} &2\lambda u(x_1) + 2(1-\lambda)u(x_0) - u(x_\mu) - u(x_\nu) \\ &\leq \ 2\left(\lambda(1-\lambda) - \frac{1}{2^{2k+2}}\right) L\sum_{h=0}^{k-1} \tilde{\omega}\left(\frac{L}{2^h}\right). \end{aligned}$$

On the other hand, we have by (2.1.4)

$$u(x_\mu) + u(x_\nu) - 2u(x_\lambda) \leq \frac{L}{2^{k+1}} \tilde{\omega}\left(\frac{L}{2^k}\right).$$

It follows that

$$\lambda u(x_1)+(1-\lambda)u(x_0)-u(x_\lambda) \leq \frac{L}{2^{k+2}} \tilde{\omega}\left(\frac{L}{2^k}\right)+\lambda(1-\lambda)L\sum_{h=0}^{k-1} \tilde{\omega}\left(\frac{L}{2^h}\right). \quad (2.1.6)$$

Our hypotheses on λ imply that

$$\lambda(1-\lambda) \geq \frac{1}{2^{k+1}}\left(1 - \frac{1}{2^{k+1}}\right) > \frac{1}{2^{k+2}}.$$

Thus, we conclude from (2.1.6)

$$\lambda u(x_1) + (1-\lambda)u(x_0) - u(x_\lambda) \leq \lambda(1-\lambda)L\sum_{h=0}^{k} \tilde{\omega}\left(\frac{L}{2^h}\right).$$

which is the induction step required to prove (2.1.5). Since the union of the sets D_j is dense in $[x_0, x_1]$ and u is supposed continuous, this proves that u is semiconcave with the desired modulus.

The assertion concerning the case of a linear $\tilde{\omega}$ is a rephrasing of the equivalence between (a) and (c) in Proposition 1.5.2. ■

We conclude this section analyzing under which conditions the composition of a semiconcave function with another function is also semiconcave.

Proposition 2.1.12 *Let $u : A \to \mathbb{R}$ be a locally semiconcave function, with A open.*

(i) *If $f : \mathbb{R} \to \mathbb{R}$ is increasing and locally semiconcave, then $f \circ u$ is locally semiconcave on A. If the moduli of semiconcavity of u and f are both linear, then so is the modulus of $f \circ u$.*

(ii) *If $\phi : V \to A$ is a C^1 function, with $V \subset \mathbb{R}^m$ open, then $u \circ \phi$ is locally semiconcave in V. If $D\phi$ is locally Lipschitz in V and the semiconcavity modulus of u is linear, then the modulus of $u \circ \phi$ is also linear.*

Proof — **(i)** Let S be any compact subset of A, and let us denote by ω_1, ω_2 the semiconcavity moduli and by L_1 and L_2 the Lipschitz constants of u in

S and of f in $u(S)$ respectively. We have, for any x, y such that $[x,y] \subset S$ and for any $\lambda \in [0,1]$,

$$\begin{aligned}
& f(u(\lambda x + (1-\lambda)y)) \\
\geq\ & f(\,\lambda u(x) + (1-\lambda)u(y) - \lambda(1-\lambda)|x-y|\omega_1(|x-y|)\,) \\
\geq\ & f(\lambda u(x) + (1-\lambda)u(y)) - L_2\lambda(1-\lambda)|x-y|\omega_1(|x-y|)
\end{aligned}$$

and therefore

$$\begin{aligned}
& \lambda f(u(x)) + (1-\lambda) f(u(y)) - f(u(\lambda x + (1-\lambda)y)) \\
\leq\ & \lambda f(u(x)) + (1-\lambda) f(u(y)) - f(\lambda u(x) + (1-\lambda)u(y)) \\
& + L_2\lambda(1-\lambda)|x-y|\omega_1(|x-y|) \\
\leq\ & \lambda(1-\lambda)|x-y|(L_1\omega_2(L_1|x-y|) + L_2\omega_1(|x-y|)).
\end{aligned}$$

Thus, $f \circ u$ is semiconcave with modulus $\omega(\rho) := L_1\omega_2(L_1\rho) + L_2\omega_1(\rho)$.
(ii) Given $x_0 \in V$, let S be a convex neighborhood of x_0 such that $\bar{S} \subset V$ and $\operatorname{co}\phi(\bar{S}) \subset A$. Let us denote by ω_1 the modulus of continuity of $D\phi$ in S and by ω_2 the semiconcavity modulus of u in $\operatorname{co}\phi(S)$. Moreover, let L_1 and L_2 be the Lipschitz constants of ϕ in S and of u in $\operatorname{co}\phi(S)$. We have, for any x, y such that $[x,y] \subset S$ and for any $\lambda \in [0,1]$,

$$\begin{aligned}
& u(\lambda\phi(x) + (1-\lambda)\phi(y)) - u(\phi(\lambda x + (1-\lambda)y)) \\
\leq\ & L_2|\lambda\phi(x) + (1-\lambda)\phi(y) - \phi(\lambda x + (1-\lambda)y))| \\
\leq\ & L_2\lambda(1-\lambda)|x-y|\omega_1(|x-y|)
\end{aligned}$$

and therefore

$$\begin{aligned}
& \lambda u(\phi(x)) + (1-\lambda)u(\phi(y)) - u(\phi(\lambda x + (1-\lambda)y)) \\
\leq\ & \lambda u(\phi(x)) + (1-\lambda)u(\phi(y)) - u(\lambda\phi(x) + (1-\lambda)\phi(y)) \\
& + u(\lambda\phi(x) + (1-\lambda)\phi(y)) - u(\phi(\lambda x + (1-\lambda)y)) \\
\leq\ & \lambda(1-\lambda)|x-y|(L_1\omega_2(L_1|x-y|) + L_2\omega_1(|x-y|)). \ \blacksquare
\end{aligned}$$

2.2 Examples

A first significant example of a semiconcave function is provided by the distance function from a given set.

Proposition 2.2.1 *Given a closed set $C \subset \mathbb{R}^n$, let d_C denote the distance function from C, defined as*

$$d_C(x) = \min_{y \in C} |y - x|.$$

Then the following properties hold.

(i) *d_C^2 is semiconcave in $\mathbb{R}^n$ with modulus $\omega(\rho) = \rho$.*

(ii) *d_C is locally semiconcave in $\mathbb{R}^n \setminus C$. More precisely, given a set S such that $\operatorname{dist}(S,C) > 0$, d_C is semiconcave in S with modulus $\omega(\rho) = (2\operatorname{dist}(S,C))^{-1}\rho$.*

(iii) *If C satisfies an interior sphere condition for some $r > 0$ (i.e., C is the union of closed spheres of radius r), then d_C is semiconcave in $\overline{\mathbb{R}^n \setminus C}$ with modulus $\omega(\rho) = (2r)^{-1}\rho$.*

Proof — **(i)** For any $x \in \mathbb{R}^n$ we have

$$d_C^2(x) - |x|^2 = \inf_{y \in C} |x-y|^2 - |x|^2 = \inf_{y \in C} |y|^2 - 2\langle x, y\rangle.$$

Since the infimum of linear functions is concave we deduce, by Proposition 1.5.2, that property (i) holds.
(ii) Let us first observe that, given $z, h \in \mathbb{R}^n$, $z \neq 0$, we have

$$\begin{aligned} &(|z+h| + |z-h|)^2 \\ \leq\ & 2(|z+h|^2 + |z-h|^2) = 4(|z|^2 + |h|^2) \leq \left(2|z| + \frac{|h|^2}{|z|}\right)^2. \end{aligned}$$

Thus

$$|z+h| + |z-h| - 2|z| \leq \frac{|h|^2}{|z|}. \tag{2.2.7}$$

Let S be a set with positive distance from C. For any x, h such that $[x-h, x+h] \subset S$, let $\bar{x} \in C$ be such that $d_C(x) = |x - \bar{x}|$. Then

$$\begin{aligned} & d_C(x+h) + d_C(x-h) - 2d_C(x) \\ \leq\ & |x+h-\bar{x}| + |x-h-\bar{x}| - 2|x-\bar{x}| \\ \leq\ & \frac{|h|^2}{|x-\bar{x}|}. \end{aligned}$$

Moreover $|x - \bar{x}| = d_C(x) \geq \operatorname{dist}(S,C)$. By Theorem 2.1.11 we conclude that d_C satisfies the desired property.
(iii) By hypothesis C is the union of closed spheres with radius r. Thus, the distance from C is the infimum of the distance from such spheres. By Proposition 2.1.5 it suffices to prove the result in the case $C = \overline{B}_r$. We have, for any $x \in \overline{\mathbb{R}^n \setminus C}$,

$$d_C(x) = |x| - r = d_{\{0\}}(x) - r.$$

Since the distance between $\{0\}$ and the complement of C is r, we obtain from part (ii) that d_C is semiconcave with the desired modulus. ∎

Unlike d_C^2, d_C is not semiconcave in the whole space $\mathbb{R}^n$; it is in general not even semiconcave in the closure of the complement of C, unless C satisfies an interior sphere condition. For instance, if $C = \{0\}$, then $d_C(x) = |x|$ is locally semiconcave in $\mathbb{R}^n \setminus \{0\}$, but not in $\mathbb{R}^n$ (see Example 2.1.4).

Another function which is not differentiable in general, but which is semiconcave, is the smallest eigenvalue of a symmetric matrix, or, in general, the sum of the smallest h eigenvalues.

Proposition 2.2.2 *Let $M(x)$ be a $k \times k$ symmetric matrix whose entries are functions in $C^1(A)$, with $A \subset \mathbb{R}^n$ open set. Then the sum of the first h eigenvalues of $M(x)$ (ordered in an increasing fashion) is a semiconcave function in A, for any $h = 1, 2, \ldots, k$.*

Proof — Let $\lambda_1(x) \leq \ldots \leq \lambda_k(x)$ denote the eigenvalues of $M(x)$. From linear algebra it is known that

$$\sum_{i=1}^h \lambda_i(x) = \min_{\xi_1,\ldots,\xi_h \in S^{k-1}} \{\sum_{i=1}^h \langle M(x)\xi_i, \xi_i\rangle \ : \ \langle \xi_i, \xi_j\rangle = \delta_{ij} \ \forall\, i, j\},$$

where S^{k-1} is the unit sphere in $\mathbb{R}^k$. Since, for any $\xi_1, \ldots, \xi_h \in S^{k-1}$, the function

$$x \to \sum_{i=1}^h \langle M(x)\xi_i, \xi_i\rangle$$

is of class C^1, our assertion follows from Propositions 2.1.5 and 2.1.2. ∎

We now give examples of partial differential equations whose solutions satisfy a semiconcavity estimate.

Proposition 2.2.3 *Let $u \in C^4([0,T] \times \mathbb{R}^n)$ be a solution of the equation*

$$\partial_t u(t,x) + H(\nabla u(t,x)) = a\Delta u(t,x) \tag{2.2.8}$$

where $a > 0$ and $H \in C^2(\mathbb{R}^n)$ is uniformly convex. Suppose that the first and second derivatives of u are bounded. Then, for any $t \in \,]0,T]$, the function $u(t,\cdot)$ is semiconcave in $\mathbb{R}^n$ with modulus $\omega(\rho) = \rho/2\lambda t$, where $\lambda > 0$ is a lower bound on the eigenvalues of $D^2 H$.

Proof — Let us set $w(t,x) = t\partial^2_{x_1 x_1} u(t,x)$. We have

$$\begin{aligned}\partial_t w &= \frac{w}{t} + t\partial^2_{x_1x_1}(a\Delta u - H) \\ &= \frac{w}{t} + a\Delta w - \langle DH, \nabla w\rangle - t\langle D^2 H(\nabla \partial_{x_1} u), \nabla \partial_{x_1} u\rangle.\end{aligned}$$

Now by assumption

$$t\langle D^2H(\nabla\partial_{x_1}u), \nabla\partial_{x_1}u\rangle \geq t\lambda|\nabla\partial_{x_1}u|^2 \geq \lambda w^2/t.$$

Therefore

$$\partial_t w \leq \frac{w}{t}(1-\lambda w) + a\Delta w - \langle DH, \nabla w\rangle.$$

Since $w(t,\cdot) \to 0$ uniformly as $t \to 0$, we deduce by the maximum principle that $\lambda w(t,x) \leq 1$ everywhere. This implies that $\partial^2_{x_1x_1}u(t,x) \leq (\lambda t)^{-1}$. The same reasoning can be applied to $\partial^2_{\nu\nu}u$ for any unit vector $\nu \in \mathbb{R}^n$. We conclude by Proposition 1.5.2. ■

Remark 2.2.4 The smoothness assumption on u in the previous proposition is not restrictive, by well known regularity properties of parabolic equations. The semiconcavity estimate obtained in the proposition is interesting because it depends only on λ and not, for instance, on the initial value of the solution. The fact that the diffusion coefficient a does also not appear in the estimate suggests that the semiconcavity of u is not a consequence of the parabolic character of the equation, and that it should hold also for the first order equation corresponding to the case $a = 0$. This agrees with the results of Section 1.5 (see in particular Corollary 1.5.4(i)) about the semiconcavity of the generalized solutions of (1.3.11) given by Hopf's formula (1.2.7). We recall that another second order equation for which a similar semiconcavity estimate holds is the equation for the pressure in the theory of porous media (see e.g. [13])

$$\partial_t v(t,x) = av(t,x)\Delta v(t,x) + |\nabla v(t,x)|^2.$$

2.3 Special properties of $SCL(A)$

In this section we focus our attention on the class SCL (A) of functions $u : A \to \mathbb{R}^n$ which are semiconcave with a linear modulus. We have seen in Proposition 1.5.2 that such functions can be regarded as C^2 perturbations of concave functions. This allows to extend immediately certain properties of concave functions, as the following ones.

Theorem 2.3.1 *Let $A \subset \mathbb{R}^n$ be an open set and let $u \in$ SCL (A). Then the following properties hold.*

(i) *(**Alexandroff's Theorem**) u is twice differentiable a.e.; that is, for a.e. every $x_0 \in A$, there exist a vector $p \in \mathbb{R}^n$ and a symmetric matrix B such that*

$$\lim_{x\to x_0} \frac{u(x) - u(x_0) - \langle p, x - x_0\rangle + \langle B(x-x_0), x - x_0\rangle}{|x-x_0|^2} = 0.$$

(ii) *The gradient of u, defined almost everywhere in A, belongs to the class* $\mathrm{BV}_{loc}(A, \mathbb{R}^n)$.

Proof — Properties (i) and (ii) hold for a convex function (see e.g. [43, Ch. 6.3]). Since u is the difference of a smooth function and a convex one, u also satisfies these properties. ■

The following result shows that semiconcave functions with linear modulus exhibit a behavior similar to C^2 functions near a minimum point.

Theorem 2.3.2 *Let $u : A \to \mathbb{R}$ be semiconcave with a linear modulus and let $x_0 \in A$ be a point of local minimum for u. Then there exist a sequence $\{x_h\} \subset A$ and an infinitesimal sequence $\{\varepsilon_h\} \subset \mathbb{R}_+$ such that u is twice differentiable in x_h and that*

$$\lim_{h\to\infty} x_h = x_0, \quad \lim_{h\to\infty} Du(x_h) = 0, \quad D^2u(x_h) \geq -\varepsilon_h I \quad \forall\, h.$$

The proof of this theorem is based on the following result.

Theorem 2.3.3 (Jensen's Lemma) *Let $u : A \to \mathbb{R}$ be semiconcave with a linear modulus and let $x_0 \in A$ be a strict local minimum point for u. Given $p \in \mathbb{R}^n$, let A_p denote the set of points where the function $x \to u(x) + \langle p, x\rangle$ has a local minimum. Then, for any $r > 0$ and $\delta > 0$, the set*

$$E = \{x \in B_r(x_0) \ : \ x \in A_p \textit{ for some } p \in B_\delta\}$$

has positive measure.

Proof — Let $\omega(\rho) = C\rho$ be the modulus of semiconcavity of u. Let us fix $r_0 \leq r$ such that $B(x_0, r_0) \subset A$ and that $\min_{\partial B(x_0,r_0)} u > u(x_0)$. Then we take $\delta_0 \leq \delta$ such that

$$r_0\delta_0 < \min_{\partial B(x_0,r_0)} u - u(x_0). \tag{2.3.9}$$

Let us first consider the case when u is of class C^∞. Given $p \in B_{\delta_0}$, we obtain from (2.3.9) that any point $x \in A_p \cap \overline{B(x_0, r_0)}$ belongs in fact to the open ball $B(x_0, r_0)$. Recalling also Proposition 1.5.2 we find, at such an x,

$$Du(x) = -p, \qquad 0 \leq D^2u(x) \leq 2CI.$$

Let us define

$$E_0 = \{x \in B_{r_0}(x_0) \ : \ x \in A_p \text{ for some } p \in B_{\delta_0}\}.$$

Then we obtain that Du is a surjective mapping from E_0 to B_{δ_0}, and in addition $|| \det D^2u|| \leq (2C)^n$ on E_0. We conclude

$$\text{meas}(B_{\delta_0}) = \text{meas}(Du(E_0)) \leq \int_{E_0} || \det D^2u(x)|| \, dx \leq \text{meas}(E_0)(2C)^n,$$

and so we obtain that E_0 has measure greater than some constant M_0 depending only on C and δ_0.

In the case where u is not smooth, we proceed by approximation. Using convolutions we obtain a family u_h of C^∞ functions, semiconcave with the same modulus of u, and converging uniformly to u. Then, for h large enough, the functions u_h also satisfy inequality (2.3.9). If we call E_h the set defined as E_0 with u replaced by u_h, we obtain that the measure of E_h is also greater than M_0. Next we observe that if some x belongs to E_h for infinitely many h, then it also belongs to E_0. This implies that E_0 has measure at least M_0. Since E contains E_0, this proves the result. ■

Proof of Theorem 2.3.2 — Let $B \subset\subset \Omega$ be a bounded open set containing x_0 such that $u(x) \leq u(x_0)$ for $x \in \bar{B}$, and let $w(x) = u(x) + |x - x_0|^4$. Then $w : B \to \mathbb{R}$ satisfies the assumptions of Jensen's Lemma; we deduce that there exist a sequence $\{p_h\} \subset \mathbb{R}^n$ converging to 0 and a sequence $\{x_h\} \subset B$ such that $w(x) + \langle p_h, x\rangle$ has a minimum at x_h. In view of Alexandroff's Theorem it is not restrictive to assume that w is twice differentiable at x_h. Then we have $Dw(x_h) = p_h$ and $D^2w(x_h) \geq 0$ for any h. Any limit point of $\{x_h\}$ is a minimum for w in $\bar{B}$; since x_0 is the only minimum point for w in $\bar{B}$, this implies that $x_h \to x_0$. Since we have $Du(x_h) = Dw(x_h) - 4|x_h - x_0|^2(x_h - x_0)$ and $D^2u(x_h) = D^2w(x_h) - 4|x_h - x_0|^2 I - 8(x_h - x_0) \otimes (x_h - x_0)$ the assertion follows with $\varepsilon_h = 12|x_h - x_0|^2$. ■

Theorem 2.3.2 plays an important role in the theory of viscosity solutions to second order degenerate elliptic equations (see [37]).

2.4 A differential Harnack inequality

Let us consider the viscous Hamilton–Jacobi equation

$$\partial_t u(t,x) + |\nabla u(t,x)|^2 = \Delta u(t,x), \qquad t \geq 0, x \in \mathbb{R}^n. \tag{2.4.10}$$

We have seen in Proposition 2.2.3 that the solutions to this equation are semiconcave. We now show how such a semiconcavity result is related to the classical Harnack inequality for the heat equation.

A remarkable feature of equation (2.4.10) is that it can be reduced to the heat equation by a change of unknown called *Cole-Hopf transformation*, or logarithmic transformation. In fact, if we set $w(t,x) = \exp(-u(t,x))$, a direct computation shows that u satisfies (2.4.10) if and only if $\partial_t w = \Delta w$. Let us investigate the properties of w which follow from the semiconcavity of u.

Proposition 2.4.1 *Let w be a positive solution of the heat equation in $[0,T] \times \mathbb{R}^n$ whose first and second derivatives are bounded. Then w satisfies*

$$\nabla^2 w + \frac{w}{2t} I - \frac{\nabla w \otimes \nabla w}{w} \geq 0; \tag{2.4.11}$$

$$\partial_t w + \frac{nw}{2t} - \frac{|\nabla w|^2}{w} \geq 0. \tag{2.4.12}$$

Here $\nabla^2 w$ denotes the hessian matrix of w with respect to the space variables; inequality (2.4.11) *means that the matrix in the left–hand side is positive semi–definite.*

Proof — It is not restrictive to assume that w is greater than some positive constant; if this is not the case, we can replace w by $w + \varepsilon$ and then let $\varepsilon \to 0^+$. Let us set $u(t,x) = -\ln(w(t,x))$. Then u is a solution of equation (2.4.10). In addition, u is bounded together with its first and second derivatives. Therefore, by Proposition 2.2.3, $u(t,\cdot)$ is semiconcave with modulus $\omega(\rho) = \rho/(4t)$. Using the equivalent formulations of Proposition 1.5.2, we can restate this property as

$$\nabla^2 u \leq \frac{1}{2t} I.$$

On the other hand, an easy computation shows that

$$\nabla^2 u = -\frac{\nabla^2 w}{w} + \frac{\nabla w \otimes \nabla w}{w^2}$$

and this proves (2.4.11). Taking the trace of the left–hand side of (2.4.11), we obtain

$$\Delta w + \frac{nw}{2t} - \frac{|\nabla w|^2}{w} \geq 0,$$

which implies (2.4.12), since w solves the heat equation. ∎

Inequality (2.4.12) is called a *differential Harnack estimate*. The connection with the classical Harnack inequality is explained by the following result.

Corollary 2.4.2 *Let w be as in the previous proposition. Then w satisfies*

$$w(t_2,y) \geq w(t_1,x)\left(\frac{t_1}{t_2}\right)^{n/2} \exp\left(-\frac{|y-x|^2}{4(t_2-t_1)}\right)$$

for all $x,y \in \mathbb{R}^n$, $t_2 > t_1 > 0$. Therefore, given any bounded set $\Omega \subset \mathbb{R}^n$ and $t_2 > t_1 > 0$, we have

$$\max_{\Omega} w(t_1,\cdot) \leq C \min_{\Omega} w(t_2,\cdot) \tag{2.4.13}$$

for some constant $C > 0$ depending on t_1, t_2 and the diameter of Ω but not on w.

Proof — Let us set

$$\gamma(s) = x + \frac{s-t_1}{t_2-t_1}(y-x), \qquad s \in [t_1,t_2].$$

Then we have, using (2.4.12),

$$\begin{aligned}
\frac{d}{ds}w(s,\gamma(s)) &= \partial_t w + \nabla w \cdot \dot\gamma \\
&\geq \partial_t w - \frac{|\nabla w|^2}{w} - \frac{w|\dot\gamma|^2}{4} \\
&\geq -\frac{nw}{2t} - \frac{w|\dot\gamma|^2}{4}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\ln\left(\frac{w(t_2,y)}{w(t_1,x)}\right) &= \int_{t_1}^{t_2} \frac{d}{ds}\ln w(s,\gamma(s))\,ds \\
&\geq \int_{t_1}^{t_2}\left(-\frac{n}{2t} - \frac{|y-x|^2}{4(t_2-t_1)^2}\right) ds \\
&= -\frac{n}{2}\ln\left(\frac{t_2}{t_1}\right) - \frac{|y-x|^2}{4(t_2-t_1)},
\end{aligned}$$

which proves our statement. ■

We recall that the Harnack estimate (2.4.13) holds in a much more general context than the one considered here; our aim was to point out the relationship with semiconcavity induced by the Hopf–Cole transformation.

2.5 A generalized semiconcavity estimate

In this section we compare the semiconcavity estimate with another one–sided estimate, a priori weaker, which was introduced in [28]. We prove here that the two estimates are in some sense equivalent, and this has applications for the study of certain Hamilton–Jacobi equations, as we will see in the following (see Theorem 5.3.7).

In the following we consider a function $u : A \to \mathbb{R}$, with $A \subset \mathbb{R}^n$ open. Given $x_0 \in A$, we set, for $0 < \delta < \mathrm{dist}(x_0, \partial A), x \in B_1$,

$$u_{x_0,\delta}(x) = \frac{u(x_0 + \delta x) - u(x_0)}{\delta}. \tag{2.5.14}$$

Definition 2.5.1 *Let $C \subset A$ be a compact set. We say that u satisfies a* generalized one–sided estimate *in C if there exist $K \geq 0$, $d \in]0, \mathrm{dist}(C, \partial A)[$ and a non-decreasing upper semicontinuous function $\widetilde{\omega} : [0,1] \to \mathbb{R}_+$, such that $\lim_{h\to 0} \widetilde{\omega}(h) = 0$ and*

$$\begin{aligned} &\lambda u_{x_0,\delta}(x) + (1-\lambda)u_{x_0,\delta}(y) - u_{x_0,\delta}(\lambda x + (1-\lambda)y) \\ &\quad \leq \ \lambda(1-\lambda)|x-y|\{K\delta + \widetilde{\omega}(|x-y|)\} \end{aligned} \tag{2.5.15}$$

for all $x_0 \in C$, $\delta \in]0, \delta_0[$, $x, y \in B_1$, $\lambda \in [0,1]$.

It is easily seen that, if u is semiconcave in A, then the above property is satisfied taking $\widetilde{\omega}$ equal to a modulus of semiconcavity of u in A and $K = 0$. Conversely, semiconcavity can be deduced from the one–sided estimate above, as the next result shows.

Theorem 2.5.2 *Let $u : A \to \mathbb{R}$, with A open and let C be a compact subset of A. If u satisfies a generalized one–sided estimate in C, then u is semiconcave in C.*

Proof — By hypothesis inequality (2.5.15) holds for some K, δ_0, $\widetilde{\omega}$. Let us take $x, y \in C$ such that $[x,y] \subset C$ and $\lambda \in [0,1]$. It is not restrictive to assume $|x-y| < \delta_0/2$. For any δ with $|x-y| < \delta < \delta_0$, we set

$$x_0 = \lambda x + (1-\lambda)y, \quad x' = \delta^{-1}(1-\lambda)(x-y), \quad y' = \delta^{-1}\lambda(y-x).$$

From (2.5.14) and (2.5.15) we obtain

$$\begin{aligned} &\lambda u(x) + (1-\lambda)u(y) - u(\lambda x + (1-\lambda)y) \\ &\quad = \ \delta\{\lambda u_{x_0,\delta}(x') + (1-\lambda)u_{x_0,\delta}(y') - u_{x_0,\delta}(\lambda x' + (1-\lambda)y')\} \\ &\quad \leq \ \delta\lambda(1-\lambda)|x'-y'|\{K\delta + \widetilde{\omega}(|x'-y'|)\} \\ &\quad = \ \lambda(1-\lambda)|x-y|\{K\delta + \widetilde{\omega}(\delta^{-1}|x-y|)\}. \end{aligned}$$

Therefore

$$\lambda u(x) + (1-\lambda)u(y) - u(\lambda x + (1-\lambda)y) \leq \lambda(1-\lambda)|x-y|\,\omega(|x-y|)$$

where $\omega(\rho) := \inf_{\delta \in]\rho,\delta_0[}\{K\delta + \widetilde{\omega}(\delta^{-1}\rho)\}$. The function ω is upper semicontinuous and non-decreasing. The conclusion will follow if we show that $\lim_{h\to 0}\omega(h) = 0$. Given $\varepsilon \in]0, 2K\delta_0[$, we choose $\eta \in]0,1[$ such that $\widetilde{\omega}(s) < \varepsilon/2$ for $0 < s < \eta$. For any $\rho < \epsilon\eta(2K)^{-1}$ we have $\varepsilon(2K)^{-1} \in]\rho, \delta_0[$; therefore

$$\omega(\rho) \leq \left\{K\frac{\varepsilon}{2K} + \widetilde{\omega}\left(\frac{2K}{\varepsilon}\rho\right)\right\} < \varepsilon.$$

This shows that $\lim_{\rho\to 0}\omega(\rho) = 0$ and concludes the proof. ■

Remark 2.5.3 From the above proof it is clear that the modulus of semiconcavity ω found for u is different from the function $\widetilde{\omega}$ that appears in the one–sided estimate (2.5.15). In some cases it is easy to compute explicitly ω. Let us suppose, for instance, that $\widetilde{\omega}(\rho) = C\rho^\alpha$, with $C, \alpha > 0$. Differentiating with respect to δ the quantity $K\delta + Ch^\alpha\delta^{-\alpha}$ we find that, for small ρ,

$$\omega(\rho) = C'\rho^{\frac{\alpha}{\alpha+1}}.$$

Thus, ω is also power–like, but with a lower exponent. Let us also observe that ω cannot be linear except in the case where $\widetilde{\omega} \equiv 0$.

Bibliographical notes.

The notion of semiconcavity has been used in several different contexts in the last decades. We have already mentioned in the previous chapter the applications to the uniqueness of solutions of Hamilton–Jacobi equations. In the framework of control theory, the role of semiconcavity for the characterization of the value function was underlined by Hrustalev [54].

Most authors have considered the notion of semiconcavity corresponding to a linear modulus in Definition 2.1.1. Moduli of a power form have been considered by Capuzzo Dolcetta and Ishii [29] and by Cannarsa and Soner [27]. The general definition used here appears in [22]. The Lipschitz continuity for a general modulus was proved in [6]. Theorem 2.1.11 generalizes a result in [27]. The counterexample of Proposition 2.1.7 is new.

The examples of Section 2.2 are well–known in the literature. More references on the distance function will be given in next chapter. For a detailed treatment of the vanishing viscosity method the reader can consult [60, 62, 67].

Theorems 2.3.2 and 2.3.3 are basically due to R. Jensen [58]. Our exposition here follows [37].

The Cole–Hopf tranformation was introduced in [52] and [33]. The idea of stating the Harnack estimate in the form of a differential inequality of the type (2.4.12) is due to Li and Yau [66] and was mainly motivated by the study of parabolic equations on manifolds. The matrix inequality (2.4.11) is due to R. Hamilton [49].

The generalized semiconcavity estimate of Section 2.5 was introduced in [28]. Theorem 2.5.2 is taken from [75].

Chapter 3

Generalized gradients and semiconcavity

Generalized differentials are the central object of the so-called nonsmooth analysis: they are generalizations of the usual notion of gradient, which allow to extend some tools of calculus to functions which are not differentiable. The best known examples are the semi-differential of convex analysis. More recently, different kinds of generalized gradients have been introduced, which allow to consider many classes of nondifferentiable functions.

For semiconcave functions, the analysis of the generalized gradients is important in view of the applications to control theory. As we have already seen in a special case (Corollary 1.4.7), if the value function of a control problem is smooth, one can find the optimal trajectories knowing the differential of the value function. In the general case, where the value function is not smooth but only semiconcave, one can try to follow a similar procedure starting from the generalized gradient.

In Section 3.1 we give the definition of the generalized differentials which are relevant for our purposes and we recall some basic properties and equivalent characterizations. Then we restrict ourselves to semiconcave functions. In Section 3.2 we show that semiconcave functions possess directional derivatives everywhere. Then, in Section 3.3, we describe the special properties of the superdifferential of a semiconcave function; in particular, we show that it is non–empty and that it is a singleton exactly at those points where the function is differentiable. These properties are well-known in the case of concave functions, but are proved here for semiconcave functions with arbitrary modulus.

Section 3.4 is devoted to the so-called marginal functions, which are obtained as the infimum of smooth functions. We show that semiconcave functions are strictly related to marginal functions. In addition, one can describe the semi-differentials of a marginal functions using the general results about semiconcave functions of the previous sections.

In Section 3.5 we study in detail the so–called inf–convolutions. These are marginal functions defined by a process which is a generalization of Hopf's formula, and provide approximations to a given function which enjoy useful properties.

3.1 Generalized differentials

We recall first some definitions and basic results from nonsmooth analysis concerning the generalized differentials. In this section u is a real–valued function defined on an open set $A \subset \mathbb{R}^n$.

Definition 3.1.1 *For any $x \in A$, the sets*

$$D^-u(x) = \left\{ p \in \mathbb{R}^n \,:\, \liminf_{y \in A, y \to x} \frac{u(y) - u(x) - \langle p, y - x \rangle}{|y - x|} \geq 0 \right\}, \tag{3.1.1}$$

$$D^+u(x) = \left\{ p \in \mathbb{R}^n \,:\, \limsup_{y \in A, y \to x} \frac{u(y) - u(x) - \langle p, y - x \rangle}{|y - x|} \leq 0 \right\} \tag{3.1.2}$$

are called, respectively, the (Fréchet) superdifferential *and* subdifferential *of u at x.*

It is easy to check that, for any $x \in A$,

$$D^-(-u)(x) = -D^+u(x)\,. \tag{3.1.3}$$

Example 3.1.2 Let $n = 1$ and $\Omega =]-1, 1[$

1. Let $u(x) = |x|$. Then it is easily seen that $D^+u(0) = \emptyset$ whereas $D^-u(0) = [-1, 1]$.
2. Let $u(x) = \sqrt{|x|}$. Then, $D^+u(0) = \emptyset$ whereas $D^-u(0) = \mathbb{R}$.
3. Let $u(x, y) = |x| - |y|$. Then, $D^+u(0, 0) = D^-u(0, 0) = \emptyset$.

Definition 3.1.3 *Let $x \in A$ and $\theta \in \mathbb{R}^n \setminus \{0\}$. The* upper and lower Dini derivatives *of u at x in the direction θ are defined as*

$$\partial^+u(x, \theta) = \limsup_{h \to 0^+, \theta' \to \theta} \frac{u(x + h\theta') - u(x)}{h}.$$

and

$$\partial^-u(x, \theta) = \liminf_{h \to 0^+, \theta' \to \theta} \frac{u(x + h\theta') - u(x)}{h},$$

respectively.

It is easy readily seen that, for any $x \in A$ and $\theta \in \mathbb{R}^n \setminus \{0\}$

$$\partial^-(-u)(x,\theta) = -\partial^+ u(x,\theta)\,. \tag{3.1.4}$$

Remark 3.1.4 Whenever u is Lipschitz continuous in a neighborhood of x, the lower Dini derivative reduces to

$$\partial^- u(x,\theta) = \liminf_{h\to 0^+} \frac{u(x+h\theta)-u(x)}{h} \tag{3.1.5}$$

for any $\theta \in \mathbb{R}^n \setminus \{0\}$. Indeed, in this case there exists $L>0$ such that, for h, θ' close to $0, \theta$,

$$\left| \frac{u(x+h\theta')-u(x)}{h} - \frac{u(x+h\theta)-u(x)}{h} \right| \le L|\theta'-\theta|\,,$$

and (3.1.5) easily follows. A similar property holds for the upper Dini derivative.

Proposition 3.1.5 *Let $u : A \to \mathbb{R}$ and $x \in A$. Then the following properties hold true.*

(a) *For any $\theta \in \mathbb{R}^n \setminus \{0\}$*

$$D^+u(x) = \{p \in \mathbb{R}^n \;:\; \partial^+ u(x,\theta) \le \langle p,\theta\rangle \quad \forall\,\theta \in \mathbb{R}^n \setminus \{0\}\ \},$$

$$D^-u(x) = \{p \in \mathbb{R}^n \;:\; \partial^- u(x,\theta) \ge \langle p,\theta\rangle \quad \forall\,\theta \in \mathbb{R}^n \setminus \{0\}\ \}.$$

(b) *$D^+u(x)$ and $D^-u(x)$ are closed convex sets (possibly empty).*

(c) *$D^+u(x)$ and $D^-u(x)$ are both non–empty if and only if u is differentiable at x; in this case we have that*

$$D^+u(x) = D^-u(x) = \{Du(x)\}.$$

Proof — **(a)** The fact that, for every vector $p \in D^+u(x)$ and any $\theta \in \mathbb{R}^n \setminus \{0\}$,

$$\partial^+ u(x,\theta) \le \langle p,\theta\rangle \tag{3.1.6}$$

is a direct consequence of (3.1.1).
The converse can be proved by contradiction. Indeed, let $p \in \mathbb{R}^n$ be a vector satisfying (3.1.6) for any $\theta \ne 0$, but suppose that $p \notin D^+u(x)$. Then, a number $\varepsilon > 0$ and a sequence $\{x_k\} \subset A$ exist such that $x_k \to x$ as $k \to \infty$, and

$$u(x_k) - u(x) \ge \langle p, x_k - x\rangle + \epsilon|x_k - x| \quad (k \in \mathbb{N}). \tag{3.1.7}$$

Moreover, possibly taking a subsequence, we may assume that the sequence of unit vectors

$$\theta_k := \frac{x_k - x}{|x_k - x|}$$

converges to some unit vector θ as $k \to \infty$. Then, (3.1.7) yields

$$\begin{aligned} \varepsilon + \langle p, \theta \rangle &\leq \limsup_{k\to\infty} \frac{u(x_k) - u(x)}{|x_k - x|} \\ &= \limsup_{k\to\infty} \frac{u(x + |x_k - x|\theta_k) - u(x)}{|x_k - x|} \leq \partial^+ u(x, \theta)\,, \end{aligned}$$

which contradicts (3.1.6). The analogous property for $D^- u(x)$ immediately follows recalling (3.1.3) and (3.1.4).
(b) This property is a consequence of (a).
(c) If u is differentiable at x, then it is easily seen that

$$Du(x) \in D^+ u(x) \cap D^- u(x).$$

Conversely, let $p_1 \in D^+ u(x)$ and $p_2 \in D^- u(x)$. Then, property (i) yields

$$\langle p_2, \theta \rangle \leq \partial^- u(x, \theta) \leq \partial^+ u(x, \theta) \leq \langle p_1, \theta \rangle\,,$$

and so $\langle p_1 - p_2, \theta \rangle \geq 0$ for any $\theta \in \mathbb{R}^n \setminus \{0\}$. This implies that $p_1 = p_2$. So, $D^+ u(x)$ and $D^- u(x)$ coincide and reduce to a singleton. Using the definition of $D^\pm$ we obtain that u is differentiable at x with gradient $Du(x) = p_1$. Hence the conclusion follows. ∎

If u is continuous, then it turns out that the super– and subdifferential can be described in terms of "test" functions. For this purpose, it is useful to introduce the following notion of contact between (the graphs of) two real–valued functions, u and ϕ, defined on A.

Definition 3.1.6 *We say that ϕ* touches u from above *at a point $x_0 \in A$ if*

$$u(x_0) = \phi(x_0) \tag{3.1.8}$$

and, for some open ball $B_r(x_0) \subset A$,

$$u(x) \leq \phi(x)\,, \qquad \forall x \in B_r(x_0)\,. \tag{3.1.9}$$

Similarly, ϕ touches u from below *at $x_0 \in A$ if the above properties hold with $\leq$ replaced by $\geq$.*

Clearly, if (3.1.8) holds, then (3.1.9) is equivalent to requiring that $u - \phi$ has a *local maximum* at x.

Proposition 3.1.7 *Let $u \in C(A), p \in \mathbb{R}^n$, and $x \in A$. Then the following properties are equivalent:*

(a) $p \in D^+u(x)$ *(resp.* $p \in D^-u(x)$ *);*

(b) $p = D\phi(x)$ *for some function* $\phi \in C^1(A)$ *touching* u *from above (resp. below) at* x*;*

(c) $p = D\phi(x)$ *for some function* $\phi \in C^1(A)$ *such that* $u - \phi$ *attains a local maximum (resp. minimum) at* x*.*

Before giving the proof, let us prove a technical lemma.

Lemma 3.1.8 *Let* $\omega :]0, +\infty[\to [0, +\infty[$ *be an upper semicontinuous function such that* $\lim_{r\to 0} \omega(r) = 0$*. Then there exists* $\omega_1 : [0, +\infty[\to [0, +\infty[$ *such that*

(i) $\omega_1(r) \to 0$ *as* $r \to 0$*,*

(ii) $\omega(r) \le \omega_1(r)$ *for any* $r \ge 0$*,*

(iii) *the function* $\gamma(r) := r\omega_1(r)$ *is in* $C^1([0, +\infty[)$ *and satisfies* $\gamma'(0) = 0$*.*

Proof — Let us first set

$$\bar{\omega}(r) = \max_{\rho \in]0,r]} \omega(\rho).$$

Then $\bar{\omega}$ is nondecreasing, not smaller than ω and tends to 0 as $r \to 0$. Next we define for $r > 0$

$$\omega_0(r) = \frac{1}{r}\int_r^{2r} \bar{\omega}(\rho)d\rho, \qquad \omega_1(r) = \frac{1}{r}\int_r^{2r} \omega_0(\rho)d\rho,$$

and we set $\omega_1(0) = 0$. We first observe that, since $\bar{\omega}$ is non decreasing, the same holds for ω_0 and ω_1. Then we obtain that $\omega(r) \le \omega_0(r) \le \bar{\omega}(2r)$, and so $\omega_0(r) \to 0$ as $r \to 0$. Arguing in the same way on ω_1 we deduce that properties (i) and (ii) hold. To prove (iii), let us set $\gamma(r) = r\omega_1(r)$. Then $\gamma \in C^1(]0, +\infty[)$ with derivative $\gamma'(r) = 2\omega_0(2r) - \omega_0(r)$. Thus $\gamma'(r) \to 0$ as $r \to 0$ and so γ is C^1 in the closed half–line $[0, +\infty[$. ■

Proof of Proposition 3.1.7 — The implications (b) ⇒ (c) and (c) ⇒ (a) are obvious; so, it is enough to prove that (a) implies (b). Given $p \in D^+u(x)$, let us define, for $r > 0$,

$$\omega(r) = \max_{y:|y-x|=r} \frac{u(y) - u(x) - \langle p, y - x\rangle}{r}.$$

The function ω is continuous and tends to 0 as $r \to 0$, by definition of D^+u. Let ω_1 be the function associated to ω given by the previous lemma. Then, setting

$$\phi(y) = u(x) + \langle p, y - x\rangle + |y - x|\omega_1(|y - x|),$$

we have that ϕ touches u from above at x. ∎

The above characterization can be used to study the super– and subdifferentiability sets of a non–smooth function u. A first result in this direction is the following.

Proposition 3.1.9 *Let $u \in C(A)$. Then the sets*

$$A^+(u) = \{x \in A : D^+u(x) \neq \emptyset\}$$

and

$$A^-(u) = \{x \in A : D^-u(x) \neq \emptyset\}$$

are both dense in A.

Proof — Given $x_0 \in A$, let us fix $R > 0$ such that $\overline{B}_R(x_0) \subset A$. Given $\varepsilon > 0$, let x_ε be a point of $\overline{B}_R(x_0)$ where the function $u(x) - |x - x_0|^2/\varepsilon$ attains its maximum. Then we have

$$u(x_\varepsilon) - \frac{|x_\varepsilon - x_0|^2}{\varepsilon} \geq u(x_0)$$

which implies that

$$|x_\varepsilon - x_0|^2 \leq \varepsilon \left[\max_{\overline{B}_R(x_0)} u - u(x_0)\right].$$

Thus, $x_\varepsilon \to x_0$ as $\varepsilon \to 0$. It also follows that x_ε lies in the interior of $\overline{B}_R(x_0)$ for ε small enough. For such an ε let us set

$$\phi_\varepsilon(x) = u(x_\varepsilon) - \frac{|x_\varepsilon - x_0|^2}{\varepsilon} + \frac{|x - x_0|^2}{\varepsilon}.$$

Then $\phi_\varepsilon \in C^1(\overline{B}_R(x_0))$ and satisfies $\phi_\varepsilon(x_\varepsilon) = u(x_\varepsilon)$, $\phi_\varepsilon(x) \geq u(x)$ for all $x \in \overline{B}_R(x_0)$. We conclude that ϕ_ε touches u from above at x_ε and so, by Proposition 3.1.7(c), $D\phi(x_\varepsilon) \in D^+u(x_\varepsilon)$. This shows that $A^+(u)$ is dense in A. ∎

Let us introduce another kind of generalized gradient which will play an important role in our future analysis.

Definition 3.1.10 *Let $u : A \to \mathbb{R}$ be locally Lipschitz. A vector $p \in \mathbb{R}^n$ is called a* reachable gradient *of u at $x \in A$, if a sequence $\{x_k\} \subset A \setminus \{x\}$ exists such that u is differentiable at x_k for each $k \in \mathbb{N}$, and*

$$\lim_{k\to\infty} x_k = x\,, \qquad \lim_{k\to\infty} Du(x_k) = p\,. \tag{3.1.10}$$

The set of all reachable gradients of u at x is denoted by $D^*u(x)$.

It is easily seen that $D^*u(x)$ is a compact set. From Rademacher's Theorem it follows that $D^*u(x) \neq \emptyset$ for every $x \in A$. Let us now recall some more definitions from nonsmooth analysis.

Definition 3.1.11 *For any fixed $x \in A$ and $\theta \in \mathbb{R}^n \setminus \{0\}$, the* generalized lower derivative *is defined as*

$$u^0_-(x,\theta) = \liminf_{h\to 0^+,\, y\to x} \frac{u(y+h\theta)-u(y)}{h}\,. \tag{3.1.11}$$

Moreover the set

$$\partial_C u(x) = \{p \in \mathbb{R}^n \ :\ u^0_-(x,\theta) \leq \langle p,\theta\rangle \quad \forall\, \theta \in \mathbb{R}^n\} \tag{3.1.12}$$

is called Clarke's gradient *of u at x.*

One can also define the *generalized upper derivative* as

$$u^0_+(x,\theta) = \limsup_{h\to 0^+,\, y\to x} \frac{u(y+h\theta)-u(y)}{h}\,.$$

Since $u^0_-(x,\theta) = -u^0_+(x,-\theta)$, an equivalent definition of Clarke's gradient is therefore

$$\partial_C u(x) = \{p \in \mathbb{R}^n \ :\ u^0_+(x,\theta) \geq \langle p,\theta\rangle \quad \forall\, \theta \in \mathbb{R}^n\}.$$

It is proved in [31] that the reachable gradients of u at x generate Clarke's gradient in the sense of convex analysis, i.e.

$$\partial_C u(x) = \mathrm{co}\ D^*u(x)\,. \tag{3.1.13}$$

Consequently, $\partial_C u(x)$ is non–empty, convex and compact. Moreover, by Remark 3.1.4,

$$u^0_-(x,\theta) \leq \partial^- u(x,\theta) \leq \partial^+ u(x,\theta) \leq u^0_+(x,\theta)$$

for any $x \in A$ and $\theta \in \mathbb{R}^n \setminus \{0\}$. Then, recalling point (a) of Proposition 3.1.5, we conclude that

$$D^+u(x) \subset \partial_C u(x) \quad \text{and} \quad D^-u(x) \subset \partial_C u(x)$$

for any $x \in A$. Let us conclude by recalling the following extension of the classical mean value theorem (see [31, Th. 2.3.7]).

Theorem 3.1.12 *Let $u : A \to \mathbb{R}$ be locally Lipschitz. Given $x, y \in A$ such that $[x, y] \subset A$, there exists $\xi \in]x, y[$ and $p \in \partial_C u(\xi)$ such that $u(y) - u(x) = p \cdot (y - x)$.*

3.2 Directional derivatives and semiconcavity

We begin our exposition of the differential properties of semiconcave functions showing that they possess (one–sided) directional derivatives

$$\partial u(x, \theta) := \lim_{h \to 0^+} \frac{u(x + h\theta) - u(x)}{h}$$

at any point x and in any direction θ.

Theorem 3.2.1 *Let $u : A \to \mathbb{R}$ be semiconcave. Then, for any $x \in A$ and $\theta \in \mathbb{R}^n$,*

$$\partial u(x, \theta) = \partial^- u(x, \theta) = \partial^+ u(x, \theta) = u^0_-(x, \theta). \tag{3.2.14}$$

Proof — Let $\delta > 0$ be fixed so that $B_{\delta|\theta|}(x) \subset A$. Then, for any pair of numbers h_1, h_2 satisfying $0 < h_1 \leq h_2 < \delta$, estimate (2.1.1) yields

$$\left(1 - \frac{h_1}{h_2}\right) u(x) + \frac{h_1}{h_2} u(x + h_2\theta) - u(x + h_1\theta) \leq h_1 \left(1 - \frac{h_1}{h_2}\right) |\theta| \omega(h_2|\theta|).$$

Hence,

$$\begin{aligned} &\frac{u(x + h_1\theta) - u(x)}{h_1} \\ \geq\quad &\frac{u(x + h_2\theta) - u(x)}{h_2} - \left(1 - \frac{h_1}{h_2}\right) |\theta| \omega(h_2|\theta|)\,. \end{aligned} \tag{3.2.15}$$

Taking the lim inf as $h_1 \to 0^+$ in both sides of the above inequality, we obtain

$$\partial^- u(x, \theta) \geq \frac{u(x + h_2\theta) - u(x)}{h_2} - |\theta| \omega(h_2|\theta|).$$

Now, taking the lim sup as $h_2 \to 0^+$, we conclude that

$$\partial^- u(x, \theta) \geq \partial^+ u(x, \theta)\,.$$

So, $\partial u(x, \theta)$ exists and coincides with the lower and upper Dini derivatives.

To complete the proof of (3.2.14) it suffices to show that

$$\partial^+ u(x, \theta) \leq u^0_-(x, \theta). \tag{3.2.16}$$

For this purpose, let $\varepsilon > 0, \lambda \in]0, \delta[$ be fixed. Since u is continuous, for some $\alpha > 0$ we have that $B_{\lambda|\theta|}(y) \subset A$ for every $y \in B_\alpha(x)$, and

$$\frac{u(x+\lambda\theta)-u(x)}{\lambda} \leq \frac{u(y+\lambda\theta)-u(y)}{\lambda} + \varepsilon\,.$$

Using inequality (3.2.15) with x replaced by y, we conclude that

$$\frac{u(x+\lambda\theta)-u(x)}{\lambda} \leq \inf_{y\in B_\alpha(x), h\in]0,\lambda[} \frac{u(y+h\theta)-u(y)}{h} + |\theta|\omega(\lambda|\theta|) + \varepsilon.$$

Therefore,

$$\frac{u(x+\lambda\theta)-u(x)}{\lambda} \leq u^0_-(x,\theta) + |\theta|\omega(\lambda|\theta|) + \varepsilon.$$

Hence, taking the limit as $\varepsilon, \lambda \to 0$, we obtain inequality (3.2.16). ∎

3.3 The superdifferential of a semiconcave function

The superdifferential of a semiconcave function enjoys many properties which are not valid for a general Lipschitz functions, and which can be regarded as extension of analogous properties of concave functions. We start with the following basic estimate.

Proposition 3.3.1 *Let $u : A \to \mathbb{R}$ be a semiconcave function with modulus ω, and let $x \in A$. Then, a vector $p \in \mathbb{R}^n$ belongs to $D^+u(x)$ if and only if*

$$u(y) - u(x) - \langle p, y-x\rangle \leq |y-x|\omega(|y-x|) \qquad (3.3.17)$$

for any point $y \in A$ such that $[y,x] \subset A$.

Proof — If $p \in \mathbb{R}^n$ satisfies (3.3.17), then, by the very definition of superdifferential, $p \in D^+u(x)$. In order to prove the converse, let $p \in D^+u(x)$. Then, dividing the semiconcavity inequality (2.1.1) (after exchanging x with y) by $\lambda|x-y|$, we have

$$\frac{u(y)-u(x)}{|y-x|} \leq \frac{u(x+\lambda(y-x))-u(x)}{\lambda|y-x|} + (1-\lambda)\omega(|x-y|), \quad \forall \lambda \in]0,1].$$

Hence, taking the limit as $\lambda \to 0^+$, we obtain

$$\frac{u(y)-u(x)}{|y-x|} \leq \frac{\langle p, y-x\rangle}{|y-x|} + \omega(|x-y|)$$

since $p \in D^+u(x)$. Estimate (3.3.17) follows. ∎

Remark 3.3.2 In particular, if u is concave, we find that $p \in D^+u(x)$ if and only if

$$u(y) \geq u(x) + \langle p, y - x \rangle, \qquad \forall\, y \in \mathbb{R}^n.$$

In convex analysis (see e.g. [72]) this property is usually taken as the definition of the superdifferential. Therefore, the Fréchet super- and subdifferential coincide with the classical semi-differentials of convex analysis in the case of a concave (resp. convex) function.

Before investigating further properties of the superdifferential, let us show how Proposition 3.3.1 easily yields a compactness property for semiconcave functions.

Theorem 3.3.3 *Let $u_n : A \to \mathbb{R}$ be a family of semiconcave functions with the same modulus ω. Given an open set $B \subset\subset A$, suppose that the u_n's are uniformly bounded in B. Then there exists a subsequence u_{n_k} converging uniformly to a function $u : B \to \mathbb{R}$ semiconcave with modulus ω. In addition, $Du_{n_k} \to Du$ a.e. in B.*

Proof — From Remark 2.1.9 we deduce that the functions u_n are uniformly Lipschitz in B; thus, by the Ascoli–Arzelà theorem, there exists a subsequence, which we for simplicity denote again by u_n, converging uniformly to some $u : B \to \mathbb{R}$. Since the semiconcavity inequality is preserved by pointwise convergence, u is semiconcave with the same modulus of the u_n's. Let us consider a point $x_0 \in B$ such that u and all u_n are differentiable at x_0, for all n; we observe that all points of B have this property except for a set of measure zero. We claim that $Du_n(x_0) \to Du(x_0)$. To prove this, we argue by contradiction. The sequence $\{Du_n(x_0)\}$ is bounded; if it does not converge to $Du(x_0)$ there exists a subsequence converging to some $p_0 \neq Du(x_0)$. But then, passing to the limit in (3.3.17), we obtain that $p_0 \in D^+u(x_0)$ and we find a contradiction. This proves that $Du_n \to Du$ a.e. in B. ∎

Let us now derive other important properties of the superdifferential of a semiconcave function.

Proposition 3.3.4 *Let $u : A \to \mathbb{R}$ be a semiconcave function with modulus ω, and let $x \in A$. Then the following properties hold true.*

(a) *If $\{x_k\}$ is a sequence in A converging to x, and if $p_k \in D^+u(x_k)$ converges to a vector $p \in \mathbb{R}^n$, then $p \in D^+u(x)$.*

(b) *$D^*u(x) \subset \partial D^+u(x)$.*

(c) $D^+u(x) \neq \emptyset$.

(d) *If $D^+u(x)$ is a singleton, then u is differentiable at x.*

(e) *If $D^+u(y)$ is a singleton for any $y \in A$, then $u \in C^1(A)$.*

Remark 3.3.5 By $\partial D^+u(x)$ in part (b) above we mean the usual notion of boundary of $D^+u(x)$ taken with respect to the topology of $\mathbb{R}^n$. We can consider another notion of boundary of $D^+u(x)$: we take the minimal affine subspace Π containing $D^+u(x)$ and consider the boundary of $D^+u(x)$ with respect to the topology of Π. We will call such a set the *relative boundary* of $D^+u(x)$. We observe that the two notions coincide only if the dimension of $D^+u(x)$ is 0 or n; in the other cases we have $D^+u(x) = \partial D^+u(x)$, whereas the relative boundary is a proper subset of $D^+u(x)$.

Proof — **(a)** This property follows from Proposition 3.3.1 since ω is upper semicontinuous.
(b) First of all we note that, in light of (a), $D^*u(x) \subset D^+u(x)$. Therefore, the nontrivial part of (b) is that all achievable gradients are boundary points of $D^+u(x)$. Indeed, let $p \in D^*u(x)$ and let $x_k \in A$ be as in (3.1.10). Without loss of generality we may assume that

$$\lim_{k\to\infty} \frac{x_k - x}{|x_k - x|} = \theta$$

for some unit vector $\theta \in \mathbb{R}^n$. We claim that

$$p - t\theta \notin D^+u(x)\,, \quad \forall\, t > 0\,, \tag{3.3.18}$$

which in turn implies that p is in the boundary of $D^+u(x)$. Indeed, by (3.3.17),

$$\begin{aligned} &u(x_k) - u(x) - \langle p - t\theta, x_k - x\rangle \\ &\quad= u(x_k) - u(x) - \langle Du(x_k), x_k - x\rangle \\ &\qquad + \langle Du(x_k) - p, x - x_k\rangle + t\langle \theta, x - x_k\rangle \\ &\quad\geq -|x_k - x|\omega(|x_k - x|) - |Du(x_k) - p||x_k - x| + t\langle \theta, x - x_k\rangle. \end{aligned}$$

Therefore,

$$\liminf_{k\to\infty} \frac{u(x_k) - u(x) - \langle p - t\theta, x_k - x\rangle}{|x_k - x|} \geq t\,,$$

and our claim (3.3.18) follows.
(c) This property is an immediate consequence of (b). Indeed, u is locally Lipschitz by Theorem 2.1.1, and so $D^*u(x) \neq \emptyset$.

(d) Suppose that $D^+u(x) = \{p\}$ for some $p \in \mathbb{R}^n$, and let $\{x_k\} \subset A$ be any sequence such that $x_k \to x$. By (c), we can pick a point $p_k \in D^+u(x_k)$ for any k. On the other hand, by (a), sequence $\{p_k\}$ can only admit p as a cluster point. Therefore, $p_k \to p$. Moreover, by Proposition 3.3.1,

$$\begin{aligned} & u(x_k) - u(x) - \langle p, x_k - x\rangle \\ &= u(x_k) - u(x) + \langle p_k, x - x_k\rangle + \langle p_k - p, x_k - x\rangle \\ &\geq -|x_k - x|\omega(|x_k - x|) - |p_k - p||x_k - x|. \end{aligned}$$

Hence,

$$\liminf_{k\to\infty} \frac{u(x_k) - u(x) - \langle p, x_k - x\rangle}{|x_k - x|} \geq 0\,.$$

Since $\{x_k\}$ is arbitrary, we conclude that $p \in D^-u(x)$. Then, Proposition 3.1.5(c) implies that u is differentiable at x.
(e) This assertion is an easy consequence of (a) and (d). ∎

We recall that property (a) above can be restated saying that D^+u is *upper semicontinuous* as a set–valued map.

Theorem 3.3.6 *Let $u : A \to \mathbb{R}$ be a semiconcave function and let $x \in A$. Then*

$$D^+u(x) = \mathrm{co}D^*u(x)\,. \tag{3.3.19}$$

Moreover,

$$\partial u(x,\theta) = \min_{p\in D^+u(x)} \langle p, \theta\rangle = \min_{p\in D^*u(x)} \langle p, \theta\rangle \tag{3.3.20}$$

for any $\theta \in \mathbb{R}^n$.

Proof — We prove (3.3.20) first. From Proposition 3.1.5(a) and Proposition 3.3.4(b) we deduce that

$$\partial u(x,\theta) \leq \min_{p\in D^+u(x)} \langle p, \theta\rangle \leq \min_{p\in D^*u(x)} \langle p, \theta\rangle\,, \quad \forall\, \theta \in \mathbb{R}^n\,.$$

Therefore, to obtain (3.3.20) it suffices to show that

$$\min_{p\in D^*u(x)} \langle p, \theta\rangle \leq \partial u(x,\theta)\,, \quad \forall\, \theta \in \mathbb{R}^n\,. \tag{3.3.21}$$

Indeed, let $\theta \in \mathbb{R}^n$ be a fixed unit vector. Since u is a.e. differentiable, one can find a sequence $\{x_k\}$ of differentiability points of u, such that

$$\theta_k := \frac{x_k - x}{|x_k - x|} \to \theta, \quad \text{as } k \to \infty\,,$$

and $Du(x_k)$ converges to some vector $p_0 \in D^*u(x)$. Let ω be a modulus of semiconcavity for u. Then, by (3.3.17),

$$\langle Du(x_k), \theta_k \rangle \leq \frac{u(x + |x_k - x|\theta_k) - u(x)}{|x_k - x|} + \omega(|x_k - x|)\,.$$

Taking the limit as $k \to \infty$, we obtain

$$\langle p_0, \theta \rangle \leq \partial u(x, \theta)\,,$$

whence (3.3.21) follows. Now, since

$$\min_{p \in D^*u(x)} \langle p, \theta \rangle = \min_{p \in coD^*u(x)} \langle p, \theta \rangle\,,$$

equality (3.3.20) implies that the two convex sets $D^+u(x)$ and $\mathrm{co}D^*u(x)$ have the same support function. So, they must be the same set and (3.3.19) is proved. ■

Corollary 3.3.7 *If $u : A \to \mathbb{R}$ is both semiconvex and semiconcave in A, then $u \in C^1(A)$.*

Proof — At any point of A both u and $-u$ have nonempty superdifferential. Since $D^+(-u) = -D^-(u)$, this implies that the subdifferential of u is also nonempty everywhere. Therefore, by Propositions 3.1.5(c) and 3.3.4(e), u belongs to $C^1(A)$. ■

Let us give an equivalent characterization of the set D^*u for a semiconcave function.

Proposition 3.3.8 *Let $u \in \mathrm{SC}\,(A)$. Then, for any $x \in A$,*

$$D^*u(x) = \Big\{ \lim_{i \to \infty} p_i \;:\; p_i \in D^+u(x_i)\,,\; x_i \to x\,,\; \mathrm{diam}(D^+u(x_i)) \to 0 \Big\}\,. \tag{3.3.22}$$

Proof — Let us denote by $D^\sharp u(x)$ the right hand side of (3.3.22). Since D^+u reduces to the gradient at any differentiability point of u, we have that $D^*u(x) \subset D^\sharp u(x)$. Now, to show the reverse inclusion, let us fix a point $p = \lim_{i\to\infty} p_i$, with $p_i \in D^+u(x_i)$ and $x_i \in A$ as in (3.3.22). Recalling the definition of D^*u, for any $i \in \mathbb{N}$ we can find a vector $p_i^* \in D^*u(x_i)$ and a point $x_i^* \in A$, at which u is differentiable, such that

$$|x_i - x_i^*| + |Du(x_i^*) - p_i^*| \leq \frac{1}{i}\,.$$

Then, $x_i^* \to x$ and

$$\begin{aligned} |Du(x_i^*) - p| &\le |Du(x_i^*) - p_i^*| + |p_i^* - p_i| + |p_i - p| \\ &\le \frac{1}{i} + \operatorname{diam}(D^+u(x_i)) + |p_i - p| \to 0, \end{aligned}$$

as $i \to \infty$. Hence, $p \in D^*u(x)$. ∎

Let us recall a standard definition from convex analysis.

Definition 3.3.9 *Given a convex set $C \subset \mathbb{R}^n$, we say that $x \in C$ is an* extreme point *of C if x cannot be expressed as the convex combination of $x_0, x_1 \in C$, with $x_0, x_1 \neq x$. The set of all the extreme points of C is denoted by* $\operatorname{Ext}(C)$.

Formula (3.3.19) trivially implies that

$$\operatorname{Ext}(D^+u(x)) \subset D^*u(x), \quad \forall x \in A. \tag{3.3.23}$$

One may wonder if, for a semiconcave function, $D^*u(x)$ always reduces to $\operatorname{Ext}(D^+u(x))$. This is not the case, as the following example shows.

Example 3.3.10 Let $u : \mathbb{R}^2 \to \mathbb{R}$ be the concave function

$$u(x, y) = -\sqrt{x^2 + y^4}\,.$$

Then u is differentiable on $\mathbb{R}^2 \setminus \{(0, 0)\}$, whereas

$$D^+u(0, 0) = [-1, 1] \times \{0\}\,.$$

Therefore, $\operatorname{Ext}(D^+u(0, 0)) = \{(-1, 0), (1, 0)\}$. On the other hand, it is easy to see that $D^*u(0, 0) = [-1, 1] \times \{0\}$. Indeed,

$$Du(x, y) = -\left(\frac{x}{\sqrt{x^2 + y^4}}, \frac{2y^3}{\sqrt{x^2 + y^4}}\right), \quad (x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}\,.$$

Hence, taking the limit in (0,0) along the half–lines

$$(x, y) = t(\cos\theta, \sin\theta), t > 0, \theta \in [0, 2\pi]\,,$$

we conclude that

$$[-1, 1] \times \{0\} = \{(\theta, 0) : \theta \in [0, 2\pi]\} \subset D^*u(0, 0)\,.$$

It is well known that the superdifferential of a concave function has a monotonicity property. Semiconcave functions satisfy a weaker form of this property, as shown by the next result.

Proposition 3.3.11 *Let $u : A \to \mathbb{R}$ be a semiconcave function with modulus ω. Given $x, y \in A$ such that $[x, y] \subset A$ we have, for any $p \in D^+u(x)$ and $q \in D^+u(y)$,*

$$\langle q - p, y - x \rangle \le 2|y - x|\omega(|y - x|). \tag{3.3.24}$$

Proof — Applying Proposition 3.3.1 twice, we find

$$-\langle q, x - y \rangle - |y - x|\omega(|y - x|) \le u(y) - u(x) \le \langle p, y - x \rangle + |y - x|\omega(|y - x|).$$

The conclusion follows. ∎

Property (3.3.24) characterizes in fact semiconcave functions, as shown by the following result.

Proposition 3.3.12 *Let $u : A \to \mathbb{R}^n$ be a Lipschitz continuous function and let $\partial_c u$ denote its Clarke gradient. Suppose that there exists a nondecreasing upper semicontinuous function $\omega : [0, \infty[\to [0, \infty[$ such that $\omega(0) = 0$ and such that*

$$\langle q - p, y - x \rangle \le |y - x|\omega(|y - x|). \tag{3.3.25}$$

for all x, y such that $[x, y] \subset A$, $p \in \partial_c u(x)$ and $q \in \partial_c u(y)$. Then u is semiconcave in A with modulus ω.

Proof — We can proceed exactly as in the proof of Proposition 2.1.2, using the nonsmooth version of the mean value theorem (Theorem 3.1.12). ∎

Before giving some further results, let us recall some more terminology from convex analysis.

Definition 3.3.13 *Let $u : A \to \mathbb{R}$ be a Lipschitz function. Given $x \in A$, $\theta \in \mathbb{R}^n \setminus \{0\}$ we set*

$$D^+u(x, \theta) = \{p \in D^+u(x) \ : \ p \cdot \theta \le q \cdot \theta, \ \forall\, q \in D^+u(x)\}.$$

This set is called the exposed face *of $D^+u(x)$ in direction θ. If $D^+u(x, \theta) = \{p\}$, then we say that θ is an* exposing direction *and p is an* exposed point *of $D^+u(x)$. We denote the set of exposed points of $D^+u(x)$ by $Eu(x)$.*

Keeping into account Proposition 3.3.4 we easily obtain that, for a semiconcave function u, the set $D^+u(x, \theta)$ is nonempty, convex and compact for any choice of x, θ.

Proposition 3.3.14 *Let $u : A \to \mathbb{R}$ be semiconcave and let $x \in A$, $\bar{p} \in \mathbb{R}^n$ be given. Suppose that there exist two sequences $\{x_k\} \subset (A \setminus \{x\})$, $\{p_k\} \subset \mathbb{R}^n$ such that*

$$\lim_{k\to\infty} x_k = x, \quad \lim_{k\to\infty} \frac{x_k - x}{|x_k - x|} = \theta, \quad p_k \in D^+u(x_k), \quad \lim_{k\to\infty} p_k = \bar{p}. \quad (3.3.26)$$

for some unit vector θ. Then $\bar{p} \in D^+u(x, \theta)$.

Proof — From Proposition 3.3.4(a) it follows that $\bar{p} \in D^+u(x)$. On the other hand we have, by Proposition 3.3.1,

$$\frac{u(x_k) - u(x)}{|x_k - x|} \geq \langle p_k, \frac{x_k - x}{|x_k - x|}\rangle - \omega(|x - x_k|), \quad \forall\, k \geq 1.$$

From property (3.1.5) and from Theorem 3.2.1 we deduce, letting $k \to \infty$,

$$\partial^+ u(x, \theta) \geq \bar{p} \cdot \theta.$$

On the other hand, any vector $p \in D^+u(x)$ satisfies the opposite inequality by Proposition 3.1.5(a). We conclude

$$\bar{p} \cdot \theta = \partial^+ u(x, \theta) = \min_{p \in D^+u(x)} p \cdot \theta. \ \blacksquare$$

Let us denote by $\Sigma(u)$ the set of points where a semiconcave function $u : A \to \mathbb{R}$ is not differentiable. As we have already remarked, such a set has zero measure by Rademacher's theorem.

Before giving our next result, we recall a notion from measure theory. Given a function $w \in L^1(A, \mathbb{R}^M)$ and $x_0 \in A$, we call a value $\bar{w} \in \mathbb{R}^M$ the *approximate limit* of w at y_0 if

$$\lim_{\rho\to 0^+} \frac{\text{meas}\,(\{y \in B_\rho(y_0) \mid |w(y) - \bar{w}| > \varepsilon\})}{\rho^N} = 0$$

for any $\varepsilon > 0$. In this case we write $\bar{w} = \text{ ap } \lim_{y\to y_0} w(y)$.

Then we have the following result showing, roughly speaking, that for the gradient of a semiconcave function the notions of limit and of approximate limit coincide.

Theorem 3.3.15 *Let $u : \Omega \to \mathbb{R}$ be a semiconcave function. Then the following properties are equivalent for a point $x \in \Omega$:*

(i) $\text{ap}\lim_{y\to x} Du(y) = p_1$*;*

(ii) $\lim\limits_{\substack{y \to x \\ y \notin \Sigma(u)}} Du(y) = p_1$;

(iii) *u is differentiable at x and $Du(x) = p_1$.*

Proof — Property (ii) can be restated as $D^*u(x) = \{p_1\}$ and so is equivalent to (iii) thanks to property (3.3.19) and to Proposition 3.3.4(d). In addition, property (ii) implies (i) because the set $\Sigma(u)$ has zero measure. Therefore it suffices to prove that property (i) implies (ii).

Let us argue by contradiction and suppose that the set $D^*u(x)$ contains a vector p_2 with $p_2 \neq p_1$. We set

$$\nu = \frac{p_2 - p_1}{|p_2 - p_1|}.$$

Given $\varepsilon > 0$, we define

$$P_\varepsilon^- = \left\{ p \in D^+u(x) \mid \langle p - p_2, \nu \rangle \geq -\varepsilon |p - p_2| \right\},$$

$$P_\varepsilon^+ = \left\{ p \in D^+u(x) \mid \langle p - p_1, \nu \rangle \leq \varepsilon |p - p_1| \right\}.$$

We claim that P_ε^+ and P_ε^- are disjoint if ε is small enough. In fact, if $q \in P_\varepsilon^+ \cap P_\varepsilon^-$,

$$\begin{aligned} |p_2 - p_1|^2 &= \langle q - p_1, p_2 - p_1 \rangle - \langle q - p_2, p_2 - p_1 \rangle \\ &= |p_2 - p_1| (\langle q - p_1, \nu \rangle - \langle q - p_2, \nu \rangle) \\ &\leq \varepsilon |p_2 - p_1| (|q - p_1| + |q - p_2|) \\ &\leq 2\varepsilon |p_2 - p_1| \operatorname{diam} D^+u(x). \end{aligned}$$

Thus, $P_\varepsilon^+ \cap P_\varepsilon^- = \emptyset$ provided

$$\varepsilon < \frac{|p_2 - p_1|}{2 \operatorname{diam} D^+v(y_0)}.$$

Now let ν' be a unit vector such that $|\nu - \nu'| \leq \varepsilon$. If $p \in D^+u(x, \nu')$ then

$$\nu' \cdot p \leq \nu' \cdot p_1,$$

and so

$$\langle \nu, p - p_1 \rangle \leq \langle \nu - \nu', p - p_1 \rangle \leq \varepsilon |p - p_1|.$$

Thus $D^+u(x, \nu') \subset P_\varepsilon^+$. Similarly we can show that $D^+u(x, -\nu') \subset P_\varepsilon^-$.

Let us now define

$$K_\varepsilon^- = \left\{ y \in \Omega \setminus \{x\} \mid \left| \frac{y - x}{|y - x|} + \nu \right| < \varepsilon \right\},$$

$$K_\varepsilon^+ = \left\{ y \in \Omega \setminus \{x\} \mid \left| \frac{y-x}{|y-x|} - \nu \right| < \varepsilon \right\}.$$

The sets $K_\varepsilon^\pm$ are the intersection of Ω with open cones of vertex x. Therefore

$$\lim_{\rho \to 0^+} \frac{\text{meas } (B_\rho(x) \cap K_\varepsilon^\pm)}{\rho^n} > 0.$$

We claim now that (i) cannot be satisfied. Suppose in fact that we have ap $\lim_{y \to x} Du(y) = p_1$. Then there exists at least a sequence $\{y_k\} \subset K_\varepsilon^-$ such that

$$y_k \to x, \qquad Du(y_k) \to p_1.$$

By passing to a subsequence, we may assume

$$\frac{y_k - y_0}{|y_k - y_0|} \to -\nu',$$

for some ν' such that $|\nu - \nu'| \leq \varepsilon$. By Proposition 3.3.14 we deduce that $p_1 \in D^+u(x, -\nu')$, which in turn implies that $p_1 \in P_\varepsilon^-$. Since $p_1 \in P_\varepsilon^+$ by definition, this implies that P_ε^- and P_ε^+ have nonempty intersection, and this is a contradiction for ε small enough. ∎

3.4 Marginal functions

A function $u : A \to \mathbb{R}$, is called a *marginal function* if it can be written in the form

$$u(x) = \inf_{s \in S} F(s, x), \tag{3.4.27}$$

where S is some topological space and the function $F : S \times A \to \mathbb{R}$ satisfies some regularity assumptions. In many cases a marginal function is semiconcave. For instance, Corollary 2.1.6 immediately implies the following.

Proposition 3.4.1 *Let $A \subset \mathbb{R}^n$ be open and let $S \subset \mathbb{R}^m$ be compact. If $F = F(s, x)$ is continuous in $C(S \times A)$ together with its partial gradient $D_x F$, then the function u defined in (3.4.27) belongs to* SC(A). *If $D^2_{xx} F$ also exists and is continuous in $S \times A$, then $u \in$* SCL(A).

We now show that the converse also holds.

Theorem 3.4.2 *Let $u : A \to \mathbb{R}$ be a semiconcave function. Then u can be locally written as the minimum of functions of class C^1. More precisely, for any $K \subset A$ compact, there exists a compact set $S \subset \mathbb{R}^{2n}$ and a continuous* *function $F : S \times K \to \mathbb{R}$ such that $F(s,\cdot)$ is C^1 for any $s \in S$, the gradients $D_xF(s,\cdot)$ are equicontinuous, and*

$$u(x) = \min_{s \in S} F(s,x), \qquad \forall\, x \in K. \tag{3.4.28}$$

If the modulus of semiconcavity of u is linear, then F can be chosen such that $F(s,\cdot)$ is C^2 for any s, with uniformly bounded C^2 norm.

Proof — Let ω be the modulus of semiconcavity of u and let ω_1 be a function such that $\omega_1(0) = 0$, that $\omega_1(r) \geq \omega(r)$ and that the function $x \to |x|\omega_1(|x|)$ belongs to $C^1(\mathbb{R}^n)$. The existence of such an ω_1 has been proved in Lemma 3.1.8. If ω is linear we simply take $\omega_1 \equiv \omega$.

Let us set $S = \{(x,p) : x \in K, p \in D^+u(x)\}$. By Proposition 3.3.4(a), S is a compact set. Then we define

$$F(y,p,x) = u(y) + \langle p, x-y\rangle + |y-x|\omega_1(|y-x|).$$

Then F has the required regularity properties and (3.4.28) follows from Proposition 3.3.1. ■

When we have a representation of the form (3.4.27) for a semiconcave function, it is interesting to relate the generalized differentials of u with the partial derivative D_xF.

Theorem 3.4.3 *Let $A \subset \mathbb{R}^n$ be open and let $S \subset \mathbb{R}^m$ be compact. Let the function $F = F(s,x)$ be continuous in $S \times A$ together with its partial derivative D_xF, and let us define $u(x) = \min_{s\in S} F(s,x)$. Given $x \in A$, let us set*

$$M(x) = \{s \in S : u(x) = F(s,x)\}, \quad Y(x) = \{D_xF(s,x) : s \in M(x)\}.$$

Then, for any $x \in A$

$$D^+u(x) = \mathrm{co}Y(x) \tag{3.4.29}$$

$$D^-u(x) = \begin{cases} \{p\} & \text{if } Y(x) = \{p\} \\ \emptyset & \text{if } Y(x) \text{ is not a singleton.} \end{cases} \tag{3.4.30}$$

Proof — From Proposition 3.4.1 we know that u is locally semiconcave in A and so we can apply the results of the previous sections. We first observe that, if $s \in M(x)$, then $F(s,\cdot)$ touches u from above at the point x, and therefore $D_xF(s,x) \in D^+u(x)$. This shows that $Y(x)$ is contained in $D^+u(x)$; since $D^+u(x)$ is convex, the same holds for the convex hull of $Y(x)$.

Let us now prove that $D^+u(x)$ is contained in the convex hull of $Y(x)$. In view of (3.3.19), it is enough to prove that $D^*u(x) \subset Y(x)$. Let us therefore pick $p \in D^*u(x)$. Let $\{x_n\}$ be a sequence of points of differentiability of u such that $x_n \to x$ and $Du(x_n) \to p$. Let $s_n \in S$ be such that $u(x_n) = F(s_n, x_n)$. By the first part of the proof, we have $D_xF(s_n, x_n) = Du(x_n)$. By possibly extracting a subsequence, we can assume that $s_n \to \bar{s}$ for some $\bar{s} \in S$. But then we have

$$u(x) = \lim u(x_n) = \lim F(s_n, x_n) = F(\bar{s}, x),$$

$$p = \lim Du(x_n) = \lim D_xF(s_n, x_n) = D_xF(\bar{s}, x).$$

This shows that $p \in Y(x)$ and concludes the proof of (3.4.29). To obtain (3.4.30), it suffices to recall Proposition 3.3.4(d) and that $D^-u(x)$ is necessarily empty if $D^+u(x)$ contains more than one element. ∎

3.5 Inf–convolutions

Given $g : \mathbb{R}^n \to \mathbb{R}$ and $\varepsilon > 0$, the functions

$$x \to \inf_{y \in \mathbb{R}^n} \left(g(y) + \frac{|x-y|^2}{2\varepsilon} \right) \qquad x \to \sup_{y \in \mathbb{R}^n} \left(g(y) - \frac{|x-y|^2}{2\varepsilon} \right)$$

are called inf– and sup–convolutions of g respectively, due to the formal analogy with the usual convolution. They have been used in various contexts as a way to approximate g; one example is the uniqueness theory for viscosity solutions of Hamilton–Jacobi equations. In some cases it is useful to consider more general expressions, where the quadratic term above is replaced by some other coercive function. In this section we analyze such general convolutions, showing that their regularity properties are strictly related with the properties of semiconcave functions studied in the previous sections.

Definition 3.5.1 *Let $g \in C(\mathbb{R}^n)$ satisfy $|g(x)| \le K(1 + |x|)$ for some $K > 0$ and let $\phi \in C(\mathbb{R}^n)$ be such that*

$$\lim_{|q| \to +\infty} \frac{\phi(q)}{|q|} = +\infty. \tag{3.5.31}$$

The inf–convolution *of g with respect to ϕ is the function*

$$g_\phi(x) = \inf_{y \in \mathbb{R}^n} [g(y) + \phi(x-y)], \tag{3.5.32}$$

while the sup–convolution *of g with respect to ϕ is defined by*

$$g^\phi(x) = \sup_{y \in \mathbb{R}^n} [g(y) - \phi(x-y)]. \tag{3.5.33}$$

We observe that the function u given by Hopf's formula (1.2.7) is an inf–convolution with respect to the x variable for any fixed t. In addition, inf–convolutions are a particular case of the marginal functions introduced in the previous section.

We give below some regularity properties of the inf–convolutions. The corresponding statements about the sup–convolutions are easily obtained observing that $g^\phi = -((-g)_\phi)$. Let us observe that our continuity and growth assumptions on g and ϕ imply that the infimum in (3.5.32) is attained for any $x \in \mathbb{R}^n$.

Theorem 3.5.2 *We have*

(i) *If ϕ is differentiable and $D\phi$ is uniformly continuous with modulus of continuity ω, then g_ϕ is semiconcave with modulus ω.*

(ii) *If g is Lipschitz continuous with constant L_g, then so is g_ϕ.*

(iii) *If g is semiconcave with modulus ω, then so is g_ϕ.*

(iv) *If g and ϕ are both convex, then so is g_ϕ.*

(v) *Suppose that ϕ is uniformly convex, that is, that there exists $C > 0$ such that $\phi(x) - C|x|^2$ is convex. Suppose also that g is semiconvex with modulus $\omega(r) = Br$ for some $B < C$. Then g_ϕ is semiconvex with modulus $r \to BC(C-B)^{-1}r$.*

Proof — Part (i) is an application of Proposition 2.1.5 and Theorem 2.1.11. To prove (ii) we observe that, for any $x, h \in \mathbb{R}^n$,

$$\begin{aligned} g_\phi(x+h) &= \min_{y\in\mathbb{R}^n} [g(y) + \phi(x+h-y)] \\ &= \min_{z\in\mathbb{R}^n} [g(z+h) + \phi(x-z)] \\ &\geq -L_g|h| + \min_{z\in\mathbb{R}^n} [g(z) + \phi(x-z)] \\ &= g_\phi(x) - L_g|h|. \end{aligned}$$

Let us now check (iii). For given $x_1, x_2 \in \mathbb{R}^n$ and $\lambda \in]0,1[$ let us set $x = \lambda x_1 + (1-\lambda)x_2$. We then call y_0 the point where the infimum in (3.5.32) is attained. If we denote by ω the semiconcavity modulus of g we obtain

$$\begin{aligned} &\lambda g_\phi(x_1) + (1-\lambda)g_\phi(x_2) - g_\phi(\lambda x_1 + (1-\lambda)x_2) \\ &\leq \lambda [\phi(x_1 - y_0 - (1-\lambda)(x_1-x_2)) + g(y_0 + (1-\lambda)(x_1-x_2))] \\ &\quad +(1-\lambda)[\phi(x_2 - y_0 - \lambda(x_2-x_1)) + g(y_0 + \lambda(x_2-x_1))] \\ &\quad -[\phi(x-y_0) + g(y_0)] \\ &= \lambda g(y_0 + (1-\lambda)(x_1-x_2)) + (1-\lambda)g(y_0+\lambda(x_2-x_1)) - g(y_0) \\ &\leq \lambda(1-\lambda)|x_2-x_1|\omega(|x_2-x_1|), \end{aligned}$$

which is the desired estimate.

We now turn to (iv) and (v). We take $x_1, x_2 \in \mathbb{R}^n$, $\lambda \in]0,1[$ and denote by y_i the points where the infimum in (3.5.32) with $x = x_i$ is attained, for $i = 1, 2$. We find by definition

$$\begin{aligned}
&\lambda g_\phi(x_1) + (1-\lambda) g_\phi(x_2) - g_\phi(\lambda x_1 + (1-\lambda)x_2) \\
&\quad \geq \lambda\phi(x_1 - y_1) + (1-\lambda)\phi(x_2 - y_2) - \phi(\lambda(x_1 - y_1) + (1-\lambda)(x_2 - y_2)) \\
&\quad\quad + \lambda g(y_1) + (1-\lambda) g(y_2) - g(\lambda y_1 + (1-\lambda) y_2). \qquad (3.5.34)
\end{aligned}$$

In case (iv) the above quantity is nonnegative, and this shows that g_ϕ is convex. In case (v) we have

$$\begin{aligned}
&\lambda\phi(x_1 - y_1) + (1-\lambda)\phi(x_2 - y_2) - \phi(\lambda(x_1 - y_1) + (1-\lambda)(x_2 - y_2)) \\
&\quad \geq C[\lambda|x_1 - y_1|^2 + (1-\lambda)|x_2 - y_2|^2 - |\lambda(x_1 - y_1) + (1-\lambda)(x_2 - y_2)|^2] \\
&\quad = C\lambda(1-\lambda)[|x_1 - x_2|^2 + |y_1 - y_2|^2 - 2\langle x_2 - x_1, y_2 - y_1\rangle], \\
&\quad \geq C\lambda(1-\lambda)\left[-\frac{C}{C-B}|x_1 - x_2|^2 + \frac{B}{C}|y_1 - y_2|^2\right]
\end{aligned}$$

and so we deduce from (3.5.34)

$$\lambda g_\phi(x_1) + (1-\lambda) g_\phi(x_2) - g_\phi(\lambda x_1 + (1-\lambda)x_2) \geq -\lambda(1-\lambda)\frac{BC}{C-B}|x_1 - x_2|^2$$

which proves (v). ∎

From the general results about marginal functions we can deduce the following description of the semidifferentials of an inf–convolution.

Proposition 3.5.3 *Let g, ϕ be as Definition 3.5.1, and suppose in addition that ϕ is of class C^1. Given $\bar{x} \in \mathbb{R}^n$, let us denote by $Y(\bar{x})$ the set of the points where the minimum in (3.5.32) with $x = \bar{x}$ is attained. Then g_ϕ is differentiable at $\bar{x}$ if and only if $Y(\bar{x})$ is a singleton. More precisely, if $Y(\bar{x}) = \{\bar{y}\}$ we have $Dg_\phi = D\phi(\bar{y} - \bar{x})$, while if $Y(\bar{x})$ contains more than one point we have*

$$D^+ g_\phi(\bar{x}) = \operatorname{co}\{D\phi(y - \bar{x}) \ : \ y \in Y(\bar{x})\}, \qquad D^- g_\phi(\bar{x}) = \emptyset.$$

Proof — The result is a consequence of Theorem 3.4.3. The deduction is not immediate, since inf–convolutions are defined by a minimization over the whole space, while in Theorem 3.4.3 the space S is supposed compact. However, the lack of compactness is only apparent. In fact, using our assumptions on g and ϕ (see Definition 3.5.1), it is not difficult to prove the following property: given $r > 0$, there exists $R > 0$ such that $Y(x) \subset B_R$

for all $x \in B_r$. Therefore, our inf–convolutions can be locally written as an infimum over a compact space and Theorem 3.4.3 can be applied. ∎

We now give an example of how inf– and sup-convolutions can be used to approximate a given function g. We restrict ourselves to the case when g is Lipschitz continuous, but we point out that such convolutions have good approximating properties also in more general cases, for instance when g is just continuous or semicontinuous, or when g is a function defined on a Hilbert space.

Definition 3.5.4 *Given $g \in \mathrm{Lip}\,(\mathbb{R}^n)$, $\phi \in C(\mathbb{R}^n)$ satisfying* (3.5.31) *and $\varepsilon > 0$ we define*

$$g_\varepsilon(x) = \min_{y \in \mathbb{R}^n} \left[g(y) + \varepsilon \phi \left(\frac{x-y}{\varepsilon} \right) \right], \qquad x \in \mathbb{R}^n, \tag{3.5.35}$$

$$g^\varepsilon(x) = \max_{y \in \mathbb{R}^n} \left[g(y) - \varepsilon \phi \left(\frac{x-y}{\varepsilon} \right) \right], \qquad x \in \mathbb{R}^n. \tag{3.5.36}$$

Remark 3.5.5 If ϕ is convex, the above definitions are related to the theory of Hamilton-Jacobi equations. To see this, let us denote by H the Legendre transform of ϕ. Then, recalling Hopf's formula, $g_\varepsilon = u(\varepsilon, \cdot)$, where u is the solution of the equation $u_t + H(\nabla u) = 0$ with initial value $u(0, \cdot) = g$. Analogously, g^ε is the solution of $u_t - H(\nabla u) = 0$, with initial value g, evaluated at time $t = \varepsilon$.

In the sequel we will also consider compositions of the form $(g^\varepsilon)_\delta, (g_\varepsilon)^\delta$, which are called *sup–inf–* and *inf–sup–convolutions*. Let us observe a preliminary property.

Lemma 3.5.6 *Let us denote by $M_\varepsilon(x)$ and $M^\varepsilon(x)$ the set of points where the minimum in* (3.5.35) *and the maximum* (3.5.36) *respectively is attained. Then there exists $R > 0$, independent of ε, such that any point $y \in M_\varepsilon(x) \cup M^\varepsilon(x)$ satisfies $|y - x| < \varepsilon R$.*

Proof — Let us denote by L_g the Lipschitz constant of g and let R be such that $(\phi(x) - \phi(0))/|x| > L_g$ for $|x| > R$. Such an R exists by assumption (3.5.31). Given $x \in \mathbb{R}^n$, suppose that $y \in M_\varepsilon(x)$. Then we have

$$\begin{aligned} g(y) + \varepsilon \phi \left(\frac{x-y}{\varepsilon} \right) &\leq g(x) + \varepsilon \phi(0) \\ &\leq g(y) + L_g |x - y| + \varepsilon \phi(0), \end{aligned}$$

which implies

$$L_g \frac{|x-y|}{\varepsilon} \geq \phi \left(\frac{x-y}{\varepsilon} \right) - \phi(0).$$

By definition of R, we obtain that $|x-y| \leq \varepsilon R$. The proof for $M^\varepsilon(x)$ is analogous. ∎

Theorem 3.5.7 *Let $g \in \mathrm{Lip}\,(\mathbb{R}^n)$, and let $\phi \in C(\mathbb{R}^n)$ satisfy (3.5.31). Then we have, with the notation of Definition 3.5.4,*

$$\lim_{\varepsilon\to 0^+}(g^\varepsilon)(x) = \lim_{\varepsilon\to 0^+}(g_\varepsilon)(x) = g(x), \qquad x \in \mathbb{R}^n \text{ uniformly} \tag{3.5.37}$$

$$\lim_{\varepsilon,\delta\to 0^+}(g^\varepsilon)_\delta(x) = \lim_{\varepsilon,\delta\to 0^+}(g_\varepsilon)^\delta(x) = g(x), \qquad x \in \mathbb{R}^n \text{ uniformly}\,. \tag{3.5.38}$$

Suppose in addition that ϕ is convex and belongs to $C^2(\mathbb{R}^n)$. Let $R>0$ be as in the previous Lemma and let $\lambda, \Lambda > 0$ be such that

$$2\lambda I \leq D^2\phi(q) \leq 2\Lambda I, \qquad \forall q \in B_R. \tag{3.5.39}$$

Then, for any $\varepsilon > 0$, g_ε (resp. g^ε) is semiconcave (resp. semiconvex) with modulus $\omega(\rho) = \Lambda\rho/\varepsilon$. In addition, if $\delta < \Lambda^{-1}\lambda\varepsilon$, the functions $(g^\varepsilon)_\delta$ and $(g_\varepsilon)^\delta$ are of class $C^{1,1}$.

Proof — For given $x \in \mathbb{R}^n$ and $\varepsilon > 0$, let us take $y_\varepsilon \in M_\varepsilon(x)$. Recalling the definition of g_ε and Lemma 3.5.6 we find

$$\begin{aligned}\varepsilon\phi(0) &\geq g_\varepsilon(x) - g(x) \geq g(y_\varepsilon) - g(x) + \varepsilon \min\phi \\ &\geq -L_g|y_\varepsilon - x| + \varepsilon\min\phi \geq -\varepsilon(L_g R - \min\phi),\end{aligned}$$

and therefore, setting $M = \max\{\phi(0), L_g R - \min\phi\}$, we have that $|g(x) - g_\varepsilon(x)| \leq \varepsilon M$ for any x. Similarly, we find $|g(x) - g^\varepsilon(x)| \leq \varepsilon M$. Since g_ε has the same Lipschitz constant of g, we also find

$$|g(x) - (g_\varepsilon)^\delta(x)| \leq |g(x) - g_\varepsilon(x)| + |g_\varepsilon(x) - (g_\varepsilon)^\delta(x)| \leq (\varepsilon+\delta)M,$$

and similarly for $(g^\varepsilon)_\delta$. This proves the first part of the statement.

The assertion on the semiconcavity of g_ε and the semiconvexity of g^ε follows from Proposition 1.5.2, from Theorem 3.5.2(ii) and from Lemma 3.5.6.

Let us now consider the inf–sup convolution $(g_\varepsilon)^\delta$. From the previous property we know that $(g_\varepsilon)^\delta$ is semiconvex with modulus $\omega(\rho) = \Lambda\delta^{-1}\rho$ and that g_ε is semiconcave with modulus $\omega(\rho) = \Lambda\varepsilon^{-1}\rho$. We now apply Theorem 3.5.2–(iv) with $\phi(x)$ replaced by $\delta\phi(\delta^{-1}x)$, and g replaced by $-(g_\varepsilon)$. Our assumption on δ implies that the hypotheses of the theorem are satisfied with $C = \delta^{-1}\lambda$ and $B = \Lambda\varepsilon^{-1}$. Thus we obtain that $(g_\varepsilon)^\delta = (-(g_\varepsilon))_\delta$ is semiconvex with linear modulus. Then, by Proposition 2.1.3, we conclude that $(g_\varepsilon)^\delta$ is of class $C^{1,1}$. ∎

Remark 3.5.8 If we take the standard kernel $\phi(x) = |x|^2/2$ then (3.5.39) holds with $\lambda = \Lambda = 1$ and so $(g_\varepsilon)^\delta$ is smooth whenever $0 < \delta < \varepsilon$. One may wonder whether the same holds for a general smooth convex superlinear ϕ. As observed in Remark 3.5.5, in this case inf- and sup-convolutions have an interpretation in terms of Hamilton-Jacobi equations. The smoothness property conjectured here would show an interesting connection between nondifferentiability of solutions and irreversibility in time. This property has been proved in [19] in the one–dimensional case, while it remains a conjecture when $n > 1$.

Bibliographical notes.

The notion of superdifferential is classical in convex analysis. Subsequently several kinds of generalized gradients have been introduced in the framework of nonsmooth analysis, see e.g. [31, 32]. The importance of the Fréchet semidifferentials for the study of Hamilton–Jacobi equations was pointed out in [36].

The results of Sections 3.2 and 3.3 are well known from convex analysis in the case of concave functions. The proofs given here for general semiconcave functions follow [27, 28, 22].

Properties of pointwise infima of functions have been widely studied in the literature (see e.g. Section 2.8 in [31] or Section II.2 in [17] and the references therein), but their treatment is usually not related with the semiconcavity property. Let us also recall that the equivalence between the monotonicity properties of the Clarke differential and the property of being the infimum of smooth functions has been analyzed by Spingarn [77] and Rockafellar [73].

A detailed treatment of inf–convolutions with a quadratic kernel in Hilbert spaces is given in [64]. These convolutions have been applied to obtain uniqueness results for viscosity solutions of Hamilton–Jacobi equations (see [58], [37], [48]). Convolutions with more general kernels have been considered in [57].

Chapter 4

Singularities of Semiconcave Functions

By a *singular point*, or *singularity* of a semiconcave function u we mean a point where u is not differentiable. This chapter is devoted to the analysis of such points. The set of all points which are singular for u is called *singular set* and is denoted here by $\Sigma(u)$. As we have already remarked, the singular set of a semiconcave function is negligible by Rademacher's theorem. However, we will see that much more detailed properties can be proved.

In Section 4.1 we study the rectifiability properties of the singular set $\Sigma(u)$. We divide $\Sigma(u)$ according to the dimension of the superdifferential of u denoting by $\Sigma^k(u)$ the set of points x such that $D^+u(x)$ has dimension at least k. Then our result (Corollary 4.1.12) states that $\Sigma^k(u)$ is countably $(n-k)$–rectifiable for all integers $k = 1, \ldots, n$.

Sections 4.2 and 4.3 are devoted to the propagation of singularities for semiconcave functions with linear modulus: given a singular point x_0, we look for conditions ensuring that x_0 belongs to a connected component of dimension $\nu \geq 1$ of the singular set. We study first the propagation along Lipschitz arcs and then along Lipschitz manifold of higher dimension. In general we find that a sufficient condition for the propagation of singularities from x_0 is that the inclusion $D^*u(x_0) \subset \partial D^+u(x_0)$ (see Proposition 3.3.4–(b)) is strict.

As an application of the results of this chapter and of the previous one, we analyze in Section 4.4 some properties of the distance function from a closed set S. We show for instance that the differentiability of the distance function at a point x_0 is equivalent to the uniqueness of the projection of x_0 on S. Using our propagation results, we show that the distance function has no isolated singularities except for the special case when the singularity is the center of a spherical connected component of the complement of S. In general, we show that a point x_0 which is singular for the distance function belongs to a connected set of singular points whose Hausdorff dimension is at least $n-k$, with $k=\dim(D^+u(x_0))$.

4.1 Rectifiability of the singular sets

Throughout this chapter $\Omega \subset \mathbb{R}^n$ is an open set and $u : \Omega \to \mathbb{R}$ is a semiconcave function. As in the previous chapter, we denote by $\Sigma(u)$ the set of points of Ω where u is not differentiable and we call it the *singular set* of u. Let us recall the following definitions.

Definition 4.1.1 *Let $k \in \{0, 1, \ldots, n\}$ and let $C \subset \mathbb{R}^n$.*

(i) *C is called a k–*rectifiable *set if there exists a Lipschitz continuous function $f : \mathbb{R}^k \to \mathbb{R}^n$ such that $C \subset f(\mathbb{R}^k)$.*

(ii) *C is called a* countably k–rectifiable *set if it is the union of a countable family of k–rectifiable sets.*

(iii) *C is called a* countably $\mathcal{H}^k$–rectifiable set *if there exists a countably k–rectifiable set $E \subset \mathbb{R}^n$ such that $\mathcal{H}^k(S \setminus E) = 0$ (see the Appendix for the definition of the Hausdorff measure $\mathcal{H}^k$).*

We have seen in Theorem 3.3.15 that, for the gradient Du of a semiconcave function, the notion of limit and of approximate limit are equivalent. Moreover, we know from Theorem 2.3.1-(ii) that Du is a function with bounded variation if u is semiconcave with a linear modulus. Therefore, general results on BV functions (Proposition A.3.4) yield immediately the following rectifiability property.

Proposition 4.1.2 *Suppose that $u : \Omega \to \mathbb{R}$ is semiconcave with a linear modulus. Then $\Sigma(u)$ coincides with the jump set S_{Du} of the gradient Du considered as a function of $BV_{loc}(\Omega, \mathbb{R}^n)$ and is a $\mathcal{H}^{n-1}$–rectifiable set.*

Using completely independent techniques, we will prove a rectifiability result which is stronger than the one above and which holds for semiconcave functions with arbitrary modulus. For this purpose, we begin with some preliminaries about rectifiable sets and give a rectifiability criterion which will be applied to $\Sigma(u)$ in the following.

In the following we call *k–planes* the k–dimensional subspaces of $\mathbb{R}^n$. Given a k–plane Π, we denote by $\Pi^\perp$ its orthogonal complement. Given $x \in \mathbb{R}^n$, we denote by $\pi(x)$ and $\pi^\perp(x)$ the orthogonal projections of x on Π and $\Pi^\perp$ respectively. We shall hereafter apply the same convention, denoting by upper-case greek letters a $k-$plane and by lower case greek the corresponding orthogonal projection.

Let $A \subset \Pi$, where $\Pi \subset \mathbb{R}^n$ is a k–plane and let $\phi : A \to \Pi^\perp$ be a Lipschitz continuous map. Then the graph

$$\Gamma_\phi := \{x \in \mathbb{R}^n \ : \ \pi(x) \in A, \pi^\perp(x) = \phi(\pi(x))\} \tag{4.1.1}$$

is clearly a k–rectifiable set.

Definition 4.1.3 *Let $\Pi \subset \mathbb{R}^n$ be a k–plane and let $M > 0$. The set*

$$K_M(\Pi) = \{x \in \mathbb{R}^n \ : \ |\pi^\perp(x)| \leq M|\pi(x)|\}$$

is called cone of axis Π and width M.

Proposition 4.1.4 *Given a set $\Gamma \subset \mathbb{R}^n$, a k–plane Π and $M > 0$, the two following properties are equivalent.*

(i) *There exists $A \subset \Pi$ and a Lipschitz function $\phi : A \to \Pi^\perp$ with $\mathrm{Lip}\,(\phi) \leq M$ such that Γ coincides with the graph Γ_ϕ defined in (4.1.1).*

(ii) *For any $x \in \Gamma$, we have $\Gamma \subset x + K_M(\Pi)$.*

Proof — Suppose that (i) holds and let x, x' be any two points of Γ. Then we have

$$|\pi^\perp(x' - x)| = |\phi(\pi(x')) - \phi(\pi(x))| \leq M|\pi(x' - x)|,$$

which means that $x' \in x + K_M(\Pi)$. Therefore, (ii) holds.

Conversely, suppose that (ii) is satisfied. Then we have, for any $x, x' \in \Gamma$,

$$|\pi^\perp(x' - x)| \leq M|\pi(x' - x)|.$$

Thus, if $x, x' \in \Gamma$ and $\pi(x) = \pi(x')$, then $x = x'$. It follows that, for any $y \in \pi(\Gamma)$, there exists a unique $z \in \Pi^\perp$ such that $y+z \in \Gamma$. Setting $A = \pi(\Gamma)$ and $z = \phi(y)$ we have that ϕ is Lipschitz continuous with Lipschitz constant M and that $\Gamma = \Gamma_\phi$. ∎

We give now a rectifiability criterion.

Theorem 4.1.5 *Let $S \subset \mathbb{R}^n$ have the following property: for all $x \in S$ there exist constants $\rho_x > 0$, $M_x \geq 0$ and a k–plane Π_x such that*

$$S \cap B_{\rho_x}(x) \subset x + K_{M_x}(\Pi_x). \tag{4.1.2}$$

Then S is countably k–rectifiable.

Proof — Let us set $\alpha = \sup_{x \in S} M_x$. It is not restrictive to assume that α is finite, since otherwise we can split S in a countable union of subsets with this property.

Let us now set, for any positive integer i,

$$S_i = \left\{ x \in S \ : \ \rho_x > \frac{1}{i} \right\},$$

and let us prove the rectifiability of S_i. Let us fix $\delta > 0$ small enough to have

$$\delta(\alpha+1) < 1, \qquad \frac{\alpha + \delta(\alpha+1)}{1 - \delta(\alpha+1)} \leq 2\alpha. \tag{4.1.3}$$

Let us now choose a family $\{\Pi_1, \ldots, \Pi_N\}$ of k–planes with the property that, given any other k–plane Π, we have

$$\min_{1 \leq j \leq N} ||\pi_j - \pi|| < \delta. \tag{4.1.4}$$

In this formula we have denoted by $||\pi_j - \pi||$ the operator norm of the linear function $\pi_j - \pi$. Let us define, for $j = 1, \ldots, N$,

$$S_{ij} = \{x \in S_i \ : \ ||\pi_x - \pi_j|| < \delta\}.$$

We claim that any set $T \subset S_{ij}$ with diameter less than $1/i$ is contained in a k–dimensional Lipschitz graph with constant not greater than 2α. This proves the theorem, since by (4.1.4) the sets S_{ij} cover S_i.

To prove our claim, let us take $x, x' \in T$. Then, using the definition of S_{ij} and assumption (4.1.2) we obtain

$$\begin{aligned}
|\pi_j^\perp(x - x')| &= |x - x' - \pi_j(x - x')| \\
&= |\pi_x^\perp(x - x') + \pi_x(x - x') - \pi_j(x - x')| \\
&\leq |\pi_x^\perp(x - x')| + \delta|x - x'| \leq \alpha|\pi_x(x - x')| + \delta|x - x'| \\
&\leq \alpha|\pi_j(x - x')| + \delta(\alpha+1)|x - x'| \\
&\leq [\alpha + \delta(\alpha+1)]|\pi_j(x - x')| + \delta(\alpha+1)|\pi_j^\perp(x - x')|.
\end{aligned}$$

Thus, by inequality (4.1.3),

$$|\pi_j^\perp(x - x')| \leq 2\alpha|\pi_j(x - x')|$$

and so $T \subset x + K_{2\alpha}(\Pi_j)$. Our claim follows from Proposition 4.1.4. ∎

We want to derive a corollary of the above criterion. Let us first recall a definition from nonsmooth analysis.

Definition 4.1.6 *Let $S \subset \mathbb{R}^n$ and $x \in \overline{S}$ be given. The* contingent cone *(or* Bouligand's tangent cone*) to S at x is the set*

$$T_S(x) = \left\{ \lim_{i \to \infty} \frac{x_i - x}{t_i} \ : \ x_i \in S, x_i \to x, t_i \in \mathbb{R}_+, t_i \downarrow 0 \right\}.$$

The vector space generated by $T_S(x)$ is called tangent space *to S at x and is denoted by* $\mathrm{Tan}_S(x)$.

Remark 4.1.7 It is easily checked that $T_S(x)$ is a cone and that $\mathrm{Tan}_S(x)$ is in general strictly larger than $T_S(x)$. For instance, if

$$S = \{(x, y) \in \mathbb{R}^2 \ : \ xy = 0\}$$

then $T_S((0,0)) = S$, while $\mathrm{Tan}_S((0,0)) = \mathbb{R}^2$.

Corollary 4.1.8 *Let $S \subset \mathbb{R}^n$ be a set such that*

$$\dim \mathrm{Tan}_S(x) \leq k, \qquad \forall\, x \in S$$

for a given integer $k \in [0, n]$. Then S is countably k–rectifiable.

Proof — Let us check that the hypotheses of Theorem 4.1.5 are satisfied. Given $x \in S$, let Π_x be a k–plane containing $\mathrm{Tan}_S(x)$. We claim that

$$\forall\, M > 0\ \exists\, \rho > 0 \ : \ S \cap B_\rho(x) \subset x + K_M(\Pi_x). \tag{4.1.5}$$

We prove this by contradiction. Suppose that (4.1.5) is not valid. Then there exists $M > 0$ and a sequence $\{x_i\} \subset S$ converging to x and such that

$$|\pi_x^\perp(x_i - x)| > M|\pi_x(x_i - x)|, \qquad \forall\, i. \tag{4.1.6}$$

By passing to a subsequence, we can assume that

$$\frac{x_i - x}{|x_i - x|} \to \theta$$

for some unit vector θ. By definition, θ belongs to $T_S(x)$ and therefore to Π_x. This implies

$$\lim_{i \to \infty} \pi_x^\perp \left(\frac{x_i - x}{|x_i - x|} \right) = 0.$$

On the other hand, by (4.1.6),

$$\left| \pi_x^\perp \left(\frac{x_i - x}{|x_i - x|} \right) \right| > M \left| \pi_x \left(\frac{x_i - x}{|x_i - x|} \right) \right| \to M \text{ as } i \to \infty.$$

The contradiction shows that our claim (4.1.5) is valid, and so we can apply Theorem 4.1.5. ■

We now turn back to the analysis of the singular set $\Sigma(u)$ of a semiconcave function $u : \Omega \to \mathbb{R}^n$. To obtain a fine description, it is convenient to introduce a hierarchy of subsets of $\Sigma(u)$ according to the dimension of the superdifferential.

Definition 4.1.9 *The* magnitude *of a point $x \in \Omega$ (with respect to u) is the integer*

$$\kappa(x) = \dim D^+u(x)\,.$$

Given an integer $k \in \{0, \ldots, n\}$ we set

$$\Sigma^k(u) = \{x \in \Omega \ : \ \kappa(x) = k\}.$$

For our purposes it is useful to introduce also the following sets. Given $\rho > 0$, we denote by $\Sigma^k_\rho(u)$ the set of all $x \in \Omega$ such that $D^+u(x)$ contains a k–dimensional sphere of radius ρ. Observe that $\Sigma^k_\rho(u)$ is contained in the union $\cup_{i=k}^n \Sigma^i(u)$, while in general is not a subset of $\Sigma^k(u)$.

Proposition 4.1.10 *If u is semiconcave in Ω, then $\Sigma^k_\rho(u)$ is closed for any $\rho > 0$ and $k = 0, \ldots, n$.*

Proof — Let $x \in \Omega$ be the limit of a sequence $\{x_i\} \subset \Sigma^k_\rho(u)$. Then, for any i we can find a k–dimensional ball B^i_ρ of radius ρ and contained in $D^+u(x_i)$. These spheres are all contained in a bounded set in $\mathbb{R}^n$ by the local Lipschitz continuity of u. Passing to a subsequence and using the upper semicontinuity of D^+u (see Proposition 3.3.4(a)) we obtain that $D^+u(x)$ contains a k–dimensional ball sphere B_ρ of radius ρ whose points are limit of points of the spheres B^i_ρ. This proves that $x \in \Sigma^k_\rho(u)$. ■

For convenience of notation in the next results, let us set $\Sigma^{n+1}(u) = \emptyset$. We also sometimes write $T(x, S)$ or $\mathrm{Tan}(x, S)$ instead of $T_S(x)$ and $\mathrm{Tan}_S(x)$.

Theorem 4.1.11 *If u is semiconcave in Ω, then*

$$\mathrm{Tan}(x, \Sigma^k_\rho(u)) \subset [D^+u(x)]^\perp, \quad \forall\, x \in \Sigma^k_\rho(u) \cap \Sigma^k(u) \tag{4.1.7}$$

for any $\rho > 0$ and $k = 0, \ldots, n$.

Proof — It suffices to prove the desired inclusion for the tangent cone $T(x, \Sigma^k_\rho(u))$. Let us take any nonzero vector $\theta \in T(x, \Sigma^k_\rho(u))$. Then there exist $x_i \in \Sigma^k_\rho(u)$ and t_i such that, as $i \to \infty$,

$$x_i \to x, \qquad t_i \downarrow 0, \qquad \frac{x_i - x}{t_i} \to \theta.$$

As in the proof of Proposition 4.1.10, we can find a k–dimensional ball $B_\rho \subset D^+u(x)$ whose elements can all be approximated by elements of $D^+u(x_i)$.

Let p, p' be two arbitrary elements of B_ρ and let $p_i \in D^+u(x_i)$ be such that $p_i \to p'$ as $i \to \infty$. From Proposition 3.3.11 we know that

$$\langle p_i - p, x_i - x\rangle \leq 2|x_i - x|\omega(|x_i - x|)$$

where ω is a modulus of semiconcavity for u. Letting $i \to \infty$ we obtain that $\langle p' - p, \theta\rangle \leq 0$. Exchanging the role of p and p' we find that $\langle p' - p, \theta\rangle = 0$ for all $p, p' \in B_\rho$. Therefore the k–plane containing B_ρ is orthogonal to θ. Since $D^+u(x)$ is k–dimensional by assumption and contains B_ρ, we conclude that $D^+u(x)$ is also orthogonal to θ. ∎

Corollary 4.1.12 *If u is semiconcave in Ω, then the set $\Sigma^k(u)$ is countably $(n-k)$–rectifiable for any integer $k = 0, \ldots, n$.*

Proof — From Theorem 4.1.11 we obtain that

$$\dim\left[\mathrm{Tan}\left(x, \Sigma^k_\rho(u)\right)\right] \leq \dim\left[D^+u(x)\right]^\perp = n - k$$

for all $x \in \Sigma^k_\rho(u) \cap \Sigma^k(u)$. We can apply Corollary 4.1.8 to conclude that $\Sigma^k_\rho(u) \cap \Sigma^k(u)$ is countably $(n-k)$–rectifiable. Therefore, $\Sigma^k(u)$ is also countably $(n-k)$–rectifiable. ∎

4.2 Propagation along Lipschitz arcs

Let u be a semiconcave function in an open domain $\Omega \subset \mathbb{R}^n$. The rectifiability properties of $\Sigma(u)$, obtained in the previous section, can be regarded as "upper bounds" for $\Sigma(u)$. From now on, we shall study the singular set of u trying to obtain "lower bounds" for such a set. In the rest of the chapter, we restrict our attention to semiconcave functions with a linear modulus.

Given a point $x_0 \in \Sigma(u)$, we are interested in conditions ensuring the existence of other singular points approaching x_0. The following example explains the nature of such conditions.

Example 4.2.1 The functions

$$u_1(x_1, x_2) = -\sqrt{x_1^2 + x_2^2}\,, \qquad u_2(x_1, x_2) = -|x_1| - |x_2|$$

are concave in $\mathbb{R}^2$, and $(0,0)$ is a singular point for both of them. Moreover, $(0,0)$ is the only singularity for u_1 while

$$\Sigma(u_2) = \{(x_1, x_2) \,:\, x_1 x_2 = 0\}.$$

So, $(0,0)$ is the intersection point of two singular lines of u_2. Notice that $(0,0)$ has magnitude 2 with respect to both functions as

$$D^+u_1(0,0) = \{(p_1,p_2) \,:\, p_1^2 + p_2^2 \le 1\}$$
$$D^+u_2(0,0) = \{(p_1,p_2) \,:\, |p_1| \le 1\,,\ |p_2| \le 1\}.$$

The different structure of $\Sigma(u_1)$ and $\Sigma(u_2)$ in a neighborhood of x_0 is captured by the reachable gradients. In fact,

$$D^*u_1(0,0) = \{(p_1,p_2) \,:\, p_1^2 + p_2^2 = 1\} = \partial D^+u_1(0,0)$$
$$D^*u_2(0,0) = \{(p_1,p_2) \,:\, |p_1| = 1\,,\ |p_2| = 1\} \neq \partial D^+u_2(0,0).$$

In other words, the inclusion $D^*u(x) \subset \partial D^+u(x)$ (see Proposition 3.3.4-(b)) is an equality for u_1, and a proper inclusion for u_2.

The above example suggests that a sufficient condition to exclude the existence of an isolated singular point $x_0 \in \Sigma(u)$ should be that $D^*u(x_0)$ fails to cover the whole boundary of $D^+u(x_0)$. As we shall see, such a condition implies a much stronger property, namely that x_0 is the initial point of a Lipschitz singular arc.

In the following we call an *arc* a continuous map $\mathbf{x} : [0,\rho] \to \mathbb{R}^n$, $\rho > 0$. We shall say that the arc $\mathbf{x}$ is *singular* for u if the support of $\mathbf{x}$ is contained in Ω and $\mathbf{x}(s) \in \Sigma(u)$ for every $s \in [0,\rho]$. The following result describes the "arc structure" of the singular set $\Sigma(u)$.

Theorem 4.2.2 *Let $x_0 \in \Omega$ be a singular point of a function $u \in SCL(\Omega)$. Suppose that*

$$\partial D^+u(x_0) \setminus D^*u(x_0) \neq \emptyset. \tag{4.2.8}$$

Then there exist a Lipschitz singular arc $\mathbf{x} : [0,\rho] \to \mathbb{R}^n$ for u, with $\mathbf{x}(0) = x_0$, and a positive number δ such that

$$\lim_{s\to 0^+} \frac{\mathbf{x}(s) - x_0}{s} \neq 0 \tag{4.2.9}$$
$$\operatorname{diam}\Big(D^+u(\mathbf{x}(s))\Big) \ge \delta \qquad \forall s \in [0,\rho]. \tag{4.2.10}$$

Moreover, $\mathbf{x}(s) \neq x_0$ for any $s \in \,]0,\rho]$.

Observe that Theorem 4.2.2 gives no information on the magnitude of $\mathbf{x}(s)$ as a singular point. However, by Corollary 4.1.12, we can exclude that the support of $\mathbf{x}$ consists entirely of singular points of magnitude n.

Remark 4.2.3 We recall that the *normal cone* to a convex set $S \subset \mathbb{R}^n$ at a point $p_0 \in S$ consists of all $v \in \mathbb{R}^n$ such that

$$\langle v, p - p_0 \rangle \le 0 \qquad \forall\, p \in S.$$

It is easy to see that the normal cone to any $p_0 \in \partial S$ contains some nonzero vector. We note that condition (4.2.8) is equivalent to the existence of two vectors, $p_0 \in \mathbb{R}^n$ and $q \in \mathbb{R}^n \setminus \{0\}$, such that

$$p_0 \in D^+u(x_0) \setminus D^*u(x_0) \tag{4.2.11}$$

$$\langle q, p - p_0 \rangle \ge 0 \qquad \forall\, p \in D^+u(x_0)\,. \tag{4.2.12}$$

Indeed, (4.2.11) and (4.2.12) imply that $p_0 \in \partial D^+u(x_0) \setminus D^*u(x_0)$. Conversely, if (4.2.8) holds true, then (4.2.11) is trivially satisfied by any vector $p_0 \in \partial D^+u(x_0) \setminus D^*u(x_0)$, and (4.2.12) follows taking $-q$ in the normal cone to the convex set $D^+u(x_0)$ at p_0.

Remark 4.2.4 It is easy to see that the support of the singular arc $\mathbf{x}$, given by Theorem 4.2.2, is a connected set of Hausdorff dimension 1. Indeed, from the Lipschitz continuity of $\mathbf{x}$ it follows that the support of $\mathbf{x}$ is 1-rectifiable, while property (4.2.9) implies that the 1-dimensional Hausdorff measure of $\mathbf{x}([0,T])$ is positive.

The idea of the proof of Theorem 4.2.2 relies on the geometric intuition that the distance between the graph of u and a suitable plane through $(x_0, u(x_0))$, transverse the graph of u, should be maximized along the singular arc we expect to find. We will be able to construct such a transverse plane using a vector of the form $p_0 - q$, where p_0 and q are chosen as in Remark 4.2.3. Indeed, condition (4.2.12) implies that $p_0 - q \notin D^+u(x_0)$, and so the graph of

$$x \mapsto u(x_0) + \langle p_0 - q, x - x_0 \rangle \qquad x \in \mathbb{R}^n \tag{4.2.13}$$

is transverse to the graph of u. We single out this step of the proof in the next lemma, as such a technique applies to any point x_0 of the domain of a concave function, regardless of the regularity of u at x_0. For $x_0 \in \Sigma(u)$, we will then show that the arc we construct in this way is singular for u.

Lemma 4.2.5 *Let u be semiconcave in $\overline{B}_R(x_0) \subset \Omega$ with modulus $\omega(\rho) = C\rho$. Fix $p_0 \in \partial D^+u(x_0)$ and let $q \in \mathbb{R}^n \setminus \{0\}$ be such that*

$$\langle q, p - p_0 \rangle \ge 0 \qquad \forall p \in D^+u(x_0)\,. \tag{4.2.14}$$

Define

$$\sigma = \min\left\{ \frac{R}{4|q|}, \frac{1}{4C} \right\}. \tag{4.2.15}$$

Then there exists a Lipschitz arc $\mathbf{x} : [0, \sigma] \to B_R(x_0)$, *with* $\mathbf{x}(0) = x_0$, *such that*

$$0 < |\mathbf{x}(s) - x_0| < 4|q|s \qquad \forall\, s \in]0, \sigma] \tag{4.2.16}$$

$$\lim_{s \downarrow 0} \frac{\mathbf{x}(s) - x_0}{s} = q \tag{4.2.17}$$

$$\mathbf{p}(s) := p_0 + \frac{\mathbf{x}(s) - x_0}{s} - q \in D^+u(\mathbf{x}(s)) \qquad \forall\, s \in]0, \sigma]\,. \tag{4.2.18}$$

Moreover,

$$\mathrm{Lip}\,(\mathbf{x}) \le 4\,L + 2\,|q| \tag{4.2.19}$$

where L *is a Lipschitz constant for* u *in* $B_R(x_0)$.

Proof — Let us define, for any $s \in]0, 1/C]$,

$$\phi_s(x) = u(x) - u(x_0) - \langle p_0 - q, x - x_0 \rangle - \frac{1}{2s}|x - x_0|^2 \qquad x \in \overline{B}_R(x_0)\,.$$

Notice that ϕ_s is the difference between the affine function in (4.2.13) and the function $x \mapsto u(x) - \frac{1}{2s}|x - x_0|^2$, which is a concave perturbation of u.

Being strictly concave, ϕ_s has a unique maximum point in $\overline{B}_R(x_0)$, that we term x_s. For technical reasons that will be clear in the sequel, we restrict our attention to the interval $0 \le s \le \sigma$, where σ is the number given by (4.2.15). Let us define $\mathbf{x}$ by

$$\mathbf{x}(s) = \begin{cases} x_0 & \text{if } s = 0 \\ \\ x_s & \text{if } s \in]0, \sigma]. \end{cases}$$

We now proceed to show that $\mathbf{x}$ possesses all the required properties.

First, we claim that $\mathbf{x}$ satisfies estimate (4.2.16). Indeed, by property (3.3.17) satisfied by the elements of D^+u, we have that

$$\phi_s(x) \le \langle q, x - x_0 \rangle + \left(C - \frac{1}{2s}\right) |x - x_0|^2 \tag{4.2.20}$$

for any $x \in \overline{B}_R(x_0)$. Moreover, $p_0 - q \notin D^+u(x_0)$ in view of condition (4.2.14). Since this fact implies that there are points in $\overline{B}_R(x_0)$ at which ϕ_s is positive, we conclude that $\phi_s(\mathbf{x}(s)) > 0$. The last estimate and (4.2.20) yield

$$0 < |\mathbf{x}(s) - x_0| < \frac{2s|q|}{1 - 2Cs} \qquad \forall s \in]0, \sigma] \tag{4.2.21}$$

and so (4.2.16) follows from (4.2.15).

Second, we proceed to check (4.2.18). For this purpose we note that, on account of estimate (4.2.21), the choice of σ forces $\mathbf{x}(s) \in B_R(x_0)$ for any $s \in [0, \sigma]$. Hence, $\mathbf{x}(s)$ is also a local maximum point of ϕ_s. This implies

$$0 \in D^+\phi_s(\mathbf{x}(s)) = D^+u(\mathbf{x}(s)) - p_0 + q - \frac{\mathbf{x}(s) - x_0}{s}$$

for any $s \in]0, \sigma]$. Clearly, the last inclusion can be recast in the desired form (4.2.18).

Third, to prove (4.2.17), we show that $\lim_{s\to 0} \mathbf{p}(s) = p_0$, where $\mathbf{p}$ is defined in (4.2.18). Let $\overline{p} = \lim_{k\to\infty} \mathbf{p}(s_k)$ for some sequence $s_k \downarrow 0$. Then, taking the scalar product of both sides of the identity

$$\mathbf{p}(s_k) - p_0 + q = \frac{\mathbf{x}(s_k) - x_0}{s_k}\,.$$

with $\mathbf{p}(s_k) - p_0$ and recalling property (3.3.24), we obtain

$$|\mathbf{p}(s_k)-p_0|^2+\langle q, \mathbf{p}(s_k)-p_0\rangle = \frac{1}{s_k}\langle \mathbf{p}(s_k)-p_0, \mathbf{x}(s_k)-x_0\rangle \leq \frac{2C}{s_k}|\mathbf{x}(s_k)-x_0|^2\,. \tag{4.2.22}$$

Now, observe that the right-hand side above tends to 0 as $k \to \infty$, in view of (4.2.21). Moreover, $\overline{p} \in D^+u(x_0)$ since D^+u is upper semicontinuous; so, $\langle q, \overline{p}-p_0\rangle \geq 0$ by assumption (4.2.14). Therefore, (4.2.22) yields $|\overline{p}-p_0|^2 \leq 0$ in the limit as $k \to \infty$. This proves that $\overline{p} = p_0$ as required.

Finally, let us derive the Lipschitz estimate (4.2.19). Let $r, s \in [0, \sigma]$. Using (4.2.18) to evaluate $\mathbf{x}(s)$ and $\mathbf{x}(r)$ one can easily compute that

$$\mathbf{x}(s) - \mathbf{x}(r) = s[\mathbf{p}(s) - \mathbf{p}(r)] + (s - r)[\mathbf{p}(r) - p_0 + q]\,. \tag{4.2.23}$$

Now, taking the scalar product of both sides of above equality with $\mathbf{x}(s) - \mathbf{x}(r)$, and recalling (3.3.24), we obtain

$$|\mathbf{x}(s) - \mathbf{x}(r)|^2 \leq 2Cs|\mathbf{x}(s) - \mathbf{x}(r)|^2 + |s - r|\,|\mathbf{x}(s) - \mathbf{x}(r)|\,|\mathbf{p}(r) - p_0 + q|\,.$$

Hence, for any $r, s \in [0, \sigma]$,

$$(1 - 2Cs)|\mathbf{x}(s) - \mathbf{x}(r)| \leq |\mathbf{p}(r) - p_0 + q|\,|s - r| \leq (2\,L + |q|)\,|s - r|\,,$$

because L provides a bound for D^+u in $B_R(x_0)$. This completes the proof. ∎

Proof of Theorem 4.2.2 — To begin, let us fix a radius $R > 0$ such that $\overline{B}_R(x_0) \subset \Omega$. We recall that, as noted in Remark 4.2.3, the geometric assumption that $\partial D^+u(x_0) \setminus D^*u(x_0)$ be nonempty is equivalent to the existence of vectors $p_0 \in \mathbb{R}^n$ and $q \in \mathbb{R}^n \setminus \{0\}$ satisfying

$$p_0 \in D^+u(x_0) \setminus D^*u(x_0) \quad \& \quad \langle q, p - p_0\rangle \geq 0 \quad \forall\, p \in D^+u(x_0)\,. \tag{4.2.24}$$

Applying Lemma 4.2.5 to such a pair, we can construct two arcs, $\mathbf{x}$ and $\mathbf{p}$, that enjoy properties (4.2.17) and (4.2.18). Moreover, the same lemma ensures that $\mathbf{x}$ is Lipschitz continuous and that $\mathbf{x}(s) \neq x_0$ for any $s \in]0, \sigma]$.

Therefore, it remains to show that the restriction of $\mathbf{x}$ to a suitable subinterval $[0, \rho]$ is singular for u, and that the diameter of $D^+u(\mathbf{x}(s))$ is bounded away from 0 for all $s \in [0, \rho]$. In fact, it suffices to check the latter point. Let us argue by contradiction: suppose that a sequence $s_k \downarrow 0$ exists such that $\mathrm{diam}(D^+u(\mathbf{x}(s_k))) \to 0$ as $k \to \infty$. Then, by Proposition 3.3.8,

$$D^+u(\mathbf{x}(s_k)) \ni \mathbf{p}(s_k) \to p_0 \quad \text{as} \quad k \to \infty \qquad \Longrightarrow \qquad p_0 \in D^*u(x_0)\,,$$

contrary to (4.2.24). ∎

4.3 Singular sets of higher dimension

In the previous section we proved that the singularities of a function $u \in SCL(\Omega)$ propagate along a Lipschitz image of an interval $[0, \rho]$, from any point $x_0 \in \Omega$ at which $D^*u(x_0)$ fails to cover the whole boundary of $D^+u(x_0)$. Such a result describes a sort of basic structure for the propagation of singularities of a concave function. However, concave functions may well present singular sets of dimension greater than 1. The next example is a case in point.

Example 4.3.1 It is easy to check that the singular set of the concave function

$$u(x) = -|x_3| \qquad x = (x_1, x_2, x_3) \in \mathbb{R}^3$$

is given by the coordinate plane

$$\Sigma(u) = \{x \in \mathbb{R}^3 : x_3 = 0\}\,.$$

Moreover,

$$D^*u(0) = \{(0,0,1), (0,0,-1)\} \qquad D^+u(0) = \{(0,0,p_3) : |p_3| \leq 1\}\,.$$

Therefore, one can apply Theorem 4.2.2 with $x_0 = 0$ and $p_0 = 0$, but this procedure only gives a Lipschitz singular arc starting at 0, whereas one would expect a 2-dimensional singular set. Actually, a more careful application of Lemma 4.2.5 suggests that a singular arc for u should correspond to any vector $q \neq 0$ satisfying (4.2.14). Moreover, such a correspondence should be 1-to-1 in light of (4.2.17). Since (4.2.14) is satisfied by all vectors $q \in \mathbb{R}^3$ such that $q_1^2 + q_2^2 = 1$, $q_3 = 0$, one can imagine to construct the whole singular plane $x_3 = 0$ in this way.

The next result generalizes Theorem 4.2.2, showing propagation of singularities along a ν-dimensional set. The integer $\nu \geq 1$ is given by the number of linearly independent directions of the normal cone to the superdifferential of u at the initial singular point, as conjectured in the above example.

In the following we write $N_S(x)$ to denote the normal cone to a convex set S at a point $x \in S$. Moreover, we use the notation

$$A \ni x \to x_0$$

to mean that $x \in A$ and $x \to x_0$. For a family $\{A_i\}_{i\in\mathbb{N}}$ of subsets of $\mathbb{R}^n$, we write

$$A_i \ni x_i \to x$$

to denote a sequence $x_i \in A_i$ converging to x.

Theorem 4.3.2 *Let $x_0 \in \Omega$ be a singular point of a function $u \in SCL(\Omega)$. Suppose that*

$$\partial D^+u(x_0) \setminus D^*u(x_0) \neq \emptyset$$

and, having fixed a point $p_0 \in \partial D^+u(x_0) \setminus D^*u(x_0)$, define

$$\nu := \dim\, N_{D^+u(x_0)}(p_0)\,.$$

Then a number $\rho > 0$ and a Lipschitz map $\mathbf{f} : N_{D^+u(x_0)}(p_0) \cap B_\rho \to \Sigma(u)$ exist such that

$$\begin{aligned} &\mathbf{f}(q) = x_0 - q + |q|\mathbf{h}(q) \\ &\textit{with}\quad \mathbf{h}(q) \to 0 \quad \textit{as} \quad N_{D^+u(x_0)} \cap B_\rho \ni q \to 0 \end{aligned} \tag{4.3.25}$$

$$\liminf_{r\to 0^+} r^{-\nu}\mathcal{H}^\nu\Big(\mathbf{f}(N_{D^+u(x_0)}(p_0) \cap B_\rho) \cap B_r(x_0)\Big) > 0\,. \tag{4.3.26}$$

Moreover,

$$\operatorname{diam}\Big(D^+u(\mathbf{f}(q))\Big) \geq \delta \qquad \forall q \in N_{D^+u(x_0)}(p_0) \cap B_\rho \tag{4.3.27}$$

for some $\delta > 0$.

Remark 4.3.3 As one can easily realize, Theorem 4.3.2 is an extension of Theorem 4.2.2. The property

$$\mathbf{f}(q) \neq x_0 \quad \forall q \in N_{D^+u(x_0)}(p_0) \cap B_\rho \setminus \{0\},$$

though absent from the statement of Theorem 4.3.2, is valid in the present case as well. In fact, it can also be derived from (4.3.25), possibly restricting the domain of $\mathbf{f}$. The reason for treating the one-dimensional case first, is that the Hausdorff estimate for the density of $\Sigma(u)$ is immediate if $\nu = 1$, and so the main ideas of the approach can be more easily understood.

As in the previous section, we first prove a preliminary result.

Lemma 4.3.4 *Let C be a semiconcavity constant for u in $\overline{B}_R(x_0) \subset \Omega$. Fix $p_0 \in \partial D^+u(x_0)$ and define*

$$\sigma = \min\left\{\frac{1}{4C}, \frac{R}{4}\right\}. \tag{4.3.28}$$

Then a Lipschitz map $\mathbf{f} : N_{D^+u(x_0)} \cap B_\sigma \to \mathbb{R}^n$ exists such that

$$\begin{aligned} &\mathbf{f}(q) = x_0 - q + |q|\mathbf{h}(q) \\ &\textit{with} \quad \mathbf{h}(q) \to 0 \quad \textit{as} \quad N_{D^+u(x_0)} \cap B_\sigma \ni q \to 0 \end{aligned} \tag{4.3.29}$$

$$p_0 + \mathbf{h}(q) \in D^+u(\mathbf{f}(q)) \qquad \forall\, q \in N_{D^+u(x_0)}(p_0) \cap B_\sigma\,. \tag{4.3.30}$$

Moreover,

$$\mathrm{Lip}\,(\mathbf{f}) \le 4\,L + 2 \tag{4.3.31}$$

where L is a Lipschitz constant for u in $B_R(x_0)$.

Proof — Let us set, for brevity,

$$N := N_{D^+u(x_0)}(p_0).$$

We proceed as in the proof of Lemma 4.2.5 and consider, for any $q \in N\setminus\{0\}$, the function

$$\phi_q(x) = u(x) - u(x_0) - \left\langle p_0 + \frac{q}{|q|}, x - x_0 \right\rangle - \frac{1}{2|q|}|x - x_0|^2 \qquad x \in \overline{B}_R(x_0)$$

and the point x_q given by the relation

$$\phi_q(x_q) = \max_{x \in \overline{B}_R(x_0)} \phi_q(x)\,.$$

Then, we define σ as in (4.3.28) and $\mathbf{f}$ by

$$\mathbf{f}(q) = \begin{cases} x_0 & \text{if } q = 0 \\ x_q & \text{if } q \in N \cap B_\sigma \setminus \{0\}\,. \end{cases}$$

Arguing as in the proof of Lemma 4.2.5, we obtain

$$0 < |\mathbf{f}(q) - x_0| < \frac{2|q|}{1 - 2C|q|} \qquad \forall q \in N \cap B_\sigma \setminus \{0\} \tag{4.3.32}$$

and

$$p_0 + \frac{\mathbf{f}(q) - x_0 + q}{|q|} \in D^+u(\mathbf{f}(q)) \qquad \forall\, q \in N_{D^+u(x_0)}(p_0) \cap B_\sigma\,. \tag{4.3.33}$$

Hence, denoting by $\mathbf{h}(q)$ the quotient in (4.3.33), assertion (4.3.30) follows.

To prove (4.3.29) we must show that

$$\mathbf{h}(q) \to 0 \qquad \text{as} \qquad N \cap B_\sigma \ni q \to 0. \tag{4.3.34}$$

For this purpose, let $\{q_i\}$ be an arbitrary sequence in $N \cap B_\sigma \setminus \{0\}$ such that $q_i \to 0$. Since $\mathbf{h}$ is bounded, we can extract a subsequence (still termed $\{q_i\}$) such that $\lim_{i\to\infty} \mathbf{h}(q_i)$ exists and

$$\lim_{i\to\infty} \frac{q_i}{|q_i|} = \overline{q}$$

for some $\overline{q} \in N$ satisfying $|\overline{q}| = 1$. We claim that $\lim_{i\to\infty} \mathbf{h}(q_i) = 0$, which in turn implies (4.3.34). Indeed, let us set

$$\overline{p} := p_0 + \lim_{i\to\infty} \mathbf{h}(q_i)$$

and observe that $\overline{p} \in D^+u(x_0)$ as D^+u is upper semicontinuous and $\mathbf{f}$ is continuous at 0. Taking the scalar product of both sides of the identity

$$\mathbf{h}(q_i) - \frac{q_i}{|q_i|} = \frac{\mathbf{f}(q_i) - x_0}{|q_i|}$$

with $\mathbf{h}(q_i)$ and applying inequality (3.3.24), we deduce that

$$\begin{aligned} |\mathbf{h}(q_i)|^2 - \left\langle \frac{q_i}{|q_i|}, \mathbf{h}(q_i) \right\rangle &= \frac{1}{|q_i|}\langle \mathbf{f}(q_i) - x_0, \mathbf{h}(q_i)\rangle \\ &\le \frac{2C}{|q_i|}|\mathbf{f}(q_i) - x_0|^2 \le \frac{2C}{|q_i|}\Big(\frac{2|q_i|}{1 - 2C|q_i|}\Big)^2 \end{aligned}$$

where the last estimate follows from (4.3.32). In the limit as $i \to \infty$, the above inequality yields

$$|\overline{p} - p_0|^2 - \langle \overline{q}, \overline{p} - p_0 \rangle \le 0\,.$$

Hence, recalling that $\overline{q} \in N$, we conclude that $\overline{p} = p_0$. Our claim is thus proved.

The reasoning that shows the Lipschitz estimate (4.3.31) is the same as in the proof of Lemma 4.2.5, and is therefore omitted. ∎

Proof of Theorem 4.3.2 — Keeping the abbreviated notation $N = N_{D^+u(x_0)}(p_0)$ as in the proof of Lemma 4.3.4, let us denote by $L(N)$ the linear subspace generated by N, and by $\pi : \mathbb{R}^n \to L(N)$ the orthogonal projection of $\mathbb{R}^n$ onto $L(N)$.

Having fixed $R > 0$ so that $\overline{B}_R(x_0) \subset \Omega$, let $\mathbf{f} : N \cap B_\sigma \to \mathbb{R}^n$ be the map given by Lemma 4.3.4. Arguing by contradiction—as in the proof of Theorem 4.2.2—the reader can easily show that a suitable restriction of $\mathbf{f}$ to $N \cap B_\rho$, $0 < \rho \leq \sigma$, satisfies (4.3.27). In particular, $\mathbf{f}(q) \in \Sigma(u)$ for any $q \in N \cap B_\rho$. Moreover, (4.3.25) is an immediate consequence of (4.3.29).

Therefore, the only point of the conclusion that needs to be demonstrated, is estimate (4.3.26) for the $\nu-$dimensional Hausdorff density of the singular set $\mathbf{f}(N \cap B_\rho)$, or, since $\mathcal{H}^\nu$ is translation invariant,

$$\liminf_{r \to 0^+} r^{-\nu} \mathcal{H}^\nu \Big([\mathbf{f}(N \cap B_\rho) - x_0] \cap B_r \Big) > 0 \,.$$

We note that the above inequality can be deduced from the lower bound

$$\liminf_{r \to 0^+} r^{-\nu} \mathcal{H}^\nu \Big(\pi([\mathbf{f}(N \cap B_\rho) - x_0] \cap B_r) \Big) > 0 \tag{4.3.35}$$

since

$$\mathcal{H}^\nu \Big([\mathbf{f}(N \cap B_\rho) - x_0] \cap B_r \Big) \geq \mathcal{H}^\nu \Big(\pi([\mathbf{f}(N \cap B_\rho) - x_0] \cap B_r) \Big) \,.$$

Now, using the shorter notation $F(q) = \pi(x_0 - \mathbf{f}(q))$, we observe that (4.3.35) can in turn be derived from the lower bound

$$\liminf_{r \to 0^+} r^{-\nu} \mathcal{H}^\nu (F(N \cap B_r)) > 0 \,. \tag{4.3.36}$$

Indeed, for any sufficiently small r, say $0 < r \leq M\rho$ where $M := \mathrm{Lip}\,(\mathbf{f})$, we have that

$$F(N \cap B_{r/M}) = \pi \Big[x_0 - \mathbf{f}(N \cap B_{r/M}) \Big] \subset \pi \Big[\Big(x_0 - \mathbf{f}(N \cap B_\rho) \Big) \cap B_r \Big] \,.$$

The rest of our reasoning will therefore be devoted to the proof of (4.3.36). To begin, we note that, in view of (4.3.25), the map $F : N \cap B_\rho \to L(N)$ introduced above can be also represented as

$$F(q) = q + H(q)$$

where $H(q)/|q| \to 0$ as $N \cap B_\rho \ni q \to 0$. Consequently, the function

$$\lambda(r) := \sup\{|H(q)| \,:\, q \in N \cap B_r\} \,.$$

satisfies

$$\lim_{r \downarrow 0} \frac{\lambda(r)}{r} = 0 \,. \tag{4.3.37}$$

Let us define

$$N_r := \Big\{ q \in N \cap B_r \ : \ d_{\Gamma_r}(q) \geq \lambda(r) \Big\} \qquad 0 < r \leq \rho \,. \tag{4.3.38}$$

where Γ_r denotes the relative boundary of $N \cap B_r$ and d_{Γ_r} the distance from Γ_r. Using the limit (4.3.37), it is easy to check that $N_r \neq \emptyset$ provided r is sufficiently small, say $0 < r < \rho_0 \leq \rho$. We claim that

$$N_r \subset F(N \cap B_r) \qquad \forall\, r \in (0, \rho_0)\,. \tag{4.3.39}$$

Indeed, having fixed $y \in N_r$, let us rewrite the equation $F(q) = y$ as

$$q = y - H(q) =: H_y(q)\,.$$

Now, observe that $\overline{B}_{\lambda(r)}(y) \subset N \cap B_r$ and that the continuous map $H_y(q)$ satisfies

$$H_y\Big(\overline{B}_{\lambda(r)}(y)\Big) \subset y - H(\overline{N \cap B_r}) \subset y + \overline{B}_{\lambda(r)} = \overline{B}_{\lambda(r)}(y)\,.$$

Therefore, applying Brouwer's Fixed Point Theorem, we conclude that $q = H_y(q)$ has a solution $q \in \overline{B}_{\lambda(r)}(y)$. So, our claim (4.3.39) follows.

Our next step is to obtain the lower bound

$$\liminf_{r \to 0^+} \frac{\mathcal{H}^\nu(N_r)}{r^\nu} > 0 \tag{4.3.40}$$

for the density of the set N_r introduced in (4.3.38). To verify the above estimate let $\overline{q}$ be a point in the relative interior of N, with $|\overline{q}| = 1$. Then, using the notation $\widehat{B}_\alpha := B_\alpha \cap L(N)$ to denote ν-dimensional balls, we have that

$$\overline{q} + \widehat{B}_{2\alpha} \subset N$$

for some $\alpha \in]0, 1/2]$. On account of (4.3.37), there exists $r_0 \in]0, \rho_0[$ such that

$$\lambda(r) \leq \frac{r\alpha}{2} \qquad \forall r \in]0, r_0]\,. \tag{4.3.41}$$

Hence,

$$\frac{r}{2}\Big(\overline{q} + \widehat{B}_{2\alpha}\Big) \subset N \cap B_r \qquad \forall r \in]0, r_0[\,. \tag{4.3.42}$$

Now, combining (4.3.41), (4.3.42) and the definition of N_r, we discover

$$\frac{r}{2}\Big(\overline{q} + \widehat{B}_\alpha\Big) \subset N_r \quad \forall r \in]0, r_0[\,.$$

Estimate (4.3.40) is an immediate consequence of the last inclusion.

Finally, to complete the proof, it suffices to observe that (4.3.36) follows from (4.3.39) and (4.3.40). ∎

Remark 4.3.5 Though clear from (4.3.26), we explicitly note that Theorem 4.3.2 ensures that, in a neighborhood of x_0, $\Sigma(u)$ covers a rectifiable set of Hausdorff dimension ν. Moreover, from formula (4.3.25) it follows that the set $\mathbf{f}(N_{D^+u(x_0)}(p_0) \cap B_\rho) \subset \Sigma(u)$ possesses tangent space at x_0 whenever the normal cone $N_{D^+u(x_0)}(p_0)$ is actually a vector space. This happens, for instance, when x_0 is a singular point of magnitude $k(x_0) < n$ and one can find a point p_0 in the relative interior of $D^+u(x_0)$, but not in $D^*u(x_0)$. Then, it is easy to check that $N_{D^+u(x_0)}(p_0)$ is a vector space of dimension $\nu = n - k(x_0)$.

4.4 Application to the distance function

In this section we examine the singular points of the distance function d_S associated to a nonempty closed subset S of $\mathbb{R}^n$.

We denote by $\mathrm{proj}_S(x)$ the set of closest points in S to x, i.e.

$$\mathrm{proj}_S(x) = \{y \in S \ : \ d_S(x) = |x-y|\} \qquad x \in \mathbb{R}^n\,.$$

The next proposition collects the properties of d_S we need for the analysis.

Proposition 4.4.1 *Let S be a nonempty closed subset of $\mathbb{R}^n$. Then the following holds:*

(i) $d_S \in SCL(\mathbb{R}^n \setminus S)$.

(ii) *d_S is differentiable at $x \notin S$ if and only if $\mathrm{proj}_S(x)$ is a singleton.*

(iii) *For any $x \notin S$*

$$D^*d_S(x) = \left\{ \frac{x-y}{|x-y|} \ : \ y \in \mathrm{proj}_S(x) \right\}\,. \tag{4.4.43}$$

(iv) *For any $x \notin S$ and any $y \in \mathrm{proj}_S(x)$, d_S is differentiable along the segment $]x,y[$.*

Proof — Property (i) is the object of Proposition 2.2.1.
STEP 1. Let us show the implication $\Rightarrow$ in (ii). Fix $\widehat{x} \in \mathrm{proj}_S(x)$. Since d_S^2 is also differentiable at x, for any $y \in \mathbb{R}^n$ we have that

$$\begin{aligned} d_S^2(x) + 2d_S(x)\langle Dd_S(x), y-x\rangle + o(|y-x|) &= d_S^2(y) \\ \le |y-\widehat{x}|^2 = d_S^2(x) + |y-x|^2 &+ 2\langle x-\widehat{x}, y-x\rangle \end{aligned}$$

where $o(|y-x|)/|y-x| \to 0$ as $y \to x$. Therefore, for any $y \in \mathbb{R}^n$,

$$\langle d_S(x)Dd_S(x) - (x-\widehat{x}), y-x\rangle \le o(|y-x|)$$

which yields in turn

$$Dd_S(x) = \frac{x - \widehat{x}}{d_S(x)}\,.$$

We have thus proved that, if d_S is differentiable at $x \notin S$, then

$$\mathrm{proj}_S(x) = \{x - d_S(x)Dd_S(x)\}\,. \tag{4.4.44}$$

STEP 2. The above formula and the definition of reachable gradient imply that

$$x - d_S(x)D^*d_S(x) \subset \mathrm{proj}_S(x) \qquad \forall\, x \notin S\,. \tag{4.4.45}$$

The inclusion we have just obtained is part of assertion (iii).
STEP 3. Another easy consequence of (4.4.45) is the fact that, if $\mathrm{proj}_S(x)$ is a singleton, say $\{\widehat{x}\}$, then

$$D^*d_S(x) = \left\{\frac{x - \widehat{x}}{|x - \widehat{x}|}\right\}\,.$$

So, d_S is differentiable at x and (ii) is completely demonstrated.
STEP 4. We proceed to show point (iv). Given $y \in \mathrm{proj}_S(x)$ and $t \in]0,1[$, let us define

$$y_t := ty + (1-t)x \tag{4.4.46}$$

Observe that

$$d_S(x) = |x - y_t| + |y_t - y| \geq |x - y_t| + d_S(y_t)\,.$$

Let $\widehat{y}_t \in \mathrm{proj}_S(y_t)$. The above inequality yields

$$d_S(x) \leq |x - \widehat{y}_t| \leq |x - y_t| + |y_t - \widehat{y}_t| = |x - y_t| + d_S(y_t) \leq d_S(x)\,.$$

Therefore, all the above inequalities are equalities, $\widehat{y}_t$ lies on the half-line through y_t with initial point x, and so $\widehat{y}_t = y$. Hence, $\mathrm{proj}_S(y_t) = \{y\}$ and d_S is differentiable at y_t by (ii).
STEP 5. To complete the proof of (iii) we have to show the reverse of inclusion (4.4.45). Let $y \in \mathrm{proj}_S(x)$ and let y_t be a point of the segment $]x,y[$, like in (4.4.46). As shown above, d_S is differentiable at y_t and so, recalling (4.4.44),

$$y = y_t - d_S(y_t)Dd_S(y_t) \qquad \forall\, t \in]0,1[\,.$$

Taking the limit as $t \downarrow 0$ we conclude that $y \in x - d_S(x)D^*d_S(x)$, as required. ∎

The following theorem characterizes the isolated singularities of d_S.

Theorem 4.4.2 *Let S be a nonempty closed subset of $\mathbb{R}^n$ and $x \notin S$ be a singular point of d_S. Then the following properties are equivalent:*

(a) *x is an isolated point of $\Sigma(d_S)$.*

(b) *$\partial D^+ d_S(x) = D^* d_S(x)$.*

(c) *$\mathrm{proj}_S(x) = \partial B_r(x)$ where $r := d_S(x)$.*

Proof — $\boxed{\text{(a)}\Rightarrow\text{(b)}}$ This implication is an immediate corollary of the propagation result of section 4.2. Indeed, should $\partial D^+ d_S(x) \setminus D^* d_S(x)$ be nonempty, then Theorem 4.2.2 would ensure the existence of a non-constant singular arc with initial point x. In particular, x could not be isolated.
$\boxed{\text{(b)}\Rightarrow\text{(c)}}$ Assume (b). First, we claim that x must be a singular point of magnitude $\kappa(x) = n$, i.e. $\dim D^+ d_S(x) = n$. For suppose the strict inequality $\kappa(x) < n$ is verified. Then, the whole superdifferential would be made of boundary points and so, owing to (b), $D^+ d_S(x) = D^* d_S(x)$. Therefore, $D^+ d_S(x) \subset \partial B_1$ as all reachable gradients of d_S are unit vectors. But the last inclusion contradicts the fact that $D^+ d_S(x)$ is a convex set of dimension at least 1. Our claim is thus proved.

Now, use the fact that $D^+ d_S(x)$ is an n-dimensional convex set with

$$\partial D^+ d_S(x) = D^* d_S(x) \subset \partial B_1$$

to conclude that $D^+ d_S(x) = \overline{B}_1$ and $D^* d_S(x) = \partial B_1$.

Finally, invoke formula (4.4.43) to discover

$$\mathrm{proj}_S(x) = x - d_S(x) D^* d_S(x) = \partial B_r(x)\,, \tag{4.4.47}$$

as stated in (c).
$\boxed{\text{(c)}\Rightarrow\text{(a)}}$ From Proposition 4.4.1 (iv) we know that d_S is differentiable along each segment $]x, y[$ with $y \in \mathrm{proj}_S(x) = \partial B_r(x)$. So, $d_S \in C^1(B_r(x) \setminus \{x\})$ and the proof is complete. ■

In other words, the previous result shows that a point x_0 is an isolated singularity for the distance function from a set S only if there exists an open sphere B centered at x_0, such that $B \cap S = \emptyset$ and $\partial B \subset S$. In particular, if S is a simply connected set in $\mathbb{R}^2$, or a set in $\mathbb{R}^n$ with trivial $n-1$ homotopy group, then the distance from S has no isolated singularities in the complement of S.

We now apply the theory of section 4.3 to obtain further results on the propagation of singularities for the distance function. In the following, we use the shorter notation

$$P_S(x) = \mathrm{co}[\mathrm{proj}_S(x)].$$

In view of the first identity in (4.4.47),

$$P_S(x) = x - d_S(x)D^+d_S(x)\,. \tag{4.4.48}$$

Theorem 4.4.3 *Let S be a nonempty closed subset of $\mathbb{R}^n$ and $x \notin S$ be a non-isolated singular point of d_S. Then $P_S(x)$ has an exposed face of dimension at least 1. Moreover, if y is in the relative interior of an exposed face E of $P_S(x)$ satisfying $\dim E \geq 1$, then $\dim N_{P_S(x)}(y)$ is a lower bound for the Hausdorff dimension of the connected component of x in $\Sigma(d_S)$.*

Proof — Since x is a non-isolated singular point of d_S, Theorem 4.4.2 ensures that

$$\partial D^+d_S(x) \setminus D^*d_S(x) \neq \emptyset\,.$$

Now, simple arguments of convex analysis show that, if $p \in \partial D^+d_S(x) \setminus D^*d_S(x)$, then p belongs to some exposed face V of $D^+d_S(x)$, with $\dim V \geq 1$. Then, recalling (4.4.47) and (4.4.48) we conclude that $x - d_S(x)V$ is an exposed face of $P_S(x)$.

By similar arguments we have that, if y is in the relative interior of an exposed face E of $P_S(x)$ with $\dim E \geq 1$, then

$$p := \frac{x-y}{d_S(x)} \in \partial D^+d_S(x) \setminus D^*d_S(x)$$

and $N_{P_S(x)}(y) = -N_{D^+d_S(x)}(p)$. Then, to complete the proof, it suffices to apply Theorem 4.3.2 to the distance function. ■

The next result immediately follows from the previous one since, owing to (4.4.48), the dimension of $P_S(x)$ coincides with $\kappa(x)$, the magnitude of x.

Corollary 4.4.4 *Let S be a nonempty closed subset of $\mathbb{R}^n$ and $x \notin S$ be a singular point of d_S. Then $n-\kappa(x)$ is a lower bound for Hausdorff dimension of the connected component of x in $\Sigma(d_S)$.*

A typical situation that is covered by Corollary 4.4.4 is when $\mathrm{proj}_S(x) = \{y_0, \ldots, y_k\}$ for some $k \in \{1, \ldots, n-1\}$ provided that the vectors $y_1 - y_0, \ldots, y_k - y_0$ are linearly independent. In this case the connected component of x in $\Sigma(d_S)$ has dimension $\geq n-k$.

Bibliographical notes

Apart from earlier contributions for specific problems like in [42], to our knowledge the first general results about the rectifiability of the singular sets of concave functions are due to Zajíček [85, 86] and Veselý [80, 81]. Similar properties were later extended to semiconcave functions in [6]. Related and improved versions of these results can be found in [1, 5, 12].

The propagation of singularities was first studied in [27] for semiconcave solutions to Hamilton-Jacobi-Bellman equations, and then in [11] for general semiconvex functions. In the first reference, singularities were shown to propagate just along a sequence of points. In the second one, on the contrary, conditions were given to derive estimates for the Hausdorff dimension of the singular set in a neighborhood of a point $x_0 \in \Sigma(u)$. The importance of a condition of type (4.2.11) was initially pointed out in [11].

The results of Sections 4.2, 4.3 and 4.4 are taken from [2]. An earlier version of the technique used in the proof of Theorem 4.2.2 was given in [1].

From the point of view of best approximation theory, the singularities of the distance function d_S have been investigated by many authors. In [20], the set $\Sigma(d_S)$ is described proving that, in the case of $n = 2$, the non-isolated singularities of the metric projection propagate along Lipschitz arcs. Such a result is extended to Hilbert spaces in [83]. Property (c) in Theorem 4.4.2 was observed by Motzkin [69] in the case of $n = 2$; it was later extended to Hilbert spaces in [83]. The propagation of the non-isolated singularities of d_S along Lipschitz arcs has been established in both Euclidean and Hilbert spaces, see [20] and [82, 83] respectively.

Chapter 5

Hamilton–Jacobi Equations

This chapter is devoted to the study of Hamilton–Jacobi equations. Our interest in these equations lies mainly in the connection with calculus of variations and optimal control. We have already analyzed in Chapter 1 a special case of such problems and a more general treatement will be givem in the following chapters. However, our point of view in this chapter will be to study Hamilton–Jacobi equations for their intrinsic interest, without referring to specific applications.

We begin by giving, in Section 5.1, a fairly general exposition of the method of characteristics. This method allows to construct smooth solutions of first order equations, and in general can be applied only locally. However, this method is interesting also for the study of solutions which are not smooth. As we will see in the following, characteristic curves (or generalized versions of them) play often an important role for generalized solutions and are related to the optimal trajectories of the associated control problem.

In Section 5.2 we recall the basic definitions and results from the theory of viscosity solutions for Hamilton–Jacobi equations. In this theory one gives a suitable definition of generalized solution which enables to obtain existence and uniqueness theorems under quite general hypotheses. In addition, in most of the cases where the equation is associated to a control problem, the viscosity solution coincides with the value function of the problem. Although this section is meant to be a collection of results whose proof can be found in the specialized monographs, we have included the proofs of some simple statements in order to give to the reader a flavor of the techniques of the theory.

In Section 5.3 we analyze the relation between semiconcavity and the viscosity property. Roughly speaking, it turns out that the two properties are equivalent when the hamiltonian is a convex function of the gradient of the solution. We also present a semiconcavity result in the stationary case due to Ishii, that requires milder assumptions on the hamiltonian.

In the remaining sections we analyze the singular set of semiconcave functions that are solutions of a Hamilton–Jacobi equation. It turns out that for such functions the general results of the previous chapter can be improved; in particular, one can give sufficient conditions for the propagation of singularities which are easier to verify than for a general semiconcave function. In addition, the singular arcs along which a singularity propagates satisfy a differential inclusion which is a generalization to the nonsmooth case of the equation satisfied by the characteristic curves.

5.1 The method of characteristics

In Section 1.4 we have introduced the method of characteristics to construct a local classical solution of the Cauchy problem for equations of the form $\partial_t u + H(\nabla u) = 0$. We now show how this method can be extended to treat general first order equations.

Before treating more general cases, let us show how the procedure of Section 1.4 can be easily extended to Cauchy problems where the hamiltonian depends also on t, x. Let us consider the problem

$$\partial_t u(t,x) + H(t,x,\nabla u(t,x)) = 0, \qquad (t,x) \in [0,\infty[\, \times \mathbb{R}^n \tag{5.1.1}$$

$$u(0,x) = u_0(x), \qquad x \in \mathbb{R}^n, \tag{5.1.2}$$

with H and u_0 of class C^2.

Let us first suppose to have a solution $u \in C^2([0,T] \times \mathbb{R}^n)$ of the above problem. Given $z \in \mathbb{R}^n$, we call *characteristic curve* associated to u starting from z the curve $X(\cdot; z)$ which solves

$$\dot{X} = H_p(t, X, \nabla u(t,X)), \qquad X(0) = z.$$

Then, if we set

$$U(t;z) = u(t, X(t;z)), \qquad P(t;z) = \nabla u(t, X(t;z)).$$

we find that

$$\dot{U} = u_t(t,X) + \nabla u(t,X) \cdot \dot{X} = -H(t,X,P) + P \cdot H_p(t,X,P),$$

$$\begin{aligned} \dot{P} &= \nabla u_t(t,X) + \nabla^2 u(t,X) H_p(t,X,\nabla u(t,X)) \\ &= \nabla[u_t(t,X) + H(t,X,\nabla u(t,X))] - H_x(t,X,\nabla u(t,X)) \\ &= -H_x(t,X,P). \end{aligned}$$

Therefore, the pair X, P solves the ordinary differential problem

$$\begin{cases} \dot{X} = H_p(t, X, P) \\ \dot{P} = -H_x(t, X, P) \end{cases} \qquad \begin{cases} X(0) = z \\ P(0) = \nabla u_0(z), \end{cases} \tag{5.1.3}$$

while U satisfies

$$\dot{U} = -H(t, X, P) + P \cdot H_p(t, X, P), \qquad U(0; z) = u_0(z). \tag{5.1.4}$$

This shows that X, P and U are uniquely determined by the initial value u_0. As in Section 1.4, the above arguments suggest that one can obtain a solution by solving (5.1.3), as long as the map $z \to X(t; z)$ is invertible. We have the following result, which can be proved by arguments similar to those of Theorem 1.4.2.

Theorem 5.1.1 *Given $z \in \mathbb{R}^n$, let $X(t; z)$, $P(t; z)$ denote the solution of problem* (5.1.3) *and let $U(t; z)$ be defined by* (5.1.4). *Suppose that there exists $T^* > 0$ such that*

(i) *the maximal interval of existence of the solution to* (5.1.3) *contains $[0, T^*[$ for all $z \in \mathbb{R}^n$;*

(ii) *the map $z \to X(t; z)$ is invertible with C^1 inverse $x \to Z(t; x)$ for all $t \in [0, T^*[$.*

Then there exists a unique solution $u \in C^2([0, T^*[\times \mathbb{R}^n)$ of problem (5.1.1), *which is given by*

$$u(t, x) = U(t; Z(t; x)), \qquad (t, x) \in [0, T^*[\times \mathbb{R}^n.$$

Let us now consider a more general situation. Let $\Omega \subset \mathbb{R}^n$ be an open set and let $\Gamma \subset \Omega$ be a $(n-1)$–dimensional surface of class C^2. We consider the problem

$$\begin{cases} H(x, u(x), Du(x)) = 0, & x \in \Omega \\ u(x) = g(x) & x \in \Gamma, \end{cases} \tag{5.1.5}$$

where $H \in C^2(\Omega \times \mathbb{R} \times \mathbb{R}^n)$ and $g \in C^2(\Omega)$. Our arguments can be easily adapted to the case of a Dirichlet problem, corresponding to $\Gamma = \partial\Omega$, under the assumptions that H, g are C^2 in $\overline{\Omega}$.

Given a point $z \in \Gamma$, let us denote by $\nu(z)$ the unit normal to Γ at z. If $u \in C^1(\Omega)$ satisfies $u = g$ on Γ, then it is known from elementary calculus that $Du(z) = Dg(z) + \lambda\nu(z)$ for some $\lambda = \lambda(z) \in \mathbb{R}$. If in addition u solves equation (5.1.5), then λ must belong to the set

$$\Lambda(z) = \{\lambda \in \mathbb{R} \ : \ H(z, g(z), Dg(z) + \lambda\nu(z)) = 0\}. \tag{5.1.6}$$

To begin with, let us first assume that we have a solution $u \in C^2(\Omega)$ of the problem (5.1.5). Given $z \in \Gamma$, we define $X(\cdot)$ to be the solution of

$$\dot{X} = H_p(X, u(X), Du(X)), \qquad X(0) = z \tag{5.1.7}$$

and we set

$$U(s) = u(X(s)), \qquad P(s) = \nabla u(X(s)). \tag{5.1.8}$$

Let us check whether the triple (X, U, P), like in the evolutionary case, satisfies an ordinary differential system with initial conditions determined by g. By a straightforward computation we obtain

Lemma 5.1.2 *Let $u \in C^2(\Omega)$ be a solution of (5.1.5). Given $z \in \Gamma$, let X, U, P defined as in (5.1.7), (5.1.8). Then there exists $\lambda \in \Lambda(z)$ such that X, U, P satisfy*

$$\begin{cases} \dot{X} = H_p(X, U, P) \\ \dot{P} = -H_x(X, U, P) - H_u(X, U, P)P \\ \dot{U} = P \cdot H_p(X, U, P) \end{cases} \tag{5.1.9}$$

with initial conditions

$$X(0) = z, \quad U(0) = g(z), \quad P(0) = Dg(z) + \lambda\nu(z). \tag{5.1.10}$$

We now no longer assume a priori the existence of a solution u, and try to use system (5.1.9), called *characteristic system*, to construct one. Before going further let us introduce the following definition. Given $z_0 \in \Gamma$, we say that the vector $p_0 \in \mathbb{R}^n$ is *characteristic* for problem (5.1.5) at z_0 if

$$H_p(z_0, g(z_0), p) \cdot \nu(z_0) = 0.$$

Let us denote by $\Lambda^*(z_0)$ the set of all $\lambda \in \Lambda(z_0)$ such that $Dg(z_0) + \lambda\nu(z_0)$ is not characteristic at z_0. It is easy to see that $\Lambda^*(z_0)$ is a discrete set.

Recall that H_p is the speed of the characteristic curves X introduced above. Therefore, a vector p_0 is characteristic if and only if the curve X obtained solving (5.1.9) with initial conditions $X(0) = z_0$, $P(0) = p_0$, $U(0) = g(z_0)$ is tangent to Γ at z_0. This is a behavior we want to avoid, because we want the characteristic curves to cover univalently a neighborhood of Γ. Therefore, the choice of λ in the initial conditions (5.1.10) will be restricted to the elements of $\Lambda^*(z_0)$.

Lemma 5.1.3 *Let $z_0 \in \Gamma$ and let $\lambda_0 \in \Lambda^*(z_0)$. Then there exists $\Gamma' \subset \Gamma$ neighborhood of z_0 (in the relative topology of Γ) and a unique map $\mu : \Gamma' \to \mathbb{R}$ such that*

$$\mu(z_0) = \lambda_0, \qquad \mu(z) \in \Lambda^*(z) \quad \forall\, z \in \Gamma'. \tag{5.1.11}$$

Moreover, if we take $\lambda = \mu(z)$ *in* (5.1.10) *and denote by* $X(t;z)$, $U(t;z)$, $P(t;z)$ *the corresponding solution of* (5.1.9)–(5.1.10)*, then the map* $(t,z) \to X(t;z)$ *is a* C^1*-diffeomorphism between a neighborhood of* $(0, z_0) \in \mathbb{R} \times \Gamma'$ *and a neighborhood of* $z_0 \in \mathbb{R}^n$*.*

Proof — Let us introduce a parametrization of Γ near z_0. More precisely, let $A \subset \mathbb{R}^{n-1}$ be an open set, let $\phi : A \to \mathbb{R}^n$ a map of class C^1 such that $\phi(A) \subset \Gamma$, $D\phi(y)$ has rank $n-1$ for all $y \in A$ and $\phi(y_0) = z_0$ for some $y_0 \in A$. Let us define, for $(y, \lambda) \in A \times \mathbb{R}$,

$$F(y, \lambda) = H(\ \phi(y),\ g(\phi(y)),\ Dg(\phi(y)) + \lambda\nu\,(\phi(y))\ ).$$

If $\lambda_0 \in \Lambda^*(z_0)$, then $F(y_0, \lambda_0) = 0$ and $F_\lambda(y_0, \lambda_0) \neq 0$. By the implicit function theorem, there is a function $\lambda = f(y)$ defined in a neighborhood of y_0 such that $F(y, f(y)) = 0$. If we set $\mu(z) = f(\phi^{-1}(z))$ for $z \in \Gamma$ close to z_0 we have that $\mu(z) \in \Lambda(z)$; if we take our neighborhoods sufficiently small we also have, by continuity, that $Dg(z) + \mu(z)\nu(z)$ is noncharacteristic.

The statement about X will follow from the inverse function theorem if we show that the jacobian of the function $\tilde{X}(t, y) := X(t, \phi(y))$ is nonsingular at $(0, y_0)$. Now, since $\tilde{X}(0, y) \equiv \phi(y)$, the partial derivatives of $\tilde{X}$ with respect to the y components at the point $(0, y_0)$ are linearly independent vectors generating the tangent space to Γ at z_0. On the other hand, the vector $\partial_t \tilde{X}(0, y_0) = H_p(z_0, g(z_0), Dg(z_0) + \lambda_0\nu(z_0))$ does not belong to this space, by the assumption of non characteristicity. Thus, the jacobian is nonsingular and the lemma is proved. ∎

Theorem 5.1.4 *Given* $z_0 \in \Gamma$ *and* $\lambda_0 \in \Lambda^*(z_0)$*, we can choose* $R > 0$ *such that there exists a unique* $u \in C^2$ *which solves problem* (5.1.5) *in* $B_R(z_0)$ *and satisfies* $Du(z_0) = Dg(z_0) + \lambda_0\nu(z_0)$*.*

Proof — We only give a sketch of the proof, referring to [42] for more details. As in the previous Lemma, we define $X(t;z)$, $U(t;z)$, $P(t;z)$ solving system (5.1.9)–(5.1.10) choosing $\lambda = \mu(z)$ in (5.1.10). By the lemma, $X(t;z)$ is locally invertible; therefore we can define in a unique way, for x near z_0, two functions $\tau(x)$ and $\zeta(x)$ such that $X(\tau(x); \zeta(x)) = x$. We then set $u(t, x) = U(\tau(x); \zeta(x))$; a long, yet straightforward, computation shows that u is of class C^2 and satisfies problem (5.1.5). The uniqueness of u follows from Lemma 5.1.2 and from the uniqueness of the function $\mu(\cdot)$ given by Lemma 5.1.3. ∎

It is clear from the above statement that neither local existence nor local uniqueness of a classical solution to problem (5.1.5) are ensured. In fact, if $\Lambda(z_0)$ is empty, then no smooth solution of (5.1.5) exists in a neighborhood of z_0. On the other hand, if $\Lambda^*(z_0)$ contains more than one element λ_0, we find different classical solutions corresponding to each choice of λ_0. Finally, nothing can be said in general about the existence and uniqueness of solutions corresponding to values of $\lambda \in \Lambda(z_0) \setminus \Lambda^*(z_0)$.

Applying the previous theorem along the whole surface Γ we easily obtain the following result.

Corollary 5.1.5 *Let $\mu \in C(\Gamma)$ be such that $\mu(z) \in \Lambda^*(z)$ for all $z \in \Gamma$. Then there exists a neighborhood $\mathcal{N}$ of Γ and a unique $u \in C^2(\mathcal{N})$ which solves problem* (5.1.5) *and satisfies $Du(z) = Dg(z) + \mu(z)\nu(z)$ for all $z \in \Gamma$.*

As in the evolutionary case, the applicability of the method of characteristics is guaranteed only in a neighborhood of Γ and we cannot expect in general the existence of a smooth solution in the whole set Ω. Observe that in this stationary case we have the additional hypothesis that $\Lambda^*(z)$ be non-empty; this can be interpreted as a compatibility condition on the problem data.

Example 5.1.6 Let us consider the *eikonal equation*

$$|Du(x)|^2 - 1 = 0, \qquad x \in \mathbb{R}^n$$

with data

$$u(z) = 0 \text{ for all } z \in \Gamma := \{z \in \mathbb{R}^n \ : \ |z| = 1\}.$$

Then

$$\Lambda(z) = \Lambda^*(z) = \{-1, +1\}$$

for all $z \in \Gamma$. The two corresponding solutions to the characteristic system are

$$X(t; z) = (1 \pm 2t)z, \qquad U(t; z) = 2t, \qquad P(t; z) = \pm z.$$

We can invert X as long as $1 \pm 2t > 0$ and find

$$z(x) = \frac{x}{|x|}, \qquad t(x) = \pm \frac{|x| - 1}{2}$$

Thus we obtain the two solutions

$$u(x) = U(t(x); z(x)) = 2t(x) = \pm(|x| - 1).$$

Observe that the two solutions are smooth everywhere except at the origin, which is the point where the characteristics lines intersect.

5.2 Viscosity solutions

Let $\Omega \subset \mathbb{R}^n$ be an open set and let $H \in C(\Omega \times \mathbb{R} \times \mathbb{R}^n)$. Let us again consider the general nonlinear first order equation

$$H(x, u, Du) = 0, \qquad x \in \Omega \subset \mathbb{R}^n, \tag{5.2.12}$$

in the unknown $u : \Omega \to \mathbb{R}$. As usual, evolution equations can be recast in this form by considering time as an additional space variable.

As we have already mentioned, equations of the above form do not possess in general global smooth solutions. A major problem is to give a notion of generalized solution such that global existence and uniqueness results can be obtained. We have seen in Chapter 1 that for certain problems a useful definition of solution is obtained requiring u to be semiconcave and to satisfy the equation almost everywhere. Here we present the notion of viscosity solution, which has a much wider range of applicability.

Definition 5.2.1 *A function $u \in C(\Omega)$ is called a* viscosity subsolution *of equation* (5.2.12) *if, for any $x \in \Omega$, it satisfies*

$$H(x, u(x), p) \leq 0, \qquad \forall\, p \in D^+u(x). \tag{5.2.13}$$

Similarly, we say that u is a viscosity supersolution *of equation* (5.2.12) *if, for any $x \in \Omega$, we have*

$$H(x, u(x), p) \geq 0, \qquad \forall\, p \in D^-u(x). \tag{5.2.14}$$

If u satisfies both the above properties, it is called a viscosity solution *of equation* (5.2.12).

Let us mention that the notion of viscosity solution can be generalized to cases where u, or H, or both, are discontinuous, or to equations of second order whose principal part is degenerate elliptic. We will not treat these topics since they are not needed for our present purposes.

Observe that, by virtue of Proposition 3.1.7, condition (5.2.13) (resp. (5.2.14) can be restated in an equivalent way by requiring

$$H(x, u(x), D\phi(x)) \leq 0 \qquad (\text{resp. } H(x, u(x), D\phi(x)) \leq 0\ ) \tag{5.2.15}$$

for any $\phi \in C^1(\Omega)$ such that $u - \phi$ has a local maximum (resp. minimum) at x.

Using the above notion of solution existence and uniqueness results for Hamilton–Jacobi have been obtained under very general hypotheses. It is beyond the scope of this book to give a detailed treatment of the theory of viscosity solutions, for which the reader may consult [67, 37, 18, 17, 35, 42] and the references therein. However, we summarize in this section some of the main results of this theory, since it is closely related to the topics treated in the rest of the book.

For the sake of the reader who is not familiar with viscosity solutions, we first prove some statements under special hypotheses, in order to present some of the basic techniques of the theory. Then we will give, without proof, some general existence and uniqueness results, which can be found in the references quoted above.

We first present the so–called *vanishing viscosity method* to prove existence of viscosity solutions as limit of solutions to a suitable approximate second–order problem.

Theorem 5.2.2 *Let* $\{\varepsilon_k\} \subset]0, +\infty[$ *be a sequence converging to* 0 *and let* $u_k \in C^2(\Omega)$ *be a solution of*

$$H(x, u_k(x), Du_k(x)) = \varepsilon_k \Delta u_k(x), \qquad x \in \Omega. \tag{5.2.16}$$

Suppose in addition that the sequence u_k *converges to some function* u *locally uniformly in* Ω. *Then* u *is a viscosity solution of equation* (5.2.12).

Proof — It is convenient to use the equivalent formulation (5.2.15) of the definition of viscosity solution. Let $\phi \in C^1(\Omega)$ be such that $u - \phi$ has a maximum at some point $x_0 \in \Omega$. We have to show that

$$H(x_0, u(x_0), D\phi(x_0)) \leq 0. \tag{5.2.17}$$

We first assume in addition that ϕ is of class C^2 and that x_0 is a strict local maximum point for $u - \phi$. Then, by the local uniform convergence of u_k to u, it is easily seen that there exists a sequence $\{x_k\}$ of points converging to x_0 such that $u_k - \phi$ has a local maximum at x_k. Since $u_k \in C^2$, we deduce

$$Du_k(x_k) = D\phi_k(x_k), \qquad \Delta u_k(x_k) \leq \Delta\phi(x_k).$$

Using the fact that u_k solves (5.2.16) we obtain

$$H(x_k, u(x_k), D\phi(x_k)) \leq \varepsilon_k \Delta\phi(x_k),$$

and so inequality (5.2.17) follows as $k \to \infty$.

In the general case, if ϕ is only of class C^1, we pick a sequence $\phi_k \in C^2(\Omega)$ such that $||D^2\phi_k||_\infty$ is uniformly bounded and $\phi_k \to \phi, D\phi_k \to D\phi$ locally uniformly in Ω. Then we apply the above argument to the sequence $u_k - \phi_k$. If the maximum at x_0 is not strict, we may replace ϕ by $\tilde{\phi}(x) = \phi + |x - x_0|^2$; then $u - \tilde{\phi}$ has a strict maximum at x_0 and so the previous argument can be applied. This proves (5.2.17) since ϕ and $\tilde{\phi}$ have the same gradient at x_0. The proof that u is a supersolution is entirely analogous. ∎

Remark 5.2.3 In the previous result one assumes the existence of a solution for the approximate equation (5.2.16). Such an equation is semilinear elliptic, and there are standard techniques (see [67]) which ensure under general hypotheses the existence of a solution as well as suitable bounds yielding the uniform convergence of some sequence $\{u_k\}$. The above result shows how powerful the notion of viscosity solution is, since the passage to the limit can be done under the only assumption that $\{u_k\}$ converges and without requiring explicitly any convergence of the derivatives. The vanishing viscosity method has played an important role at the beginning of the theory of viscosity solutions; let us mention that another common approach to prove existence results is a Perron–type technique introduced later by Ishii [56].

By an argument completely analogous to the proof of the previous theorem we can obtain the following stability result for viscosity solutions.

Theorem 5.2.4 *Let $u_k \in C(\Omega)$ be viscosity subsolutions (resp. supersolutions) of*

$$H_k(x, u_k, Du_k) = 0$$

where $H_k \in C(\Omega \times \mathbb{R} \times \mathbb{R}^n)$. Suppose that $u_k \to u$ and $H_k \to H$ locally uniformly. Then u is a viscosity subsolution (resp. supersolution) of

$$H(x, u, Du) = 0.$$

We now give a uniqueness statement for viscosity solutions of a problem on the whole space.

Theorem 5.2.5 *Let $u_1, u_2 \in C(\mathbb{R}^n)$ be viscosity solutions of*

$$u(x) + H(Du(x)) - n(x) = 0, \qquad x \in \mathbb{R}^n,$$

where H is continuous and n is uniformly continuous. Suppose in addition that $u_i(x) \to 0$ as $|x| \to +\infty$ for $i = 1, 2$. Then $u_1 \equiv u_2$.

Proof — We argue by contradiction and suppose that the difference $u_2 - u_1$ is not identically zero. Since $u_2(x) - u_1(x) \to 0$ as $|x| \to \infty$, it has either a positive maximum or a negative minimum. Let us suppose for instance that it attains a positive maximum m at some point $\bar{x}$. Let us also set $M = \sup |u_1| + \sup |u_2|$.

For given $\varepsilon > 0$, let us consider the function of $2n$ variables

$$\Phi_\varepsilon(x, y) = u_2(x) - u_1(y) - \frac{|x - y|^2}{2\varepsilon}, \qquad x, y \in \mathbb{R}^n.$$

It is easy to check that the assumption that $u_1(x) \to 0$ and $u_2(x) \to 0$ as $|x| \to \infty$ implies that $\limsup_{||(x,y)|| \to \infty} \Phi_\varepsilon(x,y) = 0$. Then, since $\Phi_\varepsilon(\bar{x}, \bar{x}) = m > 0$, we deduce that Φ_ε attains a maximum at some point $(x_\varepsilon, y_\varepsilon)$ and that $\Phi_\varepsilon(x_\varepsilon, y_\varepsilon) \geq m$. This implies

$$|x_\varepsilon - y_\varepsilon|^2 = 2\varepsilon(u_2(x_\varepsilon) - u_1(y_\varepsilon) - \Phi_\varepsilon(x_\varepsilon, y_\varepsilon)) \leq 2\varepsilon(M - m). \tag{5.2.18}$$

The property that $(x_\varepsilon, y_\varepsilon)$ is a maximum point for Φ_ε implies in particular that x_ε is a maximum point for the function $x \to u_2(x) - (|x - y_\varepsilon|^2)/(2\varepsilon)$. Therefore, the superdifferential of this function at the point $x = x_\varepsilon$ contains 0, and this is equivalent to $(x_\varepsilon - y_\varepsilon)/\varepsilon \in D^+u_2(x_\varepsilon)$. Using the definition of viscosity solution, we deduce that

$$u_2(x_\varepsilon) + H\left(\frac{x_\varepsilon - y_\varepsilon}{\varepsilon}\right) \leq n(x_\varepsilon).$$

Analogously, from the property that $y \to -u_1(y) - (|x_\varepsilon - y|^2)/(2\varepsilon)$ attains its maximum at y_ε we deduce that $(x_\varepsilon - y_\varepsilon)/\varepsilon \in D^-u_1(y_\varepsilon)$ and therefore

$$u_1(y_\varepsilon) + H\left(\frac{x_\varepsilon - y_\varepsilon}{\varepsilon}\right) \geq n(y_\varepsilon).$$

Hence

$$u_2(x_\varepsilon) - u_1(y_\varepsilon) \leq n(x_\varepsilon) - n(y_\varepsilon) \leq \omega(|y_\varepsilon - x_\varepsilon|).$$

where ω is the modulus of continuity of n. Taking into account (5.2.18) we obtain that

$$\limsup_{\varepsilon \to 0} u_2(x_\varepsilon) - u_1(y_\varepsilon) \leq 0.$$

On the other hand, we have that $u_2(x_\varepsilon) - u_1(y_\varepsilon) \geq \Phi_\varepsilon(x_\varepsilon, y_\varepsilon) \geq m > 0$ for every $\varepsilon > 0$. The contradiction shows that $u_1 \equiv u_2$. ∎

The above uniqueness result holds under less restrictive assumptions on the behavior of u_1, u_2 at infinity; the hypothesis that u_1, u_2 tend to 0 allows us to avoid some technicalities in the proof.

We now give some more general statements without proof. We begin by giving a uniqueness result for the stationary case, first in the case of a bounded domain, then in the whole space.

Theorem 5.2.6 *Let $\Omega \subset \mathbb{R}^n$ be open and bounded, and let $H \in C(\Omega \times \mathbb{R}^n)$ satisfy*

(H1) *There exists a continuous nondecreasing function $m : [0, \infty[\to [0, \infty[$ such that $m(0) = 0$ and that*

$$|H(x,p) - H(y,p)| \leq m(|x - y|(1 + |p|)), \tag{5.2.19}$$

for all $x, y \in \Omega, p \in \mathbb{R}^n$.

Suppose that $u_1, u_2 \in C(\overline{\Omega})$ are a viscosity subsolution and supersolution, respectively, of

$$u(x) + H(x, Du(x)) = 0, \qquad x \in \Omega. \tag{5.2.20}$$

If $u_1 \le u_2$ on $\partial\Omega$, then $u_1 \le u_2$ in Ω.

Proof – See e.g. [17, Theorem II.3.1] or [18, Theorem 2.4]. ∎

Theorem 5.2.7 *Let $H \in C(\mathbb{R}^n \times \mathbb{R}^n)$ satisfy assumption* (H1) *of the previous theorem with $\Omega = \mathbb{R}^n$. Suppose in addition that, for any $R > 0$, H is uniformly continuous in $\mathbb{R}^n \times B_R$. Let $u_1, u_2 \in C(\mathbb{R}^n)$ be a bounded viscosity subsolution and supersolution respectively of*

$$u(x) + H(x, Du(x)) = 0, \qquad x \in \mathbb{R}^n. \tag{5.2.21}$$

Then $u_1 \le u_2$ in $\mathbb{R}^n$.

Proof – See [18, Theorem 2.11]. ∎

The previous two theorems can be extended to hamiltonians of the form $H(x, u, Du)$ provided H is strictly increasing in the u argument. If this monotonicity assumption is violated, then the solution to the Dirichlet problem may no longer be unique, as shown by the following example (taken from [57]).

Example 5.2.8 The Dirichlet problem

$$\begin{cases} u'(x)^2 = x^2, & x \in]-1, 1[\\ u(-1) = u(1) = 0 \end{cases}$$

has infinitely many viscosity solutions, given by

$$u_c(x) = \frac{1}{2} \min\{(1 - x^2), (c + x^2)\}$$

for $c \in [-1, 1]$.

However, the strict monotonicity in u can be replaced by other assumptions, as shown by the next result.

Theorem 5.2.9 *Let $H \in C(\Omega \times \mathbb{R}^n)$ satisfy hypothesis* (H1) *in Theorem 5.2.6. Suppose in addition that*

(H2) *$H(x, \cdot)$ is convex for any $x \in \Omega$;*

(H3) *there exists* $\phi \in C^1(\Omega) \cap C(\overline{\Omega})$ *such that* $\sup_{x\in\Omega} H(x, D\phi(x)) < 0$.

Then the conclusion of Theorem 5.2.6 holds for $H(x, Du(x)) = 0$.

Proof — See [18, Theorem 2.7]. ■

We now turn to the evolutionary case and give a comparison result for the Cauchy problem.

Theorem 5.2.10 *Let* $H : [0,T] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *be given. Suppose that* H *is uniformly continuous in* $[0,T] \times \mathbb{R}^n \times B_R$ *for any* $R > 0$ *and that it satisfies*

$$|H(t,x,p) - H(t,y,p)| \leq m(|x-y|(1+|p|)), \ \forall\, x,y \in \mathbb{R}^n, p \in \mathbb{R}^n, t \in [0,T]$$

for some continuous nondecreasing function $m : [0,+\infty[\to [0,+\infty[$ *such that* $m(0) = 0$.

Suppose that u_1 *and* u_2 *are uniformly continuous in* $\mathbb{R}^n \times [0,T]$ *and are a viscosity subsolution and supersolution, respectively, of the equation*

$$u_t + H(t,x,\nabla u(t,x)) = 0, \qquad (t,x) \in]0,T[\, \times \mathbb{R}^n. \tag{5.2.22}$$

Then

$$\sup_{\mathbb{R}^n \times [0,T]} (u_1 - u_2) \leq \sup_{\mathbb{R}^n} (u_1(0) - u_2(0)).$$

Proof — See [17, Theorem II.3.7]. ■

By the above result, the viscosity solution of the Cauchy problem for equation (5.2.22) is unique and depends continuously on the initial data in the L^∞ norm. Observe that the strict monotonicity in u is no longer necessary for the comparison result as it was in the stationary case. This is a common feature of evolutionary problems.

Remark 5.2.11 The previous result can be generalized to hamiltonians depending also on u. However, the dependence on u cannot be completely arbitrary and some monotonicity assumption is needed. In particular, the case of hyperbolic conservation laws is not covered by the theory of viscosity solutions and has to be treated by other techniques. This is not surprising if one thinks of the connection described in Section 1.6, which suggests that solutions to conservation laws are less regular than the ones to Hamilton–Jacobi equations (roughly speaking, they have "one derivative less").

The interest of viscosity solutions for optimal control problems lies in the possibility of characterizing the value function as the unique viscosity solution of the associated Hamilton–Jacobi equation. Therefore, uniqueness results are significant in this context. However, for the sake of completeness, let us also mention some results about existence. We give two statements for the stationary problem, in the case $\Omega = \mathbb{R}^n$ and $\Omega \neq \mathbb{R}^n$ respectively.

Theorem 5.2.12 *Let $H : \mathbb{R}^n \times \mathbb{R}^n$ be uniformly continuous on $\mathbb{R}^n \times B_R$ for every $R > 0$ and let hypothesis* (H1) *in Theorem 5.2.6 be satisfied. Suppose in addition that $H(\cdot, 0)$ is bounded and that*

$$\lim_{|p|\to\infty} \inf_{x\in\mathbb{R}^n} H(x,p) = +\infty.$$

Then there exists a viscosity solution of $u + H(x, Du) = 0$ in $\mathbb{R}^n$ which belongs to $W^{1,\infty}(\mathbb{R}^n)$.

Proof — See [18, Theorem 2.12].

Theorem 5.2.13 *Let Ω be a proper subset of $\mathbb{R}^n$ and let $\phi \in C(\partial\Omega)$. Let $H : \Omega \times \mathbb{R}^n$ be uniformly continuous on $\mathbb{R}^n \times B_R$ for every $R > 0$ and let hypothesis* (H1) *in Theorem 5.2.6 be satisfied. Consider equation*

$$u + H(x, Du) = 0 \qquad \text{in } \Omega \tag{5.2.23}$$

with the boundary conditions

$$u = \phi \qquad \text{in } \partial\Omega. \tag{5.2.24}$$

Suppose that there exist a subsolution u_1 and a supersolution u_2 of equation (5.2.23)*, both bounded and uniformly continuous and both satisfying the boundary condition* (5.2.24)*. Then there exists also a bounded uniformly continuous solution of* (5.2.23) *satisfying* (5.2.24)*.*

Proof — See [18, Theorem 2.14].

Remark 5.2.14 One of the assumptions in the previous theorem is the existence of a subsolution and a supersolution to the equation satisfying the boundary condition. This condition may be regarded as a compatibility condition of the boundary data. It is in fact clear, at least intuitively, that a Dirichlet problem may not be solvable for general boundary data. Consider, for instance, the eikonal equation

$$|Du(x)|^2 = 1, \qquad x \in \Omega,$$

for which the comparison principle holds, by virtue of Theorem 5.2.9. It can be proved that viscosity solutions to this equation are Lipschitz continuous with constant 1. Therefore boundary data which are not Lipschitz with constant 1 are not admissible. The existence of a sub– and supersolution to the equation implies that the boundary data are admissible. More general boundary data can be treated by the theory of discontinuous viscosity solutions (see e.g. [18], [17] and the references therein).

5.3 Semiconcavity and viscosity

We now analyze the relation between the notions of semiconcave solution and viscosity solution to Hamilton–Jacobi equations. It will turn out that the two notions are strictly related when the hamiltonian is a convex function of Du. We begin with the following result.

Proposition 5.3.1 *Let u be a semiconcave function satisfying equation* (5.2.12) *almost everywhere. If H is convex in the third argument, then u is a viscosity solution of the equation.*

Proof — As a first step, we show that u satisfies the equation at all points of differentiability (our assumption is a priori slightly weaker). Let u be differentiable at some $x_0 \in \Omega$. Then there exists a sequence of points x_k converging to x_0 such that

(i) u is differentiable at x_k;

(ii) $H(x_k, u(x_k), Du(x_k)) = 0$;

(iii) $Du(x_k)$ has a limit for $k \to \infty$.

From Proposition 3.3.4(a) we deduce that the limit of $Du(x_k)$ is $Du(x_0)$. By the continuity of H, the equation holds at x_0 as well.

Let us now take an arbitrary $x \in \Omega$ and check that (5.2.13) is satisfied. We first observe that

$$H(x, u(x), p) = 0, \qquad \forall\, p \in D^*u(x). \tag{5.3.25}$$

This follows directly from the definition of D^*u, the continuity of H and the property that the equation holds at the points of differentiability. Since $D^+u(x)$ is the convex hull of $D^*u(x)$ (see Theorem 3.3.6) and H is convex, inequality (5.2.13) follows.

Now let us check inequality (5.2.14). For a given $x \in \Omega$, suppose that $D^-u(x)$ contains some vector p. Then, by Proposition 3.1.5(c) and Proposition 3.3.4(c), u is differentiable at x and $Du(x) = p$. Thus, (5.2.14) holds as an equality by the first part of the proof. ∎

Remark 5.3.2 A more careful analysis shows that the convexity of H and the semiconcavity of u play independent roles in the viscosity property. In fact, in the previous proof the deduction that u is a supersolution uses only the semiconcavity and is valid also if H is not convex. On the other hand, it is possible to prove (see see [67, p. 96] or [17, Prop. II.5.1]) that, if H is convex, any Lipschitz continuous u (not necessarily semiconcave) satisfying the equation almost everywhere is a viscosity subsolution of the equation.

The previous proposition is in general false when the hamiltonian is not convex. Consider for instance equation

$$1 - u'(x)^2 = 0, \qquad x \in]-1, 1[$$

with boundary data $u(-1) = u(1) = 0$. Then the function $u_1(x) = 1 - |x|$ is semiconcave and satisfies the equation almost everywhere, but is not a viscosity solution. On the other hand, the semiconvex function $u_2(x) = |x| - 1$ is a viscosity solution. In general, when we consider equations with a concave hamiltonian, all results of this section hold with semiconcavity replaced by semiconvexity. However, when the hamiltonian is neither convex nor concave, the viscosity property has in general no relation with semiconcavity or semiconvexity.

Remark 5.3.3 As we have seen, the fact that a semiconcave solution is also a viscosity solution is a rather direct consequence of the definitions and of the properties of generalized differentials. The converse implication is also true in many cases, as we will see, but it cannot be proved in such a direct way. In other words, semiconcavity is stronger than the viscosity property; the fact that viscosity solutions are in many cases semiconcave is nontrivial, and should be regarded as an additional regularity information. As long as one is concerned with existence and uniqueness issues, the notion of viscosity solution is the most convenient one. On the other hand, if one is also interested in the structure of the singularities, the viscosity property yields little information, while semiconcavity allows to obtain the finer results described in Chapter 4.

Before considering the question of whether viscosity solutions are semiconcave, let us consider a different issue. Let us examine the case when the hamiltonian is convex and the viscosity solution is semiconvex rather than semiconcave. This behavior is somehow unusual, as one can understand from the other results of this section, but there are interesting cases where it can occur (see for instance [25]). In such cases one can derive some consequences on the regularity of the solution, as we show here.

Proposition 5.3.4 *Let $u : \Omega \to \mathrm{I\!R}$ be a semiconvex viscosity solution of equation* (5.2.12). *If H is convex with respect to the last argument, then u satisfies*

$$H(x, u(x), p) = 0, \qquad x \in \Omega, \ p \in D^- u(x). \tag{5.3.26}$$

If the convexity of H is strict, then $u \in C^1(\Omega)$.

Proof — Since u is a viscosity solution, it satisfies the equation in the classical sense at the points of differentiability. Then, like in the proof of Proposition 5.3.1, we find that (5.3.25) holds at any point $x \in \Omega$. We recall that, by Theorem 3.3.6, the superdifferential of a semiconvex function satisfies $D^- u = \mathrm{co}\, D^* u$. Thus, the convexity of H implies

$$H(x, u(x), p) \leq 0, \qquad x \in \Omega, \ p \in D^- u(x).$$

But the converse inequality holds as well, by the definition of viscosity solution, and so we have proved (5.3.26).

If H is strictly convex, then its zero level set does not contain non-trivial convex sets. Therefore, by (5.3.26), $D^- u(x)$ reduces to a point for every $x \in \Omega$. Since u is semiconvex, Proposition 3.3.4-(e) ensures that $u \in C^1(\Omega)$. ∎

Let us now consider the converse implication to the one considered before, i.e., showing that a viscosity solution is semiconcave. Two possible approaches to the proof of such an implication are:

(i) Consider the elliptic or parabolic approximation to the Hamilton–Jacobi equation, obtained by adding a term of the form $\varepsilon \Delta u$ with $\varepsilon > 0$. Use the maximum principle to derive a semiconcavity estimate on u independent of ε. Pass to the limit as $\varepsilon \to 0$ to obtain the same estimate for the viscosity solution of the Hamilton–Jacobi equation.

(ii) Employ techniques from control theory to obtain a representation formula for the viscosity solution as the infimum of a suitable family of functions. Use this formula to prove semiconcavity by arguments similar to those used in Section 1.5 for solutions given by Hopf's formula.

We record here, without proof, some semiconcavity results for viscosity solutions which can be obtained by the two approaches described above. In the following chapters we shall give, in a detailed way, semiconcavity results obtained by approach (ii). Let us recall that we denote by $SCL(\Omega)$ the class of functions which are semiconcave with a linear modulus on every compact subset of Ω, and that we call a function $H : \mathrm{I\!R}^n \to \mathrm{I\!R}$ uniformly convex if there exists $\alpha > 0$ such that $D^2 H \geq 2\alpha I$ in the sense of distributions or, equivalently, such that $H(p) - \alpha |p|^2$ is convex.

We first give semiconcavity statements for hamiltonians which do not depend on x, u.

Theorem 5.3.5 *Let $\Omega \subset \mathbb{R}^n$ be an open set.*

(i) *If $u \in C(\Omega)$ is a viscosity solution of $u + H(Du) - n = 0$, with H uniformly convex and $n \in \mathrm{SCL}\,(\Omega)$, then $u \in \mathrm{SCL}\,(\Omega)$.*

(ii) *If $u \in C(\,]0,T[\,\times\Omega)$ is a viscosity solution of $\partial_t u + H(t, \nabla u) - n = 0$ in $]0,T[\,\times\Omega$, with H locally Lipschitz continuous, $H(t,\cdot)$ uniformly convex and $n(t,\cdot) \in \mathrm{SCL}\,(\Omega)$ for every $t \in]0,T[$, then $u(t,\cdot) \in \mathrm{SCL}\,(\Omega)$ for every $t \in]0,T[$.*

The above results are proved in [67, Prop. 3.3, Cor. 9.2] by the vanishing viscosity method (see also [62]). Observe that part (ii) of the statement, in the case when $H = H(\nabla u)$ and $n \equiv 0$, follows from Corollary 1.5.7.

These results show that the convexity of the hamiltonian induces a gain of regularity, since the solution is semiconcave, while the data for which an existence result holds are allowed to be only Lipschitz continuous. There are some semiconcavity results also for more general hamiltonians, where some regularity of H is required, in addition to convexity with respect to Du. The following result for the stationary case is given in [67, Theorem 3.3].

Theorem 5.3.6 *Let u be a viscosity solution of* (5.2.12) *and suppose that the hamiltonian $H(x,u,p)$ satisfies the following properties on any compact subset of $\Omega \times \mathbb{R} \times \mathbb{R}^n$:*

(i) *H is Lipschitz continuous and the distributional derivatives H_{xu}, H_{xp}, H_{up}, H_{pp} are bounded.*

(ii) *H is uniformly convex with respect to p.*

(iii) *H is semiconcave in x with a linear modulus independent of u, p.*

Then u is locally semiconcave with a linear modulus in Ω.

More recently, semiconcavity results have been obtained under much milder regularity assumptions on H. We remark, however, that these results are based on the generalized estimate of Section 2.5 and they yield no longer semiconcavity with a linear modulus. Let us first consider the stationary case.

Theorem 5.3.7 *Let $\Omega \subset \mathbb{R}^n$ be an open set and let $u \in \mathrm{Lip}_{loc}(\Omega)$ be a viscosity solution of $H(x,u,Du) = 0$ in Ω, where $H \in \mathrm{Lip}_{loc}(\Omega \times \mathbb{R} \times \mathbb{R}^n)$ is strictly convex in the last argument. Then u is locally semiconcave in Ω.*

Proof — Since our statement is of a local nature, it is not restrictive to consider the case when Ω is bounded and $u \in \mathrm{Lip}\,(\Omega)$. We can also assume that the Lipschitz continuity of H is uniform with respect to x and, by possibly redefining $H(x,u,p)$ for large p, that H has superlinear growth with respect to p. In addition, it is enough to consider the case when the equation satisfied by u is of the form

$$u + H(x, Du) = 0. \tag{5.3.27}$$

In fact, in the general case we can set $\tilde{H}(x,p) = -u(x) + H(x,u(x),p)$; then u is a viscosity solution of $u + \tilde{H}(x,Du) = 0$ with $\tilde{H}$ satisfying the same assumptions as H.

Now we exploit the property that the viscosity solution of an equation of the form (5.3.27) can be represented as the infimum of a variational problem. Let us first introduce the Legendre transform of H with respect to the second group of variables:

$$L(x,u) = \max_{p \in \mathbb{R}^n} [-u \cdot p - H(x,p)].$$

This transform satisfies special properties which are described in the Appendix (the reader should be warned that in the formulas in the Appendix the role of H and L is exchanged).

Let B be any ball contained in Ω. Given $\bar{x} \in B$, we consider the set of all absolutely continuous arcs starting from $\bar{x}$, which we denote by

$$\mathcal{A}(\bar{x}) = \{\xi \in W^{1,\infty}([0,\infty[\,,\mathbb{R}^n) : \xi(0) = \bar{x}\}. \tag{5.3.28}$$

Given $\xi \in \mathcal{A}(\bar{x})$, we denote by t_ξ the *exit time* of ξ from B, defined as

$$t_\xi = \min\{t > 0 \ : \ \xi(t) \in \partial B\}.$$

Let us introduce the functional

$$J(\bar{x},\xi) = \int_0^{t_\xi} e^{-t} L(\xi(t), \dot{\xi}(t))dt \, + \, e^{-t_\xi} u(\xi(t_\xi)), \qquad \forall \xi \in \mathcal{A}(\bar{x}). \tag{5.3.29}$$

Then it can be proved (see e.g. [67, Theorem 8.4]) that, for any $\bar{x} \in B$, we have

$$u(\bar{x}) = \inf_{\xi \in \mathcal{A}(\bar{x})} J(\bar{x},\xi). \tag{5.3.30}$$

In addition, in taking the above infimum it is enough to consider the arcs in $\mathcal{A}(\bar{x})$ satisfying the restriction $|\dot{\xi}(t)| \le C$, where C is a constant depending only on H and on the Lipschitz constant of u in B.

For a given $\varepsilon > 0$, let us set $\Omega_\varepsilon = \{x \in \Omega : \mathrm{dist}(x, \partial\Omega) > \varepsilon\}$. By Theorem A.1.6–(c)–(d) we can find $K > 0$ such that

$$|L(x,q) - L(y,q)| \le K|x-y|, \qquad \forall x, y \in \Omega_{\varepsilon/2}, \forall q \in B_{3C}. \tag{5.3.31}$$

In addition, $D_q L(x,q)$ exists and is continuous. Therefore, arguing as in Proposition 2.1.2, if we denote by $\omega(\cdot)$ the modulus of continuity of $D_q L(x,q)$ in $\Omega_{\varepsilon/2} \times B_{3C}$, we have

$$\begin{aligned} &\lambda L(x,q_1) + (1-\lambda)L(x,q_2) - L(x, \lambda q_1 + (1-\lambda)q_2) \qquad (5.3.32)\\ \le\ &\lambda(1-\lambda)|q_2 - q_1|\omega(|q_2 - q_1|), \qquad x \in \Omega_{\varepsilon/2},\ q_1, q_2 \in B_{3C},\ \lambda \in [0,1]. \end{aligned}$$

We now want to show that u satisfies the generalized one-sided estimate (2.5.15) in Ω_ε. Aiming at this, let us fix an arbitrary $x_0 \in \Omega_\varepsilon$ and $\delta \in\,]0, \varepsilon/8[$. We will use the representation formula (5.3.30) with $B = B_{4\delta}(x_0)$. Let us take $x, y \in B_1$, $\lambda \in [0,1]$. Let ξ be an arbitrary arc in $\mathcal{A}(x_0 + \delta(\lambda x + (1-\lambda)y))$ such that $|\dot\xi| \le C$. Then $\xi(t) \in B_{2\delta}(x_0)$ for $t \in [0, \delta C^{-1}]$ and $\xi(t) \in B$ for $t \in [0, 3\delta C^{-1}]$. This implies, in particular, that $t_\xi > 3\delta C^{-1}$. Therefore, if we set

$$\xi_1(t) = \begin{cases} \xi(t) + (1-\lambda)(\delta - Ct)(x-y) & \text{if } t \le \delta C^{-1} \\ \xi(t) & \text{if } t > \delta C^{-1}, \end{cases}$$

$$\xi_2(t) = \begin{cases} \xi(t) - \lambda(\delta - Ct)(x-y) & \text{if } t \le \delta C^{-1} \\ \xi(t) & \text{if } t > \delta C^{-1}, \end{cases}$$

we have that $\xi_1 \in \mathcal{A}(x_0 + \delta x)$, $\xi_2 \in \mathcal{A}(x_0 + \delta y)$ and $t_{\xi_1} = t_{\xi_2} = t_\xi$. Using (5.3.31) and (5.3.32) we deduce, for $t \in [0, \delta C^{-1}]$ a.e.,

$$\begin{aligned} &\lambda L(\xi_1(t), \dot\xi_1(t)) + (1-\lambda)L(\xi_2(t), \dot\xi_2(t)) - L(\xi(t), \dot\xi(t)) \\ &= \lambda[L(\xi_1(t), \dot\xi_1(t)) - L(\xi(t), \dot\xi_1(t)] \\ &\quad + (1-\lambda)[L(\xi_2(t), \dot\xi_2(t)) - L(\xi(t), \dot\xi_2(t)] \\ &\quad + \lambda L(\xi(t), \dot\xi_1(t)) + (1-\lambda)L(\xi(t), \dot\xi_2(t)) \\ &\quad - L(\xi(t), \lambda\dot\xi_1(t) + (1-\lambda)\dot\xi_2(t)) \\ &\le 2\lambda(1-\lambda)K|x-y|(\delta - Ct) + \lambda(1-\lambda)C|x-y|\omega(C|x-y|). \end{aligned}$$

Therefore,

$$\begin{aligned} &\lambda J(x_0 + \delta x, \xi_1) + (1-\lambda)J(x_0 + \delta y, \xi_2) - J(x_0 + \delta(\lambda x + (1-\lambda)y), \xi) \\ &= \int_0^{\delta C^{-1}} e^{-t}[\lambda L(\xi_1(t), \dot\xi_1(t)) + (1-\lambda)L(\xi_2(t), \dot\xi_2(t)) - L(\xi(t), \dot\xi(t)]dt \\ &\le \int_0^{\delta C^{-1}} [2\lambda(1-\lambda)K|x-y|(\delta - Ct) + \lambda(1-\lambda)C|x-y|\omega(C|x-y|)]dt \\ &= \lambda(1-\lambda)|x-y|\{K\delta^2 C^{-1} + \delta\omega(C|x-y|)\}. \end{aligned}$$

Since ξ was arbitrary we can use (5.3.30) to obtain the following estimate for the rescaled function $u_{x_0,\delta}$ (defined in (2.5.14)):

$$\begin{aligned}
&\lambda u_{x_0,\delta}(x) + (1-\lambda)u_{x_0,\delta}(y) - u_{x_0,\delta}(\lambda x + (1-\lambda)y) \\
&\quad = \delta^{-1}[\lambda u(x_0+\delta x) + (1-\lambda)u(x_0+\delta y) - u(x_0+\delta(\lambda x + (1-\lambda)y)] \\
&\quad \le \lambda(1-\lambda)|x-y|\{KC^{-1}\delta + \omega(C|x-y|)\}.
\end{aligned}$$

Then, u satisfies the generalized one–sided estimate (2.5.15) and is semiconcave, by Theorem 2.5.2. ∎

A similar result holds in the evolutionary case (see [28, Th. 3.2]).

Theorem 5.3.8 *Let $u \in \mathrm{Lip}_{loc}(]0,T[\times\Omega)$ be a viscosity solution of $\partial_t u + H(t,x,u,\nabla u) = 0$ in $]0,T[\times\Omega$, where $H \in \mathrm{Lip}_{loc}(]0,T[\times\Omega\times\mathbb{R}\times\mathbb{R}^n)$ is strictly convex in the last group of variables. Then u is locally semiconcave in $]0,T[\times\Omega$.*

There are some classes of equations, interesting for control problems, which are not covered by the above results, since $H = H(x,p)$ is convex in p, but it is neither strictly convex nor smooth. In some of these cases it is also possible to derive semiconcavity results; these are better formulated starting from the optimal control problem and will be considered in the next chapters.

In light of the previous considerations, the semiconcavity result we present to conclude this section is particularly interesting. In fact, it uses neither of the approaches (i) or (ii) outlined previously, but it is only based on the definition of viscosity solution. In addition, it allows for possibly nonconvex hamiltonians.

Theorem 5.3.9 *Let $u \in L^\infty(\mathbb{R}^n)\cap \mathrm{Lip}(\mathbb{R}^n)$ be a viscosity solution of*

$$u(x) + H(x, Du(x)) = 0, \qquad x \in \mathbb{R}^n. \tag{5.3.33}$$

Suppose that H satisfies

$$|H(x,p) - H(x,q)| \le \omega(|p-q|), \qquad \forall\, x,p,q \in \mathbb{R}^n \tag{5.3.34}$$

for some nondecreasing function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\omega(r) \to 0$ as $r \to 0$. Assume in addition that, for some $C > 0$ and $L' > 2\mathrm{Lip}(u)$,

$$H(x+h, p+Ch) - 2H(x,p) + H(x-h, p-Ch) \ge -C|h|^2 \tag{5.3.35}$$

holds for all $x, h \in \mathbb{R}^n$, $p \in B_{L'}$. Then u is semiconcave in $\mathbb{R}^n$ with linear modulus.

Proof — We consider the function $v : \mathbb{R}^{3n} \to \mathbb{R}$ defined by

$$v(x, y, z) = u(x) + u(z) - 2u(y).$$

Then v is a viscosity solution of

$$v(x, y, z) + \tilde{H}(x, y, z, D_x v, D_y v, D_z v) = 0, \qquad (x, y, z) \in \mathbb{R}^{3n}, \quad (5.3.36)$$

where

$$\tilde{H}(x, y, z, p, q, r) = H(x, p) + H(z, r) - 2H(y, -q/2).$$

Let us also define

$$\begin{aligned} \Phi(x, y, z) &= v(x, y, z) - \frac{1}{\varepsilon}|x + z - 2y|^2 - \delta(1 + |y|^2)^{1/2} \\ &\quad - \frac{C}{2}|x - y|^2 - \frac{C}{2}|z - y|^2, \end{aligned}$$

where ε, δ are positive parameters and C is as in (5.3.35). For simplicity we use in the following the notation $\langle y \rangle = (1 + |y|^2)^{1/2}$. Since u is assumed bounded, we find that Φ attains a maximum over $\mathbb{R}^{3n}$ at some point (x_0, y_0, z_0) depending on ε, δ. In addition we have, setting $p_0 = \frac{2}{\varepsilon}(x_0 + z_0 - 2y_0)$,

$$\begin{aligned} D_x(v - \Phi)(x_0, y_0, z_0) &= p_0 + C(x_0 - y_0) \\ D_y(v - \Phi)(x_0, y_0, z_0) &= -2p_0 + \delta\langle y_0 \rangle^{-1} y_0 - C(x_0 - y_0) - C(z_0 - y_0) \\ D_z(v - \Phi)(x_0, y_0, z_0) &= p_0 + C(z_0 - y_0). \end{aligned}$$

Using the fact the v is a subsolution of (5.3.36) we find

$$\begin{aligned} v(x_0, y_0, z_0) &\le -H(x_0, p_0 + C(x_0 - y_0)) - H(z_0, p_0 + C(z_0 - y_0)) \\ &\quad + 2H(y_0, p_0 - \frac{\delta}{2}\langle y_0 \rangle^{-1} y_0 + \frac{C}{2}(x_0 + z_0 - 2y_0)). \quad (5.3.37) \end{aligned}$$

By definition of (x_0, y_0, z_0) we have $\Phi(x_0, y_0, z_0) \ge \Phi(y + h, y, y - h)$ for any $y, h \in R^n$, which implies

$$u(y + h) + u(y - h) - 2u(y) \le C|h|^2 + \delta\langle y \rangle + M^*, \qquad (5.3.38)$$

where we have set

$$M^* = v(x_0, y_0, z_0) - \frac{C}{2}|x_0 - y_0|^2 - \frac{C}{2}|z_0 - y_0|^2 - \delta\langle y_0 \rangle.$$

Taking into account assumption (5.3.34), and observing that $|y_0| \le \langle y_0 \rangle$, we obtain from (5.3.37) that

$$\begin{aligned} M^* &\le -H(x_0, p_0 + C(x_0 - y_0)) - H(z_0, p_0 + C(z_0 - y_0)) \\ &\quad + 2H\left(y_0, p_0 + \frac{C}{2}(x_0 + z_0 - 2y_0)\right) + 2\omega\left(\frac{\delta}{2}\right) \qquad (5.3.39) \\ &\quad - \frac{C}{2}|x_0 - y_0|^2 - \frac{C}{2}|z_0 - y_0|^2 - \delta\langle y_0 \rangle. \end{aligned}$$

Our aim now is to show that, if ε, δ are small, then the point y_0 is close to the midpoint of x_0 and z_0. This will enable us to estimate the right–hand side of (5.3.39) by using (5.3.35). To this purpose, let us first observe that the inequality $\Phi(0,0,0) \le \Phi(x_0, y_0, z_0)$ yields

$$\delta\langle y_0\rangle + \frac{C}{2}|x_0 - y_0|^2 - L|x_0 - y_0| + \frac{C}{2}|z_0 - y_0|^2 - L|z_0 - y_0| \le \delta,$$

where we have denoted by L a Lipschitz constant of u. This inequality easily implies that there exists $R > 0$ independent of ε, δ such that

$$|x_0 - y_0| \le R, \quad |z_0 - x_0| \le R, \quad \delta\langle y_0\rangle \le R. \tag{5.3.40}$$

On the other hand, $\Phi(x_0, \frac{1}{2}(x_0 + z_0), z_0) \le \Phi(x_0, y_0, z_0)$ gives

$$\begin{aligned} \frac{1}{\varepsilon}|x_0 + z_0 - 2y_0|^2 \quad &\le \quad 2\left(u\left(\frac{x_0 + z_0}{2}\right) - u(y_0)\right) + \delta\langle\frac{x_0 + z_0}{2}\rangle - \delta\langle y_0\rangle \\ &\quad\quad L|x_0 + z_0 - 2y_0| + \frac{\delta}{2}|x_0 + z_0 - 2y_0|, \end{aligned} \tag{5.3.41}$$

since the function $t \to (1 + t^2)^{1/2}$ has Lipschitz constant 1. Therefore, if δ is enough small,

$$|p_0| \le 2L + \delta < L', \tag{5.3.42}$$

where L' is the constant appearing in assumption (5.3.35). Let us now keep δ fixed, take a sequence $\{\varepsilon_n\}$ converging to 0 and denote by x_n,y_n,z_n,p_n,M_n^* the quantities defined as x_0,y_0,z_0,p_0,M^* above with the choice $\varepsilon = \varepsilon_n$. The previous estimates (5.3.40) and (5.3.42) imply that, up to extracting a subsequence, we have

$$\begin{array}{ll} x_n - y_n \to h, & \quad y_n \to \bar{y} \\ y_n - z_n \to h', & \quad p_n \to \bar{p}, \end{array}$$

for some $h, h', \bar{y} \in \overline{B}_R$, $\bar{p} \in \overline{B}_{L'}$. In addition, from (5.3.41)

$$x_n + z_n - 2y_n \to 0,$$

and so $h = h'$. Thus, passing to the limit in (5.3.39) and recalling assumption (5.3.35),

$$\begin{aligned} \limsup_{n\to\infty} M_n^* \quad &\le \quad -H(\bar{y} + \bar{h}, \bar{p} + C\bar{h}) - H(\bar{y} - \bar{h}, \bar{p} - C\bar{h}) \\ &\quad\quad +2H(\bar{y}, \bar{p}) - C|h|^2 + 2\omega(\delta/2) - \delta\langle\bar{y}\rangle \\ &\le \quad 2\omega(\delta/2) - \delta\langle\bar{y}\rangle. \end{aligned}$$

Using this in (5.3.38), we obtain the assertion by letting $\delta \to 0$. ∎

5.4 Propagation of Singularities

We now turn to the analysis of the singular set of semiconcave solutions to Hamilton–Jacobi equations. In Chapter 3 we have obtained results which are valid for any semiconcave function; here we will prove stronger results for propagation of singularities, which exploit the additional information that the function solves a Hamilton–Jacobi equation.

As in the previous sections, we consider equations of the form

$$F(x, u(x), Du(x)) = 0 \qquad \text{a.e. in} \quad \Omega\,. \tag{5.4.43}$$

Throughout the rest of this chapter, $F : \overline{\Omega} \times \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}$ is a continuous function satisfying the following assumptions:

$(A1)$ $p \mapsto F(x, u, p)$ is convex;

$(A2)$ for any $(x, u) \in \Omega \times \mathbb{R}$ and any $p_0, p_1 \in \mathbb{R}^N$,

$$[p_0, p_1] \subset \{p \in \mathbb{R}^N \,:\, F(x, u, p) = 0\} \qquad \Longrightarrow \qquad p_0 = p_1\,.$$

Condition (A2) requires that the 0-level set $\{p \in \mathbb{R}^N \,:\, F(x, u, p) = 0\}$ contains no straight line. Clearly, such a property holds, in particular, if F is strictly convex with respect to p.

Remark 5.4.1 In Proposition 5.3.1 we have seen that, under the convexity assumption (A1), any semiconcave function u which solves equation (5.4.43) almost everywhere is also a viscosity solution of the equation. Thus, u satisfies for all $x \in \Omega$

$$F(x, u(x), p) = 0 \qquad \forall p \in D^* u(x), \tag{5.4.44}$$

$$F(x, u(x), p) \le 0 \qquad \forall p \in D^+ u(x). \tag{5.4.45}$$

The first result we prove is that condition (4.2.8) is necessary and sufficient for the propagation of a singularity at x_0, if u is a semiconcave solution of (5.4.43). Notice that, for a general semiconcave function, (4.2.8) is only a sufficient condition.

Theorem 5.4.2 *Let* (A1) *and* (A2) *be satisfied. Let* $u \in \mathrm{SCL}(\Omega)$ *be a solution of* (5.4.43) *and let* $x_0 \in \Sigma(u)$. *Then, the following properties are equivalent:*

(i) $\partial D^+ u(x_0) \setminus D^* u(x_0) \neq \emptyset$;

(ii) *there exist a sequence* $x_i \in \Sigma(u) \setminus \{x_0\}$, *converging to* x_0, *and a number* $\delta > 0$ *such that* $\operatorname{diam} D^+ u(x_i) \ge \delta$, *for any* $i \in \mathbb{N}$.

Proof — Since the implication (i)$\Rightarrow$(ii) is part of the conclusion of Theorem 4.2.2, we only need to prove that (ii)$\Rightarrow$(i). For this purpose, let us note that, in view of (5.4.44) and assumption (A2), the set $D^*u(x)$ contains no straight lines for any $x \in \Omega$. Now, let $x_i \in \Sigma(u) \setminus \{x_0\}$ be as in (ii). Without loss of generality, we can suppose that

$$\frac{x_i - x_0}{|x_i - x_0|} \to q \quad \text{as} \quad i \to \infty$$

for some unit vector q. Let us set

$$E := \left\{ p \in D^+u(x_0) \ : \ \langle p, q \rangle = \min_{\eta \in D^+u(x_0)} \langle \eta, q \rangle \right\} .$$

Then, it is immediate to see that $E \subset \partial D^+u(x_0)$.

Since $D^+u(x_i)$ is convex and diam $D^+u(x_i) \geq \delta$, we can find vectors $p_i^1, p_i^2 \in D^+u(x_i)$ such that, for any $i \in \mathbb{N}$,

$$[p_i^1, p_i^2] \subset D^+u(x_i) \qquad \text{and} \qquad |p_i^2 - p_i^1| = \delta \, .$$

Possibly extracting a subsequence, we can also suppose that

$$\lim_{i \to \infty} p_i^j = p^j \quad (j = 1, 2) \qquad \text{and} \qquad |p^2 - p^1| = \delta \, .$$

Moreover $p^1, p^2 \in D^+u(x_0)$, owing to the upper semicontinuity of D^+u.

Next, we claim that $[p^1, p^2] \subset E$. In fact, the monotonicity property of the superdifferential, (3.3.24), yields, for any $p \in D^+u(x_0)$,

$$\langle p_i^j - p, x_i - x_0 \rangle \leq 2C|x_i - x_0|^2 \qquad (i \in \mathbb{N} \, , \quad j = 1, 2)$$

for some $C > 0$. Dividing both sides of this inequality by $|x_i - x_0|$ and taking the limit as $i \to \infty$, we obtain that

$$\langle p^j - p, q \rangle \leq 0 \qquad (j = 1, 2) \quad \forall p \in D^+u(x_0) \, .$$

Now, from the above inequality we deduce that $\langle p^j, q \rangle \leq \langle p, q \rangle$ for $j = 1, 2$, and for any $p \in D^+u(x_0)$. So, $p^1, p^2 \in E$ and our claim follows.

Finally, since $[p^1, p^2] \subset E \ \subset \partial D^+u(x_0)$, and since $D^*u(x_0)$ contains no straight lines by assumption , we conclude that the set $D^*u(x_0)$ cannot cover $\partial D^+u(x_0)$. This completes our proof. ■

We can exploit further the information that u is the solution of a Hamilton–Jacobi equation by finding conditions which ensure that the crucial assumption (i) of the previous theorem is satisfied. For this purpose we prove a preliminary proposition. Let us set, for any $x \in \Omega$,

$$\begin{aligned} &\arg \min_{p \in D^+u(x)} F(x, u(x), p) := \\ &= \ \left\{ \bar{p} \in D^+u(x) \ : \ F(x, u(x), \bar{p}) = \min_{p \in D^+u(x)} F(x, u(x), p) \right\}. \end{aligned}$$

Proposition 5.4.3 *Assume* (A1) *and* (A2) *and let F be differentiable with respect to p. Let $u \in \mathrm{SCL}(\Omega)$ be a solution of equation* (5.4.43) *and let $x_0 \in \Omega$. Then,*

$$\arg \min_{p \in D^+u(x_0)} F(x_0, u(x_0), p) \neq \emptyset \tag{5.4.46}$$

and, for any $p_0 \in \arg\min_{p \in D^+u(x_0)} F(x_0, u(x_0), p)$, we have that

$$\langle F_p(x_0, u(x_0), p), p - p_0 \rangle \geq 0 \qquad \forall p \in D^+u(x_0)\,. \tag{5.4.47}$$

Moreover,

$$x_0 \in \Sigma(u) \qquad \Longleftrightarrow \qquad \min_{p \in D^+u(x_0)} F(x_0, u(x_0), p) < 0 \tag{5.4.48}$$

Furthermore, if x_0 is a singular point of u, then

$$\arg \min_{p \in D^+u(x_0)} F(x_0, u(x_0), p) \subset D^+u(x_0) \setminus D^*u(x_0) \tag{5.4.49}$$

$$F_p(x_0, u(x_0), p_0) \neq 0 \quad \Longrightarrow \quad p_0 \in \partial D^+u(x_0) \setminus D^*u(x_0) \tag{5.4.50}$$

where p_0 is any element of $\arg\min_{p \in D^+u(x_0)} F(x_0, u(x_0), p)$.

Proof — We note first that assertions (5.4.46) and (5.4.47) are a consequence of the compactness and of the convexity of the superdifferential.

Next, to prove (5.4.48), let us suppose that x_0 is singular. Then, $D^*u(x_0)$ contains at least two elements. Since $F(x_0, u(x_0), p) = 0$ for all $p \in D^*u(x_0)$, assumption (A2) implies that $\min_{p \in D^+u(x_0)} F(x_0, u(x_0), p) < 0$. The converse implication of (5.4.48) is trivial. Inclusion (5.4.49) can also be shown by similar arguments.

To prove (5.4.50), observe that $p_0 - tF_p(x_0, u(x_0), p_0) \notin D^+u(x_0)$ for every $t > 0$, in view of (5.4.47). So, $p_0 \in \partial D^+u(x_0)$ and the proof is complete. ■

The next result on propagation of singularities is a direct consequence of Theorem 4.2.2 and Proposition 5.4.3.

Theorem 5.4.4 *Assume* (A1) *and* (A2), *and let F be differentiable with respect to p. Let $x_0 \in \Omega$ be a singular point of a solution $u \in \mathrm{SCL}(\Omega)$ of equation* (5.4.43). *Suppose that*

$$F_p(x_0, u(x_0), p_0) \neq 0 \text{ for some } p_0 \in \arg \min_{p \in D^+u(x_0)} F(x_0, u(x_0), p). \tag{5.4.51}$$

Then there exist a Lipschitz arc $\mathbf{x} : [0, \rho] \to \mathrm{I\!R}^N$ for u, with $\mathbf{x}(0) = x_0$, and a positive number δ such that

$$\lim_{s \to 0^+} \frac{\mathbf{x}(s) - x_0}{s} \neq 0$$

$$\mathrm{diam}\Big(D^+u(\mathbf{x}(s))\Big) \geq \delta \qquad \forall s \in [0, \rho]\,.$$

In particular, $\mathbf{x}(s) \in \Sigma(u)$ *for any* $s \in]0,\rho]$*, and* $\mathbf{x}(s) \neq x_0$ *for* $s > 0$ *sufficiently small.*

Our next result shows that condition (5.4.51) is "almost" necessary for the propagation of singularities; in fact, each time property (i) of Theorem 5.4.2 holds, (5.4.51) is satisfied up to a possible change of the equation.

Theorem 5.4.5 *Assume* (A1) *and* (A2)*, and let* F *be differentiable with respect to* p*. Let* $x_0 \in \Omega$ *be a singular point of a solution* $u \in \mathrm{SCL}(\Omega)$ *of equation* (5.4.43)*. Fix*

$$p_0 \in \arg\min_{p \in D^+u(x_0)} F(x_0, u(x_0), p)\,.$$

If $p_0 \in \partial D^+u(x_0) \setminus D^*u(x_0)$*, then, for any open set* $\Omega' \subset\subset \Omega$*, there exists a function* $G : \Omega' \times \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}$ *such that*

(a) $G(x, u(x), Du(x)) = 0$ *a.e. in* Ω'

(b) G *is differentiable with respect to* p *and satisfies assumption* (A2);

(c) $G(x_0, u(x_0), p_0) = \min_{p \in D^+u(x_0)} G(x_0, u(x_0), p)\,;$

(d) $D_pG(x_0, u(x_0), p_0) \neq 0$.

Proof — Let $\Omega' \subset \Omega$ be a bounded open set and let L be a Lipschitz constant for u in Ω'. Suppose that

$$F_p(x_0, u(x_0), p_0) = 0\,,$$

otherwise one can take $G = F$. Let $q \neq 0$ be a normal vector to $D^+u(x_0)$ at p_0, i.e.

$$\langle p - p_0, q \rangle \leq 0 \qquad \forall p \in D^+u(x_0)\,. \tag{5.4.52}$$

Fix $c > 0$ such that

$$c + \langle q, p \rangle > 0 \qquad \forall |p| \leq L\,. \tag{5.4.53}$$

Notice that, in particular, (5.4.53) holds for every $p \in D^+u(x_0)$. We claim that

$$G(x, u, p) := (c + \langle q, p \rangle) F(x, u, p)$$

has all the required properties. In fact, properties (a) and (b) are obvious. To prove (c), it suffices to observe that (5.4.52), (5.4.53) and the fact that $F(x_0, u(x_0), p_0) < 0$ yield

$$\begin{aligned} G(x_0, u(x_0), p_0) &= (c + \langle q, p_0 \rangle) F(x_0, u(x_0), p_0) \\ &\leq (c + \langle q, p \rangle) F(x_0, u(x_0), p_0) \\ &\leq (c + \langle q, p \rangle) F(x_0, u(x_0), p) = G(x_0, u(x_0), p) \end{aligned}$$

for any $p \in D^+u(x_0)$. Finally, property (d) holds as

$$\begin{aligned} D_pG(x_0, u(x_0), p_0) &= qF(x_0, u(x_0), p_0) + (c + \langle q, p_0 \rangle) F_p(x_0, u(x_0), p_0) \\ &= qF(x_0, u(x_0), p_0) \neq 0 \,. \qquad \blacksquare \end{aligned}$$

5.5 Generalized Characteristics

In this section we will show that the singularities of a solution u to (5.4.43) propagate according to a suitable differential inclusion. For this purpose, let us introduce the following definition.

Definition 5.5.1 *We say that a Lipschitz arc,* $\mathbf{x} : [0, \rho] \to \mathbb{R}^N$, *is a generalized characteristic of equation* (5.4.43) *with initial point* x_0 *if* $\mathbf{x}(0) = x_0$ *and*

$$\mathbf{x}'(s) \in \mathrm{co} F_p(\mathbf{x}(s), u(\mathbf{x}(s)), D^+u(\mathbf{x}(s))) \quad a.e. \ in \quad [0, \rho] \,. \tag{5.5.54}$$

In order to analyze the propagation of singularities along generalized characteristics we need to require more regularity on F than in the previous sections. More precisely, we will assume that:

(A3) for any $x_1, x_2 \in \Omega$, $u_1, u_2 \in \mathbb{R}$, $p \in \mathbb{R}^N$

$$|F(x_2, u_2, p) - F(x_1, u_1, p)| \leq L_0(|x_2 - x_1| + |u_2 - u_1|),$$

for some constant $L_0 > 0$;

(A4) for any $x_1, x_2 \in \Omega$, $u_1, u_2 \in \mathbb{R}$, $p_1, p_2 \in \mathbb{R}^N$

$$|F_p(x_1, u_1, p_1) - F_p(x_2, u_2, p_2)| \leq L_1(|x_1 - x_2| + |u_1 - u_2| + |p_1 - p_2|)$$

for some constant $L_1 > 0$.

We will prove the following result.

Theorem 5.5.2 *Let* $u \in \mathrm{SCL}(\Omega)$ *be a solution of equation* (5.4.43). *Let us suppose that* (A1), (A2), (A3), (A4) *hold and let* $x_0 \in \Sigma(u)$ *be such that*

$$0 \notin \mathrm{co} F_p(x_0, u(x_0), D^+u(x_0)) \,. \tag{5.5.55}$$

Then, there exists a generalized characteristic of equation (5.4.43), $\mathbf{x} : [0, \rho] \to \Omega$, *with initial point* x_0, *such that* $\mathbf{x}(s) \in \Sigma(u)$ *and* $\mathbf{x}(\cdot)$ *is injective, for every* $s \in [0, \rho]$. *Moreover, for every* $s \in [0, \rho]$,

$$\min\{F(\mathbf{x}(s), u(\mathbf{x}(s)), p) \, : \, p \in D^+u(\mathbf{x}(s))\} \leq \frac{1}{2} F(x_0, u(x_0), p_0), \tag{5.5.56}$$

where $p_0 \in \arg\min_{p \in D^+u(x_0)} F(x_0, u(x_0), p)$.

Remark 5.5.3 Assumption (5.5.55) is a generalization of (5.4.51). In the following, we will show that in some cases (5.5.55) is necessary and sufficient for propagation of singularities along generalized characteristics. Let us also observe that, due to the upper continuity property of D^+u expressed by Proposition 3.3.4(a), assumption (5.5.55) implies that

$$0 \notin \operatorname{co} F_p(x, u(x), D^+u(x)) \qquad \forall x \in B_r(x_0)\,, \tag{5.5.57}$$

for some $r > 0$.

In order to prove Theorem 5.5.2 we need some technical preliminaries, the first of which is the following basic construction of singular arcs.

Lemma 5.5.4 *Let the assumptions of Theorem 5.5.2 be satisfied. Suppose also that* (A4) *holds with $L_1 = 1$. Let $R > 0$ be such that $\overline{B}_R(x_0) \subset \Omega$, and fix*

$$p_0 \in \arg \min_{p \in D^+u(x_0)} F(x_0, u(x_0), p)\,.$$

Let L be a Lipschitz constant for u in $\overline{B}_R(x_0)$ and let $\omega(r) = Cr$ be a modulus of semiconcavity for u in $\overline{B}_R(x_0)$ for some $C > 0$. Let us set

$$C_0 := \sup\{|F_p(x, u(x), p)| \,:\, x \in B_R(x_0) \quad |p| \le L\}\,; \tag{5.5.58}$$

$$M := 4L + 2C_0\,, \qquad C_1 := L_0(1 + L)M + 2CM^2\,; \tag{5.5.59}$$

$$\rho_0 := \min\left\{\frac{R}{4C_0}, \frac{|F(x_0, u(x_0), p_0)|}{2C_1}, \frac{1}{4C}\right\}\,. \tag{5.5.60}$$

Then, there exist two continuous arcs $\mathbf{x} : [0, \rho_0] \to B_R(x_0)$ and $\mathbf{p} : [0, \rho_0] \to \mathbb{R}^N$, with $\mathbf{x}(0) = x_0$ and $\mathbf{p}(0) = p_0$, such that

$$\mathbf{p}(s) \in D^+u(\mathbf{x}(s)) \tag{5.5.61}$$

$$\mathbf{x}(s) - x_0 = s[\mathbf{p}(s) - p_0 + F_p(x_0, u(x_0), p_0)] \tag{5.5.62}$$

$$F(\mathbf{x}(s), u(\mathbf{x}(s)), \mathbf{p}(s)) \le F(x_0, u(x_0), p_0) + C_1 s - \frac{1}{2}|\mathbf{p}(s) - p_0|^2 \tag{5.5.63}$$

$$F(\mathbf{x}(s), u(\mathbf{x}(s)), \mathbf{p}(s)) \le \frac{1}{2} F(x_0, u(x_0), p_0) \tag{5.5.64}$$

for any $s \in [0, \rho_0]$. Moreover,

$$|\mathbf{x}(t) - \mathbf{x}(s)| \le M|t - s| \qquad \forall t, s \in [0, \rho_0]. \tag{5.5.65}$$

Remark 5.5.5 Recalling property (5.4.44) it is easy to see that conditions (5.5.61) and (5.5.64) above ensure that $\mathbf{x}(s) \in \Sigma(u)$, for any $s \in [0, \rho_0]$. Moreover, we observe that $\mathbf{x}(\cdot)$ and $\mathbf{p}(\cdot)$ are related: given $\mathbf{x}(\cdot)$, arc $\mathbf{p}(\cdot)$ is uniquely defined by relation (5.5.62), and viceversa.

Proof — Applying Lemma 4.2.5 to the semiconcave function u, we can construct continuous arcs $\mathbf{x} : [0,\sigma] \to B_R(x_0)$ and $\mathbf{p} : [0,\sigma] \to \mathbb{R}^N$, with $\mathbf{x}(0) = x_0$ and $\mathbf{p}(0) = p_0$, such that, for any $s \in [0,\sigma]$,

$$\mathbf{p}(s) = p_0 - F_p(x_0, u(x_0), p_0) + \frac{\mathbf{x}(s) - x_0}{s} \in D^+u(\mathbf{x}(s)) \tag{5.5.66}$$

and

$$|\mathbf{x}(s) - x_0| \leq 4C_0 s$$

where C_0 is given by (5.5.58) and $4\rho := \min\{RC_0^{-1}, C^{-1}\}$. Moreover,

$$|\mathbf{x}(s) - \mathbf{x}(r)| \leq M|s-r| \qquad \forall s, r \in [0,\sigma] \tag{5.5.67}$$

where M is given by (5.5.59).

In order to complete the proof, we have to verify (5.5.63) and (5.5.64). Aiming at this, we observe that, for any $x \in B_R(x_0)$ and $p \in \mathbb{R}^N$,

$$\begin{aligned} F(x,u(x),p) &= F(x_0,u(x_0),p_0) + F(x,u(x),p) - F(x_0,u(x_0),p) \\ &\quad + F(x_0,u(x_0),p) - F(x_0,u(x_0),p_0) \\ &\leq F(x_0,u(x_0),p_0) + L_0(|x-x_0| + |u(x)-u(x_0)|) \\ &\quad + \int_0^1 \langle F_p(x_0,u(x_0),p_0+\lambda(p-p_0)), p-p_0\rangle \, d\lambda \end{aligned}$$

by assumption (A3). Moreover, by the Lipschitz continuity of u and assumption (A4) with $L_1 = 1$, we obtain

$$\begin{aligned} &F(x,u(x),p) \\ &\leq F(x_0,u(x_0),p_0) + L_0(1+L)|x-x_0| + \langle D_pF(x_0,u(x_0),p_0), p-p_0\rangle \\ &\quad + \int_0^1 \langle F_p(x_0,u(x_0),p_0+\lambda(p-p_0)) - D_pF(x_0,u(x_0),p_0), p-p_0\rangle \, d\lambda \\ &\leq F(x_0,u(x_0),p_0) + L_0(1+L)|x-x_0| + \langle F_p(x_0,u(x_0),p_0), p-p_0\rangle \\ &\quad + \frac{1}{2}|p-p_0|^2. \end{aligned}$$

Hence, taking in the above formula $(x,p) = (\mathbf{x}(s), \mathbf{p}(s))$,

$$\begin{aligned} F(\mathbf{x}(s), u(\mathbf{x}(s)), \mathbf{p}(s)) &\leq F(x_0,u(x_0),p_0) + L_0(1+L)|\mathbf{x}(s)-x_0| \\ &\quad + \langle F_p(x_0,u(x_0),p_0), \mathbf{p}(s)-p_0\rangle \\ &\quad + \frac{1}{2}|\mathbf{p}(s)-p_0|^2. \end{aligned}$$

Now, (5.5.66) yields

$$\begin{aligned}F(\mathbf{x}(s),u(\mathbf{x}(s)),\mathbf{p}(s)) &\leq F(x_0,u(x_0),p_0)+L_0(1+L)|\mathbf{x}(s)-x_0| \\ &\quad+\Big\langle\frac{\mathbf{x}(s)-x_0}{s}+p_0-\mathbf{p}(s),\mathbf{p}(s)-p_0\Big\rangle \\ &\quad+\frac{1}{2}|\mathbf{p}(s)-p_0|^2 \\ &= F(x_0,u(x_0),p_0)+L_0(1+L)|\mathbf{x}(s)-x_0| \\ &\quad+\frac{1}{s}\langle\mathbf{x}(s)-x_0,\mathbf{p}(s)-x_0\rangle-\frac{1}{2}|\mathbf{p}(s)-p_0|^2\,.\end{aligned}$$

Then, the semiconcavity estimate $\langle\mathbf{x}(s)-x_0,\mathbf{p}(s)-p_0\rangle\leq 2C|\mathbf{x}(s)-x_0|^2$ and (5.5.67) yield (5.5.63). Finally, (5.5.64) follows from (5.5.63) recalling the definition of ρ_0. The proof is thus complete. ∎

The next result provides a family of singular arcs satisfying suitable bounds. Such a family will be later shown to converge uniformly to a singular generalized characteristics. For brevity, let us set

$$\mathbf{p}^-(s):=\lim_{h\uparrow s}\mathbf{p}(h)\,.$$

Lemma 5.5.6 *Let $u\in\mathrm{SCL}(\Omega)$ be a solution of* (5.4.43). *Assume* (A1), (A2), (A3), *and suppose* (A4) *holds with $L_1=1$. Let $x_0\in\Omega$ be a singular point of u satisfying*

$$0\notin\operatorname{co}F_p(x,u(x),D^+u(x))\qquad\forall x\in B_R(x_0)\subset\Omega \tag{5.5.68}$$

for some $R>0$, and fix $p_0\in\arg\min_{p\in D^+u(x_0)}F(x_0,u(x_0),p)$. Then, for any $\varepsilon>0$ there exist arcs $\mathbf{x}_\varepsilon:[0,\rho_0]\to\Omega$, $\mathbf{p}_\varepsilon:[0,\rho_0]\to\mathbb{R}^N$, with $\mathbf{x}_\varepsilon(0)=x_0$ and $\mathbf{p}_\varepsilon(0)=p_0$, and a partition $0=s_0<s_1<\ldots<s_{k-1}<s_k=\rho_0$ with the following properties:

(i) $\displaystyle\max_{0\leq j\leq k-1}|s_{j+1}-s_j|<\varepsilon$

(ii) $\mathbf{p}_\varepsilon(s)\in D^+u(\mathbf{x}_\varepsilon(s))\quad\forall s\in[0,\rho_0]$

(iii) $\mathbf{p}_\varepsilon(\cdot)$ *is continuous on* $[s_j,s_{j+1}[\quad(j=0,\ldots,k-1)$

(iv) $|\mathbf{x}_\varepsilon(t)-\mathbf{x}_\varepsilon(s)|\leq M|t-s|\quad\forall t,s\in[0,\rho_0]$

(v) $F(\mathbf{x}_\varepsilon(s),u(\mathbf{x}_\varepsilon(s)),\mathbf{p}_\varepsilon(s))\leq\frac{1}{2}F(x_0,u(x_0),p_0)\quad\forall s\in[0,\rho_0]$

(vi) $|\mathbf{p}_\varepsilon(s)-\mathbf{p}_\varepsilon(s')|\leq 4\sqrt{\varepsilon}\quad\forall s,s'\in[s_j,s_{j+1}[\quad(j=0,\ldots,k-1)$, *where M and ρ_0 are defined in* (5.5.59) *and* (5.5.60). *Moreover, for any* $s \in [0, s_1[$,

$$\mathbf{x}_\varepsilon(s) - x_0 = [\mathbf{p}_\varepsilon(s) - p_0 + F_p(\mathbf{x}_\varepsilon(0), u(\mathbf{x}_\varepsilon(0)), \mathbf{p}_\varepsilon(0))]s, \tag{5.5.69}$$

while, for any $s \in [s_1, \rho_0[$,

$$\begin{aligned} &\mathbf{x}_\varepsilon(s) - x_0 \qquad\qquad (5.5.70)\\ &= \sum_{j=0}^{i(s)-1} [\mathbf{p}_\varepsilon^-(s_{j+1}) - \mathbf{p}_\varepsilon(s_j) + D_pF(\mathbf{x}_\varepsilon(s_j), u(\mathbf{x}_\varepsilon(s_j)), \mathbf{p}_\varepsilon(s_j))](s_{j+1} - s_j) \\ &\quad +[\mathbf{p}_\varepsilon(s) - \mathbf{p}_\varepsilon(s_{i(s)}) + F_p(\mathbf{x}_\varepsilon(s_{i(s)}), u(\mathbf{x}_\varepsilon(s_{i(s)})), \mathbf{p}_\varepsilon(s_{i(s)}))](s - s_{i(s)}), \end{aligned}$$

where $i(s) \in \{1, \ldots, k-1\}$ *is the unique integer such that* $s \in [s_{i(s)}, s_{i(s)+1}[$.

Proof — Let $\mathbf{x}(\cdot), \mathbf{p}(\cdot) : [0, \rho_0] \to \mathbb{R}^N$ be the continuous arcs given by Lemma 5.5.4 and let $\varepsilon > 0$ be fixed. Set

$$s_1 := \sup \left\{ t \in [0, \rho_0] \ : \ t + \frac{1}{2}|\mathbf{p}(s) - p_0|^2 < \varepsilon \quad \forall s \in [0, t] \right\}.$$

Then, property (i) is satisfied. If $s_1 = \rho_0$, define

$$\mathbf{x}_\varepsilon(s) := \mathbf{x}(s) \qquad \text{and} \qquad \mathbf{p}_\varepsilon(s) := \mathbf{p}(s) \qquad \forall s \in [0, s_1].$$

to conclude that properties (ii) – (v) and (5.5.69) can be obtained invoking Lemma 5.5.4, while (vi) follows from the trivial inequality $|\mathbf{p}_\varepsilon(s) - \mathbf{p}_\varepsilon(s')| \le |\mathbf{p}_\varepsilon(s) - p_0| + |\mathbf{p}_\varepsilon(s') - p_0|$ recalling the definition of s_1. This completes the proof in the case of $s_1 = \rho_0$.

Now, assuming that $s_1 < \rho_0$, define $\mathbf{x}_\varepsilon(\cdot)$ on $[0, s_1]$ as above, and choose

$$p_1 \in \arg \min_{p \in D^+u(x_1)} F(x_1, u(x_1), p)$$

where $x_1 := \mathbf{x}_\varepsilon(s_1)$. Let us also set

$$\mathbf{p}_\varepsilon(s) := \begin{cases} \mathbf{p}(s) & \text{for} \quad s \in [0, s_1[\\ \mathbf{p}_\varepsilon(s_1) = p_1. \end{cases}$$

Notice that $s_1 + \frac{1}{2}|\mathbf{p}(s_1) - p_0|^2 = \varepsilon$ since $\mathbf{p}(\cdot)$ is continuous. Hence, (5.5.63) yields

$$F(x_1, u(x_1), p_1) \le (C_1 + 1)s_1 + F(x_0, u(x_0), p_0) - \varepsilon . \tag{5.5.71}$$

As will be clearer later, this inequality plays a crucial role in the estimate of the number k of s_j needed to cover the interval $[0, \rho]$. Moreover, in order to prove (v), it is useful to observe that

$$F(\mathbf{x}_\varepsilon(s), u(\mathbf{x}_\varepsilon(s)), \mathbf{p}_\varepsilon(s)) \le F(x_0, u(x_0), p_0) + C_1 s \qquad \forall s \in [0, s_1], \tag{5.5.72}$$

in view of (5.5.63).

Now, let us proceed to construct s_2. Let $R_1 > 0$ be such that $B_{R_1}(x_1) \subset B_R(x_0)$, and let $\mathbf{x}^1 : [0, \rho_1] \to B_{R_1}(x_1)$, $\mathbf{p}^1 : [0, \rho_1] \to \mathrm{I\!R}^N$ be the arcs obtained applying Lemma 5.5.4 with x_0, p_0, R replaced by x_1, p_1, R_1, respectively. We note that all the properties of the conclusion of Lemma 5.5.4 hold for $\mathbf{x}^1 : [0, \rho_1] \to B_{R_1}(x_1)$ and $\mathbf{p}^1 : [0, \rho_1] \to \mathrm{I\!R}^N$ with the same constants M, C_0, C_1 computed in (5.5.58) and (5.5.59) for the base point x_0. Set

$$\begin{aligned} r_1 &= \sup\left\{ t \in [0, \rho_1] \,:\, t + \frac{1}{2}|\mathbf{p}^1(s) - p_1|^2 < \varepsilon \quad \forall s \in [0, t] \right\} \\ s_2 &= \min\{s_1 + r_1, \rho_0\}, \end{aligned}$$

and define $\mathbf{x}_\varepsilon$ on $[0, s_2]$ as

$$\mathbf{x}_\varepsilon(s) := \mathbf{x}^1(s - s_1) \qquad \forall s \in [s_1, s_2].$$

Then, either $s_2 = \rho_0$ or $s_2 < \rho_0$. In the former case, as one can easily check, the construction can be completed defining

$$\mathbf{p}_\varepsilon(s) := \mathbf{p}^1(s - s_1) \qquad \forall s \in [s_1, s_2].$$

Next, assume $s_2 < \rho_0$, set $x_2 := \mathbf{x}_\varepsilon(s_2)$, and fix

$$p_2 \in \arg \min_{p \in D^+ u(x_2)} F(x_2, u(x_2), p).$$

Let us define

$$\mathbf{p}_\varepsilon(s) := \begin{cases} \mathbf{p}^1(s - s_1) & \text{for} \quad s \in [s_1, s_2[\\ \mathbf{p}_\varepsilon(s_2) = p_2 . \end{cases}$$

We claim that $\mathbf{x}_\varepsilon(\cdot), \mathbf{p}_\varepsilon(\cdot)$ and $0 = s_0 < s_1 < s_2 < \rho_0$ satisfy all the required properties. Indeed, property (i) is a consequence of the definition of s_2 while (vi) follows arguing as in the proof of the same property in the interval $[0, s_1[$. Moreover, (ii)–(iv) hold by Lemma 5.5.4. Hence, we only need to prove conclusions (v) and (5.5.70). From properties (5.5.63) and (5.5.72) we deduce that

$$\begin{aligned} F(\mathbf{x}_\varepsilon(s), u(\mathbf{x}_\varepsilon(s)), \mathbf{p}_\varepsilon(s)) &\leq F(x_1, u(x_1), p_1) + C_1(s - s_1) \\ &\leq F(x_0, u(x_0), p_0) + C_1 s \end{aligned}$$

for any $s \in [s_1, s_2]$. Moreover, owing to (5.5.72), the last inequality holds for every $s \in [0, s_2]$. So, (v) follows as $\rho_0 \leq |F(x_0, u(x_0), p_0)|/2C_1$ by definition. In order to prove (5.5.70), we observe that

$$x_1 = x_0 + s_1[\mathbf{p}_\varepsilon^-(s_1) - p_0 + F_p(\mathbf{x}_\varepsilon(0), u(\mathbf{x}_\varepsilon(0)), \mathbf{p}_\varepsilon(0))]. \tag{5.5.73}$$

Then, by (5.5.62),

$$\mathbf{x}_\varepsilon(s) - x_1 = (s - s_1)[\mathbf{p}_\varepsilon(s) - p_1 + F_p(x_1, u(x_1), p_1)] \tag{5.5.74}$$

for all $s \in [s_1, s_2[$. So, (5.5.70) is a direct consequence of (5.5.73) and (5.5.74), and our claim is proved. Moreover, the same arguments used to prove (5.5.71), and (5.5.71) itself, yield

$$\begin{aligned} F(\mathbf{x}_\varepsilon(s_2), u(\mathbf{x}_\varepsilon(s_2)), \mathbf{p}^1(s_2)) &\leq (C_1 + 1)(s_2 - s_1) + F(x_1, u(x_1), p_1) - \varepsilon \\ &\leq (C_1 + 1)s_2 - 2\varepsilon\,. \end{aligned}$$

We will now show that the desired partition of $[0, \rho_0]$, $0 = s_0 < s_1 < \ldots < s_k = \rho_0$, can be obtained by a finite iteration of the above construction. Indeed, suppose that numbers $0 = s_0 < \ldots < s_n < \rho_0$, and arcs $\mathbf{x}_\varepsilon(s), \mathbf{p}_\varepsilon(s)$ ($s \in [0, s_n]$) are given so that conditions (i)–(vi), (5.5.69) and (5.5.70) are fulfilled. Assume in addition that

$$F(\mathbf{x}_\varepsilon(s_n), u(\mathbf{x}_\varepsilon(s_n)), \mathbf{p}^{n-1}(s_n)) \leq (C_1 + 1)s_n - n\varepsilon\,. \tag{5.5.75}$$

Define

$$C_2 := \sup\{|F(x, u(x), p)| \,:\, x \in B_R(x_0), |p| \leq L\}$$

and apply (5.5.75) to discover that $n\varepsilon - C_2 \leq (C_1+1)s_n$. The upper bound

$$n \leq \left[\frac{(C_1 + 1)\rho_0 + C_2}{\varepsilon}\right] + 1$$

follows, and the proof is complete. ∎

Lemma 5.5.7 *Under the assumptions of Lemma 5.5.6, there exists a constant $C > 0$, independent of ε, such that*

$$\left|\mathbf{x}_\varepsilon(r) - \mathbf{x}_\varepsilon(s) - \int_s^r F_p(\mathbf{x}_\varepsilon(t), u(\mathbf{x}_\varepsilon(t)), \mathbf{p}_\varepsilon(t))\,dt\right| \leq C\sqrt{\varepsilon} \tag{5.5.76}$$

for any $0 \leq s < r < \rho_0$.

Proof — From formula (5.5.70)—applied for r and s—it follows that

$$\begin{aligned} &\mathbf{x}_\varepsilon(r) - \mathbf{x}_\varepsilon(s) \qquad\qquad (5.5.77)\\ &= \sum_{j=i(s)}^{i(r)-1} [\mathbf{p}_\varepsilon^-(s_{j+1}) - \mathbf{p}_\varepsilon(s_j) + F_p(\mathbf{x}_\varepsilon(s_j), u(\mathbf{x}_\varepsilon(s_j)), \mathbf{p}_\varepsilon(s_j))](s_{j+1} - s_j) \\ &\quad + [\mathbf{p}_\varepsilon(r) - \mathbf{p}_\varepsilon(s_{i(r)}) + F_p(\mathbf{x}_\varepsilon(s_{i(r)}), u(\mathbf{x}_\varepsilon(s_{i(r)})), \mathbf{p}_\varepsilon(s_{i(r)}))](r - s_{i(r)}) \\ &\quad - [\mathbf{p}_\varepsilon(s) - \mathbf{p}_\varepsilon(s_{i(s)}) + F_p(\mathbf{x}_\varepsilon(s_{i(s)}), u(\mathbf{x}_\varepsilon(s_{i(s)})), \mathbf{p}_\varepsilon(s_{i(s)}))](s - s_{i(s)})\,. \end{aligned}$$

Now, observe that, in view of property (i) of Lemma 5.5.6,

$$\left| \int_{s_{i(s)}}^{s_{i(r)}} F_p(\mathbf{x}_\varepsilon(t), u(\mathbf{x}_\varepsilon(t)), \mathbf{p}_\varepsilon(t))\,dt - \int_s^r D_pF(\mathbf{x}_\varepsilon(t), u(\mathbf{x}_\varepsilon(t)), \mathbf{p}_\varepsilon(t))\,dt \right| \le C\varepsilon$$

for some constant C (hereafter C will denote any positive constant independent of ε). Also, by (5.5.77),

$$\begin{aligned}
&\left| \mathbf{x}_\varepsilon(r) - \mathbf{x}_\varepsilon(s) - \sum_{j=i(s)}^{i(r)-1} F_p(\mathbf{x}_\varepsilon(s_j), u(\mathbf{x}_\varepsilon(s_j)), \mathbf{p}_\varepsilon(s_j))(s_{j+1}-s_j) \right| \\
&\quad \le \sum_{j=i(s)}^{i(r)-1} |\mathbf{p}_\varepsilon^-(s_{j+1}) - \mathbf{p}_\varepsilon(s_j)|(s_{j+1}-s_j) + |\mathbf{p}_\varepsilon(s) - \mathbf{p}_\varepsilon(s_{i(s)})|(s - s_{i(s)}) \\
&\qquad + \Big[|F_p(\mathbf{x}_\varepsilon(s_{i(r)}), u(\mathbf{x}_\varepsilon(s_{i(r)})), \mathbf{p}_\varepsilon(s_{i(r)}))| \\
&\qquad + |\mathbf{p}_\varepsilon(r) - \mathbf{p}_\varepsilon(s_{i(r)})| \Big] (r - s_{i(r)}) \\
&\qquad + |F_p(\mathbf{x}_\varepsilon(s_{i(s)}), u(\mathbf{x}_\varepsilon(s_{i(s)})), \mathbf{p}_\varepsilon(s_{i(s)}))|(s - s_{i(s)})\,.
\end{aligned}$$

Hence, using properties (i) and (vi) of Lemma 5.5.6 to estimate the right-hand side above, we obtain

$$\left| \mathbf{x}_\varepsilon(r) - \mathbf{x}_\varepsilon(s) - \sum_{j=i(s)}^{i(r)-1} F_p(\mathbf{x}_\varepsilon(s_j), u(\mathbf{x}_\varepsilon(s_j)), \mathbf{p}_\varepsilon(s_j))(s_{j+1}-s_j) \right| \le C\sqrt{\varepsilon}\,.$$

Therefore, to prove (5.5.76) it suffices to show that the quantity

$$\left| \sum_{j=i(s)}^{i(r)-1} \int_{s_j}^{s_{j+1}} [F_p(\mathbf{x}_\varepsilon(s_j), u(\mathbf{x}_\varepsilon(s_j)), \mathbf{p}_\varepsilon(s_j)) - F_p(\mathbf{x}_\varepsilon(t), u(\mathbf{x}_\varepsilon(t)), \mathbf{p}_\varepsilon(t))]\,dt \right|$$

can be bounded by $C\sqrt{\varepsilon}$ for some $C > 0$. Aiming at this, we invoke assumption (A4) and properties (i), (iv) and (vi) of Lemma 5.5.6 to deduce that

$$\begin{aligned}
&|F_p(\mathbf{x}_\varepsilon(s_j), u(\mathbf{x}_\varepsilon(s_j)), \mathbf{p}_\varepsilon(s_j)) - F_p(\mathbf{x}_\varepsilon(t), u(\mathbf{x}_\varepsilon(t)), \mathbf{p}_\varepsilon(t))| \\
&\quad \le L_1[|\mathbf{x}_\varepsilon(s_j) - \mathbf{x}_\varepsilon(t)| + |u(\mathbf{x}_\varepsilon(s_j)) - u(\mathbf{x}_\varepsilon(t))| + |\mathbf{p}_\varepsilon(s_j) - \mathbf{p}_\varepsilon(t)| \\
&\quad \le L_1[M(1+L)|t - s_j| + 4\sqrt{\varepsilon}] \le C\sqrt{\varepsilon}
\end{aligned}$$

for any $t \in [s_j, s_{j+1}[$ and $i(s) \le j \le i(s+r) - 1$. The desired inequality follows, as well as (5.5.76). ∎

For the proof of Theorem 5.5.2 we also need the following measurability result for a selection of a set-valued map. We set $\mathcal{P}(\mathbb{R}^N) = \{A : A \subset \mathbb{R}^N\}$. We recall that a set–valued map $\Gamma : [0,1] \to \mathcal{P}(\mathbb{R}^N)$ is said to be measurable if for every open set $A \subset \mathbb{R}^N$ the set $\{t \in [0,1] : \Gamma(t) \cap A \ne \emptyset\}$ is measurable.

Lemma 5.5.8 *Let $F : [0,1] \to \mathcal{P}(\mathbb{R}^N)$ be convex with non empty closed images. Suppose in addition that F is upper semicontinuous and let $p : [0,1] \to \mathbb{R}^N$ be a piecewise continuous map. For $t \in [0,1]$, let $\widehat{p}(t)$ be the projection of $p(t)$ onto $F(t)$. Then, $\widehat{p}(\cdot)$ is measurable.*

Proof — The fact that F is convex yields

$$[p(t) + d_{F(t)}(p(t))\overline{B}_1] \cap F(t) = \{\widehat{p}(t)\} \qquad \forall t \in [0,1]$$

where the symbol $d_{F(t)}(\cdot)$ denotes the Euclidean distance function from $F(t)$. It is easy to see that the proof of the measurability of $\widehat{p}(\cdot)$ reduces to showing that the set-valued map $d_{F(\cdot)}(p(\cdot))\overline{B}_1$ is measurable.

We claim that the function $d_{F(\cdot)}(p(\cdot))$ is lower semicontinuous from the right, i.e. for every sequence $\{t_k\} \subset [0,1]$, with $t_k \downarrow t$,

$$\liminf_{k\to\infty} d_{F(t_k)}(p(t_k)) \geq d_{F(t)}(p(t)) \, .$$

Indeed, take a sequence $\{t_k\} \subset [0,1]$ such that $t_k \downarrow t$ as $k \to \infty$ and

$$\lim_{k\to\infty} d_{F(t_k)}(p(t_k)) = \liminf_{k\to\infty} d_{F(t_k)}(p(t_k)) \, .$$

The convexity of $F(\cdot)$ implies that, for any $p(t_k)$, there exists a unique $\widehat{p}(t_k) \in F(t_k)$ such that

$$d_{F(t_k)}(p(t_k)) = |\widehat{p}(t_k) - p(t_k)| \, .$$

Possibly extracting a subsequence, in view of the upper semicontinuity of $F(\cdot)$, we can suppose that $\widehat{p}(t_k)$ converges to a suitable $\overline{p}(t) \in F(t)$. Then, using the continuity of $p(\cdot)$, we deduce that

$$\begin{aligned} d_{F(t)}(p(t)) &\leq |\overline{p}(t) - p(t)| \\ &= \lim_{k\to\infty} d_{F(t_k)}(p(t_k)) = \liminf_{k\to\infty} d_{F(t_k)}(p(t_k)) \, , \end{aligned}$$

which proves our claim.

Now, to complete the proof we need only to verify that the set valued map $d_{F(\cdot)}(p(\cdot))\overline{B}_1$ is measurable, knowing that $d_{F(\cdot)}(p(\cdot))$ is lower semicontinuous from the right. This follows by Theorem 8.1.4, p. 310, of [15] since, for every $x \in \mathbb{R}^n$, the map

$$[0,1] \ni t \to \inf\{|x-y| \, : \, y \in d_{F(t)}(p(t))\overline{B}_1\} = (|x| - d_{F(t)}(p(t)))_+$$

where $(\cdot)_+$ denotes the positive part, is measurable. ∎

We are now ready to prove Theorem 5.5.2.

Proof of Theorem 5.5.2 — First, let us observe that, in view of Remark 5.5.3, (5.5.68) is satisfied for some $R > 0$. Also, possibly replacing F by F/L_1, we can assume that assumption (A4) holds with $L_1 = 1$ without loss of generality. The general case may indeed be recovered by a change of variable.

Let ε be a positive number, and let $\mathbf{x}_\varepsilon, \mathbf{p}_\varepsilon : [0, \rho_0] \to \mathbb{R}^N$ be the arcs given by Lemma 5.5.6. Condition (iv) of the same lemma ensures that $\{\mathbf{x}_\varepsilon(\cdot)\}_\varepsilon$ is Lipschitz continuous, uniformly with respect to ε. Moreover, $\{\mathbf{x}_\varepsilon(\cdot)\}_\varepsilon$ is easily seen to be equibounded. Therefore, there exists a sequence $\varepsilon = \varepsilon_j \downarrow 0$ and a Lipschitz continuous function $\mathbf{x}(\cdot)$ such that

$$\lim_{\varepsilon \to 0} \sup_{s \in [0, \rho_0]} |\mathbf{x}_\varepsilon(s) - \mathbf{x}(s)| = 0\,. \tag{5.5.78}$$

We claim that $\mathbf{x}(s) \in \Sigma(u)$ for any $s \in [0, \rho_0]$. Indeed, having fixed $s \in [0, \rho_0]$, let us extract a subsequence of $\mathbf{p}_\varepsilon(s)$ (still denoted by $\mathbf{p}_\varepsilon(s)$) which converges to some limit, $\mathbf{p}(s)$. By the upper-semicontinuity of the superdifferential, we have that $\mathbf{p}(s) \in D^+u(\mathbf{x}(s))$. Hence, passing to the limit as $\varepsilon \to 0$ in condition (v) of Lemma 5.5.6, we obtain

$$F(\mathbf{x}(s), u(\mathbf{x}(s)), \mathbf{p}(s)) \le \frac{1}{2} F(x_0, u(x_0), p_0) < 0$$

and the claim follows from (5.4.48).

Since $\mathbf{x}(\cdot)$ is Lipschitz continuous, by Rademacher's Theorem $\mathbf{x}(\cdot)$ is differentiable on a set of full measure $E \subset [0, \rho_0]$. We will prove that $\mathbf{x}(\cdot)$ satisfies equation (5.5.54) on E.

The main step of the proof is to show that, for any $0 \le s < r < \rho_0$,

$$\mathbf{x}(r) - \mathbf{x}(s) \in \int_s^r \operatorname{co} F_p(\mathbf{x}(t), u(\mathbf{x}(t)), D^+u(\mathbf{x}(t)))\, dt\,, \tag{5.5.79}$$

where the right-hand side denotes, as usual, the integrals of all measurable selections of the set-valued map $t \mapsto \operatorname{co} F_p(\mathbf{x}(t), u(\mathbf{x}(t)), D^+u(\mathbf{x}(t)))$. To prove (5.5.79), we observe first that (5.5.76), (5.5.78) and the Lipschitz continuity of u and F_p yield

$$\left| \mathbf{x}(r) - \mathbf{x}(s) - \int_s^r F_p(\mathbf{x}(t), u(\mathbf{x}(t)), \mathbf{p}_\varepsilon(t))\, dt \right| \le \omega(\varepsilon) \tag{5.5.80}$$

for some positive function $\omega(\cdot)$ satisfying $\lim_{\varepsilon \to 0} \omega(\varepsilon) = 0$. Now, let us denote by $\widehat{\mathbf{p}}_\varepsilon(t)$ the projection of $\mathbf{p}_\varepsilon(t)$ onto $D^+u(\mathbf{x}(t))$ for every $t \in [s, r]$, and set

$$q_\varepsilon(t) := \mathbf{p}_\varepsilon(t) - \widehat{\mathbf{p}}_\varepsilon(t)\,.$$

From Lemma 5.5.8 we deduce that $\widehat{\mathbf{p}}_\varepsilon(\cdot)$ is measurable, and so the same holds for $q_\varepsilon(\cdot)$. Moreover, $|q_\varepsilon(\cdot)|$ is bounded and converges to 0 almost everywhere in $[s,r]$, as $\varepsilon \to 0$, in view of the upper-semicontinuity of D^+u. Hence, by (5.5.80),

$$\begin{aligned} |\mathbf{x}(r) - \mathbf{x}(s) - \int_s^r F_p(\mathbf{x}(t), u(\mathbf{x}(t)), \widehat{\mathbf{p}}_\varepsilon(t))\, dt| & \\ \leq \ \omega(\varepsilon) + L_1 \int_s^r |q_\varepsilon(t)|\, dt =: \overline{\omega}(\varepsilon) & \qquad (5.5.81) \end{aligned}$$

where $\lim_{\varepsilon\to 0} \overline{\omega}(\varepsilon) = 0$ as a consequence of the Dominated Convergence Theorem. Moreover,

$$\phi_\varepsilon(t) := F_p(\mathbf{x}(t), u(\mathbf{x}(t)), \widehat{\mathbf{p}}_\varepsilon(t)) \in \mathrm{co}\; F_p(\mathbf{x}(t), u(\mathbf{x}(t)), D^+u(\mathbf{x}(t)))$$

is bounded in $L^2(s,r;\mathrm{I\!R}^N)$. Therefore, passing to a subsequence, we can assume that $\phi_\varepsilon(\cdot)$ converges to a function $\phi(\cdot)$, weakly in $L^2(s,r;\mathrm{I\!R}^N)$. Then $\phi(\cdot)$ is also a selection of $\mathrm{co}\, F_p(\mathbf{x}(\cdot), u(\mathbf{x}(\cdot)), D^+u(\mathbf{x}(\cdot)))$, and (5.5.79) follows taking the limit as $\varepsilon \to 0$ in (5.5.81).

Now, let $s \in E$ be a point of differentiability for $\mathbf{x}(\cdot)$. Then, using (5.5.79), the inclusion

$$\mathbf{x}'(s) \in \ \mathrm{co}\; F_p(\mathbf{x}(s), u(\mathbf{x}(s)), D^+u(\mathbf{x}(s)))\,, \qquad (5.5.82)$$

follows by the following argument, which is standard in control theory. Suppose that $\mathbf{x}'(s) \notin \ \mathrm{co}\; F_p(\mathbf{x}(s), u(\mathbf{x}(s)), D^+u(\mathbf{x}(s)))$. Then, there exist a positive number δ and a vector $\eta \in \mathrm{I\!R}^N$, $|\eta| = 1$, such that

$$\langle \mathbf{x}'(s), \eta\rangle \leq \langle p, \eta\rangle - \delta, \quad \forall p \in \ \mathrm{co}\; F_p(\mathbf{x}(s), u(\mathbf{x}(s)), D^+u(\mathbf{x}(s))).$$

By a continuity argument, one can find $\delta_1 > 0$ such that, for any $x \in B_{\delta_1}(\mathbf{x}(s))$,

$$\langle \mathbf{x}'(s), \eta\rangle \leq \langle p, \eta\rangle - \delta/2, \quad \forall p \in \ \mathrm{co}\; F_p(x, u(x), D^+u(x)). \qquad (5.5.83)$$

Moreover, for any $\delta_2 > 0$ there exists $r(\delta_2) > 0$ such that

$$\left|\frac{\mathbf{x}(r) - \mathbf{x}(s)}{r-s} - \mathbf{x}'(s)\right| < \delta_2\,, \qquad |\mathbf{x}(r) - \mathbf{x}(s)| < \delta_1 \qquad (5.5.84)$$

for any $r \in [s, s + r(\delta_2)]$. From (5.5.79) it follows that

$$\Big\langle \frac{\mathbf{x}(r) - \mathbf{x}(s)}{r-s}, \eta \Big\rangle \in \frac{1}{r-s}\int_s^r \langle \mathrm{co}\, F_p(\mathbf{x}(t), u(\mathbf{x}(t)), D^+u(\mathbf{x}(t))), \eta\rangle\, dt.$$

Hence, by (5.5.83),

$$\Big\langle \frac{\mathbf{x}(r) - \mathbf{x}(s)}{r-s}, \eta \Big\rangle \geq \langle \mathbf{x}'(s), \eta\rangle + \frac{\delta}{2}\,.$$

Moreover, recalling (5.5.84) we conclude that

$$\delta_2 > \left| \frac{\mathbf{x}(r) - \mathbf{x}(s)}{r - s} - \mathbf{x}'(s) \right| \geq \frac{\delta}{2},$$

contrary to the fact δ_2 is any positive number. So, (5.5.82) follows.

It remains to prove that, possibly reducing ρ_0, the map $[0, \rho_0] \ni s \mapsto \mathbf{x}(s)$ is injective. We will argue once again by contradiction, assuming the existence of two sequences $\{s_i\}$ and $\{r_i\}$, with $s_i < r_i$, such that $s_i, r_i \to 0$ and $\mathbf{x}(s_i) = \mathbf{x}(r_i)$. By (5.5.79), we have that

$$0 \in \frac{1}{r_i - s_i} \int_{s_i}^{r_i} \mathrm{co} F_p(\mathbf{x}(t), u(\mathbf{x}(t)), D^+u(\mathbf{x}(t)))\, dt\,.$$

Hence, as $i \to \infty$,

$$0 \in \mathrm{co} F_p(x_0, u(x_0), D^+u(x_0))$$

in contradiction with assumption (5.5.55). This completes the proof. ∎

5.6 Examples

In this section we apply our results for propagation of singularities to some specific examples. Let us first consider the eikonal equation

$$\begin{cases} |Du(x)|^2 = 1 & \text{a.e. in} \quad \mathbb{R}^N \setminus S \\ \\ u(x) = 0 & \text{on} \quad \partial S\,, \end{cases} \tag{5.6.85}$$

where $S \subset \mathbb{R}^N$ is a non-empty closed set.

Remark 5.6.1 It is easily seen that d_S (the euclidean distance function from S) is a viscosity solution of the above problem. In fact, d_S is semi-concave in $\mathbb{R}^N \setminus S$ (see Proposition 2.2.1(ii)) and satisfies the equation at all point of differentiability. Using standard techniques from the theory of the viscosity solutions (see e.g. [10]) one can prove that d_S is the unique *positive* solution of (5.6.85).

We start our analysis with a uniqueness result for a differential inclusion, which, up to a constant factor 2, is the equation of the generalized characteristics for (5.6.85) starting from a given point.

Lemma 5.6.2 *Let $u \in \mathrm{SCL}(\Omega)$. Then for every $x_0 \in \Omega$ there exists $\rho > 0$ such that the following problem admits a unique solution*

$$\begin{cases} \mathbf{x}'(s) \in D^+u(\mathbf{x}(s)) & a.e. \ in \quad [0, \rho] \\ \mathbf{x}(0) = x_0\,. \end{cases} \tag{5.6.86}$$

Proof — The existence of a solution of (5.6.86) is a classical result on differential inclusions. A proof can be found, for example, in [32], Section 4.1. In order to prove the uniqueness, let us suppose that problem (5.6.86) possesses two solutions, say $\mathbf{x}(\cdot)$ and $\overline{\mathbf{x}}(\cdot)$. Property (3.3.24) implies that, for a.e. s in $[0, \rho]$,

$$\frac{d}{ds}|\mathbf{x}(s) - \overline{\mathbf{x}}(s)|^2 = \langle \mathbf{x}(s) - \overline{\mathbf{x}}(s), \mathbf{x}'(s) - \overline{\mathbf{x}}'(s)\rangle \leq 2C|\mathbf{x}(s) - \overline{\mathbf{x}}(s)|^2$$

for some $C > 0$. Thus, the conclusion is an immediate consequence of the Gronwall inequality. ■

Theorem 5.6.3 *Let* $u \in \mathrm{SCL}\,(\mathbb{R}^N \setminus S)$ *be a solution of equation* (5.6.85) *and let* $x_0 \in \Sigma(u)$*. Then, the following assertions are equivalent:*

(i) $0 \notin D^+u(x_0)$*;*

(ii) *there exists a non constant singular generalized characteristic of the eikonal equation with initial point* x_0*, that is a solution,* $\mathbf{x}(\cdot)$*, of problem* (5.6.86) *satisfying* $\mathbf{x}(s) \in \sigma(u)$*, for any* $s \in [0, \rho]$*.*

Remark 5.6.4 If in the statement of the above result we add the assumption that $u > 0$, then assertion (i) possesses a clear geometrical meaning. In fact, in light of Remark 5.6.1, we have that $u = d_S$ and, using Proposition 4.4.1(iii), one can prove that $0 \notin D^+d_S(x_0)$ if $x_0 \notin \mathrm{co}\, S$.

Proof — Implication (i) $\Rightarrow$ (ii) is a direct consequence of Theorem 5.5.2. In order to prove (ii) $\Rightarrow$ (i) we will argue by contradiction. For this purpose, let us suppose that (ii) holds and that $0 \in D^+u(x_0)$. Clearly, $\mathbf{x}(s) \equiv x_0$ is a singular generalized characteristic. Moreover, Lemma 5.6.2 ensures that this constant arc is the unique generalized characteristic with initial point x_0. This contradiction completes the proof of implication (ii) $\Rightarrow$ (i). ■

We now analyze the propagation of singularities for semiconcave solutions of the Hamilton-Jacobi equation

$$u_t + H(t, x, u, \nabla u) = 0 \qquad \text{a.e. in} \qquad]0, T[\, \times U\,, \tag{5.6.87}$$

where $U \subset \mathbb{R}^n$ is an open set. We assume that $H :]0, T[\, \times U \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ is a continuous function such that

(H1) $p \mapsto H(t, x, u, p)$ is differentiable and strictly convex for any (t, x, u);

(H2) H and H_p are Lipschitz continuous.

The following propagation result is a direct consequence of Theorem 5.4.4.

Theorem 5.6.5 *Let* $u \in \mathrm{SCL}\,(\,]0,T[\,\times U)$ *be a solution of* (5.6.87). *Let us suppose that* (H1) *holds. Then, for any* $(t,x) \in \Sigma(u)$

$$\partial D^+u(t,x) \setminus D^*u(t,x) \neq \emptyset\,.$$

Consequently, all singularities propagate.

For solutions to (5.6.87) there is a natural direction for the propagation of singularities. To describe this fact, let us define

$$\nabla^+u(t,x) := \left\{\eta \in \mathbb{R}^n \,:\, \limsup_{h\to 0} \frac{u(t,x+h)-u(t,x)-\langle \eta, h\rangle}{|h|} \leq 0\right\}\,,$$

and

$$\nabla^*u(t,x) = \{p \in \mathbb{R}^n \,:\, x_i \to x\,,\ \ \nabla u(t,x_i) \to p\}\ .$$

Observe that, if $u \in SC(\,]0,T[\,\times U)$ then

$$\nabla^+u(t,x) = \mathrm{co}\,\nabla^*u(t,x)\,. \tag{5.6.88}$$

Let $\Pi_x : \mathbb{R}^{n+1} \to \mathbb{R}^n$ be the projection onto the x–space, i.e. $\Pi_x(t,x) = x$. The next result explains the relationship between $\nabla^+u(t,x)$ and $D^+u(t,x)$.

Lemma 5.6.6 *For any* $u \in SC(]0,T[\,\times U)$, *we have that*

$$\Pi_x D^+u(t,x) = \nabla^+u(t,x) \qquad \forall (t,x) \in \,]0,T[\,\times U\,. \tag{5.6.89}$$

Proof — The inclusion $\Pi_x D^+u(t,x) \subset \nabla^+u(t,x)$ is a simple consequence of the definition. Let us prove that $\nabla^+u(t,x) \subset \Pi_x D^+u(t,x)$. First, we observe that it suffices to show that

$$\nabla^*u(t,x) \subset \Pi_x D^*u(t,x)\,. \tag{5.6.90}$$

To prove this inclusion, fix $p^x := \lim_{i\to\infty} \nabla u(t,x_i) \in \nabla^*u(t,x)$. By the upper semicontinuity of D^+u we can find, for any $i \in \mathbb{N}$, a point $\{(t_i,y_i)\}$ where u is differentiable and such that

$$|(t,x_i)-(t_i,y_i)| < \frac{1}{i}, \qquad Du(t_i,y_i) \in D^+u(t,x_i) + \frac{1}{i}B_1\,.$$

The last property implies that $|\nabla u(t_i,y_i) - \nabla u(t,x_i)| \leq 1/i$. Thus, we conclude that $(t_i,x_i) \to (t,x)$ and $\nabla u(t_i,y_i) \to p^x$, which implies that $p^x \in \Pi_x D^*u(t,x)$. ∎

We can now give our propagation result for solutions of (5.6.87).

Theorem 5.6.7 *Let $u \in SC(]0,T[\times U)$ be a solution of* (5.6.87)*. Let us suppose that* (H1)*,* (H2) *and* (H3) *hold and let $(t_0, x_0) \in \Sigma(u)$. Then, there exists a generalized characteristic of equation* (5.6.87)*, $(t, x(t))$, with initial point (t_0, x_0) defined on some interval $[t_0, \rho] \subset]0, T[$, i.e. there exists a solution of*

$$\begin{cases} x'(t) \in \mathrm{co} H_p(t, x(t), u(x(t)), \nabla^+ u(t, x(t))) & a.\ e.\ in \quad (t_0, \rho) \\ x(0) = x_0 \end{cases} \tag{5.6.91}$$

such that $(t, x(t)) \in \Sigma(u)$, for every $t \in [t_0, \rho]$.

Proof — Set for any $(t, x, u, p_t, p_x) \in]0, T[\times U \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^n$

$$F(t, x, u, p_t, p_x) := p_t + H(t, x, u, p_x)\,.$$

It is immediate to the see that the above assumptions ensure that (A1), (A3) and (A4) of Theorem 5.5.2 are fulfilled. Now, to prove that (A2) holds, suppose that

$$[(p_t^1, p_x^1), (p_t^2, p_x^2)] \subset \{(p_t, p_x) \in \mathbb{R}^{n+1} \,:\, F(t, x, u, p_t, p_x) = 0\}\,. \tag{5.6.92}$$

We want to show that $(p_t^1, p_x^1) = (p_t^2, p_x^2)$. In fact, (5.6.92) yields

$$\lambda_1 p_t^1 + \lambda_2 p_t^2 + H(t, x, u, \lambda_1 p_x^1 + \lambda_2 p_x^2) = 0$$

for every $\lambda_1, \lambda_2 > 0$ such that $\lambda_1 + \lambda_2 = 1$. Hence, recalling that

$$p_t^i + H(t, x, u, p_x^i) = 0 \qquad (i = 1, 2) \tag{5.6.93}$$

we obtain

$$H(t, x, u, \lambda_1 p_x^1 + \lambda_2 p_x^2) = \lambda_1 H(t, x, u, p_x^1) + \lambda_2 H(t, x, u, p_x^2)\,.$$

So, the above equality and assumption (H1) yield $p_x^1 = p_x^2$ which in turn implies that $p_t^1 = p_t^2$ on account of (5.6.93), i.e. (A2) holds.

Finally, since

$$F_p((t, x), u, (p_t, p_x)) = (1, H_p(t, x, u, p_x)), \tag{5.6.94}$$

assumption (5.5.55) is trivially fulfilled.

Now, from Theorem 5.5.2 we deduce that there exists a generalized characteristic of equation (5.6.87) with initial point (t_0, x_0).

To complete the proof, it remains to show that the x-projection of such a generalized characteristic solves (5.6.91). This follows from the identity

$$\begin{aligned} F_p((t, x), u, D^+ u(t, x)) &= (1, H_p(t, x, u, \Pi_x D^+ u(t, x))) \\ &= (1, H_p(t, x, u, \nabla^+ u(t, x))) \end{aligned}$$

that can be easily derived from (5.6.94) applying Lemma 5.6.6. ■

We point out that in the above result, one cannot expect that $\rho = +\infty$. In other words, in general, singularities may disappear, as shown in the next example.

Example 5.6.8 Let $t_0 > 0$ be fixed and consider the function $u(t,x) = -(t-t_0)^2|x|$, for $(t,x) \in [0,+\infty[\times\mathbb{R}$. Since u is obtained as infimum of smooth functions it is easy to see that u is semiconcave. Moreover, it is clear that $\Sigma(u) = \{(t,0) \,:\, t \neq t_0, \quad t \geq 0\}$ and that u solves the following problem

$$\begin{cases} u_t + 2(t-t_0)|x| + |u_x|^2 - (t-t_0)^2 = 0 & \text{a.e. in} \quad [0,+\infty[\,\times\mathbb{R} \\ \\ u(0,x) = -t_0^2|x| & x \in \mathbb{R}\,. \end{cases}$$

Remark 5.6.9 As proved in [3], one can show that $\rho = +\infty$, if u is a concave solution of the Hamilton–Jacobi equation $u_t + H(\nabla u) = 0$. Also, if H does not depend on u and $n = 1$, then equation (5.6.87) can be transformed into a conservation law and one can apply the results of Dafermos [39] to find that $\rho = +\infty$.

Bibliographical notes

The method of characteristics is classical. It was already described in the text [30] by Carathéodory. For more details on this method the reader can consult [42] and the references therein.

Viscosity solutions have been introduced by Crandall and Lions [38] at the beginning of the 80's. Some ideas of the theory were already present in Kruzhkov's work on hyperbolic conservation laws [61]; for instance, the definition of the "good" generalized solutions by means of a family of inequalities, or the doubling of the variables in the uniqueness proof. Apart from these formal similarities, however, the two theories are substantially independent. Shortly afterwards, Lions' book [67] unified and generalized the existing results on first order equations and put in a rigorous setting the connections with control theory.

A simplified approach to the theory of viscosity solutions was given by Crandall, Evans and Lions in [36]. Our presentation of Theorems 5.2.2 and 5.2.5 follows the one of that paper. In the following years, the theory has been considerably expanded; the most significant achievement has been the treatment of second order degenerate elliptic equations at the end of the 80's. A comprehensive exposition of these developments can be found in the textbooks and survey papers listed at the beginning of Section 5.2.

The close relation between semiconcavity and viscosity solutions was clearly indicated in the first works on viscosity solutions [38, 67]. Afterwards it was recognized that the semiconcavity property could provide improved convergence rate for approximating schemes [29] and a fine description of the singularities [27]. In [28] it was shown that the mild regularity assumptions of Theorem 5.3.8 suffice to obtain the generalized one-sided estimate of Section 2.5, while in [75] it was proved that such an estimate implies semiconcavity. Theorem 5.3.9 is due to Ishii [55]; following [17], we have proved the theorem under slightly stronger hypotheses than the original ones in order to avoid some technicalities in the proof.

The results of the remaining sections are taken from [4]. Previous results on propagation of singularities for solutions of Hamilton–Jacobi equations have been obtained in [27, 28, 11]; these results are weaker than the one described here, but hold in cases where the modulus of semiconcavity of the solutions is not necessarily linear.

We recall that the notion of generalized characteristic is well known in the context of hyperbolic conservation laws from the work of Dafermos [39] (see also Wu [84]). Dafermos' analysis of the singular set (which will be summarized at the beginning of Section 6.3) is very detailed, but is restricted to the one–dimensional case.

Chapter 6

Calculus of variations

In this chapter we consider problems in calculus of variations generalizing the class analyzed in Chapter 1. In fact, we no longer restrict ourselves to the case where the integrand does not depend on t, x. In this general setting the value function cannot be represented by Hopf's formula, but it still enjoys some important properties we have seen in Chapter 1; in particular, it can be characterized as the unique semiconcave solution, or viscosity solution, of an associated Hamilton-Jacobi equation. The main purpose of our analysis here is to study the singularities of the value function and their interpretation in calculus of variations. For instance, we derive a correspondence between generalized gradients of the value function and optimal trajectories of the problem. This shows, in particular, that the singularities of the value function are exactly the endpoints for which the minimizing trajectory of the variational problem is not unique. In addition, for this problem it is possible to bound the size of the closure of the singular set Σ, proving that it enjoys the same rectifiability properties of Σ itself. This result is interesting because on the complement of $\overline{\Sigma}$ the value function has the same regularity as the data and can be computed by the method of characteristics.

The chapter is organized as follows. In Section 6.1 we recall some classical results about the existence and regularity of minimizers and give a proof of the Jacobi optimality condition for conjugate points. Then, in Section 6.2, we apply the dynamic programming approach to this problem. We show that the value function u solves in the viscosity sense the associated Hamilton–Jacobi equation and that it is semiconcave. In addition, the minimizers for the problem of calculus of variations with final endpoint (t, x) are in one–to–one correspondence with the elements of $D^*u(t, x)$ at each point; in particular, the differentiability of u is equivalent to the uniqueness of the minimizer. We show that u is as regular as the data of the problem in the complement of the closure $\overline{\Sigma}$ of its singular set Σ. Since the properties of Σ are well known from the general analysis of Chapter 4, the rest of the chapter is devoted to study $\overline{\Sigma}$ and in particular how much larger than Σ itself this set can be.

In Section 6.3 we characterize the closure of the singular set showing that $\overline{\Sigma} = \Sigma \cup \Gamma$, where Γ is the set of conjugate points. In addition, we prove some results about conjugate points showing, roughly speaking, that these are the points at which the singularities of u are generated.

The aim of Section 6.4 is to prove that $\overline{\Sigma}$ has the same rectifiability property as Σ, i.e., that it is a countably $\mathcal{H}^n$–rectifiable subset of $\mathbb{R} \times \mathbb{R}^n$. By the previous remarks, this is equivalent to proving the rectifiability of $\Gamma \setminus \Sigma$. Combining a careful analysis of the hamiltonian system satisfied by the minimizing arcs with some tools of geometric measure theory, we obtain the finer estimate

$$\mathcal{H}^{n-1+2/R}(\Gamma \setminus \Sigma) = 0,$$

where $R \geq 2$ is the differentiability class of the data. Therefore $\overline{\Sigma}$ is countably rectifiable and u is smooth in the complement of a closed rectifiable set.

6.1 A problem in calculus of variations

Let us consider again the problem in calculus of variations introduced in Chapter 1. We first recall some notation. For fixed $T \in]0, +\infty]$, we set $Q_T =]0, T[\times \mathbb{R}^n$. For any $t \in]0, T[$, let J_t be the functional

$$J_t(\xi) = \int_0^t L(s, \xi(s), \dot{\xi}(s))ds + u_0(\xi(0)) \tag{6.1.1}$$

defined on the class of all absolutely continuous arcs $\xi : [0, t] \to \mathbb{R}^n$. Here L, u_0 are given functions called *running cost* and *initial cost* respectively. Now, for a given $(t, x) \in Q_T$, consider the problem

(CV) minimize J_t over all arcs ξ satisfying the terminal condition $\xi(t) = x$.

We assume that

(L1) $L \in C^{R+1}(\overline{Q_T} \times \mathbb{R}^n)$, where $R \geq 1$ is a fixed integer;

(L2) $L(t, x, v) \geq \lambda(|v|) + \lambda_0$ for some $\lambda_0 \in \mathbb{R}$ and some nonnegative function λ such that $\lambda(r)/r \to +\infty$ as $r \to +\infty$;

(L3) $L_{vv}(t, x, v) > 0$;

(L4) for all $r > 0$ there exists $C(r) > 0$ such that

$$|L_x(t, x, v)| + |L_v(t, x, v)| < C(r)\lambda(|v|) \quad \forall\, t \in [0, T],\ x \in B_r,\ v \in \mathbb{R}^n.$$

Here we have denoted by L_{vv} the matrix of the second derivatives of L with respect to the v variable. Function λ is often referred to as a *Nagumo function* for functional J_t. We also assume that the initial cost u_0 satisfies

$$u_0 \in W^{1,\infty}(\mathbb{R}^n) \cap C^{R+1}(\mathbb{R}^n). \tag{6.1.2}$$

Then the following results hold.

Theorem 6.1.1 *Assume* (L1)—(L4), *and* (6.1.2). *Then*

(i) *for any $(t, x) \in Q_T$ there exists a minimizer ξ^* for problem* (CV);

(ii) *any minimizer ξ^* is of class $C^2([0, t])$ and satisfies the* Euler–Lagrange *equations*

$$\frac{d}{ds} L_v(s, \xi(s), \dot{\xi}(s)) = L_x(s, \xi(s), \dot{\xi}(s)) \tag{6.1.3}$$

together with the transversality condition

$$L_v(0, \xi(0), \dot{\xi}(0)) = Du_0(\xi(0)); \tag{6.1.4}$$

(iii) *for any $r > 0$, there exists $M(r) > 0$ such that, if $(t, x) \in]0, r[\, \times B_r$ and ξ^* is a minimizer for (t, x), then*

$$\sup_{s \in [0,t]} |\dot{\xi}^*(s)| \leq M(r).$$

Proof — See e.g. [48, Sect. I.9] ∎

Remark 6.1.2 In the case when $L = L(q)$, using the above results we recover Hopf's formula

$$\min J_t(\xi) = \min_{z \in \mathbb{R}^n} \left[tL\left(\frac{x - z}{t}\right) + u_0(z) \right] \tag{6.1.5}$$

derived in Theorem 1.2.1. In fact, in this case the Euler–Lagrange equations (6.1.3) are equivalent to $\ddot{\xi} = 0$. Therefore any minimizing arc is linear, and this property yields easily Hopf's formula. Thus, in this particular case the minimization problem (CV) can be reduced to problem (6.1.5), which is simpler to treat since it is on a finite dimensional space. Nevertheless, many of the properties which were proved in Chapter 1 in the case $L = L(q)$ by means of Hopf's formula can be extended to the general case, as we will see in the following.

Any arc $\xi \in C^2([0,t])$ which solves the Euler-Lagrange equations (6.1.3) together with the initial condition (6.1.4) and the final condition $\xi(t) = x$ is called an *extremal* for problem (CV) with endpoint (t,x). By the previous theorem, a minimizing arc is also an extremal. This property is usually not enough in practice to determine minimizers. A first reason is that being an extremal is only a necessary condition for being a minimizer. Another difficulty is that finding all extremals requires solving a boundary value problem for Euler-Lagrange system, since the value of $\xi(s)$ is prescribed at $s = t$ while the value of $\dot{\xi}(s)$ is prescribed at $s = 0$. Since finding all solutions to this problem may not be an easy task, it is important to have additional tools to study minimizers.

Let us set

$$H(t,x,p) = \max_{v \in \mathbb{R}^n} [p \cdot v - L(t,x,v)]. \tag{6.1.6}$$

Function H is the Legendre–Fenchel transform of L with respect to the third argument (see Appendix) and is called the *hamiltonian* associated with L.

Euler's equations (6.1.3) can be equivalently restated as a first order system of $2n$ equations, as it is shown below. Such a formulation, called *hamiltonian*, is more convenient to use in most of our subsequent computations.

Theorem 6.1.3 *Assume* (L1)—(L4), *and* (6.1.2). *Let $\xi^* \in C^2([0,t])$ be an extremal arc for problem* (CV) *and let us set*

$$\eta^*(s) = L_v(s, \xi^*(s), \dot{\xi}^*(s)), \qquad s \in [0,t].$$

Then $\eta^*(0) = Du_0(\xi^*(0))$ and the pair (ξ^*, η^*) satisfies

$$\begin{cases} \dot{\xi}^*(s) = H_p(s, \xi^*(s), \eta^*(s)) \\ \dot{\eta}^*(s) = -H_x(s, \xi^*(s), \eta^*(s)). \end{cases} \tag{6.1.7}$$

Conversely, suppose that $\xi^*, \eta^* \in C^2([0,t])$ solve system (6.1.7) *together with the condition $\eta^*(0) = Du_0(\xi^*(0))$. Then ξ^* is an extremal arc for problem* (CV).

Proof — Since $L_v(t,x,\cdot)$ and $H_p(t,x,\cdot)$ are reciprocal inverse (see (A.1.13)), the definition of η^* implies that

$$\dot{\xi}^*(s) = H_p(s, \xi^*(s), \eta^*(s)).$$

On the other hand, since ξ^* is an extremal, we have

$$\dot{\eta}^*(s) = L_x(s, \xi^*(s), \dot{\xi}^*(s)) \qquad \eta^*(0) = Du_0(\xi^*(0)). \tag{6.1.8}$$

Recalling that $-H_x(t,x,L_v(t,x,v)) = L_x(t,x,v)$ (see (A.1.11)) we obtain the first part of the assertion. The converse part is obtained by similar arguments. ∎

Following the terminology of control theory, we will call η^* the *dual arc* associated with ξ^*.

For any $z \in \mathbb{R}^n$, we denote by $(\xi(\cdot,z), \eta(\cdot,z))$ the solution of

$$\begin{cases} \dot{\xi}(s) = H_p(s,\xi(s),\eta(s)) \\ \dot{\eta}(s) = -H_x(s,\xi(s),\eta(s)) \end{cases} \qquad s \geq 0, \tag{6.1.9}$$

with initial data

$$\begin{cases} \xi(0) = z \\ \eta(0) = Du_0(z). \end{cases} \tag{6.1.10}$$

Our regularity assumptions imply that ξ, η and their time derivatives are of class C^R. In addition, since $\xi(\cdot,z)$ is also solution of the Euler–Lagrange equations (6.1.3), it can be equivalently characterized as the solution of

$$\ddot{\xi} = L_{vv}^{-1}(-L_{vx}\dot{\xi} - L_{vt} + L_x)$$

(where the derivatives of L are evaluated at $(s,\xi(s),\dot{\xi}(s))$) with initial data

$$\begin{cases} \xi(0) = z \\ \dot{\xi}(0) = H_p(0,z,Du_0(z)). \end{cases} \tag{6.1.11}$$

Observe that, by Theorem 6.1.3, an arc is an extremal for problem (CV) if and only if it is of the form $\xi = \xi(\cdot,z)$ for some $z \in \mathbb{R}^n$.

By differentiating (6.1.9)–(6.1.10) with respect to z we obtain that ξ_z and η_z satisfy the linearized system

$$\begin{cases} \dot{\xi}_z = H_{px}\xi_z + H_{pp}\eta_z \\ \dot{\eta}_z = -H_{xx}\xi_z - H_{xp}\eta_z, \end{cases} \tag{6.1.12}$$

with initial conditions

$$\begin{cases} \xi_z(0) = I \\ \eta_z(0) = D_{zz}u_0(z). \end{cases} \tag{6.1.13}$$

By well known properties of linear systems, we have for any z

$$\operatorname{rank}\begin{pmatrix} \xi_z(t,z) \\ \eta_z(t,z) \end{pmatrix} = \operatorname{rank}\begin{pmatrix} \xi_z(0,z) \\ \eta_z(0,z) \end{pmatrix} = n, \qquad t \geq 0. \tag{6.1.14}$$

Therefore, for any $t > 0$, $z, \theta \in \mathbb{R}^n$, $\theta \neq 0$,

$$\xi_z(t,z)\theta = 0 \implies \eta_z(t,z)\theta \neq 0. \tag{6.1.15}$$

We now recall further classical notions from calculus of variations.

Definition 6.1.4 (Conjugate and irregular points)

(i) *A point* $(t, x) \in Q_T$ *is called* regular *if there exists a unique minimizer of* (CV) *with endpoint* (t, x)*. All other points are called* irregular.

(ii) *A point* $(t, x) \in Q_T$ *is called* conjugate *if* $z \in \mathbb{R}^n$ *exists such that* $\xi(t, z) = x$, $\xi(\cdot, z)$ *is a minimizer of* (CV)*, and*

$$\det \xi_z(t, z) = 0.$$

We denote by Σ *the set of irregular points and by* Γ *the set of conjugate points.*

It is easily checked that, for any $R > 0$, one can find $T > 0$ such that, for all $t \le T$, $\xi(t, z_1) \ne \xi(t, z_2)$ for all $z_1, z_2 \in B_R$ and $\det \xi_z(t, z) \ne 0$ for all $z \in B_R$. Thus, in an initial time interval (whose size may depend on R) all points are regular and not conjugate.

Before continuing our analysis, let us consider a specific example to illustrate the meaning of the notions introduced above.

Example 6.1.5 Let us consider the case when $n = 1$ and $L(q) = q^2/2$. We make the following assumption on the initial cost.

(A) The second derivative derivative u_0'' has the following behavior: it attains a minimum value $c^* < 0$ at some point z^*, is strictly decreasing in $]-\infty, z^*]$, strictly increasing in $[z^*, +\infty[$ and tends to 0 as $z \to \pm\infty$.

The solution to system (6.1.9)–(6.1.10) is given by

$$\xi(t, z) = z + u_0'(z)t, \qquad \eta(t, z) = u_0'(z), \qquad t \ge 0, z \in \mathbb{R}.$$

For a fixed $(\bar{t}, \bar{x})$, we want to find the minimizers of problem (CV). We know that we can restrict our attention to the arcs of the form $\xi(\cdot, z)$ for some z. Therefore we will first find the values of z such that $\xi(\bar{t}, z) = \bar{x}$ and then find among them the minimizing one ones. Let us set $T^* = -1/c^*$. Since $\xi_z(t, z) = 1 + u_0''(z)t$, we see that the analysis is different depending on whether $\bar{t}$ is smaller or greater than T^*.

Case (i): $\bar{t} \le T^*$. In this case the function $z \to \xi(\bar{t}, z)$ is strictly increasing in $\mathbb{R}$ and so there exists a unique $\bar{z}$ such that $\xi(\bar{t}, \bar{z}) = \bar{x}$. We deduce that $\xi(\cdot, \bar{z})$ is the unique minimizer for $(\bar{t}, \bar{x})$. Thus, all points $(\bar{t}, \bar{x})$ with $\bar{t} \le T^*$ are regular. In addition, $\xi_z(t, z) > 0$ for all $z \in \mathbb{R}$ and $t \le T^*$ except for the case $z = z^*, t = T^*$. Therefore, setting $x^* = z^* + u_0'(z^*)T^* = \xi(z^*, T^*)$, we have that the point (T^*, x^*) is conjugate, while all other points $(\bar{t}, \bar{x})$ with $\bar{t} \le T^*$ are not.

Case (ii): $\bar t > T^*$. This case requires a longer analysis. By assumption (A), there are z_1 and z_2, with $z_1 < z^* < z_2$, such that $u_0''(z_1) = u_0''(z_2) = -1/\bar t$. In addition, the function $z \to \xi(\bar t, z)$ is increasing for $z \le z_1$, decreasing for $z \in [z_1, z_2]$, increasing for $z \ge z_2$. Let us set $x_- = \xi(\bar t, z_2)$, $x_+ = \xi(\bar t, z_1)$. If $\bar x \notin [x_-, x_+]$, there exists a unique value $\bar z \in \mathbb{R}$ such that $\xi(\bar t, \bar z) = \bar x$; thus $\xi(\cdot, \bar z)$ is the unique minimizer for $(\bar t, \bar x)$ and $(\bar t, \bar x)$ is regular. In addition, $\xi_z(\bar t, \bar z) \neq 0$ and so $(\bar t, \bar x)$ is not conjugate in this case, too. If $\bar x \in [x_-, x_+]$ there is instead more than one value of z such that $\xi(\bar t, z) = \bar x$ and a further analysis is required to determine which value yields a minimizer.

We call z_0 the value to the left of z_1 such that $\xi(\bar t, z_0) = x_-$ and z_3 the value to the right of z_2 such that $\xi(\bar t, z_3) = x_+$. Let us first consider the case $x = x_-$. Then there are two values of z such that $\xi(\bar t, z) = x_-$, namely $z = z_0$ and $z = z_2$. To determine whether $\xi(\cdot, z_0)$ or $\xi(\cdot, z_2)$ is optimal for $(\bar t, x_-)$, let us compare directly the two values of the functional. Setting $\xi_i := \xi(\bar t, z_i)$ we obtain

$$\begin{aligned} J_{\bar t}(\xi_2) - J_{\bar t}(\xi_0) &= u_0(z_2) - u_0(z_0) + \frac{(x_- - z_2)^2}{2\bar t} - \frac{(x_- - z_0)^2}{2\bar t} \\ &= \int_{z_0}^{z_2} \left(u_0'(z) + \frac{z - x_-}{t} \right) dz > 0 \end{aligned} \tag{6.1.16}$$

because, by our definitions, $u_0'(z) + (z - x_-)/t$ vanishes for $z = z_0$, $z = z_2$ and is positive if $z_0 < z < z_2$. It follows that $\xi(\cdot, z_0)$ is the unique optimal trajectory for x_-. By a similar computation, one obtains that $\xi(\cdot, z_3)$ is the unique optimal trajectory for x_+.

Let us consider instead the case when $\bar x \in]x_-, x_+[$. Then there are three values w_1, w_2, w_3 such that $\xi(\bar t, w_i) = \bar x$; we label them in such a way that $w_i \in]z_{i-1}, z_i[$ for $i = 1, 2, 3$. We can observe that the trajectory starting from w_2 is not a local minimizer. In fact

$$\frac{d^2}{dz^2}\bigg|_{z=w_2} \left(u_0(z) + \frac{(\bar x - z)^2}{2\bar t} \right) = u_0''(w_2) + \frac{1}{\bar t} < 0$$

since $z_1 < w_2 < z_2$. Therefore only the trajectories starting from w_1 and w_3 are candidates to be a minimizer.

Let us first observe that the values w_1 and w_3 are smooth functions of $x \in]x_-, x_+[$, both strictly increasing. They satisfy

$$w_1(x) \to z_0, w_3(x) \to z_2 \text{ as } x \to x_-$$

$$w_1(x) \to z_1, w_3(x) \to z_3 \text{ as } x \to x_+$$

Let us set

$$I(x) = u_0(w_3(x)) - u_0(w_1(x)) + \frac{(x - w_3(x))^2}{2\bar t} - \frac{(x - w_1(x))^2}{2\bar t}.$$

Then we have, keeping into account (6.1.16),

$$\lim_{x \to x_-} I(x) > 0, \qquad \lim_{x \to x_+} I(x) < 0.$$

We claim that I is strictly decreasing. Indeed,

$$\begin{aligned} I'(x) &= \frac{w_1(x) - w_3(x)}{\bar{t}} + \left(u_0'(w_3(x)) + \frac{w_3(x) - x}{\bar{t}} \right) w_3'(x) \\ &\quad - \left(u_0'(w_1(x)) + \frac{w_1(x) - x}{\bar{t}} \right) w_1'(x). \end{aligned}$$

Since by definition $u_0'(w_i(x)) + (w_i(x) - x)/\bar{t} = 0$ and $w_1(x) < w_3(x)$ our claim follows. We deduce that $I(x)$ changes sign only once at some value $\hat{x} \in]x_-, x_+[$. By the definition of $I(x)$, we see that the sign of $I(x)$ tells us whether the functional is greater when evaluated at $\xi(\cdot, w_1)$ or at $\xi(\cdot, w_3)$. Therefore, if $\bar{x} \in]x_-, \hat{x}[$ the unique minimizer is the arc $\xi(\cdot, w_1(\bar{x}))$, while if $\bar{x} \in]\hat{x}, x_+[$ the unique minimizer is the arc $\xi(\cdot, w_3(\bar{x}))$. If $x = \hat{x}$ then both arcs are minimizers and so $(\bar{t}, \hat{x})$ is an irregular point.

Observe that x_-, x_+ and $\hat{x}$ depend smoothly on t for $t > T^*$. In addition $x_-(t)$ and $x_+(t)$ both tend to x^* as $t \to T^{*+}$ and so does $\hat{x}(t)$ by comparison. In conclusion, there is one conjugate point, i.e. (T^*, x^*), a smooth curve $x = \hat{x}(t)$ of irregular points starting from the conjugate point (T^*, x^*) and all other points are regular. It is easy to check that the solution is smooth on the set of regular points, is not differentiable at the irregular points, is differentiable but not twice differentiable at the conjugate point.

One could analyze in a similar way the cases when $\mathbb{R}$ can be partitioned in a finite number of intervals such that u_0'' is strictly monotone on every interval. One would find that the solution is smooth outside a finite number of curves of irregular points, each one starting from a conjugate point. Hence the solution is piecewise smooth, and so it is far from being as irregular as a general Lipschitz continuous or semiconcave function can be. In the remainder of this chapter we investigate how much of the behavior of this special example can be found in the general case. ■

We now show how the notions of irregular and conjugate points play a role in optimality conditions.

Theorem 6.1.6 *Assume* (L1)—(L4), *and* (6.1.2). *If ξ^* is a minimizer for problem* (CV) *then $(s, \xi^*(s))$ is regular and not conjugate for any $0 \le s < t$.*

Remark 6.1.7 The above result is classical in calculus of variations (see [51, 46]). We give a proof of it below, because in the textbooks it is usually proved for functionals where both endpoints of the arc are fixed. The statement that along an optimal trajectory there are no conjugate points is known as *Jacobi optimality condition.*

Before giving the proof of Theorem 6.1.6, we introduce another functional. Given an extremal arc $\xi^* \in C^2([0,t])$, we set

$$\begin{aligned}\Lambda^*(s,x,v) &= \langle L_{xx}(s,\xi^*(s),\dot{\xi}^*(s))x,x\rangle \\ &\quad +2\langle L_{xv}(s,\xi^*(s),\dot{\xi}^*(s))x,v\rangle + \langle L_{vv}(s,\xi^*(s),\dot{\xi}^*(s))v,v\rangle, \end{aligned} \tag{6.1.17}$$

$$\Phi^*(v) = \frac{\partial^2 u_0}{\partial v^2}(\xi^*(0)) = \langle D^2 u_0(\xi^*(0))v,v\rangle, \tag{6.1.18}$$

for all $s \in [0,t]$ and $x, v \in \mathbb{R}^n$. Then we define, for $\alpha \in W^{1,1}([0,t])$,

$$J^*(\alpha) = \int_0^t \Lambda^*(s,\alpha(s),\dot{\alpha}(s))\,ds + \Phi^*(\alpha(0)). \tag{6.1.19}$$

Observe that, since L_{vv} is positive definite, this is a functional of the form (6.1.1) . It is easily checked that J^* gives the second variation of J_t at ξ^*, that is

$$\frac{d^2}{d\varepsilon^2}_{|\varepsilon=0} J_t(\xi^* + \varepsilon\alpha) = J^*(\alpha),$$

for any $\alpha \in W^{1,1}([0,t])$ such that $\alpha(t) = 0$. As an immediate consequence we obtain

Proposition 6.1.8 *Let ξ^* be a minimizer for problem* (CV) *and let J^* be the functional associated with ξ^* as in* (6.1.17)–(6.1.19). *Then $J^*(\alpha) \geq 0$ for all $\alpha \in W^{1,1}([0,t])$ such that $\alpha(t) = 0$.*

We are now ready to give the proof of Theorem 6.1.6.

Proof of Theorem 6.1.6 — Given a minimizer ξ^* for the point (t,x), let us first show that $(t_0, \xi^*(t_0))$ is regular for all $t_0 \in]0,t[$. Given $t_0 \in]0,t[$, let ξ_0 be a minimizer for the point $(t_0, \xi^*(t_0))$. Then it is easily seen that the arc

$$\gamma(s) = \begin{cases} \xi_0(s) & 0 \leq s \leq t_0 \\ \xi^*(s) & t_0 < s \leq t \end{cases}$$

is a minimizer for (t,x). Since minimizers are of class C^2, we obtain that $\dot{\xi}^*(t_0) = \dot{\xi}_0(s)$. But then ξ^* and ξ_0 coincide on $[0,t_0]$, since they solve Euler's equation with the same terminal conditions. This shows that there are no minimizers for the point $(t_0, \xi^*(t_0))$ different from ξ^*, and so the point is regular.

To prove the second part of the theorem, let us argue by contradiction and suppose that there exists $t_0 \in]0,t[$ such that the point $(t_0, \xi^*(t_0))$ is conjugate. Let us set for simplicity $x_0 = \xi^*(t_0)$ and $z^* = \xi^*(0)$. By Theorem 6.1.3, ξ^* coincides with $\xi(\cdot, z^*)$. By the first part of the proof, ξ^* is the unique minimizer for (t_0, x_0). Then, by the definition of conjugate point, there exists a nonzero vector θ such that

$$\xi_z(t_0, z^*)\theta = 0.$$

Let us define $\alpha : [0,t] \to \mathbb{R}^n$ as follows:

$$\alpha(s) = \begin{cases} \xi_z(s, z^*)\theta & 0 \leq s \leq t_0 \\ 0 & t_0 < s \leq t. \end{cases}$$

Since α is piecewise smooth and continuous at t_0, we have $\alpha \in W^{1,1}([0,t])$. However, α it is not differentiable at t_0. In fact, its left derivative at t_0 is nonzero, by (6.1.12) and (6.1.15).

Let J^* be the functional associated with ξ^* defined in (6.1.17)–(6.1.19). We claim that $J^*(\alpha) = 0$. If this is true, then Proposition 6.1.8 implies that α is a minimizer for J^*. This is impossible, because any minimizer for J^* is smooth by Theorem 6.1.1(ii), and the contradiction proves that (t_0, x_0) cannot be conjugate.

Therefore, to conclude the proof of the theorem we only need to prove our claim that $J^*(\alpha) = 0$. This is done by a direct computation, as we show below. In the following formulas, to simplify notation, we write L_{xx} instead of $L_{xx}(s, \xi(s,z^*), \dot{\xi}(s,z^*))$, ξ instead of $\xi(s,z^*)$, etc. We obtain

$$\begin{aligned}
J^*(\alpha) &= \int_0^t \Lambda^*(s, \alpha(s), \dot{\alpha}(s))\, ds + \Phi^*(\alpha(0)) \\
&= \int_0^{t_0} \Big(\langle L_{xx}\xi_z\theta, \xi_z\theta\rangle + 2\langle L_{xv}\xi_z\theta, \dot{\xi}_z\theta\rangle + \langle L_{vv}\dot{\xi}_z\theta, \dot{\xi}_z\theta\rangle \Big)\, ds \\
&\quad + \langle D^2u_0(z^*)\theta, \theta\rangle \\
&= \int_0^{t_0} \langle \partial_z \big|_{z=z^*} L_x(s, \xi(s,z), \dot{\xi}(s,z))\, \theta, \xi_z\theta\rangle ds \\
&\quad + \int_0^{t_0} \langle \partial_z \big|_{z=z^*} L_v(s, \xi(s,z), \dot{\xi}(s,z))\, \theta, \dot{\xi}_z\theta\rangle ds + \langle D^2u_0(z^*)\theta, \theta\rangle \\
&= \int_0^{t_0} \langle \partial_z \big|_{z=z^*} (L_x - \dot{L}_v)\theta, \xi_z\theta\rangle ds + \langle \partial_z \,|_{z=z^*} (L_v)\theta, \xi_z\theta\rangle|_{s=0}^{s=t_0} \\
&\quad + \langle D^2u_0(z^*)\theta, \theta\rangle \\
&= -\langle (L_{vx}\xi_z + L_{vv}\dot{\xi}_z)\theta, \xi_z\theta\rangle|_{s=0} + \langle D^2u_0(z^*)\theta, \theta\rangle \\
&= \langle (-(L_{vx} + L_{vv}\dot{\xi}_z)|_{s=0} + D^2u_0(z^*))\theta, \theta\rangle.
\end{aligned}$$

Here we have used the fact that $L_x - \dot{L}_v \equiv 0$ along ξ since ξ satisfies Euler's equations, the property that $\xi_z = I$ for $s = 0$ and the assumption

$\xi_z(s, z^*)\theta = 0$ for $s = t_0$. Now, if we differentiate with respect to z the transversality condition $L_v(0, z, \dot{\xi}(0, z)) = Du_0(z)$ we obtain

$$(L_{vx} + L_{vv}\dot{\xi}_z)|_{s=0} = D^2u_0(z),$$

and so the right–hand side in the previous formula vanishes, proving our claim. ■

6.2 The value function

As in Chapter 1, we develop further the study of problem (CV) looking at the *value function* associated to the problem. We define

$$u(t, x) = \min_{\xi(t)=x} J_t(\xi). \tag{6.2.20}$$

We recall that u satisfies the dynamic programming principle (Theorem 1.1.2). We begin our analysis by proving the semiconcavity of u.

Theorem 6.2.1 *Let L satisfy properties* (L1)—(L4) *and let u_0 be bounded and belong to the class* $\mathrm{SCL}(\mathbb{R}^n)$ *of the functions which are locally semiconcave with linear modulus. Then the value function u defined in* (6.2.20) *belongs to* $\mathrm{SCL}(\overline{Q}_T)$.

It is convenient to prove first the semiconcavity inequality in a special case, as in the next lemma.

Lemma 6.2.2 *Under the hypotheses of the previous theorem, for any $\rho > 0$ there exists C such that*

$$u(2h, x + z) + u(0, x - z) - 2u(h, x) \leq C(h^2 + |z|^2)$$

for all $x, z \in B_\rho$ and $h \in [0, \rho]$.

Proof — Let x, z, h be as in the statement of the lemma and let $\xi : [0, t] \to \mathbb{R}^n$ be an optimal trajectory for the point (h, x). We define an admissible trajectory ξ_0 for $(2h, x + z)$ in the following way:

$$\xi_0(s) = z - x + 2\xi\left(\frac{s}{2}\right), \qquad s \in [0, 2h].$$

It is immediately checked that $\xi_0(2h) = x + z$. It follows that

$$\begin{aligned}
&u(2h, x+z) + u(0, x-z) - 2u(h, x) \\
&\quad\le \int_0^{2h} L(s, \xi_0(s), \dot\xi_0(s))\, ds - 2\int_0^h L(s, \xi(s), \dot\xi(s))\, ds \\
&\qquad + u_0(x-z) + u_0(\xi_0(0)) - 2u_0(\xi(0)) \\
&\quad= 2\int_0^h \{L(2s, z - x + 2\xi(s), \dot\xi(s)) - L(s, \xi(s), \dot\xi(s))\}\, ds \\
&\qquad + u_0(x-z) + u_0(2\xi(0) - x + z) - 2u_0(\xi(0)).
\end{aligned}$$

By Theorem 6.1.1(iii), $|\xi|$ and $|\dot\xi|$ are bounded by some constant M depending only on ρ. Thus we obtain

$$|\xi(s) - x + z| = |\xi(s) - \xi(h) + z| \le Mh + |z|, \qquad s \in [0, h].$$

Hence, using the Lipschitz continuity of L with respect to the first two arguments and the semiconcavity of u_0 we obtain

$$\begin{aligned}
&u(2h, x+z) + u(0, x-z) - 2u(h, x) \\
&\quad\le 2C_1 \int_0^h (s + |\xi(s) - x + z|)\, ds + C_2|\xi(0) - x + z|^2 \\
&\quad\le 2C_1 h(h + Mh + |z|) + C_2(Mh + |z|)^2. \quad \blacksquare
\end{aligned}$$

Proof of Theorem 6.2.1 — For fixed $r > 0$, we take any $x, z \in \mathbb{R}^n$, $t, h \ge 0$ such that $x \pm z \in B_r$ and that $t \pm h \in [0, r]$. Let $\xi : [0, t] \to \mathbb{R}^n$ an optimal trajectory for the point (t, x). Let us define

$$\xi_-(s) = -z + \xi(s+h), \qquad s \in [0, t-h],$$

$$\xi_+(s) = z + \xi(s-h), \qquad s \in [2h, t+h].$$

Since $\xi_\pm(t \pm h) = x \pm z$, we obtain from the dynamic programming principle

$$\begin{aligned}
&u(t+h, x+z) + u(t-h, x-z) - 2u(t, x) \\
&\quad\le \int_{2h}^{t+h} L(s, \xi_+(s), \dot\xi_+(s))\, ds + u(2h, \xi_+(2h)) \\
&\qquad + \int_0^{t-h} L(s, \xi_-(s), \dot\xi_-(s))\, ds + u(0, \xi_-(0)) \\
&\qquad - 2\int_h^t L(s, \xi(s), \dot\xi(s))\, ds - 2u(h, \xi(h)) \\
&\quad= u(2h, \xi(h) + z) + u(0, \xi(h) - z) - 2u(h, \xi(h)) \\
&\qquad + \int_h^t \{L(s-h, \xi(s) - z, \dot\xi(s)) + L(s+h, \xi(s) + z, \dot\xi(s)) \\
&\qquad\quad - 2L(s, \xi(s), \dot\xi(s)\}\, ds.
\end{aligned}$$

By Theorem 6.1.1(iii), $|\xi(s)|$ and $|\dot{\xi}(s)|$ are uniformly bounded by some constant depending only on r. We can therefore use the semiconcavity of L in the first two arguments and the estimate given by Lemma 6.2.2 to obtain, for suitable constants C, C_1,

$$\begin{aligned} &u(t+h, x+z) + u(t-h, x-z) - 2u(t,x) \\ &\leq\; C(h^2+|z|^2) + \int_h^t C_1(h^2+|z|^2)\,ds \\ &\leq\; C(h^2+|z|^2) + rC_1(h^2+|z|^2). \;\blacksquare \end{aligned}$$

In the previous proof it was tacitly assumed that $T=+\infty$. For $T<+\infty$ one should consider sets of the form $[0,T]\times B_r$ with $r>0$ arbitrary, rather than the sets $[0,r]\times B_r$ considered above, and keep the remainder of the proof as is. Since such minor modifications can be easily done by the reader, we will restrict ourselves to the case $T=+\infty$ in the following proofs as well.

Theorem 6.2.3 *Let properties* (L1)—(L4) *hold and let u_0 be bounded and lower semicontinuous on* $\mathbb{R}^n$. *Then, for any* $t\in\,]0,T[$, $u(\cdot,t)$ *belongs to* $\mathrm{SCL}(\mathbb{R}^n)$.

Proof — For fixed $t, r>0$, let us consider $x, z\in\mathbb{R}^n$ such that $x\pm z\in B_r$ and let ξ be an optimal arc for the point (t,x). Define

$$\xi_\pm(s) = \xi(s) \pm \frac{s}{t}z.$$

Then we have, using the semiconcavity of L in the second and third argument,

$$\begin{aligned} &u(t,x+z)+u(t,x-z)-2u(t,x) \\ &\leq \int_0^t L(s,\xi_+(s),\dot{\xi}_+(s))\,ds + u_0(\xi_+(0)) \\ &\quad + \int_0^t L(s,\xi_-(s),\dot{\xi}_-(s))\,ds + u_0(\xi_-(0)) \\ &\quad -2\int_0^t L(s,\xi(s),\dot{\xi}(s))\,ds - 2u_0(\xi(0)) \\ &= \int_0^t \left\{ L\left(s,\xi(s)+\frac{s}{t}z,\dot{\xi}(s)+\frac{z}{t}\right) + L\left(s,\xi(s)-\frac{s}{t}z,\dot{\xi}(s)-\frac{z}{t}\right) \right. \\ &\qquad \left. -\,2L(s,\xi(s),\dot{\xi}(s)) \right\} ds \\ &\leq \int_0^t C\frac{s^2+1}{t^2}|z|^2\,ds = C\left(\frac{t}{3}+\frac{1}{t}\right)|z|^2. \;\blacksquare \end{aligned}$$

Corollary 6.2.4 *Assume* (L1)—(L4) *and let* u_0 *be bounded and lower semicontinuous on* $\mathbb{R}^n$. *Then* $u \in \mathrm{SCL}\,(Q_T)$.

Proof — Let us fix $t' \in \,]0, T[\,$. Using Theorem 6.2.3 and formula (1.1.5), we see that the restriction of u to $[t', T[\, \times \mathbb{R}^n$ can be represented as the value function of a problem with a semiconcave initial cost. Therefore it is semiconcave by Theorem 6.2.1. By the arbitrariness of t' we obtain the assertion. ■

Comparing the previous corollary with Theorem 6.2.1, we observe that the semiconcavity requirement on the initial u_0 has been dropped. Therefore, the value function u not only inherits the semiconcavity of the initial cost u_0, but is semiconcave even if u_0 is not. In the latter case, of course, the semiconcavity of u holds only in the open set Q_T which does not contain the plane $t = 0$.

Theorem 6.2.5 *The value function u defined in* (6.2.20) *is the unique viscosity solution of the* Dynamic Programming equation

$$\begin{cases} \partial_t u(t,x) + H(t,x,\nabla u(t,x)) = 0 & \text{in } Q_T \\ u(0,x) = u_0(x) & \text{in } \mathbb{R}^n\ . \end{cases} \tag{6.2.21}$$

Proof — Let us take $(t,x) \in Q_T$ and $(p_t, p_x) \in D^+u(t,x)$. Then we have, for any $v \in \mathbb{R}^n$,

$$\limsup_{h \to 0^+} \frac{u(t-h, x-hv) - u(t,x) + h(p_t + v \cdot p_x)}{h\sqrt{1+|v|^2}} \le 0$$

which is equivalent to

$$\limsup_{h \to 0^+} \frac{u(t-h, x-hv) - u(t,x)}{h} \le -p_t - v \cdot p_x.$$

Now let us set $\zeta(\sigma) = x + (\sigma - t)v$ for $\sigma \le t$. We obtain from the dynamic programming principle

$$\begin{aligned} u(t,x) &\le u(t-h, \zeta(t-h)) + \int_{t-h}^{t} L(\sigma, \zeta(\sigma), \dot\zeta(\sigma))d\sigma \\ &= u(t-h, x-hv) + \int_{t-h}^{t} L(\sigma, x + (\sigma - t)v, v)d\sigma, \end{aligned}$$

which yields

$$\begin{aligned} &\limsup_{h \to 0^+} \frac{u(t-h, x-hv) - u(t,x)}{h} \\ &\ge \lim_{h \to 0^+} -\frac{1}{h}\int_{t-h}^{t} L(\sigma, x + (\sigma - t)v, v)d\sigma = -L(t,x,v). \end{aligned}$$

We conclude that

$$p_t + v \cdot p_x - L(t,x,v) \le 0, \qquad \forall\, v \in \mathbb{R}^n,$$

and so $p_t + H(t,x,p_x) \le 0$, which proves that u is a viscosity subsolution of the equation.

To prove that u is a supersolution, let us take $(p_t,p_x) \in D^-u(t,x)$. Let $\xi \in C^2([0,T])$ be an optimal trajectory for (t,x) and let us set $w = \dot{\xi}(t)$. Then we have, using the Lipschitz continuity of u,

$$\begin{aligned} &\liminf_{h\to 0^+} \frac{u(t-h,\xi(t-h)) - u(t,x)}{h} \\ = \ &\liminf_{h\to 0^+} \frac{u(t-h,x-hw) - u(t,x)}{h} \ge -p_t - w \cdot p_x. \end{aligned}$$

On the other hand, by the optimality of ξ,

$$u(t,x) = u(t-h,\xi(t-h)) + \int_{t-h}^{t} L(s,\xi(s),\dot{\xi}(s))\,ds,$$

for all $0 \le h \le t$, which implies

$$\lim_{h\to 0^+} \frac{u(t-h,\xi(t-h)) - u(t,x)}{h} = -L(t,x,w).$$

We conclude that $p_t + w \cdot p_x - L(t,x,w) \ge 0$, which implies that $p_t + H(t,x,p_x) \ge 0$. This shows that u is a viscosity supersolution as well. By definition, u satisfies the initial condition and, by the uniqueness statement in Theorem 5.2.10, is the unique viscosity solution of our Cauchy problem. ∎

Theorem 6.2.6 *Let $(t,x) \in Q_T$ be given, and let ξ be an optimal trajectory for (t,x). Then u is differentiable at $(s,\xi(s))$ for all $s \in]0,t[$.*

Proof — Let $s \in]0,t[$ be given, and let us set for simplicity $y = \xi(s)$. We know from the previous theorem that the elements of $D^+u(s,y)$ satisfy the Hamilton–Jacobi equation as an inequality. We claim that the fact that (s,y) is not an endpoint of the optimal trajectory ξ implies that the converse inequality holds as well. In fact we have, by the dynamic programming principle,

$$u(s+h,\xi(s+h)) = u(s,y) + \int_{s}^{s+h} L(\sigma,\xi(\sigma),\dot{\xi}(\sigma))d\sigma$$

for all $0 \leq h \leq t - s$. Therefore, setting $w = \dot{\xi}(s)$, we obtain

$$\lim_{h \to 0^+} \frac{u(s+h, \xi(s+h)) - u(s,y)}{h} = L(s,y,w).$$

Arguing as in the previous proof, we obtain that $p_t + H(s,y,p_x) \geq 0$ for any $(p_t, p_x) \in D^+u(s,y)$. Since we already know that the converse inequality holds, we conclude that

$$p_t + H(s, \xi(s), p_x) = 0, \qquad s \in]0,t[\,, (p_t,p_x) \in D^+u(s,\xi(s)). \tag{6.2.22}$$

Since H is strictly convex in the third argument and $D^+u(s,\xi(s))$ is a convex set, we deduce that $D^+u(s,\xi(s))$ consists of a single element. Recalling that u is semiconcave, we obtain that u differentiable at $(s,\xi(s))$, by Proposition 3.3.4-(d) ∎

Theorem 6.2.7 *Let $(t,x) \in Q_T$ be given, and let ξ be an optimal trajectory for (t,x). Let η be the arc associated with ξ as in Theorem 6.1.3. Then we have*

$$\eta(t) \in \nabla^+ u(t,x); \tag{6.2.23}$$

$$\eta(s) = \nabla u(s, \xi(s)), \qquad s \in]0,t[\,. \tag{6.2.24}$$

Proof — To prove (6.2.23), it suffices to show that, for any nonzero vector $v \in \mathbb{R}^n$, we have

$$\limsup_{h \to 0} \frac{u(t, x+hv) - u(t,x) - hp(t) \cdot v}{h} \leq 0. \tag{6.2.25}$$

Let us set $x_0 = \xi(0)$. For fixed h, the arc $s \to \xi(s) + hv$ is admissible for our functional with endpoint $(t, x+hv)$ and therefore,

$$u(t,x+hv) \leq u_0(x_0 + hv) + \int_0^t L(s, \xi(s) + hv, \dot{\xi}(s))\, ds.$$

Recalling also (6.1.8), it follows that

$$\begin{aligned}
&\limsup_{h \to 0} \frac{u(t,x+hv) - u(t,x)}{h} \\
&\leq \lim_{h \to 0} \frac{u_0(x_0+hv) - u_0(x_0)}{h} \\
&\quad + \lim_{h \to 0} \int_0^t \frac{L(s,\xi(s)+hv,\dot{\xi}(s)) - L(s,\xi(s),\dot{\xi}(s))}{h}\, ds \\
&= v \cdot \left(Du_0(x_0) + \int_0^t L_x(s,\xi(s),\dot{\xi}(s))\, ds \right) = v \cdot \eta(t),
\end{aligned}$$

which proves (6.2.25).

If we now consider $s \in]0,t[$ we have that the trajectory ξ is optimal for the point $(s,\xi(s))$. By the first part of the proof we deduce that $\eta(s) \in \nabla^+ u(s,\xi(s))$. But u is differentiable at $(s,\xi(s))$ by Theorem 6.2.6 and so (6.2.24) follows. ∎

Theorem 6.2.8 *For any $(t,x) \in Q_T$ the map that associates with any $(p_t,p_x) \in D^*u(t,x)$ the arc ξ obtained by solving problem* (6.1.9) *with the terminal conditions*

$$\begin{cases} \xi(t) = x \\ \eta(t) = p_x \end{cases} \tag{6.2.26}$$

provides a one–to–one correspondence between $D^*u(t,x)$ and the set of minimizers of (CV) *with endpoint (t,x).*

Proof — Let us first prove the conclusion when (t,x) is a point of differentiability for u. In this case the set $D^+u(t,x)$ contains only the gradient $(u_t(t,x),\nabla u(t,x))$. Since u is semiconcave, the same is true for $D^*u(t,x)$. We must therefore show that problem (CV) has a unique minimizer, obtained by solving problem (6.1.9) with final conditions

$$\xi(t) = x, \qquad \eta(t) = \nabla u(t,x). \tag{6.2.27}$$

To see this, let us suppose that ξ^* is a minimizer, and let η^* be the associated arc defined as in Theorem 6.1.3. Then the pair (ξ^*,η^*) solves system (6.1.9). In addition, by Theorem 6.2.7 and by the differentiability of u at (t,x), the pair (ξ^*,η^*) also satisfies conditions (6.2.27). This proves the theorem in this case.

Let us now consider the case of a general (t,x). Let (p_t,p_x) be any point in $D^*u(t,x)$. Then there exists a sequence $\{(t_n,x_n)\}$ of points where u is differentiable which converges to (t,x) and is such that $(p_t,p_x) = \lim_{n\to\infty} Du(t_n,x_n)$. By Theorem 6.2.5, we have that

$$\partial_t u(t_n,x_n) + H(t_n,x_n,\nabla u(t_n,x_n)) = 0,$$

which implies, by letting $n \to \infty$, that $p_t + H(t,x,p_x) = 0$. For fixed n, by the first part of the proof, there exists a unique minimizer ξ_n for the point (t_n,x_n), obtained by solving system (6.1.9) with the conditions $\xi_n(t_n) = x_n$, $\eta_n(t_n) = \nabla u(t_n,x_n)$. Then $\xi_n \to \xi$ as $n \to \infty$, where ξ is the first component of the solution of problem (6.1.9)–(6.2.26). It follows, by the continuity of u and the optimality of ξ_n,

$$\begin{aligned} u(t,x) &= \lim_{n\to\infty} u(t_n,x_n) = \lim_{n\to\infty}\left(\int_0^{t_n} L(s,\xi_n(s),\dot\xi_n(s))\,ds + u_0(\xi_n(0))\right) \\ &= \int_0^t L(s,\xi(s),\dot\xi(s))\,ds + u_0(\xi(0)), \end{aligned}$$

and so the arc ξ is a minimizer for the point (t,x). This proves that the correspondence $(p_t,p_x) \to \xi$ is indeed a map from $D^*u(t,x)$ to the set of minimizers for (t,x).

Next, we prove the injectivity of our map. Let (p_t^1,p_x^1) and (p_t^2,p_x^2) be two distinct elements of $D^*u(t,x)$. As observed before, $p_t^i = -H(t,x,p_x^i)$, and so p_x^1 and p_x^2 must be distinct. Then the corresponding solutions (ξ_1,η_1) and (ξ_2,η_2) of problem (6.1.9)–(6.2.26) are also distinct. This implies that $\xi_1 \neq \xi_2$, since otherwise η_1 and η_2 would also coincide, as a consequence of the definition of η (see Theorem 6.1.3).

Finally, we prove surjectivity. Let ξ^* be any minimizer for (t,x) and let η^* be the associated dual arc. Then the pair (ξ^*,η^*) solves system (6.1.9). In addition, by Theorems 6.2.5 and 6.2.7, u is differentiable at the points of the form $(s,\xi^*(s))$ with $0<s<t$ and satisfies

$$\begin{aligned}&\nabla u(s,\xi^*(s)) = \eta^*(s),\\ &\partial_t u(s,\xi^*(s)) = -H(s,\xi^*(s),\nabla u(s,\xi^*(s))) = -H(s,\xi^*(s),\eta^*(s)).\end{aligned}$$

Therefore, setting

$$\begin{aligned}p_x &= \eta^*(t) = \lim\nolimits_{s\to t} \nabla u(s,\xi^*(s)),\\ p_t &= -H(t,x,p_x) = \lim\nolimits_{s\to t} \partial_t u(s,\xi^*(s)),\end{aligned}$$

we have that $(p_t,p_x) \in D^*u(t,x)$. This shows that ξ^* belongs to the range of the map defined in the statement of the theorem. ∎

In particular, we can relate the differentiability of u to the uniqueness of minimizers for problem (CV).

Corollary 6.2.9 *A point (t,x) is regular if and only if u is differentiable at x.*

Let us denote by Σ the set of irregular points (equivalently, the set of points where u is not differentiable) and by Γ the set of conjugate points.

Theorem 6.2.10 *Assume* (L1)—(L4), *and* (6.1.2). *Then $\Sigma \cup \Gamma$ is a closed set and $u \in C^{R+1}(Q_T \setminus (\Sigma \cup \Gamma))$.*

Proof — Let us consider a point $(\bar t,\bar x)$ in the complement of $\Sigma \cup \Gamma$. By Corollary 6.2.9 there exists a unique minimizer for problem (CV) with endpoint $(\bar t,\bar x)$, which has the form $\xi(\cdot,\bar z)$ for some $\bar z \in \mathbb{R}^n$; in addition $\det \xi_z(\bar t,\bar z) \neq 0$. Then we can find a neighborhood W of $(\bar t,\bar z)$ such that the map $(t,z) \to (t,\xi(t,z))$ is a C^R diffeomorphism.

We claim now that there exists a neighborhood N of $(\bar{t},\bar{x})$ with the following property: if $(t,x) \in N$ and ξ^* is a minimizer for (t,x), then $\xi^* = \xi(\cdot,z)$ for some z such that $(t,z) \in W$.

We prove the claim by contradiction. If no such neighborhood exists, we can find a sequence $\{(t_n,x_n)\}$ converging to (t,x) with the following property: each (t_n,x_n) has a minimizer ξ_n of the form $\xi_n = \xi(\cdot,z_n)$ with $(t_n,z_n) \notin W$. Now let η_n be the dual arc associated with ξ_n. By Theorem 6.2.7, $\eta_n(t_n) \in \nabla^+ u(t_n,x_n)$; since $(t_n,x_n) \to (\bar{t},\bar{x})$ and u is differentiable at $(\bar{t},\bar{x})$, we have that $\eta_n(t_n) \to \nabla u(\bar{t},\bar{x})$. By Theorem 6.2.8, $\xi_n \to \xi$, where ξ is obtained by solving system (6.1.9) with terminal conditions $\xi(\bar{t}) = \bar{x}, \eta(\bar{t}) = D_x u(\bar{t},\bar{x})$. But then ξ is the unique minimizer for $(\bar{t},\bar{x})$ and coincides with $\xi(\cdot,\bar{z})$. In particular, this implies that $z_n = \xi_n(0) \to \xi(0) = \bar{z}$. But this contradicts the assumption that $(z_n,t_n) \notin W$ for all n.

From the claim we deduce that, for $(t,x) \in N$, we have

$$u(t,x) = u_0(z) + \int_0^t L(s,\xi(s,z),\dot{\xi}(s,z))\,ds,$$

where z is uniquely determined by the conditions $(t,z) \in W$, $x = \xi(t,z)$. This shows that u is of class C^R in W and that all points of W are regular and not conjugate. In addition, by Theorem 6.2.7 we have $\nabla u(t,x) = \eta(t,z)$ and $\partial_t u(t,x) = -H(t,x,\nabla u(t,x))$; therefore the derivatives of u are also of class C^R and $u \in C^{R+1}(N)$. ■

We conclude the section giving a characterization of the extreme points of D^+u (see Definitions 3.3.9 and 3.3.13).

Theorem 6.2.11 *For any $(t,x) \in Q_T$ we have*

$$\begin{aligned} \mathrm{E}u(t,x) &= \mathrm{Ext}(D^+u(t,x)) = D^*u(t,x) \\ &= \{(q,p) \in D^+u(t,x) \mid p_t + H(t,x,p_x) = 0\}. \end{aligned}$$

Proof — The inclusion $\mathrm{E}u(t,x) \subset \mathrm{Ext}(D^+u(t,x))$ follows from the definition, while the inclusions $\mathrm{E}u(t,x) \subset D^*u(t,x)$ and $D^*u(t,x) \subset D^+u(t,x)$ hold for a general semiconcave function (see (3.3.23) and Proposition 3.3.4-(b)). The fact that any $(p_t,p_x) \in D^*u(t,x)$ satisfies $p_t + H(t,x,p_x) = 0$ has been already observed during the proof of Theorem 6.2.8. To complete the proof we only need to show that the any $(p_t,p_x) \in D^+u(t,x)$ which satisfies $p_t + H(t,x,p_x) = 0$ is an exposed vector for $D^+u(t,x)$. This is a consequence of the strict convexity of H. If we set $h = H_p(t,x,p_x)$ we have in fact

$$H(t,x,q_x) > H(t,x,p_x) + h \cdot (q_x - p_x), \qquad q_x \in \mathbb{R}^n, q_x \neq p_x.$$

On the other hand, since u is a viscosity solution, we have

$$q_t + H(t,x,q_x) \le 0, \qquad (q_t,q_x) \in D^+u(t,x).$$

Therefore

$$-p_t - h\cdot p_x < -q_t - h\cdot q_x, \qquad (q_t,q_x)\in D^+u(t,x), q_x \neq p_x,$$

and therefore (p_t,p_x) is an exposed vector with respect to the direction $(-1,-h)$. ∎

6.3 The singular set of u

In this section we begin the study of the singular set of the value function u of problem (CV). We assume that conditions (L1)—(L4) and (6.1.2) are satisfied throughout the section. Equivalently, we may think of u as the unique viscosity solution of problem (6.2.21), assuming that $H \in C^{R+1}$ and that H_{pp} is positive definite. For brevity, we omit recalling these hypotheses in the statements below.

Before starting our analysis, let us briefly review some of the results of Dafermos [39]. He considers the Cauchy problem for the one-dimensional scalar conservation law

$$\begin{cases} \partial_t v(t,x) + \partial_x f(t,x,v(t,x)) = 0 & (t,x)\in \mathbb{R}_+ \times \mathbb{R} \\ v(0,x) = v_0(x) & x \in \mathbb{R}, \end{cases} \tag{6.3.28}$$

where f, v_0 satisfy suitable regularity assumptions and f is strictly convex in the third variable. We have considered this equation in Section 1.6 in the case where $f = f(v)$. We have seen that, if $f = H$, then $u(t,x)$ is a semiconcave solution of (6.2.21) if and only if $v(t,x) = \partial_x u(t,x)$ is an entropy solution of (6.3.28) with initial data $v_0(x) = u_0'(x)$. It can be proved that the same equivalence holds also when f depends on t,x, and so in the one dimensional case our equation belongs to the class considered by Dafermos.

The results of [39], rephrased according to our terminology, are the following. The set $\Sigma \cup \Gamma$ is the union of an at most countable family of arcs of the form $\{(t,x)\,|\,t \ge t_i, x = \gamma_i(t)\}$, with γ_i Lipschitz and $t_i > 0$. All points of the form $(t,\gamma_i(t))$ with $t > t_i$ are irregular. The initial point $(t_i,\gamma(t_i))$ is called a *shock generation point*. It is a conjugate point, and may be either regular or irregular. In the latter case it is called the center of a *compression wave.* The function γ_i may be characterized as the unique solution of the differential inclusion

$$\begin{cases} \gamma'(t) \in [H_p(t,\gamma(t),u_x(t,\gamma(t)+)),H_p(t,\gamma(t),u_x(t,\gamma(t)-))], & t \geq t_0 \\ \gamma(t_0) = \gamma_i(t_0). \end{cases}$$

Here we have denoted by $u_x(t,x\pm)$ the right and left limit of u_x in the x variable. It follows from the semiconcavity of u that for any $(t,x) \in Q$ these limits exist and satisfy $H_p(t,x,u_x(t,x+)) \leq H_p(t,x,u_x(t,x-))$. Observe that the arcs γ_i solutions of the above differential inclusion are generalized characteristics according to our definition (5.5.54).

The above results require the special structure of the one–dimensional case. We will see, however, that some of them possess a multidimensional analogue, although proofs must be based on a different method.

To begin with, we study the structure of the set $\Sigma \cup \Gamma$. The following lemma states that any point (t_0, x_0) which is singular or conjugate is the cluster point of a family of singular points, propagating forward in time. The interesting case is when (t_0, x_0) is conjugate, since when it is singular the result is a consequence of Theorem 5.6.7.

Lemma 6.3.1 *Let $(t_0,x_0) \in \Sigma \cup \Gamma$ be given. Then there exist $M > 0$, $h > 0$ with the following property: given any $\varepsilon \in (0,h]$ there exists $x_\varepsilon \in \overline{B_{\varepsilon M}(x_0)}$ such that $(t_0+\varepsilon, x_\varepsilon) \in \Sigma$.*

Proof — By Theorem 6.1.1(iii) there exists a constant M such that, if $(t,x) \in [t_0,t_0+1] \times B_1(x_0)$ and X^* is a minimizer for problem (CV), then $|\dot{X}^*(s)| < M$ for any s. We then define $h = \min\{1, M^{-1}\}$.

We claim that M and h have the desired property. Let us suppose, on the contrary, that there exists $\varepsilon \in (0,h]$ such that all points of the form $(t_0+\varepsilon, x)$ with $x \in \overline{B_{\varepsilon M}(x_0)}$ are regular. Then, by Corollary 6.2.9 and Proposition 3.3.4-(e), $Du(t_0+\varepsilon,x)$ exists for any $x \in \overline{B_{\varepsilon M}(x_0)}$ and depends continuously on x.

Let us consider the hamiltonian system (6.1.9) with final conditions (6.2.26) and denote by $\xi(s;t,x,p)$ the first component of the solution. By Theorem 6.2.8, for any $x \in B_{\varepsilon M}(x_0)$ the arc $\xi(\cdot;t_0+\varepsilon,x,\nabla u(t_0+\varepsilon,x))$ is the unique minimizer for problem (CV) with endpoint $(t_0+\varepsilon,x)$. We now define a map $\Lambda : \overline{B_{\varepsilon M}(x_0)} \to \mathbb{R}^n$ in the following way:

$$\Lambda(x) = x_0 + x - \xi(t_0; t_0+\varepsilon, x, \nabla u(t_0+\varepsilon,x)).$$

The previous remarks show that Λ is continuous. In addition,

$$\begin{aligned} &|x - \xi(t_0;t_0+\varepsilon,x,\nabla u(t_0+\varepsilon,x))| \\ &\quad = \ |\xi(t_0+\varepsilon;t_0+\varepsilon,x,\nabla u(t_0+\varepsilon,x)) - \xi(t_0;t_0+\varepsilon,x,\nabla u(t_0+\varepsilon,x))| \\ &\quad \leq \ \varepsilon \sup|\dot{\xi}| \leq \varepsilon M \end{aligned}$$

and so Λ maps $\overline{B_{\varepsilon M}(x_0)}$ into itself. Hence, by Brouwer's theorem, $x_\varepsilon \in \overline{B_{\varepsilon M}(x_0)}$ exists such that $\Lambda(x_\varepsilon) = x_\varepsilon$. Recalling the definition of Λ, we conclude that the minimizer for $(t_0+\varepsilon, x_\varepsilon)$ passes through the point (t_0, x_0). This fact contradicts Theorem 6.1.6, since we are assuming that (t_0, x_0) is irregular or conjugate. ∎

From the above lemma we recover the converse of the inclusion $\overline{\Sigma} \subset \Sigma \cup \Gamma$ (see Theorem 6.2.10) and obtain the following

Corollary 6.3.2 $\overline{\Sigma} = \Sigma \cup \Gamma$.

We now give a sufficient condition in terms of reachable gradients to detect conjugate points.

Theorem 6.3.3 *If $D^*u(t,x)$ is an infinite set, then $(t,x) \in \Gamma$.*

Proof — Let us assume that (t,x) is not conjugate. Then, if we set

$$Z = \{z \in \mathbb{R}^n \,|\, \xi(\cdot, z) \text{ is a minimizer for } (t,x)\},$$

we have that the points of Z are isolated. In fact, if $z \in Z$, then $\xi_z(t,z)$ is nonsingular, and so $X(t,z') \neq X(t,z) = x$ for z' in a neighborhood of z. Moreover, by Theorem 6.1.1(iii), Z is bounded. We deduce that Z is finite.

On the other hand, there are infinitely many minimizers of (CV) with endpoint (t,x), by Theorem 6.2.10. In addition, Theorem 6.1.3 implies that two different minimizers have different initial points. Thus Z is infinite. We have reached a contradiction. ∎

The next result says, roughly speaking, that the second derivative of u is infinite at a conjugate point.

Theorem 6.3.4 *Let $(t_0, x_0) \in \Gamma$ be given, and let $\xi(\cdot, z_0)$ be a minimizer for (t_0, x_0) such that* $\det \xi_z(t_0, z_0) = 0$. *Then*

$$\lim_{t \to t_0^-} \|\nabla^2 u(t, \xi(t, z_0))\| = +\infty. \tag{6.3.29}$$

If in addition (t_0, x_0) is regular and $\{x_k\}$, $\{p_k\}$ satisfy

$$\begin{cases} \lim\limits_{k\to\infty} x_k = x_0, \\[2ex] \lim\limits_{k\to\infty} \dfrac{x_k - x_0}{|x_k - x_0|} = v \textit{ for some } v \notin \mathit{Im}\, \xi_z(t_0, z_0), \\[2ex] p_k \in \nabla^* u(t_0, x_k) \end{cases} \tag{6.3.30}$$

then

$$\lim_{k\to\infty} \frac{|p_k - \nabla u(t_0, x_0)|}{|x_k - x_0|} = \infty.$$

Proof — Let $\theta \in (\mathbb{R}^n \setminus \{0\})$ be given such that $\xi_z(t_0, z_0)\theta = 0$. Then $\eta_z(t_0, z_0)\theta \neq 0$ by (6.1.14). Given $t < t_0$, the point $(t, \xi(t, z_0))$ is regular and not conjugate by Theorem 6.1.6, and so are the points in a neighborhood, by Theorem 6.2.10. Then we have, by Theorem 6.2.8

$$\nabla u(t, \xi(t, z)) = \eta(t, z)$$

for z in a neighborhood of z_0. It follows

$$\nabla^2 u(t, \xi(t, z_0))\xi_z(t, z_0) = \eta_z(t, z_0), \qquad t < t_0.$$

Since $\eta_z(t, z_0)\theta \neq 0$ we conclude that

$$\|D^2_{xx} u(t, \xi(t, z_0))\| \geq \frac{|\eta_z(t, z_0)\theta|}{|\xi_z(t, z_0)\theta|} \to \infty \text{ as } t \to t_0^-.$$

Now suppose (t_0, x_0) is regular. Then $\xi(\cdot, z_0)$ is the unique minimizer for (t_0, x_0) and we have $\nabla u(t_0, x_0) = \eta(z_0, t_0)$ by Theorem 6.2.8. Let $\{x_k\}$, $\{p_k\}$ satisfy assumptions (6.3.30). Again by Theorem 6.2.8 there exists a sequence $\{z_k\}$ such that $x_k = \xi(t_0, z_k)$, $p_k = \eta(t_0, z_k)$. The sequence $\{z_k\}$ is bounded by Theorem 6.1.1(iii); moreover, any cluster point $\bar{z}$ of $\{z_k\}$ is such that the trajectory $\xi(\cdot, \bar{z})$ is a minimizer for (t_0, x_0). Since such a minimizer is unique, we deduce that $\lim z_k = z_0$.

By passing to a subsequence we may assume

$$\lim_{k\to\infty} \frac{z_k - z_0}{|z_k - z_0|} = \nu$$

for some unit vector ν. Since

$$x_k = x_0 + \xi_z(t_0, z_0)(z_k - z_0) + o(|z_k - z_0|)$$

we deduce

$$\lim_{k\to\infty} \frac{x_k - x_0}{|z_k - z_0|} = \xi_z(t_0, z_0)\nu.$$

If $\xi_z(t_0, z_0)\nu \neq 0$ then the above relation implies that $v \in \operatorname{Im} \xi_z(t_0, z_0)$, in contradiction with our assumptions. Therefore we have $\xi_z(t_0, z_0)\nu = 0$, which implies $\eta_z(t_0, z_0)\nu \neq 0$ by (6.1.14). We conclude

$$\begin{aligned}
&\lim_{k\to\infty} \frac{|p_k - \nabla u(t_0, x_0)|}{|x_k - x_0|} \\
&= \lim_{k\to\infty} \frac{|\eta(t_0, z_k) - \eta(t_0, z_0)|}{|z_k - z_0|} \frac{|z_k - z_0|}{|\xi(t_0, z_k) - \xi(t_0, z_0)|} \\
&= \lim_{k\to\infty} \frac{|\eta_z(t_0, z_0)\nu|z_k - z_0| + o(|z_k - z_0|)|}{|z_k - z_0|} \frac{|z_k - z_0|}{o(|z_k - z_0|)} = \infty. \quad \blacksquare
\end{aligned}$$

Finally, we describe the structure of u about a singular point which is not conjugate.

Theorem 6.3.5 *Let $(\bar{t},\bar{x}) \in \Sigma \setminus \Gamma$ be given. Then there exists a neighborhood N of $(\bar{t},\bar{x})$ and a finite number of C^{R+1} functions $v_1,\ldots,v_k : N \to \mathbb{R}$ such that $Dv_i \neq Dv_j$ if $i \neq j$ and $u = \min\{v_1,\ldots,v_k\}$ in N.*

Proof — Let $(\bar{t},\bar{x}) \in \Sigma \setminus \Gamma$. Then, by Theorem 6.3.3, $D^*u(\bar{t},\bar{x})$ contains a finite number of distinct elements $(q_1,p_1),\ldots,(q_k,p_k) \in \mathbb{R} \times \mathbb{R}^n$. By Theorem 6.2.11 we have $H(\bar{t},\bar{x},p_i)+q_i = 0$ for all i and therefore $p_1,\ldots,p_k$ are all different. By Theorem 6.2.8 the minimizers of (CV) with endpoint $(\bar{t},\bar{x})$ are obtained solving (6.1.9) with the terminal conditions $\xi(\bar{t}) = \bar{x}$, $\eta(\bar{t}) = p_i$ with $i = 1,\ldots,k$. Let us denote with $z_1,\ldots,z_k$ the initial point of these trajectories; then the minimizers can be written as $\xi(\cdot,z_1)$,..., $\xi(\cdot,z_k)$. Moreover $\eta(t,z_i) = p_i$ for $i = 1,\ldots,k$.

We argue now like in the proof of Theorem 6.2.10, with the difference that we have here k minimizers instead of one. Using the assumption that the point $(\bar{t},\bar{x})$ is not conjugate we can find $W_1,\ldots,W_k$ neighborhoods of $(\bar{t},z_i)$ and V neighborhood of $(\bar{t},\bar{x})$ such that the map $(t,z) \to (t,\xi(t,z))$ is a diffeomorphism between W_i and V for all $i = 1,\ldots,n$. We choose V small enough in such a way that $W_i \cap W_j = \emptyset$ if $i \neq j$ and that $\eta(t,z) \neq \eta(t',z')$ for any $(t,z) \in W_i$, $(t',z') \in W_j$. For every $i = 1,\ldots,n$, we define $\zeta_i : V \to \mathbb{R}^n$ in the following way: $\zeta_i(t,x) = z$ where z is the unique value such that $(t,z) \in W_i$ and that $\xi(t,z) = x$. Let us set

$$v_i(t,x) = u_0(z) + \int_0^t L(s,\xi(s,z),\dot{\xi}(s,z))ds, \text{ with } z = \zeta_i(t,x).$$

Then, in a neighborhood $N \subset V$ of $(\bar{t},\bar{x})$ we have that $u = \min\{v_1,\ldots,v_k\}$. Otherwise we could find points (t,x) arbitrarily close to $(\bar{t},\bar{x})$ which have a minimizer of the form $\xi(\cdot,z)$ with (t,z) not belonging to any W_i and we would obtain a contradiction like in the proof of Theorem 6.2.10. In addition, the functions v_i are of class C^{R+1} and satisfy $\nabla v_i(t,x) = \eta(t,\zeta_i(t,x))$ for all i, and so $Dv_i(t,x) \neq Dv_j(t,x)$ if $i \neq j$. ∎

From the above result we can deduce the following representation of the set $\Sigma \setminus \Gamma$.

Corollary 6.3.6 *Given $(t,x) \in \Sigma \setminus \Gamma$, there exists a neighborhood V of (t,x) such that $N \cap \Sigma$ is contained in a finite union of $n-$ dimensional hypersurfaces of class C^{R+1}.*

Proof — Let $v_1,\ldots,v_k$ be the functions given by Theorem 6.3.5. Since $u = \min_{i=1,\ldots,k} v_i$, u is not differentiable at a point $(s,y) \in N$ if and only if there exist $h,j \in \{1,\ldots,k\}$ such that $v_h(s,y) = v_j(s,y)$ and $v_i(s,y) \geq v_h(s,y)$ for any $i = 1,\ldots,k$. But since $Dv_i \neq Dv_j$ in N, the set $\{(s,y) \mid v_h(s,y) = v_j(s,y)\}$ is a C^{R+1} hypersurface for any pair h,j. ∎

6.4 Rectifiability of $\overline{\Sigma}$

We continue in this section the analysis of the regularity of the value function u of problem (CV). We have seen that u is semiconcave with linear modulus (see Theorem 6.2.1); therefore, its gradient Du belongs to the class BV (see Theorem 2.3.1–(ii)). In addition, its singular set Σ is countably n–rectifiable (Corollary 4.1.12). It is natural to investigate whether u enjoys further regularity properties. We recall that in Example 6.1.5 we have analyzed a class of data for which the value function is piecewise smooth. Such a type of regularity has been investigated by Schaeffer [74] in the one–dimensional case: he has shown that the value function is piecewise smooth for a generic class of data, but that there exist C^∞ data for which this property fails.

Therefore, we cannot expect the value function to be piecewise smooth in general. However, we will prove a regularity property which is in some sense close to piecewise smoothness, showing that u is regular on the complement of a closed set of codimension one. Our starting point is the result, stated in Theorem 6.2.10, that u has the same smoothness of the initial data in the complement of the set $\overline{\Sigma} = \Sigma \cup \Gamma$. We will prove that this set is countably $\mathcal{H}^n$–rectifiable. Since we already know that Σ is countably n–rectifiable, we only need to prove the rectifiability of Γ, or at least of $\Gamma \setminus \Sigma$. This will be done in this section; actually, we also give a Hausdorff estimate on $\Gamma \setminus \Sigma$ which is much sharper than the one needed for the rectifiability of Γ. In fact we prove that $\mathcal{H}^{n-1+2/R}(\Gamma \setminus \Sigma) = 0$, showing that the set of conjugate points (at least the regular ones) is very "small" compared with the set of irregular points.

Let us first outline the main ideas of our approach. We introduce the set

$$G = \{(t,z) \mid \det \xi_z(t,z) = 0\}.$$

and the function

$$\overline{\xi}(t,z) = (t, \xi(t,z)).$$

Then, by definition, $\overline{\xi} \in C^R$ and $\Gamma \subset \overline{\xi}(G)$. We can use this property to obtain upper bounds on the size of Γ. A possible way, pursued in [46], is to apply the following Sard–type result ([45, Theorem 3.4.3]):

Theorem 6.4.1 *Let $F : \mathbb{R}^N \to \mathbb{R}^N$ be a map of class C^R for some $R \geq 1$. For any $k \in \{0, 1, \ldots, N-1\}$ set*

$$A_k = \{x \in \mathbb{R}^N \mid \operatorname{rk} DF(x) \leq k\},$$

where $\operatorname{rk} DF$ *denotes the rank of DF. Then $\mathcal{H}^{k+(N-k)/R}(F(A_k)) = 0$.*

This result, applied to $\overline{\xi}$, yields

$$\mathcal{H}^{n+1/R}(\Gamma) = 0.$$

However, the above estimate does not imply the rectifiability of Γ even if the data of our problem are C^∞.

Another approach is to try and prove the rectifiability of G using the implicit function theorem. In the case of one space dimension this idea can be implemented in a very simple way.

Theorem 6.4.2 *Suppose $n = 1$. Then Γ is a countably 1–rectifiable set.*

Proof — Since $\Gamma \subset \overline{\xi}(G)$, it suffices to prove the rectifiability of G. We recall that $H_{pp} > 0$ by the convexity of H. Thus, if $(t_0, z_0) \in G$ we have, by (6.1.12) and (6.1.15),

$$\xi_{zt}(t_0, z_0) = H_{pp}(\xi(t_0, z_0), \eta(t_0, z_0))\eta_z(t_0, z_0) \neq 0.$$

By the implicit function theorem, we can cover G by a countable union of arcs of class C^1. ∎

When we try to extend this argument to the case of $n > 1$ we find two difficulties. The first one is technical, and consists of the fact that we have to deal with $\det \xi_z(t, z)$ instead of $\xi_z(t, z)$. A more substantial problem is the fact that $\det \xi_z(t, z)$ may well vanish simultaneously with its time derivative. This fact is explained by the following result.

Theorem 6.4.3 *Let $n > 1$. Then, for any $(t, z) \in G$,*

$$\partial_t(\det \xi_z(t, z)) \neq 0 \quad \Longleftrightarrow \quad \operatorname{rk} \xi_z(t, z) = n - 1.$$

Before giving the proof of Theorem 6.4.3, let us show how this result can be used to obtain the rectifiability of Γ.

Theorem 6.4.4 *Suppose $n > 1$. Then Γ is a countably $\mathcal{H}^n$–rectifiable set.*

Proof — Let us define

$$G' = \{(t, z) \mid \operatorname{rk}(\xi_z(t, z)) = n - 1\}, \tag{6.4.31}$$
$$G'' = \{(t, z) \mid \operatorname{rk}(\xi_z(t, z)) < n - 1\}. \tag{6.4.32}$$

Then, by Theorem 6.4.3 and the implicit function theorem, we deduce that G' is a countably n–rectifiable set. On the other hand, by Theorem 6.4.1,

$$\mathcal{H}^n(\overline{\xi}(G'')) \leq \mathcal{H}^{n-1+2/R}(\overline{\xi}(G'')) = 0.$$

Since $\Gamma \subset \overline{\xi}(G' \cup G'')$, our assertion follows. ∎

The proof of Theorem 6.4.3 requires many intermediate steps as we show below. The first one consists of giving a suitable expression for the derivative of a determinant. For this purpose, we recall below some elementary facts from linear algebra.

Let $A = (a_{ij})$ be an $n \times n$ matrix (we assume $n > 1$). We denote by A^+ the transpose of the matrix of the cofactors of A. Equivalently, if we denote by a^+_{ij} the generic element of A^+ and by $\hat{A}_{ij}$ the matrix obtained removing from A the i–th row and the j–th column, then $a^+_{ij} = (-1)^{i+j} \det \hat{A}_{ji}$. Then, Laplace's identity states that

$$\sum_{j=1}^{n} a_{ij} a^+_{jk} = \delta_{ik} \det A,$$

or

$$AA^+ = A^+ A = (\det A)\, I. \tag{6.4.33}$$

Therefore

$$\frac{\partial \det A}{\partial a_{ij}} = \frac{\partial}{\partial a_{ij}} \left(\sum_{k=1}^{n} a_{ik} a^+_{ki} \right) = a^+_{ji}. \tag{6.4.34}$$

Moreover, (6.4.33) implies that

$$\text{rk } A = n \quad \Rightarrow \quad \text{rk } A^+ = n. \tag{6.4.35}$$

Similarly, it is easy to check that

$$\text{rk } A = n-1 \quad \Rightarrow \quad \text{rk } A^+ = 1 \tag{6.4.36}$$

$$\text{rk } A < n-1 \quad \Rightarrow \quad \text{rk } A^+ = 0 \tag{6.4.37}$$

Indeed, recalling the estimate

$$\text{rk } BC \geq \text{rk } B + \text{rk } C - n,$$

which holds for any two $n \times n$ matrices B and C, from (6.4.33) we conclude that $\text{rk } A^+ \leq 1$ if $\text{rk } A = n-1$. On the other hand, by definition, $A^+ = 0$ if and only if $\text{rk } A < n-1$.

Finally, if $A(\cdot)$ is time–dependent and of class C^1, then (6.4.34) yields

$$\begin{aligned} \frac{d}{dt} \det A(t) &= \sum_{i,j=1}^{n} \frac{\partial \det A(t)}{\partial a_{ij}} a'_{ij}(t) \\ &= \sum_{i,j=1}^{n} a^+_{ji}(t) a'_{ij}(t) = \text{tr}\,(A'(t) A^+(t)). \end{aligned} \tag{6.4.38}$$

Let us now consider the hamiltonian system (6.1.9) again. In the following we denote by A^* the transpose of a given matrix A.

Lemma 6.4.5 *We have $\xi_z^*(t,z)\eta_z(t,z) = \eta_z^*(t,z)\xi_z(t,z)$ for any (t,z).*

Proof — Let us set $R(t) = \xi_z^*(t,z)\eta_z(t,z)$. Since

$$R(0) = D^2u_0(z) = R^*(0),$$

$$R'(t) = \eta_z^*(t,z)H_{pp}\eta_z(t,z) - \xi_z^*(t,z)H_{xx}\xi_z(t,z) = (R^*)'(t),$$

we have $R(t) = R^*(t)$ for any $t \geq 0$, which proves our assertion. ∎

Lemma 6.4.6 *Suppose $n > 1$. Let* rk $\xi_z(t_0,z_0) = n-1$ *for some (t_0,z_0) and let θ be a generator of* Ker $\xi_z(t_0,z_0)$. *Then $\eta_z(t_0,z_0)\theta$ generates the orthogonal complement of* Im $\xi_z(t_0,z_0)$ *and*

$$\xi_z^+(t_0,z_0) = c\theta \otimes \eta_z(t_0,z_0)\theta \tag{6.4.39}$$

for some $c \in \mathbb{R} \setminus \{0\}$.

Proof — We know from (6.1.15) that $\eta_z(t_0,z_0)\theta \neq 0$. Given any $w \in \mathbb{R}^n$, we have

$$\begin{aligned}\langle \eta_z(t_0,z_0)\theta, \xi_z(t_0,z_0)w\rangle &= \langle \xi_z^*(t_0,z_0)\eta_z(t_0,z_0)\theta, w\rangle \\ &= \langle \eta_z^*(t_0,z_0)\xi_z(t_0,z_0)\theta, w\rangle = 0.\end{aligned}$$

Since w is arbitrary, we obtain the first assertion of the Lemma.

To prove (6.4.39) we first observe that, by (6.4.36), $\xi_z^+(t_0,z_0)$ has rank one, and so $\xi_z^+(t_0,z_0) = v_1 \otimes v_2$ for some nonzero vectors v_1, v_2. Moreover,

$$v_1\mathbb{R} = \mathrm{Im}\,\xi_z^+(t_0,z_0) = \mathrm{Ker}\,\xi_z(t_0,z_0).$$

Indeed, equality (6.4.33) yields $\mathrm{Im}\,\xi_z^+(t_0,z_0) \subset$ Ker $\xi_z(t_0,z_0)$; then, the second equality above follows from the fact that both $\mathrm{Im}\,\xi_z^+(t_0,z_0)$ and Ker $\xi_z(t_0,z_0)$ are one–dimensional spaces. Similarly,

$$\{v_2\}^\perp = \mathrm{Ker}\,\xi_z^+(t_0,z_0) = \mathrm{Im}\,\xi_z(t_0,z_0).$$

Therefore v_1 and v_2 must be parallel to θ and $\eta_z(t_0,z_0)\theta$ respectively. ∎

Proof of Theorem 6.4.3 — By (6.4.38) we have, for any (t,z), (we omit for simplicity the dependence of H on (t,ξ,η))

$$\begin{aligned}\frac{\partial}{\partial t}\det\xi_z(t,z) &= \mathrm{tr}\left(\xi_{zt}(t,z)\xi_z^+(t,z)\right) \\ &= \mathrm{tr}\left(H_{px}\xi_z(t,z)\xi_z^+(t,z)\right) + \mathrm{tr}\left(H_{pp}\eta_z(t,z)\xi_z^+(t,z)\right) \\ &= \det\xi_z(t,z)\mathrm{tr}\left(H_{px}\right) + \mathrm{tr}\left(H_{pp}\eta_z(t,z)\xi_z^+(t,z)\right).\end{aligned}$$

Now let us suppose that $\operatorname{rk}\xi_z(t,z) = n-1$. Then, if θ is such that $\operatorname{Ker}\xi_z(t_0,z_0) = \theta\mathbb{R}$ and $q = \eta_z(t_0,z_0)\theta$, we have by Lemma 6.4.6,

$$\begin{aligned}\frac{\partial}{\partial t}\det\xi_z(t_0,z_0) &= \operatorname{tr}(cH_{pp}\eta_z(t,z)\theta\otimes q)\\ &= cH_{pp}q\cdot q \neq 0,\end{aligned}$$

since H_{pp} is positive definite and $q \neq 0$. On the other hand, if $\operatorname{rk}\xi_z(t,z) < n-1$, then $\xi_z^+(t,z) = 0$ by (6.4.37) and thus the derivative of $\det\xi_z$ is zero. ∎

Remark 6.4.7 As a consequence of the implicit function theorem and of Theorem 6.4.3, we conclude that the set G' defined in (6.4.31) is locally a graph, that is for any $(t_0,z_0) \in G'$, there exist $r,\varepsilon > 0$ and a map $\phi : B_r(z_0) \to]t_0-\varepsilon, t_0+\varepsilon[$ of class C^{R-1}, such that

$$\det\xi_z(t,z) = 0 \quad \text{if and only if } t = \phi(z) \tag{6.4.40}$$

for any $(t,z) \in]t_0-\varepsilon, t_0+\varepsilon[\times B_r(z_0)$. If we choose r small enough, we may also assume that $\xi_z(\phi(z),z)$ has rank $n-1$ for any $z \in B_r(z_0)$. Then, there also exists a vector field $\theta : B_r(z_0) \to (\mathbb{R}^n \setminus \{0\})$ of class C^{R-1} such that, for any $z \in B_r(z_0)$, $\theta(z)$ is a generator of $\operatorname{Ker}\xi_z(\phi(z),z)$. Indeed, for θ one can take the exterior product of $n-1$ linearly independent rows of $\xi_z(\phi(z),z)$.

The geometric relationship between vector fields $D\phi$ and θ, about a regular conjugate point, is made clear by the following result.

Proposition 6.4.8 *Let G' be as in* (6.4.31) *and, given $(t_0,z_0) \in G'$, let r,ε and ϕ,θ be defined as in Remark 6.4.7. If $\bar z \in B_r(z_0)$ is such that $(\phi(\bar z), \xi(\phi(\bar z),\bar z)) \notin \Sigma$, then*

$$D\phi(\bar z)\cdot\theta(\bar z) = 0.$$

Before proving this result, let us show how it can be used to improve the information about $\Gamma \setminus \Sigma$.

Theorem 6.4.9 *Assume conditions* (L1)—(L4) *and* (6.1.2). *Then*

$$\mathcal{H}^{n-1+2/R}(\Gamma\setminus\Sigma) = 0.$$

In particular, if

$$L \in C^\infty(\overline{Q_T}\times\mathbb{R}^n), \quad u_0 \in C^\infty(\mathbb{R}^n),$$

then

$$\mathcal{H}-\dim(\Gamma\setminus\Sigma) \leq n-1.$$

Proof — Let G' and G'' be defined as in (6.4.31) and (6.4.32), respectively, and let $(z_0, t_0) \in G'$ be fixed. Then, by Remark 6.4.7, there exist $r, \varepsilon > 0$ such that $(]t_0 - \varepsilon, t_0 + \varepsilon[\times B_r(z_0)) \cap G$ coincides with the graph of some function $\phi : B_r(z_0) \to]t_0 - \varepsilon, t_0 + \varepsilon[$. In other words, $\overline{\xi}((]t_0 - \varepsilon, t_0 + \varepsilon[\times B_r(z_0)) \cap G) = \tilde{\xi}(B_r(z_0))$, where

$$\tilde{\xi}(z) = \overline{\xi}(\phi(z), z), \qquad z \in B_r(z_0).$$

Moreover, Proposition 6.4.8 implies that, for any $z \in B_r(z_0)$ satisfying $\tilde{\xi}(z) \notin \Sigma$, there exists $\theta \in (\mathbb{R}^n \setminus \{0\})$ such that $\xi_z(\phi(z), z)\theta = D\phi(z) \cdot \theta = 0$. For such a z,

$$D\tilde{\xi}(z)\theta = 0,$$

and so

$$\text{rk } D\tilde{\xi}(z) = \text{rk } [\xi_t(\phi(z), z) \otimes D\phi(z) + \xi_z(\phi(z), z)] \leq n - 1.$$

By Theorem 6.4.1 we conclude that $\mathcal{H}^{n-1+2/R}(\overline{\xi}(G') \setminus \Sigma) = 0$. On the other hand, from the proof of Theorem 6.4.4 we know that $\mathcal{H}^{n-1+2/R}(\overline{\xi}(G'')) = 0$. Since

$$\Gamma \setminus \Sigma \subset \overline{\xi}(G'') \cup (\overline{\xi}(G') \setminus \Sigma),$$

the conclusion follows. ∎

For the proof of Proposition 6.4.8 we need the following result.

Lemma 6.4.10 *Let $(t_0, z_0) \in G'$, and let r, ε and ϕ, θ be defined as in Remark 6.4.7. Then, for any $z \in B_r(z_0)$,*

$$D\phi(z) \cdot \theta(z) = 0 \quad \Longleftrightarrow \quad \frac{\partial^2 \xi}{\partial \theta(z)^2}(\phi(z), z) \cdot (\eta_z(\phi(z), z)\theta(z)) = 0. \tag{6.4.41}$$

Proof — By definition we have, for any $z \in B_r(z_0)$ and $i = 1, \ldots, n$,

$$\sum_{j=1}^{n} \frac{\partial \xi_i}{\partial z_j}(\phi(z), z)\theta_j(z) = 0. \tag{6.4.42}$$

Let us set for simplicity $q(z) = \eta_z(\phi(z), z)\theta(z)$. Differentiating (6.4.42) with respect to z_k and using (6.1.12) we obtain

$$\sum_{h=1}^{n} \frac{\partial^2 H}{\partial p_i \partial p_h} q_h \frac{\partial \phi}{\partial z_k} + \sum_{j=1}^{n} \left\{ \frac{\partial^2 \xi_i}{\partial z_j \partial z_k}\theta_j + \frac{\partial \xi_i}{\partial z_j}\frac{\partial \theta_j}{\partial z_k} \right\} = 0 \tag{6.4.43}$$

for any $i, k = 1, \ldots, n$. We then multiply equality (6.4.43) by $q_i\theta_k$ and sum over $i, k = 1, \ldots, n$ to obtain

$$\sum_{i,h=1}^{n} \frac{\partial^2 H}{\partial p_i \partial p_h} q_h q_i \sum_{k=1}^{n} \frac{\partial \phi}{\partial z_k}\theta_k + \sum_{i,j,k=1}^{n} \frac{\partial^2 \xi_i}{\partial z_j \partial z_k}\theta_j\theta_k q_i = 0 \tag{6.4.44}$$

recalling that, in light of Lemma 6.4.6, $q \perp \mathrm{Im}\xi_z$. This last identity may be rewritten as

$$(H_{pp}q \cdot q)(D\phi \cdot \theta) + \frac{\partial^2 \xi}{\partial \theta^2} \cdot q = 0.$$

Since H_{pp} is positive definite and $q \neq 0$, we have that $H_{pp}q \cdot q \neq 0$. Therefore, $D\phi \cdot \theta = 0$ if and only if $\frac{\partial^2 \xi}{\partial \theta^2} \cdot q = 0$. ∎

In the proof of Proposition 6.4.8 we will apply a known lemma on local invertibility of mappings, which we recall below. For its proof see e.g. [7, Lemma 2.5].

Lemma 6.4.11 *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a map of class C^2. Let DF have rank $n-1$ at some $\bar{z} \in \mathbb{R}^n$ and set $\bar{x} = F(\bar{z})$. Let θ be a generator of* $\mathrm{Ker}DF(\bar{z})$*, and w be a nonzero vector orthogonal to* $\mathrm{Im}DF(\bar{z})$*. Suppose that*

$$\frac{\partial^2 F}{\partial \theta^2}(\bar{z}) \cdot w > 0.$$

Then there exist $\rho, \sigma > 0$ such that the equation

$$F(z) = \bar{x} + sw, \quad z \in B_\rho(\bar{z})$$

has two solutions if $0 < s < \sigma$, and no solutions if $-\sigma < s < 0$.

Proof of Proposition 6.4.8. Let us take $\bar{z} \in B_r(z_0)$ and set

$$\bar{t} = \phi(\bar{z})\bar{x} = \xi(\bar{t}, \bar{z}).$$

Suppose that $D\phi(\bar{z}) \cdot \theta(\bar{z}) \neq 0$. Then, in light of Lemmas 6.4.6 and 6.4.10 the map $F = \xi(\bar{t}, \cdot)$ satisfies all assumptions of Lemma 6.4.11 with

$$\theta = \theta(\bar{z}), \quad w = \pm \eta_z(\bar{t}, \bar{z})\theta(\bar{z}).$$

For $x_k = \bar{x} - w/k$, let ξ_k be a minimizer of functional $J_{\bar{t}}$ in (6.1.1) with terminal condition $\xi_k(\bar{t}) = x_k$. Let us consider the sequence $\{z_k\}$ of all initial points of such minimizers, i.e. $z_k = \xi_k(0)$. Then ξ_k coincides with the characteristic starting at z_k and in particular $x_k = \xi(\bar{t}, z_k)$. So, Lemma 6.4.11 implies that $z_k \notin B_\rho(\bar{z})$ if k is large enough.

On the other hand, $\{z_k\}$ is bounded. Hence, we conclude that there exists a subsequence of $\{z_k\}$ converging to some $\hat{z} \neq \bar{z}$. Then $\xi(\cdot, \hat{z})$ is a minimizing trajectory for $J_{\bar{t}}$ at $(\bar{t}, \bar{x})$, different from $\xi(\cdot, \bar{z})$. This fact contradicts our assumption that $(\bar{t}, \bar{x})$ be a regular point. ∎

To conclude, we give an example showing that, if the smoothness of the data of (CV) is below a certain threshold, then $\overline{\Sigma}$ can fail to be countably $\mathcal{H}^n$–rectifiable.

Example 6.4.12 *There exists $u_0 \in W^{2,\infty}(\mathbb{R}) \setminus C^2(\mathbb{R})$ with the following property: the solution of equation*

$$\partial_t u(t,x) + \frac{1}{2}(\partial_x u(t,x))^2 = 0 \qquad t \in \mathbb{R}_+, x \in \mathbb{R} \tag{6.4.45}$$

with initial value $u(0,x) = u_0(x)$ is such that $\mathcal{H}$–$\dim(\overline{\Sigma}) \geq 1 + \log 2/\log 3$.

Proof — We introduce some notation. Let i, j be integers with $i \geq 0$ and $0 \leq j \leq 2^i - 1$. Then j can be written in the form $j = \sum_{h=0}^{i-1} e_h 2^h$ with $e_h \in \{0,1\}$. We define

$$c_{i,j} = \sum_{h=1}^{i}(2e_{i-h})3^{-h} + 1/3^i 2, \quad I_{i,j} =]c_{i,j} - 1/(3^i 6), c_{i,j} + 1/(3^i 6)[\,.$$

We then consider the *Cantor set C*

$$C = [0,1] \setminus \bigcup_{i=0,\ldots,\infty,\ j=0,\ldots,2^i-1} I_{i,j}.$$

It is well known that C consists of all numbers $x \in [0,1]$ that can be written in the form $x = \sum_{h=1}^{\infty} d_h 3^{-h}$ with $d_h \in \{0,2\}$. Thus C is uncountable; moreover $\mathcal{H}$–$\dim(C) = \log 2/\log 3 > 0$ (see [44, example 2.7]).

A property we will need in the sequel is the fact that $C \subset \overline{\{c_{i,j}\}}$. Indeed, take any $x = \sum_{h=1}^{\infty} d_h 3^{-h}$ with $d_h \in \{0,2\}$. Given a positive integer i, define $j(i) = \sum_{h=0}^{i-1}(d_{i-h}/2)2^h$. Then

$$|c_{i,j(i)} - x| = \left| \sum_{h=1}^{\infty} d_h 3^{-h} - \sum_{h=1}^{i-1} d_h 3^{-h} \right| \leq 3^{-i+1}.$$

Since i is arbitrary, we have proved our claim.

We now define, for $\lambda \in]0,1[\,$,

$$b(\lambda, z) = \begin{cases} \left(\dfrac{z^2}{\lambda} - \lambda\right)^2 & -\lambda < z < \lambda \\ 0 & z \geq \lambda \text{ or } z \leq -\lambda. \end{cases}$$

Note that $b(\lambda, z) \in W^{2,\infty}(\mathbb{R})$. The second derivative $b_{zz}(\lambda, z)$ exists for $z \neq \pm\lambda$ and satisfies $b_{zz}(\lambda, z) = b_{zz}(1, z/\lambda)$. It follows that $|b_{zz}(\lambda, z)| \leq K$ for some constant K independent of λ. This implies that $b_z(\lambda, \cdot)$ has Lipschitz constant K and that

$$\sup_{z \in [-\lambda, -\lambda+\varepsilon[} b(\lambda, z) \leq \int_{-\lambda}^{-\lambda+\varepsilon} \left(\int_{-\lambda}^{z} |b_{zz}(\lambda, w)| dw \right) dz \leq K\varepsilon^2/2 \tag{6.4.46}$$

for every $\lambda > 0, \varepsilon > 0$. Let us first consider equation (6.4.45) with initial value

$$u(0,x) = b(\lambda, x) \qquad x \in \mathbb{R}.$$

The solution is the value function of the associated problem of calculus of variations

$$u(t,x) = \min_{\begin{cases} \xi \in W^{1,1}([0,t]) \\ \xi(t) = x \end{cases}} \left[\int_0^t \frac{1}{2}(\dot{\xi}(s))^2 dt + b(\lambda, \xi(0)) \right]. \tag{6.4.47}$$

We know that any optimal path for the point (t,x) is of the form $\xi(\cdot, z)$ for some z. In our case the hamiltonian H only depends on p and the solution of system (6.1.9) is the following

$$\xi(t,z) = z + tb_z(\lambda, z) = z + 4t\left(\left(\frac{z}{\lambda}\right)^2 - 1\right) z\chi_{[-\lambda,\lambda]}(z). \tag{6.4.48}$$

Let us compute the minimizers in the case $x = 0$. If $t \leq 1/4$ then equality (6.4.48) shows that $\xi(t,z) = 0$ only if $z = 0$, while if $t > 1/4$ this equality holds if $z = 0$ and if $z = \pm z(t,\lambda)$ where

$$z(t,\lambda) = \lambda\sqrt{1 - \frac{1}{4t}}.$$

A direct computation shows that if $t > 1/4$ the minimum in (6.4.47) is attained for $z = \pm z(t,\lambda)$, while the trajectory $\xi(\cdot, 0)$ is not minimal. Thus, if $t > 1/4$ the minimizer for the point $(t,0)$ is not unique, and so the half line $]1/4, \infty[\times\{0\}$ is contained in the set Σ of irregular points.

The initial value for our counterexample is obtained repeating this construction on each interval $I_{i,j}$. We define

$$u_0(z) = \begin{cases} 0 & z \in C \text{ or } z > 1 \text{ or } z < -1 \\ b(1/3^i 6, z - c_{i,j}). & \text{if } z \in I_{i,j}. \end{cases} \tag{6.4.49}$$

Let us show that $u_0 \in W^{2,\infty}(\mathbb{R})$. By construction u_0 is differentiable in the complement of C; on the other hand we claim that if $z \in C$ then $|u_0(y) - u_0(z)| \leq K|y - z|^2/2$, which implies that $u_0'(z) = 0$. In fact, let us take for instance $y > z$. If $y \in C$ then $u(y) = u(z) = 0$. Otherwise let y' be the left endpoint of the interval $I_{i,j}$ to which y belongs; we have, by estimate (6.4.46),

$$u(y) - u(z) = u(y) - u(y') \leq K|y - y'|^2/2 \leq K|y - z|^2/2,$$

which proves our claim. In a similar way, we can prove that u_0' is Lipschitz continuous by using the uniform Lipschitz continuity of $b_z(\lambda,\cdot)$.

However, u_0 is not of class C^2. In fact, u_0'' attains in each interval $I_{i,j}$ a maximum and a minimum value which are independent of i,j. Since the length of $I_{i,j}$ tends to zero as $i\to\infty$, this shows that u_0'' cannot be uniformly continuous in $[-1,1]$.

Let us now consider equation (6.4.45) with initial value u_0 given by (6.4.49). It is easy to see by the definition of u_0 that the solution is given by

$$u(t,x)=\begin{cases}0 & x\in C \text{ or } x>1 \text{ or } x<-1\\ u^{1/3^i6}(t,x-c_{i,j}) & \text{if } x\in I_{i,j},\end{cases}$$

where we have denoted by $u^\lambda(t,x)$ the solution of equation (6.4.45) with initial value $u_0(z)=b(\lambda,z)$. Our previous analysis implies that $\Sigma\supset\,]1/4,\infty[\,\times\{c_{i,j}\mid i\geq 0, 0\leq j\leq 2^i-1\}$. But since $C\subset\overline{\{c_{i,j}\}}$ we obtain that $\overline{\Sigma}\supset\,]1/4,\infty[\,\times C$. Thus Σ has Hausdorff dimension at least $1+\log 2/\log 3$. ∎

Bibliographical notes.

The results of Section 6.1 are classical; our presentation partly follows [47, 48]. General references about the dynamic programming approach considered in Section 6.2 can be found among the bibliographical notes to Chapter 1. The result of Corollary 6.2.9 was first obtained by Kuznetsov and Šiškin (see [63]), while Theorem 6.2.8 was given in [28]. Fleming derived Theorem 6.2.10 in [46] and pointed out that it could provide a basis for studying the fine regularity properties of the value function.

The presentations of Section 6.3 and 6.4 follows [24]. Some of the results of Section 6.3 were more or less explicitly present in the previous literature. For instance, the property that conjugate points are related to a blowup of the second derivatives was well known. Also, Corollary 6.3.6 was already given in [46]. The rectifiability results of Section 6.4 are instead an original contribution of [24]. The impulse for studying that kind of regularity was given by some recent works about image segmentation (see e.g. [8, 9]), where the class SBV of special functions of bounded variation was introduced.

Appendix A

A.1 The Legendre transform

The *Legendre transform* (sometimes also called *Fenchel transform* or *convex conjugate*) is a classical topic of convex analysis. A comprehensive treatment of this transform can be found in many textbooks (e.g. [72]); here we give, for the convenience of the reader, a short and self–contained exposition of the properties which are used in the rest of our work.

Let us first recall some basic notions about convex functions. Given $L : \mathbb{R}^n \to \mathbb{R}$ convex and $u \in \mathbb{R}^n$, we recall that the subdifferential of L at u is the set of all $p \in \mathbb{R}^n$ such that

$$L(v) \geq L(u) + p \cdot (v - u) \qquad \forall\, v \in \mathbb{R}^n.$$

The subdifferential of L at u is denoted by $D^- L(u)$. As observed in Remark 3.3.2, for a convex function this definition is equivalent to the one of Fréchet subdifferential given in Definition 3.1.1. Therefore, from Proposition 3.3.4 we immediately deduce the following

Proposition A.1.1 *Let $L : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then $D^- L(u)$ is nonempty for every $u \in \mathbb{R}^n$. In addition, L is differentiable at u if and only if $D^- L(u)$ is a singleton. If L is differentiable at all points, then its differential $DL(u)$ is continuous.*

Let us now give the definition of the Legendre transform. We restrict ourselves to convex functions defined in the whole space and superlinear, a class which is enough general for the purposes of this book.

Definition A.1.2 *Let $L : \mathbb{R}^n \to \mathbb{R}$ be a convex function which satisfies*

$$\lim_{|u| \to \infty} \frac{L(u)}{|u|} = +\infty. \tag{A.1.1}$$

The Legendre transform *of L is the function*

$$L^*(p) = \sup_{u\in\mathbb{R}^n} [u \cdot p - L(u)], \qquad p \in \mathbb{R}^n. \tag{A.1.2}$$

Theorem A.1.3 *Let $L : \mathbb{R}^n \to \mathbb{R}$ be a convex function satisfying* (A.1.1).

(a) *For every p there exists at least one point u_p where the supremum in* (A.1.2) *is attained. In addition, for every bounded set $C \subset \mathbb{R}^n$ there exists $R > 0$ such that $|u_p| < R$ for all $p \in C$.*

(b) *The function L^* is convex and satisfies $\displaystyle\lim_{|p|\to\infty} \frac{L^*(p)}{|p|} = +\infty$.*

(c) $L^{**} = L$.

(d) *Given $\bar{u}, \bar{p} \in \mathbb{R}^n$ we have*

$$\bar{p} \in D^- L(\bar{u}) \iff \bar{u} \in D^- L^*(\bar{p}) \iff L(\bar{u}) + L^*(\bar{p}) = \bar{u} \cdot \bar{p}.$$

Proof — **(a)** The claimed properties are a straightforward consequence of the convexity of L and of assumption (A.1.1).
(b) The conjugate function L^* is convex since it is the supremum of linear functions. In addition, for all $M > 0$ and $p \in \mathbb{R}^n$ we have

$$L^*(p) \geq M \frac{p}{|p|} \cdot p - L\left(M \frac{p}{|p|}\right) \geq M|p| - \max_{|u|=M} L(u),$$

and so, for all $M > 0$,

$$\liminf_{|p|\to\infty} \frac{L^*(p)}{|p|} \geq M.$$

(c) We have by definition $L(u) \geq u \cdot p - L^*(p)$ for all u, p, which implies that $L \geq L^{**}$. To prove the converse inequality, fix any $\bar{v} \in \mathbb{R}^n$ and take $\bar{p} \in D^- L(\bar{v})$. Let $\bar{u}$ be such that $L^*(\bar{p}) = \bar{u} \cdot \bar{p} - L(\bar{u})$. Since $\bar{p} \in D^- L(\bar{v})$, we have $L(\bar{u}) \geq L(\bar{v}) + \bar{p} \cdot (\bar{u} - \bar{v})$ and therefore

$$L(\bar{v}) \leq L(\bar{u}) - \bar{p} \cdot (\bar{u} - \bar{v}) = -L^*(\bar{p}) + \bar{p} \cdot \bar{v},$$

which implies that $L(\bar{v}) \leq L^{**}(\bar{v})$.
(d) By the definition of subdifferential,

$$\bar{p} \in D^- L(\bar{u}) \quad \iff \quad \bar{p} \cdot \bar{u} - L(\bar{u}) = \max_{u\in\mathbb{R}^n} [\bar{p} \cdot u - L(u)].$$

Therefore $\bar{p} \in D^- L(\bar{u})$ if and only if $L^*(\bar{p}) = \bar{u} \cdot \bar{p} - L(\bar{u})$. Analogously, $\bar{u} \in D^- L^*(\bar{p})$ if and only if $L^{**}(\bar{u}) = \bar{u} \cdot \bar{p} - L^*(\bar{p})$. Since $L^{**} = L$, we obtain the desired equivalence. ■

Theorem A.1.4 *L is strictly convex if and only if L^* is of class C^1.*

Proof — It is easy to see that L is not strictly convex if and only if there exist two distinct points u_1, u_2 such that $D^-L(u_1) \cap D^-L(u_2)$ contains some element $\bar{p}$. By part (c) of the previous theorem, this is equivalent to the property that u_1, u_2 both belong to $D^-L^*(\bar{p})$, that is, to the nondifferentiability of L^* at $\bar{p}$. ∎

Theorem A.1.5 *Suppose that $L \in C^k(\mathbb{R}^n)$, with $k \geq 2$, satisfies assumption* (A.1.1) *and*

$$D^2L(u) \text{ is positive definite for all } u \in \mathbb{R}^n. \tag{A.1.3}$$

Then L^* belongs to $C^k(\mathbb{R}^n)$, DL is a C^{k-1} diffeomorphism from $\mathbb{R}^n$ to $\mathbb{R}^n$ and

$$L^*(p) = p \cdot DL^{-1}(p) - L(DL^{-1}(p)) \tag{A.1.4}$$

$$DL^*(p) = (DL)^{-1}(p), \tag{A.1.5}$$

$$D^2L^*(p) = \left[D^2L\left((DL)^{-1}(p)\right)\right]^{-1} \tag{A.1.6}$$

Proof — By the previous theorem and by the property $L^{**} = L$, if L is differentiable and strictly convex, then so is L^*; in this case, by Theorem (c), the differentials DL and DL^* are reciprocal inverse. Therefore DL is a diffeomorphism provided D^2L is everywhere non degenerate, which is ensured by (A.1.3). This proves (A.1.5) and (A.1.6). To obtain (A.1.4) we observe that, under our assumptions, the maximum in (A.1.2) is achieved at a unique point u_p which satisfies $DL(u_p) = p$. ∎

Let now $\Omega \subset \mathbb{R}^m$ be an open set and let $L \in C(\Omega \times \mathbb{R}^n)$. Suppose that L is convex in the second argument and that it satisfies

$$L(x,u) \geq l(|u|) - l_0, \qquad (x,u) \in \Omega \times \mathbb{R}^n \tag{A.1.7}$$

where l_0 is a suitable constant and $l : \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing function such that $l(r)/r \to +\infty$ as $r \to +\infty$. Under these hypotheses we can define the Legendre transform of L with respect to the second group of variables as follows

$$H(x,p) = \sup_{u \in \mathbb{R}^n} \left[p \cdot u - L(x,u)\right], \qquad (x,u) \in \Omega \times \mathbb{R}^n. \tag{A.1.8}$$

Theorem A.1.6

(a) *For every* $(x,p) \in \Omega \times \mathbb{R}^n$ *there exists at least one point* $u(x,p)$ *where the supremum in* (A.1.8) *is attained. In addition, for every bounded set* $C \subset \mathbb{R}^n$ *there exists* $R > 0$ *such that every* u *associated to* $(x,p) \in \Omega \times C$ *satisfies* $|u| < R$.

(b) *The transformed function* H *is continuous, convex in the second argument and satisfies*

$$\lim_{|p|\to\infty} \inf_{x\in\Omega} \frac{H(x,p)}{|p|} = +\infty.$$

(c) *If* L *is strictly convex in* u, *then* $D_pH(x,p)$ *exists and is continuous for* $(x,p) \in \Omega \times \mathbb{R}^n$.

(d) *Suppose that for every* $R > 0$ *there exists* $\alpha_R > 0$ *such that*

$$|L(x,u) - L(y,u)| \leq \alpha_R |x-y|, \qquad x,y \in \Omega, u \in B_R.$$

Then for every $\rho > 0$ *there exists* $\beta_\rho > 0$ *such that*

$$|H(x,p) - H(y,p)| \leq \beta_\rho |x-y|, \qquad x,y \in \Omega, u \in B_\rho.$$

Proof — **(a)** It is a direct consequence of the properties of L.
(b) Convexity and superlinearity are proved like in the case without x–dependence. To prove continuity, let us consider a sequence $(x_n, p_n) \subset \Omega \times \mathbb{R}^n$ converging to some $(\bar{x}, \bar{p})$. We have

$$H(x_n,p_n) = p_n \cdot u_n - L(x_n,u_n), \qquad H(\bar{x},\bar{p}) = \bar{p} \cdot \bar{u} - L(\bar{x},\bar{u})$$

for suitable $u_n, \bar{u}$. Then

$$H(x_n,p_n) \geq p_n \cdot \bar{u} - L(x_n,\bar{u}).$$

Letting $n \to \infty$ we obtain that $\liminf_{n\to\infty} H(x_n,p_n) \geq H(\bar{x},\bar{p})$. To prove the converse inequality, let us choose a subsequence (x_{n_k}, p_{n_k}) such that $\lim_{k\to\infty} H(x_{n_k},p_{n_k}) = \limsup_{n\to\infty} H(x_n,p_n)$. Since, by part (i), $\{u_n\}$ is bounded, we can assume that the subsequence u_{n_k} converges to some value u^*. Then we have

$$\begin{aligned} H(\bar{x},\bar{p}) &\geq \bar{p} \cdot u^* - L(\bar{x},u^*) = \lim_{k\to\infty} p_{n_k} \cdot u_{n_k} - L(x_{n_k},u_{n_k}) \\ &= \limsup_{n\to\infty} H(x_n,p_n). \end{aligned}$$

This proves the continuity of H.

(c) If L is strictly convex, then the supremum in (A.1.8) is attained for a unique value of u, which we will denote by $u(x,p)$. From Theorem A.1.4 and Theorem A.1.3–(d) it follows that H is differentiable with respect to p and that $D_pH(x,p) = u(x,p)$. Let us prove the continuity of $u(x,p)$. Recalling (a), it suffices to show that, if $(x_n,p_n) \to (\bar{x},\bar{p})$ and $u(x_n,p_n) \to \bar{u}$, then $\bar{u} = u(\bar{x},\bar{p})$. This follows from the continuity of H, since we can pass to the limit in the equality $H(x_n,p_n) = p_n \cdot u_n - L(x_n,u_n)$ to obtain $H(\bar{x},\bar{p}) = \bar{p}\cdot\bar{u} - L(\bar{x},\bar{u})$, which implies that $\bar{u} = u(\bar{x},\bar{p})$ by the uniqueness of the maximizer in (A.1.8).

(d) Let us fix $\rho > 0$. By part (a) there exists $R > 0$ with the following property: for every $(x,p) \in \Omega \times B_\rho$, any u which maximizes the expression in (A.1.8) belongs to B_R. Let us now consider $(x,p),(y,p) \in \Omega \times B_\rho$ and suppose, for instance, that $H(x,p) \geq H(y,p)$. Let $\bar{u}$ be such that $H(y,p) = p\cdot\bar{u} - L(y,\bar{u})$. Then $\bar{u} \in B_R$ and we can use our hypothesis on L to obtain

$$\begin{aligned} |H(x,p) - H(y,p)| &= H(x,p) - H(y,p) \\ &\leq p\cdot\bar{u} - L(x,\bar{u}) - p\cdot\bar{u} + L(y,\bar{u}) \leq \alpha_R|x-y|. \end{aligned}$$

Choosing $\beta_\rho = \alpha_R$ yields our assertion. ■

To conclude we consider the case when L is uniformly convex and smooth.

Corollary A.1.7 *Let $L \in C^k(\Omega \times \mathbb{R}^n)$ with $k \geq 2$. Assume that L satisfies* (A.1.7) *and*

$$\frac{\partial^2 L}{\partial u}(x,u) \ \text{ is positive definite for all } (x,u) \in \Omega \times \mathbb{R}^n. \tag{A.1.9}$$

Then, H belongs to $C^k(\Omega \times \mathbb{R}^n)$. Moreover, if we denote by $u(x,p)$ the unique value of u at which the infimum in (A.1.8) *is attained, we have*

$$\frac{\partial H}{\partial p}(x,p) = u(x,p), \tag{A.1.10}$$

$$\frac{\partial H}{\partial x}(x,p) = -\frac{\partial L}{\partial x}(x,u(x,p)) \tag{A.1.11}$$

$$\frac{\partial^2 H}{\partial p^2}(x,p) = \left[\frac{\partial^2 L}{\partial u^2}(x,u(x,p))\right]^{-1} \tag{A.1.12}$$

In addition,

$$v = \frac{\partial H}{\partial p}(x,p) \text{ if and only if } p = \frac{\partial L}{\partial v}(x,v) \tag{A.1.13}$$

Proof — For any $x \in \Omega$, Theorem A.1.5 implies that $H(x, \cdot) \in C^k(\mathbb{R}^n)$ and that (A.1.13) holds. Since $u(x,p)$ satisfies

$$p - \frac{\partial L}{\partial u}(x, u(x,p)) = 0 \tag{A.1.14}$$

we see that (A.1.10) follows from (A.1.13) and that (A.1.12) is a consequence of (A.1.6). In addition, we obtain from the implicit function theorem and from assumption (A.1.9) that $u(x,p) \in C^{k-1}(\Omega \times \mathbb{R}^n)$. Then, recalling that $H(x,p) = p \cdot u(x,p) - L(x, u(x,p))$, we see that H is at least of class $C^{k-1}(\Omega \times \mathbb{R}^n)$, with $k - 1 \geq 1$. Consider now $x \in \Omega$, $y \in \mathbb{R}^m$ and $\lambda > 0$ small enough. We have

$$\begin{aligned}
&\frac{H(x+\lambda y, p) - H(x,p)}{\lambda} \\
&\quad \geq \frac{p \cdot u(x,p) - L(x+\lambda y, u(x,p)) - p \cdot u(x,p) + L(x, u(x,p))}{\lambda} \\
&\quad = \frac{L(x, u(x,p)) - L(x+\lambda y, u(x,p))}{\lambda}.
\end{aligned}$$

Letting $\lambda \to 0$ we obtain

$$\frac{\partial H}{\partial x}(x,p) \cdot y \geq -\frac{\partial L}{\partial x}(x, u(x,p)) \cdot y$$

for all $y \in \mathbb{R}^m$, which implies (A.1.11). The C^k regularity of H is then a consequence of (A.1.10), (A.1.11). ∎

A.2 Hausdorff measure and rectifiable sets

We recall the definition the definition of the ρ-dimensional Hausdorff measure of a set $S \subset \mathbb{R}^n$ for a given $\rho \in [0, \infty[$. First we define, for $\delta > 0$,

$$\mathcal{H}^\rho_\delta(S) = \frac{\alpha(\rho)}{2^\rho} \inf \left\{ \sum_{i=1}^{+\infty} (\operatorname{diam} C_i)^\rho \ : \ C_i \subset \mathbb{R}^n, \ S \subset \bigcup_{i=1}^{+\infty} C_i, \ \operatorname{diam} C_i < \delta \right\},$$

where

$$\alpha(\rho) = \frac{\pi^{\rho/2}}{\Gamma\left(\frac{\rho}{2} + 1\right)},$$

and $\Gamma(t) = \int_0^\infty e^{-x} x^{t-1}\, dx$ is Euler's Gamma function. Observe that, if ρ is an integer, then $\alpha(\rho)$ is the measure of a ρ–dimensional unit sphere.

Definition A.2.1 *Given $S \subset \mathrm{I\!R}^n$ and $\rho \geq 0$, the* (outer) ρ–dimensional Hausdorff measure *of S is defined as*

$$\mathcal{H}^\rho(S) = \lim_{\delta > 0} \mathcal{H}^\rho_\delta(S) = \sup_{\delta > 0} \mathcal{H}^\rho_\delta(S).$$

It can be proved (see e.g. [43]) that the measure defined above satisfies the following properties.

Theorem A.2.2

(i) *The measure $\mathcal{H}^\rho$ is Borel regular for all $\rho > 0$.*

(ii) *$\mathcal{H}^n$ coincides with the Lebesgue measure.*

(iii) *$\mathcal{H}^o$ is the counting measure.*

(iv) *If $A \subset \mathrm{I\!R}^n$ is a k–dimensional smooth manifold and $\mu(A)$ is the k–dimensional area of A, then $\mathcal{H}^k(A) = \mu(A)$.*

(v) *If $\mathcal{H}^\rho(S) < +\infty$ for some $\rho < n$, then $\mathcal{H}^\sigma(S) = 0$ for any $\sigma > \rho$.*

(vi) *If $\mathcal{H}^\rho(S) > 0$ for some $\rho > 0$, then $\mathcal{H}^\sigma(S) = +\infty$ for any $\sigma \in [0, \rho[$.*

(vii) *If $f \in \mathrm{Lip}\,(\mathrm{I\!R}^n; \mathrm{I\!R}^N)$, then*

$$\mathcal{H}^\rho(f(A)) \leq (\mathrm{Lip}\,(f))^\rho \mathcal{H}^\rho(A) \qquad \forall A \subset \mathrm{I\!R}^n\,. \tag{A.2.15}$$

The previous properties justify the following definition.

Definition A.2.3 *Given $S \subset \mathrm{I\!R}^n$, we define its* Hausdorff dimension *to be*

$$\mathcal{H}\text{–}\dim(S) = \inf\{\rho \geq 0 \ : \ \mathcal{H}^\rho(S) = 0\}.$$

Observe that $\mathcal{H}\text{–}\dim(S) \leq n$ by Theorem A.2.2(ii)-(v). Moreover

$$\mathcal{H}\text{–}\dim(S) = d \quad \Longrightarrow \quad \mathcal{H}^\rho(S) = 0,\ \forall \rho > s, \quad \mathcal{H}^\rho(S) = +\infty,\ \forall \rho < s.$$

Therefore

$$\begin{aligned} \mathcal{H}\text{–}\dim(S) &= \inf\{\rho \geq 0 \ : \ \mathcal{H}^\rho(S) < +\infty\} \\ &= \sup\{\rho \geq 0 \ : \ \mathcal{H}^\rho(S) > 0\}. \end{aligned}$$

A.3 BV functions

We recall here the definition and some basic properties of the class BV of the functions of bounded variation. Given an open set $\Omega \subset \mathbb{R}^N$ and a vector–valued function $w \in L^1(\Omega, \mathbb{R}^M)$, we denote by $D^{wk}w$ the distributional derivative of w, while $Dw(y)$ denotes, as usual, the gradient of w at a point y at which w is differentiable.

Definition A.3.1 *We say that a function $w \in L^1(\Omega, \mathbb{R}^M)$ belongs to the class $BV(\Omega, \mathbb{R}^M)$ of functions of bounded variation in Ω if the first partial derivatives of w in the sense of distributions are measures with finite total variation in Ω.*

The derivative of a function of bounded variation is usually decomposed into three parts. To introduce this decomposition we first need to recall some definitions from measure theory.

Definition A.3.2 *Given $w \in L^1(\Omega, \mathbb{R}^M)$ and $x_0 \in \Omega$, we say that $\bar{w} \in \mathbb{R}^M$ is the* approximate limit *of w at y_0 if*

$$\lim_{\rho \to 0^+} \frac{\mathrm{meas}\left(\{y \in B_\rho(y_0) \mid |w(y) - \bar{w}| > \varepsilon\}\right)}{\rho^N} = 0$$

for any $\varepsilon > 0$. In this case we write $\bar{w} = \mathrm{ap}\ \lim_{y \to y_0} w(y)$. More generally, given $\nu \in \mathbb{R}^N$, $|\nu| = 1$ we denote by $w_\nu(y_0)$ the value $\bar{w} \in \mathbb{R}^M$ (if it exists) such that

$$\lim_{\rho \to 0^+} \frac{\mathrm{meas}\left(\{y \in B_\rho(y_0) \cap \Omega_\nu(y_0) \mid |w(y) - \bar{w}| > \varepsilon\}\right)}{\rho^N} = 0$$

for any $\varepsilon > 0$, where

$$\Omega_\nu(y_0) = \{y \in \Omega \mid \langle y - y_0, \nu \rangle \geq 0\}.$$

Clearly that, if $\mathrm{ap}\lim_{y \to y_0} w(y)$ exists, then $w_\nu(y_0) = \mathrm{ap}\lim_{y \to y_0} w(y)$ for all ν. An interesting case is when the approximate limit does not exist, but we can find ν such that the two limits $w_\nu(y_0)$ and $w_{-\nu}(y_0)$ exist and are different. Such a behavior means, roughly speaking, that w has an approximate jump discontinuity at y_0 perpendicular to the direction ν.

Definition A.3.3 *If $w \in BV(\Omega, \mathbb{R}^M)$ we call the* jump set *of w the set*

$$S_w = \{y_0 \in \Omega \mid \ \nexists\ \mathrm{ap} \lim_{y \to y_0} w(y)\}.$$

Let $D^{wk}w = D^a w + D^s w$ be the Radon–Nikodym decomposition of $D^{wk}w$ in its absolutely continuous and singular part with respect to Lebesgue measure.

We denote by Jw the restriction of $D^s w$ to the jump set S_w and define $Cw = D^s w - Jw = D^{wk} w - D^a w - Jw$. The terms $D^a w$, Jw and Cw are called respectively the regular part, *the* jump part *and the* Cantor part *of $D^{wk} w$.*

Proposition A.3.4 *Let $w \in BV(\Omega, \mathrm{I\!R}^M)$. Then S_w is countably $\mathcal{H}^{N-1}$–rectifiable. Furthermore:*

(i) *if $Dw(y)$ exists for a.e. $y \in \Omega$, then it coincides with the density of $D^a w$;*

(ii) *there exists $\hat{\nu} : S_w \to \partial B_1$ such that $w_{\hat{\nu}(y)}(y)$, $w_{-\hat{\nu}(y)}(y)$ exist for $\mathcal{H}^{N-1}$–a.e. $y \in S_w$ and*

$$Jw = (w_{\hat{\nu}} - w_{-\hat{\nu}}) \otimes \hat{\nu}\, d\mathcal{H}^{N-1}|_{S_w}.$$

The unit vector field $\hat{\nu}$ is uniquely defined up to a set of zero $\mathcal{H}^{N-1}$ measure, and may be regarded as an approximate normal to S_w in the sense that, if G is a C^1 hypersurface with normal vector ν, then $\nu = \hat{\nu}$ on $S_w \cap G$ ($\mathcal{H}^{N-1}$ a.e.);

(iii) *if $F \subset \Omega$ is a countably $\mathcal{H}^{N-1}$–rectifiable set, then the restriction of Cw to F is zero.*

Proof — See [45], [43], [8], [9]. ∎

Bibliography

[1] ALBANO P., CANNARSA P., Singularities of semiconcave functions in Banach spaces, in *Stochastic Analysis, Control, Optimization and Applications*, W.M.McEneaney, G.G.Yin, Q.Zhang (eds.), Birkhäuser, Boston, 1999, 171–190.

[2] ALBANO P., CANNARSA P., Structural properties of singularities of semiconcave functions, *Annali Scuola Norm. Sup. Pisa Sci. Fis. Mat.* **28** (1999), 719–740.

[3] ALBANO P., CANNARSA P., Propagation of singularities for concave solutions of Hamilton–Jacobi equations, in *Equadiff 99*, World Scientific, Singapore, 2000.

[4] ALBANO P., CANNARSA P., Propagation of singularities for solutions of nonlinear first order PDEs, preprint, 2000.

[5] G. ALBERTI, On the structure of singular sets of convex functions, *Calc. Var. Partial Differential Equations* **2** (1994), 17–27.

[6] ALBERTI G., AMBROSIO L., CANNARSA P., On the singularities of convex functions, *Manuscripta Math.* **76** (1992), 421–435.

[7] AMBROSETTI A., PRODI G., *A primer of nonlinear analysis*, Cambridge University Press, Cambridge (1993).

[8] AMBROSIO L., Variational problems in SBV and image segmentation, *Acta Appl. Math.* **17** (1989), 1–40.

[9] AMBROSIO L., Existence theory for a new class of variational problems, *Arch. Rat. Mech. Anal.* **111** (1990), 291–322.

[10] AMBROSIO L., Geometric evolution problems, distance function and viscosity solutions, in *Calculus of variations and partial differential equations (Pisa, 1996)* (G. Buttazzo, A. Marino, M.K.V. Murthy eds.) Springer, Berlin, (2000).

[11] Ambrosio L., Cannarsa P., Soner H.M., On the propagation of singularities of semi–convex functions, *Ann. Scuola Normale Sup. Pisa Ser. IV* **20** (1993), 597–616.

[12] Anzellotti G., Ossanna E., Singular sets of convex bodies and surfaces with generalized curvatures, *Manuscripta Math.* **86** (1995), 417–433.

[13] Aronson D.G., The porous medium equation, in *Some problems on nonlinear diffusion* (Fasano A. and Primicerio M., Eds.), Lect. Notes Math. **1224**, Springer, 1986, 1–46.

[14] Aubin J.P., Cellina A., *Differential Inclusions*, Springer-Verlag, Grundlehren der Math. Wiss., (1984).

[15] Aubin J.P., Frankowska H., *Set-valued analysis*, Birkhäuser, Boston, 1990.

[16] Bardi M., A boundary value problem for the minimum time function, *SIAM J. Control Optim.* **27** (1989), 776–785.

[17] Bardi M., Capuzzo Dolcetta I., *Optimal control and viscosity solutions of Hamilton–Jacobi equations*, Birkhäuser, 1997.

[18] Barles G., *Solutions de viscosité des équations de Hamilton–Jacobi*, Springer Verlag, Berlin, 1994.

[19] Barron E.N., Cannarsa P., Jensen R., Sinestrari C., Regularity of Hamilton–Jacobi equations when forward is backward, *Indiana Univ. Math. J.* **48** (1999), 385–409.

[20] Bartke K., Berens H., Eine Beschreibung der Nichteindeutigkeitsmenge für die beste Approximation in der euklidischen Ebene, *J. Approx. Theory* **47** (1986), 54–74.

[21] Bellman R., *Dynamic programming*, Princeton University Press (1957).

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

[23] Cannarsa P., Frankowska H., Sinestrari C., Optimality conditions and synthesis for the minimum time problem, *Set-Valued Anal.*, **8** (2000) 127–148.

[24] Cannarsa P., Mennucci A., Sinestrari C., Regularity results for solutions of a class of Hamilton–Jacobi equations, *Arch. Rat. Mech. Analysis* **140**, 197–223 (1997).

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

[26] Cannarsa P., Sinestrari C., On a class of nonlinear time optimal control problems, *Discrete and Continuous Dynamical Systems* **1** (1995), 285–300.

[27] Cannarsa P., Soner H.M., On the singularities of the viscosity solutions to Hamilton–Jacobi–Bellman equations, *Indiana Univ. Math. J.* **36** (1987), 501–524.

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

[29] Capuzzo Dolcetta I., Ishii I., Approximate solutions of the Bellman equation of deterministic control theory, *Appl. Math. Optimization* **11** (1984), 161–181.

[30] Carathéodory C., *Calculus of variations and partial differential equations of the first order*, Teubner, Berlin, (1935).

[31] Clarke F. H.,*Optimization and nonsmooth analysis*, Wiley, New York, 1983.

[32] Clarke F.H., Ledyaev Y.S., Stern R.J., Wolenski P.R., *Nonsmooth analysis and control theory*, Graduate Texts in Mathematics, Springer, New York, 1998.

[33] Cole J., On a quasilinear parabolic equation occurring in aerodynamics, *Q. Appl. Math.* **9** (1951), 225–236.

[34] Conway E.D., Hopf E., Hamilton's theory and generalized solutions of the Hamilton-Jacobi equation, *J. Math. Mech.* **13** (1964), 939–986.

[35] Crandall M.G., Viscosity solutions: A primer, in *Viscosity solutions and applications* (I. Capuzzo Dolcetta, P.L. Lions editors), Springer Verlag, Berlin, 1997.

[36] Crandall M.G., Evans L.C., Lions P.L., Some properties of viscosity solutions of Hamilton–Jacobi equations, *Trans. Amer. Math. Soc.* **282** (1984), 487–502.

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

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

[39] Dafermos C.M., Generalized characteristics and the structure of solutions of hyperbolic conservation laws, *Indiana Univ. Math. J.* **26** (1977), 1097–1119.

[40] Douglis A., The continuous dependence of generalized solutions of non–linear partial differential equations upon initial data, *Comm. Pure Appl. Math.* **14** (1961), 267–284.

[41] Erdös P., Some remarks on the measurability of certain sets, *Bull. Amer. Math. Soc.* **51** (1945), 728–731.

[42] Evans L.C., *Partial Differential Equations*, A.M.S., (1998).

[43] Evans L.C., Gariepy R.F., *Measure theory and fine properties of functions*, Studies in Advanced Mathematics, CRC Press, Ann Arbor (1992).

[44] Falconer K., *Fractal Geometry*, Mathematical foundations and applications, Wiley, New York (1990).

[45] Federer H., *Geometric measure theory*, Springer Verlag, Berlin (1969).

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

[47] Fleming, W.H., Rishel, R. W., *Deterministic and stochastic optimal control*, Springer Verlag, New York (1975).

[48] Fleming W.H., Soner H.M., *Controlled Markov processes and viscosity solutions*, Springer Verlag, Berlin (1993).

[49] Hamilton R.S., A matrix Harnack estimate for the heat equation, *Comm. Anal. Geom.* **1** (1993), 113–126.

[50] HERMES H., LASALLE J.P., *Functional analysis and time optimal control*, New York, Academic Press (1969).

[51] HESTENES M.R., *Calculus of variations and optimal control theory*, Wiley, New York, (1966).

[52] HOPF E., The partial differential equation $u_t + uu_x = \mu u_{xx}$, *Comm. Pure. Appl. Math.* **3** (1950), 201–230.

[53] HOPF E., Generalized solutions of nonlinear equations of first order, *J. Math. Mech.* **14** (1965), 951–973.

[54] HRUSTALEV M.M., Necessary and sufficient optimality conditions in the form of Bellman's equation, *Soviet Math. Dokl.* **19** (1978), 1262–1266.

[55] ISHII H., Uniqueness of unbounded viscosity solutions of Hamilton–Jacobi equations, *Indiana Univ. Math. J.* **33** (1984), 721–748.

[56] ISHII H., Perron's method for Hamilton–Jacobi equations, *Duke Math. J.*, **55** (1987), 369–384.

[57] ISHII H., RAMASWAMY M., Uniqueness results for a class of Hamilton–Jacobi equations with singular coefficients, *Comm. Partial Differential Equat.* **20** (1995), 2187–2213.

[58] JENSEN R., The maximum principle for viscosity solutions of fully nonlinear second order partial differential equations, *Arch. Rat. Mech. Anal.* **101** (1988), 1–27.

[59] KRUZHKOV S.N., The Cauchy problem in the large for certain nonlinear first order differential equations, *Soviet. Math. Dokl.* **1** (1960), 474–477.

[60] KRUZHKOV S.N., The Cauchy problem in the large for nonlinear equations and for certain quasilinear systems of the first order with several variables, *Soviet. Math. Dokl.* **5** (1964), 493–496.

[61] KRUZHKOV S.N., First order quasilinear equations in several independent variables, *Mat. Sb.* **81** (1970), 228-255 (english tr.: *Math. USSR-Sb.* **10** (1970), 217-243).

[62] KRUZHKOV S.N., Generalized solutions of the Hamilton–Jacobi equations of the eikonal type I, *Math. USSR Sb.* **27** (1975), 406–445.

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

[64] Lasry J.M., Lions P.L., A remark on regularization in Hilbert spaces, *Israel J. Math* **55** (1986), 257–266.

[65] Lax P.D., Hyperbolic systems of conservation laws II, *Comm. Pure Appl. Math.* **10** (1957), 537–566.

[66] Li P., Yau S.-T., On the parabolic kernel of the Schrödinger operator, *Acta Math.* **156** (1968), 153–201.

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

[68] Lee E.B., Markus L., *Foundations of optimal control theory*, New York, John Wiley 1968.

[69] Motzkin T., Sur quelques propriétés caractéristiques des ensembles convexes, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.* **21** (1935), 562–567.

[70] Oleinik O., Discontinuous solutions of nonlinear differential equations. *Usp. Mat. Nauk* **12** (1957), 3–73. English transl. in *Amer. Math. Soc. Transl. Ser. 2* **26**, 95–172.

[71] Petrov N.N., On the Bellman function for the time-optimal process problem, *J. Appl. Math. Mech.* **34** (1970), 785–791.

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

[73] Rockafellar R.T., Favorable classes of Lipschitz continuous functions in subgradient optimization, in *Progress in Nondifferential Optimization*, (Nurminski E., Ed.), IIASA Collaborative Proceedings Series, Laxenburg, **125** (1982).

[74] Schaeffer D., A regularity theorem for conservation laws, *Adv. in Math.* **11** (1973), 368–386.

[75] Sinestrari C., Semiconcavity of solutions of stationary Hamilton–Jacobi equations, *Nonlinear Anal.* **24** (1995), 1321–1326.

[76] Sinestrari C., Local regularity properties of the minimum time function, in *PDE methods in control and shape analysis*, (Da Prato G., Zolesio J.P., Eds.), Marcel Dekker, New York (1996).

[77] Spingarn J., Submonotone subdifferentials of Lipschitz functions, *Trans. Amer. Math. Soc.* **264** (1981), 77–89.

[78] Stein E.M., *Singular integrals and differentiability properties of functions*, Princeton University Press, Princeton, NJ (1970).

[79] Subbotin A.I., *Generalized solutions of first order PDEs: the dynamic optimization perspective*, Birkhäuser, Boston, (1995).

[80] Veselý L., On the multiplicity points of monotone operators on separable Banach spaces, *Comment. Math. Univ. Carolin.* **27** (1986), 551–570.

[81] Veselý L., On the multiplicity points of monotone operators on separable Banach spaces II, *Comment. Math. Univ. Carolin.* **28** (1987), 295–299.

[82] Veselý L., A connectedness property of maximal monotone operators and its application to approximation theory, *Proc. Amer. Math. Soc.* **115** (1992), 663–667.

[83] Westphal U., Frerking J., On a property of metric projections onto closed subsets of Hilbert spaces, *Proc. Amer. Math. Soc.* **105** (1989), 644–651.

[84] Wu Z.-Q., The ordinary differential equation with discontinuous right–hand members and the discontinuous solutions of the quasilinear partial differential equations, *Acta Math. Sinica* **13** (1963), 515–530. English translation: *Scientia Sinica* **13** (1964), 1901–1907.

[85] Zajíček L., On the points of multiplicity of monotone operators, *Comment. Math. Univ. Carolin.* **19** (1978), 179–189.

[86] Zajíček L., On the differentiation of convex functions in finite and infinite dimensional spaces, *Czechoslovak Math. J.* **29** (1979), 340–348.