

## 5 Lipschitz Nonsmooth Optimization

---

Smooth and convex functions, which we discussed in the previous two chapters, are also (locally) Lipschitz. This chapter aims to introduce a more general framework, which provides optimality conditions for problems with Lipschitz data. We first present the theory of generalized gradients for locally Lipschitz functions, developed by Frank H. Clarke in the 1970's, and we then pass to a more general framework, by giving some comprehensive elements of the modern theory developed by Boris S. Mordukhovich and his collaborators in the last three decades. Of course, we limit our exposition to the finite dimensional spaces framework, but both of these theories can be developed in much more general situations. It is beyond the scope of this book to discuss the case of multifunctions (i.e., mappings with values given by sets), and of the associated generalized differentiation objects (derivatives and coderivatives, associated by means of corresponding tangent and normal cones), although many interesting results, which link the whole machinery of calculus for functions, multifunctions and tangent cones exist in literature.

---

### 5.1 Clarke Generalized Calculus

Consider the following optimization problem:

$$\min f(x_1, x_2) \quad \text{such that } x_1^2 + x_2^2 \leq 1, \quad (5.1.1)$$

where  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  is the function  $f(x_1, x_2) = \max \{ \min \{ 2x_1, x_2 \}, x_1 + 2x_2 \}$ . Remark that this function is neither convex nor smooth on the admissible set of points given by the unit ball. Therefore, the theory developed until now cannot apply. Remark that  $f$ , as well as the function which gives the restrictions (i.e.,  $g(x_1, x_2) := x_1^2 + x_2^2 - 1$ ), do exhibit nice regularity properties: they are both Lipschitz functions.

It is interesting to develop a theory which could solve such a problem. This is the main purpose of this chapter.

#### 5.1.1 Clarke Subdifferential

Consider a function  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  and suppose that  $f$  is locally Lipschitz with modulus  $L > 0$  around a point  $\bar{x} \in \mathbb{R}^p$ , which means that there is an  $\varepsilon > 0$  such that  $f$  is  $L$ -Lipschitz on  $D(\bar{x}, \varepsilon)$ , i.e.,

$$|f(x) - f(u)| \leq L \cdot \|x - u\|, \quad \forall x, u \in D(\bar{x}, \varepsilon).$$

One can define the (Clarke) generalized directional derivative of  $f$  at  $\bar{x}$  in the direction  $u$ , denoted  $f^\circ(\bar{x}, u)$ , as follows:

$$f^\circ(\bar{x}, u) := \limsup_{x \rightarrow \bar{x}, t \downarrow 0} \frac{f(x + tu) - f(x)}{t}. \tag{5.1.2}$$

Some useful basic properties of this function are given bellow.

**Proposition 5.1.1.** *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be a locally Lipschitz function with modulus  $L > 0$  around a point  $\bar{x} \in \mathbb{R}^p$ . Then*

(i) *The function  $u \mapsto f^\circ(\bar{x}, u)$  is finite, sublinear and  $L$ -Lipschitz on  $\mathbb{R}^p$ , and it satisfies*

$$|f^\circ(\bar{x}, u)| \leq L \|u\|, \quad \forall u \in \mathbb{R}^p. \tag{5.1.3}$$

Moreover,

$$f^\circ(\bar{x}, -u) = (-f)^\circ(\bar{x}, u), \quad \forall u \in \mathbb{R}^p. \tag{5.1.4}$$

(ii) *For every  $u \in \mathbb{R}^p$ , the function  $(x, v) \mapsto f^\circ(x, v)$  is upper semicontinuous at  $(\bar{x}, u)$ .*

*Proof* (i) By the Lipschitz condition of  $f$  around  $\bar{x}$ , it follows that for every  $x$  and positive  $t$  sufficiently close to  $\bar{x}$  and 0, respectively, one has

$$\left| \frac{f(x + tu) - f(x)}{t} \right| \leq \frac{L \|tu\|}{t} = L \|u\|,$$

which shows (5.1.3) and the fact that  $f^\circ(\bar{x}, \cdot)$  is everywhere finite.

Moreover, for every  $\lambda \geq 0$ , one has

$$\frac{f(x + t\lambda u) - f(x)}{t} = \lambda \frac{f(x + t\lambda u) - f(x)}{\lambda t},$$

which shows the positive homogeneity of  $f^\circ(\bar{x}, \cdot)$ , i.e.,

$$f^\circ(\bar{x}, \lambda u) = \lambda f^\circ(\bar{x}, u), \quad \forall \lambda \geq 0.$$

Also, for every  $u, v \in \mathbb{R}^p$ ,

$$\begin{aligned} f^\circ(\bar{x}, u + v) &= \limsup_{x \rightarrow \bar{x}, t \downarrow 0} \frac{f(x + tu + tv) - f(x)}{t} \\ &\leq \limsup_{x \rightarrow \bar{x}, t \downarrow 0} \frac{f(x + tu + tv) - f(x + tu)}{t} + \limsup_{x \rightarrow \bar{x}, t \downarrow 0} \frac{f(x + tu) - f(x)}{t}. \end{aligned}$$

But, since for any  $r, \delta > 0$ ,

$$\sup_{x \in B(\bar{x}, r), t \in (0, \delta)} \frac{f(x + tu + tv) - f(x + tu)}{t} \leq \sup_{y \in B(\bar{x}, r + \delta \|u\|), t \in (0, \delta)} \frac{f(y + tv) - f(y)}{t},$$

it follows, by letting  $r \rightarrow 0$ ,  $\delta \rightarrow 0$ , that

$$\limsup_{x \rightarrow \bar{x}, t \downarrow 0} \frac{f(x + tu + tv) - f(x + tu)}{t} \leq \limsup_{y \rightarrow \bar{x}, t \downarrow 0} \frac{f(y + tv) - f(y)}{t}$$

and, finally, that  $f^\circ(\bar{x}, \cdot)$  is subadditive, i.e.,

$$f^\circ(\bar{x}, u + v) \leq f^\circ(\bar{x}, u) + f^\circ(\bar{x}, v), \quad \forall u, v \in \mathbb{R}^p.$$

Concerning relation (5.1.4), observe that for every  $u \in \mathbb{R}^p$ ,

$$\begin{aligned} f^\circ(\bar{x}, -u) &= \limsup_{x \rightarrow \bar{x}, t \downarrow 0} \frac{f(x) - f(x - tu)}{t} = \limsup_{x \rightarrow \bar{x}, t \downarrow 0} \frac{(-f)(x) - (-f)(x - tu)}{t} \\ &= \limsup_{x \rightarrow \bar{x}, t \downarrow 0} \frac{(-f)((x - tu) + tu) - (-f)(x - tu)}{t} \\ &= \limsup_{y \rightarrow \bar{x}, t \downarrow 0} \frac{(-f)(y + tu) - (-f)(y)}{t} = (-f)^\circ(\bar{x}, u). \end{aligned}$$

Now, by the subadditivity and relations (5.1.3), (5.1.4), one gets that

$$f^\circ(\bar{x}, u) - f^\circ(\bar{x}, v) \leq f^\circ(\bar{x}, u - v) \leq L \|u - v\|, \quad \forall u, v \in \mathbb{R}^p,$$

which implies the expected Lipschitz property of  $f^\circ(\bar{x}, \cdot)$ .

(ii) Concerning the upper semicontinuity of the function  $(x, v) \mapsto f^\circ(x, v)$  at  $(\bar{x}, u)$ , take an arbitrary sequence  $(x_n, u_n)$  converging to  $(\bar{x}, u)$ . It follows that for each  $n$ ,

$$f^\circ(x_n, u_n) - \frac{1}{n} < \limsup_{x \rightarrow x_n, t \downarrow 0} \frac{f(x + tu_n) - f(x)}{t}.$$

Using the definition of the upper limit, one can find  $y_n \in \mathbb{R}^p$  with  $\|y_n - x_n\| < \frac{1}{n}$  and  $t_n \in (0, \frac{1}{n})$  such that

$$\begin{aligned} f^\circ(x_n, u_n) - \frac{1}{n} &< \frac{f(y_n + t_n u_n) - f(y_n)}{t_n} \\ &= \frac{f(y_n + t_n u) - f(y_n)}{t_n} + \frac{f(y_n + t_n u_n) - f(y_n + t_n u)}{t_n}. \end{aligned}$$

By the Lipschitz property of  $f$ , we get that the final term is smaller than  $L \|u_n - u\|$ . Passing to  $\limsup$  for  $n \rightarrow \infty$ , it follows that

$$\limsup_{n \rightarrow \infty} f^\circ(x_n, u_n) \leq f^\circ(\bar{x}, u),$$

which shows the desired upper semicontinuity.  $\square$

Given a nonempty subset  $A \subset \mathbb{R}^p$ , its support function is the function  $h_A : \mathbb{R}^p \rightarrow (-\infty, \infty]$  given as

$$h_A(u) := \sup \{ \langle \xi, u \rangle \mid \xi \in A \}.$$

Some properties of the support function, useful in the sequel, are given in the next result.

**Proposition 5.1.2.** (i) Let  $A$  be a nonempty subset of  $\mathbb{R}^p$ . Then  $h_A$  is positively homogeneous, subadditive, hence convex and continuous.

(ii) If  $A$  is convex and closed, then

$$\xi \in A \iff \langle \xi, u \rangle \leq h_A(u), \forall u \in \mathbb{R}^p.$$

(iii) If  $A$  and  $B$  are two nonempty, convex and closed subsets of  $\mathbb{R}^p$ , then

$$A \subset B \iff h_A(u) \leq h_B(u), \forall u \in \mathbb{R}^p. \tag{5.1.5}$$

(iv) If  $\varphi : \mathbb{R}^p \rightarrow \mathbb{R}$  is positively homogeneous, subadditive and bounded on the unit ball, then there is a uniquely defined nonempty, convex and compact subset  $A$  of  $\mathbb{R}^p$  such that  $\varphi = h_A$ , and the supremum is realized at every  $u \in \mathbb{R}^p$ .

*Proof* (i) It is straightforward from the definition that  $h_A$  is positively homogeneous and subadditive.

(ii) It is obvious that  $\xi \in A$  implies  $\langle \xi, u \rangle \leq h_A(u)$  for every  $u \in \mathbb{R}^p$ . Consider now  $\xi \notin A$ . Since  $\{\xi\}$  is convex, and  $A$  is convex and closed, it follows by applying the first argument of the proof of the separation Theorem 4.1.3, that there exist  $u \in \mathbb{R}^p$  and  $\alpha \in \mathbb{R}$  such that  $\langle \xi, u \rangle > \alpha > \langle \chi, u \rangle$ , for any  $\chi \in A$ . This means  $\langle \xi, u \rangle > h_A(u)$ , hence the converse implication is proved.

(iii) Again, the direct implication is straightforward. For the converse, take  $\xi \in A$  such that  $\xi \notin B$ . Using the proof of (ii), one gets that there is  $u \in \mathbb{R}^p$  such that  $\langle \xi, u \rangle > h_B(u)$ , which means that  $h_A(u) \geq \langle \xi, u \rangle > h_B(u)$ .

(iv) Given  $\varphi$ , define

$$A := \{ \xi \in \mathbb{R}^p \mid \langle \xi, u \rangle \leq \varphi(u), \forall u \in \mathbb{R}^p \}.$$

Clearly,  $A$  is a convex set. Now, because  $A$  can be written as

$$A = \bigcap_{u \in \mathbb{R}^p} \{ \xi \in \mathbb{R}^p \mid \langle \xi, u \rangle \leq \varphi(u) \},$$

it follows that  $A$  is closed as intersection of closed sets. Moreover, denoting by  $M > 0$  the boundedness constant of  $\varphi$  on the unit ball, one has, for any  $\xi \in A$  and any  $u \in D(0, 1)$ , that  $\langle \xi, u \rangle \leq \varphi(u) \leq M$ , hence  $\|\xi\| \leq M$ . So  $A$  is bounded, therefore compact. The inequality  $h_A \leq \varphi$  is obvious from the definition of  $A$ . Let us show the equality. Take arbitrary  $u \in \mathbb{R}^p$  and consider the linear subspace  $H := \{ \lambda u \mid \lambda \in \mathbb{R} \}$ . Pick  $\zeta \in \mathbb{R}^p$  such that  $\langle \zeta, u \rangle = \varphi(u)$ . For instance, one can take  $\zeta := 0$  if  $u = 0$ , and  $\zeta := \|u\|^{-2} \varphi(u)u$  if  $u \neq 0$ . Then, if  $\lambda \geq 0$ , one has  $\langle \zeta, \lambda u \rangle = \varphi(\lambda u)$ , and if  $\lambda < 0$ ,

$$\langle \zeta, \lambda u \rangle = -(-\lambda)\varphi(u) = -\varphi(-\lambda u) \leq \varphi(\lambda u).$$

In any case, the linear functional  $\langle \zeta, \cdot \rangle$  is majorized on  $H$  by the sublinear functional  $\varphi$ . By applying the Hahn-Banach Theorem, one gets that there is  $\xi \in \mathbb{R}^p$  such

that  $\langle \xi, v \rangle \leq \varphi(v)$  for any  $v \in \mathbb{R}^p$  and  $\langle \zeta, u \rangle = \langle \xi, u \rangle = \varphi(u)$ . It means that  $\xi \in A$  and  $\varphi(u) \leq h_A(u)$ . Since  $u$  was chosen arbitrarily, we have  $\varphi = h_A$ . The proof of the uniqueness follows by the use of (iii).  $\square$

Now, consider a locally Lipschitz function  $f$ , and take instead of the function  $\varphi$  from the previous proposition the generalized directional derivative  $f^\circ(\bar{x}, \cdot)$ , Clarke defined the generalized gradient as the nonempty compact set whose support function is  $f^\circ(\bar{x}, \cdot)$ .

**Definition 5.1.3.** Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be a locally Lipschitz function with modulus  $L > 0$  around a point  $\bar{x} \in \mathbb{R}^p$ . The Clarke generalized gradient, or the Clarke subdifferential of  $f$  at  $\bar{x}$  is the set

$$\partial_C f(\bar{x}) = \{ \xi \in \mathbb{R}^p \mid f^\circ(\bar{x}, u) \geq \langle \xi, u \rangle, \forall u \in \mathbb{R}^p \}. \tag{5.1.6}$$

Some basic properties of the Clarke subdifferential are given next.

**Proposition 5.1.4.** Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be a locally Lipschitz function with modulus  $L > 0$  around a point  $\bar{x} \in \mathbb{R}^p$ . Then

- (i)  $\partial_C f(\bar{x})$  is a nonempty, compact and convex set;
- (ii) One has

$$f^\circ(\bar{x}, u) = \max \{ \langle \xi, u \rangle \mid \xi \in \partial_C f(\bar{x}) \}. \tag{5.1.7}$$

*Proof* Both (i) and (ii) follow from Proposition 5.1.2.  $\square$

**Example 5.1.5.** Consider the function  $f : \mathbb{R}^p \rightarrow \mathbb{R}, f(x) = \|x\|$ . Of course, this function is Lipschitz with modulus 1. We know by Proposition 5.1.1 that  $f^\circ(0, u) \leq \|u\|$  for any  $u \in \mathbb{R}^p$ . Also, since

$$\frac{f(x + tu) - f(x)}{t} = \frac{\|x + tu\| - \|x\|}{t}, \forall x, u \in \mathbb{R}^p, \forall t > 0,$$

it follows, by passing to  $\limsup$  for  $x \rightarrow 0$ , that  $f^\circ(0, u) \geq \|u\|$ , for any  $u \in \mathbb{R}^p$ , hence  $f^\circ(0, u) = \|u\|$ , for any  $u \in \mathbb{R}^p$ . Then

$$\partial_C f(0) = \{ \xi \in \mathbb{R}^p \mid \|u\| \geq \langle \xi, u \rangle, \forall u \in \mathbb{R}^p \} = D(0, 1).$$

The next proposition provides a convergence result, which is useful in the subsequent sections.

**Proposition 5.1.6.** Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be locally Lipschitz around  $\bar{x}$ , and take the sequences  $(x_n)$  and  $(\xi_n)$  such that

$$x_n \rightarrow \bar{x} \text{ and } \xi_n \in \partial_C f(x_n) \text{ for every } n.$$

If  $\xi$  is a cluster point of  $(\xi_n)$  (in particular, if  $\xi_n \rightarrow \xi$ ), then  $\xi \in \partial_C f(\bar{x})$ .

*Proof* Consider  $u \in \mathbb{R}^p$ . Then, for every  $n$ , one has  $f^\circ(x_n, u) \geq \langle \xi_n, u \rangle$ . Denote the subsequence of  $(\xi_n)$  which converges to  $\xi$  by  $(\chi_n)$ . It follows that  $\langle \chi_n, u \rangle \rightarrow \langle \xi, u \rangle$  and, moreover, using the upper semicontinuity of  $f^\circ$  proven in Proposition 5.1.1, we deduce that  $f^\circ(x, u) \geq \langle \xi, u \rangle$ . Since  $u \in \mathbb{R}^p$  was arbitrarily chosen, we get the conclusion.  $\square$

Usually, the generalized directional derivative will not be computed directly, but, rather by taking the upper limit of the difference quotients (keep in mind the usual case of the derivative, where the same applies). This is quite clear from the definition: one must compute the upper limit of a quotient which contains two variables changing simultaneously. Nevertheless, we emphasize two cases where the situation is simpler, and which also justify the name of generalized gradient: the cases of smooth and convex functions.

**Theorem 5.1.7.** Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be a function.

- (i) If  $f$  is  $C^1$ , then  $f^\circ(x, u) = \nabla f(x)(u)$  for every  $x, u \in \mathbb{R}^p$ , and  $\partial_C f(x) = \{\nabla f(x)\}$ .
- (ii) If  $f$  is convex, then  $f^\circ(x, u) = f'(x, u)$  for every  $x, u \in \mathbb{R}^p$ , and  $\partial_C f(x) = \partial f(x)$ .

*Proof* (i) Remark, firstly, that because  $f$  is  $C^1$ , it is locally Lipschitz around any  $x \in \mathbb{R}^p$  (by the mean value theorem). Consider  $x, u \in \mathbb{R}^p$  and take  $x_n \rightarrow x$  and  $t_n \downarrow 0$  (that is,  $(t_n) \subset (0, \infty)$  and  $t_n \rightarrow 0$ ) two sequences for which  $\limsup$  in the definition of  $f^\circ(x, u)$  is attained, that is

$$\lim_{n \rightarrow \infty} \frac{f(x_n + t_n u) - f(x_n)}{t_n} = f^\circ(x, u).$$

By Lagrange Theorem, one can find  $y_n \in [x_n, x_n + t_n u]$  such that

$$f(x_n + t_n u) - f(x_n) = \nabla f(y_n)(t_n u) = t_n \nabla f(y_n)(u).$$

As  $f$  is  $C^1$ ,  $\nabla f(y_n) \rightarrow \nabla f(x)$  for  $n \rightarrow \infty$ , which finally gives

$$f^\circ(x, u) = \nabla f(x)(u), \quad \forall x, u \in \mathbb{R}^p.$$

Then

$$\partial_C f(x) = \{ \xi \in \mathbb{R}^p \mid \langle \xi, u \rangle \leq \nabla f(x)(u), \quad \forall u \in \mathbb{R}^p \},$$

which shows, due to the linearity of  $\nabla f(x)(u)$  (by changing  $u \rightarrow -u$ ), that in fact one has equality above, hence the conclusion.

(ii) Consider now the case of a convex function  $f$ . We know in this case that  $f$  is locally Lipschitz (Theorem 4.2.5), and also its directional derivative  $f'(x, u)$  always exists and is finite (Proposition 4.2.1). Because

$$f'(x, u) = \lim_{t \downarrow 0} \frac{f(x + tu) - f(x)}{t},$$

immediately follows that  $f'(x, u) \leq f^\circ(x, u)$ . On the other hand, since  $t \rightarrow \frac{f(y + tu) - f(y)}{t}$  is always increasing (see the proof of Proposition 4.2.1), using the definition of  $f^\circ(x, u)$ , one has

$$\begin{aligned} f^\circ(x, u) &= \inf_{r, \delta > 0} \sup_{\|y-x\| < r} \sup_{t \in (0, \delta)} \frac{f(y + tu) - f(y)}{t} \\ &= \inf_{r, \delta > 0} \sup_{\|y-x\| < r} \frac{f(y + \delta u) - f(y)}{\delta}. \end{aligned}$$

Due to the Lipschitz property of  $f$  (with modulus  $L$ ), it follows that

$$\begin{aligned} f(y + \delta u) - f(y) &= f(x + \delta u) - f(x) + [f(y + \delta u) - f(y) - f(x + \delta u) + f(x)] \\ &\leq f(x + \delta u) - f(x) + 2L \|y - x\| \leq f(x + \delta u) - f(x) + 2Lr, \end{aligned}$$

hence

$$f^\circ(x, u) \leq \inf_{r, \delta > 0} \left( \frac{f(x + \delta u) - f(x)}{\delta} + 2Lr \right) = f'(x, u).$$

It means that  $f'(x, u) = f^\circ(x, u)$ . Then

$$\begin{aligned} \partial_C f(x) &= \{ \xi \in \mathbb{R}^p \mid f^\circ(x, u) \geq \langle \xi, u \rangle \ \forall u \in \mathbb{R}^p \} \\ &= \{ \xi \in \mathbb{R}^p \mid f'(x, u) \geq \langle \xi, u \rangle \ \forall u \in \mathbb{R}^p \} \\ &= \partial f(x). \end{aligned} \quad \square$$

We now consider the general case, where the next theorem will provide a useful method to compute the generalized gradient of a locally Lipschitz function. We will use the celebrated theorem of Rademacher, which asserts that locally Lipschitz functions are differentiable almost everywhere (in the sense of the Lebesgue measure), i.e., the points where the function is not differentiable form a set whose Lebesgue measure is zero.

**Theorem 5.1.8.** *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be a locally Lipschitz function around a point  $x \in \mathbb{R}^p$ . Let  $\Omega \subset \mathbb{R}^p$  be any set of zero measure in  $\mathbb{R}^p$ , and  $\Omega_f$  be the set of points where  $f$  fails to be differentiable. Then*

$$\partial_C f(x) = \text{conv} \left\{ \lim_{n \rightarrow \infty} \nabla f(x_n) \mid x_n \rightarrow x, x_n \notin \Omega \cup \Omega_f \right\}. \quad (5.1.8)$$

*Proof* Remark that, by the Rademacher Theorem, the measure of  $\Omega \cup \Omega_f$  is zero, hence there are sequences which satisfy the conditions from (5.1.8). Since  $f$  is locally Lipschitz around  $x$  (we denote its Lipschitz modulus by  $L$ ), we deduce that its differential satisfies

$$\nabla f(y)(u) = \lim_{t \rightarrow 0} \frac{f(y + tu) - f(y)}{t} \leq L \|u\|$$

for every  $y$  near  $x$  where  $f$  is differentiable, hence

$$\|\nabla f(y)\| \leq L,$$

for any such  $y$ . Consequently, if  $x_n \rightarrow x$ ,  $x_n \notin \Omega \cup \Omega_f$ , one can extract a subsequence of  $(x_n)$ , denoted also by  $(x_n)$  for simplicity, for which  $(\nabla f(x_n))$  converges to a limit  $\xi$ . By the very definitions of the differential and of the generalized directional derivative, it follows that

$$\begin{aligned} \langle \nabla f(x_n), u \rangle &= \nabla f(x_n)(u) = \lim_{t \rightarrow 0} \frac{f(x_n + tu) - f(x_n)}{t} \\ &\leq \limsup_{y \rightarrow x_n, t \downarrow 0} \frac{f(y + tu) - f(y)}{t} = f^\circ(x_n, u), \end{aligned}$$

for any  $n \in \mathbb{N}$  and  $u \in \mathbb{R}^p$ . Using the upper semicontinuity of  $f^\circ(x_n, u)$  at  $(x, u)$ , and passing to the limit for  $n \rightarrow \infty$  above, it follows that

$$\langle \xi, u \rangle \leq f^\circ(x, u)$$

for any  $u \in \mathbb{R}^p$ . Hence  $\xi \in \partial_c f(x)$ , and in fact

$$A := \left\{ \lim_{n \rightarrow \infty} \nabla f(x_n) \mid x_n \rightarrow x, x_n \notin \Omega \cup \Omega_f \right\} \subset \partial_c f(x), \quad (5.1.9)$$

since  $\xi$  is an arbitrary limit of the kind taken in  $A$ . Since  $\partial_c f(x)$  is a convex set, it contains also the convex hull of  $A$ .

For the reverse inclusion, we will show that for any  $u \neq 0$  in  $\mathbb{R}^p$ ,

$$f^\circ(x, u) \leq \limsup_{y \rightarrow x, y \notin \Omega \cup \Omega_f} \nabla f(y)(u).$$

For this, denote the lim sup in the right-hand side by  $\ell$ , and take arbitrary  $\varepsilon > 0$ . Since

$$\limsup_{y \rightarrow x, y \notin \Omega \cup \Omega_f} \nabla f(y)(u) = \inf_{r > 0} \sup_{\substack{y \in B(x, r) \\ y \notin \Omega \cup \Omega_f}} \nabla f(y)(u) < \ell + \varepsilon,$$

it follows that there exists  $r > 0$  such that

$$\nabla f(y)(u) < \ell + \varepsilon, \quad \forall y \in B(x, r), y \notin \Omega \cup \Omega_f.$$

Without loss of generality, we may suppose that  $r$  is sufficiently small such that  $f$  is Lipschitz on  $B(x, r)$ . By the definition of the differential, we know that

$$\nabla f(y)(u) = \lim_{t \rightarrow 0} \frac{f(y + tu) - f(y)}{t} < \ell + \varepsilon, \quad \forall y \in B(x, r), y \notin \Omega \cup \Omega_f.$$

Then there exists  $\delta > 0$  such that, for any  $t \in (0, \delta)$ , and any  $y \in B(x, r), y \notin \Omega \cup \Omega_f$ , one has

$$f(y + tu) - f(y) < t(\ell + \varepsilon).$$



Since  $f$  is continuous, and the above inequality is true for any  $y$  in  $B(x, r)$  with the exception of a set of zero measure, it follows that the same inequality, but nonstrict, is satisfied by the all points  $y$  in  $B(x, r)$ . Consequently,

$$f(y + tu) - f(y) \leq t(\ell + \varepsilon), \quad \forall t \in (0, \delta), y \in B(x, r),$$

hence

$$f^\circ(x, u) \leq \ell + \varepsilon.$$

As  $\varepsilon$  was taken arbitrarily, it follows  $f^\circ(x, u) \leq \ell = \limsup_{y \rightarrow x, y \in \Omega \cup \Omega_f} \nabla f(y)(u)$ . Then for any realizing sequence (for which the lim sup is attained)  $x_n \rightarrow x, x_n \notin \Omega \cup \Omega_f$ , and for any  $u \in \mathbb{R}^p \setminus \{0\}$ , one has that

$$f^\circ(x, u) \leq \left\langle \lim_{n \rightarrow \infty} \nabla f(x_n), u \right\rangle \leq h_A(u) = h_{\text{conv } A}(u),$$

where the last equality is easy to prove for any set  $A$ . Since  $f^\circ(x, \cdot)$  is  $h_{\partial_c f(x)}$ , the conclusion follows from Proposition 5.1.2 (iii). □

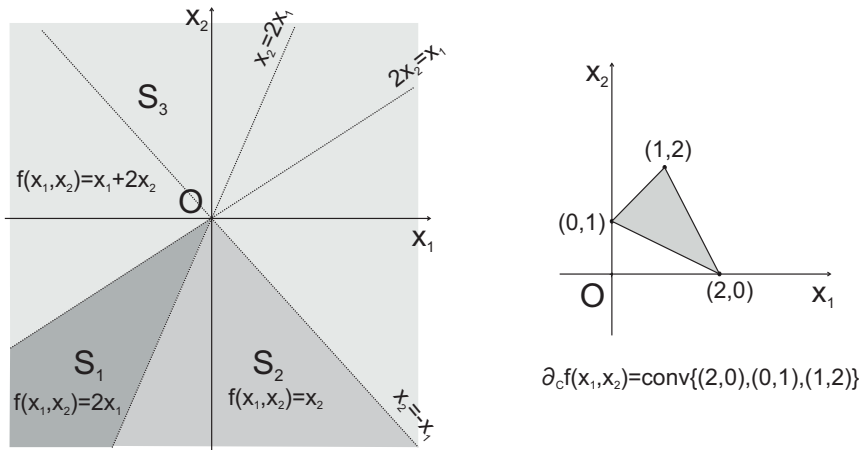


Figure 5.1: The domain of  $f$  and its Clarke subdifferential at  $(0, 0)$ .

**Example 5.1.9.** Coming back to the problem (5.1.1) from the beginning of this chapter, observe that one can use the gradient formula in order to compute  $\partial_c f(0, 0)$  for the function  $f(x_1, x_2) = \max \{ \min \{ 2x_1, x_2 \}, x_1 + 2x_2 \}$ . The function  $f$  can equivalently be written as

$$f(x_1, x_2) = \begin{cases} 2x_1, & \text{if } (x_1, x_2) \in S_1 := \{(x_1, x_2) \mid 2x_1 \leq x_2 \text{ and } x_1 \geq 2x_2\} \\ x_2, & \text{if } (x_1, x_2) \in S_2 := \{(x_1, x_2) \mid 2x_1 \geq x_2 \text{ and } x_1 \leq -x_2\} \\ x_1 + 2x_2, & \text{if } (x_1, x_2) \in S_3 := \{(x_1, x_2) \mid x_1 \leq 2x_2 \text{ or } x_1 \geq -x_2\}. \end{cases} \tag{5.1.10}$$

Remark, moreover, that  $S_1 \cup S_2 \cup S_3 = \mathbb{R}^2$ , and the union of their boundaries,  $A$ , is a set of zero measure. Moreover,  $f$  is differentiable on  $\mathbb{R}^2 \setminus A$ , the gradient having three possible values:  $(2, 0)$ ,  $(0, 1)$  and  $(1, 2)$ . According to the gradient formula, one gets that  $\partial_C f(0, 0)$  is the triangle obtained by the convex hull of these three points (see Figure 5.1).

**Remark 5.1.10.** Observe that  $\partial_C f(x)$  can fail to be a singleton even  $f$  is differentiable. An example is provided by the function  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x}, & \text{if } x \neq 0 \\ 0, & \text{if } x = 0. \end{cases} \tag{5.1.11}$$

One can see that  $f$  is differentiable at 0 and  $f'(0) = 0$ . Moreover,

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

which is not continuous. Observe also that, if  $x_n \rightarrow 0$ , then  $2x_n \sin \frac{1}{x_n} \rightarrow 0$ , hence for every such sequence, if it exists,  $\lim_{n \rightarrow \infty} f'(x_n) \in [-1, 1]$ . Furthermore, for  $y_n := \frac{1}{2n\pi} \rightarrow 0$  and  $z_n := \frac{1}{(2n+1)\pi} \rightarrow 0$ , one gets respectively  $\lim_{n \rightarrow \infty} f'(y_n) = -1$ , and  $\lim_{n \rightarrow \infty} f'(z_n) = 1$ . In conclusion, by the gradient formula, we obtain that  $\partial_C f(0) = [-1, 1]$ . This example shows that the generalized gradient extends the notion of continuous differentiability.

The next property will be useful in the development of the generalized gradient calculus.

**Definition 5.1.11.** A function  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  is said to be regular at  $x$  if it is locally Lipschitz around  $x$  and admits directional derivatives  $f'(x, u)$  (formally defined as in the convex case) satisfying  $f'(x, u) = f^\circ(x, u)$  for any  $u \in \mathbb{R}^p$ .

Theorem 5.1.7 shows that convex and  $C^1$  functions are regular at any point. One can see that if  $f, g$  are regular functions at  $x$  and  $\lambda \geq 0$ , then  $\lambda f$  and  $f+g$  are regular at  $x$ . The first assertion easily follows, and the second one will be also proven (see the proof of Theorem 5.1.13). A first example of a function which is not regular is given by (5.1.11). This is because  $f'(0, 1) = \nabla f(0)(1) = 0$ , but  $f^\circ(0, 1) = \max \{ \xi \mid \xi \in \partial_C f(0) = [-1, 1] \} = 1$ . Let us now provide another example of a function which is not regular.

**Example 5.1.12.** Consider  $f : \mathbb{R} \rightarrow \mathbb{R}$  given by  $f(x) := -|x|$  for any  $x \in \mathbb{R}$ . This function is not regular at 0. Observe first that it is Lipschitz and  $f'(0, u) = -|u|$  for any  $u \in \mathbb{R}$ . Moreover, since  $f$  is differentiable on  $\mathbb{R} \setminus \{0\}$ , with  $\nabla f(x)$  being either  $-1$  or  $1$  at any point different from 0, one gets that  $\partial_C f(0) = [-1, 1]$  in view of the gradient formula. Formula (5.1.7) shows that  $f^\circ(0, 1) = 1 \neq -1 = f'(0, 1)$ .

In fact, the example above is a particular instance of the following general case: every concave function that has a corner at  $x$  fails to be regular at  $x$ . This is because, for such a situation,

$$f'(x, u) \neq -f'(x, -u) = (-f)'(x, -u) = (-f)^\circ(x, -u) = f^\circ(x, u).$$

An important calculus rule is provided next.

**Theorem 5.1.13** (sum rule). *Let  $f, g : \mathbb{R}^p \rightarrow \mathbb{R}$  be two functions which are Lipschitz around  $x$ . Then*

$$\partial_C(f + g)(x) \subset \partial_C f(x) + \partial_C g(x). \quad (5.1.12)$$

*Equality holds if  $f$  and  $g$  are regular at  $x$ .*

*Proof* Observe that the definition of the upper limit allows us to write

$$\begin{aligned} (f + g)^\circ(x, u) &= \limsup_{y \rightarrow x, t \downarrow 0} \frac{(f + g)(y + tu) - (f + g)(y)}{t} \\ &\leq \limsup_{y \rightarrow x, t \downarrow 0} \frac{f(y + tu) - f(y)}{t} + \limsup_{y \rightarrow x, t \downarrow 0} \frac{g(y + tu) - g(y)}{t} \\ &= f^\circ(x, u) + g^\circ(x, u), \quad \forall u \in \mathbb{R}^p. \end{aligned} \quad (5.1.13)$$

Now, take  $\chi \in \partial_C(f + g)(x)$ . This means that  $\langle \chi, u \rangle \leq (f + g)^\circ(x, u) \leq f^\circ(x, u) + g^\circ(x, u)$ , for any  $u$ . Using (5.1.7), there exists  $\xi \in \partial_C f(x)$  such that  $\langle \xi, u \rangle = f^\circ(x, u)$ . Therefore,  $\langle \chi - \xi, u \rangle \leq g^\circ(x, u)$  for any  $u$ , which proves that  $\chi - \xi \in \partial_C g(x)$ , i.e.,  $\chi \in \partial_C f(x) + \partial_C g(x)$ .

Suppose now that  $f$  and  $g$  are regular. Then, using the regularity and the definitions of the Clarke generalized derivative and of the directional derivative, one gets

$$f^\circ(x, u) + g^\circ(x, u) = f'(x, u) + g'(x, u) = (f + g)'(x, u) \leq (f + g)^\circ(x, u), \quad \forall u \in \mathbb{R}^p.$$

Combined with (5.1.13), this proves that  $f + g$  is regular at  $x$ . Now, for any  $\xi \in \partial_C f(x)$ ,  $\zeta \in \partial_C g(x)$ , we have from definitions that  $\langle \xi, u \rangle \leq f^\circ(x, u)$ ,  $\langle \zeta, u \rangle \leq g^\circ(x, u)$  for any  $u$ , hence  $\langle \xi + \zeta, u \rangle \leq f^\circ(x, u) + g^\circ(x, u) \leq (f + g)^\circ(x, u)$ . It follows that  $\xi + \zeta \in \partial_C(f + g)(x)$ , which ends the proof.  $\square$

**Remark 5.1.14.** *Since for any scalar  $\lambda \in \mathbb{R}$ , we have  $\partial_C(\lambda f)(x) = \lambda \partial_C f(x)$ , it follows that for any two locally Lipschitz functions around  $x$ , and for any scalars  $\alpha, \beta \in \mathbb{R}$ , one has*

$$\partial_C(\alpha f + \beta g)(x) \subset \alpha \partial_C f(x) + \beta \partial_C g(x). \quad (5.1.14)$$

*Moreover, since the regularity is preserved by summation and multiplication with positive scalars, we deduce that for any two regular functions  $f, g$ , and any positive scalars  $\alpha, \beta \geq 0$ ,*

$$\partial_C(\alpha f + \beta g)(x) = \alpha \partial_C f(x) + \beta \partial_C g(x).$$

**Remark 5.1.15.** An easy example of strict containment in (5.1.12) is given by the functions  $f, g : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) := -|x|$ ,  $g(x) := |x|$  for any  $x$ . We saw before that  $f$  is not regular at 0 and  $\partial_c f(0) = [-1, 1]$ . One can easily prove (see Example 5.1.5) that  $\partial_c g(0) = [-1, 1]$ . Then

$$\{0\} = \partial_c(f + g)(0) \subsetneq \partial_c f(0) + \partial_c g(0) = [-2, 2].$$

The next important result is a generalized Fermat rule.

**Theorem 5.1.16** (Fermat rule for Clarke calculus). *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be a locally Lipschitz function. If  $f$  has a local minimum or maximum at  $x$ , then*

$$0 \in \partial_c f(x).$$

*Proof* One has, if  $x$  is a local minimum, that

$$f^\circ(x, u) = \limsup_{y \rightarrow x, t \downarrow 0} \frac{f(y + tu) - f(y)}{t} \geq \limsup_{t \downarrow 0} \frac{f(x + tu) - f(x)}{t} \geq 0.$$

On the other hand, if  $x$  is a local maximum,

$$f^\circ(x, u) = \limsup_{y \rightarrow x, t \downarrow 0} \frac{f(y + tu) - f(y)}{t} \geq \limsup_{t \downarrow 0} \frac{f(x) - f(x - tu)}{t} \geq 0.$$

Hence, in both cases,  $f^\circ(x, u) \geq 0$  for any  $u$ , which gives the conclusion. □

The next result is a mean value theorem for locally Lipschitz functions due to Gérard Lebourg, see (Lebourg, 1975).

**Theorem 5.1.17** (Lebourg). *Let  $x, y \in \mathbb{R}^p$ , and suppose that  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  is locally Lipschitz on a neighborhood of the line segment joining the points  $x$  and  $y$ . Then there exists  $z \in (x, y)$  such that*

$$f(y) - f(x) \in \langle \partial_c f(z), y - x \rangle.$$

*Proof* Denote by  $x_t := x + t(y - x)$  and consider the function  $g : [0, 1] \rightarrow \mathbb{R}$  given by

$$g(t) := f(x_t).$$

Observe that for any  $t_1, t_2 \in [0, 1]$ ,

$$\begin{aligned} |g(t_1) - g(t_2)| &= |f(x + t_1(y - x)) - f(x + t_2(y - x))| \\ &\leq \ell \|y - x\| \cdot |t_1 - t_2|, \end{aligned}$$

where  $\ell$  is the Lipschitz constant of  $f$ , hence  $g$  is Lipschitz on  $[0, 1]$ .

Now, for every  $v \in \mathbb{R}$ ,

$$\begin{aligned} g^\circ(t, v) &= \limsup_{s \rightarrow t, \lambda \downarrow 0} \frac{g(s + \lambda v) - g(s)}{\lambda} \\ &= \limsup_{s \rightarrow t, \lambda \downarrow 0} \frac{f(x + (s + \lambda v)(y - x)) - f(x + s(y - x))}{\lambda} \\ &\leq \limsup_{z \rightarrow x_t, \lambda \downarrow 0} \frac{f(z + \lambda v(y - x)) - f(z)}{\lambda} = f^\circ(x_t, v(y - x)). \end{aligned}$$

But this means that for any  $\xi \in \partial_C g(x)$  and any  $v \in \mathbb{R}$ ,

$$\xi v \leq g^\circ(t, v) \leq f^\circ(x_t, v(y - x)) = \max v \cdot \langle \partial_C f(x_t), y - x \rangle.$$

Observe that  $\langle \partial_C f(x_t), y - x \rangle$  is a compact interval of reals. Take above  $v = \pm 1$  to deduce that there is  $\chi \in \partial_C f(x_t)$  such that  $\xi = \langle \chi, y - x \rangle$ . This gives  $\partial_C g(x) \subset \langle \partial_C f(x_t), y - x \rangle$ .

Consider now the function  $h : [0, 1] \rightarrow \mathbb{R}$  given as

$$h(t) := g(t) - t(f(y) - f(x)).$$

One easily gets that  $h$  is locally Lipschitz and, by Theorem 5.1.13 and Remark 5.1.14, that

$$\partial_C h(t) \subset \partial_C g(t) - (f(y) - f(x)) \subset \langle \partial_C f(x_t), y - x \rangle - f(y) + f(x). \quad (5.1.15)$$

Since  $h(0) = h(1) = f(x)$ , one gets by the continuity of  $h$  that there is  $t_0 \in (0, 1)$  such that  $t_0$  is a local minimum or a local maximum for  $h$ . In view of the Fermat rule (Theorem 5.1.16), one has

$$0 \in \partial_C h(t_0),$$

which gives, by the use of (5.1.15), the conclusion of the theorem with  $z := x_{t_0}$ .  $\square$

For a function  $f : \mathbb{R}^p \rightarrow \mathbb{R}^k$  and a vector  $\xi \in \mathbb{R}^k$ , define the function  $\langle \xi, f \rangle : \mathbb{R}^p \rightarrow \mathbb{R}$  as

$$\langle \xi, f \rangle(x) := \langle \xi, f(x) \rangle. \quad (5.1.16)$$

**Theorem 5.1.18** (chain rule 1). *Let  $f := (f_1, \dots, f_k) : \mathbb{R}^p \rightarrow \mathbb{R}^k$  and  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  be locally Lipschitz around  $x$  and  $f(x)$ , respectively. Then the function  $h := g \circ f : \mathbb{R}^p \rightarrow \mathbb{R}$  is locally Lipschitz around  $x$ , and*

$$\partial_C h(x) \subset \text{cl conv} \{ \partial_C \langle y, f \rangle(x) \mid y \in \partial_C g(f(x)) \}. \quad (5.1.17)$$

*Moreover, if  $f_1, \dots, f_k$  are regular at  $x$ ,  $g$  is also regular at  $f(x)$ , and  $\partial_C g(f(x)) \subset \mathbb{R}_+^k$ , then  $h$  is regular at  $x$  and equality holds in (5.1.17).*

*Proof* The fact that  $h$  is locally Lipschitz around  $x$  is straightforward.

Pick  $\xi \in \partial_C h(x)$ , and take arbitrary  $u \in \mathbb{R}^p$ . Take some realizing sequences  $y_n \rightarrow x$ ,  $t_n \downarrow 0$  for  $h^\circ(x, u)$ , that is

$$h^\circ(x, u) = \lim_{n \rightarrow \infty} \frac{g(f(y_n + t_n u)) - g(f(y_n))}{t_n}.$$

By Theorem 5.1.17, for every  $n$  sufficiently large, there is  $z_n \in [f(y_n), f(y_n + t_n u)]$  such that

$$g(f(y_n + t_n u)) - g(f(y_n)) \in \langle \partial_C g(z_n), f(y_n + t_n u) - f(y_n) \rangle.$$

It follows that there is  $\chi_n \in \partial_C g(z_n)$  such that

$$\frac{h(y_n + t_n u) - h(y_n)}{t_n} = \frac{\langle \chi_n, f(y_n + t_n u) - f(y_n) \rangle}{t_n} = \left\langle \chi_n, \frac{f(y_n + t_n u) - f(y_n)}{t_n} \right\rangle.$$

Denote by  $B(f(x), 2\varepsilon)$  a neighborhood of  $f(x)$  where  $g$  is Lipschitz and by  $L_g$  its Lipschitz modulus. As  $z_n \rightarrow f(x)$  for  $n \rightarrow \infty$ , it follows that for every  $n$  sufficiently large,  $z_n \in B(f(x), \varepsilon)$ , and  $g$  is Lipschitz with modulus  $L_g$  on  $B(z_n, \varepsilon)$ . Then, by relation (5.1.3) applied for  $z_n$  and the definition of the generalized gradient, one has for any  $n$  sufficiently large and any  $v \in \mathbb{R}^k$  that

$$\langle \chi_n, v \rangle \leq g^\circ(z_n, v) \leq L_g \|v\|.$$

It follows that  $\|\chi_n\| \leq L_g$  for any  $n$  sufficiently large, hence  $(\chi_n)$  is bounded. Without loss of generality, we may suppose that  $\chi_n \rightarrow \chi$ . As  $z_n \rightarrow f(x)$  and  $\chi_n \in \partial_C g(z_n)$  for any  $n$  sufficiently large, we obtain that  $\chi \in \partial_C g(f(x))$ .

Also by Theorem 5.1.17, applied for  $\langle \chi, f \rangle$ , we get that there is  $p_n \in [y_n, y_n + t_n u]$  such that

$$\langle \chi, f \rangle(y_n + t_n u) - \langle \chi, f \rangle(y_n) \in \langle \partial_C \langle \chi, f \rangle(p_n), t_n u \rangle,$$

hence there is  $q_n \in \partial_C \langle \chi, f \rangle(p_n)$  such that

$$\left\langle \chi, \frac{f(y_n + t_n u) - f(y_n)}{t_n} \right\rangle = \frac{\langle \chi, f \rangle(y_n + t_n u) - \langle \chi, f \rangle(y_n)}{t_n} = \langle q_n, u \rangle.$$

Since  $y_n \rightarrow x$  and  $t_n \downarrow 0$ , it follows that  $p_n \rightarrow x$ . Reasoning as above for the locally Lipschitz function  $\langle \chi, f \rangle$ , it follows that  $(q_n)$  is bounded, so we may suppose again, without loss of generality, that  $q_n \rightarrow q$ . Therefore,  $q \in \partial_C \langle \chi, f \rangle(x)$ .

Combining the previous relations, we get

$$\begin{aligned} \frac{h(y_n + t_n u) - h(y_n)}{t_n} &= \left\langle \chi_n, \frac{f(y_n + t_n u) - f(y_n)}{t_n} \right\rangle \\ &= \left\langle \chi_n - \chi, \frac{f(y_n + t_n u) - f(y_n)}{t_n} \right\rangle + \left\langle \chi, \frac{f(y_n + t_n u) - f(y_n)}{t_n} \right\rangle \\ &= \left\langle \chi_n - \chi, \frac{f(y_n + t_n u) - f(y_n)}{t_n} \right\rangle + \langle q_n, u \rangle. \end{aligned}$$

Observe that the Lipschitz property of  $f$  implies that  $(f(y_n + t_n u) - f(y_n))/t_n$  is bounded, hence passing to the limit we get

$$h^\circ(x, u) = \langle q, u \rangle,$$

where  $q \in \partial_C \langle \chi, f \rangle(x)$  and  $\chi \in \partial_C g(f(x))$ .

This means that for any  $u \in \mathbb{R}^p$ ,  $h^\circ(x, u) \leq h_A(u)$ , where by  $A$  we have denoted the set  $\text{cl conv} \{ \partial_C \langle y, f \rangle(x) \mid y \in \partial_C g(f(x)) \}$ , and by  $h_A$  its support function. The conclusion follows by taking into account the equivalence (5.1.5) from Proposition 5.1.2.

Now, suppose that  $f_1, \dots, f_k$  are regular at  $x$ ,  $g$  is regular at  $f(x)$ , and  $\partial_C g(f(x)) \subset \mathbb{R}_+^k$ . Take arbitrary  $u \in \mathbb{R}^p$  and observe that

$$\begin{aligned} h'(x, u) &= \lim_{t \downarrow 0} \frac{g(f(x + tu)) - g(f(x))}{t} \\ &= - \lim_{t \downarrow 0} \frac{g(f(x) + tf'(x, u)) - g(f(x + tu))}{t} + g'(f(x), f'(x, u)), \end{aligned}$$

where by  $f'(x, u)$  we have denoted the vector whose components are  $f'_i(x, u)$ ,  $i = \overline{1, k}$ .

Since

$$\begin{aligned} \left| \frac{g(f(x) + tf'(x, u)) - g(f(x + tu))}{t} \right| &\leq \frac{1}{t} L_g \|tf'(x, u) - (f(x + tu) - f(x))\| \\ &= L_g \left\| f'(x, u) - \frac{f(x + tu) - f(x)}{t} \right\| \rightarrow 0 \end{aligned}$$

when  $t \downarrow 0$ , it means that

$$h'(x, u) = g'(f(x), f'(x, u)). \quad (5.1.18)$$

Now, take  $y \in \partial_C g(f(x)) \subset \mathbb{R}_+^k$  and remark that the function  $\langle y, f \rangle$  is regular at  $x$ , since

$$\langle y, f \rangle(x) = y_1 f_1(x) + \dots + y_k f_k(x),$$

i.e.,  $\langle y, f \rangle$  is a positive linear combination of regular functions. Then

$$\begin{aligned} \max \{ \langle y, f \rangle^\circ(x, u) \mid y \in \partial_C g(f(x)) \} &= \max \{ \langle y, f \rangle'(x, u) \mid y \in \partial_C g(f(x)) \} \\ &= \max \{ \langle y, f'(x, u) \rangle \mid y \in \partial_C g(f(x)) \} \\ &= g^\circ(f(x), f'(x, u)) \\ &= g'(f(x), f'(x, u)) \text{ (since } g \text{ is regular)} \\ &= h'(x, u) \text{ (from (5.1.18))} \\ &\leq h^\circ(x, u). \end{aligned}$$

It means also from (5.1.5) that for any  $y \in \partial_C g(f(x))$ ,

$$\partial_C \langle y, f \rangle(x) \subset \partial_C h(x).$$

But this shows that

$$\{\partial_C \langle y, f \rangle (x) \mid y \in \partial_C g(f(x))\} \subset \partial_C h(x),$$

and, as the set from the right-hand side is convex and closed, one has equality in (5.1.17).  $\square$

**Theorem 5.1.19** (chain rule 2). *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}^k$  be locally Lipschitz around  $x$  and  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  be continuously differentiable at  $f(x)$ . Then the function  $h := g \circ f : \mathbb{R}^p \rightarrow \mathbb{R}$  is locally Lipschitz around  $x$ , and*

$$\partial_C h(x) = \partial_C \langle \nabla g(f(x)), f \rangle (x). \tag{5.1.19}$$

*Proof* Take  $y$  and  $t > 0$  near  $x$  and 0, respectively, and observe that the Lagrange Theorem applied for  $g$  asserts that there is  $z \in [f(y), f(y + tu)]$  such that

$$g(f(y + tu)) = g(f(y)) + \langle \nabla g(z), f(y + tu) - f(y) \rangle.$$

This gives us, for arbitrary  $u \in \mathbb{R}^p$ , that

$$\begin{aligned} h^\circ(x, u) &= \limsup_{y \rightarrow x, t \downarrow 0} \frac{g(f(y + tu)) - g(f(y))}{t} \\ &= \limsup_{y \rightarrow x, t \downarrow 0} \left\langle \nabla g(z), \frac{f(y + tu) - f(y)}{t} \right\rangle. \end{aligned}$$

Observe that the Lipschitz property of  $f$  implies that  $(f(y + tu) - f(y)) / t$  is bounded. Moreover,  $y \rightarrow x, t \downarrow 0$ , combined with the same property of  $f$ , gives us  $z \rightarrow f(x)$ . Using also the  $C^1$  property of  $g$ , one gets from above that

$$\begin{aligned} h^\circ(x, u) &= \limsup_{y \rightarrow x, t \downarrow 0} \left\langle \nabla g(f(x)), \frac{f(y + tu) - f(y)}{t} \right\rangle \\ &= \limsup_{y \rightarrow x, t \downarrow 0} \frac{\langle \nabla g(f(x)), f(y + tu) \rangle - \langle \nabla g(f(x)), f(y) \rangle}{t} \\ &= \langle \nabla g(f(x)), f \rangle^\circ(x, u). \end{aligned}$$

The relation (5.1.19) follows now from Proposition 5.1.2 (iii).  $\square$

**Remark 5.1.20.** *The inclusion “ $\subset$ ” from (5.1.19) immediately follows from Theorem 5.1.18.*

Recall that for a differentiable function, its differential  $\nabla f(x)$  can be identified with the Jacobian matrix.

**Theorem 5.1.21** (chain rule 3). *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}^k$  be continuously differentiable near  $x$  and  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  be locally Lipschitz around  $f(x)$ . Then the function  $h := g \circ f : \mathbb{R}^p \rightarrow \mathbb{R}$  is locally Lipschitz around  $x$ , and*

$$\partial_C h(x) \subset \langle \partial_C g(f(x)), \nabla f(x) \rangle := \{ \langle \xi, \nabla f(x) \rangle \mid \xi \in \partial_C g(f(x)) \}. \tag{5.1.20}$$



Moreover, if  $\nabla f(x) : \mathbb{R}^p \rightarrow \mathbb{R}^k$  is surjective, then equality holds.

*Proof* The proof of (5.1.20) follows taking into account in Theorem 5.1.18 that for any  $y \in \partial_C g(f(x))$ ,

$$\partial_C \langle y, f \rangle (x) = \langle y, \nabla f(x) \rangle,$$

hence, since  $\partial_C g(f(x))$  is compact and convex and  $\nabla f(x)$  is linear and bounded,

$$\text{cl conv} \{ \partial_C \langle y, f \rangle (x) \mid y \in \partial_C g(f(x)) \} = \langle \partial_C g(f(x)), \nabla f(x) \rangle.$$

Suppose now that  $\nabla f(x)$  is surjective. By the Lyusternik-Graves Theorem (i.e., Theorem 2.4.4), it follows that  $f$  is open, i.e., for any neighborhood  $U$  of  $x$ ,  $f(U)$  is a neighborhood of  $f(x)$ . This implies

$$\limsup_{y \rightarrow f(x), t \downarrow 0} \frac{g(y + t\nabla f(x)(u)) - g(y)}{t} = \limsup_{z \rightarrow x, t \downarrow 0} \frac{g(f(z) + t\nabla f(x)(u)) - g(f(z))}{t}.$$

Also, since  $f$  is  $C^1$  near  $x$ ,

$$\begin{aligned} \lim_{z \rightarrow x, t \downarrow 0} \frac{f(z + tu) - f(z) - t\nabla f(x)(u)}{t} &= \lim_{z \rightarrow x, t \downarrow 0} \frac{f(z + tu) - f(z) - t\nabla f(z)(u)}{t} \\ &\quad + \lim_{z \rightarrow x} (\nabla f(z) - \nabla f(x))(u) = 0. \end{aligned}$$

Therefore, it follows that

$$\begin{aligned} g^\circ(f(x), \nabla f(x)(u)) &= \limsup_{y \rightarrow f(x), t \downarrow 0} \frac{g(y + t\nabla f(x)(u)) - g(y)}{t} \\ &= \limsup_{z \rightarrow x, t \downarrow 0} \frac{g(f(z) + t\nabla f(x)(u)) - g(f(z))}{t} \\ &= \limsup_{z \rightarrow x, t \downarrow 0} \frac{g(f(z + tu)) - g(f(z))}{t} = h^\circ(x, u), \quad \forall u \in \mathbb{R}^p, \end{aligned}$$

where for the third equality the local Lipschitz of  $g$  around  $f(x)$  is used. Take now arbitrary  $\chi \in \partial_C g(f(x))$  and  $u \in \mathbb{R}^p$ . Then

$$\langle \chi, \nabla f(x) \rangle (u) = \langle \chi, \nabla f(x)(u) \rangle \leq g^\circ(f(x), \nabla f(x)(u)) = h^\circ(x, u).$$

It means that  $\langle \chi, \nabla f(x) \rangle \in \partial_C h(x)$ . Therefore, the conclusion follows. □

Finally, we present a result which will be useful in deriving optimality conditions.

**Proposition 5.1.22.** Consider  $\varphi : \mathbb{R}^p \rightarrow \mathbb{R}^k$  locally Lipschitz around  $x$ ,  $K \subset \mathbb{R}^k$  a compact set, and define

$$F(x) := \max_{y \in K} \langle y, \varphi(x) \rangle.$$

Then  $F$  is locally Lipschitz around  $x$ . For any  $u$  around  $x$ , denote by

$$K(u) := \{ y \in K \mid F(u) = \langle y, \varphi(u) \rangle \}$$

the set of active indices at  $u$ . If  $K(x)$  is a singleton set (i.e.,  $K(x) = \{y_0\}$ ), then

$$\partial_C F(x) \subset \partial_C \langle y_0, \varphi \rangle (x).$$

*Proof* Denote by  $L$  the Lipschitz constant of  $\varphi$ . The fact that  $F$  is locally Lipschitz around  $x$  is straightforward.

Pick  $\xi \in \partial_C F(x)$  and take arbitrary  $u \in \mathbb{R}^p$ . Take some realizing sequences  $y_n \rightarrow x$ ,  $t_n \downarrow 0$  for  $F^\circ(x, u)$ . Moreover, take  $y_n \in K(y_n + t_n u)$ . Since  $K$  is compact, we may suppose, without loss of generality, that  $y_n$  converges to some  $\bar{y}$ . As  $F(y_n + t_n u) = \langle y_n, \varphi(y_n + t_n u) \rangle$  for every  $n$  and  $y_n + t_n u \rightarrow x$ , we obtain by passing to the limit that  $F(x) = \langle \bar{y}, \varphi(x) \rangle$ , i.e.,  $\bar{y} \in K(x)$ , which means that  $\bar{y} = y_0$ . It follows that

$$\begin{aligned} F^\circ(x, u) &= \lim_{n \rightarrow \infty} \frac{F(y_n + t_n u) - F(y_n)}{t_n} \leq \limsup_{n \rightarrow \infty} \frac{\langle y_n, \varphi(y_n + t_n u) \rangle - \langle y_n, \varphi(y_n) \rangle}{t_n} \\ &= \limsup_{n \rightarrow \infty} \frac{\langle y_0, \varphi \rangle (y_n + t_n u) - \langle y_0, \varphi \rangle (y_n)}{t_n} \\ &\quad + \limsup_{n \rightarrow \infty} \frac{\langle y_n - y_0, \varphi(y_n + t_n u) - \varphi(y_n) \rangle}{t_n} \\ &\leq \langle y_0, \varphi \rangle^\circ(x, u) + \limsup_{n \rightarrow \infty} \|y_n - y_0\| \frac{\|\varphi(y_n + t_n u) - \varphi(y_n)\|}{t_n} \\ &\leq \langle y_0, \varphi \rangle^\circ(x, u) + \limsup_{n \rightarrow \infty} \|y_n - y_0\| L \|u\| = \langle y_0, \varphi \rangle^\circ(x, u). \end{aligned}$$

Hence, we obtained  $F^\circ(x, u) \leq \langle y_0, \varphi \rangle^\circ(x, u)$ , which shows the inclusion  $\partial_C F(x) \subset \partial_C \langle y_0, \varphi \rangle (x)$  by Proposition 5.1.2.  $\square$

### 5.1.2 Clarke Tangent and Normal Cones

Throughout this section, we will discuss the notions of tangent and normal cones, as defined by Clarke. In the sequel,  $A$  denotes a nonempty closed subset of  $\mathbb{R}^p$ . Recall that the distance function associated to the set  $A$  is given by

$$d_A(x) = \inf_{y \in A} \|y - x\|,$$

and has the important property (in our context) that is 1–Lipschitz. The announced tangency notion will be constructed through the generalized gradient of this function, and the normal cone will be defined by polarity. One of the main reasons for which these notions were introduced by Clarke in the early 1980’s is given by the next penalization principle. Notice that a local version of the first item of the next result was proved and used in the previous chapter (see Theorem 4.3.2).

**Theorem 5.1.23** (penalization principle). *Let  $U \subset \mathbb{R}^p$  be an open set which contains  $A$ , and  $f : U \rightarrow \mathbb{R}$  be a Lipschitz function with modulus  $L > 0$ . Consider the minimization*

problem

$$(P) \min f(x), \quad \text{subject to } x \in A.$$

(i) If  $\bar{x} \in A$  is a global solution of (P), then for every  $K \geq L$ , the function  $x \mapsto f(x) + Kd_A(x)$  attains an (unconstrained) minimum on  $U$  at  $\bar{x}$ .

(ii) Conversely, suppose that, for some  $K > L$ , the function  $x \mapsto f(x) + Kd_A(x)$  attains its minimum on  $U$  at  $\bar{x}$ . Then  $\bar{x} \in A$  and solves (P).

*Proof* Suppose first that  $\bar{x} \in A$  is a solution of (P). Take  $x \in U$  and  $\varepsilon > 0$  arbitrarily. Pick  $y_\varepsilon \in A$  such that  $\|y_\varepsilon - x\| \leq d_A(x) + \varepsilon$ . Using the Lipschitz property of  $f$ , one gets

$$f(\bar{x}) \leq f(y_\varepsilon) \leq f(x) + L \|y_\varepsilon - x\| \leq f(x) + Ld_A(x) + L\varepsilon.$$

Letting  $\varepsilon$  to tend at 0, one gets the conclusion of the first part for  $K = L$ . If  $K > L$ , the conclusion is now obvious.

For the converse, suppose by contradiction that  $\bar{x} \notin A$ , hence  $x \in U \setminus A$ . Since  $A$  is closed, it follows that  $d_A(\bar{x}) > 0$ . Then one can find  $y \in A$  such that

$$\|y - \bar{x}\| < d_A(\bar{x}) + \left(\frac{K}{L} - 1\right) d_A(\bar{x}) = \frac{K}{L} d_A(\bar{x}).$$

Using the minimality assumption and the Lipschitz property of  $f$ , one gets

$$f(\bar{x}) + Kd_A(\bar{x}) \leq f(y) \leq f(\bar{x}) + L \|y - \bar{x}\| < f(\bar{x}) + Kd_A(\bar{x}),$$

a contradiction which shows that  $\bar{x} \in A$ . Then for every  $x \in A$ ,

$$f(\bar{x}) = f(\bar{x}) + Kd_A(\bar{x}) \leq f(x) + Kd_A(x) = f(x),$$

hence  $\bar{x}$  solves (P). □

In view of the generalized Fermat rule (Theorem 5.1.16) and of sum rule (5.1.14), solutions of (P) satisfy

$$0 \in \partial_C(f + Kd_A)(\bar{x}) \subset \partial_C f(\bar{x}) + K\partial_C d_A(\bar{x}). \tag{5.1.21}$$

It is interesting to study the geometric interpretation of the term  $\partial_C d_A(x)$ .

**Definition 5.1.24.** Let  $\bar{x} \in A$ . The Clarke tangent and normal cones to  $A$  at  $\bar{x}$  are respectively the sets

$$T_C(A, \bar{x}) := \{u \in \mathbb{R}^p \mid d_A^\circ(\bar{x}, u) = 0\} = \{u \in \mathbb{R}^p \mid d_A^\circ(\bar{x}, u) \leq 0\} \tag{5.1.22}$$

$$N_C(A, \bar{x}) := T_C(A, \bar{x})^\circ = \{\xi \in \mathbb{R}^p \mid \langle \xi, x \rangle \leq 0, \forall x \in T_C(A, \bar{x})\}. \tag{5.1.23}$$

Remark that the second equality from (5.1.22) follows from the fact that one always has  $d_A^\circ(\bar{x}, u) \geq 0$ , since  $d_A$  attains a minimum at  $\bar{x}$ .

One may easily observe that  $T_C(A, \bar{x}) = \mathbb{R}^p$  and  $N_C(A, \bar{x}) = \{0\}$  if  $\bar{x} \in \text{int } A$ . The following theorem collects some other important properties of the Clarke tangent and normal cones.

**Theorem 5.1.25.** *Let  $\bar{x} \in A$ . Then:*

- (i)  $T_C(A, \bar{x})$  is a closed convex cone;
- (ii)  $u \in T_C(A, \bar{x})$  if and only if, for every  $(x_n) \subset A$ ,  $x_n \rightarrow \bar{x}$  and every  $(t_n) \subset (0, \infty)$ ,  $(t_n) \rightarrow 0$ , there exists  $(u_n) \rightarrow u$  such that  $x_n + t_n u_n \in A$  for any  $n$ ; as consequence,  $T_C(A, \bar{x}) \subset T_B(A, \bar{x})$ ;
- (iii)  $N_C(A, \bar{x}) = \text{cl cone } \partial_C d_A(\bar{x})$ ;
- (iv)  $T_C(A, \bar{x}) = N_C(A, \bar{x})^\circ$ ;
- (v)  $N_C(A, \bar{x})$  is a closed convex cone containing  $N_B(A, \bar{x})$ .

*Proof* (i) We know by Proposition 5.1.1 that the function  $u \mapsto d_A^\circ(\bar{x}, u)$  is a positively homogeneous convex function, which shows, in light of (5.1.22), that  $T_C(A, \bar{x})$  is a convex cone. The continuity of the same function proves the closedness of  $T_C(A, \bar{x})$ .

(ii) Take  $u \in T_C(A, \bar{x})$  and take arbitrary sequences  $(x_n) \subset A$ ,  $(x_n) \rightarrow \bar{x}$  and  $(t_n) \subset (0, \infty)$ ,  $(t_n) \rightarrow 0$ . Since  $d_A^\circ(\bar{x}, u) = 0$ , one gets that

$$\inf_{r>0} \sup_{x \in B(\bar{x}, r), t \in (0, r)} \frac{d_A(y + tu) - d_A(y)}{t} = 0.$$

Take  $\varepsilon > 0$ . One can find  $r > 0$  such that  $\sup_{x \in B(\bar{x}, r), t \in (0, r)} \frac{d_A(y + tu) - d_A(y)}{t} < \varepsilon$ . For this  $r$ , there is  $n_0 \in \mathbb{N}$  such that, for any  $n \geq n_0$ ,  $x_n \in B(\bar{x}, r)$  and  $t_n \in (0, r)$ . This means that

$$\begin{aligned} \varepsilon &> \sup_{x \in B(\bar{x}, r), t \in (0, r)} \frac{d_A(y + tu) - d_A(y)}{t} \geq \frac{d_A(x_n + t_n u) - d_A(x_n)}{t_n} \\ &= \frac{d_A(x_n + t_n u)}{t_n} \geq 0, \end{aligned}$$

for any  $n \geq n_0$ , which shows that

$$\lim_{n \rightarrow \infty} \frac{d_A(x_n + t_n u)}{t_n} = 0.$$

Take  $y_n \in A$  such that  $\|x_n + t_n u - y_n\| \leq d_A(x_n + t_n u) + \frac{t_n}{n}$ , and define  $u_n := \frac{y_n - x_n}{t_n}$ . Using the relation above, we obtain that  $u_n \rightarrow u$  and  $x_n + t_n u_n = y_n \in A$ , as claimed.

For the converse inclusion, take  $u \in \mathbb{R}^p$  such that for every  $(x_n) \subset A$ ,  $(x_n) \rightarrow \bar{x}$  and every  $(t_n) \subset (0, \infty)$ ,  $(t_n) \rightarrow 0$ , there exists  $(u_n) \rightarrow u$  such that  $x_n + t_n u_n \in A$  for any  $n$ . Fix some sequences  $(y_n) \rightarrow \bar{x}$  and  $(t_n) \downarrow 0$  such that

$$\lim_{n \rightarrow \infty} \frac{d_A(y_n + t_n u) - d_A(y_n)}{t_n} = d_A^\circ(\bar{x}, u).$$

Let  $x_n \in A$  such that  $\|x_n - y_n\| \leq d_A(y_n) + \frac{t_n}{n}$ . It follows that  $(x_n) \rightarrow \bar{x}$ , hence from the property of  $u$  there is  $(u_n) \rightarrow u$  such that  $x_n + t_n u_n \in A$  for any  $n$ . But then, since  $d_A$  is 1-Lipschitz, one gets

$$d_A(y_n + t_n u) \leq d_A(x_n + t_n u_n) + \|x_n - y_n\| + t_n \|u_n - u\| \leq d_A(y_n) + t_n \left( \frac{1}{n} + \|u_n - u\| \right),$$

which implies that  $d_A^\circ(\bar{x}, u) \leq 0$ , i.e., the conclusion.

(iii) Take nonzero  $\xi_1, \xi_2 \in \partial_C d_A(\bar{x})$ ,  $\lambda_1, \lambda_2 > 0$ , and choose  $t \in [0, 1]$ . Define

$$\xi := \frac{(1-t)\lambda_1}{(1-t)\lambda_1 + t\lambda_2} \xi_1 + \frac{t\lambda_2}{(1-t)\lambda_1 + t\lambda_2} \xi_2 \in \partial_C d_A(\bar{x}),$$

and observe that  $(1-t)\lambda_1 \xi_1 + t\lambda_2 \xi_2 = [(1-t)\lambda_1 + t\lambda_2] \xi \in \text{cone } \partial_C d_A(\bar{x})$ , which means that cone  $\partial_C d_A(\bar{x})$  is a convex cone. It follows that  $\text{cl cone } \partial_C d_A(\bar{x})$  is a closed convex cone.

Take  $u \in \partial_C d_A(\bar{x})^-$ . If one would have  $d_A^\circ(\bar{x}, u) = \sup \{ \langle \xi, u \rangle \mid \xi \in \partial_C d_A(\bar{x}) \} > 0$ , then there exists  $\xi \in \partial_C d_A(\bar{x})$  such that  $\langle \xi, u \rangle > 0$ , which is a contradiction that shows that  $d_A^\circ(\bar{x}, u) \leq 0$ , i.e.,  $u \in T_C(A, \bar{x})$ . The reverse inclusion is easy to prove in the same manner, hence

$$\partial_C d_A(\bar{x})^- = [\text{cl cone } \partial_C d_A(\bar{x})]^- = T_C(A, \bar{x}). \tag{5.1.24}$$

By applying the bipolar theorem (Theorem 2.1.6) in the last equality, one gets that  $N_C(A, \bar{x}) = \text{cl cone } \partial_C d_A(\bar{x})$ .

(iv) Follows from (5.1.24) and (iii).

(v) This is straightforward, taking into account that if  $A \subset B$ , then  $A^- \supset B^-$ .  $\square$

The following proposition continues the idea from (5.1.21).

**Proposition 5.1.26.** *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be Lipschitz with modulus  $L$  around  $x$  and suppose  $x$  is a solution of (P). Then*

$$0 \in \partial_C(f + Ld_A)(x) \subset \partial_C f(x) + L\partial_C d_A(x) \subset \partial_C f(x) + N_C(A, x). \tag{5.1.25}$$

*Proof* It follows from (5.1.21) and (iv) from the previous theorem.  $\square$

It is interesting to further study the structure of  $N_C(A, x)$  in some special situations. To this end, a useful notion is introduced next.

**Definition 5.1.27.** *A set  $A$  is said to be regular at  $x \in A$  if  $T_C(A, x) = T_B(A, x)$ .*

It admits the following characterization.

**Proposition 5.1.28.**  *$A$  is regular if and only if  $N_C(A, x) = N_B(A, \bar{x})$ .*

*Proof* For the necessity, observe that if  $A$  is regular, then

$$N_C(A, x) = T_C(A, x)^- = T_B(A, x)^- = N_B(A, x).$$

For the sufficiency, suppose that  $N_C(A, x) = N_B(A, \bar{x})$ . Then

$$T_C(A, x) = N_C(A, x)^- = N_B(A, x)^- = (T_B(A, x)^-)^- \supset T_B(A, x) \supset T_C(A, x),$$

which shows the desired conclusion.  $\square$

In the case of closed convex sets, we have the following result.

**Theorem 5.1.29.** *Suppose  $A$  is a closed convex set, and  $x \in A$ . Then  $A$  is regular and*

$$\begin{aligned} T_B(A, x) &= T_C(A, x) = \text{cl cone}(A - x), \\ N_B(A, x) &= N_C(A, x) = \{u \in \mathbb{R}^p \mid \langle u, y - x \rangle \leq 0, \forall y \in A\}. \end{aligned}$$

*Proof* In view of the Propositions 2.1.15 and 5.1.28, we only have to prove that  $A$  is regular.

Take  $u \in A - x$ , and choose arbitrary  $(x_n) \subset A$ ,  $(x_n) \rightarrow x$  and  $(t_n) \downarrow 0$ . Define  $u_n := (x - x_n) + u$  for every  $n$  and remark that  $(u_n) \rightarrow u$ . Moreover, observe that

$$x_n + t_n u_n = (1 - t_n)x_n + t_n(u + x) \in A,$$

since  $A$  is convex. It follows from Theorem 5.1.25 that  $u \in T_C(A, x)$ . As  $T_C(A, x)$  is a convex cone, we have  $\text{cl cone}(A - x) \subset T_C(A, x)$ . But we already know that

$$T_C(A, x) \subset T_B(A, x) = \text{cl cone}(A - x),$$

which ends the proof.  $\square$

Recall that if  $A$  is a nonempty closed set in  $\mathbb{R}^p$ , we denote by  $\text{pr}_A x$  the nonempty projection set  $\{y \in A \mid d_A(x) = \|x - y\|\}$ .

**Theorem 5.1.30.** *Let  $A$  be a nonempty closed subset of  $\mathbb{R}^p$ , and  $\bar{x} \in \text{bd } A$ . Then*

$$N_C(A, \bar{x}) = \text{cl conv} \left\{ \lambda \lim_{n \rightarrow \infty} \frac{x_n - y_n}{\|x_n - y_n\|} \mid \lambda \geq 0, x_n \notin A, x_n \rightarrow \bar{x}, y_n \in \text{pr}_A x_n \right\}. \quad (5.1.26)$$

*Proof* Observe first that if  $x \in A$  is such that  $\nabla d_A(x)$  exists, it follows that

$$\nabla d_A(x)(u) = \lim_{t \rightarrow 0} \frac{d_A(x + tu) - d_A(x)}{t} = \lim_{t \rightarrow 0} \frac{d_A(x + tu)}{t} \geq 0, \quad \forall u \in \mathbb{R}^p,$$

hence  $\nabla d_A(x) = 0$ .

Suppose  $x \notin A$  and  $\nabla d_A(x)$  exists. Take  $y \in \text{pr}_A x$  and  $u \in \mathbb{R}^p$ , and denote  $g(x) := \|x - y\|$ . Then

$$\begin{aligned} g^\circ(x, u) &\geq \limsup_{t \downarrow 0} \frac{g(x + tu) - g(x)}{t} = \limsup_{t \downarrow 0} \frac{\|x + tu - y\| - \|x - y\|}{t} \\ &\geq \limsup_{t \downarrow 0} \frac{d_A(x + tu) - d_A(x)}{t} = \nabla d_A(x)(u), \end{aligned}$$

since  $\|x - y\| = d_A(x)$  and  $\nabla d_A(x)$  exists. But this means that  $\nabla d_A(x) \in \partial_C g(x) = \left\{ \frac{x - y}{\|x - y\|} \right\}$ , i.e.,  $\nabla d_A(x) = \frac{x - y}{\|x - y\|}$ . Suppose now there are two points  $y_1, y_2 \in \text{pr}_A x$ .

It means that  $\nabla d_A(x) = \frac{x - y_1}{d_A(x)} = \frac{x - y_2}{d_A(x)}$ , i.e.,  $\text{pr}_A x$  consists in only one point.

Take now  $\bar{x} \in \text{bd } A$ . Using the gradient formula (5.1.8) for  $d_A$ , we have that

$$\partial_C d_A(\bar{x}) = \text{conv} \left\{ \lim_{n \rightarrow \infty} \nabla d_A(x_n) \mid x_n \rightarrow \bar{x}, x_n \notin \Omega_{d_A} \right\},$$

where by  $\Omega_{d_A}$  one denotes the set of points at which  $d_A$  fails to be differentiable. Taking into account the discussion above, one gets

$$\partial_C d_A(\bar{x}) \subset \text{conv} \left\{ 0, \lim_{n \rightarrow \infty} \frac{x_n - y_n}{\|x_n - y_n\|} \mid x_n \notin A, x_n \rightarrow \bar{x}, y_n \in \text{pr}_A x_n \right\}. \quad (5.1.27)$$

Now, observe that for any  $x \in A$ , the function  $d_A$  attains a minimum at  $x$ , hence by applying the generalized Fermat rule (Theorem 5.1.16), one gets that  $0 \in \partial_C d_A(x)$ .

Take now  $x \notin A$  and  $y \in \text{pr}_A x$ . Consider the function  $f(z) := \|x - z\|$ . This function, which is 1-Lipschitz, attains its minimum on  $A$  at  $y$ , hence by applying Proposition 5.1.26 it follows that

$$0 \in \partial_C f(y) + \partial_C d_A(y).$$

As  $x \neq y$ ,  $f$  is  $C^1$  around  $y$ , hence  $\partial_C f(y) = \{\nabla f(y)\} = \left\{ -\frac{x - y}{\|x - y\|} \right\}$ . It follows that  $\frac{x - y}{\|x - y\|} \in \partial_C d_A(y)$ . Hence,

$$\left\{ 0, \lim_{n \rightarrow \infty} \frac{x_n - y_n}{\|x_n - y_n\|} \mid x_n \notin A, x_n \rightarrow \bar{x}, y_n \in \text{pr}_A x_n \right\} \subset \partial_C d_A(\bar{x})$$

and, as  $\partial_C d_A(\bar{x})$  is a convex set, one has equality in (5.1.27). The result now follows from the fact that  $N_C(A, \bar{x}) = \text{cl cone } \partial_C d_A(\bar{x})$ . □

The next example emphasizes the utility of the formula (5.1.26), and at the same time exhibits a case where the Bouligand and the Clarke tangent cones are different.

**Example 5.1.31.** Consider the set  $A := \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 = |x_1|\}$ . Then it is easy to see, using Definition 2.1.9, that  $T_B(A, (0, 0)) = A$ . Moreover, using the polarity, one gets

$$N_B(A, (0, 0)) = A^- = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 \leq -|x_1|\}.$$

Now, for the Clarke normal cone, one gets, by using formula (5.1.26), that

$$N_C(A, (0, 0)) = \text{cl conv} \left\{ (x_1, x_2) \in \mathbb{R}^2 \mid x_2 = |x_1| \text{ or } x_2 \leq -|x_1| \right\} = \mathbb{R}^2,$$

hence  $T_C(A, (0, 0)) = \{(0, 0)\}$ .

The next theorem shows that the Clarke tangent and normal cones to the graphs of Lipschitz mappings are always linear subspaces. Among other things, this result emphasizes the importance of studying also other types of normal cones, such as the ones in the next chapter. In what follows,  $\text{gr } f$  denotes the graph of a function  $f$ .

**Theorem 5.1.32** (Rockafellar). *Let  $U \subset \mathbb{R}^p$  be an open set, and  $f : U \rightarrow \mathbb{R}^k$  be a Lipschitz function. Consider  $\bar{x} \in U$  and  $\bar{y} := f(\bar{x})$ . Then  $T_C(\text{gr } f, (\bar{x}, \bar{y}))$  and  $N_C(\text{gr } f, (\bar{x}, \bar{y}))$  are linear subspaces of  $\mathbb{R}^p \times \mathbb{R}^k$ .*

*Proof* According to Theorem 5.1.25 (ii),  $(\xi, \eta) \in T_C(\text{gr } f, (\bar{x}, \bar{y}))$  if and only if for every  $(x_n, y_n) \in \text{gr } f$ ,  $(x_n, y_n) \rightarrow (\bar{x}, \bar{y})$  and every  $(t_n) \downarrow 0$ , there exists  $(\xi_n, \eta_n) \rightarrow (\xi, \eta)$  such that  $(x_n, y_n) + t_n(\xi_n, \eta_n) \in \text{gr } f$  for any  $n$ . Since  $f$  is continuous, this condition reduces to the following: for every  $(x_n) \rightarrow \bar{x}$  and every  $(t_n) \downarrow 0$ , there is  $\xi_n \rightarrow \xi$  such that

$$\frac{f(x_n + t_n \xi_n) - f(x_n)}{t_n} \rightarrow \eta. \tag{5.1.28}$$

Since  $f$  is Lipschitz, it follows that

$$\|f(x_n + t_n \xi_n) - f(x_n + t_n \xi)\| \leq L t_n \|\xi_n - \xi\|,$$

where by  $L$  we have denoted the Lipschitz modulus of  $f$ . The limit from (5.1.28) is the same when one replaces  $\xi_n$  by  $\xi$ . It follows that  $(\xi, \eta) \in T_C(\text{gr } f, (\bar{x}, \bar{y}))$  if and only if for every  $(x_n) \rightarrow \bar{x}$  and every  $(t_n) \downarrow 0$ , there is  $\xi_n \rightarrow \xi$  such that

$$\frac{f(x_n + t_n \xi) - f(x_n)}{t_n} \rightarrow \eta,$$

that is,

$$\lim_{\substack{t \downarrow 0 \\ x \rightarrow \bar{x}}} \frac{f(x + t\xi) - f(x)}{t} = \eta,$$

i.e., the generalized directional derivative of  $f$  at  $\bar{x}$  in the direction  $\xi$  is in fact a usual lim and equals  $\eta$ . This is also equivalent to

$$f^\circ(\bar{x}, \xi) = \eta = \lim_{\substack{t \downarrow 0 \\ x' \rightarrow \bar{x}}} \frac{f(x') - f(x' - t\xi)}{t},$$

where  $x' := x + t\xi$ , hence if  $(\xi, \eta) \in T_C(\text{gr } f, (\bar{x}, \bar{y}))$ , one necessarily has also that  $(-\xi, -\eta) \in T_C(\text{gr } f, (\bar{x}, \bar{y}))$ . Since we also know from Theorem 5.1.25 (i) that  $T_C(\text{gr } f, (\bar{x}, \bar{y}))$  is a closed convex cone, it follows that  $T_C(\text{gr } f, (\bar{x}, \bar{y}))$  is actually a (closed) linear subspace of  $\mathbb{R}^p \times \mathbb{R}^k$ . Since the dual set of a linear subspace is also a linear subspace, it follows that  $N_C(\text{gr } f, (\bar{x}, \bar{y}))$  is also a (closed) linear subspace of  $\mathbb{R}^p \times \mathbb{R}^k$ .  $\square$

We have seen that the distance function serves as a link between the analytical theory of generalized gradients and the geometric constructions given by the tangent



and normal cones. The next result shows other interesting connections, this time for a general function; the final equivalence in the section about the Fréchet and Morukhovich normal cones arises from very definitions, proving that in fact all the normal constructions we use can be seen in a unifying way.

**Theorem 5.1.33.** *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be Lipschitz around  $\bar{x}$ . Then:*

- (i)  $T_C(\text{epi } f, (\bar{x}, f(\bar{x}))) = \text{epi } f^\circ(\bar{x}, \cdot)$ ;
- (ii)  $\xi \in \partial_C f(\bar{x}) \Leftrightarrow (\xi, -1) \in N_C(\text{epi } f, (\bar{x}, f(\bar{x})))$ .

*Proof* (i) Take  $(u, r) \in T_C(\text{epi } f, (\bar{x}, f(\bar{x})))$ . Moreover, take some realizing sequences for  $f^\circ(\bar{x}, u)$  : we have  $(x_n) \rightarrow \bar{x}$  and  $(t_n) \downarrow 0$  such that

$$\lim_{n \rightarrow \infty} \frac{f(x_n + t_n u) - f(x_n)}{t_n} = f^\circ(\bar{x}, u).$$

Since  $(x_n, f(x_n))$  belongs to  $\text{epi } f$  and converges to  $(\bar{x}, f(\bar{x}))$ , we have, due to Theorem 5.1.25 (ii), that there is  $(u_n, r_n) \rightarrow (u, r)$  such that  $(x_n, f(x_n)) + t_n(u_n, r_n) \in \text{epi } f$  for every  $n$ , i.e.,

$$f(x_n + t_n u_n) \leq f(x_n) + t_n r_n,$$

and using the Lipschitz property of  $f$  (with constant  $\ell$ ), it follows that

$$\frac{f(x_n + t_n u) - f(x_n)}{t_n} \leq r_n + \ell \|u_n - u\|.$$

Passing to the limit for  $n \rightarrow \infty$ , we get  $f^\circ(\bar{x}, u) \leq r$ , i.e.,  $(u, r) \in \text{epi } f^\circ(\bar{x}, \cdot)$ .

For the converse inclusion, we prove that for every  $\delta \geq 0$ , we have  $(u, f^\circ(\bar{x}, u) + \delta) \in T_C(\text{epi } f, (\bar{x}, f(\bar{x})))$ . Take arbitrary  $(x_n, r_n) \in \text{epi } f$ ,  $(x_n, r_n) \rightarrow (\bar{x}, f(\bar{x}))$ , and  $(t_n) \downarrow 0$ . Define  $u_n := u$  for every  $n$ , and take

$$s_n := \max \left\{ f^\circ(\bar{x}, u) + \delta, \frac{f(x_n + t_n u) - f(x_n)}{t_n} \right\}.$$

Since  $\limsup_{n \rightarrow \infty} \frac{f(x_n + t_n u) - f(x_n)}{t_n} \leq f^\circ(\bar{x}, u)$ , it follows that  $s_n \rightarrow f^\circ(\bar{x}, u) + \delta$ , hence  $(u_n, s_n) \rightarrow (u, f^\circ(\bar{x}, u) + \delta)$ . In order to show by Theorem 5.1.25 (ii) the desired assertion, we must prove that  $(x_n, r_n) + t_n(u_n, s_n) \in \text{epi } f$  for every  $n$ , i.e.,

$$f(x_n + t_n u) \leq r_n + t_n s_n, \quad \forall n.$$

The definition of  $s_n$  and the fact that  $(x_n, r_n) \in \text{epi } f$  for every  $n$ , imply that

$$r_n + t_n s_n \geq r_n + f(x_n + t_n u) - f(x_n) \geq f(x_n + t_n u),$$

so the proof of (i) is now finished.

(ii) Remark that  $\xi \in \partial_C f(\bar{x})$  if and only if  $f^\circ(\bar{x}, u) \geq \langle \xi, u \rangle$  for every  $u$ , that is, for every  $u \in \mathbb{R}^p$  and every  $r \geq f^\circ(\bar{x}, u)$ , one has  $\langle (\xi, -1), (u, r) \rangle \leq 0$ . Using (i), the last assertion is equivalent to: for every  $(u, r) \in \text{epi } f^\circ(\bar{x}, \cdot) = T_C(\text{epi } f, (\bar{x}, f(\bar{x})))$ , one has  $\langle (\xi, -1), (u, r) \rangle \leq 0$ , i.e.,  $(\xi, -1) \in T_C(\text{epi } f, (\bar{x}, f(\bar{x})))^\circ = N_C(\text{epi } f, (\bar{x}, f(\bar{x})))$ .  $\square$

### 5.1.3 Optimality Conditions in Lipschitz Optimization

In this section, we will derive necessary optimality conditions in Fritz John form for the minimization problem

$$(MP) \quad \min f(x) \text{ subject to } g(x) \leq 0, \quad h(x) = 0, \quad x \in A,$$

where  $f, g = (g_1, \dots, g_n)$  and  $h = (h_1, \dots, h_m)$  are locally Lipschitz functions which map from  $\mathbb{R}^p$  into  $\mathbb{R}, \mathbb{R}^n$  and  $\mathbb{R}^m$ , respectively, and the set  $A$  is closed. Notice that the (MP) problem combines both functional and geometrical constraints, which were separately analyzed in the smooth case, since the presence of the set  $A$  along the smooth functions introduces a nonsmooth behaviour, which cannot be investigated with the tools of classical differentiation.

**Theorem 5.1.34** (Fritz John conditions for Clarke calculus). *Let  $\bar{x}$  be a solution of (MP), where  $f, g$  and  $h$  are Lipschitz around  $\bar{x}$ . Then there exist  $\lambda_0 \geq 0, \lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$  and  $\mu = (\mu_1, \dots, \mu_m) \in \mathbb{R}^m$ , with  $\lambda_0 + \|\lambda\| + \|\mu\| \neq 0$ , such that*

$$0 \in \lambda_0 \partial_C f(\bar{x}) + \sum_{i=1}^n \lambda_i \partial_C g_i(\bar{x}) + \sum_{j=1}^m \mu_j \partial_C h_j(\bar{x}) + N_C(A, \bar{x}) \tag{5.1.29}$$

and

$$\lambda_i \geq 0, \quad \lambda_i g_i(\bar{x}) = 0 \text{ for any } i \in \overline{1, n}. \tag{5.1.30}$$

*Proof* Suppose, without loss of generality, that the functions  $f, g$  and  $h$  are Lipschitz on a neighborhood of the set  $A$  (otherwise, replace  $A$  by  $A \cap B(\bar{x}, \delta)$  for some  $\delta > 0$ , and observe that neither the assumptions, nor the conclusion change).

Fix  $\varepsilon \in (0, 1)$ , denote the compact set  $S(0, 1) \cap [\mathbb{R}_+ \times \mathbb{R}_+^n \times \mathbb{R}^m]$  by  $S$ , and define the function

$$F_\varepsilon(x) := \max_{(\lambda_0, \lambda, \mu) \in S} \left\{ \lambda_0(f(x) - f(\bar{x}) + \varepsilon) + \sum_{i=1}^n \lambda_i g_i(x) + \sum_{j=1}^m \mu_j h_j(x) \right\}.$$

Suppose there is an  $x$  such that  $F_\varepsilon(x) \leq 0$ . It follows that

$$\lambda_0(f(x) - f(\bar{x}) + \varepsilon) + \sum_{i=1}^n \lambda_i g_i(\bar{x}) + \sum_{j=1}^m \mu_j h_j(\bar{x}) \leq 0,$$

for any  $(\lambda_0, \lambda, \mu) \in S$ . By taking successively one of the scalars  $\lambda_0, \lambda_1, \dots, \lambda_n, \mu_1, \dots, \mu_m$  equal to 1, and the rest equal to 0, one gets

$$f(x) \leq f(\bar{x}) - \varepsilon, \quad g(\bar{x}) = 0, \quad h(\bar{x}) = 0,$$

a contradiction which shows that the assumption made was false. Consequently,  $F_\varepsilon(x) > 0$  for any  $x \in A$ . Moreover, being the max of Lipschitz functions,  $F_\varepsilon$  is Lipschitz on a neighborhood of  $A$ , with a constant we denote by  $L$  which may be chosen

such that it does not depend on  $\varepsilon$ . Consider the closed set  $A$  and apply the Ekeland Variational Principle (see Remark 3.1.13) to get the existence of  $x_\varepsilon \in A$  such that

$$F_\varepsilon(x_\varepsilon) \leq F_\varepsilon(\bar{x}) - \sqrt{\varepsilon} \|x_\varepsilon - \bar{x}\|$$

and

$$F_\varepsilon(x_\varepsilon) \leq F_\varepsilon(x) + \sqrt{\varepsilon} \|x_\varepsilon - x\| \text{ for any } x \in A.$$

Observe that  $F_\varepsilon(\bar{x}) = \varepsilon$ . One gets from the first relation that  $\|x_\varepsilon - \bar{x}\| \leq \sqrt{\varepsilon}$ . On the other hand, from the second relation, one sees that the Lipschitz function  $x \mapsto F_\varepsilon(x) + \sqrt{\varepsilon} \|x_\varepsilon - x\|$  attains its minimum on  $A$  at  $x_\varepsilon$ , hence by applying Proposition 5.1.26, one gets the existence of  $L > 0$  such that

$$0 \in \partial_C F_\varepsilon(x_\varepsilon) + \sqrt{\varepsilon} D(0, 1) + L \partial_C d_A(x_\varepsilon). \tag{5.1.31}$$

Denote by

$$S(x_\varepsilon) := \{y \in S \mid F_\varepsilon(x_\varepsilon) = \langle y, \varphi(x_\varepsilon) \rangle\},$$

where  $y := (\lambda_0, \lambda, \mu)$  and  $\varphi(x) := (f(x) - f(\bar{x}) + \varepsilon, g(x), h(x))$ . In order to apply Proposition 5.1.22, we want to prove that  $S(x_\varepsilon)$  is a singleton. Suppose the contrary. Then there exist  $y^l := (\lambda_0^l, \lambda^l, \mu^l)$  ( $l = 1, 2$ ), two different points in  $S(x_\varepsilon)$ . Remember that  $S$  is a subset of the unit sphere in  $\mathbb{R}^{n+m+1}$ , and choose  $t > 1$  such that  $\bar{y} := \frac{t}{2}(y^1 + y^2) \in S$ . Then, since  $F_\varepsilon(x_\varepsilon) > 0$ ,

$$F_\varepsilon(x_\varepsilon) \geq \langle \bar{y}, \varphi(x_\varepsilon) \rangle = \frac{t}{2} \left( \langle y^1, \varphi(x_\varepsilon) \rangle + \langle y^2, \varphi(x_\varepsilon) \rangle \right) = t F_\varepsilon(x_\varepsilon) > F_\varepsilon(x_\varepsilon),$$

a contradiction. Hence,  $S(x_\varepsilon) = \{(\lambda_0^\varepsilon, \lambda^\varepsilon, \mu^\varepsilon)\}$ .

We can apply now Proposition 5.1.22, to get that

$$\partial_C F_\varepsilon(x_\varepsilon) \subset \partial_C \left( \lambda_0^\varepsilon f + \sum_{i=1}^n \lambda_i^\varepsilon g_i + \sum_{j=1}^m \mu_j^\varepsilon h_j \right) (x_\varepsilon).$$

By the use of the sum rule (Theorem 5.1.13), we get from (5.1.31) that

$$0 \in \lambda_0^\varepsilon \partial_C f(x_\varepsilon) + \sum_{i=1}^n \lambda_i^\varepsilon \partial_C g_i(x_\varepsilon) + \sum_{j=1}^m \mu_j^\varepsilon \partial_C h_j(x_\varepsilon) + \sqrt{\varepsilon} D(0, 1) + L \partial_C d_A(x_\varepsilon).$$

Set now  $\varepsilon$  from the above equal to  $\frac{1}{k}$  for any  $k \geq 2$ . One gets hence  $x_k$  such that  $\|x_k - \bar{x}\| \leq \frac{1}{\sqrt{k}}$ , and also  $(\lambda_0^k, \lambda^k, \mu^k) \in S$ , as well as  $\xi^k \in D\left(0, \frac{1}{\sqrt{k}}\right)$  such that

$$\xi^k \in \lambda_0^k \partial_C f(x_k) + \sum_{i=1}^n \lambda_i^k \partial_C g_i(x_k) + \sum_{j=1}^m \mu_j^k \partial_C h_j(x_k) + N_C(A, x_k).$$

Observe that we may suppose again, since  $S$  is compact, that  $(\lambda_0^k, \lambda^k, \mu^k)$  converges to some  $(\lambda_0, \lambda, \mu) \in S$ . As  $x_k \rightarrow \bar{x}$  and  $\xi^k \rightarrow 0$ , by passing to the limit in the relation above (see Proposition 5.1.6), one gets (5.1.29).

Suppose now that there is  $i_0 \in \overline{1, n}$  such that  $g_{i_0}(\bar{x}) < 0$ . If  $\lambda_{i_0} > 0$ , we will have

$$F(x_k) = \lambda_0^k \left( f(x_k) - f(\bar{x}) + \frac{1}{k} \right) + \sum_{i=1}^n \lambda_i^k g_i(x_k) + \sum_{j=1}^m \mu_j^k h_j(x_k) \rightarrow \sum_{i=1}^n \lambda_i g_i(\bar{x}) < 0,$$

but this contradicts the fact that  $F(x_k) > 0$  for every  $k$ . □

Notice again that specific constraint qualification conditions could be imposed in order to avoid the situation  $\lambda_0 = 0$ . We mention that, for this case of nonsmooth calculus, the qualification conditions are beyond the aim of this book, but the interested reader is invited to consult (Clarke, 1983).

**Example 5.1.35.** *Let us examine again the problem (5.1.1). We are now close to our initial aim, which was to provide optimality conditions and, furthermore, to find the minimum.*

*Observe first that, by Weierstrass Theorem, this problem admits at least one solution, since the function  $f$  is continuous, and the set of restrictions is compact. If  $(x_1, x_2)$  is a minimum point, we know by Theorem 5.1.34 that there exists  $(\lambda_0, \lambda_1) \neq (0, 0)$  such that*

$$\begin{aligned} (0, 0) &\in \lambda_0 \partial_C f(x_1, x_2) + \lambda_1 \partial_C g(x_1, x_2) \\ &= \lambda_0 \partial_C f(x_1, x_2) + \lambda_1 \nabla g(x_1, x_2) \\ &= \lambda_0 \partial_C f(x_1, x_2) + 2\lambda_1 \cdot (x_1, x_2), \end{aligned}$$

as  $g(x_1, x_2) = x_1^2 + x_2^2 - 1$  is continuously differentiable and  $\nabla g(x_1, x_2) = 2(x_1, x_2)$ .

Suppose  $\lambda_0 = 0$ . Then, again by Theorem 5.1.34,  $\lambda_1 > 0$  and  $g(x_1, x_2) = 0$ , i.e.,  $x_1^2 + x_2^2 = 1$ . On the other hand, one must have  $(x_1, x_2) = (0, 0)$ , which gives us a contradiction which means that we can take  $\lambda_0 = 1$ .

Suppose  $\lambda_1 = 0$ . Then  $(0, 0) \in \partial_C f(x_1, x_2)$ . We know from Example 5.1.9 that  $\partial_C f(0, 0) = \text{conv} \{ (2, 0), (0, 1), (1, 2) \}$ , hence  $(x_1, x_2)$  cannot be  $(0, 0)$ . Recall that the function  $f$  can equivalently be written as in (5.1.10). Suppose  $(x_1, x_2)$  lies in the interior of one of the sets  $S_1, S_2$  and  $S_3$ . But in this case  $\partial_C f(x_1, x_2) = \{ \nabla f(x_1, x_2) \}$ , and the gradient  $\nabla f(x_1, x_2)$  equals, respectively,  $(2, 0), (0, 1)$  and  $(1, 2)$ , so  $(0, 0) \in \partial_C f(x_1, x_2)$  cannot hold. Suppose now the nonzero vector  $(x_1, x_2)$  is on the boundary between two of the sets  $S_1, S_2$  and  $S_3$ . For instance, take  $(x_1, x_2) \in S_1 \cap S_2 \setminus \{(0, 0)\}$ . Then  $\partial_C f(x_1, x_2) = \text{conv} \{ (2, 0), (0, 1) \}$ , so the inclusion  $(0, 0) \in \partial_C f(x_1, x_2)$  cannot hold. The same happens in the other two similar cases. In conclusion, our assumption, that  $\lambda_1 = 0$ , must be false. Therefore, we can write:

$$(x_1, x_2) \in -\frac{1}{2\lambda_1} \partial_C f(x_1, x_2), \quad \lambda_1 > 0.$$

We know from Theorem 5.1.34, that  $x_1^2 + x_2^2 = 1$ . Suppose now  $(x_1, x_2)$  lies in the interior of one of the sets  $S_1, S_2$  and  $S_3$ . If  $(x_1, x_2) \in \text{int } S_1 =$

$\{(x_1, x_2) \mid 2x_1 < x_2 \text{ and } x_1 > 2x_2\}$ , one must have, since  $\partial_C f(x_1, x_2) = \{(2, 0)\}$ , that

$$(x_1, x_2) = \left(-\frac{1}{\lambda_1}, 0\right),$$

which is obviously a contradiction. If  $(x_1, x_2) \in \text{int } S_2 = \{(x_1, x_2) \mid 2x_1 > x_2 \text{ and } x_1 < -x_2\}$ , one would obtain  $(0, -\frac{1}{2\lambda_1}) \in \text{int } S_2$ , again a contradiction. Finally, if  $(x_1, x_2) \in \text{int } S_3 = \{(x_1, x_2) \mid x_1 < 2x_2 \text{ or } x_1 > -x_2\}$ , it should happen that  $-\frac{1}{2\lambda_1}(1, 2) \in \text{int } S_3$  for some positive  $\lambda_1$ , which is false. Only three points remain as candidates for the minimum points: the points from the unit sphere which lie on the boundary between two of the sets  $S_1, S_2$  and  $S_3$ . Suppose  $(x_1, x_2) \in S_2 \cap S_3 \setminus \{(0, 0)\}$ . Then  $(x_1, x_2)$  must be of the form  $(\varepsilon, -\varepsilon)$ , with  $\varepsilon > 0$ , and should be an element of the set  $-\frac{1}{2\lambda_1} \text{conv}\{(0, 1), (1, 2)\}$  for some positive  $\lambda_1$ , which is false. If  $(x_1, x_2) \in S_1 \cap S_3 \setminus \{(0, 0)\}$ , then  $(x_1, x_2) = (-2\varepsilon, -\varepsilon)$  for some  $\varepsilon > 0$ . It should satisfy  $(-2\varepsilon, -\varepsilon) \in -\frac{1}{2\lambda_1} \text{conv}\{(2, 0), (1, 2)\}$ , and taking into account that  $x_1^2 + x_2^2 = 1$ , one gets the point  $(-\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}})$  as a candidate for the minimum. Finally, if  $(x_1, x_2) \in S_1 \cap S_2 \setminus \{(0, 0)\}$ , then  $(x_1, x_2)$  must be of the form  $(-\varepsilon, -2\varepsilon)$ , for some  $\varepsilon > 0$ , and should be an element of  $-\frac{1}{2\lambda_1} \text{conv}\{(2, 0), (0, 1)\}$  for some positive  $\lambda_1$ , which set equals in fact  $-\mathbb{R}_+^2$ . The equality  $x_1^2 + x_2^2 = 1$  gives the point  $(-\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}})$  as another candidate for providing us the minimum. Observe, though, that for sufficiently small  $\delta$ ,  $(-\frac{1}{\sqrt{5}} + 3\delta, -\frac{2}{\sqrt{5}} - \delta) \in S_2$  and is feasible. Moreover, for  $\delta \rightarrow 0$ , it tends to  $(-\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}})$ , and  $f(-\frac{1}{\sqrt{5}} + 3\delta, -\frac{2}{\sqrt{5}} - \delta) = -\frac{2}{\sqrt{5}} - \delta < -\frac{2}{\sqrt{5}} = f(-\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}})$ , hence  $(-\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}})$  is not a local minimum. Analogously, one can prove that  $(-\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}})$  is not a local maximum, too. It means that the minimum point we are looking for is  $(-\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}})$ .

## 5.2 Mordukhovich Generalized Calculus

We saw in the previous section an approach which provided us with useful generalized differentiation tools, in order to solve some nonconvex and nonsmooth optimization problems. We saw that the Clarke tangent and normal cones, as well as the associated subdifferential, have the property that they are convex. In order to pass to a fully nonconvex case, we present in this section some other (even more general) differentiation objects, which have the advantage, besides allowing us to work in the fully nonconvex case, to exhibit robust (exact) calculus rules.

### 5.2.1 Fréchet and Mordukhovich Normal Cones

Let us begin our presentation by introducing two important constructions.

**Definition 5.2.1.** Let  $A$  be a nonempty subset of  $\mathbb{R}^p$ .

(i) Given  $x \in A$ , the Fréchet normal cone to  $A$  at  $x$  is the set

$$N_F(A, x) := \left\{ \xi \in \mathbb{R}^p \mid \limsup_{u \xrightarrow{A} x} \frac{\langle \xi, u - x \rangle}{\|u - x\|} \leq 0 \right\}, \quad (5.2.1)$$

where by  $u \xrightarrow{A} x$  we understand  $u \rightarrow x$  with  $u \in A$ . An element  $\xi \in N_F(A, x)$  is called Fréchet normal to  $A$  at  $x$ . If  $x \notin A$ , we define  $N_F(A, x) := \emptyset$ .

(ii) Let  $\bar{x} \in A$ . The Mordukhovich (or limiting, or basic) normal cone to  $A$  at  $\bar{x}$  is

$$N_M(A, \bar{x}) := \left\{ \xi \in \mathbb{R}^p \mid \exists x_n \xrightarrow{A} \bar{x}, \xi_n \rightarrow \xi, \xi_n \in N_F(A, x_n), \forall n \in \mathbb{N} \right\}. \quad (5.2.2)$$

If  $\bar{x} \notin A$ , we put  $N_M(A, \bar{x}) := \emptyset$ .

**Remark 5.2.2.** Observe that  $\xi \in N_F(A, \bar{x})$  if and only if for every  $\varepsilon > 0$ , there exists a neighborhood  $V$  of  $\bar{x}$  such that, for every  $x \in A \cap V$ , one has

$$\langle \xi, x - \bar{x} \rangle \leq \varepsilon \|x - \bar{x}\|. \quad (5.2.3)$$

In the next proposition we collect some basic properties of the Fréchet and Mordukhovich normal cones.

**Proposition 5.2.3.** Let  $A$  be a nonempty subset of  $\mathbb{R}^p$  and  $\bar{x} \in A$ . Then:

(i) One has

$$N_F(A, \bar{x}) = N_F(\text{cl } A, \bar{x}) \text{ and } N_M(A, \bar{x}) \subset N_M(\text{cl } A, \bar{x}). \quad (5.2.4)$$

(ii) If  $\bar{x} \in \text{int } A$ , one has  $N_F(A, \bar{x}) = N_M(A, \bar{x}) = \{0\}$ .

(iii)  $N_F(A, \bar{x})$  is a closed convex cone, and  $N_M(A, \bar{x})$  is a closed cone.

(iv) If  $A$  is a convex set, then

$$N_F(A, \bar{x}) = N_M(A, \bar{x}) = \left\{ \xi \in \mathbb{R}^p \mid \langle \xi, x - \bar{x} \rangle \leq 0 \text{ for every } x \in A \right\}, \quad (5.2.5)$$

i.e., it coincides with the normal cone  $N(A, \bar{x})$ .

(v) If  $B$  is a nonempty subset of  $\mathbb{R}^k$  and  $\bar{y} \in B$ , then

$$N_F(A \times B, (\bar{x}, \bar{y})) = N_F(A, \bar{x}) \times N_F(B, \bar{y}),$$

$$N_M(A \times B, (\bar{x}, \bar{y})) = N_M(A, \bar{x}) \times N_M(B, \bar{y}).$$

*Proof* (i) Observe first that for every two sets  $A \subset B$  with  $x \in X$ , one has  $N_F(B, x) \subset N_F(A, x)$ . This follows immediately from the fact that  $A \subset B$  implies

$\left[ u \xrightarrow{A} x \Rightarrow u \xrightarrow{B} x \right]$ , which gives us that  $\limsup_{u \xrightarrow{A} x} \frac{\langle \xi, u - x \rangle}{\|u - x\|} \leq \limsup_{u \xrightarrow{B} x} \frac{\langle \xi, u - x \rangle}{\|u - x\|}$ . Consequently, one has that  $N_F(\text{cl} A, \bar{x}) \subset N_F(A, \bar{x})$ .

For the reverse inclusion, take  $\xi \in N_F(A, \bar{x})$ . Then the equivalence from Remark 5.2.2 is valid, where the neighborhood  $V$  can be taken as being open, without loss of generality. Fix  $y \in \text{cl} A \cap V$ . If  $y \in A$ , then it satisfies (5.2.3). If  $y \in \text{cl} A \setminus A$ , then there exists  $(y_n) \subset A$ ,  $y_n \rightarrow y$ . Since  $V$  is open, one has that  $y_n \in A \cap V$  for every  $n$  sufficiently large. Then  $y_n$  satisfies (5.2.3), which implies when passing to the limit that  $\langle \xi, y - \bar{x} \rangle \leq \varepsilon \|y - \bar{x}\|$ . Consequently,  $\limsup_{y \xrightarrow{\text{cl} A} \bar{x}} \frac{\langle \xi, y - \bar{x} \rangle}{\|y - \bar{x}\|} \leq \varepsilon$  for every  $\varepsilon > 0$ , which shows that  $\xi \in N_F(\text{cl} A, \bar{x})$ .

For the second relation from (5.2.4), observe that

$$\left[ x_n \xrightarrow{A} \bar{x} \Rightarrow x_n \xrightarrow{\text{cl} A} \bar{x} \right] \text{ and } \left[ N_F(A, x_n) = N_F(\text{cl} A, x_n) \right]$$

give us the desired inclusion.

(ii) If  $\bar{x} \in \text{int} A$  and  $\xi \in N_F(A, \bar{x})$ , then for every  $\varepsilon > 0$ , there exists  $\delta > 0$  such that  $D(\bar{x}, \delta) \subset A$  and  $\langle \xi, x - \bar{x} \rangle < \varepsilon \|x - \bar{x}\|$  for every  $x \in D(\bar{x}, \delta)$ . It follows that  $\langle \xi, u \rangle < \varepsilon \|u\|$ , for every  $u \in D(0, 1)$ , hence  $\|\xi\| < \varepsilon$ . Since  $\varepsilon$  was arbitrarily chosen, we have that  $\xi = 0$ . The formula (5.2.2) and the fact that  $x_n \xrightarrow{A} \bar{x}$  imply that  $x_n \in \text{int} A$  for every  $n$  sufficiently large gives us that  $N_M(A, \bar{x}) = \{0\}$ .

(iii) The fact that  $N_F(A, \bar{x})$  and  $N_M(A, \bar{x})$  are cones easily follows from their definitions. Taking into account, for instance, the equivalence from Remark 5.2.2, one easily gets the convexity and closedness of  $N_F(A, \bar{x})$ .

Consider now  $(\xi_n) \subset N_M(A, \bar{x})$ ,  $\xi_n \rightarrow \xi$ . Then, for every  $\varepsilon > 0$ , there is  $n_0 \in \mathbb{N}$  such that  $\|\xi_n - \xi\| < 2^{-1}\varepsilon$  for every  $n \geq n_0$ . Fix  $n \geq n_0$ . Since  $\xi_n \in N_M(A, \bar{x})$ , there exist  $(x_k^n)_k \rightarrow \bar{x}$  and  $(\xi_k^n)_k \rightarrow \xi$  such that  $\xi_k^n \in N_F(A, x_k^n)$ , for every  $k$ . Consequently, one can find  $k_n^1, k_n^2 \geq n \geq n_0$  such that

$$\begin{aligned} \|x_{k_n^1}^n - \bar{x}\| &< \varepsilon, \forall k \geq k_n^1 \text{ and} \\ \|\xi_{k_n^2}^n - \xi_n\| &< 2^{-1}\varepsilon, \forall k \geq k_n^2. \end{aligned}$$

Define  $k_n := \max \{k_n^1, k_n^2\}$ ,  $y_n := x_{k_n}^n$  and  $\eta_n := \xi_{k_n}^n$ . One has

$$\begin{aligned} \|y_n - \bar{x}\| &< \varepsilon \text{ and} \\ \|\eta_n - \xi\| &\leq \|\eta_n - \xi_n\| + \|\xi_n - \xi\| < \varepsilon, \end{aligned}$$

hence, as  $n$  was taken arbitrary such that  $n \geq n_0$ , it follows that  $y_n \rightarrow \bar{x}$  and  $\eta_n \rightarrow \xi$ . Moreover, since  $\eta_n \in N_F(A, y_n)$  for every  $n \geq n_0$ , we obtain that  $\xi \in N_M(A, \bar{x})$ . Hence,  $N_M(A, \bar{x})$  is a closed set.

(iv) We prove first that

$$N_F(A, \bar{x}) = \left\{ \xi \in \mathbb{R}^p \mid \langle \xi, x - \bar{x} \rangle \leq 0 \text{ for every } x \in A \right\}. \tag{5.2.6}$$

The “ $\supset$ ” inclusion from (5.2.6) immediately follows for an arbitrary set  $A$ . Suppose now  $A$  is a convex set, and take  $\xi \in N_F(A, \bar{x})$  and  $x \in A$ . Then  $x_\lambda := (1 - \lambda)\bar{x} + \lambda x \in A$  for every  $\lambda \in [0, 1]$ , and  $x_\lambda \rightarrow \bar{x}$  for  $\lambda \downarrow 0$ . It follows that for every  $\varepsilon > 0$ , there exists  $\lambda > 0$  sufficiently small such that

$$\langle \xi, x_\lambda - \bar{x} \rangle < \varepsilon \|x_\lambda - \bar{x}\|,$$

which means, successively, that

$$\begin{aligned} \langle \xi, \lambda(x - \bar{x}) \rangle &< \varepsilon \|\lambda(x - \bar{x})\|, \\ \frac{\langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} &< \varepsilon, \quad \forall \varepsilon > 0, \forall x \in A \setminus \{\bar{x}\}, \\ \langle \xi, x - \bar{x} \rangle &\leq 0, \quad \forall x \in A. \end{aligned}$$

Now, since  $N_F(A, \bar{x}) \subset N_M(A, \bar{x})$ , take any  $\xi \in N_M(A, \bar{x})$ . There exist  $(x_n) \xrightarrow{A} \bar{x}$  and  $(\xi_n) \rightarrow \xi$  such that  $\xi_n \in N_F(A, x_n)$  for every  $n$ . It means, in view of (5.2.6), that

$$\langle \xi_n, x - x_n \rangle \leq 0, \quad \forall n \in \mathbb{N}, \forall x \in A.$$

Passing to the limit for  $n \rightarrow \infty$ , one gets the conclusion.

(v) Both equalities follow straightforward. □

The following representation of the Mordukhovich normal cone can be useful in computations.

**Theorem 5.2.4.** *Suppose  $A \subset \mathbb{R}^p$  is a closed set and let  $\bar{x} \in A$ . Then*

$$N_M(A, \bar{x}) = \left\{ \xi \in \mathbb{R}^p \mid \begin{array}{l} \exists (\lambda_n) \subset [0, \infty), (x_n) \rightarrow \bar{x}, (y_n) \subset \mathbb{R}^p \text{ such that} \\ y_n \in \text{pr}_A x_n \text{ for every } n \text{ and } \lambda_n(x_n - y_n) \rightarrow \xi \end{array} \right\} \quad (5.2.7)$$

*Proof* We will prove first that for every  $x \in A$ , the next inclusion holds:

$$N_F(A, x) \subset \left\{ \xi \in \mathbb{R}^p \mid \begin{array}{l} \exists (\lambda_n) \subset [0, \infty), (x_n) \rightarrow x, (y_n) \subset \mathbb{R}^p \text{ such that} \\ y_n \in \text{pr}_A x_n \text{ for every } n \text{ and } \lambda_n(x_n - y_n) \rightarrow \xi \end{array} \right\}. \quad (5.2.8)$$

Take  $x \in A$  and  $\xi \in N_F(A, x)$ . Set  $x_n := x + \frac{1}{n}\xi$  and pick  $y_n \in \text{pr}_A x_n$  for any  $n$ . Observe that  $y_n \in \text{pr}_A x_n$  if and only if

$$\|x_n - y_n\| \leq \|x_n - v\|, \quad \forall v \in A,$$

that is

$$\begin{aligned} 0 &\leq \|x_n - v\|^2 - \|x_n - y_n\|^2 = \langle x_n - v, x_n - y_n \rangle \\ &\quad + \langle x_n - v, y_n - v \rangle - \langle x_n - y_n, v - y_n \rangle - \langle x_n - y_n, x_n - v \rangle \\ &= \langle 2x_n - v - y_n, y_n - v \rangle = -2 \langle x_n - y_n, v - y_n \rangle + \|v - y_n\|^2. \end{aligned}$$



By replacing  $v$  with  $x$  and using the form of  $x_n$ , one gets

$$2 \left\langle x + \frac{1}{n} \xi - y_n, x - y_n \right\rangle \leq \|x - y_n\|^2, \\ n \|x - y_n\| \leq \frac{2 \langle \xi, y_n - x \rangle}{\|y_n - x\|}.$$

Moreover, because  $x \in A$  and  $y_n \in \text{pr}_A x_n$ , we know that

$$\|y_n - x\| \leq \|y_n - x_n\| + \|x - x_n\| \\ \leq \|x - x_n\| + \|x - x_n\| = \frac{2}{n} \|\xi\| \rightarrow 0,$$

hence, since  $y_n \xrightarrow{A} x$  and  $\xi \in N_F(A, x)$ , one has that  $\frac{2 \langle \xi, y_n - x \rangle}{\|y_n - x\|} \rightarrow 0$ . Consequently,  $n \|x - y_n\| \rightarrow 0$  and

$$n(x_n - y_n) = n(x - y_n) + \xi \rightarrow \xi,$$

which proves (5.2.8).

To prove now the “ $\subset$ ” in (5.2.7), take  $\xi \in N_M(A, \bar{x})$ . Then there exists  $x_n \xrightarrow{A} \bar{x}$ ,  $\xi_n \rightarrow \xi$  such that  $\xi_n \in N_F(A, x_n)$  for every  $n$ . For every  $\varepsilon > 0$ , there exists then  $n_0 \in \mathbb{N}$  such that, for every  $n \geq n_0$ ,

$$\|x_n - \bar{x}\| < 2^{-1} \varepsilon, \\ \|\xi_n - \xi\| < 2^{-1} \varepsilon.$$

Fix  $n \geq n_0$ . Since  $x_n \in A$ , one has using (5.2.8) that there exist  $(\lambda_k^n)_k \subset [0, \infty)$ ,  $(x_k^n)_k \rightarrow x_n$ ,  $(y_k^n)_k$  such that  $y_k^n \in \text{pr}_A x_k^n$  for every  $k$ , and  $\lambda_k^n (x_k^n - y_k^n) \rightarrow \xi_n$  for  $k \rightarrow \infty$ . For the  $\varepsilon$  chosen before, there are  $k_n^1, k_n^2 \geq n \geq n_0$  such that

$$\|x_k^n - x_n\| < 2^{-1} \varepsilon, \forall k \geq k_n^1, \\ \|\lambda_k^n (x_k^n - y_k^n) - \xi_n\| < 2^{-1} \varepsilon, \forall k \geq k_n^2.$$

Take  $k_n := \max \{k_n^1, k_n^2\}$  and set  $u_n := x_{k_n}^n$ ,  $v_n := y_{k_n}^n$ ,  $\alpha_n := \lambda_{k_n}^n$ . One has then

$$\|u_n - \bar{x}\| \leq \|u_n - x_n\| + \|x_n - \bar{x}\| < \varepsilon, \\ \|\alpha_n (u_n - v_n) - \xi\| \leq \|\alpha_n (u_n - v_n) - \xi_n\| + \|\xi_n - \xi\| < \varepsilon,$$

and since  $v_n \in \text{pr}_A u_n$ , we have proved the desired inclusion in (5.2.8).

Let us prove the “ $\supset$ ” inclusion in (5.2.8). Now consider  $\xi \in \mathbb{R}^p$  such that there exist  $(\lambda_n) \subset [0, \infty)$  and  $(x_n) \rightarrow \bar{x}$  such that  $\lambda_n (x_n - y_n) \rightarrow \xi$ , where  $y_n \in \text{pr}_A x_n$  for every  $n$ .

Take arbitrary  $v \in A$  and use the above characterization of the projection  $y_n \in \text{pr}_A x_n$  to deduce successively that

$$\langle x_n - y_n, v - y_n \rangle \leq \frac{1}{2} \|v - y_n\|^2,$$

$$\frac{\langle \lambda_n(x_n - y_n), v - y_n \rangle}{\|v - y_n\|} \leq \frac{\lambda_n}{2} \|v - y_n\|, \quad \forall v \in A.$$

Taking the  $\limsup$  for fixed  $n$  in the previous inequality and setting  $\xi_n := \lambda_n(x_n - y_n)$ , we get that

$$\limsup_{v \xrightarrow{A} y_n} \frac{\langle \xi_n, v - y_n \rangle}{\|v - y_n\|} \leq 0,$$

hence  $\xi_n \in N_F(A, y_n)$  for every  $n$ . Moreover, since  $x_n \rightarrow \bar{x}$  and  $y_n \in \text{pr}_A x_n$ , one can prove as above that  $y_n \rightarrow \bar{x}$ . It follows that  $\xi \in N_M(A, \bar{x})$  and the reverse inclusion is now proved.  $\square$

Let us provide some examples which contain the calculus of the Fréchet and Mordukhovich normal cones to some sets.

**Example 5.2.5.** Consider the set  $A := \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 = -|x_1|\}$ . Let us compute first  $N_F(A, (0, 0))$ . Take arbitrary  $\varepsilon > 0$ . One should find a neighborhood  $V$  of  $(0, 0)$  such that, for every  $(u, v) \in A \cap V$ ,

$$\langle (\xi, \eta), (u, v) \rangle < \varepsilon \|(u, v)\|.$$

The form of the set  $A$  allows us to deduce that

$$\xi u - \eta |u| < \sqrt{2}\varepsilon |u|,$$

for every  $u$  close to 0. One gets from the above inequality that  $\eta \geq |\xi|$ . As for every such element  $(\xi, \eta)$  and  $(u, v) \in A$  one has  $\langle (\xi, \eta), (u, v) \rangle \leq 0$ , we deduce that

$$N_F(A, (0, 0)) = \{(\xi, \eta) \in \mathbb{R}^2 \mid \eta \geq |\xi|\}.$$

Let us use now formula (5.2.7) to calculate  $N_M(A, (0, 0))$ . For this, let us observe first how the projection set of different elements from  $\mathbb{R}^2$  on  $A$  looks. If  $(x_1, x_2) \in \mathbb{R}^2$  is such that  $x_2 \geq |x_1|$ , then  $\text{pr}_A(x_1, x_2) = \{(0, 0)\}$ . If  $(x_1, x_2) \in \{(u, v) \mid u \geq 0, v \leq u\}$ , then  $\text{pr}_A(x_1, x_2)$  is the unique point from  $\{(\xi, -\xi) \mid \xi \geq 0\}$  such that the line passing through  $(x_1, x_2)$  is orthogonal on it. Similarly, if  $(x_1, x_2) \in \{(u, v) \mid u \leq 0, v \leq -u\}$ , then  $\text{pr}_A(x_1, x_2)$  is the unique point from  $\{(\xi, \xi) \mid \xi \leq 0\}$  such that the line passing through  $(x_1, x_2)$  is orthogonal to this line. It follows that when  $(x_n, y_n) \rightarrow (0, 0)$  and  $\lambda_n \geq 0$ , the all possible limits of sequences of the type  $\lambda_n((x_n, y_n) - \text{pr}_A(x_n, y_n))$  lie in the set

$$\{(\xi, \eta) \in \mathbb{R}^2 \mid \eta \geq |\xi| \text{ or } \eta = -|\xi|\},$$

which is equal to  $N_M(A, (0, 0))$ .

It can be observed from the previous example that the Mordukhovich normal cone is nonconvex in general, which shows that it cannot be the polar to any tangent cone, since polarity always implies convexity. It is not the case with the Fréchet normal cone, as the next result shows.

**Theorem 5.2.6.** *Let  $A \subset \mathbb{R}^p$  be a nonempty set and  $\bar{x} \in A$ . Then*

$$N_F(A, \bar{x}) = T_B(A, \bar{x})^-.$$

*Proof* Take  $\xi \in N_F(A, \bar{x})$  and choose arbitrary  $v \in T_B(A, \bar{x})$ . There exist sequences  $(t_n) \downarrow 0$  and  $(v_n) \rightarrow v$  such that  $\bar{x} + t_n v_n \in A$  for every  $n$ . Using the definition of the Fréchet normal cone, it follows that for every  $\varepsilon > 0$ , exists  $n_0 \in \mathbb{N}$  such that

$$t_n \langle \xi, v_n \rangle < \varepsilon t_n \|v_n\|, \quad \forall n \geq n_0.$$

This shows, after dividing by  $t_n$  and passing to the limit above, that  $\langle \xi, v \rangle \leq \varepsilon \|v\|$ . Letting  $\varepsilon \rightarrow 0$ , one gets that  $\langle \xi, v \rangle \leq 0$  and, furthermore, taking into account the arbitrariness of  $v$ , that  $\xi \in T_B(A, \bar{x})^-$ .

For the reverse inclusion, suppose  $\langle \xi, v \rangle \leq 0$  for every  $v \in T_B(A, \bar{x})$ . Now, choose an sequence  $x_n \xrightarrow{A} \bar{x}$  for which the lim sup from the definition of the Fréchet normal cone is attained:

$$\limsup_{x \xrightarrow{A} \bar{x}} \frac{\langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} = \lim_{n \rightarrow \infty} \frac{\langle \xi, x_n - \bar{x} \rangle}{\|x_n - \bar{x}\|}.$$

Since  $\bar{x} + \|x_n - \bar{x}\| \frac{x_n - \bar{x}}{\|x_n - \bar{x}\|} = x_n \in A$  for every  $n$ ,  $\|x_n - \bar{x}\| \downarrow 0$ , and without loss of generality we may suppose that  $\frac{x_n - \bar{x}}{\|x_n - \bar{x}\|}$  converges to some  $u$ . It follows that  $u \in T_B(A, \bar{x})$ , and hence

$$\lim_{n \rightarrow \infty} \left\langle \xi, \frac{x_n - \bar{x}}{\|x_n - \bar{x}\|} \right\rangle = \langle \xi, u \rangle \leq 0,$$

which shows that  $\xi \in N_F(A, \bar{x})$  and ends the proof. □

Theorem 5.1.32 shows that, in case of a Lipschitz function  $f$ ,  $N_C(\text{gr} f, \bar{x})$  is always a linear subspace, which drives us to the idea that, in many situations, especially when dealing with graphical sets, considering the Mordukhovich normal cone could be more appropriate.

Let us now formulate a result which gives us a smooth variational description of Fréchet normals, useful in many instances, including calculus rules for the Fréchet subdifferential of the difference of two functions.

**Theorem 5.2.7.** *Let  $A$  be a nonempty subset of  $\mathbb{R}^p$  and  $\bar{x} \in A$ . For  $\xi \in \mathbb{R}^p$ , suppose that there exists a function  $s : U \rightarrow \mathbb{R}$  defined on a neighborhood of  $\bar{x}$ , which is Fréchet differentiable at  $\bar{x}$ , and such that  $\nabla s(\bar{x}) = \xi$  and  $s$  achieves a local maximum relative to  $A$  at  $\bar{x}$ . Then  $\xi \in N_F(A, \bar{x})$ . Conversely, for every  $\xi \in N_F(A, \bar{x})$ , there is a function  $s : \mathbb{R}^p \rightarrow \mathbb{R}$  such that  $s(x) \leq s(\bar{x}) = 0$  for every  $x \in A$  and  $s$  is Fréchet differentiable at  $\bar{x}$  with  $\nabla s(\bar{x}) = \xi$ .*

*Proof* For the direct implication, we know that there is a continuous function  $\alpha$  with  $\lim_{x \rightarrow \bar{x}} \frac{\alpha(\|x - \bar{x}\|)}{\|x - \bar{x}\|} = 0$  such that

$$s(x) = s(\bar{x}) + \langle \xi, x - \bar{x} \rangle + \alpha(\|x - \bar{x}\|) \leq s(\bar{x}),$$

for every  $x \in A$  near  $\bar{x}$ . It means that  $\langle \xi, x - \bar{x} \rangle + \alpha(\|x - \bar{x}\|) \leq 0$  for every  $x \in A$  near  $\bar{x}$ , which implies for such  $x$  that

$$\frac{\langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} \leq -\frac{\alpha(\|x - \bar{x}\|)}{\|x - \bar{x}\|}.$$

Passing to  $\limsup$  for  $x \xrightarrow{A} \bar{x}$ , we deduce that  $\xi \in N_F(A, \bar{x})$ .

For the converse, consider the function  $s : \mathbb{R}^p \rightarrow \mathbb{R}$  given by

$$s(x) := \begin{cases} \min \{0, \langle \xi, x - \bar{x} \rangle\}, & \text{if } x \in A \\ \langle \xi, x - \bar{x} \rangle, & \text{otherwise.} \end{cases}$$

Remark that  $s(x) \leq 0 = s(\bar{x})$  for every  $x \in A$  and  $s(x) \leq \langle \xi, x - \bar{x} \rangle$  for every  $x \in \mathbb{R}^p$ . Then

$$\begin{aligned} \left| \frac{s(x) - s(\bar{x}) - \langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} \right| &= \frac{\langle \xi, x - \bar{x} \rangle - s(x)}{\|x - \bar{x}\|} \\ &= \begin{cases} 0, & \text{if } x \notin A \text{ or } x \in A \text{ and } \langle \xi, x - \bar{x} \rangle \leq 0 \\ \frac{\langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|}, & \text{if } x \in A \text{ and } \langle \xi, x - \bar{x} \rangle > 0. \end{cases} \end{aligned}$$

Since  $\xi \in N_F(A, \bar{x})$ , for every  $\varepsilon > 0$  there exists a neighborhood  $V$  of  $\bar{x}$  such that  $\frac{\langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} < \varepsilon$  for every  $x \in A \cap V$ , which means that

$$\left| \frac{s(x) - s(\bar{x}) - \langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} \right| < \varepsilon, \quad \forall x \in V.$$

But this means that  $s$  is Fréchet differentiable at  $\bar{x}$  with  $\nabla s(\bar{x}) = \xi$ . □

We close this section with a result useful for the next.

**Proposition 5.2.8.** *Suppose  $A \subset \mathbb{R}^p$  is a closed set and  $\bar{x} \in \text{bd } A$ . Then*

$$N_M(A, \bar{x}) \subset N_M(\text{bd } A, \bar{x}).$$

*Proof* Pick  $\xi \in N_M(A, \bar{x}) \setminus \{0\}$ . According to the definition of the Mordukhovich normal cone, there exist  $(x_n) \xrightarrow{A} \bar{x}$  and  $(\xi_n) \rightarrow \xi$  such that  $\xi_n \in N_F(A, x_n)$  for every  $n$ . Due to the continuity of the norm, it follows that  $\|\xi_n\| \rightarrow \|\xi\| > 0$ , hence  $x_n \notin \text{int } A$ , i.e.,  $x_n \in \text{bd } A$  for every  $n$  sufficiently large, because otherwise one would have  $\xi_n = 0$ . From this,

$$\limsup_{x \xrightarrow{\text{bd } A} x_n} \frac{\langle \xi_n, x - x_n \rangle}{\|x - x_n\|} \leq \limsup_{x \xrightarrow{A} x_n} \frac{\langle \xi_n, x - x_n \rangle}{\|x - x_n\|} \leq 0,$$

hence  $\xi_n \in N_F(\text{bd } A, x_n)$  for every  $n$ , so  $\xi \in N_M(\text{bd } A, \bar{x})$ . □

### 5.2.2 Fréchet and Mordukhovich Subdifferentials

Let us introduce now the subdifferentials associated to the normal cones defined before. We begin with the Fréchet subdifferential.

**Definition 5.2.9.** Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ . The set

$$\partial_F f(\bar{x}) := \{ \xi \in \mathbb{R}^p \mid (\xi, -1) \in N_F(\text{epi } f, (\bar{x}, f(\bar{x}))) \}$$

is called the Fréchet subdifferential of  $f$  at  $\bar{x}$ , and its elements are called Fréchet subgradients of  $f$  at  $\bar{x}$ .

Observe that, since the Fréchet normal cone is a closed convex cone, the Fréchet subdifferential is a closed convex set in  $\mathbb{R}^p$ . The next example shows that the subdifferential can be empty, even for simple Lipschitz functions.

**Example 5.2.10.** Consider the function  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = -|x|$ . Then  $\text{epi } f = \{(x_1, x_2) \mid x_2 \geq -|x_1|\}$ . One can easily see that  $T_B(\text{epi } f, (0, 0)) = \text{epi } f$ , hence

$$N_F(\text{epi } f, (0, 0)) = (\text{epi } f)^- = \{(0, 0)\}.$$

It follows that  $\partial_F f(0, 0) = \emptyset$ .

Let us provide an analytical characterization of the Fréchet subgradients.

**Theorem 5.2.11.** Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ . Then

$$\partial_F f(\bar{x}) = \left\{ \xi \in \mathbb{R}^p \mid \liminf_{x \rightarrow \bar{x}} \frac{f(x) - f(\bar{x}) - \langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0 \right\}. \tag{5.2.9}$$

*Proof* Take  $\xi$  from the right-hand side set. We want to prove that  $\xi \in \partial_F f(\bar{x})$ . Observe that, from the choice of  $\xi$ , for every  $\varepsilon > 0$ , there exists a neighborhood  $V$  of  $\bar{x}$  such that, for every  $x \in V$ , one has

$$f(x) - f(\bar{x}) - \langle \xi, x - \bar{x} \rangle \geq -\varepsilon \|x - \bar{x}\|.$$

Pick such  $x$  and fix  $\alpha \geq f(x)$ . Then

$$\begin{aligned} \langle (\xi, -1), (x, \alpha) - (\bar{x}, f(\bar{x})) \rangle &= \langle \xi, x - \bar{x} \rangle + f(\bar{x}) - \alpha \\ &\leq \langle \xi, x - \bar{x} \rangle + f(\bar{x}) - f(x) \\ &\leq \varepsilon \|x - \bar{x}\| \\ &\leq \varepsilon \|(x, \alpha) - (\bar{x}, f(\bar{x}))\|, \end{aligned}$$

which proves that  $(\xi, -1) \in N_F(\text{epi } f, (\bar{x}, f(\bar{x})))$ .

For the converse inclusion, take  $\xi \in \partial_F f(\bar{x})$  and suppose that there is  $\varepsilon > 0$  such that

$$\sup_{V \in \mathcal{V}(\bar{x})} \inf_{x \in V} \frac{f(x) - f(\bar{x}) - \langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} < -\varepsilon.$$

It means that one can construct a sequence  $x_n \rightarrow \bar{x}$  such that

$$\frac{f(x_n) - f(\bar{x}) - \langle \xi, x_n - \bar{x} \rangle}{\|x_n - \bar{x}\|} < -\varepsilon,$$

which implies that  $(x_n, \alpha_n) \in \text{epi } f$ , where  $\alpha_n := f(\bar{x}) + \langle \xi, x_n - \bar{x} \rangle - \varepsilon \|x_n - \bar{x}\|$ . We have

$$\begin{aligned} \|(x_n, \alpha_n) - (\bar{x}, f(\bar{x}))\| &\leq \|x_n - \bar{x}\| + |\langle \xi, x_n - \bar{x} \rangle - \varepsilon \|x_n - \bar{x}\|| \\ &\leq (1 + \|\xi\| + \varepsilon) \|x_n - \bar{x}\|, \end{aligned}$$

whence

$$\begin{aligned} \frac{\langle (\xi, -1), (x_n, \alpha_n) - ((\bar{x}, f(\bar{x}))) \rangle}{\|(x_n, \alpha_n) - ((\bar{x}, f(\bar{x})))\|} &= \frac{\varepsilon \|x_n - \bar{x}\|}{\|(x_n, \alpha_n) - ((\bar{x}, f(\bar{x})))\|} \\ &\geq \frac{\varepsilon}{1 + \|\xi\| + \varepsilon}. \end{aligned}$$

It follows that  $(\xi, -1) \notin N_F(\text{epi } f, (\bar{x}, f(\bar{x})))$ , a contradiction. The proof is now complete.  $\square$

The next proposition partially justifies the name of Fréchet subgradients for the elements of  $\partial_F f(\bar{x})$ .

**Proposition 5.2.12.** *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be Fréchet differentiable at  $\bar{x}$ . Then  $\partial_F f(\bar{x}) = \{\nabla f(\bar{x})\}$ .*

*Proof* It is obvious from the definition of Fréchet differentiability and the analytical characterization of the Fréchet subgradients that  $\nabla f(\bar{x}) \in \partial_F f(\bar{x})$ .

Suppose  $\xi \in \partial_F f(\bar{x})$ . Using (5.2.9), one gets that

$$\liminf_{x \rightarrow \bar{x}} \frac{\langle \nabla f(\bar{x}) - \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0,$$

which means that for every  $\varepsilon > 0$ , there exists  $\delta > 0$  such that, for every  $x \in D(\bar{x}, \delta)$ ,

$$\langle \xi - \nabla f(\bar{x}), x - \bar{x} \rangle \leq \varepsilon \|x - \bar{x}\|.$$

It follows that

$$\|\xi - \nabla f(\bar{x})\| \leq \varepsilon$$

for every positive  $\varepsilon$ , i.e.,  $\xi = \nabla f(\bar{x})$ , and hence the conclusion.  $\square$

The next smooth variational description of Fréchet subgradients follows from Theorem 5.2.7.

**Theorem 5.2.13.** *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be a function. Then  $\xi \in \partial_F f(\bar{x})$  if and only if there exists a function  $s : \mathbb{R}^p \rightarrow \mathbb{R}$ , Fréchet differentiable at  $\bar{x}$ , which satisfies:*

$$s(\bar{x}) = f(\bar{x}), \quad s(x) \leq f(x), \quad \forall x \in \mathbb{R}^p, \quad \text{and } \nabla s(\bar{x}) = \xi. \tag{5.2.10}$$

*Proof* Suppose  $\xi \in \partial_F f(\bar{x})$ . Consider the function  $s : \mathbb{R}^p \rightarrow \mathbb{R}$  given as

$$s(x) := \min \{f(x), f(\bar{x}) + \langle \xi, x - \bar{x} \rangle\}, \quad \forall x \in \mathbb{R}^p.$$

Then, obviously,  $s(\bar{x}) = f(\bar{x})$  and  $s(x) \leq f(x)$  for every  $x \in \mathbb{R}^p$  from the definition. Also from the definition of  $s$ , it immediately follows that

$$\limsup_{x \rightarrow \bar{x}} \frac{s(x) - s(\bar{x}) - \langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} \leq 0. \tag{5.2.11}$$

Moreover, one has

$$\frac{s(x) - s(\bar{x}) - \langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} = \begin{cases} 0, & \text{if } f(x) \geq f(\bar{x}) + \langle \xi, x - \bar{x} \rangle \\ \frac{f(x) - f(\bar{x}) - \langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|}, & \text{otherwise,} \end{cases}$$

which implies, using  $\xi \in \partial_F f(\bar{x})$ , that

$$\liminf_{x \rightarrow \bar{x}} \frac{s(x) - s(\bar{x}) - \langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0. \tag{5.2.12}$$

Relations (5.2.11) and (5.2.12) mean that  $s$  is Fréchet differentiable at  $\bar{x}$  and  $\nabla s(\bar{x}) = \xi$ .

Conversely, suppose  $\xi \in \mathbb{R}^p$  and there exists an  $s : \mathbb{R}^p \rightarrow \mathbb{R}$ , Fréchet differentiable at  $\bar{x}$  and satisfying (5.2.10). Then

$$\liminf_{x \rightarrow \bar{x}} \frac{f(x) - f(\bar{x}) - \langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq \liminf_{x \rightarrow \bar{x}} \frac{s(x) - s(\bar{x}) - \langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} = 0,$$

hence  $\xi \in \partial_F f(\bar{x})$ . □

If  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  is convex, then the Fréchet subdifferential reduces to the usual convex subdifferential given in the previous chapter.

**Proposition 5.2.14.** *Suppose  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  is convex. Then*

$$\partial_F f(\bar{x}) = \partial_M f(\bar{x}) = \partial f(\bar{x}).$$

*Proof* This immediately follows from the fact that  $\text{epi } f$  is a convex set and from the formula (5.2.5). □

Let us proceed now in introducing the Mordukhovich subdifferential. For this, we need first the next result.

**Proposition 5.2.15.** *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be a function and  $(\bar{x}, \bar{\alpha}) \in \text{epi } f$ . Then  $\lambda \geq 0$  for every  $(\xi, -\lambda) \in N_M(\text{epi } f, (\bar{x}, \bar{\alpha}))$ , hence there exist the subsets  $D$  and  $D^\infty$  of  $\mathbb{R}^p$  such that*

$$N_M(\text{epi } f, (\bar{x}, \bar{\alpha})) = \{\lambda(\xi, -1) \mid \xi \in D, \lambda > 0\} \cup \{(\xi, 0) \mid \xi \in D^\infty\}.$$

*Proof* Choose arbitrary  $(\xi, -\lambda) \in N_M(\text{epi } f, (\bar{x}, \bar{\alpha}))$ . Then by formula (5.2.2), it follows that  $\exists(x_n, \alpha_n) \xrightarrow{\text{epi } f} (\bar{x}, \bar{\alpha})$ ,  $\xi_n \rightarrow \xi$ , and  $\lambda_n \rightarrow \lambda$ , such that

$$\limsup_{(x, \alpha) \xrightarrow{\text{epi } f} (x_n, \alpha_n)} \frac{\langle \xi_n, x - x_n \rangle - \lambda_n(\alpha - \alpha_n)}{\|(x, \alpha) - (x_n, \alpha_n)\|} \leq 0, \quad \forall n.$$

Then, for every  $\varepsilon > 0$ , there exists  $\delta_{\varepsilon, n} > 0$  such that, for every  $(x, \alpha) \in D((x_n, \alpha_n), \delta)$ , one has

$$\frac{\langle \xi_n, x - x_n \rangle - \lambda_n(\alpha - \alpha_n)}{\|(x, \alpha) - (x_n, \alpha_n)\|} < \varepsilon.$$

By taking  $x := x_n$  and  $\alpha := \alpha_n + \delta$ , it follows that  $(x, \alpha) \in D((x_n, \alpha_n), \delta)$ , hence

$$\frac{-\lambda_n(\alpha - \alpha_n)}{|\alpha - \alpha_n|} = -\lambda_n < \varepsilon.$$

Passing to the limit for  $n \rightarrow \infty$ , one gets that  $-\lambda \leq \varepsilon$  for every positive  $\varepsilon$ , hence the conclusion. □

**Definition 5.2.16.** Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be a function and  $\bar{x} \in \mathbb{R}^p$ .

(i) The set

$$\partial_M f(\bar{x}) := \{ \xi \in \mathbb{R}^p \mid (\xi, -1) \in N_M(\text{epi } f, (\bar{x}, f(\bar{x}))) \}$$

is called the Mordukhovich (or limiting, or basic) subdifferential of  $f$  at  $\bar{x}$ , and its elements are called the basic subgradients of  $f$  at  $\bar{x}$ .

(ii) The set

$$\partial^\infty f(\bar{x}) := \{ \xi \in \mathbb{R}^p \mid (\xi, 0) \in N_M(\text{epi } f, (\bar{x}, f(\bar{x}))) \}$$

is called the singular subdifferential of  $f$  at  $\bar{x}$ , and its elements are called the singular subgradients of  $f$  at  $\bar{x}$ .

It follows from the definitions that for every function  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ , and every  $\bar{x} \in \mathbb{R}^p$ ,

$$\partial_F f(\bar{x}) \subset \partial_M f(\bar{x}). \tag{5.2.13}$$

**Example 5.2.17.** Consider again the function  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) := -|x|$ . Since  $\text{epi } f = \{(x_1, x_2) \mid x_2 \geq -|x_1|\}$ , one easily deduces, using the formula (5.2.7), that

$$N_M(\text{epi } f, (0, 0)) = \{(x, x) \mid x \leq 0\} \cup \{(x, -x) \mid x > 0\}.$$

Hence,  $\partial_M f(0) = \{-1, 1\}$  and  $\partial^\infty f(0) = \{0\}$ . Recall that  $\partial_F f(0) = \emptyset$ , hence the inclusion (5.2.13) can be strict. Note that the Mordukhovich subdifferential is a nonconvex set, and this should not be surprising, since its definition is based on the Mordukhovich normal cone, which is nonconvex in general.



The next result emphasizes estimates for the subdifferential of locally Lipschitz functions.

**Theorem 5.2.18.** *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ . If  $f$  is locally Lipschitz around  $\bar{x}$  with modulus  $\ell \geq 0$ , then*

$$\|\xi\| \leq \ell, \forall \xi \in \partial_M f(\bar{x}) \tag{5.2.14}$$

and

$$\partial^\infty f(\bar{x}) = \{0\}. \tag{5.2.15}$$

*Proof* Take  $(\xi, \lambda) \in N_M(\text{epi } f, (\bar{x}, f(\bar{x})))$ . Since  $f$  is continuous,  $\text{epi } f$  is closed, and since  $\text{gr } f = \text{bd epi } f$ , it follows from Proposition 5.2.8 that  $(\xi, \lambda) \in N_M(\text{gr } f, (\bar{x}, f(\bar{x})))$ . Then there exist  $(x_n) \rightarrow \bar{x}$ ,  $(\xi_n) \rightarrow \xi$  and  $(\lambda_n) \rightarrow \lambda$  such that  $(\xi_n, \lambda_n) \in N_F(\text{gr } f, (x_n, f(x_n)))$  for every  $n$ .

Now, from the Lipschitz property of  $f$ , we know that there exists  $\delta > 0$  such that, for every  $x, u \in D(\bar{x}, 2\delta)$ , one has

$$|f(x) - f(u)| \leq \ell \|x - u\|. \tag{5.2.16}$$

Moreover, for every  $\varepsilon > 0$ , and for every  $n$ , there is  $y \leq \min \{\delta, \ell\delta\}$  such that, for every  $u \in D(x_n, y)$ ,

$$\langle \xi_n, u - x_n \rangle + \lambda_n(f(u) - f(x_n)) \leq \varepsilon (\|u - x_n\| + |f(u) - f(x_n)|). \tag{5.2.17}$$

Take  $n$  sufficiently large such that  $x_n \in D(\bar{x}, \delta)$  and  $u \in D(x_n, y\ell^{-1})$ . It follows that

$$\|u - \bar{x}\| \leq \|u - x_n\| + \|x_n - \bar{x}\| \leq y\ell^{-1} + \delta \leq 2\delta.$$

One can employ now (5.2.17) and (5.2.16) for  $u$  and  $x_n$  in order to get that

$$\begin{aligned} \langle \xi_n, u - x_n \rangle &\leq -\lambda_n(f(u) - f(x_n)) + \varepsilon (\|u - x_n\| + |f(u) - f(x_n)|) \\ &\leq (|\lambda_n| \ell + \varepsilon + \varepsilon\ell) \|u - x_n\| \\ &\leq |\lambda_n| y + \varepsilon y\ell^{-1} + \varepsilon y. \end{aligned}$$

Since the previous relation holds for every  $u \in D(x_n, y\ell^{-1})$ , it follows that

$$y\ell^{-1} \|\xi_n\| \leq y (|\lambda_n| + \varepsilon\ell^{-1} + \varepsilon).$$

Dividing by  $y$ , passing to the limit for  $n \rightarrow \infty$ , and then for  $\varepsilon \rightarrow 0$ , one gets that

$$\ell^{-1} \|\xi\| \leq |\lambda|. \tag{5.2.18}$$

Now, if  $\xi \in \partial_M f(\bar{x})$ , it means that  $\lambda$  from above equals  $-1$ , hence from (5.2.18) one gets (5.2.14). If  $\xi \in \partial^\infty f(\bar{x})$ , it means that  $\lambda$  from above is 0 and (5.2.15) follows.  $\square$

**Remark 5.2.19.** *In view of relation (5.2.13), the previous theorem also shows that for every locally Lipschitz function  $f$  around a point  $\bar{x}$ , having the Lipschitz modulus  $\ell \geq 0$ , one has*

$$\|\xi\| \leq \ell, \forall \xi \in \partial_F f(\bar{x}). \tag{5.2.19}$$

We now emphasize a case when the Mordukhovich subdifferential reduces to the usual Fréchet differential.

**Proposition 5.2.20.** *Suppose  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  is  $C^1$  around  $\bar{x}$ . Then*

$$\partial_M f(\bar{x}) = \{ \nabla f(\bar{x}) \}.$$

*Proof* Pick  $\xi \in \partial_M f(\bar{x})$ . Then  $(\xi, -1) \in N_M(\text{epi } f, (\bar{x}, f(\bar{x}))) \subset N_M(\text{gr } f, (\bar{x}, f(\bar{x})))$ , which means that there exist  $(x_n) \rightarrow \bar{x}$ ,  $(\xi_n) \rightarrow \xi$  and  $(\lambda_n) \rightarrow 1$  such that for every  $\varepsilon > 0$ , there exists  $\delta > 0$  such that for every  $x \in D(x_n, \delta)$ ,

$$\langle \xi_n, x - x_n \rangle - \lambda_n(f(x) - f(x_n)) \leq \varepsilon (\|x - x_n\| + |f(x) - f(x_n)|).$$

Using the Taylor expansion for the function  $f$  at every such point  $x_n$ , one gets that there is  $y_n$  on the segment  $[x, x_n]$  such that

$$f(x) - f(x_n) = \langle \nabla f(y_n), x - x_n \rangle.$$

It means that

$$\begin{aligned} \langle \xi_n - \lambda_n \nabla f(y_n), x - x_n \rangle &\leq \varepsilon (\|x - x_n\| + |f(x) - f(x_n)|) \\ &\leq (\varepsilon + \varepsilon \ell) \|x - x_n\| \\ &\leq (\varepsilon + \varepsilon \ell) \delta, \forall x \in D(x_n, \delta), \end{aligned}$$

where  $\ell > 0$  is a Lipschitz modulus for the  $C^1$  function  $f$ . Hence,

$$\delta \|\xi_n - \lambda_n \nabla f(y_n)\| \leq (\varepsilon + \varepsilon \ell) \delta.$$

Passing to the limit for  $\varepsilon \rightarrow 0$  and  $n \rightarrow \infty$ , and taking into account that  $y_n \rightarrow \bar{x}$ , hence  $\nabla f(y_n) \rightarrow \nabla f(\bar{x})$  since  $f$  is  $C^1$ , one gets that  $\xi = \nabla f(\bar{x})$ . □

The representation of singular subgradients will be useful in the following sections. The notation  $x_n \xrightarrow{f} \bar{x}$  means  $x_n \rightarrow \bar{x}$  and  $f(x_n) \rightarrow f(\bar{x})$ .

**Theorem 5.2.21** (representation of singular subgradients). *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be lower semicontinuous around  $\bar{x}$ . Then*

$$\partial_M f(\bar{x}) = \left\{ \xi \in \mathbb{R}^p \mid \exists x_n \xrightarrow{f} \bar{x}, \xi_n \rightarrow \xi \text{ such that } \xi_n \in \partial_F f(x_n), \forall n \in \mathbb{N} \right\}. \tag{5.2.20}$$

*Proof* The “ $\supset$ ” inclusion follows immediately. Suppose  $\xi \in \partial_M f(\bar{x})$ , i.e.,  $(\xi, -1) \in N_M(\text{epi } f, (\bar{x}, f(\bar{x})))$ . Since  $f$  is lower semicontinuous around  $\bar{x}$ ,  $\text{epi } f$  is locally closed around  $(\bar{x}, f(\bar{x}))$ , hence by employing Proposition 5.2.8,  $(\xi, -1) \in N_M(\text{gr } f, (\bar{x}, f(\bar{x})))$ . There exist then  $(x_n, f(x_n)) \rightarrow (\bar{x}, f(\bar{x}))$  and  $(\xi_n, \lambda_n) \rightarrow (\xi, -1)$  such that  $(\xi_n, -\lambda_n) \in N_F(\text{gr } f, (x_n, f(x_n)))$  for every  $n$ . It means that for every  $\varepsilon > 0$ , there exists  $\delta > 0$  such that

$$\langle \xi_n, x - x_n \rangle - \lambda_n(f(x) - f(x_n)) \leq \varepsilon (\|x - x_n\| + |f(x) - f(x_n)|)$$

for every  $x \in D(x_n, \delta)$  and  $f(x) \in D(f(x_n), \delta)$ . Since  $\lambda_n \rightarrow 1$ , for every  $\varepsilon > 0$ , and every  $n$  such that  $|\lambda_n - 1| < \varepsilon$ , one has from above that

$$\begin{aligned} \langle \xi_n, x - x_n \rangle - (f(x) - f(x_n)) &\leq \varepsilon (\|x - x_n\| + |f(x) - f(x_n)|) + |\lambda_n - 1| \cdot |f(x) - f(x_n)| \\ &\leq 2\varepsilon (\|x - x_n\| + |f(x) - f(x_n)|), \end{aligned}$$

for every  $x \in D(x_n, \delta)$  and  $f(x) \in D(f(x_n), \delta)$ . It means that  $(\xi_n, -1) \in N_F(\text{gr } f, (x_n, f(x_n)))$  for every  $n$  sufficiently large. Suppose, by contradiction, that  $(\xi_n, -1) \notin N_F(\text{epi } f, (x_n, f(x_n)))$ . Pick  $y \in (0, 1)$  and sequences  $(u_j, \alpha_j) \xrightarrow{\text{epi } f} (x_n, f(x_n))$  such that

$$\langle \xi_n, u_j - x_n \rangle - (\alpha_j - f(x_n)) > y \|(u_j, \alpha_j) - (x_n, f(x_n))\|. \tag{5.2.21}$$

Since  $f$  is lower semicontinuous at  $x_n$ , we have

$$f(x_n) = \lim_{j \rightarrow \infty} \alpha_j \geq \limsup_{j \rightarrow \infty} f(u_j) \geq \liminf_{j \rightarrow \infty} f(u_j) \geq f(x_n),$$

hence  $f(u_j) \rightarrow f(x_n)$ . Moreover,

$$\|(u_j, f(u_j)) - (x_n, f(x_n))\| \leq \|(u_j, \alpha_j) - (x_n, f(x_n))\| + \alpha_j - f(u_j).$$

It follows from (5.2.21) that

$$\begin{aligned} \langle \xi_n, u_j - x_n \rangle - (\alpha_j - f(x_n)) &> y \|(u_j, f(u_j)) - (x_n, f(x_n))\| - y(\alpha_j - f(u_j)) \\ &\geq y \|(u_j, f(u_j)) - (x_n, f(x_n))\| - (\alpha_j - f(u_j)), \end{aligned}$$

i.e.,

$$\langle \xi_n, u_j - x_n \rangle - (f(u_j) - f(x_n)) > y \|(u_j, f(u_j)) - (x_n, f(x_n))\|,$$

with  $u_j \rightarrow x_n$  and  $f(u_j) \rightarrow f(x_n)$ , which contradicts  $(\xi_n, -1) \in N_F(\text{gr } f, (x_n, f(x_n)))$ . Hence,  $(\xi_n, -1) \in N_F(\text{epi } f, (x_n, f(x_n)))$ , i.e.,  $\xi_n \in \partial_F f(x_n)$ , and the desired inclusion is now proved.  $\square$

**Example 5.2.22.** Let us consider the function  $f : \mathbb{R} \rightarrow \mathbb{R}$  given as

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x}, & \text{if } x \neq 0 \\ 0, & \text{if } x = 0. \end{cases}$$

Using formula (5.2.20), one gets that  $\partial_M f(0) = [-1, 1]$ , while  $\nabla f(0) = 0$ . This shows that the Mordukhovich subdifferential does not reduce in general to the gradient, in case of a differentiable function.

The following result is yet another nonsmooth version of the Fermat rule.

**Theorem 5.2.23** (Fermat rule for Mordukhovich calculus). *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ . If  $f$  has a local minimum at  $\bar{x}$ , then*

$$0 \in \partial_F f(\bar{x}) \subset \partial_M f(\bar{x}).$$

*Proof* Since  $f(x) \geq f(\bar{x})$  for every  $x$  close to  $\bar{x}$ , it follows that

$$\liminf_{x \rightarrow \bar{x}} \frac{f(x) - f(\bar{x}) - \langle 0, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0,$$

hence  $0 \in \partial_F f(\bar{x})$  due to the analytical characterization of Fréchet subgradients. The inclusion between the two subdifferentials is true in general.  $\square$

We continue by presenting some considerations about the links between the Fréchet and Mordukhovich normal cones and the corresponding subdifferentials of the distance function.

**Proposition 5.2.24.** *For any set  $A \subset \mathbb{R}^p$  and  $\bar{x} \in A$ , one has*

$$\partial_F d_A(\bar{x}) = N_F(A, \bar{x}) \cap D(0, 1), \quad N_F(A, \bar{x}) = \bigcup_{\lambda > 0} \lambda \partial_F d_A(\bar{x}). \quad (5.2.22)$$

*Proof* Take  $\xi \in \partial_F d_A(\bar{x})$ . Using the analytical description of the Fréchet subgradients given by (5.2.9), we get that for every  $\varepsilon > 0$ , there is  $\delta > 0$  such that, for any  $x \in D(\bar{x}, \delta)$ , one has

$$d_A(x) - d_A(\bar{x}) - \langle \xi, x - \bar{x} \rangle \geq -\varepsilon \|x - \bar{x}\|. \quad (5.2.23)$$

This means that for any  $x \in A \cap D(\bar{x}, \delta)$ , one has  $\langle \xi, x - \bar{x} \rangle \leq \varepsilon \|x - \bar{x}\|$ , which shows that  $\xi \in N_F(A, \bar{x})$ . Moreover, since  $d_A$  is 1-Lipschitz, we get from (5.2.23) that

$$\langle \xi, x - \bar{x} \rangle \leq (\varepsilon + 1) \|x - \bar{x}\| \leq (\varepsilon + 1)\delta, \quad \forall x \in D(\bar{x}, \delta),$$

hence  $\delta \|\xi\| \leq (\varepsilon + 1)\delta$ , i.e.,  $\|\xi\| \leq 1 + \varepsilon$  for every  $\varepsilon > 0$ . The “ $\subset$ ” inclusion from the first relation in (5.2.22) holds.

Take now  $\xi \in N_F(A, \bar{x})$  such that  $\|\xi\| \leq 1$ . For  $x \notin A$ , we find  $u \in A$  such that

$$\|x - u\| \leq d_A(x) + \|x - \bar{x}\|^2.$$

Then

$$\begin{aligned} \|u - \bar{x}\| &\leq \|u - x\| + \|x - \bar{x}\| \\ &\leq d_A(x) + \|x - \bar{x}\|^2 + \|x - \bar{x}\| \\ &\leq 3 \|x - \bar{x}\|, \end{aligned}$$

for any  $x$  close to  $\bar{x}$ . Hence

$$\begin{aligned} & \liminf_{\substack{x \rightarrow \bar{x} \\ x \notin A}} \frac{d_A(x) - d_A(\bar{x}) - \langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq \liminf_{\substack{x \rightarrow \bar{x} \\ x \notin A}} \frac{\|x - u\| - \|x - \bar{x}\|^2 - \langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} \\ & \geq \liminf_{\substack{x \rightarrow \bar{x} \\ x \notin A}} \frac{\|x - u\| - \langle \xi, x - u \rangle}{\|x - \bar{x}\|} - \limsup_{\substack{x \rightarrow \bar{x} \\ x \notin A}} \frac{\langle \xi, u - \bar{x} \rangle}{\|x - \bar{x}\|} \\ & \geq \liminf_{\substack{x \rightarrow \bar{x} \\ x \notin A}} \frac{(1 - \|\xi\|)\|x - u\|}{\|x - \bar{x}\|} - \limsup_{\substack{x \rightarrow \bar{x} \\ x \notin A}} \frac{\langle \xi, u - \bar{x} \rangle}{\|x - \bar{x}\|} \geq - \limsup_{\substack{x \rightarrow \bar{x} \\ x \notin A}} \frac{\langle \xi, u - \bar{x} \rangle}{\|x - \bar{x}\|}. \end{aligned}$$

If  $\langle \xi, u - \bar{x} \rangle > 0$  (hence  $u \neq \bar{x}$ ), then

$$\frac{\langle \xi, u - \bar{x} \rangle}{\|x - \bar{x}\|} \leq \frac{3 \langle \xi, u - \bar{x} \rangle}{\|u - \bar{x}\|}.$$

This implies that  $-\limsup_{\substack{x \rightarrow \bar{x} \\ x \notin A}} \frac{\langle \xi, u - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0$ , since  $\xi \in N_F(A, \bar{x})$ . Because in case that  $\langle \xi, u - \bar{x} \rangle \leq 0$  one trivially has the same inequality, we have hence

$$\liminf_{\substack{x \rightarrow \bar{x} \\ x \notin A}} \frac{d_A(x) - d_A(\bar{x}) - \langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0.$$

Moreover, from  $\xi \in N_F(A, \bar{x})$ , we get that

$$\liminf_{\substack{x \rightarrow \bar{x} \\ x \in A}} \frac{d_A(x) - d_A(\bar{x}) - \langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} = - \limsup_{\substack{x \rightarrow \bar{x} \\ x \in A}} \frac{\langle \xi, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0.$$

This implies that  $\xi \in \partial_F d_A(\bar{x})$ , and the first equality from (5.2.22) is completely proved.

To finish the proof, observe that the second equality from (5.2.22) easily follows from the first one. □

**Theorem 5.2.25.** *Let  $A \subset \mathbb{R}^p$  be a closed set and  $\bar{x} \in A$ . Then*

$$N_M(A, \bar{x}) = \bigcup_{\lambda > 0} \lambda \partial_M d_A(\bar{x}). \tag{5.2.24}$$

*Proof* For the direct inclusion, take  $\xi \in N_M(A, \bar{x}) \setminus \{0\}$ , and find by definition sequences  $(x_n) \xrightarrow{A} \bar{x}$  and  $(\xi_n) \rightarrow \xi$  such that  $\xi_n \in N_F(A, x_n)$  for any  $n$ . Since  $(\xi_n)$  is bounded, there is a bounded sequence of  $\lambda_n > 0$  such that  $\frac{\|\xi_n\|}{\lambda_n} \leq 1$  for any  $n$ . It means that  $\frac{\xi_n}{\lambda_n} \in \partial_F d_A(x_n)$  for any  $n$ , by the first relation from (5.2.22). Without loss of generality, because  $\lambda_n \geq \|\xi_n\|$ , we may suppose that  $\lambda_n$  converges to some  $\lambda > 0$ . Moreover, using the representation of the Mordukhovich subgradients (5.2.20), we get that  $\xi \in \lambda \partial_M d_A(\bar{x})$ .

Suppose now  $A$  is closed and prove the opposite inclusion in (5.2.24). Take  $\xi \in \partial_M d_A(\bar{x})$ , and find by (5.2.20) some sequences  $(x_n) \rightarrow \bar{x}$  and  $(\xi_n) \rightarrow \xi$  such that  $\xi_n \in$

$\partial_F d_A(x_n)$  for any  $n$ . If there is  $(x_k)$ , a subsequence of  $(x_n)$ , which belongs to  $A$ , then the desired inclusion follows by passing to the limit for  $k \rightarrow \infty$  and taking into account that  $\xi_k \in \partial_F d_A(x_k) = N_F(A, x_k) \cap D(0, 1)$  for any  $k$ . Suppose next that  $x_n \notin A$  for all  $n \in \mathbb{N}$ . Since

$$\liminf_{x \rightarrow x_n} \frac{d_A(x) - d_A(x_n) - \langle \xi_n, x - x_n \rangle}{\|x - x_n\|} \geq 0 > -\frac{1}{n}, \quad (5.2.25)$$

it follows that there is  $\eta_n \downarrow 0$  such that

$$\langle \xi_n, x - x_n \rangle \leq \frac{1}{n} \|x - x_n\|, \quad \text{for any } x \in D(x_n, \eta_n) \cap A, \quad n \in \mathbb{N}.$$

Choose  $\rho_n \downarrow 0$  such that  $\rho_n < \min \{ \eta_n^2, \frac{1}{n} d_A(x_n) \}$ , and take  $\nu_n \downarrow 1$  such that  $(\nu_n - 1)d_A(x_n) < \rho_n^2$ . We can find then  $\tilde{x}_n \in A$  such that  $\|\tilde{x}_n - x_n\| \leq \nu_n d_A(x_n)$ , and because of (5.2.25) we get that

$$\begin{aligned} \langle \xi_n, u \rangle &\leq d_A(x_n + u) - d_A(x_n) + \frac{1}{n} \|u\| \\ &\leq d_A(x_n + u) - \nu_n^{-1} \|\tilde{x}_n - x_n\| + \frac{1}{n} \|u\| \\ &\leq d_A(\tilde{x}_n + u) + (1 - \nu_n^{-1}) \|\tilde{x}_n - x_n\| + \frac{1}{n} \|u\| \end{aligned}$$

whenever  $\|u\| \leq \eta_n$ . It follows that

$$\langle \xi_n, x - \tilde{x}_n \rangle \leq (1 - \nu_n^{-1}) \|\tilde{x}_n - x_n\| + \frac{1}{n} \|x - \tilde{x}_n\|$$

whenever  $x \in D(\tilde{x}_n, \eta_n) \cap A$ , hence

$$0 \leq \varphi_n(x) := -\langle \xi_n, x - \tilde{x}_n \rangle + \frac{1}{n} \|x - \tilde{x}_n\| + \gamma_n^2, \quad x \in D(\tilde{x}_n, \eta_n) \cap A,$$

where  $\gamma_n^2 := (1 - \nu_n^{-1}) \|\tilde{x}_n - x_n\|$ . This implies that

$$\gamma_n^2 = \varphi_n(\tilde{x}_n) \leq \inf_{x \in D(\tilde{x}_n, \eta_n) \cap A} \varphi_n(x) + \gamma_n^2$$

for any  $n$ . We can apply hence the Ekeland Variational Principle to the continuous function  $\varphi_n$  on the closed set  $D(\tilde{x}_n, \eta_n) \cap A$  (see Remark 3.1.13), in order to get  $\hat{x}_n \in D(\tilde{x}_n, \eta_n) \cap A$  such that  $\|\hat{x}_n - \tilde{x}_n\| \leq \gamma_n$  and

$$-\langle \xi_n, \hat{x}_n - \tilde{x}_n \rangle + \frac{1}{n} \|\hat{x}_n - \tilde{x}_n\| \leq -\langle \xi_n, x - \tilde{x}_n \rangle + \frac{1}{n} \|x - \tilde{x}_n\| + \gamma_n \|x - \hat{x}_n\| \quad (5.2.26)$$

for all  $x \in D(\tilde{x}_n, \eta_n) \cap A$ . By the fact that  $\gamma_n^2 \leq \nu_n(1 - \nu_n^{-1})d_A(x_n) < \rho_n^2$ , defining  $r_n := \rho_n - \gamma_n > 0$ , we get that

$$\|x - \hat{x}_n\| \leq r_n \Rightarrow \|x - \tilde{x}_n\| \leq \|x - \hat{x}_n\| + \gamma_n \leq \rho_n \leq \eta_n.$$

It follows from (5.2.26) that

$$\langle \xi_n, x - \hat{x}_n \rangle \leq \frac{1}{n} (\|x - \tilde{x}_n\| - \|\hat{x}_n - \tilde{x}_n\|) + \gamma_n \|x - \hat{x}_n\| \leq \left( \frac{1}{n} + \gamma_n \right) \|x - \hat{x}_n\|$$

whenever  $x \in D(\widehat{x}_n, r_n) \cap A$ .

Since  $A$  is a closed set, for each  $n$  we form  $\widehat{x}_n + \alpha \xi_n$  with some parameter  $\alpha > 0$ , and select  $w_n \in \text{pr}_A(\widehat{x}_n + \alpha \xi_n)$ . It means that

$$\|\widehat{x}_n + \alpha \xi_n - w_n\|^2 \leq \alpha^2 \|\xi_n\|^2,$$

because  $\widehat{x}_n \in A$ . We have that

$$\|\widehat{x}_n + \alpha \xi_n - w_n\|^2 = \|\widehat{x}_n - w_n\|^2 + 2\alpha \langle \xi_n, \widehat{x}_n - w_n \rangle + \alpha^2 \|\xi_n\|^2,$$

hence we obtain that

$$\|\widehat{x}_n - w_n\|^2 \leq 2\alpha \langle \xi_n, w_n - \widehat{x}_n \rangle \text{ for any } \alpha > 0. \tag{5.2.27}$$

Using the convergence  $w_n \rightarrow \widehat{x}_n$  if  $\alpha \downarrow 0$ , we find from above a sequence  $(\alpha_n) \downarrow 0$  such that

$$\langle \xi_n, w_n - \widehat{x}_n \rangle \leq \left( \frac{1}{\alpha_n} + y_n \right) \|w_n - \widehat{x}_n\|.$$

But this implies that  $\|w_n - \widehat{x}_n\| \leq 2\alpha_n \left( \frac{1}{\alpha_n} + y_n \right)$  due to (5.2.27). Hence,  $w_n \rightarrow \bar{x}$  for  $n \rightarrow \infty$ . Moreover, by defining

$$\chi_n := \xi_n + \frac{1}{\alpha_n}(\widehat{x}_n - w_n),$$

we get that  $\|\chi_n - \xi_n\| \leq 2 \left( \frac{1}{\alpha_n} + y_n \right)$ , hence  $\chi_n \rightarrow \xi$  for  $n \rightarrow \infty$ .

To end the proof of the theorem, it remains to show that  $\chi_n \in N_F(A, w_n)$  for any  $n$ . Indeed, for every fixed  $x \in A$  we have that

$$\begin{aligned} 0 &\leq \|\widehat{x}_n + \alpha_n \xi_n - x\|^2 - \|\widehat{x}_n + \alpha_n \xi_n - w_n\|^2 \\ &= \langle \alpha_n \xi_n + \widehat{x}_n - x, \alpha_n \xi_n + \widehat{x}_n - w_n \rangle + \langle \alpha_n \xi_n + \widehat{x}_n - x, w_n - x \rangle \\ &\quad - \langle \alpha_n \xi_n + \widehat{x}_n - w_n, x - w_n \rangle - \langle \alpha_n \xi_n + \widehat{x}_n - w_n, \alpha_n \xi_n + \widehat{x}_n - x \rangle \\ &= -2\alpha_n \langle \chi_n, x - w_n \rangle + \|x - w_n\|^2. \end{aligned}$$

It means that

$$\langle \chi_n, x - w_n \rangle \leq \frac{1}{2\alpha_n} \|x - w_n\|^2 \text{ for all } x \in A,$$

which shows that  $\chi_n \in N_F(A, w_n)$  for any  $n$ , and this ends the proof of the theorem.  $\square$

### 5.2.3 The Extremal Principle

We dedicate this section to an important concept concerning the local extremality of sets, which will provide us, by different variants of the associated Extremal Principle, powerful tools of understanding and proving further results.

**Definition 5.2.26** (local extremality of sets). *Let  $A_1, \dots, A_k$  be nonempty subsets of  $\mathbb{R}^p$  for  $k \geq 2$ , and let  $\bar{x} \in A_1 \cap \dots \cap A_k$ . We say that  $\bar{x}$  is a local extremal point of the set system  $\{A_1, \dots, A_k\}$  if there are sequences  $(x_{in}) \subset \mathbb{R}^p$ ,  $i = 1, \dots, k$ , and a neighborhood  $U$  of  $\bar{x}$  such that  $x_{in} \rightarrow 0$  for  $n \rightarrow \infty$  and*

$$\bigcap_{i=1}^k (A_i - x_{in}) \cap U = \emptyset \text{ for sufficiently large } n \in \mathbb{N}.$$

*In this case  $\{A_1, \dots, A_k, \bar{x}\}$  is called an extremal system in  $\mathbb{R}^p$ .*

The local extremality of sets at a common point means, intuitively, that these sets can be locally “pushed apart” by a small translation of one of them. If  $n = 2$ , the local extremality of  $\{A_1, A_2, \bar{x}\}$  can be equivalently written as follows: there exists a neighborhood  $U$  of  $\bar{x}$  such that for any  $\varepsilon > 0$ , there is  $x \in D(0, \varepsilon)$  such that  $(A_1 + x) \cap A_2 \cap U = \emptyset$ . For instance, if one takes the sets  $A_1 := \{(x, x) \mid x \in \mathbb{R}\}$  and  $A_2 := \{(x, -x) \mid x \in \mathbb{R}\}$ , then  $\{A_1, A_2, (0, 0)\}$  is not an extremal system in  $\mathbb{R}^2$ . In this example, the condition  $A_1 \cap A_2 = \{\bar{x}\}$  does not mean in general that  $\{A_1, A_2, \bar{x}\}$  is an extremal system. It is the case, though, if  $A_2 = \{\bar{x}\}$  and  $\bar{x} \in \text{bd } A_1$ .

Another important example of extremal system follows. Consider the constrained minimization problem

$$\min f(x) \text{ for } x \in S \subset \mathbb{R}^p,$$

where  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ , and suppose  $\bar{x}$  is a local solution of this problem. If one takes  $A_1 := \text{epi } f$  and  $A_2 := S \times \{f(\bar{x})\}$ , then  $\{A_1, A_2, (\bar{x}, f(\bar{x}))\}$  is an extremal system in  $\mathbb{R}^{p+1}$ . This can be seen from the fact that  $V$ , a neighborhood of  $\bar{x}$ , exists and  $f(x) \geq f(\bar{x})$ , for every  $x \in V \cap S$ , hence if  $(\alpha_n) \subset (-\infty, 0)$  converges to 0,  $\alpha - \alpha_n \geq f(x) - \alpha_n > f(\bar{x})$  for every  $(x, \alpha) \in \text{epi } f \cap [V \cap S \times \mathbb{R}]$ . Consequently, taking  $x_{1n} := (0, \alpha_n)$ ,  $x_{2n} := (0, 0)$  and  $U := V \times \mathbb{R}$ , one has

$$\begin{aligned} (A_1 - x_{1n}) \cap (A_2 - x_{2n}) \cap U &= \left\{ (x, \alpha - \alpha_n) \mid \begin{array}{l} (x, \alpha) \in \text{epi } f \cap [V \cap S \times \mathbb{R}] \\ \alpha - \alpha_n = f(\bar{x}) \end{array} \right\} \\ &= \emptyset. \end{aligned}$$

We proceed by introducing two versions of the extremal principle, which can be seen, as we will justify in the next, as local extensions of the separation theorems for nonconvex sets around the extremal points (see Theorems 4.1.3, 4.1.4). We say that a set  $A \subset \mathbb{R}^p$  is closed around  $\bar{x} \in A$  if there is  $\varepsilon > 0$  such that  $A \cap D(\bar{x}, \varepsilon)$  is closed. Notice that every closed set, as well as every open set, is locally closed.

**Definition 5.2.27.** *Let  $\{A_1, \dots, A_k, \bar{x}\}$  be an extremal system in  $\mathbb{R}^p$ .*

*(i)  $\{A_1, \dots, A_k, \bar{x}\}$  satisfies the approximate extremal principle if for every  $\varepsilon > 0$ , there are  $x_i \in A_i \cap D(\bar{x}, \varepsilon)$  and  $\xi_i \in N_F(A_i, x_i) + D(0, \varepsilon)$ ,  $i = 1, \dots, k$  such that*

$$\xi_1 + \dots + \xi_k = 0, \quad \|\xi_1\| + \dots + \|\xi_k\| = 1. \tag{5.2.28}$$



(ii)  $\{A_1, \dots, A_k, \bar{x}\}$  satisfies the exact extremal principle if there are  $\xi_i \in N_M(A_i, \bar{x})$ ,  $i = 1, \dots, k$  such that (5.2.28) holds.

We say that the corresponding version of the extremal principle holds in  $\mathbb{R}^p$  if it holds for every extremal system  $\{A_1, \dots, A_k, \bar{x}\}$ , where all the sets  $A_i$  are closed around  $\bar{x}$ .

If  $\{A_1, \dots, A_k, \bar{x}\}$  satisfies the exact extremal principle, then it also satisfies the approximate extremal principle. This follows immediately: from the definition of the Mordukhovich normal cones, since  $\xi_i \in N_M(A_i, \bar{x})$ , there are the sequences  $(x_{in}) \xrightarrow{A_i} \bar{x}$  and  $(\xi_{in}) \rightarrow \xi_i$  such that  $\xi_{in} \in N_F(A_i, x_i), \forall i = 1, \dots, k$ . Then for every  $\varepsilon > 0$ , one can choose  $n$  sufficiently large such that  $x_{in} \in A_i \cap D(\bar{x}, \varepsilon)$ , and also  $\xi_i = \xi_{in} + (\xi_i - \xi_{in}) \in N_F(A_i, x_i) + D(0, \varepsilon)$ , for every  $i = 1, \dots, k$ .

Observe also that in case  $n = 2$ , and  $A_1$  and  $A_2$  are convex sets, due to the form of the normal cones in the convex case, if the exact extremal principle holds, then it reduces to the existence of

$$\xi \in N_M(A_1, \bar{x}) \cap (-N_M(A_2, \bar{x})) \setminus \{0\},$$

which yields

$$\langle \xi, x_1 \rangle \leq \langle \xi, x_2 \rangle \text{ for every } x_1 \in A_1, \text{ and } x_2 \in A_2,$$

i.e., exactly the separation property of  $A_1$  and  $A_2$ .

The main result of this section, proven by Mordukhovich by the method of metric approximations, is the following:

**Theorem 5.2.28** (exact extremal principle). *The exact extremal principle holds in  $\mathbb{R}^p$ .*

*Proof* Consider the extremal system  $\{A_1, \dots, A_k, \bar{x}\}$ , where we suppose without loss of generality that all the sets  $A_i$  are closed. Also, without loss of generality, suppose that the neighborhood  $U$  of  $\bar{x}$  from Definition 5.2.27 is  $\mathbb{R}^p$ , and consider the sequences  $(x_{in}) \subset \mathbb{R}^p, i = 1, \dots, k$  such that  $x_{in} \rightarrow 0$  and

$$\bigcap_{i=1}^k (A_i - x_{in}) = \emptyset \text{ for all } n \in \mathbb{N}. \tag{5.2.29}$$

Moreover, for every  $n$ , consider the minimization problem

$$\min d_n(x) := \left[ \sum_{i=1}^k d_{A_i}^2(x + x_{in}) \right]^{1/2} + \|x - \bar{x}\|^2, \quad x \in \mathbb{R}^p.$$

By using Theorem 3.1.7, the above problem admits a solution  $x_n$ . Denote

$$\alpha_n := \left[ \sum_{i=1}^k d_{A_i}^2(x_n + x_{in}) \right]^{1/2}.$$

If  $\alpha_n$  would be 0, this would mean, due to the closedness of the sets  $A_i$ , that  $x_n \in \bigcap_{i=1}^k (A_i - x_{in})$ , which obviously contradicts (5.2.29). Hence,

$$\begin{aligned} 0 < d_n(x_n) = \alpha_n + \|x_n - \bar{x}\|^2 &\leq \left[ \sum_{i=1}^k d_{A_i}^2(\bar{x} + x_{in}) \right]^{1/2} \\ &\leq \left[ \sum_{i=1}^k x_{in}^2 \right]^{1/2} \downarrow 0, \end{aligned}$$

which shows that  $x_n \rightarrow \bar{x}$  and  $\alpha_n \downarrow 0$ .

Now, since the sets  $A_i$  are closed, consider the elements  $u_{in}$  from the nonempty projection sets  $\text{pr}_{A_i}(x_n + x_{in}) : u_{in} \in A_i$  and  $d_{A_i}(x_n + x_{in}) = \|x_n + x_{in} - u_{in}\|$ . Also, for all  $n$ , consider the minimization problem

$$\min \rho_n(x) := \left[ \sum_{i=1}^k \|x + x_{in} - u_{in}\|^2 \right]^{1/2} + \|x - \bar{x}\|^2, \quad x \in \mathbb{R}^p. \quad (5.2.30)$$

Because

$$\rho_n(x) \geq d_n(x) \geq d_n(x_n) = \rho_n(x_n),$$

it follows that the problem (5.2.30) has the optimal solution  $x_n$ . Moreover, since  $\alpha_n > 0$ , the functions  $\rho_n$  are  $C^1$  around  $x_n$ . Due to the classical Fermat rule, this means that

$$\nabla \rho_n(x_n) = \sum_{i=1}^k \xi_{in} + 2(x_n - \bar{x}) = 0, \quad (5.2.31)$$

where  $\xi_{in} := \frac{1}{\alpha_n}(x_n + x_{in} - u_{in})$ ,  $i = 1, \dots, k$ . Remark that

$$\|\xi_{1n}\|^2 + \dots + \|\xi_{kn}\|^2 = 1. \quad (5.2.32)$$

It follows that all  $\xi_{in}$  belong to the unit ball of  $\mathbb{R}^p$ , which is a compact set. We may suppose then, without loss of generality, that there are elements  $\xi_i$  in  $D(0, 1)$  such that  $\xi_{in} \rightarrow \xi_i$  for every  $i = 1, \dots, k$ . Passing to the limit in (5.2.31), we get that  $\xi_1 + \dots + \xi_k = 0$ . Moreover, due to the fact that  $u_{in} \in \text{pr}_{A_i}(x_n + x_{in})$ ,  $\frac{1}{\alpha_n} \subset (0, \infty)$ ,  $x_n + x_{in} \rightarrow \bar{x}$  and  $\xi_{in} \rightarrow \xi_i$ , it means, based on formula (5.2.7), that  $\xi_i \in N_M(A_i, \bar{x})$  for all  $i = 1, \dots, k$ . Passing to the limit in (5.2.32), it means that not all  $\xi_i$  are 0, which gives us

$$\|\xi_1\| + \dots + \|\xi_k\| = 1,$$

eventually by replacing  $\xi_i$  with  $(\|\xi_1\| + \dots + \|\xi_k\|)^{-1} \cdot \xi_i$ . □

As consequence, we get the next result. We say that a set is proper if it is nonempty and different from the whole space.

**Corollary 5.2.29.** *For every proper closed set  $A \subset \mathbb{R}^p$  and  $\bar{x} \in A$ , one has*

$$N_M(A, \bar{x}) \neq \{0\} \iff \bar{x} \in \text{bd } A.$$

*Proof* It follows immediately that if  $\bar{x} \notin \text{bd } A$ , then  $\bar{x} \in \text{int } A$ , hence  $N_M(A, \bar{x}) = \{0\}$ . If  $\bar{x} \in \text{bd } A$ , then  $\{A, \{\bar{x}\}, \bar{x}\}$  is an extremal system in  $\mathbb{R}^p$ , hence using Theorem 5.2.28, one has that there are  $\xi_1 \in N_M(A, \bar{x})$  and  $\xi_2 \in N_M(\{\bar{x}\}, \bar{x}) = \mathbb{R}^p$ , not equal both to 0, such that  $\xi_1 = -\xi_2$ . It follows that  $0 \neq \xi_1 \in N_M(A, \bar{x})$ .  $\square$

### 5.2.4 Calculus Rules

One of the main advantages of the generalized differentiation objects (i.e., normal cones, subdifferentials) presented in this chapter is their rich associated calculus. As we presented in the introduction of this chapter, since it would be outside the aim of this book, we do not present generalized differentiation objects related to set-valued mappings (derivatives and coderivatives), which are closely interrelated in the calculus rules mentioned before. In what follows, we will concentrate on rules involving subdifferentials of sums and differences of functions. Moreover, such sum rules have the advantage to fully employ some important results previously given, in such a way that one can draw an intuitive picture on how the whole machinery of calculus associated to Mordukhovich differentiation objects works. For many other results involving calculus rules for Mordukhovich generalized differentiation objects, the reader could consult (Mordukhovich, 2006).

We begin our exposition with a sum rule where one of the functions is differentiable.

**Proposition 5.2.30.** *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ .*

(i) *For any  $g : \mathbb{R}^p \rightarrow \mathbb{R}$  Fréchet differentiable at  $\bar{x}$ , one has*

$$\partial_F(f + g)(\bar{x}) = \partial_F f(\bar{x}) + \nabla g(\bar{x}). \tag{5.2.33}$$

(ii) *If  $f$  is lower semicontinuous around  $\bar{x}$ , for any  $g : \mathbb{R}^p \rightarrow \mathbb{R}$  continuously differentiable around  $\bar{x}$ , one has*

$$\partial_M(f + g)(\bar{x}) = \partial_M f(\bar{x}) + \nabla g(\bar{x}). \tag{5.2.34}$$

(iii) *For any  $g : \mathbb{R}^p \rightarrow \mathbb{R}$  Lipschitz continuous around  $\bar{x}$ , one has*

$$\partial^\infty(f + g)(\bar{x}) = \partial^\infty f(\bar{x}). \tag{5.2.35}$$

*Proof* (i) Let us prove the “ $\subset$ ” inclusion in (5.2.33). Suppose  $\xi \in \partial_F(f + g)(\bar{x})$ . Fix arbitrary  $\varepsilon > 0$ . Using the analytical characterization of the Fréchet subgradients (i.e., Theorem 5.2.11), there exists  $\delta > 0$  such that, for every  $x \in D(\bar{x}, \delta)$ ,

$$f(x) + g(x) - f(\bar{x}) - g(\bar{x}) - \langle \xi, x - \bar{x} \rangle \geq -\frac{\varepsilon}{2} \|x - \bar{x}\|.$$

Also, from the Fréchet differentiability of  $g$  at  $\bar{x}$ , there exists  $y \in (0, \delta)$  such that, for every  $x \in D(x, y)$ ,

$$-g(x) + g(\bar{x}) + \langle \nabla g(\bar{x}), x - \bar{x} \rangle \geq -\frac{\varepsilon}{2} \|x - \bar{x}\|.$$

Adding the previous relations, one gets that  $\xi - \nabla g(\bar{x}) \in \partial_F f(\bar{x})$ .

For the reverse inclusion, take  $\xi \in \partial_F f(\bar{x})$ . Then  $\xi \in \partial_F ((f + g) - g)(\bar{x}) \subset \partial_F (f + g)(\bar{x}) - \nabla g(\bar{x})$  from the already proven inclusion. It means that  $\xi + \nabla g(\bar{x}) \in \partial_F (f + g)(\bar{x})$ .

(ii) Consider now  $\xi \in \partial_M (f + g)(\bar{x})$ , i.e.,  $(\xi, -1) \in N_M(\text{epi}(f + g), (\bar{x}, (f + g)(\bar{x})))$ . Since  $f + g$  is a lower semicontinuous function, the set  $\text{epi } f$  is closed, hence using Proposition 5.2.8 we know that  $(\xi, -1) \in N_M(\text{gr}(f + g), (\bar{x}, (f + g)(\bar{x})))$ . This means, based on the definition of basic subgradients, that there exist  $(x_n) \rightarrow \bar{x}$ ,  $(\xi_n) \rightarrow \xi$  and  $(\lambda_n) \rightarrow 1$  such that  $(\xi_n, -\lambda_n) \in N_F(\text{gr}(f + g), (x_n, (f + g)(x_n)))$  for every  $n$ . This means that for every  $\varepsilon > 0$ , there exists  $\delta > 0$  such that for every  $(x, (f + g)(x)) \in D((x_n, (f + g)(x_n)), \delta)$ , one has

$$\begin{aligned} \langle \xi_n, x - x_n \rangle - \lambda_n (f(x) + g(x) - f(x_n) - g(x_n)) &\leq \varepsilon (\|x - x_n\| \\ &+ |f(x) - f(x_n)| + |g(x) - g(x_n)|). \end{aligned} \tag{5.2.36}$$

On the other hand, for every fixed such  $x$  and  $x_n$ , due to the fact that  $g$  is  $C^1$ , there exists  $y_n$  on the segment which joins  $x$  and  $x_n$  such that

$$g(x) - g(x_n) = \langle \nabla g(y_n), x - x_n \rangle.$$

Combining relation (5.2.36) with the Lipschitz property of  $g$  (with modulus  $\ell \geq 0$ ), one gets that

$$\langle \xi_n - \lambda_n \nabla g(y_n), x - x_n \rangle - \lambda_n (f(x) - f(x_n)) \leq \varepsilon (1 + \ell) (\|x - x_n\| + |f(x) - f(x_n)|). \tag{5.2.37}$$

Since if  $x$  is close to  $x_n$ , and  $(f + g)(x)$  is close to  $(f + g)(x_n)$ , one gets from the Lipschitz property of  $g$  that  $f(x)$  must be close to  $f(x_n)$ . The previous relation shows that  $(\xi_n - \lambda_n \nabla g(y_n), -\lambda_n) \in N_F(\text{gr } f, (x_n, f(x_n)))$ . Set  $\eta_n := \xi_n - \lambda_n \nabla g(y_n)$  and suppose, by contradiction, that  $(\eta_n, -\lambda_n) \notin N_F(\text{epi } f, (x_n, f(x_n)))$ . Then there exist  $y \in (0, \lambda_n)$  and sequences  $(u_j, \alpha_j) \xrightarrow{\text{epi } f} (x_n, f(x_n))$  such that

$$\langle \eta_n, u_j - x_n \rangle - \lambda_n (\alpha_j - f(x_n)) > y \|(u_j, \alpha_j) - (x_n, f(x_n))\|.$$

On the other hand, since  $\alpha_j \geq f(u_j)$ , one has

$$\|(u_j, f(u_j)) - (x_n, f(x_n))\| \leq \|(u_j, \alpha_j) - (x_n, f(x_n))\| + (\alpha_j - f(u_j)),$$

hence

$$\begin{aligned} \langle \eta_n, u_j - x_n \rangle - \lambda_n (\alpha_j - f(x_n)) &> y \|(u_j, f(u_j)) - (x_n, f(x_n))\| - y (\alpha_j - f(u_j)) \\ &> y \|(u_j, f(u_j)) - (x_n, f(x_n))\| - \lambda_n (\alpha_j - f(u_j)). \end{aligned}$$

It means that

$$\langle \eta_n, u_j - x_n \rangle - \lambda_n (f(u_j) - f(x_n)) > y \| (u_j, f(u_j)) - (x_n, f(x_n)) \|,$$

which contradicts the fact that  $(\eta_n, -\lambda_n) \in N_F(\text{gr } f, (x_n, f(x_n)))$ . It follows that  $(\xi_n - \lambda_n \nabla g(y_n), -\lambda_n) \in N_F(\text{epi } f, (x_n, f(x_n)))$  for every  $n$ . Since  $\xi_n - \lambda_n \nabla g(y_n) \rightarrow \xi - \nabla g(\bar{x})$  due to the fact that  $g$  is  $C^1$ , and  $(x_n, f(x_n)) \xrightarrow{\text{epi } f} (\bar{x}, f(\bar{x}))$  due to the fact that  $x_n \rightarrow \bar{x}$ ,  $g(x_n) \rightarrow g(\bar{x})$  and  $(f + g)(x_n) \rightarrow (f + g)(\bar{x})$ , it follows that  $\xi - \nabla g(\bar{x}) \in \partial_M f(\bar{x})$ . We proved the “ $\subset$ ” inclusion in (5.2.34). The other inclusion can be deduced as in (i).

(iii) Let us prove now the “ $\subset$ ” inclusion in (5.2.35). Take  $\xi \in \partial^\infty(f + g)(\bar{x})$ . There exist the sequences  $(x_n, \alpha_n) \xrightarrow{\text{epi}(f+g)} (\bar{x}, (f + g)(\bar{x}))$ ,  $(\xi_n) \rightarrow \xi$  and  $(\lambda_n) \rightarrow 0$  such that, for every  $\varepsilon > 0$ , there exists  $\delta > 0$  such that

$$\langle \xi_n, x - x_n \rangle - \lambda_n (\alpha - \alpha_n) \leq \varepsilon (\|x - x_n\| + |\alpha - \alpha_n|)$$

for every  $(x, \alpha) \in \text{epi}(f + g)$  such that  $x \in D(x_n, \delta)$  and  $\alpha$  such that  $|\alpha - \alpha_n| < \delta$ . Suppose  $\ell$  is the Lipschitz modulus of  $g$  around  $\bar{x}$ , and take  $\beta_n := \alpha_n - g(x_n)$ . We have that  $(x_n, \beta_n) \xrightarrow{\text{epi } f} (\bar{x}, f(\bar{x}))$  and if  $(x, \alpha) \in \text{epi } f$ ,  $x \in D\left(x_n, \frac{\delta}{2(\ell+1)}\right)$ , and  $|\alpha - \beta_n| \leq \frac{\delta}{2(\ell+1)}$ , one has  $(x, g(x) + \alpha) \in \text{epi}(f + g)$ , and

$$\begin{aligned} |(g(x) + \alpha) - \alpha_n| &\leq |g(x) - g(x_n)| + |\alpha - (\alpha_n - g(x_n))| \\ &\leq \ell \|x - x_n\| + |\alpha - \beta_n| \leq \frac{\delta}{2} < \delta, \end{aligned}$$

hence

$$\begin{aligned} \langle \xi_n, x - x_n \rangle - \lambda_n (\alpha - \beta_n) &= \langle \xi_n, x - x_n \rangle - \lambda_n ((\alpha + g(x)) - \alpha_n) - \lambda_n (g(x_n) - g(x)) \\ &\leq \varepsilon (\|x - x_n\| + |(\alpha + g(x)) - \alpha_n|) + \lambda_n |g(x_n) - g(x)| \\ &\leq \varepsilon (\|x - x_n\| + |(\alpha + g(x_n)) - \alpha_n|) + (\varepsilon + \lambda_n)\ell \|x - x_n\| \\ &\leq (\varepsilon + \varepsilon\ell + \lambda_n\ell) (\|x - x_n\| + |\alpha - \beta_n|), \end{aligned}$$

for every  $(x, \alpha) \in \text{epi } f$ ,  $x \in D\left(x_n, \frac{\delta}{2(\ell+1)}\right)$ , and  $|\alpha - \beta_n| \leq \frac{\delta}{2(\ell+1)}$ . Since  $\varepsilon' := \varepsilon + \varepsilon\ell + \lambda_n\ell$  can be taken positive and arbitrarily small due to the fact that  $\lambda_n \rightarrow 0$  and  $\varepsilon > 0$  is arbitrary, it follows from above that  $(\xi_n, -\lambda_n) \in N_F(\text{epi } f, (x_n, \beta_n))$  for every  $n$ , hence  $(\xi, 0) \in N_M(\text{epi } f, (\bar{x}, f(\bar{x})))$ . The other inclusion can be proved similarly to the final part of (i). □

**Lemma 5.2.31.** Consider  $f_1, f_2 : \mathbb{R}^p \rightarrow \mathbb{R}$  such that  $f_1$  is Lipschitz continuous around  $\bar{x}$ , and  $f_2$  is lower semicontinuous around this point. If  $f_1 + f_2$  attains a local minimum at  $\bar{x}$ , then for every  $\varepsilon > 0$ , there are  $x_i \in D(\bar{x}, \varepsilon)$  with  $|f_i(x_i) - f_i(\bar{x})| \leq \varepsilon$ ,  $i = 1, 2$ , such that

$$0 \in \partial_F f_1(x_1) + \partial_F f_2(x_2) + D(0, \varepsilon). \tag{5.2.38}$$

*Proof* Fix  $\varepsilon > 0$ . Without loss of generality, suppose  $\bar{x} = 0$ ,  $f_1(\bar{x}) = f_2(\bar{x}) = 0$ ,  $f_1$  is Lipschitz on  $D(0, \varepsilon)$  with modulus  $\ell > 0$ , and  $f_2$  is lower semicontinuous on  $D(0, \varepsilon)$ . Then the sets

$$A_1 := \text{epi } f_1 \text{ and } A_2 := \{(x, \alpha) \mid f_2(x) \leq -\alpha\}$$

are closed around  $(0, 0)$ . Moreover, since  $f_1 + f_2$  attains a local minimum at  $0$ , and  $(f_1 + f_2)(0) = 0$ , it follows that  $(0, 0)$  is an extremal point for  $\{A_1, A_2\}$ . Using Theorem 5.2.28, it follows that the exact extremal principle, hence also the approximate extremal principle hold for the system  $\{A_1, A_2, (0, 0)\}$ .

Then for every  $y > 0$ , there are  $(x_i, \alpha_i) \in A_i$ , and  $(\xi_i, \lambda_i) \in \mathbb{R}^{p+1}$ ,  $i = 1, 2$ , such that

$$(\xi_1, -\lambda_1) \in N_F(A_1, (x_1, \alpha_1)), \quad (-\xi_2, \lambda_2) \in N_F(A_2, (x_2, \alpha_2)), \quad (5.2.39)$$

$$\|(x_i, \alpha_i)\| < y, \quad \frac{1}{2} - y < \|(\xi_i, \lambda_i)\| < \frac{1}{2} + y, \quad i = 1, 2 \quad (5.2.40)$$

$$\|(\xi_1, -\lambda_1) + (-\xi_2, \lambda_2)\| \leq y. \quad (5.2.41)$$

Due to the form of the sets  $A_1, A_2$  and (5.2.39), it follows that  $\lambda_1, \lambda_2 \geq 0$ . Choose  $y \in \left(0, \min \left\{ \frac{1}{4(2+\ell)}, \frac{\varepsilon}{4(1+\ell)^2} \right\}\right)$  such that relations (5.2.39)–(5.2.41) hold. Using the fact that  $f_1$  is Lipschitz continuous on  $D(0, \varepsilon)$ , one gets (see the proof of Theorem 5.2.18) from  $(\xi_1, -\lambda_1) \in N_F(A_1, (x_1, \alpha_1))$  with  $\|(x_1, \alpha_1)\| < y < \varepsilon$  that  $\|\xi_1\| \leq \ell \lambda_1$ . Hence,

$$\|(\xi_1, \lambda_1)\| \leq (\ell + 1)\lambda_1.$$

One can now employ (5.2.40) and (5.2.41) to get

$$\lambda_1 \geq \frac{1}{\ell + 1} \|(\xi_1, \lambda_1)\| > \frac{1}{2(\ell + 1)} - \frac{y}{\ell + 1} > 0$$

and

$$\lambda_1 - \lambda_2 \leq |\lambda_1 - \lambda_2| \leq \|(\xi_1, -\lambda_1) + (-\xi_2, \lambda_2)\| \leq y,$$

hence

$$\lambda_2 \geq \lambda_1 - y > \frac{1}{2(\ell + 1)} - \frac{y}{\ell + 1} - y > \frac{1}{4(\ell + 1)} > 0$$

due to the choice of  $y$ . Since  $\alpha_1 \geq f_1(x_1)$ , suppose  $\alpha_1 > f_1(x_1)$ . From  $\left(\frac{\xi_1}{\lambda_1}, -1\right) \in N_F(A_1, (x_1, \alpha_1))$ , one gets that for every  $\delta \in (0, 1)$ , there exists  $\tau > 0$  such that for every  $(u, \alpha) \in D((x_1, \alpha_1), \tau) \cap \text{epi } f_1$ , one has

$$\left\langle \frac{\xi_1}{\lambda_1}, u - x_1 \right\rangle - (\alpha - \alpha_1) < \delta \|(u - x_1, \alpha - \alpha_1)\|. \quad (5.2.42)$$

Choosing  $u := x_1$ , one can find from  $\alpha_1 > f_1(x_1)$  that there is  $\alpha \in (f(x_1), \alpha_1)$  arbitrarily close to  $\alpha_1$ . For such  $(u, \alpha)$ , it follows from (5.2.42) that  $\alpha_1 - \alpha < \delta(\alpha_1 - \alpha)$ ,

hence  $\delta > 1$ , a contradiction. Then,  $\alpha_1 = f_1(x_1)$ . Similarly, one can prove that  $\alpha_2 = -f_2(x_2)$ . Hence,

$$\eta_1 := \frac{\xi_1}{\lambda_1} \in \partial_F f_1(x_1) \text{ and } \eta_2 := -\frac{\xi_2}{\lambda_2} \in \partial_F f_2(x_2).$$

By (5.2.40), we have that

$$\|x_i\| \leq y < \varepsilon \text{ and } |f_i(x_i) - f_i(\bar{x})| = |\alpha_i| \leq y < \varepsilon, \quad i = 1, 2.$$

Moreover,

$$\begin{aligned} \|\eta_1 + \eta_2\| &= \left\| \frac{\xi_1(\lambda_1 - \lambda_2)}{\lambda_1 \lambda_2} - \frac{\xi_1 - \xi_2}{\lambda_2} \right\| \leq \frac{\|\xi_1\|}{\lambda_1} \left( \frac{|\lambda_1 - \lambda_2|}{\lambda_2} \right) + \frac{\|\xi_1 - \xi_2\|}{\lambda_2} \\ &\leq \ell \frac{y}{\lambda_2} + \frac{y}{\lambda_2} = \frac{y}{\lambda_2} (\ell + 1) \leq 4y(\ell + 1)^2 < \varepsilon. \end{aligned}$$

We have hence (5.2.38). □

**Theorem 5.2.32** (sum rule). *Consider  $f_1, f_2 : \mathbb{R}^p \rightarrow \mathbb{R}$  such that  $f_1$  is Lipschitz continuous around  $\bar{x}$ , and  $f_2$  is lower semicontinuous around this point. Then:*

(i) *for every  $\varepsilon > 0$ , one has*

$$\begin{aligned} \partial_F(f_1 + f_2)(\bar{x}) \subset \bigcup \{ \partial_F f_1(x_1) + \partial_F f_2(x_2) \mid x_i \in D(\bar{x}, \varepsilon), \\ |f_i(x_i) - f_i(\bar{x})| \leq \varepsilon, \quad i = 1, 2 \} + D(0, \varepsilon). \end{aligned} \tag{5.2.43}$$

(ii) *one has*

$$\partial_M(f_1 + f_2)(\bar{x}) \subset \partial_M f_1(\bar{x}) + \partial_M f_2(\bar{x}). \tag{5.2.44}$$

*Proof* (i) Fix  $\varepsilon > 0$  and take  $\xi \in \partial_F(f_1 + f_2)(\bar{x})$ . Then one can find  $\eta$  such that

$$0 < \eta < \min \left\{ \frac{\varepsilon}{4}, \frac{\varepsilon}{\|\xi\| + \varepsilon + 2} \right\}.$$

By using the analytical characterization of the Fréchet subgradients given by (5.2.9), one deduces that the function

$$(f_1(x) - \langle \xi, x - \bar{x} \rangle + \varepsilon \|x - \bar{x}\|) + f_2(x)$$

attains a local minimum at  $\bar{x}$ . One may apply Lemma 5.2.31 for the chosen  $\eta$  to find  $x_i \in D(\bar{x}, \eta)$  and  $\xi_i \in \mathbb{R}^p, i = 1, 2$ , such that

$$\begin{aligned} |f_1(x_1) - \langle \xi, x_1 - \bar{x} \rangle + \varepsilon \|x_1 - \bar{x}\| - f_1(\bar{x})| &\leq \eta, \quad |f_2(x_2) - f_2(\bar{x})| \leq \eta \\ \xi_1 \in \partial_F(f_1 - \langle \xi, \cdot \rangle + \eta \|\cdot - \bar{x}\|)(x_1), \quad \xi_2 \in \partial_F f_2(x_2), \\ -\xi_1 - \xi_2 &\in D(0, \eta). \end{aligned}$$

By (5.2.33), it follows that  $\xi'_1 := \xi + \xi_1 \in \partial_F(f_1 + \eta \|\cdot - \bar{x}\|)(x_1)$ , and hence  $\xi - \xi'_1 - \xi_2 \in D(0, \eta)$ . Moreover,

$$|f_1(x_1) - f_1(\bar{x})| \leq \eta + (\|\xi\| + \varepsilon) \|x_1 - \bar{x}\| < \eta (\|\xi\| + \varepsilon + 1).$$

Now, due to the analytical characterization of the Fréchet subgradients given by (5.2.9), we have from  $\xi'_1 \in \partial_F(f_1 + \eta \|\cdot - \bar{x}\|)(x_1)$  that the function  $f_1 + \varphi$  with

$$\varphi(x) := \eta \|x - \bar{x}\| - \langle \xi'_1, x - x_1 \rangle + \eta \|x - x_1\|$$

attains a local minimum at  $x_1$ . Remark that, because  $\varphi$  is the sum of three convex (hence, continuous) functions on  $\mathbb{R}^p$ , one has by Proposition 5.2.14 that  $\partial_M \varphi(x) = \partial \varphi(x) \subset -\xi'_1 + D(0, 2\eta)$  for any  $x \in \mathbb{R}^p$ . Using again Lemma 5.2.31 for the chosen  $\eta$  and for the sum  $f_1 + \varphi$ , we find  $x'_1 \in D(x_1, \eta)$  such that

$$|f_1(x'_1) - f_1(x_1)| \leq \eta \text{ and } \xi'_1 \in \partial_F f_1(x'_1) + D(0, 3\eta).$$

Concluding, we found

$$\xi \in \partial_F f_1(x'_1) + \partial_F f_2(x_2) + D(0, 4\eta),$$

with  $\|x'_1 - \bar{x}\| \leq 2\eta < \varepsilon$ ,  $\|x_2 - \bar{x}\| \leq \eta < \varepsilon$ ,  $|f_1(x'_1) - f_1(\bar{x})| \leq \eta (\|\xi\| + \varepsilon + 2) < \varepsilon$ ,  $|f_2(x_2) - f_2(\bar{x})| \leq \eta < \varepsilon$ . Hence, (5.2.43) is completely proved.

(ii) Take arbitrary  $\xi \in \partial_M(f_1 + f_2)(\bar{x})$ . Since  $f_1 + f_2$  is lower semicontinuous, it follows by (5.2.20) that there are sequences  $(x_n) \rightarrow \bar{x}$  and  $(\xi_n) \rightarrow \xi$  such that  $f_1(x_n) + f_2(x_n) \rightarrow f_1(\bar{x}) + f_2(\bar{x})$  and  $\xi_n \in \partial_F(f_1 + f_2)(x_n)$  for every  $n$ . Using now (5.2.43) for  $\varepsilon := \frac{1}{n}$ , we can find sequences  $(x_{in}) \rightarrow \bar{x}$  and  $(\xi_{in}) \in \mathbb{R}^p$  with  $f_i(x_{in}) \rightarrow f_i(\bar{x})$ ,  $\xi_{in} \in \partial_F f_i(x_{in})$ ,  $i = 1, 2$ , and

$$\|\xi_n - \xi_{1n} - \xi_{2n}\| \leq \frac{1}{n}, \quad \forall n \geq 1. \tag{5.2.45}$$

Because  $f_1$  is Lipschitz continuous around  $\bar{x}$ , we know that the sequence  $(\xi_{1n})$  is bounded by the Lipschitz modulus of  $f_1$  due to (5.2.19). Also,  $(\xi_n)$  is bounded because it is convergent. It follows from (5.2.45) that  $(\xi_{2n})$  is also bounded. Without loss of generality, we may suppose hence that there are  $\xi_1, \xi_2 \in \mathbb{R}^p$  such that  $\xi_{in} \rightarrow \xi_i$ ,  $i = 1, 2$ . From (5.2.45), we know that  $\xi = \xi_1 + \xi_2$ . Moreover, due to (5.2.20), we have that  $\xi_i \in \partial_M f_i(x_i)$ , which ends the proof of (5.2.44).  $\square$

We continue our exposition dedicated to calculus rules by providing a difference rule for the Fréchet subgradients. As mentioned before, its proof uses the smooth variational description of the Fréchet subgradients, given in Theorem 5.2.13.

**Theorem 5.2.33.** *Let  $f, g : \mathbb{R}^p \rightarrow \mathbb{R}$  be two functions, and  $\bar{x} \in \mathbb{R}^p$ . Then*

$$\partial_F(f - g)(\bar{x}) \subset \bigcap_{\xi \in \partial_F g(\bar{x})} [\partial_F f(\bar{x}) - \xi] \subset \partial_F f(\bar{x}) - \partial_F g(\bar{x}), \tag{5.2.46}$$

*provided  $\partial_F g(\bar{x}) \neq \emptyset$ .*



*Proof* Take  $\chi \in \partial_F(f - g)(\bar{x})$  arbitrarily and  $\xi \in \partial_F g(\bar{x})$ . Using Theorem 5.2.13 for  $\xi \in \partial_F g(\bar{x})$ , one gets the existence of  $s : \mathbb{R}^p \rightarrow \mathbb{R}$ , Fréchet differentiable at  $\bar{x}$ , which satisfies:

$$s(\bar{x}) = g(\bar{x}), \quad s(x) \leq g(x) \quad \forall x \in \mathbb{R}^p, \quad \text{and} \quad \nabla s(\bar{x}) = \xi. \tag{5.2.47}$$

For any  $\varepsilon > 0$ , by applying the definition of the Fréchet subgradients for  $\chi \in \partial_F(f - g)(\bar{x})$ , there is  $\delta > 0$  such that

$$\begin{aligned} \langle \chi, x - \bar{x} \rangle &\leq f(x) - g(x) - (f(\bar{x}) - g(\bar{x})) + \varepsilon \|x - \bar{x}\| \\ &\leq f(x) - s(x) - (f(\bar{x}) - s(\bar{x})) + \varepsilon \|x - \bar{x}\| \end{aligned}$$

whenever  $\|x - \bar{x}\| < \delta$ . Using now (5.2.47) and (5.2.33), one gets that

$$\chi \in \partial_F(f - s)(\bar{x}) = \partial_F f(\bar{x}) - \nabla s(\bar{x}) = \partial_F f(\bar{x}) - \xi,$$

which ends the proof. □

By Proposition 5.2.30, (5.2.46) becomes equality if one of the functions  $f, g$  is Fréchet differentiable at  $\bar{x}$ . There exist as well other cases when the equality holds: take  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  given as  $f(x) = \sqrt{|x|}$  and  $g(x) = |x|$ . Then  $\partial_F f(0) = \mathbb{R}$ ,  $\partial_F g(0) = [-1, 1]$ , and  $\partial_F(f - g)(0) = \mathbb{R}$ , hence equality holds in (5.2.46). Also, observe that the assumption that  $\partial_F g(\bar{x}) \neq \emptyset$  is essential for the fulfillment of the result: take, for instance,  $f, g : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = |x|$  and  $g(x) = -|x|$ . Then  $\partial_F(f - g)(0) = [-2, 2]$ , but  $\partial_F g(\bar{x}) = \emptyset$ .

As corollaries of the Theorem 5.2.33, we obtain some other interesting facts. The first one concerns the so-called DC-functions, i.e., differences of convex functions. We denote, as always, by  $\partial f(\bar{x})$  the convex subdifferential.

**Corollary 5.2.34.** *Let  $f, g : \mathbb{R}^p \rightarrow \mathbb{R}$  be convex functions, and  $\bar{x} \in \mathbb{R}^p$ . Then*

$$\partial_F(f - g)(\bar{x}) = \bigcap_{\xi \in \partial g(\bar{x})} [\partial f(\bar{x}) - \xi]$$

*Proof* It follows from Theorem 5.2.33 and Proposition 5.2.14. □

The second corollary gives us a calculus rule for the Mordukhovich subdifferential of the difference of two functions.

**Corollary 5.2.35.** *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be lower semicontinuous around  $\bar{x} \in \mathbb{R}^p$ , and  $g : \mathbb{R}^p \rightarrow \mathbb{R}$  be locally Lipschitz around  $\bar{x}$ , and such that  $\partial_M g(x)$  is nonempty for every  $x$  close to  $\bar{x}$ . Then*

$$\partial_M(f - g)(\bar{x}) \subset \partial_M f(\bar{x}) - \partial_M g(\bar{x}).$$

*Proof* Take  $\xi \in \partial_M(f - g)(\bar{x})$  arbitrarily. By using Theorem 5.2.21, there exist sequences  $(x_n) \xrightarrow{f-g} \bar{x}$  and  $(\xi_n) \rightarrow \xi$  such that  $\xi_n \in \partial_F(f - g)(x_n)$  for every  $n$ . Applying Theorem 5.2.33 for the previous relation, one gets that there are sequences  $(\chi_n), (\eta_n)$  such that  $\chi_n \in \partial_F f(x_n)$  and  $\eta_n \in \partial_F g(x_n)$  for every  $n$  and  $\xi_n = \chi_n - \eta_n$ , for every  $n$ . Since  $g$  is locally Lipschitz around  $\bar{x}$ , we know by (5.2.19) that the sequence  $(\eta_n)$  is bounded, so we may suppose, without loss of generality, that there is  $\eta \in \mathbb{R}^p$  such that  $\eta_n \rightarrow \eta$ . The Lipschitz property of  $g$  implies  $g(x_n) \rightarrow g(\bar{x})$ , and by  $f(x_n) - g(x_n) \rightarrow f(\bar{x}) - g(\bar{x})$  one gets that  $f(x_n) \rightarrow f(\bar{x})$ . Moreover, we have that  $\eta \in \partial_M g(\bar{x})$ . Furthermore, since  $\xi_n = \chi_n - \eta_n$ , for every  $n$ ,  $\xi_n \rightarrow \xi$ , and  $\eta_n \rightarrow \eta$ , we get that  $\chi_n \rightarrow \xi + \eta$ , hence  $\xi + \eta \in \partial_M f(\bar{x})$ . But this means that  $\xi \in \partial_M f(\bar{x}) - \partial_M g(\bar{x})$ , and the proof is finished.  $\square$

Observe that all the assumptions on  $g$  from the previous corollary are automatically satisfied if  $g$  is convex.

Recall that for a function  $f : \mathbb{R}^p \rightarrow \mathbb{R}^k$  and a vector  $\xi \in \mathbb{R}^k$ , the function  $\langle \xi, f \rangle : \mathbb{R}^p \rightarrow \mathbb{R}$  is given by

$$\langle \xi, f \rangle (x) = \langle \xi, f(x) \rangle.$$

We can present a chain rule for the Fréchet subdifferential.

**Theorem 5.2.36** (chain rule). *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}^k$  be Lipschitz around  $\bar{x} \in \mathbb{R}^p$ ,  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  be a function, and denote  $\bar{y} := f(\bar{x})$ . Then*

$$\partial_F(g \circ f)(\bar{x}) \subset \bigcap_{-\xi \in \partial_F(-g)(\bar{y})} \partial_F \langle \xi, f \rangle (\bar{x}), \tag{5.2.48}$$

*provided  $\partial_F(-g)(\bar{y}) \neq \emptyset$ . Moreover, (5.2.48) holds as equality if  $g$  is Fréchet differentiable at  $\bar{y}$ .*

*Proof* Take  $\chi \in \partial_F(g \circ f)(\bar{x})$  and  $-\xi \in \partial_F(-g)(\bar{y})$ . Using Theorem 5.2.13, one gets the existence of  $s : \mathbb{R}^p \rightarrow \mathbb{R}$ , Fréchet differentiable at  $\bar{y}$ , which satisfies:

$$s(\bar{y}) = g(\bar{y}), \quad g(y) \leq s(y) \quad \forall y \in \mathbb{R}^k, \quad \text{and} \quad \nabla s(\bar{y}) = \xi. \tag{5.2.49}$$

Now, from the definition of the Fréchet subgradients, applied for  $\chi \in \partial_F(g \circ f)(\bar{x})$ , and then from (5.2.49), one gets that for any  $\varepsilon > 0$ , there is  $\delta > 0$  such that

$$\begin{aligned} \langle \chi, x - \bar{x} \rangle &\leq g(f(x)) - g(f(\bar{x})) + \varepsilon \|x - \bar{x}\| \\ &\leq s(f(x)) - s(f(\bar{x})) + \varepsilon \|x - \bar{x}\| \\ &= \langle \xi, f(x) - f(\bar{x}) \rangle + \alpha(\|x - \bar{x}\| + \|f(x) - f(\bar{x})\|) + \varepsilon \|x - \bar{x}\|, \end{aligned}$$

whenever  $\|x - \bar{x}\| < \delta$ , where  $\alpha$  is a function such that  $\lim_{x \rightarrow \bar{x}} \frac{\alpha(\|x - \bar{x}\| + \|f(x) - f(\bar{x})\|)}{\|x - \bar{x}\| + \|f(x) - f(\bar{x})\|} = 0$ . Hence, for every  $\varepsilon > 0$ ,

$$\langle \chi, x - \bar{x} \rangle - \langle \xi, f(x) - f(\bar{x}) \rangle \leq 2\varepsilon (\|x - \bar{x}\| + \|f(x) - f(\bar{x})\|),$$

whenever  $\|x - \bar{x}\| < \delta$ , which implies that

$$\langle \xi, f \rangle(x) - \langle \xi, f \rangle(\bar{x}) - \langle \chi, x - \bar{x} \rangle \geq -2\varepsilon(\ell + 1) \|x - \bar{x}\|$$

whenever  $\|x - \bar{x}\| < \delta$ , since  $f$  is Lipschitz around  $\bar{x}$  (with modulus  $\ell \geq 0$ ). This implies that

$$\liminf_{x \rightarrow \bar{x}} \frac{\langle \xi, f \rangle(x) - \langle \xi, f \rangle(\bar{x}) - \langle \chi, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq -2\varepsilon(\ell + 1).$$

Since  $\varepsilon > 0$  was arbitrarily chosen, we have that  $\chi \in \partial_F \langle \xi, f \rangle(\bar{x})$ , and (5.2.48) is proved.

Suppose now that  $g$  is Fréchet differentiable. Then  $\partial_F(-g)(\bar{y}) = -\nabla g(\bar{y}) =: -\xi$ , so the intersection from the right-hand side of (5.2.48) reduces to an element. We have then from (5.2.48)

$$\partial_F(g \circ f)(\bar{x}) \subset \partial_F \langle \xi, f \rangle(\bar{x}).$$

Suppose  $\chi \notin \partial_F(g \circ f)(\bar{x})$ , and prove that  $\chi \notin \partial_F \langle \xi, f \rangle(\bar{x})$ . Observe, moreover, that  $\eta \in \partial_F \langle \xi, f \rangle(\bar{x})$  is equivalent to  $(\eta, -1) \in N_F(\text{epi} \langle \xi, f \rangle, (\bar{x}, \langle \xi, f \rangle(\bar{x})))$ . By Proposition 5.2.8, it is sufficient to prove that  $(\chi, -1) \notin N_F(\text{gr} \langle \xi, f \rangle, (\bar{x}, \langle \xi, f \rangle(\bar{x})))$ . But since if  $x \rightarrow \bar{x}$ , due to the continuity of the function  $x \mapsto \langle \xi, f \rangle(x)$ , this is equivalent to  $x \xrightarrow{\text{gr} \langle \xi, f \rangle} \bar{x}$ , it means that we must prove that

$$\limsup_{x \rightarrow \bar{x}} \frac{\langle \chi, x - \bar{x} \rangle - \langle \xi, f(x) - f(\bar{x}) \rangle}{\|x - \bar{x}\| + |\langle \xi, f(x) - f(\bar{x}) \rangle|} > 0.$$

But then, since

$$\|x - \bar{x}\| + |\langle \xi, f(x) - f(\bar{x}) \rangle| \leq (1 + \|\xi\| \ell) \|x - \bar{x}\|,$$

it is sufficient to prove that

$$\limsup_{x \rightarrow \bar{x}} \frac{\langle \chi, x - \bar{x} \rangle - \langle \xi, f(x) - f(\bar{x}) \rangle}{\|x - \bar{x}\|} > 0. \tag{5.2.50}$$

But from  $\chi \notin \partial_F(g \circ f)(\bar{x})$ , we know that

$$\liminf_{x \rightarrow \bar{x}} \frac{\langle \xi, f \rangle(x) - \langle \xi, f \rangle(\bar{x}) - \langle \chi, x - \bar{x} \rangle}{\|x - \bar{x}\|} < 0,$$

which means exactly (5.2.50). The theorem is completely proved. □

The next result is an approximate mean value theorem for lower semicontinuous functions.

**Theorem 5.2.37.** *Let  $\varphi : \mathbb{R}^p \rightarrow \mathbb{R}$  be a lower semicontinuous function and  $a \neq b$ . Consider any point  $c \in [a, b]$  at which the function*

$$\psi(x) := \varphi(x) - \frac{\varphi(b) - \varphi(a)}{\|b - a\|} \|x - a\|$$

attains its minimum on  $[a, b]$ ; such a point always exists. Then there are sequences  $(x_n) \xrightarrow{\varphi} c$  and  $\xi_n \in \partial_F \varphi(x_n)$  such that

$$\liminf_{n \rightarrow \infty} \langle \xi_n, b - x_n \rangle \geq \frac{\varphi(b) - \varphi(a)}{\|b - a\|} \|c - a\|, \quad (5.2.51)$$

$$\liminf_{n \rightarrow \infty} \langle \xi_n, b - a \rangle \geq \varphi(b) - \varphi(a). \quad (5.2.52)$$

Moreover, when  $c \neq a$  one has

$$\lim_{n \rightarrow \infty} \langle \xi_n, b - a \rangle = \varphi(b) - \varphi(a).$$

*Proof* The function  $\psi$  is lower semicontinuous, and therefore attains its minimum on  $[a, b]$  at some point  $c$ . Because  $\psi(a) = \psi(b)$ , we may suppose  $c \in [a, b]$ . Without loss of generality, suppose  $\varphi(a) = \varphi(b)$ , which means that  $\psi(x) = \varphi(x)$  on  $[a, b]$ . Because  $\varphi$  is lower semicontinuous, there exists  $r > 0$  and  $y \in \mathbb{R}$  such that  $\varphi(x) \geq y$  for every  $x \in P := [a, b] + D(0, r)$ . Then for every  $n \in \mathbb{N}$ , one can find  $r_n \in (0, r)$  such that

$$\varphi(x) \geq \varphi(c) - \frac{1}{n^2}, \quad \text{for all } x \in [a, b] + D(0, r_n).$$

Moreover, choose  $t_n \geq n$  such that  $y + t_n r_n \geq \varphi(c) - \frac{1}{n^2}$ . We have then

$$\varphi(c) \leq \inf_{x \in P} \varphi_n(x) + \frac{1}{n^2},$$

where  $\varphi_n(x) := \varphi(x) + t_n d_{[a,b]}(x)$  is lower semicontinuous. By the use of Ekeland Variational Principle on the closed set  $P$  (see Remark 3.1.13), we get  $x_n \in P$  such that

$$\|x_n - c\| \leq \frac{1}{n}, \quad \varphi_n(x_n) \leq \varphi(c), \quad \text{and}$$

$$\varphi_n(x_n) \leq \varphi_n(x) + \frac{1}{n} \|x - x_n\| \quad \text{for any } x \in P.$$

The last relation shows that the function  $\varphi_n(x) + \frac{1}{n} \|x - x_n\|$  attains its minimum on  $P$  at  $x = x_n$ . Taking into account that  $x_n \in \text{int } P$  for large  $n$ , we can conclude that this function attains a local minimum (without restrictions) at  $x_n$ . Using Lemma 5.2.31 for  $\varepsilon = \varepsilon_n \downarrow 0$ , we find sequences  $(u_n) \xrightarrow{\varphi} c$ ,  $(v_n) \rightarrow c$ ,  $\xi_n \in \partial_F \varphi(u_n)$ ,  $\eta_n \in \partial_F d_{[a,b]}(v_n)$ , and  $e_n \in D(0, 1)$  such that

$$\left\| \xi_n + t_n \eta_n + n^{-1} e_n \right\| \leq \varepsilon_n, \quad n \in \mathbb{N}. \quad (5.2.53)$$

Since  $\eta_n \in \partial_F d_{[a,b]}(v_n) = \partial d_{[a,b]}(v_n)$ , it follows that  $\|\eta_n\| \leq 1$ , and

$$\langle \eta_n, b - v_n \rangle \leq d_{[a,b]}(b) - d_{[a,b]}(v_n) \leq 0, \quad n \in \mathbb{N}.$$

Picking  $w_n \in \text{pr}_{[a,b]}(v_n)$ , we get

$$\langle \eta_n, b - w_n \rangle = \langle \eta_n, b - v_n \rangle + \langle \eta_n, v_n - w_n \rangle \leq d_{[a,b]}(b) - d_{[a,b]}(v_n) + \|\eta_n\| \cdot \|v_n - w_n\|$$

$$\leq -d_{[a,b]}(v_n) + \|v_n - w_n\| = 0.$$

It means that  $\langle \eta_n, b - c \rangle \leq 0$  for large  $n$ , since  $w_n \rightarrow c \neq b$ , and hence, since  $a, b, c$  are colinear,

$$\langle \eta_n, b - a \rangle = \langle \eta_n, b - c \rangle \cdot \frac{\|a - b\|}{\|c - b\|} \leq 0 \text{ for large } n.$$

By using now (5.2.53), we get that there exists  $b_n \in D(0, \varepsilon_n)$  such that

$$\xi_n = -t_n \eta_n - n^{-1} e_n - \varepsilon_n b_n,$$

hence,

$$\liminf_{n \rightarrow \infty} \langle \xi_n, b - u_n \rangle \geq 0, \quad \liminf_{n \rightarrow \infty} \langle \xi_n, b - a \rangle \geq 0,$$

which prove (5.2.51) and (5.2.52). Suppose that  $c \neq a$ . Then  $v_n \neq a$  for large  $n$ , hence  $\langle \eta_n, b - c \rangle = 0$ . This implies that

$$\langle \xi_n, b - a \rangle = \langle \xi_n, b - c \rangle \cdot \frac{\|a - b\|}{\|c - b\|} = \langle -t_n \eta_n - n^{-1} e_n - \varepsilon_n b_n, b - c \rangle \cdot \frac{\|a - b\|}{\|c - b\|} \rightarrow 0,$$

where  $b_n \in D(0, 1)$  exists from (5.2.53). This concludes the proof. □

We end this subsection with a result which links the Clarke and Mordukhovich constructions presented throughout this chapter.

**Theorem 5.2.38.** *The following assertions hold:*

(i) *Let  $A \in \mathbb{R}^p$  be a closed set and  $\bar{x} \in A$ . Then*

$$N_C(A, \bar{x}) = \text{cl conv } N_M(A, \bar{x}). \tag{5.2.54}$$

(ii) *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be locally Lipschitz around  $\bar{x}$ . Then*

$$\partial_C f(x) = \text{cl conv } \partial_M f(\bar{x}). \tag{5.2.55}$$

*Proof* We prove first that

$$\partial_M f(\bar{x}) \subset \partial_C f(\bar{x}), \tag{5.2.56}$$

which will show that  $\text{cl conv } \partial_M f(\bar{x}) \subset \partial_C f(\bar{x})$ , due to the fact that  $\partial_C f(\bar{x})$  is a closed convex set. Take  $\chi \in \partial_M f(\bar{x})$ . Then there are  $(x_n) \xrightarrow{f} \bar{x}$  and  $(\chi_n) \rightarrow \chi$  such that  $\chi_n \in \partial_F f(x_n)$  for every  $n$ . This means, due to the definition of the Fréchet subgradients, that for every  $n$  there is a neighborhood  $U_n$  of  $x_n$  such that

$$f(x) - f(x_n) - \langle \chi_n, x - x_n \rangle \geq -\frac{1}{n} \|x - x_n\|, \quad \forall x \in U_n.$$

But this shows that the function

$$\psi_n(x) := f(x + x_n) - \langle \chi_n, x \rangle + \frac{1}{n} \|x\|$$

attains a local minimum at 0. This proves, by Theorem 5.1.16, that

$$0 \in \partial_C \left[ f(\cdot + x_n) - \langle \chi_n, \cdot \rangle + \frac{1}{n} \|\cdot\| \right] (0).$$

Since all the involved functions are locally Lipschitz, it follows from the sum rule for the Clarke subdifferential that

$$\chi_n \in \partial_C f(x_n) + D \left( 0, \frac{1}{n} \right).$$

Using Proposition 5.1.6, we get that  $\chi \in \partial_C f(\bar{x})$ . Hence, the desired inclusion holds. For proving the reverse inclusion in (5.2.55), we will show the representation

$$\begin{aligned} f^\circ(\bar{x}, u) &= \sup \{ \langle \xi, u \rangle \mid \xi \in \partial_M f(\bar{x}) \} \\ &= \max \{ \langle \xi, u \rangle \mid \xi \in \text{cl } \partial_M f(\bar{x}) \}. \end{aligned} \tag{5.2.57}$$

This will prove that  $f^\circ(\bar{x}, u)$  is the support function of the set  $P := \partial_M f(\bar{x})$ . Observe, by definition, that the support function of a set coincides with the support function of the closed convex hull of this set, hence

$$f^\circ(\bar{x}, u) = \sup \{ \langle \xi, u \rangle \mid \xi \in \text{cl conv } \partial_M f(\bar{x}) \}.$$

Moreover, we know that two convex and closed sets coincide if and only if their support functions coincide. Since  $\partial_C f(x)$  and  $\text{cl conv } \partial_M f(\bar{x})$  are both closed and convex, we get (5.2.55).

Take some realizing sequences for  $f^\circ(\bar{x}, u)$ . This means that there are  $(t_n) \downarrow 0$  and  $(x_n) \rightarrow \bar{x}$  such that

$$\frac{f(x_n + t_n u) - f(x_n)}{t_n} \rightarrow f^\circ(\bar{x}, u) \text{ for } n \rightarrow \infty.$$

By applying Theorem 5.2.37 to  $f$  on the interval  $[x_n, x_n + t_n u]$  for every  $n$ , we get  $(v_k) \rightarrow c_n \in [x_n, x_n + t_n u]$  for  $k \rightarrow \infty$  and  $\xi_k \in \partial_F f(v_k)$  such that

$$f(x_n + t_n u) - f(x_n) \leq t_n \liminf_{k \rightarrow \infty} \langle \xi_k, u \rangle, \quad \forall n \in \mathbb{N}.$$

Since  $f$  is locally Lipschitz around  $\bar{x}$  and  $\xi_k \in \partial_F f(v_k)$ , we know that  $(\xi_k)$  is bounded, hence without loss of generality we may suppose that  $(\xi_k)$  is convergent to some  $\xi$ . Passing to the limit above first for  $k \rightarrow \infty$ , and then for  $n \rightarrow \infty$ , we get that

$$f^\circ(\bar{x}, u) \leq \langle \xi, u \rangle,$$

for some  $\xi \in \partial_M f(\bar{x})$ . Since  $\langle \xi', u \rangle \leq f^\circ(\bar{x}, u)$  for any  $\xi' \in \partial_C f(\bar{x}) \supset \partial_M f(\bar{x})$ , we get the representation (5.2.57). Hence, the proof of (5.2.55) is complete.

Let us prove now (5.2.54). Since due to (5.2.24) and (5.2.56) we have that

$$N_M(A, \bar{x}) = \bigcup_{\lambda > 0} \lambda \partial_M d_A(\bar{x}) \subset \bigcup_{\lambda > 0} \lambda \partial_C d_A(\bar{x}) = N_C(A, \bar{x}),$$

we also have  $\text{cl conv } N_M(A, \bar{x}) = N_C(A, \bar{x})$ , since  $N_C(A, \bar{x})$  is closed and convex. For the reverse inclusion, we use (5.2.55) to get that

$$\bigcup_{\lambda > 0} \lambda \partial_C d_A(\bar{x}) = \bigcup_{\lambda > 0} \lambda [\text{cl conv } \partial_M d_A(\bar{x})] \subset \text{cl conv } \bigcup_{\lambda > 0} \lambda \partial_M d_A(\bar{x}).$$

The theorem is now completely proved. □

### 5.2.5 Optimality Conditions

In this section, we make use of the exact version of the extremal principle in order to get necessary optimality conditions for the minimization problem considered in the previous section:

$$(MP) \quad \min f(x), \text{ subject to } g(x) \leq 0, \ h(x) = 0, \ x \in A, \tag{MP}$$

where the functions  $f, g = (g_1, \dots, g_n)$  and  $h = (h_1, \dots, h_m)$  are locally Lipschitz functions which map from  $\mathbb{R}^p$  into  $\mathbb{R}, \mathbb{R}^n$  and  $\mathbb{R}^m$ , respectively, and the set  $A \subset \mathbb{R}^p$  is closed. The final aim is to get Fritz John necessary optimality conditions for problem (MP).

**Theorem 5.2.39.** *Let  $\bar{x}$  be a solution of (MP), where  $f, g$  and  $h$  are Lipschitz around  $\bar{x}$ . Then there exist  $\lambda_0 \geq 0, \lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$  and  $\mu = (\mu_1, \dots, \mu_m) \in \mathbb{R}^m$ , with  $\lambda_0 + \|\lambda\| + \|\mu\| \neq 0$ , such that*

$$0 \in \lambda_0 \partial_M f(\bar{x}) + \sum_{i=1}^n \lambda_i \partial_M g_i(\bar{x}) + \sum_{j=1}^m \mu_j [\partial_M h_j(\bar{x}) \cup \partial_M (-h_j)(\bar{x})] + N_M(A, \bar{x}) \tag{5.2.58}$$

and

$$\begin{aligned} \lambda_i &\geq 0, \ \lambda_i g_i(\bar{x}) = 0, \ \forall i \in \overline{1, n}, \\ \mu_j &\geq 0, \ \forall j \in \overline{1, m}. \end{aligned} \tag{5.2.59}$$

*Proof* Suppose, without loss of generality, that  $f(\bar{x}) = 0$ . Then it is easy to observe that the point  $(\bar{x}, 0)$  is a local extremal point of the the following system of closed sets in  $\mathbb{R}^{p+n+m+1}$ :

$$\begin{aligned} A_0 &:= \{(x, \lambda_0, \lambda_1, \dots, \lambda_n, \mu_1, \dots, \mu_m) \mid \lambda_0 \geq f(x)\}, \\ A_i &:= \{(x, \lambda_0, \lambda_1, \dots, \lambda_n, \mu_1, \dots, \mu_m) \mid \lambda_i \geq g_i(x)\}, \ i \in \overline{1, n}, \\ A_{n+j} &:= \{(x, \lambda_0, \lambda_1, \dots, \lambda_n, \mu_1, \dots, \mu_m) \mid \mu_j = h_j(x)\}, \ j \in \overline{1, m}, \\ A_{n+m+1} &:= A \times \{0\}. \end{aligned}$$

Since by Theorem 5.2.28 the exact extremal principle holds in  $\mathbb{R}^{p+n+m+1}$ , we find elements

$$(\xi_0, -\lambda_0) \in N_M(\text{epi } f, (\bar{x}, 0)),$$

$$\begin{aligned} (\xi_i, -\lambda_i) &\in N_M(\text{epi } g_i, (\bar{x}, 0)), \quad i \in \overline{1, n} \\ (\xi_{n+j}, -\lambda_{n+j}) &\in N_M(\text{gr } h_j, (\bar{x}, 0)), \quad j \in \overline{1, m} \\ \widehat{\xi} &\in N_M(A, \bar{x}) \end{aligned}$$

such that

$$\xi_0 + \dots + \xi_{n+m} + \widehat{\xi} = 0, \tag{5.2.60}$$

$$\|(\xi_0, -\lambda_0)\| + \dots + \|(\xi_{n+j}, -\lambda_{n+j})\| + \|\widehat{\xi}\| = 1. \tag{5.2.61}$$

Using Proposition 5.2.15 on basic normals to epigraphs, we deduce that  $\lambda_i \geq 0$  for  $i \in \overline{0, n}$ .

If  $g_i(\bar{x}) < 0$  for some fixed  $i \in \overline{1, n}$ , then  $g_i(x) < 0$  for all  $x$  around  $\bar{x}$  due to the Lipschitz property of  $g_i$ . But this implies that  $(\bar{x}, 0)$  is an interior point of  $\text{epi } g_i$ , which means by  $N_M(\text{epi } g_i, (\bar{x}, 0)) = \{0\}$  that  $\xi_i = 0$  and  $\lambda_i = 0$ .

Now, by Proposition 5.2.15 and Theorem 5.2.18, we know that

$$(\xi, -\lambda) \in N_M(\text{epi } \varphi, (\bar{x}, \varphi(\bar{x}))) \iff \xi \in \lambda \partial_M \varphi(\bar{x}), \quad \lambda \geq 0$$

if  $\varphi$  is Lipschitz around  $\bar{x}$ . Also, by Proposition 5.2.8, for such function  $\varphi$ , one has that

$$(\xi, -\lambda) \in N_M(\text{gr } \varphi, (\bar{x}, \varphi(\bar{x}))) \iff \xi \in \partial_M(\lambda\varphi)(\bar{x}).$$

Taking into account also that

$$\partial_M(\lambda\varphi)(\bar{x}) \subset |\lambda| [\partial_M \varphi(\bar{x}) \cup \partial_M(-\varphi)(\bar{x})] \text{ for all } \lambda \in \mathbb{R},$$

we have, by denoting  $\mu_j := |\lambda_{n+j}|$  for  $j \in \overline{1, m}$  and (5.2.60), that (5.2.58) holds, and also  $\mu_j \geq 0$  for any  $j \in \overline{1, m}$ . This completes the proof of the theorem.  $\square$

**Remark 5.2.40.** Observe that the set  $\partial_M \varphi(\bar{x}) \cup \partial_M(-\varphi)(\bar{x})$  reduces to  $\{\nabla \varphi(\bar{x}), -\nabla \varphi(\bar{x})\}$  when  $\varphi$  is  $C^1$ , so Theorem 5.2.39 is indeed a generalization to the nonsmooth case of the known results.

Moreover, remark that, since  $\partial_C \varphi(\bar{x})$  is always larger than  $\partial_M \varphi(\bar{x})$ , the previous theorem may provide more precise necessary optimality conditions than Theorem 5.1.34 in case that the equality constraints are not present. To observe this, consider for instance the minimization problem

$$\min f(x) := -|x|, \quad x \in \mathbb{R}.$$

Then  $\bar{x} = 0$  is not a minimum point of this problem, while  $0 \in \partial_C f(\bar{x}) = [-1, 1]$ . On the other hand,  $0 \notin \partial_M f(\bar{x}) = \{-1, 1\}$ .

Another example is given by the minimization problem

$$\min x_1 \text{ subject to } \varphi(x_1, x_2) := |x_1| - |x_2| \leq 0.$$



Then  $\partial_M \varphi(0, 0) = \{(y_1, y_2) \mid y_1 \in [-1, 1], y_2 = \pm 1\}$ . It follows by Theorem 5.2.39 that the point  $(0, 0)$  cannot be optimal for the considered problem. Since

$$\partial_C \varphi(0, 0) = \text{cl conv } \partial_M \varphi(0, 0) = [-1, 1] \times [-1, 1],$$

it means that we cannot decide that  $(0, 0)$  is not an optimal candidate by Theorem 5.1.34.