

## 4 Convex Nonsmooth Optimization

In this chapter we study optimization problems involving convex functions that are not necessarily differentiable. Naturally, several new tools are needed in order to compensate for the lack of differentiability. On one hand, we need to study convex sets which were briefly defined and studied in Section 2.1. A new object to replace the differential is introduced and studied. With all these tools in hand, we will be able to derive a generalized Karush-Kuhn-Tucker theorem in the case of convex nonsmooth optimization.

### 4.1 Further Properties and Separation of Convex Sets

We start with some results concerning the fundamental topological properties of convex sets in  $\mathbb{R}^p$ .

**Theorem 4.1.1.** *Let  $C \subset \mathbb{R}^p$  be a convex set. Then*

- (i)  $\text{cl } C$  is convex;
- (ii) if  $x \in \text{int } C$  and  $y \in \text{cl } C$ , then  $[x, y) \subset \text{int } C$ ;
- (iii)  $\text{int } C$  is convex;
- (iv) if  $\text{int } C \neq \emptyset$ , then  $\text{cl } C = \text{cl}(\text{int } C)$  and  $\text{int } C = \text{int}(\text{cl } C)$ .

*Proof* (i) Let us take  $x, y \in \text{cl } C$  and  $\alpha \in (0, 1)$ , and a neighborhood  $V$  of  $0 \in \mathbb{R}^p$ . It is well known that there is a neighborhood  $U$  of  $0$  such that  $\alpha U + (1 - \alpha)U \subset V$ . Since  $x, y \in \text{cl } C$ , there are  $x_U, y_U \in C$  such that  $x_U \in (x + U) \cap C$  and  $y_U \in (y + U) \cap C$ . Consequently, from the convexity of  $C$ , one has

$$\begin{aligned} C &\ni \alpha x_U + (1 - \alpha)y_U \in \alpha(x + U) + (1 - \alpha)(y + U) \\ &= \alpha x + (1 - \alpha)y + \alpha U + (1 - \alpha)U \subset \alpha x + (1 - \alpha)y + V, \end{aligned}$$

whence  $C \cap (\alpha x + (1 - \alpha)y + V) \neq \emptyset$ . Since  $V$  is an arbitrary neighborhood of  $0$ , this shows that  $\alpha x + (1 - \alpha)y \in \text{cl } C$ . Of course, an argument based on a characterization of  $\text{cl } C$  that used sequences is also possible.

(ii) Take  $\alpha \in (0, 1)$ . It is enough to show that  $\alpha x + (1 - \alpha)y \in \text{int } C$ . Since  $x \in \text{int } C$ , there is a neighborhood  $V$  of  $0 \in \mathbb{R}^p$  with  $x + V + V \subset C$ . On the other hand,  $y \in \text{cl } C$  implies that  $C \cap (y - \alpha(1 - \alpha)^{-1}V) \neq \emptyset$ , whence  $y$  can be written as  $c + \alpha(1 - \alpha)^{-1}v$ , with  $c \in C$  and  $v \in V$ . We get that

$$\begin{aligned} \alpha x + (1 - \alpha)y + \alpha V &= \alpha x + (1 - \alpha)c + \alpha v + \alpha V = (1 - \alpha)c + \alpha(x + v + V) \\ &\subset (1 - \alpha)c + \alpha C \subset C. \end{aligned}$$

Since  $\alpha V$  is still a neighborhood of  $0$ , we conclude that  $\alpha x + (1 - \alpha)y \in \text{int } C$ .

(iii) If  $x, y \in \text{int } C$ , then the above implication means that  $[x, y] \subset \text{int } C$ , whence  $\text{int } C$  is convex.

(iv) Clearly,  $\text{cl}(\text{int } C) \subset \text{cl } C$ . Consider  $x \in \text{cl } C$ . By (ii), for any  $y \in \text{int } C$ ,  $(x, y] \subset \text{int } C$ , which means that  $x$  can be approached by a sequence of points in  $\text{int } C$ , that is  $x \in \text{cl}(\text{int } C)$ . For the second part, one inclusion always holds:  $\text{int } C \subset \text{int}(\text{cl } C)$ . Consider  $x \in \text{int}(\text{cl } C)$ , which means that there exists a neighborhood  $V$  of  $0 \in \mathbb{R}^p$  such that  $x + V \subset \text{cl } C$ , whence, once again from (ii), for any  $y \in \text{int } C$ ,  $\alpha \in (0, 1)$  and  $v \in V$ ,

$$\alpha(x + v) + (1 - \alpha)y \in \text{int } C.$$

But  $V$  is absorbing (that is, for every  $x \in \mathbb{R}^p$ , there is  $\lambda > 0$  such that  $\lambda x \in V$ ), so for  $\alpha$  sufficiently close to 1,

$$\bar{v} := \frac{(1 - \alpha)(x - y)}{\alpha} \in V$$

and for such an  $\alpha$ ,

$$\begin{aligned} x &= \alpha \left( x + \frac{(1 - \alpha)(x - y)}{\alpha} \right) + (1 - \alpha)y \\ &= \alpha(x + \bar{v}) + (1 - \alpha)y \in \text{int } C. \end{aligned}$$

The proof is complete. □

Now we need a supplementary investigation into the projection of a point on a closed convex set (see Theorem 2.1.5 (iii)).

**Proposition 4.1.2.** *Let  $C \subset \mathbb{R}^p$  be a nonempty closed and convex set. Then the application  $\mathbb{R}^p \ni x \mapsto \text{pr}_C x$  is 1-Lipschitz. In particular, it is continuous.*

*Proof* As proved in Theorem 2.1.5 (iii), the projection  $\text{pr}_C x$  of a point  $x$  on  $C$  is characterized by the properties

$$\begin{cases} \text{pr}_C x \in C \\ \langle x - \text{pr}_C x, u - \text{pr}_C x \rangle \leq 0, \forall u \in C. \end{cases}$$

Take  $x_1, x_2 \in \mathbb{R}^p$ . Then the above system gives

$$\begin{aligned} \langle x_1 - \text{pr}_C x_1, \text{pr}_C x_2 - \text{pr}_C x_1 \rangle &\leq 0 \\ \langle x_2 - \text{pr}_C x_2, \text{pr}_C x_1 - \text{pr}_C x_2 \rangle &\leq 0. \end{aligned}$$

This means that  $\langle x_1 - \text{pr}_C x_1 - x_2 + \text{pr}_C x_2, \text{pr}_C x_2 - \text{pr}_C x_1 \rangle \leq 0$ , whence

$$\langle x_1 - x_2 + \text{pr}_C x_2 - \text{pr}_C x_1, \text{pr}_C x_2 - \text{pr}_C x_1 \rangle \leq 0.$$

We get that

$$\|\text{pr}_C x_2 - \text{pr}_C x_1\|^2 \leq \langle x_2 - x_1, \text{pr}_C x_2 - \text{pr}_C x_1 \rangle \leq \|x_2 - x_1\| \cdot \|\text{pr}_C x_2 - \text{pr}_C x_1\|,$$

so the desired inequality holds. Since every Lipschitz function is continuous, so the second part follows as well.  $\square$

Now we are able to prove a separation result for convex sets.

**Theorem 4.1.3.** *Let  $C \subset \mathbb{R}^p$  be a nonempty convex set and  $\bar{x} \notin C$ . Then there is an  $a \in \mathbb{R}^p \setminus \{0\}$  such that for all  $c \in C$  one has*

$$\langle a, \bar{x} \rangle \leq \langle a, c \rangle.$$

*Proof* First, suppose that  $\bar{x} \notin \text{cl } C$ . In fact, in this case we can prove a stronger inequality. Notice that Theorem 4.1.1 (i) ensures that  $\text{cl } C$  is a closed convex set, so there exists  $\text{pr}_{\text{cl } C} \bar{x} \in \text{cl } C$ , which is not equal to  $\bar{x}$ . The properties of projection ensure that

$$\langle \text{pr}_{\text{cl } C} \bar{x} - \bar{x}, x - \text{pr}_{\text{cl } C} \bar{x} \rangle \geq 0, \quad \forall x \in \text{cl } C.$$

Define  $a := \text{pr}_{\text{cl } C} \bar{x} - \bar{x} \in \mathbb{R}^p \setminus \{0\}$  and rewrite the above inequality as

$$\langle a, x - \bar{x} - a \rangle \geq 0, \quad \forall x \in \text{cl } C.$$

This means that

$$\langle a, c - \bar{x} \rangle \geq \|a\|^2 > 0, \quad \forall c \in C,$$

and the desired inequality follows.

Take now that case when  $\bar{x} \in \text{cl } C \setminus C$ . Observe (by the use of Theorem 4.1.1 (iv)) that

$$\text{cl}(\mathbb{R}^p \setminus \text{cl } C) = \mathbb{R}^p \setminus \text{int } \text{cl } C = \mathbb{R}^p \setminus \text{int } C \supset \text{cl } C \setminus C,$$

so  $\bar{x} \in \text{cl}(\mathbb{R}^p \setminus \text{cl } C)$ , which means that there exists a sequence  $(x_k)_{k \in \mathbb{N}}$  of points outside  $\text{cl } C$  with  $x_k \rightarrow \bar{x}$  as  $k \rightarrow \infty$ . Then, for all  $k \in \mathbb{N}$ ,  $\text{pr}_{\text{cl } C} x_k \neq x_k$  and we can define

$$a_k := \frac{\text{pr}_{\text{cl } C} x_k - x_k}{\|\text{pr}_{\text{cl } C} x_k - x_k\|}$$

with the property that for every  $k \in \mathbb{N}$  and  $x \in \text{cl } C$ ,

$$\langle a_k, x - \text{pr}_{\text{cl } C} x_k \rangle \geq 0.$$

The sequence  $(a_k)$  is bounded, whence it has a convergent subsequence. As usual, without relabeling, we assume that  $(a_k)$  converges to a limit  $a \in \mathbb{R}^p$ . Since  $\|a_k\| = 1$  for all  $k$ , we deduce that  $a \neq 0$ . By means of Proposition 4.1.2,

$$\text{pr}_{\text{cl } C} x_k \rightarrow \text{pr}_{\text{cl } C} \bar{x} = \bar{x},$$

and passing to the limit in the previous inequality, one gets

$$\langle a, c - \bar{x} \rangle \geq 0, \quad \forall c \in C.$$

The result is proved.  $\square$

**Theorem 4.1.4.** *Let  $A$  and  $B$  be two nonempty convex sets in  $\mathbb{R}^p$ . If  $A \cap B = \emptyset$ , then there exists  $a \in \mathbb{R}^p \setminus \{0\}$  such that*

$$\langle a, x \rangle \leq \langle a, y \rangle, \quad \forall x \in A, \forall y \in B.$$

*Proof* Consider the convex set

$$C := A - B.$$

Since  $A \cap B = \emptyset$ , we deduce that  $\bar{x} := 0 \notin C$ , and from Theorem 4.1.3 we get the existence of an element  $a \in \mathbb{R}^p \setminus \{0\}$  such that for all  $c \in C$  one has

$$\langle a, \bar{x} \rangle \leq \langle a, c \rangle.$$

This proves the result. □

## 4.2 The Subdifferential of a Convex Function

We have seen (Theorem 2.2.10) that if  $f$  is a differentiable function on an open convex set  $D \subset \mathbb{R}^p$ , then  $f$  is convex if and only if for every  $x, y \in D$ ,

$$f(y) \geq f(x) + \nabla f(x)(y - x).$$

As explained before, the differential  $\nabla f(x)$  is a linear functional from  $\mathbb{R}^p$  to  $\mathbb{R}$  and it can actually be identified to an element of  $\mathbb{R}^p$ . By this identification, the expression  $\nabla f(x)(y - x)$  can be written as well as  $\langle \nabla f(x), y - x \rangle$ .

Consider now the case of a convex function defined on a convex subset  $D$  of  $\mathbb{R}^p$ , i.e.,  $f : D \rightarrow \mathbb{R}$  which is not necessarily differentiable. Fix  $x \in D$ . Then it is natural to consider those elements  $u \in \mathbb{R}^p$  having the property that for all  $y \in D$ ,

$$\langle u, y - x \rangle \leq f(y) - f(x). \quad (4.2.1)$$

In this way, it is possible to replace in many results the missing differential at  $x$  with the set consisting of such elements. More precisely, an element satisfying (4.2.1) is called a subgradient of  $f$  at  $x$ , and the set of all such elements is denoted by  $\partial f(x)$  and is called the subdifferential of  $f$  at  $x$ .

We will observe that, in the case of a differentiable convex function,  $\partial f(x)$  reduces to  $\{\nabla f(x)\}$  for any  $x \in \text{int } D$ . Before that, we investigate some generalized differentiation properties of a convex function.

**Proposition 4.2.1.** *Let  $D \subset \mathbb{R}^p$  be a convex set,  $f : D \rightarrow \mathbb{R}$  be a convex function, and let  $x \in \text{int } D$ . Then for every  $v \in \mathbb{R}^p$ , then a directional derivative of  $f$  at  $x$  in the direction  $v$  exists and is defined as*

$$f'(x, v) := \lim_{t \rightarrow 0^+} \frac{f(x + tv) - f(x)}{t},$$

which can be written as

$$f'(x, v) = \inf_{t>0} \frac{f(x + tv) - f(x)}{t}.$$

*Proof* Fix  $x \in \text{int } D$ . For any  $v \in \mathbb{R}^p$  there exists  $t > 0$  such that  $x + tv \in D$ , so by the convexity of  $D$ ,  $x + sv \in D$  for all  $s \in [0, t]$ . It is enough to prove that the map

$$t \mapsto \frac{f(x + tv) - f(x)}{t}$$

is increasing (since every monotone function admits lateral limits, the conclusions follow). Indeed, if  $0 < s < t$ , then

$$\frac{f(x + sv) - f(x)}{s} \leq \frac{f(x + tv) - f(x)}{t}$$

means that

$$f(x + sv) \leq \frac{s}{t}f(x + tv) + \left(1 - \frac{s}{t}\right)f(x),$$

that is

$$f\left(\frac{s}{t}(x + tv) + \left(1 - \frac{s}{t}\right)x\right) \leq \frac{s}{t}f(x + tv) + \left(1 - \frac{s}{t}\right)f(x),$$

an inequality which follows from the convexity of  $f$ . □

**Proposition 4.2.2.** *Let  $D \subset \mathbb{R}^p$  be a convex set,  $f : D \rightarrow \mathbb{R}$  be a convex function, and let  $x \in \text{int } D$ . Then*

$$\partial f(x) = \{u \in \mathbb{R}^p \mid f'(x, v) \geq \langle u, v \rangle, \forall v \in \mathbb{R}^p\}.$$

*Proof* Take  $u \in \partial f(x)$ . Then

$$f(x + tv) - f(x) \geq \langle u, tv \rangle,$$

for all  $v \in \mathbb{R}^p$  and  $t > 0$  with  $x + tv \in D$ . This implies that

$$f'(x, v) \geq \langle u, v \rangle, \forall v \in \mathbb{R}^p.$$

Conversely, suppose that the above inequality holds. Take  $y \in D$ . Then  $v := y - x$  satisfies the inclusion  $x + v \in D$ . The monotonicity of

$$t \mapsto \frac{f(x + tv) - f(x)}{t}$$

(see the proof of Proposition 4.2.1) allows us to write

$$f(y) - f(x) = f(x + v) - f(x) \geq \inf_{t>0} \frac{f(x + tv) - f(x)}{t} \geq \langle u, v \rangle = \langle u, y - x \rangle,$$

which confirms that  $u \in \partial f(x)$ . □

**Proposition 4.2.3.** *Let  $D \subset \mathbb{R}^p$  be a convex set,  $f : D \rightarrow \mathbb{R}$  be a convex function, and let  $\bar{x} \in \text{int } D$ . If  $f$  is differentiable at  $\bar{x}$ , then  $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$ .*

*Proof* Theorem 2.2.10 yields the inclusion  $\nabla f(\bar{x}) \in \partial f(\bar{x})$ . Now, if  $u \in \partial f(\bar{x})$  then, by Proposition 4.2.2,

$$f'(x, v) \geq \langle u, v \rangle, \forall v \in \mathbb{R}^p.$$

But, since  $f$  is differentiable it is easy to observe that  $f'(x, v) = \langle \nabla f(x), v \rangle$ , so we have that

$$\langle \nabla f(x), v \rangle \geq \langle u, v \rangle, \forall v \in \mathbb{R}^p.$$

This implies that  $u = \nabla f(x)$  and the proof is complete. □

It is important to see that at a point of nondifferentiability, the subdifferential is not, in general, a singleton set, as the next example illustrates.

**Example 4.2.4.** Consider  $f : \mathbb{R} \rightarrow \mathbb{R}, f(x) = |x|$ . It is easy to see that this function is convex and it is differentiable on  $\mathbb{R} \setminus \{0\}$ . An easy computation reveals that the subdifferential of this convex function is

$$\partial f(x) = \begin{cases} -1, & \text{if } x < 0 \\ [-1, 1], & \text{if } x = 0 \\ 1, & \text{if } x > 0. \end{cases}$$

Another important fact about convex functions refers to their local Lipschitz behaviour, which induces important properties of the subdifferential. One can say that a function is locally Lipschitz if for every point in its domain there exists a neighborhood where  $f$  is Lipschitz. More precisely,  $f : D \subset \mathbb{R}^p \rightarrow \mathbb{R}$  is locally Lipschitz on  $D$  if for every point  $x \in D$ , there exist  $\varepsilon > 0$  and  $L > 0$  such that

$$|f(x) - f(y)| \leq L \|x - y\|$$

for every  $x, y \in B(x, \varepsilon) \cap D$ .

**Theorem 4.2.5.** Let  $D \subset \mathbb{R}^p$  be a convex set and  $f : D \rightarrow \mathbb{R}$  be a convex function. Then  $f$  is locally Lipschitz on  $\text{int } D$ .

*Proof* As shown in the first part of the proof of Theorem 2.2.4, for every point in  $\text{int } D$ , there is a neighborhood on which  $f$  is bounded. Take arbitrary  $\bar{x} \in \text{int } D$  and  $r > 0$ ,  $M > 0$  such that for every  $x \in D(\bar{x}, r) \subset \text{int } D$ ,  $f(x) \leq f(\bar{x}) + M$ . We show that for any  $r' \in (0, r)$  and any  $x, y \in D(\bar{x}, r')$ ,

$$|f(x) - f(y)| \leq \frac{M}{r} \cdot \frac{r + r'}{r - r'} \cdot \|x - y\|,$$

which implies that  $f$  is Lipschitz around  $\bar{x}$ . We can suppose, without loss of generality, that  $\bar{x} = 0$  and  $f(\bar{x}) = 0$  (otherwise, we consider instead of  $f$  the function  $g(x) := f(x + \bar{x}) - f(\bar{x})$ ). Consider  $r' \in (0, r)$  and  $x, y \in D(\bar{x}, r')$ ,  $x \neq y$ . There exist  $z$  with  $\|z\| = r$

and  $\alpha \in (0, 1)$  such that  $y = (1 - \alpha)x + \alpha z$ . Indeed, taking  $\varphi(t) := ty + (1 - t)x$ , the map  $t \mapsto \|\varphi(t)\|$  is continuous on  $[1, \infty)$ ,  $\|\varphi(1)\| = \|y\| < r'$  and  $\lim_{t \rightarrow \infty} \|\varphi(t)\| \rightarrow \infty$ , since  $\|\varphi(t)\| \geq t\|x - y\| - \|x\|$ . Therefore, there exists  $\bar{t} > 1$  such that  $\|\varphi(\bar{t})\| = r \in (r', \infty)$ . Take  $z := \varphi(\bar{t})$ , i.e.,

$$z = \bar{t}y + (1 - \bar{t})x.$$

Taking  $\alpha := \bar{t}^{-1}$ , we obtain that

$$y = (1 - \alpha)x + \alpha z.$$

Moreover,  $\alpha = \|z - x\|^{-1} \|y - x\|$ . Since

$$f(y) \leq (1 - \alpha)f(x) + \alpha f(z),$$

we get

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

From

$$0 = f(0) = f\left(2^{-1}x + 2^{-1}(-x)\right) \leq 2^{-1}f(x) + 2^{-1}f(-x),$$

we obtain  $-f(x) \leq f(-x)$ . But, if  $x \neq 0$ ,  $u := r\|x\|^{-1}(-x) \in D(0, r)$ , so

$$f(-x) = f\left(\frac{\|x\|}{r}u\right) \leq \frac{\|x\|}{r}f(u) \leq M\frac{\|x\|}{r}.$$

Clearly, this is true also for  $x = 0$ . Since  $\|z - x\| \geq r - r'$  and  $r + \|x\| \leq r + r'$ , we finally obtain

$$f(y) - f(x) \leq \frac{\|y - x\|}{\|z - x\|} \left(M + M\frac{\|x\|}{r}\right) \leq \frac{M}{r} \cdot \frac{r + r'}{r - r'} \cdot \|x - y\|.$$

Interchanging  $x$  and  $y$ , the theorem is proved. □

A cornerstone for the development of a subdifferential calculus (i.e., calculus with subdifferentials) is to observe that, in general, the subdifferential at a given point is a nonempty set.

**Theorem 4.2.6.** *Let  $D \subset \mathbb{R}^p$  be a convex set and  $f : D \rightarrow \mathbb{R}$  be a convex function. Then for every  $x \in \text{int } D$ , the subdifferential  $\partial f(x)$  is a nonempty compact set.*

*Proof* Fix  $x \in \text{int } D$ . First we show that  $\partial f(x)$  is nonempty. From Theorem 2.2.4, the function  $f$  is continuous at  $x$ . In particular, this shows that the convex set  $\text{int epi } f$  is nonempty (one applies here both Proposition 2.2.3 (ii) and Theorem 4.1.1 (iii)). Clearly,  $(x, f(x)) \notin \text{int epi } f$ , so, from Theorem 4.1.3, there exists  $(u, \alpha) \in \mathbb{R}^p \times \mathbb{R} \setminus \{(0, 0)\}$  such that for every  $(y, t) \in \text{int epi } f$ ,

$$\langle u, y \rangle + \alpha t \leq \langle u, x \rangle + \alpha f(x).$$

Moreover, this is true for every  $(y, t) \in \text{cl}(\text{int epi } f) = \text{cl epi } f = \text{epi } f$ . If we suppose that  $\alpha = 0$ , then the inequality  $\langle u, x - y \rangle \geq 0$  holds for all  $y \in D$ , and this means that  $u = 0$ , which contradicts the fact that the pair  $(u, \alpha)$  is not zero. Consequently,  $\alpha \neq 0$ . But, letting  $t \rightarrow \infty$  in the above inequality, one deduces that  $\alpha < 0$ . Therefore, we can take  $\alpha = -1$ , and then we get

$$\langle u, y \rangle - t \leq \langle u, x \rangle - f(x), \quad \forall (y, t) \in \text{epi } f.$$

In particular, for every  $y \in D$ ,

$$\langle u, y \rangle - f(y) \leq \langle u, x \rangle - f(x),$$

that is

$$\langle u, y - x \rangle \leq f(y) - f(x),$$

i.e.,  $u \in \partial f(x)$ .

Let us notice now that the fact that  $\partial f(x)$  is always a closed set is an easy remark. It remains to show that  $\partial f(x)$  is a bounded set. By virtue of Theorem 4.2.5,  $f$  is Lipschitz on a neighborhood of  $x$ . Denote by  $L$  the corresponding Lipschitz constant on the given neighborhood  $V := D(x, \varepsilon) \subset D$  of  $x$ . Take  $u \in \partial f(x)$ . Then for any  $y \in V$ ,

$$\langle u, y - x \rangle \leq f(y) - f(x) \leq |f(y) - f(x)| \leq L \|y - x\|.$$

Since  $y = x + \varepsilon v \in V$  for any  $v$  in the unit ball, this implies that

$$\langle u, v \rangle \leq L \|v\|, \quad \forall v \in D(0, 1),$$

which means that  $\|u\| \leq L$ . Then  $\partial f(x)$  is bounded. □

The next calculus rule is fundamental.

**Theorem 4.2.7** (sum rule).  *$D \subset \mathbb{R}^p$  be a convex set with nonempty interior and let  $f, g : D \rightarrow \mathbb{R}$  be convex functions. Then for any  $x \in D$ ,*

$$\partial(f + g)(x) = \partial f(x) + \partial g(x).$$

*Proof* The inclusion  $\partial f(x) + \partial g(x) \subset \partial(f + g)(x)$  easily follows from the definition of the subdifferential. Consider  $u \in \partial(f + g)(x)$ . We can consider the case  $x = 0, f(0) = g(0) = 0$  directly, simply by taking  $f_1(y) := f(y + x) - f(x)$  and  $g_1(y) := g(y + x) - g(x)$ . Denote

$$A := \text{int epi } f \text{ and } B := \{(v, t) \in D \times \mathbb{R} \mid t \leq \langle u, v \rangle - g(v)\}$$

Both these sets are nonempty and convex. If  $(v, t) \in A \cap B$  did exist, then, on one hand,  $f(v) < t$ , and, on the other hand, since  $u \in \partial(f + g)(0)$ ,

$$f(v) < t \leq \langle u, v \rangle - g(v) \leq f(v),$$



which is not possible. Therefore,  $A \cap B = \emptyset$ . Theorem 4.1.4 ensures the existence of an element  $(a, \alpha) \in \mathbb{R}^p \times \mathbb{R} \setminus \{(0, 0)\}$  such that for all  $(u, s) \in A$  and  $(v, t) \in B$ ,

$$\langle a, v \rangle + \alpha t \leq \langle a, u \rangle + \alpha s.$$

It follows that this inequality holds for any  $(u, s) \in \text{cl } A = \text{epi } f$  and any  $(v, t) \in B$ . If  $\alpha = 0$ , then  $\langle a, v \rangle \leq \langle a, u \rangle$  for every  $u, v \in D$ , which means that  $a = 0$  (otherwise,  $a$  would have positive scalar product with all the elements of a ball centered at  $0$ , because  $D$  has nonempty interior), an impossible situation. Therefore,  $\alpha \neq 0$ . Letting  $s \rightarrow \infty$ , we infer that  $\alpha > 0$  and therefore we can take  $\alpha = 1$ , so

$$\langle a, v \rangle + t \leq \langle a, u \rangle + s, \quad \forall (u, s) \in \text{epi } f, \quad \forall (v, t) \in B.$$

For  $(v, t) = (0, 0)$  and  $s = f(u)$ , we get

$$\langle -a, u \rangle \leq f(u), \quad \forall u \in \text{epi } f,$$

i.e.,  $-a \in \partial f(0)$ . On the other hand, for  $(u, s) = (0, 0)$ , we obtain

$$\langle a, v \rangle + t \leq 0, \quad \forall (v, t) \in B,$$

which implies that

$$\langle a + u, v \rangle \leq g(v), \quad \forall v \in D,$$

i.e.,  $a + u \in \partial g(0)$ , so  $u = -a + a + u \in \partial f(0) + \partial g(0)$ . □

### 4.3 Optimality Conditions

We give now optimality conditions for convex nondifferentiable data for the basic optimization problems already studied in the smooth case. We start with the case of optimization without constraints.

**Theorem 4.3.1.** *Let  $D \subset \mathbb{R}^p$  be a convex set, and  $f : D \rightarrow \mathbb{R}$  be a convex function. An element  $\bar{x} \in D$  is a (local) minimum point of  $f$  if and only if  $0 \in \partial f(\bar{x})$ .*

*Proof* If  $\bar{x}$  is a local minimum point, then it is a global one (see Proposition 3.1.23), and therefore

$$0 = \langle 0, y - \bar{x} \rangle \leq f(y) - f(\bar{x})$$

for all  $y \in D$ . Then,  $0 \in \partial f(\bar{x})$ .

Conversely, suppose that  $0 \in \partial f(\bar{x})$ . From the definition of the subgradients, this means that  $\langle 0, y - \bar{x} \rangle \leq f(y) - f(\bar{x})$  for all  $y \in D$ , whence  $\bar{x}$  is a minimum point for  $f$ . □

We now consider the optimization problems with geometric constraints. In order to tackle this case, we need a general penalization result. Generally speaking, to penalize an optimization problem means to transform it in such a way that a (local) minimum point of the initial constrained problem becomes a (local) minimum point of the

unconstrained problem. This can be done by adding to the objective function a penalty term which somehow contains the constraints of the initial problem. We present here a method to penalize a given optimization problem (not necessarily convex) with geometric constraints due to Canadian–French mathematician Frank H. Clarke, who proposed it in 1983.

**Theorem 4.3.2.** *Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  be a function and  $M \subset \mathbb{R}^p$  a closed set. Consider the problem*

$$\min f(x), \quad x \in M.$$

*Suppose that  $\bar{x}$  is a (local) solution of this problem and  $f$  is locally Lipschitz of constant  $L \geq 0$  around  $\bar{x}$ . Then  $\bar{x}$  is a (local) minimum without constraints of the mapping  $x \mapsto f(x) + Ld_M(x)$ .*

*Proof* Let  $U$  be the neighborhood of  $\bar{x}$  such that  $f$  is  $L$ -Lipschitz on  $U$  and  $f(\bar{x}) \leq f(x)$  for every  $x \in U \cap M$ . Then there exists  $\alpha > 0$  such that  $B(\bar{x}, \alpha) \subset U$ . Consider  $V = B(\bar{x}, 3^{-1}\alpha)$  and take  $x \in V$  arbitrarily. First, if  $x \in V \cap M \subset U \cap M$ , it is clear that

$$f(\bar{x}) + Ld_M(\bar{x}) \leq f(x) + Ld_M(x).$$

Then, if  $x \in V \setminus M$ , for every  $\varepsilon \in (0, 3^{-1}\alpha)$ , there exists  $x_\varepsilon \in M$  such that

$$\begin{aligned} \|x - x_\varepsilon\| &< d_M(x) + \varepsilon \\ &\leq \|x - \bar{x}\| + \varepsilon \\ &\leq 3^{-1}\alpha + \varepsilon < 2 \cdot 3^{-1}\alpha. \end{aligned}$$

Then,

$$\begin{aligned} \|x_\varepsilon - \bar{x}\| &\leq \|x_\varepsilon - x\| + \|x - \bar{x}\| \\ &< 2 \cdot 3^{-1}\alpha + 3^{-1}\alpha = \alpha. \end{aligned}$$

Consequently,  $x_\varepsilon \in U \cap M$ , whence,

$$\begin{aligned} f(\bar{x}) + Ld_M(\bar{x}) &\leq f(x_\varepsilon) \leq f(x) + L\|x - x_\varepsilon\| \\ &\leq f(x) + L(d_M(x) + \varepsilon) \\ &= f(x) + Ld_M(x) + L\varepsilon. \end{aligned}$$

Letting  $\varepsilon \rightarrow 0$ , we get the conclusion. □

We now connect the normal cone to a convex set to the subdifferential of the (non-differentiable) distance function: see Propositions 2.1.15 and 2.2.9. We shall see that similar relations hold in more general settings, as shown in Chapter 5.

**Proposition 4.3.3.** *Let  $C \subset \mathbb{R}^p$  be a nonempty closed convex set and  $\bar{x} \in C$ . Then*

$$[0, \infty) \partial d_C(\bar{x}) = N(C, \bar{x}).$$

*Proof* Take  $u \in \partial d_C(\bar{x})$ . Then

$$\langle u, y - \bar{x} \rangle \leq d_C(y) - d_C(\bar{x}), \quad \forall y \in \mathbb{R}^p.$$

In particular, for  $y \in C$ ,

$$\langle u, y - \bar{x} \rangle \leq 0,$$

which proves that  $u \in N(C, \bar{x})$ . Consequently,  $\partial d_C(\bar{x}) \subset N(C, \bar{x})$  and since the latter set is a cone, one infers that  $[0, \infty) \partial d_C(\bar{x}) \subset N(C, \bar{x})$ .

Conversely, take  $u \in N(C, \bar{x})$  with  $\|u\| \leq 1$ , i.e.,

$$\langle u, c - \bar{x} \rangle \leq 0, \quad \forall c \in C.$$

Then

$$\langle -u, c \rangle \geq \langle -u, \bar{x} \rangle, \quad \forall c \in C,$$

which means that  $\bar{x}$  is a minimum point of the mapping  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ ,  $f(x) := \langle -u, x \rangle$  on the set  $C$ . It is clear that  $f$  is 1-Lipschitz (since  $\|u\| \leq 1$ ), and therefore, by virtue of Theorem 4.3.2,  $\bar{x}$  is a minimum without restrictions of the convex function  $f + d_C$ . Therefore, Theorem 4.3.1 gives

$$0 \in \partial(f + d_C)(\bar{x}),$$

and, by means of Theorem 4.2.7, we get

$$0 \in \partial f(\bar{x}) + \partial d_C(\bar{x}).$$

Since  $f$  is differentiable and  $\nabla f(\bar{x}) = -u$ , from Proposition 4.2.3,  $\partial f(\bar{x}) = \{-u\}$ . Consequently,  $u \in \partial d_C(\bar{x})$ , so we have shown that

$$N(C, \bar{x}) \cap D(0, 1) \subset \partial d_C(\bar{x}),$$

and the conclusion follows. □

Now we are able to present the optimality condition for convex geometric constraint problems.

**Theorem 4.3.4** (Pshenichnyi-Rockafellar). *Let  $D \subset \mathbb{R}^p$  be an open convex set,  $f : D \rightarrow \mathbb{R}$  be a convex function, and  $C \subset D$  be a nonempty closed convex set. A point  $\bar{x} \in C$  is a local minimum point of  $f$  on  $C$  if and only if*

$$0 \in \partial f(\bar{x}) + N(C, \bar{x}).$$

*Proof* Suppose that  $\bar{x} \in C$  is a local minimum point of  $f$  on  $C$ . Since  $f$  is locally Lipschitz (Theorem 4.2.5) on  $D$  (notice that we assume that  $\text{int } D = D$ ), we can apply Theorem 4.3.2 to deduce that  $\bar{x}$  is a local minimum (without constraints) of the convex function

$f + Ld_C$  defined on  $D$ , where  $L > 0$  is the Lipschitz constant of  $f$  around  $\bar{x}$ . Using Theorem 4.3.1, we can write

$$0 \in \partial(f + Ld_C)(\bar{x}),$$

and by virtue of Theorem 4.2.7,

$$0 \in \partial f(\bar{x}) + \partial(Ld_C)(\bar{x}).$$

It is easy to see that  $\partial(Ld_C)(\bar{x}) = L\partial(d_C)(\bar{x})$  and, finally, Proposition 4.3.3 allows us to write

$$0 \in \partial f(\bar{x}) + N(C, \bar{x}).$$

Conversely, suppose that the above condition holds. This means that there exists  $a \in -N(C, \bar{x}) \cap \partial f(\bar{x})$ , i.e., for every  $y \in C$ ,

$$0 \leq \langle a, y - \bar{x} \rangle \leq f(y) - f(\bar{x}),$$

which confirms that  $\bar{x}$  is a minimum point of  $f$  on  $C$ . □

Let now us consider an open convex set  $D \subset \mathbb{R}^p$ . We study the case of convex optimization with functional constraints, that is the problem (P) of minimizing a convex objective function  $f : D \rightarrow \mathbb{R}$  under the constraint  $x \in C := \{x \in D \mid g(x) \leq 0, h(x) = 0\}$  where  $g : D \rightarrow \mathbb{R}^n$  has all the component functions  $g_1, g_2, \dots, g_n : D \rightarrow \mathbb{R}$  convex and  $h : D \rightarrow \mathbb{R}^m$  is affine. Due to the assumptions on  $D$ ,  $g$  and  $h$ , the set  $C$  is convex. We discussed this problem in Section 3.2 with smoothness assumptions on  $f$ ,  $g$  and  $h$ . Now, this problem is investigated without the differentiability hypotheses, but using subdifferential calculus.

To begin with, we give a Fritz John type result. We use the notations  $\mathbb{R}_+^p := (\mathbb{R}_+)^p$  and  $\mathbb{R}^p := (-\mathbb{R}_+)^p$ .

**Theorem 4.3.5.** *Suppose that  $\bar{x}$  is a solution of the convex problem (P). Then there exist  $\lambda_0 \geq 0$ ,  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n) \in \mathbb{R}^n$ ,  $\mu = (\mu_1, \mu_2, \dots, \mu_m) \in \mathbb{R}^m$ , with  $\lambda_0 + \|\lambda\| + \|\mu\| \neq 0$ , such that*

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

and

$$\lambda_i \geq 0, \lambda_i g_i(\bar{x}) = 0, \forall i \in \overline{1, n}.$$

*Proof* As discussed in the smooth case, we can suppose, without loss of generality, that  $A(\bar{x}) = \overline{1, n}$ . Take

$$A := \{(f(x) - f(\bar{x}) + t, g(x) + s, h(x)) \mid x \in D, t \in [0, \infty), s \in \mathbb{R}_+^p\},$$

and

$$B := (-\infty, 0) \times \mathbb{R}^p \times \{0\}^m.$$

It is easy to see that the properties of  $f$ ,  $g$  and  $h$  (i.e.,  $f$  convex,  $g_i$  convex for all  $i \in \overline{1, n}$ , and  $h$  affine) ensure the convexity of  $A$ , while the convexity of  $B$  is obvious. On the other hand, if a common element of  $A$  and  $B$  exists, then there also exists  $x \in D$  with  $f(x) - f(\bar{x}) < 0$ ,  $g(x) \leq 0$ ,  $h(x) = 0$ , which would contradict the (global) minimality of  $\bar{x}$ . Consequently,  $A \cap B = \emptyset$ . We apply Theorem 4.1.4 in order to deduce the existence of some elements  $\lambda_0 \in \mathbb{R}$ ,  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n) \in \mathbb{R}^n$ ,  $\mu = (\mu_1, \mu_2, \dots, \mu_m) \in \mathbb{R}^m$ , with  $|\lambda_0| + \|\lambda\| + \|\mu\| \neq 0$  such that

$$\lambda_0 a + \langle \lambda, u \rangle \leq \lambda_0(f(x) - f(\bar{x}) + t) + \langle \lambda, g(x) + s \rangle + \langle \mu, h(x) \rangle$$

for all  $x \in D$ ,  $t \in [0, \infty)$ ,  $s \in \mathbb{R}_+^p$ ,  $a \in (-\infty, 0)$ ,  $u \in \mathbb{R}^p$ . It is not possible to have  $\lambda_0 < 0$ . Indeed, if we suppose so, then for fixed  $x$ ,  $t$ ,  $s$  and  $u$ , letting  $a \rightarrow -\infty$ , we arrive at a contradiction, since the right-hand side is fixed, while the left-hand side goes to  $+\infty$ . A similar argument employed for  $s_i \rightarrow \infty$  (for all  $i \in \overline{1, n}$ ) allows us to conclude that  $\lambda_i \geq 0$ , for all  $i \in \overline{1, n}$ . Letting  $a \rightarrow 0$ ,  $t \rightarrow 0$ ,  $u \rightarrow 0$  and  $s \rightarrow 0$ , we actually get that

$$0 \leq \lambda_0(f(x) - f(\bar{x})) + \langle \lambda, g(x) \rangle + \langle \mu, h(x) \rangle$$

for all  $x \in D$ . Since  $g(\bar{x}) = 0$ , we deduce that  $\lambda_i g_i(\bar{x}) = 0$ , for all  $i \in \overline{1, n}$ . Finally, we can write

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

which means that

$$0 \in \partial \left( \lambda_0 f + \sum_{i=1}^n \lambda_i g_i + \sum_{j=1}^m \mu_j h_j \right) (\bar{x}).$$

Theorem 4.2.7 allows us to write

$$0 \in \partial(\lambda_0 f)(\bar{x}) + \sum_{i=1}^n \partial(\lambda_i g_i)(\bar{x}) + \sum_{j=1}^m \partial(\mu_j h_j)(\bar{x}).$$

Since  $\lambda_0 \geq 0$ , one has that  $\partial \lambda_0 f(\bar{x}) = \lambda_0 \partial f(\bar{x})$ . The same argument is applicable to the equality  $\partial \lambda_i g_i(\bar{x}) = \lambda_i \partial g_i(\bar{x})$  for all  $i \in \overline{1, n}$ . The equality  $\partial \mu_j h_j(\bar{x}) = \mu_j \partial h_j(\bar{x})$  for all  $j \in \overline{1, m}$  is true by the fact that  $h$  is affine. The proof is complete.  $\square$

As we already saw several times, in order to pass from Fritz John type conditions to Karush-Kuhn-Tucker type condition (i.e., to ensure  $\lambda_0 \neq 0$ ) one needs a supplementary (constraint qualification) assumption. Since we are now in the convex case, we use here a Slater type condition, which for the nondifferentiable case looks as follows: there exists  $u \in D$  such that  $g(u) < 0$ ,  $h(u) = 0$ , and the set  $\{\nabla h_j(u) \mid j \in \overline{1, m}\}$  is linearly independent (for  $u \in D$ ; notice that  $\nabla h_j(u)$  is the same for any  $u \in D$ ). Under this condition, we can deduce from Theorem 4.3.5 a Karush-Kuhn-Tucker theorem, as follows.

**Theorem 4.3.6.** *Suppose that  $\bar{x}$  is a solution of the convex problem (P) and that the Slater condition holds. Then there exist  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n) \in \mathbb{R}^n$ ,  $\mu = (\mu_1, \mu_2, \dots, \mu_m) \in \mathbb{R}^m$  such that*

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

and

$$\lambda_i \geq 0, \lambda_i g_i(\bar{x}) = 0, \forall i \in \overline{1, n}.$$

*Proof* It is sufficient to show that in Theorem 4.3.5 is not possible to have  $\lambda_0 = 0$ . Suppose, by way of contradiction, that  $\lambda_0 = 0$ . Then

$$0 \in \sum_{i=1}^n \lambda_i \partial g_i(\bar{x}) + \sum_{j=1}^m \mu_j \partial h_j(\bar{x}) = \partial \left( \sum_{i=1}^n \lambda_i g_i + \sum_{j=1}^m \mu_j h_j \right) (\bar{x})$$

that is, for every  $x \in D$ ,

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

i.e.,

$$\sum_{i=1}^n \lambda_i g_i(x) + \sum_{j=1}^m \mu_j h_j(x) \geq 0.$$

For  $x = u$  (the element from Slater condition), this becomes

$$\sum_{i=1}^n \lambda_i g_i(u) \geq 0,$$

which is possible only if  $\lambda_i = 0$  for all  $i \in \overline{1, n}$ . So, in fact,

$$\sum_{j=1}^m \mu_j h_j(x) \geq \sum_{j=1}^m \mu_j h_j(\bar{x}) = 0, \forall x \in D.$$

This means that  $\bar{x}$  is a minimum point (without constraint) for the function  $x \mapsto \sum_{j=1}^m \mu_j h_j(x)$ , so, from the Fermat Theorem,

$$\sum_{j=1}^m \mu_j \nabla h_j(\bar{x}) = 0.$$

The linear independence condition says that the above relation holds only if  $\mu_j = 0$  for all  $j \in \overline{1, m}$ . The facts obtained up to now, collectively imply that  $\lambda_0 + \|\lambda\| + \|\mu\| = 0$ , which is a contradiction. The proof is complete.  $\square$