Chapter 2
Convex Analysis

The theory of nonsmooth analysis is based on convex analysis. Thus, we start this chapter by giving basic concepts and results of convexity (for further readings see also [202, 204]). We take a geometrical viewpoint by examining the tangent and normal cones of convex sets. Then we generalize the concepts of differential calculus for convex, not necessarily differentiable functions [204]. We define subgradients and subdifferentials and present some basic results. At the end of this chapter, we link these analytical and geometrical concepts together.

2.1 Convex Sets

We start this section by recalling the definition of a convex set.

Definition 2.1 Let S be a subset of $\mathbb{R}^n$. The set S is said to be *convex* if

$$\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y} \in S,$$

for all $\boldsymbol{x}, \boldsymbol{y} \in S$ and $\lambda \in [0, 1]$.

Geometrically this means that the set is convex if the closed line-segment $[\boldsymbol{x}, \boldsymbol{y}]$ is entirely contained in S whenever its endpoints $\boldsymbol{x}$ and $\boldsymbol{y}$ are in S (see Fig. 2.1).

Example 2.1 (*Convex sets*). Evidently the empty set $\emptyset$, a singleton $\{\boldsymbol{x}\}$, the whole space $\mathbb{R}^n$, linear subspaces, open and closed balls and halfspaces are convex sets. Furthermore, if S is a convex set also $\operatorname{cl} S$ and $\operatorname{int} S$ are convex.

Theorem 2.1 *Let $S_i \subseteq \mathbb{R}^n$ be convex sets for $i = 1, \ldots, m$. Then their intersection*

$$\bigcap_{i=1}^{m} S_i \tag{2.1}$$

A. Bagirov et al., *Introduction to Nonsmooth Optimization*,
DOI: 10.1007/978-3-319-08114-4_2

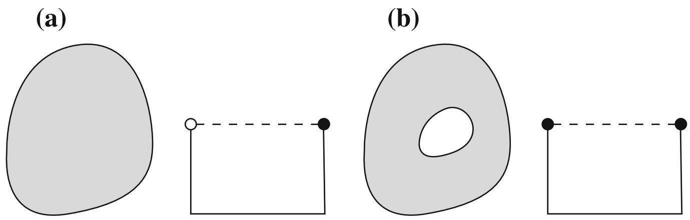


Fig. 2.1 Illustration of convex and nonconvex sets. (**a**) Convex. (**b**) Not convex

is also convex.

Proof Let $\boldsymbol{x}, \boldsymbol{y} \in \bigcap_{i=1}^{m} S_i$ and $\lambda \in [0, 1]$ be arbitrary. Because $\boldsymbol{x}, \boldsymbol{y} \in S_i$ and S_i is convex for all $i = 1, \ldots, m$, we have $\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in S_i$ for all $i = 1, \ldots, m$. This implies that

$$\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in \bigcap_{i=1}^{m} S_i$$

and the proof is complete. □

Example 2.2 (*Intersection of convex sets*). The hyperplane

$$H(\boldsymbol{p}, \alpha) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{p}^T(\boldsymbol{x} - \boldsymbol{x}_0) = 0\},$$

where $\boldsymbol{x}_0, \boldsymbol{p} \in \mathbb{R}^n$ and $\boldsymbol{p} \neq \boldsymbol{0}$ is convex, since it can be represent as an intersection of two convex closed halfspaces as

$$\begin{aligned} H(\boldsymbol{p}, \alpha) &= H^+(\boldsymbol{p}, \alpha) \cap H^-(\boldsymbol{p}, \alpha) \\ &= \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{p}^T(\boldsymbol{x} - \boldsymbol{x}_0) \geq 0\} \cap \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{p}^T(\boldsymbol{x} - \boldsymbol{x}_0) \leq 0\}. \end{aligned}$$

The next theorem shows that the space of convex sets has some linear properties. This is due to fact that the space of convex sets is a subspace of the power set $\mathcal{P}(\mathbb{R}^n)$ consisting of all subsets of $\mathbb{R}^n$.

Theorem 2.2 *Let $S_1, S_2 \subseteq \mathbb{R}^n$ be nonempty convex sets and $\mu_1, \mu_2 \in \mathbb{R}$. Then the set $\mu_1 S_1 + \mu_2 S_2$ is also convex.*

Proof Let the points $\boldsymbol{x}, \boldsymbol{y} \in \mu_1 S_1 + \mu_2 S_2$ and $\lambda \in [0, 1]$. Then $\boldsymbol{x}$ and $\boldsymbol{y}$ can be written

$$\begin{cases} \boldsymbol{x} &= \mu_1 \boldsymbol{x}_1 + \mu_2 \boldsymbol{x}_2, \quad \text{where } \boldsymbol{x}_1 \in S_1 \text{ and } \boldsymbol{x}_2 \in S_2 \\ \boldsymbol{y} &= \mu_1 \boldsymbol{y}_1 + \mu_2 \boldsymbol{y}_2, \quad \text{where } \boldsymbol{y}_1 \in S_1 \text{ and } \boldsymbol{y}_2 \in S_2 \end{cases}$$

and

$$\begin{aligned}\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} &= \lambda(\mu_1 \boldsymbol{x}_1 + \mu_2 \boldsymbol{x}_2) + (1-\lambda)(\mu_1 \boldsymbol{y}_1 + \mu_2 \boldsymbol{y}_2)\\ &= \mu_1(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{y}_1) + \mu_2(\lambda \boldsymbol{x}_2 + (1-\lambda)\boldsymbol{y}_2)\\ &\in \mu_1 S_1 + \mu_2 S_2.\end{aligned}$$

Thus the set $\mu_1 S_1 + \mu_2 S_2$ is convex. □

2.1.1 Convex Hulls

A linear combination $\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i$ is called a *convex combination* of elements $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}^n$ if each $\lambda_i \geq 0$ and $\sum_{i=1}^{k} \lambda_i = 1$. The convex hull generated by a set is defined as a set of convex combinations as follows.

Definition 2.2 The *convex hull* of a set $S \subseteq \mathbb{R}^n$ is

$$\text{conv } S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i,\ \sum_{i=1}^{k} \lambda_i = 1,\ \boldsymbol{x}_i \in S,\ \lambda_i \geq 0, k > 0\}.$$

The proof of the next lemma is left as an exercise.

Lemma 2.1 *If $S \subseteq \mathbb{R}^n$, then* conv S *is a convex set and S is convex if and only if*

$$S = \text{conv } S.$$

Proof Exercise. □

The next theorem shows that the convex hull is actually the intersection of all the convex sets containing the set, in other words, it is the smallest convex set containing the set itself (see Fig. 2.2).

Theorem 2.3 *If $S \subseteq \mathbb{R}^n$, then*

$$\text{conv } S = \bigcap_{\substack{S \subseteq \hat{S} \\ \hat{S} \text{ convex}}} \hat{S}.$$

Proof Let $\hat{S}$ be convex such that $S \subseteq \hat{S}$. Then due to Lemma 2.1 we have conv $S \subseteq$ conv $\hat{S} = \hat{S}$ and thus we have

$$\text{conv } S \subseteq \bigcap_{\substack{S \subseteq \hat{S} \\ \hat{S} \text{ convex}}} \hat{S}.$$

On the other hand, it is evident that $S \subseteq \text{conv } S$ and due to Lemma 2.1 conv S is a convex set. Then conv S is one of the sets $\hat{S}$ forming the intersection and thus

$$\bigcap_{\substack{S\subseteq \hat{S}\\ \hat{S} \text{ convex}}} \hat{S} = \bigcap_{\substack{S\subseteq \hat{S}\\ \hat{S} \text{ convex}}} \hat{S} \cap \text{conv } S \subseteq \text{conv } S$$

and the proof is complete. $\square$

2.1.2 Separating and Supporting Hyperplanes

Next we consider some nice properties of hyperplanes. Before those we need the concept of distance function.

Definition 2.3 Let $S \subseteq \mathbb{R}^n$ be a nonempty set. The *distance function* $d_S\colon \mathbb{R}^n \to \mathbb{R}$ to the set S is defined by

$$d_S(\boldsymbol{x}) := \inf \{\|\boldsymbol{x} - \boldsymbol{y}\| \mid \boldsymbol{y} \in S\} \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^n. \tag{2.2}$$

The following lemma shows that a closed convex set always has a unique closest point.

Lemma 2.2 *Let $S \subset \mathbb{R}^n$ be a nonempty, closed convex set and $\boldsymbol{x}^* \notin S$. Then there exists a unique $\boldsymbol{y}^* \in \text{bd } S$ minimizing the distance to $\boldsymbol{x}^*$. In other words*

$$d_S(\boldsymbol{x}^*) = \|\boldsymbol{x}^* - \boldsymbol{y}^*\|.$$

Moreover, a necessary and sufficient condition for a such $\boldsymbol{y}^*$ is that

$$(\boldsymbol{x}^* - \boldsymbol{y}^*)^T(\boldsymbol{x} - \boldsymbol{y}^*) \leq 0 \quad \text{for all } \boldsymbol{x} \in S. \tag{2.3}$$

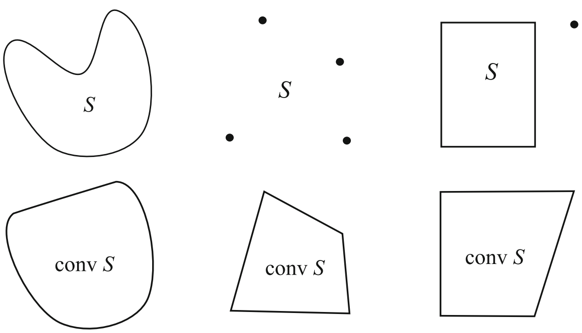


Fig. 2.2 Examples of convex hulls

Proof First we prove the existence of a closest point. Since $S \neq \emptyset$, there exists $\hat{\boldsymbol{x}} \in S$ and we can define $\hat{S} := S \cap \text{cl } B(\boldsymbol{x}^*; r)$, where $r := \|\boldsymbol{x}^* - \hat{\boldsymbol{x}}\| > 0$. Then $\hat{S} \neq \emptyset$ since $\hat{\boldsymbol{x}} \in \hat{S}$. Moreover, $\hat{S}$ is closed, since both S and cl $B(\boldsymbol{x}^*; r)$ are closed, and bounded, since $\hat{S} \subseteq \text{cl } B(\boldsymbol{x}^*; r)$, thus $\hat{S}$ is a nonempty compact set. Then, due to Weierstrass' Theorem 3.1 the continuous function

$$g(\boldsymbol{y}) := \|\boldsymbol{x}^* - \boldsymbol{y}\|$$

attains its minimum over $\hat{S}$ at some $\boldsymbol{y}^* \in \hat{S}$ and we have

$$d_{\hat{S}}(\boldsymbol{x}^*) = g(\boldsymbol{y}^*) = \|\boldsymbol{x}^* - \boldsymbol{y}^*\|.$$

If $\boldsymbol{y} \in S \setminus \hat{S}$, it means that $\boldsymbol{y} \notin \text{cl } B(\boldsymbol{x}^*; r)$, in other words

$$g(\boldsymbol{y}) > r \geq g(\boldsymbol{y}^*)$$

and thus

$$d_S(\boldsymbol{x}^*) = g(\boldsymbol{y}^*) = \|\boldsymbol{x}^* - \boldsymbol{y}^*\|.$$

In order to show the uniqueness, suppose that there exists another $\boldsymbol{z}^* \in S$ such that $\boldsymbol{z}^* \neq \boldsymbol{y}^*$ and $g(\boldsymbol{z}^*) = g(\boldsymbol{y}^*)$. Then due to convexity we have $\frac{1}{2}(\boldsymbol{y}^* + \boldsymbol{z}^*) \in S$ and by triangle inequality

$$\begin{aligned} g\left(\tfrac{1}{2}(\boldsymbol{y}^* + \boldsymbol{z}^*)\right) &= \|\boldsymbol{x}^* - \tfrac{1}{2}(\boldsymbol{y}^* + \boldsymbol{z}^*)\| \leq \tfrac{1}{2}\|\boldsymbol{x}^* - \boldsymbol{y}^*\| + \tfrac{1}{2}\|\boldsymbol{x}^* - \boldsymbol{z}^*\| \\ &= \tfrac{1}{2}g(\boldsymbol{y}^*) + \tfrac{1}{2}g(\boldsymbol{z}^*) = g(\boldsymbol{y}^*). \end{aligned}$$

The strict inequality cannot hold since g attains its minimum over S at $\boldsymbol{y}^*$. Thus we have

$$\|(\boldsymbol{x}^* - \boldsymbol{y}^*) + (\boldsymbol{x}^* - \boldsymbol{z}^*)\| = \|\boldsymbol{x}^* - \boldsymbol{y}^*\| + \|\boldsymbol{x}^* - \boldsymbol{z}^*\|,$$

which is possible only if the vectors $\boldsymbol{x}^* - \boldsymbol{y}^*$ and $\boldsymbol{x}^* - \boldsymbol{z}^*$ are collinear. In other words $\boldsymbol{x}^* - \boldsymbol{y}^* = \lambda(\boldsymbol{x}^* - \boldsymbol{z}^*)$ for some $\lambda \in \mathbb{R}$. Since

$$\|\boldsymbol{x}^* - \boldsymbol{y}^*\| = \|\boldsymbol{x}^* - \boldsymbol{z}^*\|$$

we have $\lambda = \pm 1$. If $\lambda = -1$ we have

$$\boldsymbol{x}^* = \tfrac{1}{2}(\boldsymbol{y}^* + \boldsymbol{z}^*) \in S,$$

which contradicts the assumption $\boldsymbol{x}^* \notin S$, and if $\lambda = 1$, we have $\boldsymbol{z}^* = \boldsymbol{y}^*$, thus $\boldsymbol{y}^*$ is a unique closest point.

Next we show that $\boldsymbol{y}^* \in \text{bd } S$. Suppose, by contradiction, that $\boldsymbol{y}^* \in \text{int } S$. Then there exists $\varepsilon > 0$ such that $B(\boldsymbol{y}^*; \varepsilon) \subset S$. Because $g(\boldsymbol{y}^*) = \|\boldsymbol{x}^* - \boldsymbol{y}^*\| > 0$ we can define

$$\boldsymbol{w}^* := \boldsymbol{y}^* + \frac{\varepsilon}{2g(\boldsymbol{y}^*)}(\boldsymbol{x}^* - \boldsymbol{y}^*)$$

and we have $\boldsymbol{w}^* \in B(\boldsymbol{y}^*; \varepsilon)$ since

$$\begin{aligned}\|\boldsymbol{w}^* - \boldsymbol{y}^*\| &= \Big\|\boldsymbol{y}^* + \frac{\varepsilon}{2g(\boldsymbol{y}^*)}(\boldsymbol{x}^* - \boldsymbol{y}^*) - \boldsymbol{y}^*\Big\| \\ &= \frac{\varepsilon}{2g(\boldsymbol{y}^*)}\|\boldsymbol{x}^* - \boldsymbol{y}^*\| = \frac{\varepsilon}{2}.\end{aligned}$$

Thus $\boldsymbol{w}^* \in S$ and, moreover

$$\begin{aligned}g(\boldsymbol{w}^*) &= \Big\|\boldsymbol{x}^* - \boldsymbol{y}^* - \frac{\varepsilon}{2g(\boldsymbol{y}^*)}(\boldsymbol{x}^* - \boldsymbol{y}^*)\Big\| \\ &= (1 - \frac{\varepsilon}{2g(\boldsymbol{y}^*)})g(\boldsymbol{y}^*) = g(\boldsymbol{y}^*) - \frac{\varepsilon}{2} < g(\boldsymbol{y}^*),\end{aligned}$$

which is impossible, since g attains its minimum over S at $\boldsymbol{y}^*$. Thus we have $\boldsymbol{y}^* \in \text{bd } S$.

In order to prove that (2.3) is a sufficient condition, let $\boldsymbol{x} \in S$. Then (2.3) implies

$$\begin{aligned}g(\boldsymbol{x})^2 &= \|\boldsymbol{x}^* - \boldsymbol{y}^* + \boldsymbol{y}^* - \boldsymbol{x}\|^2 \\ &= \|\boldsymbol{x}^* - \boldsymbol{y}^*\|^2 + \|\boldsymbol{y}^* - \boldsymbol{x}\|^2 + 2(\boldsymbol{x}^* - \boldsymbol{y}^*)^T(\boldsymbol{y}^* - \boldsymbol{x}) \\ &\geq \|\boldsymbol{x}^* - \boldsymbol{y}^*\|^2 \\ &= g(\boldsymbol{y}^*)^2,\end{aligned}$$

which means that $\boldsymbol{y}^*$ is the closest point.

On the other hand, if $\boldsymbol{y}^*$ is the closest point, we have

$$g(\boldsymbol{x}) \geq g(\boldsymbol{y}^*) \quad \text{for all } \boldsymbol{x} \in S.$$

Let $\boldsymbol{x} \in S$ be arbitrary. The convexity of S implies that

$$\boldsymbol{y}^* + \lambda(\boldsymbol{x} - \boldsymbol{y}^*) = \lambda\boldsymbol{x} + (1 - \lambda)\boldsymbol{y}^* \in S \quad \text{for all } \lambda \in [0, 1]$$

and thus

$$g\left(\boldsymbol{y}^* + \lambda(\boldsymbol{x} - \boldsymbol{y}^*)\right) \geq g(\boldsymbol{y}^*). \tag{2.4}$$

Furthermore, we have

$$g\left(\boldsymbol{y}^* + \lambda(\boldsymbol{x} - \boldsymbol{y}^*)\right)^2 = \|\boldsymbol{x}^* - \boldsymbol{y}^* - \lambda(\boldsymbol{x} - \boldsymbol{y}^*)\|^2$$
$$= g(\boldsymbol{y}^*)^2 + \lambda^2\|\boldsymbol{x} - \boldsymbol{y}^*\|^2 - 2\lambda(\boldsymbol{x}^* - \boldsymbol{y}^*)^T(\boldsymbol{x} - \boldsymbol{y}^*)$$

and combining this with (2.4) we get

$$2\lambda(\boldsymbol{x}^* - \boldsymbol{y}^*)^T(\boldsymbol{x} - \boldsymbol{y}^*) \leq \lambda^2\|\boldsymbol{x} - \boldsymbol{y}^*\|^2 \quad \text{for all } \lambda \in [0, 1]. \tag{2.5}$$

Dividing (2.5) by $\lambda > 0$ and letting $\lambda \downarrow 0$ we get (2.3). □

Next we define separating and supporting hyperplanes.

Definition 2.4 Let $S_1, S_2 \subset \mathbb{R}^n$ be nonempty sets. A hyperplane

$$H(\boldsymbol{p}, \alpha) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{p}^T(\boldsymbol{x} - \boldsymbol{x}_0) = 0\},$$

where $\boldsymbol{p} \neq \boldsymbol{0}$ and $\boldsymbol{p}^T\boldsymbol{x}_0 = \alpha$, *separates* S_1 and S_2 if $S_1 \subseteq H^+(\boldsymbol{p}, \alpha)$ and $S_2 \subseteq H^-(\boldsymbol{p}, \alpha)$, in other words

$$\boldsymbol{p}^T(\boldsymbol{x} - \boldsymbol{x}_0) \geq 0 \quad \text{for all } \boldsymbol{x} \in S_1 \quad \text{and}$$
$$\boldsymbol{p}^T(\boldsymbol{x} - \boldsymbol{x}_0) \leq 0 \quad \text{for all } \boldsymbol{x} \in S_2.$$

Moreover, the separation is *strict* if $S_1 \cap H(\boldsymbol{p}, \alpha) = \emptyset$ and $S_2 \cap H(\boldsymbol{p}, \alpha) = \emptyset$.

Example 2.3 (*Separation of convex sets*). Let $S_1 := \{\boldsymbol{x} \in \mathbb{R}^2 \mid \frac{1}{4}x_1^2 + x_2^2 \leq 1\}$ and $S_2 := \{\boldsymbol{x} \in \mathbb{R}^2 \mid (x_1 - 4)^2 + (x_2 - 2)^2 \leq 1\}$. Then the hyperplane $H((1, 1)^T, 3\frac{1}{2})$, in other words the line $x_2 = -x_1 + 3\frac{1}{2}$ separates S_1 and S_2 (see Fig. 2.3). Notice that $H((1, 1)^T, 3\frac{1}{2})$ is not unique but there exist infinitely many hyperplanes separating S_1 and S_2.

Definition 2.5 Let $S \subset \mathbb{R}^n$ be a nonempty set and $\boldsymbol{x}_0 \in \text{bd } S$. A hyperplane

$$H(\boldsymbol{p}, \alpha) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{p}^T(\boldsymbol{x} - \boldsymbol{x}_0) = 0\},$$

where $\boldsymbol{p} \neq \boldsymbol{0}$ and $\boldsymbol{p}^T\boldsymbol{x}_0 = \alpha$, *supports* S at $\boldsymbol{x}_0$ if either $S \subseteq H^+(\boldsymbol{p}, \alpha)$, in other words

$$\boldsymbol{p}^T(\boldsymbol{x} - \boldsymbol{x}_0) \geq 0 \quad \text{for all } \boldsymbol{x} \in S$$

or $S \subseteq H^-(\boldsymbol{p}, \alpha)$, in other words

$$\boldsymbol{p}^T(\boldsymbol{x} - \boldsymbol{x}_0) \leq 0 \quad \text{for all } \boldsymbol{x} \in S.$$

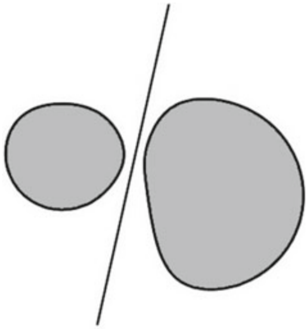


Fig. 2.3 Separation of convex sets

Example 2.4 (*Supporting hyperplanes*). Let $S := \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1\}$. Then the hyperplane $H((0, 1)^T, 1)$, in other words the line $x_2 = 1$ supports S at $\boldsymbol{x}_0 = (0, 1)^T$. Notice that $H((0, 1)^T, 1)$ is the unique supporting hyperplane of S at $\boldsymbol{x}_0 = (0, 1)^T$.

Theorem 2.4 *Let $S \subset \mathbb{R}^n$ be a nonempty, closed convex set and $\boldsymbol{x}^* \notin S$. Then there exists a hyperplane $H(\boldsymbol{p}, \alpha)$ supporting S at some $\boldsymbol{y}^* \in$ bd S and separating S and $\{\boldsymbol{x}^*\}$.*

Proof According to Lemma 2.2 there exists a unique $\boldsymbol{y}^* \in$ bd S minimizing the distance to $\boldsymbol{x}^*$. Let $\boldsymbol{p} := \boldsymbol{x}^* - \boldsymbol{y}^* \neq \boldsymbol{0}$ and $\alpha := \boldsymbol{p}^T \boldsymbol{y}^*$. Then due to (2.3) we have

$$\boldsymbol{p}^T(\boldsymbol{x} - \boldsymbol{y}^*) = (\boldsymbol{x}^* - \boldsymbol{y}^*)^T(\boldsymbol{x} - \boldsymbol{y}^*) \leq 0 \quad \text{for all } \boldsymbol{x} \in S, \tag{2.6}$$

in other words $S \subseteq H^-(\boldsymbol{p}, \alpha)$. This means that $H(\boldsymbol{p}, \alpha)$ supports S at $\boldsymbol{y}^*$. Moreover, we have

$$\boldsymbol{p}^T\boldsymbol{x}^* = \boldsymbol{p}^T(\boldsymbol{x}^* - \boldsymbol{y}^*) + \boldsymbol{p}^T\boldsymbol{y}^* = \|\boldsymbol{p}\|^2 + \alpha > \alpha \tag{2.7}$$

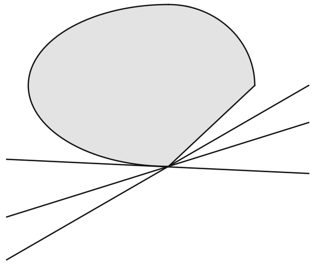
in other words $\{\boldsymbol{x}^*\} \subset H^+(\boldsymbol{p}, \alpha)$ and thus $H(\boldsymbol{p}, \alpha)$ separates S and $\{\boldsymbol{x}^*\}$. □

Next we prove a little bit stronger result, namely that there always exists a hyperplane strictly separating a point and a closed convex set.

Theorem 2.5 *Let $S \subset \mathbb{R}^n$ be a nonempty, closed convex set and $\boldsymbol{x}^* \notin S$. Then there exists a hyperplane $H(\boldsymbol{p}, \beta)$ strictly separating S and $\{\boldsymbol{x}^*\}$.*

Proof Using Lemma 2.2 we get a unique $\boldsymbol{y}^* \in$ bd S minimizing the distance to $\boldsymbol{x}^*$. As in the previous proof let $\boldsymbol{p} := \boldsymbol{x}^* - \boldsymbol{y}^* \neq \boldsymbol{0}$ but choose now $\beta := \boldsymbol{p}^T\boldsymbol{w}^*$, where $\boldsymbol{w}^* = \frac{1}{2}(\boldsymbol{x}^* + \boldsymbol{y}^*)$. Then due to (2.3) we have

Fig. 2.4 Supporting hyperplanes



$$\begin{aligned}\boldsymbol{p}^T(\boldsymbol{x} - \boldsymbol{w}^*) &= \boldsymbol{p}^T(\boldsymbol{x} - \boldsymbol{y}^* - \tfrac{1}{2}\boldsymbol{p}) \\ &= (\boldsymbol{x}^* - \boldsymbol{y}^*)^T(\boldsymbol{x} - \boldsymbol{y}^*) - \tfrac{1}{2}\boldsymbol{p}^T\boldsymbol{p} \\ &\le -\tfrac{1}{2}\|\boldsymbol{p}\|^2 < 0 \quad \text{for all } \boldsymbol{x} \in S,\end{aligned}$$

in other words $S \subset H^-(\boldsymbol{p}, \beta)$ and $S \cap H(\boldsymbol{p}, \beta) = \emptyset$. Moreover, we have

$$\begin{aligned}\boldsymbol{p}^T(\boldsymbol{x}^* - \boldsymbol{w}^*) &= \boldsymbol{p}^T(\boldsymbol{x}^* - \tfrac{1}{2}\boldsymbol{x}^* - \tfrac{1}{2}\boldsymbol{y}^*) \\ &= \tfrac{1}{2}\boldsymbol{p}^T(\boldsymbol{x}^* - \boldsymbol{y}^*) \\ &= \tfrac{1}{2}\|\boldsymbol{p}\|^2 > 0,\end{aligned}$$

which means that $\{\boldsymbol{x}^*\} \subset H^+(\boldsymbol{p}, \beta)$ and $\{\boldsymbol{x}^*\} \cap H(\boldsymbol{p}, \beta) = \emptyset$. Thus $H(\boldsymbol{p}, \beta)$ strictly separates S and $\{\boldsymbol{x}^*\}$. □

Replacing S by cl conv S in Theorem 2.5 we obtain the following result.

Corollary 2.1 *Let $S \subset \mathbb{R}^n$ be a nonempty set and $\boldsymbol{x}^* \notin$ cl conv S. Then there exists a hyperplane $H(\boldsymbol{p}, \beta)$ strictly separating S and $\{\boldsymbol{x}^*\}$.*

The next theorem is very similar to Theorem 2.3 showing that the closure of convex hull is actually the intersection of all the closed halfspaces containing the set.

Theorem 2.6 *If $S \subset \mathbb{R}^n$, then*

$$\text{cl conv } S = \bigcap_{\substack{S \subseteq H^-(\boldsymbol{p}, \alpha) \\ \boldsymbol{p} \neq \boldsymbol{0},\, \alpha \in \mathbb{R}}} H^-(\boldsymbol{p}, \alpha).$$

Proof Due to Theorem 2.3 we have

$$\text{conv } S = \bigcap_{\substack{S \subseteq \hat{S} \\ \hat{S} \text{ convex}}} \hat{S} \subseteq \bigcap_{\substack{S \subseteq H^-(\boldsymbol{p}, \alpha) \\ \boldsymbol{p} \neq \boldsymbol{0},\, \alpha \in \mathbb{R}}} H^-(\boldsymbol{p}, \alpha) =: T.$$

Since T is closed as an intersection of closed sets, we have

$$\text{cl conv } S \subseteq \text{cl } T = T.$$

Next we show that also $T \subseteq \text{cl conv } S$. To the contrary suppose that there exists $\boldsymbol{x}^* \in T$ but $\boldsymbol{x}^* \notin \text{cl conv } S$. Then due to Corollary 2.1 there exists a closed halfspace $H^-(\boldsymbol{p}, \beta)$ such that $S \subseteq H^-(\boldsymbol{p}, \beta)$ and $\boldsymbol{x}^* \notin H^-(\boldsymbol{p}, \beta)$, thus $\boldsymbol{x}^* \notin T \subseteq H^-(\boldsymbol{p}, \beta)$, which is a contradiction and the proof is complete. □

We can also strengthen the supporting property of Theorem 2.4, namely there exists actually a supporting hyperplane at every boundary point.

Theorem 2.7 *Let $S \subset \mathbb{R}^n$ be a nonempty convex set and $\boldsymbol{x}_0 \in \text{bd } S$. Then there exists a hyperplane $H(\boldsymbol{p}, \alpha)$ supporting* cl S *at $\boldsymbol{x}_0$.*

Proof Since $\boldsymbol{x}_0 \in \text{bd } S$ there exists a sequence $(\boldsymbol{x}_k)$ such that $\boldsymbol{x}_k \notin \text{cl } S$ and $\boldsymbol{x}_k \to \boldsymbol{x}_0$. Then due to Theorem 2.4 for each $\boldsymbol{x}_k$ there exists $\boldsymbol{y}_k \in \text{bd } S$ such that the hyperplane $H(\boldsymbol{q}_k, \beta_k)$, where $\boldsymbol{q}_k := \boldsymbol{x}_k - \boldsymbol{y}_k$ and $\beta_k := \boldsymbol{q}_k^T \boldsymbol{y}_k$ supports cl S at $\boldsymbol{y}_k$. Then inequality (2.6) implies that

$$0 \geq \boldsymbol{q}_k^T(\boldsymbol{x} - \boldsymbol{y}_k) = \boldsymbol{q}_k^T \boldsymbol{x} - \beta_k \quad \text{for all } \boldsymbol{x} \in \text{cl } S,$$

and thus

$$\boldsymbol{q}_k^T \boldsymbol{x} \leq \beta_k \quad \text{for all } \boldsymbol{x} \in \text{cl } S.$$

On the other hand, according to (2.7) we get $\boldsymbol{q}_k^T \boldsymbol{x}_k > \beta_k$, thus we have

$$\boldsymbol{q}_k^T \boldsymbol{x} < \boldsymbol{q}_k^T \boldsymbol{x}_k \quad \text{for all } \boldsymbol{x} \in \text{cl } S. \tag{2.8}$$

Next we normalize vectors $\boldsymbol{q}_k$ by defining $\boldsymbol{p}_k := \boldsymbol{q}_k / \|\boldsymbol{q}_k\|$. Then $\|\boldsymbol{p}_k\| = 1$, which means that the sequence $(\boldsymbol{p}_k)$ is bounded having a convergent subsequence $(\boldsymbol{p}_{k_j})$, in other words there exists a limit $\boldsymbol{p} \in \mathbb{R}^n$ such that $\boldsymbol{p}_{k_j} \to \boldsymbol{p}$ and $\|\boldsymbol{p}\| = 1$. It is easy to verify, that (2.8) holds also for $\boldsymbol{p}_{k_j}$, in other words

$$\boldsymbol{p}_{k_j}^T \boldsymbol{x} < \boldsymbol{p}_{k_j}^T \boldsymbol{x}_{k_j} \quad \text{for all } \boldsymbol{x} \in \text{cl } S. \tag{2.9}$$

Fixing now $\boldsymbol{x} \in \text{cl } S$ in (2.9) and letting $j \to \infty$ we get $\boldsymbol{p}^T \boldsymbol{x} \leq \boldsymbol{p}^T \boldsymbol{x}_0$. In other words

$$\boldsymbol{p}^T(\boldsymbol{x} - \boldsymbol{x}_0) \leq 0,$$

which means that cl $S \subseteq H^-(\boldsymbol{p}, \alpha)$, where $\alpha := \boldsymbol{p}^T \boldsymbol{x}_0$ and thus $H(\boldsymbol{p}, \alpha)$ supports cl S at $\boldsymbol{x}_0$. □

Finally we consider a nice property of convex sets, namely two disjoint convex sets can always be separated by a hyperplane. For strict separation it is not enough to suppose the closedness of the sets, but at least one of the sets should be bounded as well.

Theorem 2.8 *Let $S_1, S_2 \subset \mathbb{R}^n$ be nonempty convex sets. If $S_1 \cap S_2 = \emptyset$, then there exists a hyperplane $H(\boldsymbol{p}, \alpha)$ separating S_1 and S_2. If, in addition, S_1 and S_2 are closed and S_1 is bounded, then the separation is strict.*

Proof It follows from Theorem 2.2, that the set

$$S := S_1 - S_2 = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = \boldsymbol{x}_1 - \boldsymbol{x}_2,\ \boldsymbol{x}_1 \in S_1, \boldsymbol{x}_2 \in S_2\}$$

is convex. Furthermore, $\mathbf{0} \notin S$, since otherwise there would exist $\boldsymbol{x}_1 \in S_1$ and $\boldsymbol{x}_2 \in S_2$ such that $\mathbf{0} = \boldsymbol{x}_1 - \boldsymbol{x}_2$, in other words $\boldsymbol{x}_1 = \boldsymbol{x}_2 \in S_1 \cap S_2 = \emptyset$, which is impossible.

If $\mathbf{0} \notin \operatorname{cl} S$, then due to Corollary 2.1 there exists a hyperplane $H(\boldsymbol{p}, \alpha)$ strictly separating S and $\{\mathbf{0}\}$, in other words

$$\boldsymbol{p}^T\boldsymbol{x} < \alpha < \boldsymbol{p}^T\mathbf{0} = 0 \quad \text{for all } \boldsymbol{x} \in S.$$

Since $\boldsymbol{x} = \boldsymbol{x}_1 - \boldsymbol{x}_2$, where $\boldsymbol{x}_1 \in S_1$ and $\boldsymbol{x}_2 \in S_2$, we get

$$\boldsymbol{p}^T\boldsymbol{x}_1 < \alpha < \boldsymbol{p}^T\boldsymbol{x}_2 \quad \text{for all } \boldsymbol{x}_1 \in S_1, \boldsymbol{x}_2 \in S_2,$$

and thus $H(\boldsymbol{p}, \alpha)$ strictly separates S_1 and S_2.

On the other hand, if $\mathbf{0} \in \operatorname{cl} S$ it must hold that $\mathbf{0} \in \operatorname{bd} S$ (since $\mathbf{0} \notin \operatorname{int} S$). Then due to Theorem 2.7 there exists a hyperplane $H(\boldsymbol{p}, \beta)$ supporting $\operatorname{cl} S$ at $\mathbf{0}$, in other words

$$\boldsymbol{p}^T(\boldsymbol{x} - \mathbf{0}) \leq 0 \quad \text{for all } \boldsymbol{x} \in \operatorname{cl} S.$$

Denoting again $\boldsymbol{x} = \boldsymbol{x}_1 - \boldsymbol{x}_2$, where $\boldsymbol{x}_1 \in S_1$ and $\boldsymbol{x}_2 \in S_2$, we get

$$\boldsymbol{p}^T\boldsymbol{x}_1 \leq \boldsymbol{p}^T\boldsymbol{x}_2 \quad \text{for all } \boldsymbol{x}_1 \in S_1, \boldsymbol{x}_2 \in S_2.$$

Since the set of real numbers $\{\boldsymbol{p}^T\boldsymbol{x}_1 \mid \boldsymbol{x}_1 \in S_1\}$ is bounded above by some number $\boldsymbol{p}^T\boldsymbol{x}_2$, where $\boldsymbol{x}_2 \in S_2 \neq \emptyset$ it has a finite supremum. Defining $\alpha := \sup\{\boldsymbol{p}^T\boldsymbol{x}_1 \mid \boldsymbol{x}_1 \in S_1\}$ we get

$$\boldsymbol{p}^T\boldsymbol{x}_1 \leq \alpha \leq \boldsymbol{p}^T\boldsymbol{x}_2 \quad \text{for all } \boldsymbol{x}_1 \in S_1, \boldsymbol{x}_2 \in S_2,$$

and thus $H(\boldsymbol{p}, \alpha)$ separates S_1 and S_2.

Suppose next, that S_1 and S_2 are closed and S_1 is bounded. In order to show that S is closed suppose, that there exists a sequence $(\boldsymbol{x}_k) \subset S$ and a limit $\boldsymbol{x} \in \mathbb{R}^n$ such that $\boldsymbol{x}_k \to \boldsymbol{x}$. Then due to the definition of S we have $\boldsymbol{x}_k = \boldsymbol{x}_{1_k} - \boldsymbol{x}_{2_k}$, where $\boldsymbol{x}_{1_k} \in S_1$ and $\boldsymbol{x}_{2_k} \in S_2$. Since S_1 is compact, there exists a convergent subsequence $(\boldsymbol{x}_{1_{k_j}})$ and a limit $\boldsymbol{x}_1 \in S_1$ such that $\boldsymbol{x}_{1_{k_j}} \to \boldsymbol{x}_1$. Then we have

$\boldsymbol{x}_{2_{kj}} = \boldsymbol{x}_{1_{kj}} - \boldsymbol{x}_{k_j} \to \boldsymbol{x}_1 - \boldsymbol{x} := \boldsymbol{x}_2$. Since S_2 is closed $\boldsymbol{x}_2 \in S_2$. Thus $\boldsymbol{x} \in S$ and S is closed. Now the case $\mathbf{0} \notin \text{cl } S = S$ given above is the only possibility and thus we can find $H(\boldsymbol{p}, \alpha)$ strictly separating S_1 and S_2. $\square$

The next two examples show that both closedness and compactness assumptions actually are essential for strict separation.

Example 2.5 (*Strict separation, counter example 1*). Let $S_1 := \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1 > 0 \text{ and } x_2 \geq 1/x_1\}$ and $S_2 := \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_2 = 0\}$. Then both S_1 and S_2 are closed but neither of them is bounded. It follows that $S_1 - S_2 = \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_2 > 0\}$ is not closed and there does not exist any strictly separating hyperplane.

Example 2.6 (*Strict separation, counter example 2*). Let $S_1 := \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1\}$ and $S_2 := \{\boldsymbol{x} \in \mathbb{R}^2 \mid (x_1 - 2)^2 + x_2^2 < 1\}$. Then both S_1 and S_2 are bounded but S_2 is not closed and it follows again that $S_1 - S_2 = \{\boldsymbol{x} \in \mathbb{R}^2 \mid (x_1 + 2)^2 + x_2^2 < 4\}$ is not closed and there does not exist any strictly separating hyperplane.

2.1.3 Convex Cones

Next we define the notion of a cone, which is a set containing all the rays passing through its points emanating from the origin.

Definition 2.6 A set $C \subseteq \mathbb{R}^n$ is a *cone* if $\lambda \boldsymbol{x} \in C$ for all $\boldsymbol{x} \in C$ and $\lambda \geq 0$. Moreover, if C is convex, then it is called a *convex cone*.

Example 2.7 (*Convex cones*). It is easy to show that a singleton $\{\mathbf{0}\}$, the whole space $\mathbb{R}^n$, closed halfspaces $H^+(\boldsymbol{p}, 0)$ and $H^-(\boldsymbol{p}, 0)$, the nonnegative orthant $\mathbb{R}^n_+ = \{\boldsymbol{x} \in \mathbb{R}^n \mid x_i \geq 0,\ i = 1 \ldots, n\}$ and halflines starting from the origin are examples of closed convex cones.

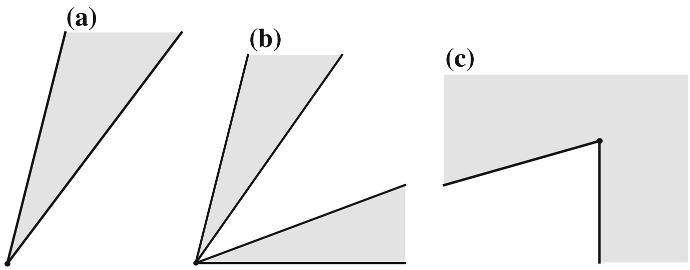


Fig. 2.5 Illustration of convex and nonconvex cones. (**a**) Convex. (**b**) Not convex. (**c**) Not convex

Theorem 2.9 *A set $C \subseteq \mathbb{R}^n$ is a convex cone if and only if*

$$\lambda \boldsymbol{x} + \mu \boldsymbol{y} \in C \quad for\ all\ \boldsymbol{x}, \boldsymbol{y} \in C\ and\ \lambda, \mu \geq 0. \tag{2.10}$$

Proof Evidently (2.10) implies that C is a convex cone.

Next, let C be a convex cone and suppose that $\boldsymbol{x}, \boldsymbol{y} \in C$ and $\lambda, \mu \geq 0$. Since C is a cone we have $\lambda \boldsymbol{x} \in C$ and $\mu \boldsymbol{y} \in C$. Furthermore, since C is convex we have

$$\tfrac{1}{2}\lambda \boldsymbol{x} + (1 - \tfrac{1}{2})\mu \boldsymbol{y} \in C \tag{2.11}$$

and again using the cone property we get

$$\lambda \boldsymbol{x} + \mu \boldsymbol{y} = 2\left(\tfrac{1}{2}\lambda \boldsymbol{x} + (1 - \tfrac{1}{2})\mu \boldsymbol{y}\right) \in C \tag{2.12}$$

and the proof is complete. □

Via the next definition we get a connection between sets and cones. Namely a set generates a cone, when every point of the set is replaced by a ray emanating from the origin.

Definition 2.7 The *ray* of a set $S \subseteq \mathbb{R}^n$ is

$$\text{ray } S = \bigcup_{\lambda \geq 0} \lambda S = \{\lambda \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} \in S,\ \lambda \geq 0\}.$$

The proof of the next lemma is left as an exercise.

Lemma 2.3 *If $S \subseteq \mathbb{R}^n$, then* ray S *is a cone and $C \subseteq \mathbb{R}^n$ is cone if and only if*

$$C = \text{ray } C.$$

Proof Exercise. □

The next theorem shows that the ray of a set is actually the intersection of all the cones containing S, in other words, it is the smallest cone containing S (see Fig. 2.6).

Theorem 2.10 *If $S \subset \mathbb{R}^n$, then*

$$\mathrm{ray}\ S = \bigcap_{\substack{S \subseteq C \\ C \text{ cone}}} C.$$

Proof Let C be a cone such that $S \subseteq C$. Then due to Lemma 2.3 we have ray $S \subseteq$ ray $C = C$ and thus we have

$$\mathrm{ray}\ S \subseteq \bigcap_{\substack{S \subseteq C \\ C \text{ cone}}} C.$$

On the other hand, it is evident that $S \subseteq \mathrm{conv}\ S$ and due to Lemma 2.3 ray S is a cone. Then ray S is one of the cones C forming the intersection and thus

$$\bigcap_{\substack{S \subseteq C \\ C \text{ cone}}} C = \bigcap_{\substack{S \subseteq C \\ C \text{ cone}}} C \cap \mathrm{ray}\ S \subseteq \mathrm{ray}\ S$$

and the proof is complete. □

It can be seen from Fig. 2.6 that a ray is not necessarily convex. However, if the set is convex, then also its ray is convex.

Theorem 2.11 *If $S \subseteq \mathbb{R}^n$ is convex, then* ray S *is a convex cone.*

Proof Due to Lemma 2.3 ray S is a cone. For convexity let $\boldsymbol{x}, \boldsymbol{y} \in$ ray S and $\lambda, \mu \geq 0$. Then $\boldsymbol{x} = \alpha \boldsymbol{u}$ and $\boldsymbol{y} = \beta \boldsymbol{v}$, where $\boldsymbol{u}, \boldsymbol{v} \in S$ and $\alpha, \beta \geq 0$. Since S is convex we have

$$\boldsymbol{z} := \frac{\lambda\alpha}{\lambda\alpha + \mu\beta}\boldsymbol{u} + \left(1 - \frac{\lambda\alpha}{\lambda\alpha + \mu\beta}\right)\boldsymbol{v} \in S.$$

The fact that ray S is cone implies that $(\lambda\alpha + \mu\beta)\boldsymbol{z} \in$ ray S, in other words

$$(\lambda\alpha + \mu\beta)\boldsymbol{z} = \lambda\alpha\boldsymbol{u} + \mu\beta\boldsymbol{v} = \lambda\boldsymbol{x} + \mu\boldsymbol{y} \in \mathrm{ray}\ S.$$

According to Theorem 2.9 this means, that ray S is convex. □

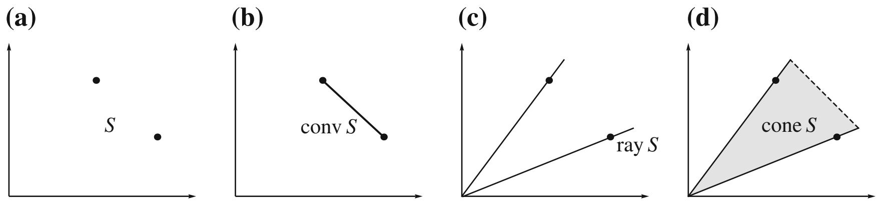


Fig. 2.6 Convex hull, ray and conic hull of a set. (**a**) Set. (**b**) Convex hull. (**c**) Ray. (**d**) Conic hull

It is also easy to show that a ray is not necessarily closed. However, if the set is compact not including the origin its ray is closed.

Theorem 2.12 *If* $S \subset \mathbb{R}^n$ *is compact such that* $\mathbf{0} \notin S$*, then* ray S *is closed.*

Proof Let $(\boldsymbol{x}_j) \subset \text{ray } S$ be a sequence such that $\boldsymbol{x}_j \to \boldsymbol{x}$. Next we show that $\boldsymbol{x} \in \text{ray } S$. The fact that $\boldsymbol{x}_j \in \text{ray } S$ means that $\boldsymbol{x}_j = \lambda_j \boldsymbol{y}_j$ where $\lambda_j \geq 0$ and $\boldsymbol{y}_j \in S$ for all $j \in \mathbb{N}$. Since S is compact the sequence $\boldsymbol{y}_j$ is bounded, thus there exists a subsequence $(\boldsymbol{y}_{j_i}) \subset S$ such that $\boldsymbol{y}_{j_i} \to \boldsymbol{y}$. Because S is closed, it follows that $\boldsymbol{y} \in S$. Furthermore, since $\mathbf{0} \notin S$ one has $\boldsymbol{y} \neq \mathbf{0}$, thus the sequence λ_{j_i} is also converging to some $\lambda \geq 0$. Then $\lambda_{j_i} \boldsymbol{y}_{j_i} \to \lambda \boldsymbol{y} = \boldsymbol{x}$, which means that $\boldsymbol{x} \in \text{ray } S$, in other words S is closed. □

Similarly to the convex combination we say that the linear combination $\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i$ is a *conic combination* of elements $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}^n$ if each $\lambda_i \geq 0$ and the conic hull generated by a set is defined as a set of conic combinations as follows.

Definition 2.8 The *conic hull* of a set $S \subseteq \mathbb{R}^n$ is

$$\text{cone } S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i,\ \boldsymbol{x}_i \in S,\ \lambda_i \geq 0, k > 0\}.$$

The proof of the next lemma is again left as an exercise.

Lemma 2.4 *If* $S \subseteq \mathbb{R}^n$*, then* cone S *is a convex cone and* $C \subseteq \mathbb{R}^n$ *is convex cone if and only if*

$$C = \text{cone } C.$$

Proof Exercise. □

The next theorem shows that the conic hull cone S is actually the intersection of all the convex cones containing S, in other words, it is the smallest convex cone containing S (see Fig. 2.6).

Theorem 2.13 *If* $S \subset \mathbb{R}^n$*, then*

$$\text{cone } S = \bigcap_{\substack{S \subseteq C \\ C \text{ convex cone}}} C.$$

Proof Let C be a convex cone such that $S \subseteq C$. Then due to Lemma 2.4 we have cone $S \subseteq$ cone $C = C$ and thus we have

$$\text{cone } S \subseteq \bigcap_{\substack{S \subseteq C \\ C \text{ convex cone}}} C.$$

On the other hand, it is evident that $S \subseteq \text{cone } S$ and due to Lemma 2.4 cone S is a convex cone. Then cone S is one of the convex cones forming the intersection and thus

$$\bigcap_{\substack{S \subseteq C \\ C \text{ convex cone}}} C = \bigcap_{\substack{S \subseteq C \\ C \text{ convex cone}}} C \cap \text{cone } S \subseteq \text{cone } S$$

and the proof is complete. □

Note, that according to Lemma 2.1, Theorems 2.10 and 2.13, and Definitions 2.7 and 2.8 we get the following result.

Corollary 2.2 *If $S \subseteq \mathbb{R}^n$, then*

$$\text{cone } S = \text{conv ray } S.$$

Finally we get another connection between sets and cones. Namely, every set generates also so called polar cone.

Definition 2.9 The *polar cone* of a nonempty set $S \subseteq \mathbb{R}^n$ is

$$S^\circ = \{\boldsymbol{y} \in \mathbb{R}^n \mid \boldsymbol{y}^T \boldsymbol{x} \leq 0 \text{ for all } \boldsymbol{x} \in S\}.$$

The polar cone $\emptyset^0$ of the empty set $\emptyset$ is the whole space $\mathbb{R}^n$.

The next lemma gives some basic properties of polar cones (see Fig. 2.7). The proof is left as an exercise.

Lemma 2.5 *If $S \subseteq \mathbb{R}^n$, then S° is a closed convex cone and $S \subseteq S^{\circ\circ}$.*

Proof Exercise. □

Theorem 2.14 *The set $C \subseteq \mathbb{R}^n$ is a closed convex cone if and only if*

$$C = C^{\circ\circ}.$$

Proof Suppose first that $C = C^{\circ\circ} = (C^\circ)^\circ$. Then due to Lemma 2.5 C is a closed convex cone.

Suppose next, that C is a closed convex cone. Lemma 2.5 implies that $C \subseteq C^{\circ\circ}$. We shall prove next that $C^{\circ\circ} \subseteq C$. Clearly $\emptyset^{\circ\circ} = (\mathbb{R}^n)^\circ = \emptyset$ and thus we can assume that C is nonempty. Suppose, by contradiction, that there exists $\boldsymbol{x} \in C^{\circ\circ}$ such that $\boldsymbol{x} \notin C$. Then due to Theorem 2.4 there exists a hyperplane $H(\boldsymbol{p}, \alpha)$ separating C and $\{\boldsymbol{x}\}$, in other words there exist $\boldsymbol{p} \neq \boldsymbol{0}$ and $\alpha \in \mathbb{R}$ such that

$$\boldsymbol{p}^T \boldsymbol{y} \leq \alpha \text{ for all } \boldsymbol{y} \in C \quad \text{and} \quad \boldsymbol{p}^T \boldsymbol{x} > \alpha.$$

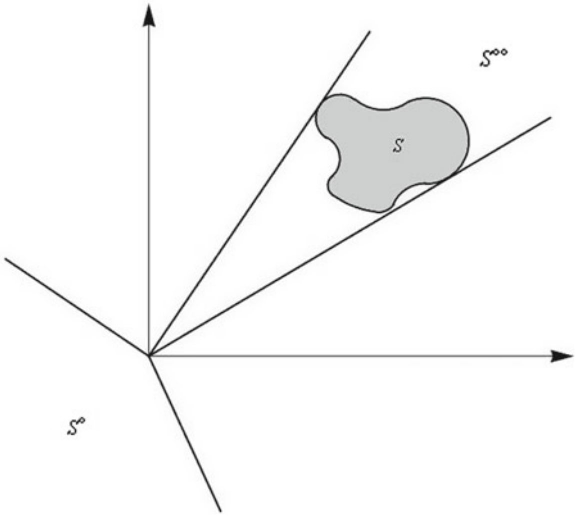


Fig. 2.7 Polar cones of the set

Since $\mathbf{0} \in C$ we have $\alpha \geq \boldsymbol{p}^T\mathbf{0} = 0$ and thus

$$\boldsymbol{p}^T\boldsymbol{x} > 0. \tag{2.13}$$

If $\boldsymbol{p} \notin C^{\circ}$ then due to the definition of the polar cone there exists $\boldsymbol{z} \in C$ such that $\boldsymbol{p}^T\boldsymbol{z} > 0$. Since C is cone we have $\lambda\boldsymbol{z} \in C$ for all $\lambda \geq 0$. Then $\boldsymbol{p}^T(\lambda\boldsymbol{z}) > 0$ can grow arbitrary large when $\lambda \to \infty$, which contradicts the fact that $\boldsymbol{p}^T\boldsymbol{y} \leq \alpha$ for all $\boldsymbol{y} \in C$. Therefore we have $\boldsymbol{p} \in C^{\circ}$. On the other hand

$$\boldsymbol{x} \in C^{\circ\circ} = \{\boldsymbol{y} \in \mathbb{R}^n \mid \boldsymbol{y}^T\boldsymbol{v} \leq 0 \text{ for all } \boldsymbol{v} \in C^{\circ}\}$$

and thus $\boldsymbol{p}^T\boldsymbol{x} \leq 0$, which contradicts (2.13). We conclude that $\boldsymbol{x} \in C$ and the proof is complete. □

2.1.4 Contingent and Normal Cones

In this subsection we consider tangents and normals of convex sets. First we define a classical notion of contingent cone consisting of the tangent vectors (see Fig. 2.8).

Definition 2.10 The *contingent cone* of the nonempty set S at $\boldsymbol{x} \in S$ is given by the formula

$$K_S(\boldsymbol{x}) := \{\boldsymbol{d} \in \mathbb{R}^n \mid \text{there exist } t_i \downarrow 0 \text{ and } \boldsymbol{d}_i \to \boldsymbol{d} \text{ such that } \boldsymbol{x} + t_i \boldsymbol{d}_i \in S\}. \tag{2.14}$$

The elements of $K_S(\boldsymbol{x})$ are called *tangent vectors*.

Several elementary facts about the contingent cone will now be listed.

Theorem 2.15 *The contingent cone $K_S(\boldsymbol{x})$ of the nonempty convex set S at $\boldsymbol{x} \in S$ is a closed convex cone.*

Proof We begin by proving that $K_S(\boldsymbol{x})$ is closed. To see this, let $(\boldsymbol{d}_i)$ be a sequence in $K_S(\boldsymbol{x})$ converging to $\boldsymbol{d} \in \mathbb{R}^n$. Next we show that $\boldsymbol{d} \in K_S(\boldsymbol{x})$. The fact that $\boldsymbol{d}_i \to \boldsymbol{d}$ implies that for all $\varepsilon > 0$ there exists $i_0 \in \mathbb{N}$ such that

$$\|\boldsymbol{d} - \boldsymbol{d}_i\| < \varepsilon/2 \qquad \text{for all} \quad i \geq i_0.$$

On the other hand, $\boldsymbol{d}_i \in K_S(\boldsymbol{x})$, thus for each $i \in \mathbb{N}$ there exist sequences $(\boldsymbol{d}_{i_j}) \subset \mathbb{R}^n$ and $(t_{i_j}) \subset \mathbb{R}$ such that $\boldsymbol{d}_{i_j} \to \boldsymbol{d}_i$, $t_{i_j} \downarrow 0$ and $\boldsymbol{x} + t_{i_j}\boldsymbol{d}_{i_j} \in S$ for all $j \in \mathbb{N}$. Then there exist $j_i^{\boldsymbol{y}} \in \mathbb{N}$ and $j_i^t \in \mathbb{N}$ such that for all $i \in \mathbb{N}$

$$\|\boldsymbol{d}_i - \boldsymbol{d}_{i_j}\| < \varepsilon/2 \qquad \text{for all} \quad j \geq j_i^{\boldsymbol{y}}$$

and

$$|t_{i_j}| < 1/i \qquad \text{for all} \quad j \geq j_i^t.$$

Let us choose $j_i := \max\{j_i^{\boldsymbol{y}}, j_i^t\}$. Then $t_{i_{j_i}} \downarrow 0$ and for all $i \geq i_0$

$$\|\boldsymbol{d} - \boldsymbol{d}_{i_{j_i}}\| \leq \|\boldsymbol{d} - \boldsymbol{d}_i\| + \|\boldsymbol{d}_i - \boldsymbol{d}_{i_{j_i}}\| \leq \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

which implies that $\boldsymbol{d}_{i_{j_i}} \to \boldsymbol{d}$ and, moreover, $\boldsymbol{x} + t_{i_{j_i}}\boldsymbol{d}_{i_{j_i}} \in S$. By the definition of the contingent cone, this means that $\boldsymbol{d} \in K_S(\boldsymbol{x})$ and thus $K_S(\boldsymbol{x})$ is closed.

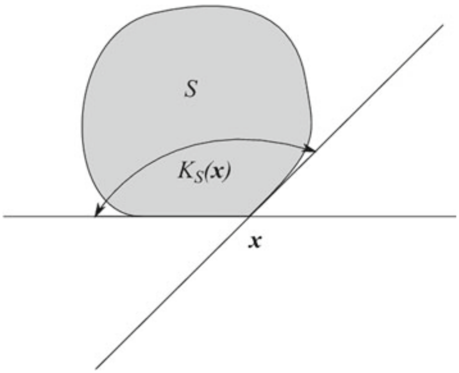


Fig. 2.8 Contingent cone $K_S(\boldsymbol{x})$ of a convex set

We continue by proving that $K_S(\boldsymbol{x})$ is a cone. If $\boldsymbol{d} \in K_S(\boldsymbol{x})$ is arbitrary then there exist sequences $(\boldsymbol{d}_j) \subset \mathbb{R}^n$ and $(t_j) \subset \mathbb{R}$ such that $\boldsymbol{d}_j \to \boldsymbol{d}$, $t_j \downarrow 0$ and $\boldsymbol{x}+t_j\boldsymbol{d}_j \in S$ for all $j \in \mathbb{N}$. Let $\lambda > 0$ be fixed and define $\boldsymbol{d}'_j := \lambda\boldsymbol{d}_j$ and $t'_j := t_j/\lambda$. Since $t'_j \downarrow 0$,

$$\|\boldsymbol{d}'_j - \lambda\boldsymbol{d}\| = \lambda\|\boldsymbol{d}_j - \boldsymbol{y}\| \longrightarrow 0 \quad \text{whenever} \quad j \to \infty$$

and

$$\boldsymbol{x} + t'_j\boldsymbol{d}'_j = \boldsymbol{x} + \frac{t_j}{\lambda} \cdot \lambda\boldsymbol{d}_j \in S$$

it follows that $\lambda\boldsymbol{d} \in K_S(\boldsymbol{x})$. Thus $K_S(\boldsymbol{x})$ is a cone.

For convexity let $\lambda \in [0, 1]$ and $\boldsymbol{d}^1, \boldsymbol{d}^2 \in K_S(\boldsymbol{x})$. We need to show that $\boldsymbol{d} := (1-\lambda)\boldsymbol{d}^1 + \lambda\boldsymbol{d}^2$ belongs to $K_S(\boldsymbol{x})$. By the definition of $K_S(\boldsymbol{x})$ there exist sequences $(\boldsymbol{d}^1_j), (\boldsymbol{d}^2_j) \subset \mathbb{R}^n$ and $(t^1_j), (t^2_j) \subset \mathbb{R}$ such that $\boldsymbol{d}^i_j \to \boldsymbol{d}^i$, $t^i_j \downarrow 0$ and $\boldsymbol{x} + t^i_j\boldsymbol{d}^i_j \in S$ for all $j \in \mathbb{N}$ and $i = 1, 2$. Define

$$\boldsymbol{d}_j := (1-\lambda)\boldsymbol{d}^1_j + \lambda\boldsymbol{d}^2_j \qquad \text{and} \qquad t_j := \min\{t^1_j, t^2_j\}.$$

Then we have

$$\boldsymbol{x} + t_j\boldsymbol{d}_j = (1-\lambda)(\boldsymbol{x} + t_j\boldsymbol{d}^1_j) + \lambda(\boldsymbol{x} + t_j\boldsymbol{d}^2_j) \in S$$

because S is convex and

$$\boldsymbol{x} + t_j\boldsymbol{d}^i_j = (1 - \frac{t_j}{t^i_j})\boldsymbol{x} + \frac{t_j}{t^i_j}(\boldsymbol{x} + t^i_j\boldsymbol{d}^i_j) \in S$$

because $\frac{t_j}{t^i_j} \in [0, 1]$ and S is convex. Moreover, we have

$$\begin{aligned}\|\boldsymbol{d}_j - \boldsymbol{d}\| &= \|(1-\lambda)\boldsymbol{d}^1_j + \lambda\boldsymbol{d}^2_j - (1-\lambda)\boldsymbol{d}^1 - \lambda\boldsymbol{d}^2\| \\ &\le (1-\lambda)\|\boldsymbol{d}^1_j - \boldsymbol{d}^1\| + \lambda\|\boldsymbol{d}^2_j - \boldsymbol{d}^2\| \longrightarrow 0,\end{aligned}$$

when $j \to \infty$, in other words $\boldsymbol{d}_j \to \boldsymbol{d}$. Since $t_j \downarrow 0$ we have $\boldsymbol{d} \in K_S(\boldsymbol{x})$ and thus $K_S(\boldsymbol{x})$ is convex. □

The following cone of feasible directions is very useful in optimization when seeking for feasible search directions.

Definition 2.11 The *cone of globally feasible directions* of the nonempty set S at $\boldsymbol{x} \in S$ is given by the formula

$$G_S(\boldsymbol{x}) := \{\boldsymbol{d} \in \mathbb{R}^n \mid \text{there exists } t > 0 \text{ such that } \boldsymbol{x} + t\boldsymbol{d} \in S\}.$$

The cone of globally feasible directions has the same properties as the contingent cone but it is not necessarily closed. The proof of the next theorem is very similar to that of Theorem 2.15 and it is left as an exercise.

Theorem 2.16 *The cone of globally feasible directions $G_S(\boldsymbol{x})$ of the nonempty convex set S at $\boldsymbol{x} \in S$ is a convex cone.*

Proof Exercise. □

We have the following connection between the contingent cone and the cone of feasible directions.

Theorem 2.17 *If S is a nonempty set and $\boldsymbol{x} \in S$, then*

$$K_S(\boldsymbol{x}) \subseteq \text{cl } G_S(\boldsymbol{x}).$$

If, in addition, S is convex then

$$K_S(\boldsymbol{x}) = \text{cl } G_S(\boldsymbol{x}).$$

Proof If $\boldsymbol{d} \in K_S(\boldsymbol{x})$ is arbitrary, then there exist sequences $\boldsymbol{d}_j \to \boldsymbol{d}$ and $t_j \downarrow 0$ such that $\boldsymbol{x} + t_j \boldsymbol{d}_j \in S$ for all $j \in \mathbb{N}$, thus $\boldsymbol{d} \in \text{cl } G_S(\boldsymbol{x})$.

To see the equality, let S be convex and $\boldsymbol{d} \in \text{cl } G_S(\boldsymbol{x})$. Then there exist sequences $\boldsymbol{d}_j \to \boldsymbol{d}$ and $t_j > 0$ such that $\boldsymbol{x} + t_j \boldsymbol{d}_j \in S$ for all $j \in \mathbb{N}$. It suffices now to find a sequence t'_j such that $t'_j \downarrow 0$ and $\boldsymbol{x} + t'_j \boldsymbol{d}_j \in S$. Choose $t'_j := \min\{\frac{1}{j}, t_j\}$, which implies that

$$|t'_j| \leq \frac{1}{j} \longrightarrow 0$$

and by the convexity of S it follows that

$$\boldsymbol{x} + t'_j \boldsymbol{d}_j = (1 - \frac{t'_j}{t_j})\boldsymbol{x} + \frac{t'_j}{t_j}(\boldsymbol{x} + t_j \boldsymbol{d}_j) \in S,$$

which proves the assertion. □

Next we shall define the concept of normal cone (see Fig. 2.9). As we already have the contingent cone, it is natural to use polarity to define the normal vectors.

Definition 2.12 The *normal cone* of the nonempty set S at $\boldsymbol{x} \in S$ is the set

$$N_S(\boldsymbol{x}) := K_S(\boldsymbol{x})^\circ = \{\boldsymbol{z} \in \mathbb{R}^n \mid \boldsymbol{z}^T \boldsymbol{d} \leq 0 \text{ for all } \boldsymbol{d} \in K_S(\boldsymbol{x})\}. \tag{2.15}$$

The elements of $N_S(\boldsymbol{x})$ are called *normal vectors*.

The natural corollary of the polarity is that the normal cone has the same properties as the contingent cone.

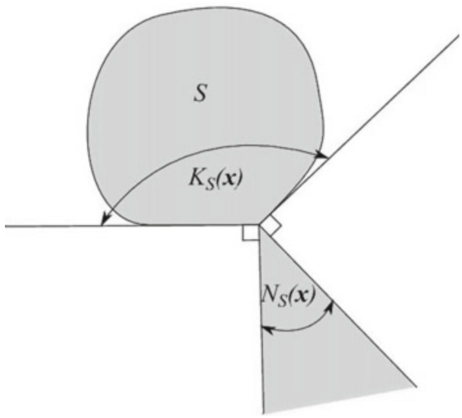


Fig. 2.9 Contingent and normal cones of a convex set

Theorem 2.18 *The normal cone $N_S(\boldsymbol{x})$ of the nonempty convex set S at $\boldsymbol{x} \in S$ is a closed convex cone.*

Proof Follows directly from Lemma 2.5. □

Notice that if $\boldsymbol{x} \in \text{int } S$, then clearly $K_S(\boldsymbol{x}) = \mathbb{R}^n$ and $N_S(\boldsymbol{x}) = \emptyset$. Thus the only interesting cases are those when $\boldsymbol{x} \in \text{bd } S$.

Next we present the following alternative characterization to the normal cone.

Theorem 2.19 *The normal cone of the nonempty convex set S at $\boldsymbol{x} \in S$ can also be written as follows*

$$N_S(\boldsymbol{x}) = \{\boldsymbol{z} \in \mathbb{R}^n \mid \boldsymbol{z}^T(\boldsymbol{y} - \boldsymbol{x}) \leq 0 \;\; for\ all\ \boldsymbol{y} \in S\}. \tag{2.16}$$

Proof Let us denote

$$Z := \{\boldsymbol{z} \in \mathbb{R}^n \mid \boldsymbol{z}^T(\boldsymbol{y} - \boldsymbol{x})^T \leq 0 \text{ for all } \boldsymbol{y} \in S\}.$$

If $\boldsymbol{z} \in N_S(\boldsymbol{x})$ is an arbitrary point, then by the definition of the normal cone we have

$$\boldsymbol{z}^T\boldsymbol{d} \leq 0 \quad \text{for all} \quad \boldsymbol{d} \in K_S(\boldsymbol{x}).$$

Now let $\boldsymbol{y} \in S$, set $\boldsymbol{d} := \boldsymbol{y} - \boldsymbol{x}$ and choose $t := 1$. Then

$$\boldsymbol{x} + t\boldsymbol{d} = \boldsymbol{x} + t\boldsymbol{y} - t\boldsymbol{x} = \boldsymbol{y} \in S,$$

thus $\boldsymbol{d} \in G_S(\boldsymbol{x}) \subseteq \text{cl } G_S(\boldsymbol{x}) = K_S(\boldsymbol{x})$ by Theorem 2.17. Since $\boldsymbol{z} \in N_S(\boldsymbol{x})$ one has

$$\boldsymbol{z}^T(\boldsymbol{y} - \boldsymbol{x})^T = \boldsymbol{z}^T\boldsymbol{d} \leq 0,$$

thus $\boldsymbol{z} \in Z$ and we have $N_S(\boldsymbol{x}) \subseteq Z$.

On the other hand, if $\boldsymbol{z} \in Z$ and $\boldsymbol{d} \in K_S(\boldsymbol{x})$ then there exist sequences $(\boldsymbol{d}_j) \subset \mathbb{R}^n$ and $(t_j) \subset \mathbb{R}$ such that $\boldsymbol{d}_j \to \boldsymbol{d}$, $t_j > 0$ and $\boldsymbol{x} + t_j\boldsymbol{d}_j \in S$ for all $j \in \mathbb{N}$. Let us set $\boldsymbol{y}_j := \boldsymbol{x} + t_j\boldsymbol{d}_j \in S$. Since $\boldsymbol{z} \in Z$ we have

$$t_j\boldsymbol{z}^T\boldsymbol{d}_j = \boldsymbol{z}^T(\boldsymbol{y}_j - \boldsymbol{x}) \leq 0.$$

Because t_j is positive, it implies that $\boldsymbol{z}^T\boldsymbol{d}_j \leq 0$ for all $j \in \mathbb{N}$. Then

$$\begin{aligned} \boldsymbol{z}^T\boldsymbol{d} &= \boldsymbol{z}^T\boldsymbol{d}_j + \boldsymbol{z}^T(\boldsymbol{d} - \boldsymbol{d}_j) \\ &\leq \|\boldsymbol{z}\|\, \|\boldsymbol{d} - \boldsymbol{d}_j\|, \end{aligned}$$

where $\|\boldsymbol{d} - \boldsymbol{d}_j\| \longrightarrow 0$ as $j \to \infty$. This means that

$$\boldsymbol{z}^T\boldsymbol{d} \leq 0 \quad \text{for all} \quad \boldsymbol{d} \in K_S(\boldsymbol{x}).$$

In other words, we have $\boldsymbol{z} \in N_S(\boldsymbol{x})$ and thus $Z \subseteq N_S(\boldsymbol{x})$, which completes the proof. □

The main difference between the groups of cones ray S, cone S, S° and $K_S(\boldsymbol{x})$, $G_S(\boldsymbol{x})$, $N_S(\boldsymbol{x})$ is, that the origin is the vertex point of the cone in the first group and the point $\boldsymbol{x} \in S$ in the second group. If we shift $\boldsymbol{x}$ to the origin, we get the following connections between these two groups.

Theorem 2.20 *If S is a nonempty convex set such that $\mathbf{0} \in S$, then*

(i) $G_S(\mathbf{0}) = \text{ray } S$,
(ii) $K_S(\mathbf{0}) = \text{cl ray } S$,
(iii) $N_S(\mathbf{0}) = S^\circ$.

Proof Exercise. □

2.2 Convex Functions

A function $f: \mathbb{R}^n \to \mathbb{R}$ is said to be *convex* if

$$f(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda)f(\boldsymbol{y}) \tag{2.17}$$

whenever $\boldsymbol{x}$ and $\boldsymbol{y}$ are in $\mathbb{R}^n$ and $\lambda \in [0, 1]$. If a strict inequality holds in (2.17) for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ such that $\boldsymbol{x} \neq \boldsymbol{y}$ and $\lambda \in (0, 1)$, the function f is said to be *strictly convex*. A function $f: \mathbb{R}^n \to \mathbb{R}$ is *(strictly) concave* if $-f$ is (strictly) convex (see Fig. 2.10).

Next we give an equivalent definition of a convex function.

Theorem 2.21 *(Jensen's inequality) A function $f: \mathbb{R}^n \to \mathbb{R}$ is convex if and only if*

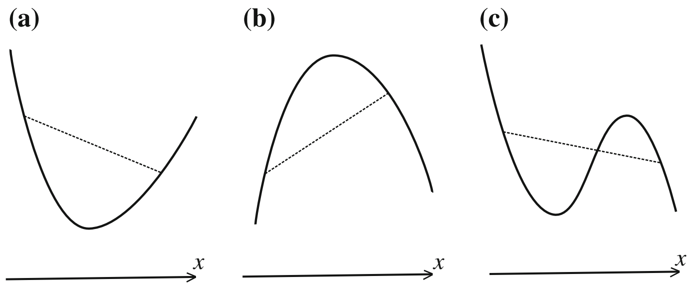


Fig. 2.10 Examples of different functions. (**a**) Convex. (**b**) Concave. (**c**) Neither convex or concave

$$f\left(\sum_{i=1}^{m} \lambda_i \boldsymbol{x}_i\right) \leq \sum_{i=1}^{m} \lambda_i f(\boldsymbol{x}_i), \tag{2.18}$$

whenever $\boldsymbol{x}_i \in \mathbb{R}^n$, $\lambda_i \in [0, 1]$ *for all* $i = 1, \ldots, m$ *and* $\sum_{i=1}^{m} \lambda_i = 1$.

Proof Follows by induction from the definition of convex function. □

Next we show that a convex function is always locally Lipschitz continuous.

Theorem 2.22 *Let* $f\colon \mathbb{R}^n \to \mathbb{R}$ *be a convex function. Then for any* $\boldsymbol{x}$ *in* $\mathbb{R}^n$, f *is locally Lipschitz continuous at* $\boldsymbol{x}$.

Proof Let $\boldsymbol{u} \in \mathbb{R}^n$ be arbitrary. We begin by proving that f is bounded on a neighborhood of $\boldsymbol{u}$. Let $\varepsilon > 0$ and define the hypercube

$$S_\varepsilon := \{\boldsymbol{y} \in \mathbb{R}^n \mid |y_i - u_i| \leq \varepsilon \text{ for all } i = 1, \ldots, n\}.$$

Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ denote the $m = 2^n$ vertices of S_ε and let

$$M := \max \{f(\boldsymbol{u}_i) \mid i = 1, \ldots, m\}.$$

Since each $\boldsymbol{y} \in S_\varepsilon$ can be expressed as $\boldsymbol{y} = \sum_{i=1}^{m} \lambda_i \boldsymbol{u}_i$ with $\lambda_i \geq 0$ and $\sum_{i=1}^{m} \lambda_i = 1$, by Theorem 2.21. we obtain

$$f(\boldsymbol{y}) = f\left(\sum_{i=1}^{m} \lambda_i \boldsymbol{u}_i\right) \leq \sum_{i=1}^{m} \lambda_i f(\boldsymbol{u}_i) \leq M \sum_{i=1}^{m} \lambda_i = M.$$

Since $B(\boldsymbol{u}; \varepsilon) \subset S_\varepsilon$, we have an upper bound M of f on an ε-neighborhood of $\boldsymbol{u}$, that is

$$f(\boldsymbol{x}') \leq M \quad \text{for all} \quad \boldsymbol{x}' \in B(\boldsymbol{u}; \varepsilon).$$

Now let $\boldsymbol{x} \in \mathbb{R}^n$, choose $\rho > 1$ and $\boldsymbol{y} \in \mathbb{R}^n$ so that $\boldsymbol{y} = \rho \boldsymbol{x}$. Define

$$\begin{aligned}
&\lambda := 1/\rho \quad \text{and} \\
&V := \{\boldsymbol{v} \mid \boldsymbol{v} = (1-\lambda)(\boldsymbol{x}' - \boldsymbol{u}) + \boldsymbol{x}, \ \text{where } \boldsymbol{x}' \in B(\boldsymbol{u}; \varepsilon)\}.
\end{aligned}$$

The set V is a neighborhood of $\boldsymbol{x} = \lambda \boldsymbol{y}$ with radius $(1-\lambda)\varepsilon$. By convexity one has for all $\boldsymbol{v} \in V$

$$\begin{aligned}
f(\boldsymbol{v}) &= f((1-\lambda)(\boldsymbol{x}' - \boldsymbol{u}) + \lambda \boldsymbol{y}) \\
&= f((1-\lambda)\boldsymbol{x}' + \lambda(\boldsymbol{y} + \boldsymbol{u} - \tfrac{1}{\lambda}\boldsymbol{u})) \\
&\leq (1-\lambda) f(\boldsymbol{x}') + \lambda f(\boldsymbol{y} + \boldsymbol{u} - \tfrac{1}{\lambda}\boldsymbol{u}).
\end{aligned}$$

Now $f(\boldsymbol{x}') \leq M$ and $f(\boldsymbol{y} + \boldsymbol{u} - \frac{1}{\lambda}\boldsymbol{u}) = \text{constant} =: K$ and thus

$$f(\boldsymbol{v}) \leq M + \lambda K.$$

In other words, f is bounded above on a neighborhood of $\boldsymbol{x}$.

Let us next show that f is also bounded below. Let $\boldsymbol{z} \in B(\boldsymbol{x}; (1-\lambda)\varepsilon)$ and define $\boldsymbol{z}' := 2\boldsymbol{x} - \boldsymbol{z}$. Then

$$\|\boldsymbol{z}' - \boldsymbol{x}\| = \|\boldsymbol{x} - \boldsymbol{z}\| \leq (1-\lambda)\varepsilon.$$

Thus $\boldsymbol{z}' \in B(\boldsymbol{x}; (1-\lambda)\varepsilon)$ and $\boldsymbol{x} = (\boldsymbol{z} + \boldsymbol{z}')/2$. The convexity of f implies that

$$f(\boldsymbol{x}) = f((\boldsymbol{z} + \boldsymbol{z}')/2) \leq \tfrac{1}{2} f(\boldsymbol{z}) + \tfrac{1}{2} f(\boldsymbol{z}'),$$

and

$$f(\boldsymbol{z}) \geq 2 f(\boldsymbol{x}) - f(\boldsymbol{z}') \geq 2 f(\boldsymbol{x}) - M - \lambda K$$

so that f is also bounded below on a neighborhood of $\boldsymbol{x}$. Thus we have proved that f is bounded on a neighborhood of $\boldsymbol{x}$.

Let $N > 0$ be a bound of $|f|$ so that

$$|f(\boldsymbol{x}')| \leq N \quad \text{for all} \quad \boldsymbol{x}' \in B(\boldsymbol{x}; 2\delta),$$

where $\delta > 0$, and let $\boldsymbol{x}_1, \boldsymbol{x}_2 \in B(\boldsymbol{x}; \delta)$ with $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$. Define

$$\boldsymbol{x}_3 := \boldsymbol{x}_2 + (\delta/\alpha)(\boldsymbol{x}_2 - \boldsymbol{x}_1),$$

where $\alpha := \|\boldsymbol{x}_2 - \boldsymbol{x}_1\|$. Then

$$
\begin{aligned}
\|\boldsymbol{x}_3 - \boldsymbol{x}\| &= \|\boldsymbol{x}_2 + (\delta/\alpha)(\boldsymbol{x}_2 - \boldsymbol{x}_1) - \boldsymbol{x}\| \\
&\leq \|\boldsymbol{x}_2 - \boldsymbol{x}\| + (\delta/\alpha)\|\boldsymbol{x}_2 - \boldsymbol{x}_1\| \\
&< \delta + \frac{\delta}{\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|}\|\boldsymbol{x}_2 - \boldsymbol{x}_1\| \\
&= 2\delta,
\end{aligned}
$$

thus $\boldsymbol{x}_3 \in B(\boldsymbol{x}; 2\delta)$. Solving for $\boldsymbol{x}_2$ gives

$$
\boldsymbol{x}_2 = \frac{\delta}{\alpha + \delta}\boldsymbol{x}_1 + \frac{\alpha}{\alpha + \delta}\boldsymbol{x}_3,
$$

and by the convexity we get

$$
f(\boldsymbol{x}_2) \leq \frac{\delta}{\alpha + \delta} f(\boldsymbol{x}_1) + \frac{\alpha}{\alpha + \delta} f(\boldsymbol{x}_3).
$$

Then

$$
\begin{aligned}
f(\boldsymbol{x}_2) - f(\boldsymbol{x}_1) &\leq \frac{\alpha}{\alpha + \delta}[f(\boldsymbol{x}_3) - f(\boldsymbol{x}_1)] \\
&\leq \frac{\alpha}{\delta}|f(\boldsymbol{x}_3) - f(\boldsymbol{x}_1)| \\
&\leq \frac{\alpha}{\delta}(|f(\boldsymbol{x}_3)| + |f(\boldsymbol{x}_1)|).
\end{aligned}
$$

Since $\boldsymbol{x}_1, \boldsymbol{x}_3 \in B(\boldsymbol{x}; 2\delta)$ we have $|f(\boldsymbol{x}_3)| < N$ and $|f(\boldsymbol{x}_1)| < N$, thus

$$
f(\boldsymbol{x}_2) - f(\boldsymbol{x}_1) \leq \frac{2N}{\delta}\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|.
$$

By changing the roles of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ we have

$$
|f(\boldsymbol{x}_2) - f(\boldsymbol{x}_1)| \leq \frac{2N}{\delta}\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|,
$$

showing that the function f is locally Lipschitz continuous at $\boldsymbol{x}$. $\square$

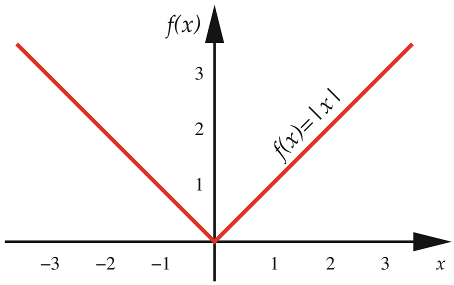


Fig. 2.11 Absolute-value function $f(x) = |x|$

The simplest example of nonsmooth function is the absolute-value function on reals (see Fig. 2.11).

Example 2.8 (*Absolute-value function*). Let us consider the absolute-value function

$$f(x) = |x|$$

on reals.

The gradient of function f is

$$\nabla f(x) = \begin{cases} 1, & \text{when } x > 0, \\ -1, & \text{when } x < 0. \end{cases}$$

Function f is not differentiable at $x = 0$.

We now show that function f is both convex and (locally) Lipschitz continuous. Let $\lambda \in [0, 1]$ and $x, y \in \mathbb{R}$. By triangle inequality we have

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= |\lambda x + (1-\lambda)y| \\ &\le |\lambda x| + |(1-\lambda)y| \\ &= |\lambda||x| + |1-\lambda||y| \\ &= \lambda|x| + (1-\lambda)|y| \\ &= \lambda f(x) + (1-\lambda)f(y). \end{aligned}$$

Thus, function f is convex. Furthermore, by triangle inequality we also have

$$|f(x) - f(y)| = ||x| - |y|| \le |x - y|$$

for all $x, y \in \mathbb{R}$. In space $\mathbb{R}$, the right-hand side equals to the norm $\|\boldsymbol{x} - \boldsymbol{y}\|$. Thus, we have the Lipschitz constant $K = 1 > 0$ and function f is Lipschitz continuous.

2.2.1 Level Sets and Epigraphs

Next we consider two sets, namely level sets and epigraphs, closely related to convex functions.

Definition 2.13 The *level set* of a function $f: \mathbb{R}^n \to \mathbb{R}$ with a parameter $\alpha \in \mathbb{R}$ is defined as

$$\operatorname{lev}_\alpha f := \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) \le \alpha\}.$$

We have the following connection between the convexity of functions and level sets.

Theorem 2.23 *If $f: \mathbb{R}^n \to \mathbb{R}$ is a convex function, then the level set* $\operatorname{lev}_\alpha f$ *is a convex set for all $\alpha \in \mathbb{R}$.*

Proof If $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{lev}_\alpha f$ and $\lambda \in [0, 1]$ we have $f(\boldsymbol{x}) \le \alpha$ and $f(\boldsymbol{y}) \le \alpha$. Let $\boldsymbol{z} := \lambda\boldsymbol{x} + (1 - \lambda)\boldsymbol{y}$ with some $\lambda \in [0, 1]$. Then the convexity of f implies that

$$f(\boldsymbol{z}) \le \lambda f(\boldsymbol{x}) + (1 - \lambda) f(\boldsymbol{y}) \le \lambda\alpha + (1 - \lambda)\alpha = \alpha,$$

in other words $\boldsymbol{z} \in \operatorname{lev}_\alpha f$ and thus $\operatorname{lev}_\alpha f$ is convex. □

The previous result can not be inverted since there exist nonconvex functions with convex level sets (see Fig. 2.12). The equivalence can be achieved by replacing the level set with the so called epigraph being a subset of $\mathbb{R}^n \times \mathbb{R}$ (see Fig. 2.13).

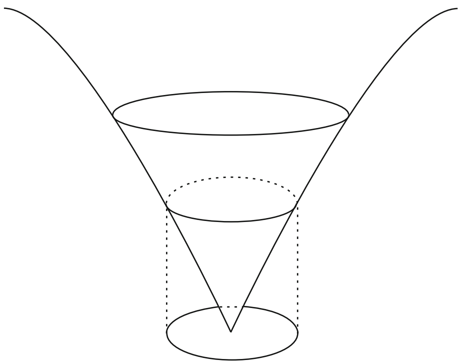


Fig. 2.12 Nonconvex function with convex level sets

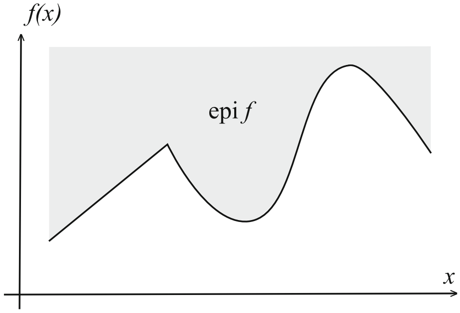


Fig. 2.13 Epigraph of the function

Definition 2.14 The *epigraph* of a function $f: \mathbb{R}^n \to \mathbb{R}$ is the following subset of $\mathbb{R}^n \times \mathbb{R}$:

$$\text{epi } f := \{(\boldsymbol{x}, r) \in \mathbb{R}^n \times \mathbb{R} \mid f(\boldsymbol{x}) \leq r\}. \tag{2.19}$$

Theorem 2.24 *The function $f: \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the epigraph* epi f *is a convex set.*

Proof Exercise. □

Notice, that we have the following connection between the epigraph and level sets of a function $f: \mathbb{R}^n \to \mathbb{R}$ at $\boldsymbol{x} \in \mathbb{R}^n$

$$\text{lev}_{f(\boldsymbol{x})} f = \{\boldsymbol{y} \in \mathbb{R}^n \mid (\boldsymbol{y}, f(\boldsymbol{x})) \in \text{epi } f\}.$$

2.2.2 Subgradients and Directional Derivatives

In this section we shall generalize the classical notion of gradient for convex but not necessarily differentiable functions. Before that we consider some properties related to the directional derivative of convex functions.

Theorem 2.25 *If $f: \mathbb{R}^n \to \mathbb{R}$ is a convex function, then the directional derivative $f'(\boldsymbol{x}; \boldsymbol{d})$ exists in every direction $\boldsymbol{d} \in \mathbb{R}^n$ and it satisfies*

$$f'(\boldsymbol{x}; \boldsymbol{d}) = \inf_{t>0} \frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})}{t}. \tag{2.20}$$

Proof Let $\boldsymbol{d} \in \mathbb{R}^n$ be an arbitrary direction. Define $\varphi: \mathbb{R} \to \mathbb{R}$ by

$$\varphi(t) := \frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})}{t}.$$

We begin by proving that φ is bounded below at t when $t \downarrow 0$. Let $\varepsilon > 0$ and let constants t_1 and t_2 be such that $0 < t_1 < t_2 < \varepsilon$. By the convexity of f we have

$$\begin{aligned}
\varphi(t_2) - \varphi(t_1) &= \frac{1}{t_1 t_2}[t_1 f(\boldsymbol{x} + t_2\boldsymbol{d}) - t_2 f(\boldsymbol{x} + t_1\boldsymbol{d}) + (t_2 - t_1) f(\boldsymbol{x})] \\
&= \frac{1}{t_1}\Big\{\Big(\frac{t_1}{t_2} f(\boldsymbol{x} + t_2\boldsymbol{d}) + (1 - \frac{t_1}{t_2}) f(\boldsymbol{x})\Big) \\
&\quad - f\Big(\frac{t_1}{t_2}(\boldsymbol{x} + t_2\boldsymbol{d}) + (1 - \frac{t_1}{t_2})\boldsymbol{x}\Big)\Big\} \\
&\geq 0,
\end{aligned}$$

thus the function $\varphi(t)$ decreases as $t \downarrow 0$. Then for all $0 < t < \varepsilon$ one has

$$\begin{aligned}
\varphi(t) - \varphi(-\varepsilon/2) &= \frac{\frac{1}{2} f(\boldsymbol{x} + t\boldsymbol{d}) + \frac{1}{2} f(\boldsymbol{x}) + \frac{t}{\varepsilon} f(\boldsymbol{x} - \frac{\varepsilon}{2}\boldsymbol{d}) + (1 - \frac{t}{\varepsilon}) f(\boldsymbol{x}) - 2 f(\boldsymbol{x})}{t/2} \\
&\geq \frac{\frac{1}{2} f(\boldsymbol{x} + \frac{t}{2}\boldsymbol{d}) + \frac{1}{2} f(\boldsymbol{x} - \frac{t}{2}\boldsymbol{d}) - f(\boldsymbol{x})}{t/4} \\
&\geq \frac{f(\boldsymbol{x}) - f(\boldsymbol{x})}{t/4} = 0,
\end{aligned}$$

which means that the function φ is bounded below for $0 < t < \varepsilon$. This implies that there exists the limit

$$\lim_{t \downarrow 0} \varphi(t) = f'(\boldsymbol{x}; \boldsymbol{d}) \quad \text{for all} \quad \boldsymbol{d} \in \mathbb{R}^n$$

and since the function $\varphi(t)$ decreases as $t \downarrow 0$ we deduce that

$$f'(\boldsymbol{x}; \boldsymbol{d}) = \inf_{t > 0} \frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})}{t}.$$

□

Theorem 2.26 *Let $f: \mathbb{R}^n \to \mathbb{R}$ be a convex function with a Lipschitz constant K at $\boldsymbol{x} \in \mathbb{R}^n$. Then the function $\boldsymbol{d} \mapsto f'(\boldsymbol{x}; \boldsymbol{d})$ is positively homogeneous and subadditive on $\mathbb{R}^n$ with*

$$|f'(\boldsymbol{x}; \boldsymbol{d})| \leq K\|\boldsymbol{d}\|.$$

Proof We start by proving the inequality. From the Lipschitz condition we obtain

$$\begin{aligned}
|f'(\boldsymbol{x}; \boldsymbol{d})| &\leq \lim_{t \downarrow 0} \frac{|f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})|}{t} \\
&\leq \lim_{t \downarrow 0} \frac{K\|\boldsymbol{x} + t\boldsymbol{d} - \boldsymbol{x}\|}{t} \\
&\leq K\|\boldsymbol{d}\|.
\end{aligned}$$

Next we show that $f'(\boldsymbol{x};\cdot)$ is positively homogeneous. To see this, let $\lambda > 0$. Then

$$\begin{aligned} f'(\boldsymbol{x};\lambda\boldsymbol{d}) &= \lim_{t\downarrow 0}\frac{f(\boldsymbol{x}+t\lambda\boldsymbol{d})-f(\boldsymbol{x})}{t} \\ &= \lim_{t\downarrow 0}\lambda\cdot\left\{\frac{f(\boldsymbol{x}+t\lambda\boldsymbol{d})-f(\boldsymbol{x})}{t\lambda}\right\} \\ &= \lambda\cdot\lim_{t\downarrow 0}\frac{f(\boldsymbol{x}+t\lambda\boldsymbol{d})-f(\boldsymbol{x})}{t\lambda} \\ &= \lambda\cdot f'(\boldsymbol{x};\boldsymbol{d}). \end{aligned}$$

We turn now to the subadditivity. Let $\boldsymbol{d},\boldsymbol{p}\in\mathbb{R}^n$ be arbitrary directions, then by convexity

$$\begin{aligned} f'(\boldsymbol{x};\boldsymbol{d}+\boldsymbol{p}) &= \lim_{t\downarrow 0}\frac{f(\boldsymbol{x}+t(\boldsymbol{d}+\boldsymbol{p}))-f(\boldsymbol{x})}{t} \\ &= \lim_{t\downarrow 0}\frac{f(\frac{1}{2}(\boldsymbol{x}+2t\boldsymbol{d})+\frac{1}{2}(\boldsymbol{x}+2t\boldsymbol{p}))-f(\boldsymbol{x})}{t} \\ &\le \lim_{t\downarrow 0}\frac{f(\boldsymbol{x}+2t\boldsymbol{d})-f(\boldsymbol{x})}{2t}+\lim_{t\downarrow 0}\frac{f(\boldsymbol{x}+2t\boldsymbol{p})-f(\boldsymbol{x})}{2t} \\ &= f'(\boldsymbol{x};\boldsymbol{d})+f'(\boldsymbol{x};\boldsymbol{p}). \end{aligned}$$

Thus $\boldsymbol{d}\mapsto f'(\boldsymbol{x};\boldsymbol{d})$ is subadditive. □

From the previous theorem we derive the following consequence.

Corollary 2.3 *If $f:\mathbb{R}^n\to\mathbb{R}$ is a convex function, then the function $\boldsymbol{d}\mapsto f'(\boldsymbol{x};\boldsymbol{d})$ is convex, its epigraph* epi $f'(\boldsymbol{x};\cdot)$ *is a convex cone and we have*

$$f'(\boldsymbol{x};-\boldsymbol{d})\ge -f'(\boldsymbol{x};\boldsymbol{d})\quad\textit{for all } \boldsymbol{x}\in\mathbb{R}^n.$$

Proof Exercise. □

Next we define the subgradient and the subdifferential of a convex function. Note the analogy to the smooth differential theory, namely if a function $f:\mathbb{R}^n\to\mathbb{R}$ is both convex and differentiable, then for all $\boldsymbol{y}\in\mathbb{R}^n$ we have

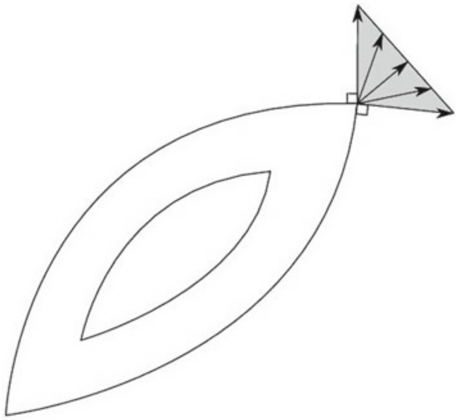
$$f(\boldsymbol{y})\ge f(\boldsymbol{x})+\nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}).$$

Figure 2.14 illustrates the meaning of the definition of the subdifferential.

Definition 2.15 The *subdifferential of a convex function* $f:\mathbb{R}^n\to\mathbb{R}$ at $\boldsymbol{x}\in\mathbb{R}^n$ is the set $\partial_c f(\boldsymbol{x})$ of vectors $\boldsymbol{\xi}\in\mathbb{R}^n$ such that

$$\partial_c f(\boldsymbol{x})=\left\{\boldsymbol{\xi}\in\mathbb{R}^n \mid f(\boldsymbol{y})\ge f(\boldsymbol{x})+\boldsymbol{\xi}^T(\boldsymbol{y}-\boldsymbol{x})\text{ for all }\boldsymbol{y}\in\mathbb{R}^n\right\}.$$

Fig. 2.14 Subdifferential



Each vector $\boldsymbol{\xi} \in \partial_c f(\boldsymbol{x})$ is called a *subgradient* of f at $\boldsymbol{x}$.

Example 2.9 (*Absolute-value function*). As noted in Example 2.8 function $f(x) = |x|$ is convex and differentiable when $x \neq 0$. By the definition of subdifferential we have

$$\begin{aligned} \xi \in \partial_c f(0) &\iff |y| \geq |0| + \xi \cdot (y - 0) \quad \text{for all } y \in \mathbb{R} \\ &\iff |y| \geq \xi \cdot y \quad \text{for all } y \in \mathbb{R} \\ &\iff \xi \leq 1 \quad \text{and} \quad \xi \geq -1. \end{aligned}$$

Thus, $\partial_c f(0) = [-1, 1]$.

Theorem 2.27 *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function with a Lipschitz constant K at $\boldsymbol{x}^* \in \mathbb{R}^n$. Then the subdifferential $\partial_c f(\boldsymbol{x}^*)$ is a nonempty, convex, and compact set such that*

$$\partial_c f(\boldsymbol{x}^*) \subseteq B(\mathbf{0}; K).$$

Proof We show first that there exists a subgradient $\boldsymbol{\xi} \in \partial_c f(\boldsymbol{x}^*)$, in other words $\partial_c f(\boldsymbol{x}^*)$ is nonempty. By Theorem 2.24 epi f is a convex set and by Theorem 2.22 and Exercise 2.29 it is closed. Since $(\boldsymbol{x}^*, f(\boldsymbol{x}^*)) \in \text{epi } f$ it is also nonempty, furthermore we have $(\boldsymbol{x}^*, f(\boldsymbol{x}^*)) \in \text{bd epi } f$. Then due to Theorem 2.7 there exists a hyperplane supporting epi f at $(\boldsymbol{x}^*, f(\boldsymbol{x}^*))$. In other words there exists $(\boldsymbol{\xi}^*, \mu) \neq (\mathbf{0}, 0)$ where $\boldsymbol{\xi}^* \in \mathbb{R}^n$ and $\mu \in \mathbb{R}$ such that for all $(\boldsymbol{x}, r) \in \text{epi } f$ we have

$$(\boldsymbol{\xi}^*, \mu)^T \big((\boldsymbol{x}, r) - (\boldsymbol{x}^*, f(\boldsymbol{x}^*))\big) = (\boldsymbol{\xi}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) + \mu(r - f(\boldsymbol{x}^*)) \leq 0. \quad (2.21)$$

In the above inequality r can be chosen as large as possible, thus μ must be nonpositive. If $\mu = 0$ then (2.21) reduces to

$$(\boldsymbol{\xi}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) \leq 0 \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^n.$$

If we choose $\boldsymbol{x} := \boldsymbol{x}^* + \boldsymbol{\xi}^*$ we get $(\boldsymbol{\xi}^*)^T\boldsymbol{\xi}^* = \|\boldsymbol{\xi}^*\|^2 \leq 0$. This means that $\boldsymbol{\xi}^* = \mathbf{0}$, which is impossible because $(\boldsymbol{\xi}^*, \mu) \neq (\mathbf{0}, 0)$, thus we have $\mu < 0$. Dividing the inequality (2.21) by $|\mu|$ and noting $\boldsymbol{\xi} := \boldsymbol{\xi}^*/|\mu|$ we get

$$\boldsymbol{\xi}^T(\boldsymbol{x} - \boldsymbol{x}^*) - r + f(\boldsymbol{x}^*) \leq 0 \quad \text{for all } (\boldsymbol{x}, r) \in \text{epi } f.$$

If we choose now $r := f(\boldsymbol{x})$ we get

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \boldsymbol{\xi}^T(\boldsymbol{x} - \boldsymbol{x}^*) \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^n,$$

which means that $\boldsymbol{\xi} \in \partial_c f(\boldsymbol{x}^*)$.

To see the convexity let $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \in \partial_c f(\boldsymbol{x}^*)$ and $\lambda \in [0, 1]$. Then we have

$$\begin{aligned} f(\boldsymbol{y}) &\geq f(\boldsymbol{x}^*) + \boldsymbol{\xi}_1^T(\boldsymbol{y} - \boldsymbol{x}^*) \quad \text{for all } \boldsymbol{y} \in \mathbb{R}^n \quad \text{and} \\ f(\boldsymbol{y}) &\geq f(\boldsymbol{x}^*) + \boldsymbol{\xi}_2^T(\boldsymbol{y} - \boldsymbol{x}^*) \quad \text{for all } \boldsymbol{y} \in \mathbb{R}^n. \end{aligned}$$

Multiplying the above two inequalities by λ and $(1 - \lambda)$, respectively, and adding them together, we obtain

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}^*) + \big(\lambda\boldsymbol{\xi}_1 + (1-\lambda)\boldsymbol{\xi}_2\big)^T(\boldsymbol{y} - \boldsymbol{x}^*) \quad \text{for all } \boldsymbol{y} \in \mathbb{R}^n,$$

in other words

$$\lambda\boldsymbol{\xi}_1 + (1-\lambda)\boldsymbol{\xi}_2 \in \partial_c f(\boldsymbol{x}^*)$$

and thus $\partial_c f(\boldsymbol{x}^*)$ is convex.

If $\boldsymbol{d} \in \mathbb{R}^n$ we get from the definition of the subdifferential

$$\varphi(t) := \frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})}{t} \geq \frac{\boldsymbol{\xi}^T(t\boldsymbol{d})}{t} = \boldsymbol{\xi}^T\boldsymbol{d} \quad \text{for all } \boldsymbol{\xi} \in \partial_c f(\boldsymbol{x}^*).$$

Since $\varphi(t) \to f'(\boldsymbol{x}^*; \boldsymbol{d})$ when $t \downarrow 0$ we obtain

$$f'(\boldsymbol{x}^*; \boldsymbol{d}) \geq \boldsymbol{\xi}^T\boldsymbol{d} \quad \text{for all } \boldsymbol{\xi} \in \partial_c f(\boldsymbol{x}^*). \tag{2.22}$$

Thus for an arbitrary $\boldsymbol{\xi} \in \partial_c f(\boldsymbol{x}^*)$ we get

$$\|\boldsymbol{\xi}\|^2 = |\boldsymbol{\xi}^T\boldsymbol{\xi}| \leq |f'(\boldsymbol{x}^*; \boldsymbol{\xi})| \leq K\|\boldsymbol{\xi}\|$$

by Theorem 2.26. This means that $\partial_c f(\boldsymbol{x}^*)$ is bounded and we have

$$\partial_c f(\boldsymbol{x}^*) \subseteq B(\mathbf{0}; K).$$

Thus, for compactness it suffices to show that $\partial_c f(\boldsymbol{x}^*)$ is closed. To see this let $(\boldsymbol{\xi}_i) \subset \partial_c f(\boldsymbol{x}^*)$ such that $\boldsymbol{\xi}_i \to \boldsymbol{\xi}$. Then for all $\boldsymbol{y} \in \mathbb{R}^n$ we have

$$f(\boldsymbol{y}) - f(\boldsymbol{x}^*) \geq \boldsymbol{\xi}_i^T(\boldsymbol{y} - \boldsymbol{x}) \to \boldsymbol{\xi}^T(\boldsymbol{y} - \boldsymbol{x}),$$

whenever $i \to \infty$, thus $\boldsymbol{\xi} \in \partial_c f(\boldsymbol{x}^*)$ and $\partial_c f(\boldsymbol{x}^*)$ is closed. $\square$

The next theorem shows the relationship between the subdifferential and the directional derivative. It turns out that knowing $f'(\boldsymbol{x}; \boldsymbol{d})$ is equivalent to knowing $\partial_c f(\boldsymbol{x})$.

Theorem 2.28 *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then for all $\boldsymbol{x} \in \mathbb{R}^n$*

(i) $\partial_c f(\boldsymbol{x}) = \{\boldsymbol{\xi} \in \mathbb{R}^n \mid f'(\boldsymbol{x}, \boldsymbol{d}) \geq \boldsymbol{\xi}^T \boldsymbol{d}$ *for all* $\boldsymbol{d} \in \mathbb{R}^n\}$, *and*
(ii) $f'(\boldsymbol{x}; \boldsymbol{d}) = \max\,\{\boldsymbol{\xi}^T \boldsymbol{d} \mid \boldsymbol{\xi} \in \partial_c f(\boldsymbol{x})\}$ *for any* $\boldsymbol{d} \in \mathbb{R}^n$.

Proof (i) Set

$$S := \{\boldsymbol{\xi} \in \mathbb{R}^n \mid f'(\boldsymbol{x}; \boldsymbol{d}) \geq \boldsymbol{\xi}^T \boldsymbol{d} \text{ for all } \boldsymbol{d} \in \mathbb{R}^n\}$$

and let $\boldsymbol{\xi} \in S$ be arbitrary. Then it follows from convexity that, for all $\boldsymbol{d} \in \mathbb{R}^n$, we have

$$\begin{aligned}
\boldsymbol{\xi}^T \boldsymbol{d} &\leq f'(\boldsymbol{x}; \boldsymbol{d}) \\
&= \lim_{t \downarrow 0} \frac{f\big((1-t)\boldsymbol{x} + t(\boldsymbol{x} + \boldsymbol{d})\big) - f(\boldsymbol{x})}{t} \\
&\leq \lim_{t \downarrow 0} \frac{(1-t)f(\boldsymbol{x}) + tf(\boldsymbol{x} + \boldsymbol{d}) - f(\boldsymbol{x})}{t} \\
&= f(\boldsymbol{x} + \boldsymbol{d}) - f(\boldsymbol{x}),
\end{aligned}$$

whenever $t \leq 1$. By choosing $\boldsymbol{d} := \boldsymbol{y} - \boldsymbol{x}$ we derive $\boldsymbol{\xi} \in \partial_c f(\boldsymbol{x})$. On the other hand, if $\boldsymbol{\xi} \in \partial_c f(\boldsymbol{x})$ then due to (2.22) we have

$$f'(\boldsymbol{x}; \boldsymbol{d}) \geq \boldsymbol{\xi}^T \boldsymbol{d} \quad \text{for all } \boldsymbol{d} \in \mathbb{R}^n.$$

Thus $\boldsymbol{\xi} \in S$, which establishes (i).

(ii) First we state that since the subdifferential is compact and nonempty set (Theorem 2.27) the maximum of the linear function $\boldsymbol{d} \mapsto \boldsymbol{\xi}^T \boldsymbol{d}$ is well-defined due to the Weierstrass' Theorem 1.1. Again from (2.22) we deduce that for each $\boldsymbol{d} \in \mathbb{R}^n$ we have

$$f'(\boldsymbol{x}; \boldsymbol{d}) \geq \max\,\{\boldsymbol{\xi}^T \boldsymbol{d} \mid \boldsymbol{\xi} \in \partial_c f(\boldsymbol{x})\}.$$

Suppose next that there were $\boldsymbol{d}^* \in \mathbb{R}^n$ for which

$$f'(\boldsymbol{x}; \boldsymbol{d}^*) > \max \{\boldsymbol{\xi}^T \boldsymbol{d}^* \mid \boldsymbol{\xi} \in \partial_c f(\boldsymbol{x})\}. \tag{2.23}$$

By Corollary 2.3 function $\boldsymbol{d} \mapsto f'(\boldsymbol{x}; \boldsymbol{d})$ is convex and thus by Theorem 2.24 epi $f'(\boldsymbol{x}; \cdot)$ is a convex set and by Theorem 2.22 and Exercise 2.29 it is closed. Since $(\boldsymbol{d}^*, f'(\boldsymbol{x}; \boldsymbol{d}^*)) \in$ epi $f'(\boldsymbol{x}; \cdot)$ it is also nonempty, furthermore we have $(\boldsymbol{d}^*, f'(\boldsymbol{x}; \boldsymbol{d}^*)) \in$ bd epi $f'(\boldsymbol{x}; \cdot)$. Then due to Theorem 2.7 there exists a hyperplane supporting epi $f'(\boldsymbol{x}; \cdot)$ at $(\boldsymbol{d}^*, f'(\boldsymbol{x}; \boldsymbol{d}^*))$, in other words there exists $(\boldsymbol{\xi}^*, \mu) \neq (\boldsymbol{0}, 0)$ where $\boldsymbol{\xi}^* \in \mathbb{R}^n$ and $\mu \in \mathbb{R}$ such that for all $(\boldsymbol{d}, r) \in$ epi $f'(\boldsymbol{x}; \cdot)$ we have

$$\begin{aligned}(\boldsymbol{\xi}^*, \mu)^T \big((\boldsymbol{d}, r) - (\boldsymbol{d}^*, f'(\boldsymbol{x}; \boldsymbol{d}^*))\big) &= (\boldsymbol{\xi}^*)^T (\boldsymbol{d} - \boldsymbol{d}^*) + \mu(r - f'(\boldsymbol{x}; \boldsymbol{d}^*)) \\ &\leq 0.\end{aligned} \tag{2.24}$$

Just like in the proof of Theorem 2.27 we can deduce that $\mu < 0$. Again dividing the inequality (2.24) by $|\mu|$ and noting $\boldsymbol{\xi} := \boldsymbol{\xi}^*/|\mu|$ we get

$$\boldsymbol{\xi}^T (\boldsymbol{d} - \boldsymbol{d}^*) - r + f'(\boldsymbol{x}; \boldsymbol{d}^*) \leq 0 \quad \text{for all } (\boldsymbol{d}, r) \in \text{epi } f'(\boldsymbol{x}; \cdot).$$

If we choose now $r := f'(\boldsymbol{x}; \boldsymbol{d})$ we get

$$f'(\boldsymbol{x}; \boldsymbol{d}) - f'(\boldsymbol{x}; \boldsymbol{d}^*) \geq \boldsymbol{\xi}^T (\boldsymbol{d} - \boldsymbol{d}^*) \quad \text{for all } \boldsymbol{d} \in \mathbb{R}^n. \tag{2.25}$$

Then from the subadditivity of the directional derivative (Theorem 2.26) we obtain

$$f'(\boldsymbol{x}; \boldsymbol{d} - \boldsymbol{d}^*) \geq \boldsymbol{\xi}^T (\boldsymbol{d} - \boldsymbol{d}^*) \quad \text{for all } \boldsymbol{d} \in \mathbb{R}^n,$$

which by assertion (i) means that $\boldsymbol{\xi} \in \partial_c f(\boldsymbol{x})$. On the other hand from Eqs. (2.25) and (2.23) we get

$$f'(\boldsymbol{x}; \boldsymbol{d}) - \boldsymbol{\xi}^T \boldsymbol{d} \geq f'(\boldsymbol{x}; \boldsymbol{d}^*) - \boldsymbol{\xi}^T \boldsymbol{d}^* > 0 \quad \text{for all } \boldsymbol{d} \in \mathbb{R}^n,$$

in other words we have

$$f'(\boldsymbol{x}; \boldsymbol{d}) > \boldsymbol{\xi}^T \boldsymbol{d} \quad \text{for all } \boldsymbol{d} \in \mathbb{R}^n.$$

Now by choosing $\boldsymbol{d} := \boldsymbol{0}$ we get '$0 > 0$', which is impossible, thus by the contradiction (2.23) is wrong and we have the equality

$$f'(\boldsymbol{x}; \boldsymbol{d}) = \max \{\boldsymbol{\xi}^T \boldsymbol{d} \mid \boldsymbol{\xi} \in \partial_c f(\boldsymbol{x})\} \quad \text{for all } \boldsymbol{d} \in \mathbb{R}^n$$

and the proof is complete. □

Example 2.10 (*Absolute-value function*). By Theorem 2.28 (i) we have

$$\xi \in \partial_c f(0) \iff f'(0, d) \geq \xi \cdot d \quad \text{for all } d \in \mathbb{R}.$$

Now

$$f'(0, d) = \lim_{t \downarrow 0} \frac{|0 + td| - |0|}{t} = \lim_{t \downarrow 0} \frac{t|d|}{t} = |d|$$

and, thus,

$$\begin{aligned} \xi \in \partial_c f(0) &\iff |d| \geq \xi \cdot d \quad \text{for all } d \in \mathbb{R} \\ &\iff \xi \in [-1, 1]. \end{aligned}$$

The next theorem shows that the subgradients really are generalizations of the classical gradient.

Theorem 2.29 *If* $f\colon \mathbb{R}^n \to \mathbb{R}$ *is convex and differentiable at* $\boldsymbol{x} \in \mathbb{R}^n$, *then*

$$\partial_c f(\boldsymbol{x}) = \{\nabla f(\boldsymbol{x})\}. \tag{2.26}$$

Proof According to Theorem 2.25 the directional derivative $f'(\boldsymbol{x}; \boldsymbol{d})$ of a convex function exists in every direction $\boldsymbol{d} \in \mathbb{R}^n$. From the definition of differentiability we have

$$f'(\boldsymbol{x}; \boldsymbol{d}) = \nabla f(\boldsymbol{x})^T \boldsymbol{d} \quad \text{for all } \boldsymbol{d} \in \mathbb{R}^n,$$

which implies, by Theorem 2.28 (i) that $\nabla f(\boldsymbol{x}) \in \partial_c f(\boldsymbol{x})$. Suppose next that there exists another $\boldsymbol{\xi} \in \partial_c f(\boldsymbol{x})$ such that $\boldsymbol{\xi} \neq \nabla f(\boldsymbol{x})$. Then by Theorem 2.28 (i) we have

$$\boldsymbol{\xi}^T \boldsymbol{d} \leq f'(\boldsymbol{x}; \boldsymbol{d}) = \nabla f(\boldsymbol{x})^T \boldsymbol{d} \quad \text{for all } \boldsymbol{d} \in \mathbb{R}^n,$$

in other words

$$(\boldsymbol{\xi} - \nabla f(\boldsymbol{x}))^T \boldsymbol{d} \leq 0 \quad \text{for all } \boldsymbol{d} \in \mathbb{R}^n.$$

By choosing $\boldsymbol{d} := \boldsymbol{\xi} - \nabla f(\boldsymbol{x})$ we get

$$\|\boldsymbol{\xi} - \nabla f(\boldsymbol{x}))\|^2 \leq 0,$$

implying that $\boldsymbol{\xi} = \nabla f(\boldsymbol{x})$, which contradicts the assumption. Thus

$$\partial_c f(\boldsymbol{x}) = \{\nabla f(\boldsymbol{x})\}. \qquad \square$$

Example 2.11 (*Absolute-value function*). Let us define the whole subdifferential $\partial f(x)$ of the function $f(x) = |x|$. Function f is differentiable in everywhere except in $x = 0$, and

$$\nabla f(x) = \begin{cases} 1, & \text{when } x > 0 \\ -1, & \text{when } x < 0. \end{cases}$$

In Examples 2.9 and 2.10, we have computed the subdifferential at $x = 0$, that is, $\partial_c f(0) = [-1, 1]$. Thus, the subdifferential of f is

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

(see also Fig. 2.15).

We are now ready to present a very useful result in developing optimization methods. It gives a representation to a convex function by using subgradients.

Theorem 2.30 *If* $f \colon \mathbb{R}^n \to \mathbb{R}$ *is convex then for all* $\boldsymbol{y} \in \mathbb{R}^n$

$$f(\boldsymbol{y}) = \max\,\{f(\boldsymbol{x}) + \boldsymbol{\xi}^T(\boldsymbol{y} - \boldsymbol{x}) \mid \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{\xi} \in \partial_c f(\boldsymbol{x})\}. \tag{2.27}$$

Proof Suppose that $\boldsymbol{y} \in \mathbb{R}^n$ is an arbitrary point and $\boldsymbol{\zeta} \in \partial f(\boldsymbol{y})$. Let

$$S := \{f(\boldsymbol{x}) + \boldsymbol{\xi}^T(\boldsymbol{y} - \boldsymbol{x}) \mid \boldsymbol{\xi} \in \partial_c f(\boldsymbol{x}),\ \boldsymbol{x} \in \mathbb{R}^n\}.$$

By the definition of subdifferential of a convex function we have

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \boldsymbol{\xi}^T(\boldsymbol{y} - \boldsymbol{x}) \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^n \text{and } \boldsymbol{\xi} \in \partial_c f(\boldsymbol{x})$$

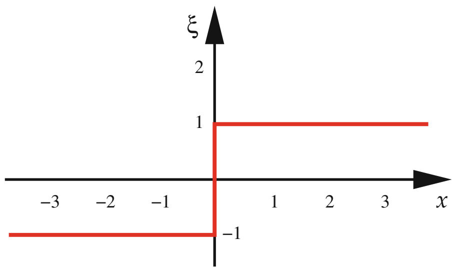


Fig. 2.15 Subdifferential $\partial_c f(x)$ of $f(x) = |x|$

implying that the set S is bounded from above and

$$\sup S \leq f(\boldsymbol{y}).$$

On the other hand, we have

$$f(\boldsymbol{y}) = f(\boldsymbol{y}) + \boldsymbol{\zeta}^T(\boldsymbol{y} - \boldsymbol{y}) \in S,$$

which means that $f(\boldsymbol{y}) \leq \sup S$. Thus

$$f(\boldsymbol{y}) = \max\,\{f(\boldsymbol{x}) + \boldsymbol{\xi}^T(\boldsymbol{y} - \boldsymbol{x}) \mid \boldsymbol{\xi} \in \partial_c f(\boldsymbol{x}),\ \boldsymbol{x} \in \mathbb{R}^n\}. \qquad \square$$

2.2.3 *ε-Subdifferentials*

In nonsmooth optimization, so called bundle methods are based on the concept of ε-subdifferential, which is an extension of the ordinary subdifferential. Therefore we now give the definition of ε-subdifferential and present some of its basic properties.

We start by generalizing the ordinary directional derivative. Note the analogy with the property (2.20).

Definition 2.16 Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex. The ε-*directional derivative* of f at $\boldsymbol{x}$ in the direction $\boldsymbol{d} \in \mathbb{R}^n$ is defined by

$$f'_\varepsilon(\boldsymbol{x}; \boldsymbol{d}) = \inf_{t>0} \frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x}) + \varepsilon}{t}. \tag{2.28}$$

Now we can reach the same results as in Theorem 2.26 and Corollary 2.3 also for the ε-directional derivative.

Theorem 2.31 *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function with a Lipschitz constant K at $\boldsymbol{x} \in \mathbb{R}^n$. Then the function $\boldsymbol{d} \mapsto f'_\varepsilon(\boldsymbol{x}; \boldsymbol{d})$ is*

(i) *positively homogeneous and subadditive on $\mathbb{R}^n$ with*

$$|f'_\varepsilon(\boldsymbol{x}; \boldsymbol{d})| \leq K\|\boldsymbol{d}\|,$$

(ii) *convex, its epigraph* epi $f'_\varepsilon(\boldsymbol{x};\ \cdot)$ *is a convex cone and*

$$f'_\varepsilon(\boldsymbol{x}; -\boldsymbol{d}) \geq -f'_\varepsilon(\boldsymbol{x}; \boldsymbol{d}) \quad \textit{for all } \boldsymbol{x} \in \mathbb{R}^n.$$

Proof These results follow immediately from Theorem 2.26, Corollary 2.3 and the fact that for all $\varepsilon > 0$ we have $\inf_{t>0} \varepsilon/t = 0$. $\square$

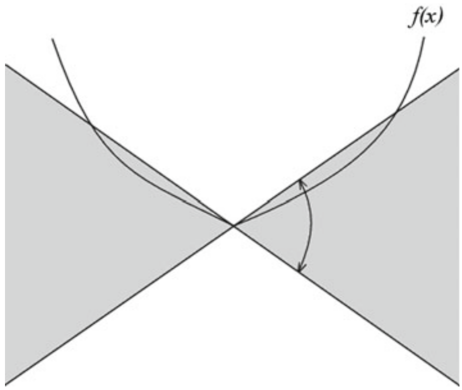


Fig. 2.16 Illustration of ε-subdifferential

As before we now generalize the subgradient and the subdifferential of a convex function. We illustrate the meaning of the definition in Fig. 2.16.

Definition 2.17 Let $\varepsilon \geq 0$, then the ε- *subdifferential of the convex function* $f: \mathbb{R}^n \to \mathbb{R}$ at $\boldsymbol{x} \in \mathbb{R}^n$ is the set

$$\partial_\varepsilon f(\boldsymbol{x}) = \{\boldsymbol{\xi} \in \mathbb{R}^n \mid f(\boldsymbol{x}') \geq f(\boldsymbol{x}) + \boldsymbol{\xi}^T(\boldsymbol{x}' - \boldsymbol{x}) - \varepsilon \quad \text{for all} \quad \boldsymbol{x}' \in \mathbb{R}^n\}. \tag{2.29}$$

Each element $\boldsymbol{\xi} \in \partial_\varepsilon f(\boldsymbol{x})$ is called an ε-*subgradient of* f *at* $\boldsymbol{x}$.

The following summarizes some basic properties of the ε-subdifferential.

Theorem 2.32 *Let* $f: \mathbb{R}^n \to \mathbb{R}$ *be convex function with a Lipschitz constant* K *at* $\boldsymbol{x} \in \mathbb{R}^n$. *Then*

(i) $\partial_0 f(\boldsymbol{x}) = \partial_c f(\boldsymbol{x})$.
(ii) *If* $\varepsilon_1 \leq \varepsilon_2$, *then* $\partial_{\varepsilon_1} f(\boldsymbol{x}) \subseteq \partial_{\varepsilon_2} f(\boldsymbol{x})$.
(iii) $\partial_\varepsilon f(\boldsymbol{x})$ *is a nonempty, convex, and compact set such that* $\partial_\varepsilon f(\boldsymbol{x}) \subseteq B(\boldsymbol{0}; K)$.
(iv) $\partial_\varepsilon f(\boldsymbol{x}) = \{\boldsymbol{\xi} \in \mathbb{R}^n \mid f'_\varepsilon(\boldsymbol{x}; \boldsymbol{d}) \geq \boldsymbol{\xi}^T \boldsymbol{d}$ *for all* $\boldsymbol{d} \in \mathbb{R}^n\}$.
(v) $f'_\varepsilon(\boldsymbol{x}; \boldsymbol{d}) = \max\{\boldsymbol{\xi}^T \boldsymbol{d} \mid \boldsymbol{\xi} \in \partial_\varepsilon f(\boldsymbol{x})\}$ *for all* $\boldsymbol{d} \in \mathbb{R}^n$.

Proof The definition of the ε-subdifferential implies directly the assertions (i) and (ii) and the proofs of assertions (iv) and (v) are the same as for $\varepsilon = 0$ in Theorem 2.28 (i) and (ii), respectively. By Theorem 2.27 $\partial_c f(\boldsymbol{x})$ is nonempty which implies by assertion (i) that $\partial_\varepsilon f(\boldsymbol{x})$ is also nonempty. The proofs of the convexity and compactness are also same as in Theorem 2.27. □

The following shows that the ε-subdifferential contains in a compressed form the subgradient information from the whole neighborhood.

Theorem 2.33 *Let* $f: \mathbb{R}^n \to \mathbb{R}$ *be convex with Lipschitz constant* K *at* $\boldsymbol{x}$. *Then for all* $\varepsilon \geq 0$ *we have*

$$\partial_c f(\boldsymbol{y}) \subseteq \partial_\varepsilon f(\boldsymbol{x}) \qquad \text{for all} \quad \boldsymbol{y} \in B\left(\boldsymbol{x}; \tfrac{\varepsilon}{2K}\right). \tag{2.30}$$

Proof Let $\boldsymbol{\xi} \in \partial_c f(\boldsymbol{y})$ and $\boldsymbol{y} \in B\left(\boldsymbol{x}; \frac{\varepsilon}{2K}\right)$. Then for all $\boldsymbol{z} \in \mathbb{R}^n$ it holds

$$\begin{aligned} f(\boldsymbol{z}) &\geq f(\boldsymbol{y}) + \boldsymbol{\xi}^T(\boldsymbol{z}-\boldsymbol{y}) \\ &= f(\boldsymbol{x}) + \boldsymbol{\xi}^T(\boldsymbol{z}-\boldsymbol{x}) - (f(\boldsymbol{x}) - f(\boldsymbol{y}) + \boldsymbol{\xi}^T(\boldsymbol{z}-\boldsymbol{x}) - \boldsymbol{\xi}^T(\boldsymbol{z}-\boldsymbol{y})) \end{aligned}$$

and, using the Lipschitz condition and Theorem 2.27 we calculate

$$\begin{aligned} |f(\boldsymbol{x}) - f(\boldsymbol{y}) + &\boldsymbol{\xi}^T(\boldsymbol{z}-\boldsymbol{x}) - \boldsymbol{\xi}^T(\boldsymbol{z}-\boldsymbol{y})| \\ &\leq |f(\boldsymbol{x}) - f(\boldsymbol{y})| + |\boldsymbol{\xi}^T(\boldsymbol{z}-\boldsymbol{x}) - \boldsymbol{\xi}^T(\boldsymbol{z}-\boldsymbol{y})| \\ &\leq K\,\|\boldsymbol{x}-\boldsymbol{y}\| + \|\boldsymbol{\xi}\|\,\|\boldsymbol{x}-\boldsymbol{y}\| \\ &\leq 2K\,\|\boldsymbol{x}-\boldsymbol{y}\| \\ &\leq 2K \cdot \frac{\varepsilon}{2K} = \varepsilon, \end{aligned}$$

which gives $\boldsymbol{\xi} \in \partial_\varepsilon f(\boldsymbol{x})$. □

2.3 Links Between Geometry and Analysis

In this section we are going to show that the analytical and geometrical concepts defined in the previous sections are actually equivalent. We have already showed that the level sets of a convex function are convex, the epigraph of the directional derivative is a convex cone and a function is convex if and only if its epigraph is convex. In what follows we give some more connections, on the one hand, between directional derivatives and contingent cones, and on the other hand, between subdifferentials and normal cones in terms of epigraph, level sets and the distance function.

2.3.1 Epigraphs

The next two theorems describe how one could equivalently define tangents and normals by using the epigraph of a convex function (see Figs. 2.17 and 2.18). First result show that the contingent cone of the epigraph is the epigraph of the directional derivative.

Theorem 2.34 *If the function $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex, then*

$$K_{\text{epi } f}(\boldsymbol{x}, f(\boldsymbol{x})) = \text{epi } f'(\boldsymbol{x};\, \cdot). \tag{2.31}$$

Proof Suppose first that $(\boldsymbol{d}, r) \in K_{\text{epi } f}(\boldsymbol{x}, f(\boldsymbol{x}))$. By the definition of the contingent cone there exist sequences $(\boldsymbol{d}_j, r_j) \to (\boldsymbol{d}, r)$ and $t_j \downarrow 0$ such that

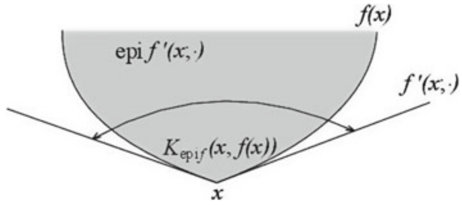


Fig. 2.17 Contingent cone of the epigraph

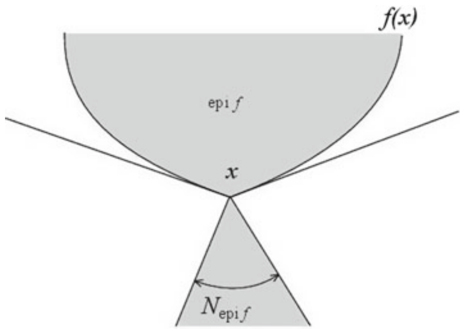


Fig. 2.18 Normal cone of the epigraph

$$(\boldsymbol{x}, f(\boldsymbol{x})) + t_j(\boldsymbol{d}_j, r_j) \in \text{epi } f \quad \text{for all} \quad j \in \mathbb{N},$$

in other words

$$f(\boldsymbol{x} + t_j\boldsymbol{d}_j) \leq f(\boldsymbol{x}) + t_j r_j.$$

Now by using (2.20) we can calculate

$$\begin{aligned} f'(\boldsymbol{x}; \boldsymbol{d}) &= \inf_{t>0} \frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})}{t} \\ &= \lim_{j\to\infty} \frac{f(\boldsymbol{x} + t_j\boldsymbol{d}_j) - f(\boldsymbol{x})}{t_j} \\ &\leq \lim_{j\to\infty} r_j = r, \end{aligned}$$

which implies that $(\boldsymbol{d}, r) \in \text{epi } f'(\boldsymbol{x};\ \cdot)$.

Suppose, next, that $(\boldsymbol{d}, r) \in \text{epi } f'(\boldsymbol{x};\ \cdot)$, which means that

$$f'(\boldsymbol{x}; \boldsymbol{d}) = \lim_{t\downarrow 0} \frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})}{t} \leq r.$$

Then there exists a sequence $t_j \downarrow 0$ such that

$$\frac{f(\boldsymbol{x} + t_j\boldsymbol{d}) - f(\boldsymbol{x})}{t_j} \leq r + \frac{1}{j},$$

which yields

$$f(\boldsymbol{x} + t_j \boldsymbol{d}) \le f(\boldsymbol{x}) + t_j (r + \frac{1}{j})$$

and thus $(\boldsymbol{x}, f(\boldsymbol{x})) + t_j(\boldsymbol{d}, r + \frac{1}{j}) \in \text{epi } f$. This and the fact that $(\boldsymbol{d}, r + \frac{1}{j}) \to (\boldsymbol{d}, r)$ shows that $(\boldsymbol{d}, r) \in K_{\text{epi } f}(\boldsymbol{x}, f(\boldsymbol{x}))$ and we obtain the desired conclusion. □

Next we show that the subgradient is essentially a normal vector of the epigraph.

Theorem 2.35 *If the function* $f \colon \mathbb{R}^n \to \mathbb{R}$ *is convex, then*

$$\partial_c f(\boldsymbol{x}) = \{\boldsymbol{\xi} \in \mathbb{R}^n \mid (\boldsymbol{\xi}, -1) \in N_{\text{epi } f}(\boldsymbol{x}, f(\boldsymbol{x}))\}. \tag{2.32}$$

Proof By Theorem 2.28 (i) we know that $\boldsymbol{\xi} \in \partial_c f(\boldsymbol{x})$ if and only if for any $\boldsymbol{d} \in \mathbb{R}^n$ we have $f'(\boldsymbol{x}; \boldsymbol{d}) \ge \boldsymbol{\xi}^T \boldsymbol{d}$. This is equivalent to the condition that for any $\boldsymbol{d} \in \mathbb{R}^n$ and $r \ge f'(\boldsymbol{x}; \boldsymbol{d})$ we have $r \ge \boldsymbol{\xi}^T \boldsymbol{d}$, that is, for any $\boldsymbol{d} \in \mathbb{R}^n$ and $r \ge f'(\boldsymbol{x}; \boldsymbol{d})$ we have

$$(\boldsymbol{\xi}, -1)^T (\boldsymbol{d}, r) \le 0.$$

By the definition of the epigraph and Theorem 2.34 we have $(\boldsymbol{d}, r) \in \text{ epi } f'(\boldsymbol{x}; \cdot) = K_{\text{epi } f}(\boldsymbol{x}; f(\boldsymbol{x}))$. This and the last inequality means, by the definition of the normal cone, that $(\boldsymbol{\xi}, -1)$ lies in $N_{\text{epi } f}(\boldsymbol{x}; f(\boldsymbol{x}))$. □

2.3.2 Level Sets

In the following theorem we give the relationship between the directional derivative and the contingent cone via the level sets.

Theorem 2.36 *If the function* $f \colon \mathbb{R}^n \to \mathbb{R}$ *is convex, then*

$$K_{\text{lev}_{f(\boldsymbol{x})} f}(\boldsymbol{x}) \subseteq \text{lev}_0 f'(\boldsymbol{x}; \cdot). \tag{2.33}$$

If, in addition, $\boldsymbol{0} \notin \partial_c f(\boldsymbol{x})$*, then*

$$K_{\text{lev}_{f(\boldsymbol{x})} f}(\boldsymbol{x}) = \text{lev}_0 f'(\boldsymbol{x}; \cdot). \tag{2.34}$$

Proof Suppose first that $\boldsymbol{d} \in K_{\text{lev}_{f(\boldsymbol{x})} f}(\boldsymbol{x})$. By the definition of the contingent cone there exist sequences $\boldsymbol{d}_j \to \boldsymbol{d}$ and $t_j \downarrow 0$ such that

$$\boldsymbol{x} + t_j \boldsymbol{d}_j \in \text{lev}_{f(\boldsymbol{x})} f \quad \text{for all} \quad j \in \mathbb{N},$$

in other words

$$f(\boldsymbol{x} + t_j \boldsymbol{d}_j) \le f(\boldsymbol{x}).$$

Now by using (2.20) we can calculate

$$\begin{aligned} f'(\boldsymbol{x}; \boldsymbol{d}) &= \inf_{t>0} \frac{f(\boldsymbol{x}+t\boldsymbol{d}) - f(\boldsymbol{x})}{t} \\ &= \lim_{j\to\infty} \frac{f(\boldsymbol{x}+t_j\boldsymbol{d}_j) - f(\boldsymbol{x})}{t_j} \\ &\le \lim_{j\to\infty} r_j = r, \end{aligned}$$

which implies that $\boldsymbol{d} \in \text{lev}_0\, f'(\boldsymbol{x};\ \cdot)$.

Suppose, next, that $\boldsymbol{0} \notin \partial_c f(\boldsymbol{x})$ and $\boldsymbol{d} \in \text{lev}_0\, f'(\boldsymbol{x};\ \cdot)$, which means that $f'(\boldsymbol{x}; \boldsymbol{d}) \le 0$. Since $\boldsymbol{0} \notin \partial_c f(\boldsymbol{x})$ by Theorem 2.28 (i) we have

$$f'(\boldsymbol{x}; \boldsymbol{d}) = \lim_{t\downarrow 0} \frac{f(\boldsymbol{x}+t\boldsymbol{d}) - f(\boldsymbol{x})}{t} < 0.$$

Then there exists a sequence $t_j \downarrow 0$ such that

$$\frac{f(\boldsymbol{x}+t_j\boldsymbol{d}) - f(\boldsymbol{x})}{t_j} \le 0,$$

which yields

$$f(\boldsymbol{x}+t_j\boldsymbol{d}) \le f(\boldsymbol{x})$$

and thus $\boldsymbol{x} + t_j\boldsymbol{d} \in \text{lev}_{f(\boldsymbol{x})}\, f$. This means that $\boldsymbol{d} \in K_{\text{lev}_{f(\boldsymbol{x})} f}(\boldsymbol{x})$ and the proof is complete. □

Next theorem shows the connection between subgradients and normal vectors of the level sets.

Theorem 2.37 *If the function $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex, then*

$$N_{\text{lev}_{f(\boldsymbol{x})} f}(\boldsymbol{x}) \supseteq \text{ray}\ \partial_c f(\boldsymbol{x}).$$

If, in addition, $\boldsymbol{0} \notin \partial_c f(\boldsymbol{x})$, then

$$N_{\text{lev}_{f(\boldsymbol{x})} f}(\boldsymbol{x}) = \text{ray}\ \partial_c f(\boldsymbol{x}).$$

Proof If $\boldsymbol{z} \in \text{ray}\ \partial_c f(\boldsymbol{x})$ then $\boldsymbol{z} = \lambda\boldsymbol{\xi}$, where $\lambda \ge 0$ and $\boldsymbol{\xi} \in \partial_c f(\boldsymbol{x})$. Let now $\boldsymbol{d} \in K_{\text{lev}_{f(\boldsymbol{x})} f}(\boldsymbol{x})$, which means due to Theorem 2.36 that $\boldsymbol{d} \in \text{lev}_0\, f'(\boldsymbol{x};\ \cdot)$. Then using Theorem 2.28 (i) we get

$$\boldsymbol{z}^T\boldsymbol{d} = \lambda\boldsymbol{\xi}^T\boldsymbol{d} \le \lambda f'(\boldsymbol{x}; \boldsymbol{d}) \le 0,$$

in other words $\boldsymbol{z} \in N_{\text{lev}_{f(\boldsymbol{x})} f}(\boldsymbol{x})$.

Suppose next that $\boldsymbol{0} \notin \partial_c f(\boldsymbol{x})$ and there exists $\boldsymbol{z} \in N_{\text{lev}_{f(\boldsymbol{x})} f}(\boldsymbol{x})$ such that $\boldsymbol{z} \notin \text{ray}\ \partial_c f(\boldsymbol{x})$. According to Theorem 2.27 $\partial_c f(\boldsymbol{x})$ is a convex and compact set. Since $\boldsymbol{0} \notin \partial_c f(\boldsymbol{x})$ Theorems 2.11 and 2.12 implies that ray $\partial_c f(\boldsymbol{x})$ is closed and convex, respectively. As a cone it is nonempty since $\mathbf{0} \in \text{ray } \partial_c f(\boldsymbol{x})$. Then by Theorem 2.4 there exists a hyperplane separating $\{\boldsymbol{z}\}$ and $\text{ray } \partial_c f(\boldsymbol{x})$, in other words there exist $\boldsymbol{p} \neq \mathbf{0}$ and $\alpha \in \mathbb{R}$ such that

$$\boldsymbol{y}^T \boldsymbol{p} \leq \alpha \quad \text{for all} \quad \boldsymbol{y} \in \text{ray } \partial_c f(\boldsymbol{x}) \tag{2.35}$$

and

$$\boldsymbol{z}^T \boldsymbol{p} > \alpha. \tag{2.36}$$

Since $\text{ray } \partial_c f(\boldsymbol{x})$ is cone the components of $\boldsymbol{y}$ can be chosen as large as possible in (2.35), thus $\alpha \leq 0$. On the other hand $\mathbf{0} \in \text{ray } \partial_c f(\boldsymbol{x})$ implying $\alpha \geq \boldsymbol{p}^T \mathbf{0} = 0$, thus $\alpha = 0$. Since $\partial_c f(\boldsymbol{x}) \subseteq \text{ray } \partial_c f(\boldsymbol{x})$ Theorem 2.28 (ii) and (2.35) imply

$$f'(\boldsymbol{x}; \boldsymbol{p}) = \max_{\boldsymbol{\xi} \in \partial_c f(\boldsymbol{x})} \boldsymbol{\xi}^T \boldsymbol{p} \leq \max_{\boldsymbol{y} \in \text{ray } \partial_c f(\boldsymbol{x})} \boldsymbol{y}^T \boldsymbol{p} \leq 0.$$

This means that $\boldsymbol{p} \in \text{lev}_0 f'(\boldsymbol{x}; \cdot)$ and thus due to Theorem 2.36 we have $\boldsymbol{p} \in K_{\text{lev}_{f(\boldsymbol{x})} f}(\boldsymbol{x})$. Since $\boldsymbol{z} \in N_{\text{lev}_{f(\boldsymbol{x})} f}(\boldsymbol{x})$ it follows from the definition of the normal cone that

$$\boldsymbol{z}^T \boldsymbol{p} \leq 0$$

contradicting with inequality (2.36). Thus, $\boldsymbol{z} \in \text{ray } \partial_c f(\boldsymbol{x})$ and the theorem is proved. □

2.3.3 Distance Function

Finally we study the third link between analysis and geometry, namely the distance function defined by (2.2). First we give some important properties of the distance function.

Theorem 2.38 *If $S \subseteq \mathbb{R}^n$ is a nonempty set, then the distance function d_S is Lipschitz continuous with constant $K = 1$, in other words*

$$|d_S(\boldsymbol{x}) - d_S(\boldsymbol{y})| \leq \|\boldsymbol{x} - \boldsymbol{y}\| \quad \text{for all} \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n. \tag{2.37}$$

If in addition the set S is convex then the function d_S is also convex.

Proof Let any $\varepsilon > 0$ and $\boldsymbol{y} \in \mathbb{R}^n$ be given. By definition, there exists a point $\boldsymbol{z} \in S$ such that

$$d_S(\boldsymbol{y}) \geq \|\boldsymbol{y} - \boldsymbol{z}\| - \varepsilon.$$

Now we have

$$\begin{aligned} d_S(\boldsymbol{x}) &\le \|\boldsymbol{x}-\boldsymbol{z}\| \le \|\boldsymbol{x}-\boldsymbol{y}\| + \|\boldsymbol{y}-\boldsymbol{z}\| \\ &\le \|\boldsymbol{x}-\boldsymbol{y}\| + d_S(\boldsymbol{y}) + \varepsilon \end{aligned}$$

which establishes the Lipschitz condition as $\varepsilon > 0$ is arbitrary.

Suppose now that S is a convex set and let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $\lambda \in [0, 1]$ and $\varepsilon > 0$ be given. Choose points $\boldsymbol{z}_{\boldsymbol{x}}, \boldsymbol{z}_{\boldsymbol{y}} \in S$ such that

$$\|\boldsymbol{z}_{\boldsymbol{x}} - \boldsymbol{x}\| \le d_S(\boldsymbol{x}) + \varepsilon \quad \text{and} \quad \|\boldsymbol{z}_{\boldsymbol{y}} - \boldsymbol{x}\| \le d_S(\boldsymbol{y}) + \varepsilon$$

and define $\boldsymbol{z} := (1-\lambda)\boldsymbol{z}_{\boldsymbol{x}} + \lambda \boldsymbol{z}_{\boldsymbol{y}} \in S$. Then

$$\begin{aligned} d_S\big((1-\lambda)\boldsymbol{x} + \lambda\boldsymbol{y}\big) &\le \|\boldsymbol{c} - (1-\lambda)\boldsymbol{x} - \lambda\boldsymbol{y}\| \\ &\le (1-\lambda)\|\boldsymbol{z}_{\boldsymbol{x}} - \boldsymbol{x}\| + \lambda\|\boldsymbol{z}_{\boldsymbol{y}} - \boldsymbol{y}\| \\ &\le (1-\lambda)d_S(\boldsymbol{x}) + \lambda d_S(\boldsymbol{y}) + \varepsilon. \end{aligned}$$

Since ε is arbitrary, d_S is convex. □

Lemma 2.6 *If $S \subseteq \mathbb{R}^n$ is closed, then*

$$\boldsymbol{x} \in S \iff d_S(\boldsymbol{x}) = 0. \tag{2.38}$$

Proof Let $\boldsymbol{x} \in Z$ be arbitrary. Then

$$0 \le d_S(\boldsymbol{x}) \le \|\boldsymbol{x} - \boldsymbol{x}\| = 0$$

and thus $d_S(\boldsymbol{x}) = 0$.

On the other hand if $d_S(\boldsymbol{x}) = 0$, then there exists a sequence $(\boldsymbol{y}_j) \subset S$ such that

$$\|\boldsymbol{x} - \boldsymbol{y}_j\| < 1/j \longrightarrow 0, \qquad \text{when } j \to \infty.$$

Thus, the sequence $(\boldsymbol{y}_j)$ converges to $\boldsymbol{x}$ and $\boldsymbol{x} \in \operatorname{cl} S = S$. □

The next two theorems show how one could equivalently define tangents and normals by using the distance function.

Theorem 2.39 *The contingent cone of the convex set S at $\boldsymbol{x} \in S$ can also be written as*

$$K_S(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{R}^n \mid d'_S(\boldsymbol{x}; \boldsymbol{y}) = 0\}. \tag{2.39}$$

Proof Let $Z := \{\boldsymbol{y} \in \mathbb{R}^n \mid d'_S(\boldsymbol{x}; \boldsymbol{y}) = 0\}$ and let $\boldsymbol{y} \in K_S(\boldsymbol{x})$ be arbitrary. Then there exist sequences $(\boldsymbol{y}_j) \subset \mathbb{R}^n$ and $(t_j) \subset \mathbb{R}$ such that $\boldsymbol{y}_j \to \boldsymbol{y}$, $t_j \downarrow 0$ and $\boldsymbol{x} + t_j\boldsymbol{y}_j \in S$ for all $j \in \mathbb{N}$. It is evident that $d'_S(\boldsymbol{x}; \boldsymbol{y})$ is always nonnegative thus it suffices to show that $d'_S(\boldsymbol{x}; \boldsymbol{y}) \leq 0$. Since $\boldsymbol{x} \in S$ we have

$$\begin{aligned} d'_S(\boldsymbol{x}; \boldsymbol{y}) &= \lim_{t\downarrow 0} \frac{d_S(\boldsymbol{x} + t\boldsymbol{y}) - d_S(\boldsymbol{x})}{t} \\ &= \lim_{t\downarrow 0} \frac{\inf_{\boldsymbol{z}\in S}\{\|\boldsymbol{x} + t\boldsymbol{y} - \boldsymbol{z}\|\}}{t} \\ &\leq \lim_{t\downarrow 0} \frac{\inf_{\boldsymbol{z}\in S}\{\|\boldsymbol{x} + t\boldsymbol{y}_j - \boldsymbol{z}\|\} + \|t\boldsymbol{y} - t\boldsymbol{y}_j\|}{t} \end{aligned}$$

and

$$\inf_{\boldsymbol{z}\in S}\{\|\boldsymbol{x} + t\boldsymbol{y}_j - \boldsymbol{z}\|\} = \inf_{\boldsymbol{z}\in S}\{\|(1 - \frac{t}{t_j})\boldsymbol{x} + \frac{t}{t_j}(\boldsymbol{x} + t_j\boldsymbol{y}_j) - \boldsymbol{c}\|\}.$$

Since $\boldsymbol{x} \in S$, $\boldsymbol{x} + t_j\boldsymbol{y}_j \in S$ and $t/t_j \in [0, 1]$ whenever $0 \leq t \leq t_j$, the convexity of S implies that

$$(1 - \frac{t}{t_j})\boldsymbol{x} + \frac{t}{t_j}(\boldsymbol{x} + t_j\boldsymbol{y}_j) \in S,$$

and thus

$$\inf_{\boldsymbol{z}\in S} \|(1 - \frac{t}{t_j})\boldsymbol{x} + \frac{t}{t_j}(\boldsymbol{x} + t_j\boldsymbol{y}_j) - \boldsymbol{z}\| = 0.$$

Therefore

$$d'_S(\boldsymbol{x}; \boldsymbol{y}) \leq t\|\boldsymbol{y} - \boldsymbol{y}_j\| \longrightarrow 0,$$

when $j \to \infty$. Thus $d'_S(\boldsymbol{x}; \boldsymbol{y}) = 0$ and $K_S(\boldsymbol{x}) \subseteq Z$.

For the converse let $\boldsymbol{y} \in Z$ and $(t_j) \subset \mathbb{R}$ be such that $t_j \downarrow 0$. By the definition of Z we get

$$d'_S(\boldsymbol{x}; \boldsymbol{y}) = \lim_{t_j\downarrow 0} \frac{d_S(\boldsymbol{x} + t_j\boldsymbol{y})}{t_j} = 0.$$

By the definition of d_S we can choose points $\boldsymbol{z}_j \in S$ such that

$$\|\boldsymbol{x} + t_j\boldsymbol{y} - \boldsymbol{z}_j\| \leq d_S(\boldsymbol{x} + t_j\boldsymbol{y}) + \frac{t_j}{j}.$$

By setting

$$\boldsymbol{y}_j := \frac{\boldsymbol{z}_j - \boldsymbol{x}}{t_j},$$

we have

$$\boldsymbol{x} + t_j\boldsymbol{y}_j = \boldsymbol{x} + t_j\frac{\boldsymbol{z}_j - \boldsymbol{x}}{t_j} = \boldsymbol{z}_j \in S$$

and

$$\begin{aligned}\|\boldsymbol{y}-\boldsymbol{y}_j\| &= \|\boldsymbol{y}-\frac{\boldsymbol{z}_j-\boldsymbol{x}}{t_j}\| \\ &= \frac{\|\boldsymbol{x}+t_j\boldsymbol{y}-\boldsymbol{z}_j\|}{t_j} \\ &\le \frac{d_S(\boldsymbol{x}+t_j\boldsymbol{y})}{t_j}+\frac{1}{j} \longrightarrow 0,\end{aligned}$$

as $j \to \infty$. Thus $\boldsymbol{y} \in K_S(\boldsymbol{x})$ and $Z = K_S(\boldsymbol{x})$. □

Theorem 2.40 *The normal cone of the convex set S at $\boldsymbol{x} \in S$ can also be written as*

$$N_S(\boldsymbol{x}) = \text{cl ray } \partial_c d_S(\boldsymbol{x}). \tag{2.40}$$

Proof First, let $\boldsymbol{z} \in \partial_c d_S(\boldsymbol{x})$. Then by Theorem 2.28 (i)

$$\boldsymbol{z}^T\boldsymbol{y} \le d'_S(\boldsymbol{x};\boldsymbol{y}) \quad \text{for all} \quad \boldsymbol{y} \in \mathbb{R}^n.$$

If one has $\boldsymbol{y} \in K_S(\boldsymbol{x})$ then by Theorem 2.39 $d'_S(\boldsymbol{x};\boldsymbol{y}) = 0$. Thus $\boldsymbol{z}^T\boldsymbol{y} \le 0$ for all $\boldsymbol{y} \in K_S(\boldsymbol{x})$ which implies that $\boldsymbol{z} \in N_S(\boldsymbol{x})$. By Theorem 2.27 $\partial_c d_S(\boldsymbol{x})$ is a convex set and then by Theorem 2.11 ray $\partial_c d_S(\boldsymbol{x})$ is a convex cone. Furthermore, by Theorem 2.10 ray $\partial_c d_S(\boldsymbol{x})$ is the smallest cone containing $\partial_c d_S(\boldsymbol{x})$. Then, because $N_S(\boldsymbol{x})$ is also a convex cone (Theorem 2.18), we have

$$\text{ray } \partial_c d_S(\boldsymbol{x}) \subseteq N_S(\boldsymbol{x}).$$

On the other hand, if $N_S(\boldsymbol{x}) = \{\boldsymbol{0}\}$ we have clearly $N_S(\boldsymbol{x}) \subseteq \text{ray } \partial_c d_S(\boldsymbol{x})$. Suppose next that $N_S(\boldsymbol{x}) \ne \{\boldsymbol{0}\}$ and let $\boldsymbol{z} \in N_S(\boldsymbol{x})\setminus\{\boldsymbol{0}\}$ be arbitrary. Since S is convex due to Theorem 2.19 we have

$$\boldsymbol{z}^T(\boldsymbol{y}-\boldsymbol{x}) \le 0 \quad \text{for all} \quad \boldsymbol{y} \in S$$

and hence $S \subseteq H^-(\boldsymbol{z}, \boldsymbol{z}^T\boldsymbol{x})$. Since $d_S(\boldsymbol{y}) \ge 0$ for all $\boldsymbol{y} \in \mathbb{R}^n$ we have

$$\lambda\boldsymbol{z}^T(\boldsymbol{y}-\boldsymbol{x}) \le 0 \le d_S(\boldsymbol{y}) \quad \text{for all} \quad \boldsymbol{y} \in H^-(\boldsymbol{z}, \boldsymbol{z}^T\boldsymbol{x}) \text{ and } \lambda \ge 0.$$

Suppose next that $\boldsymbol{y} \in H^+(\boldsymbol{z}, \boldsymbol{z}^T\boldsymbol{x})$. Since $S \subseteq H^-(\boldsymbol{z}, \boldsymbol{z}^T\boldsymbol{x})$ we have clearly $d_{H^-(\boldsymbol{z},\boldsymbol{z}^T\boldsymbol{x})}(\boldsymbol{y}) \le d_S(\boldsymbol{y})$ for all $\boldsymbol{y} \in \mathbb{R}^n$. On the other hand (see Exercise 2.3)

$$d_{H^-(\boldsymbol{z},\boldsymbol{z}^T\boldsymbol{x})}(\boldsymbol{y}) = \frac{1}{\|\boldsymbol{z}\|}\boldsymbol{z}^T(\boldsymbol{y}-\boldsymbol{x}) \quad \text{for all} \quad \boldsymbol{y} \in H^+(\boldsymbol{z}, \boldsymbol{z}^T\boldsymbol{x}).$$

Thus, for any $\boldsymbol{y} \in \mathbb{R}^n = H^-(\boldsymbol{z}, \boldsymbol{z}^T\boldsymbol{x}) \cup H^+(\boldsymbol{z}, \boldsymbol{z}^T\boldsymbol{x})$ we have

$$\frac{1}{\|\boldsymbol{z}\|}\boldsymbol{z}^T(\boldsymbol{y}-\boldsymbol{x}) \le d_S(\boldsymbol{y}) = d_S(\boldsymbol{y}) - d_S(\boldsymbol{x}).$$

Then the definition of subdifferential of convex function and the convexity of d_S imply that

$$\frac{1}{\|\boldsymbol{z}\|}\boldsymbol{z} \in \partial_c d_S(\boldsymbol{x}),$$

thus $N_S(\boldsymbol{x}) \subseteq \text{ray } \partial_c d_S(\boldsymbol{x})$ and the proof is complete. □

Note that since $N_S(\boldsymbol{x})$ is always closed, we deduce that also ray $\partial_c d_S(\boldsymbol{x})$ is closed if S is convex.

2.4 Summary

This chapter contains the basic results from convex analysis. First we have concentrated on geometrical concepts and started by considering convex sets and cones. The main results are the existence of separating and supporting hyperplanes (Theorems 2.4, 2.7 and 2.8). We have defined tangents and normals in the form of contingent and normal cones. Next we moved to analytical concepts and defined subgradients and subdifferentials of convex functions. Finally we showed that everything is one by connecting these geometrical and analytical concepts via epigraphs, level sets and the distance functions. We have proved, for example, that the contingent cone of the epigrapf is the epigrapf of the directional derivative (Theorem 2.34), the contingent cone of the zero level set is zero level set of the directional derivative (Theorem 2.36), and the contigent cone of a convex set consist of the points where the directional derivative of the distance function vanish (Theorem 2.39).

Exercises

2.1 Show that open and closed balls and halfspaces are convex sets.

2.2 (Lemma 2.1) Prove that if $S \subseteq \mathbb{R}^n$, then conv S is a convex set and S is convex if and only if $S = \text{conv } S$.

2.3 Let $\boldsymbol{p} \in \mathbb{R}^n$, $\boldsymbol{p} \neq \boldsymbol{0}$ and $\alpha \in \mathbb{R}$. Prove that

$$d_{H^-(\boldsymbol{p},\alpha)}(\boldsymbol{y}) = \frac{1}{\|\boldsymbol{p}\|}(\boldsymbol{p}^T\boldsymbol{y} - \alpha) \quad \text{for all} \quad \boldsymbol{y} \in H^+(\boldsymbol{p},\alpha).$$

2.4 (Farkas' Lemma) Let $A \in \mathbb{R}^{n\times n}$ and $\boldsymbol{c} \in \mathbb{R}^n$. Prove that either

$$A\boldsymbol{x} \le \boldsymbol{0} \quad \text{and} \quad \boldsymbol{c}^T\boldsymbol{x} > 0 \quad \text{for some } \boldsymbol{x} \in \mathbb{R}^n$$

or

$$A^T \boldsymbol{y} = \boldsymbol{c} \quad \text{and} \quad \boldsymbol{y} \geq 0 \quad \text{for some } \boldsymbol{y} \in \mathbb{R}^n.$$

2.5 (Gordan's Lemma) Let $A \in \mathbb{R}^{n \times n}$. Prove that either

$$A\boldsymbol{x} < \boldsymbol{0} \quad \text{and} \quad \boldsymbol{c}^T \boldsymbol{x} > 0 \quad \text{for some } \boldsymbol{x} \in \mathbb{R}^n$$

or

$$A^T \boldsymbol{y} = \boldsymbol{0} \quad \text{and} \quad \boldsymbol{0} \neq \boldsymbol{y} \geq 0 \quad \text{for some } \boldsymbol{y} \in \mathbb{R}^n.$$

2.6 Show that closed halfspaces $H^+(\boldsymbol{p}, 0)$ and $H^-(\boldsymbol{p}, 0)$, the nonnegative orthant $\mathbb{R}^n_+ = \{\boldsymbol{x} \in \mathbb{R}^n \mid x_i \geq 0,\ i = 1 \ldots, n\}$ and halflines starting from the origin are closed convex cones.

2.7 (Lemma 2.3) Prove that if $S \subseteq \mathbb{R}^n$, then ray S is a cone and $C \subseteq \mathbb{R}^n$ is cone if and only if $C =$ ray C.

2.8 (Lemma 2.4) Prove that if $S \subseteq \mathbb{R}^n$, then cone S is a convex cone and $C \subseteq \mathbb{R}^n$ is convex cone if and only if $C =$ cone C.

2.9 (Corollary 2.2) Prove that if $S \subseteq \mathbb{R}^n$, then cone $S =$ conv ray S.

2.10 Show that $S_1 \subseteq S_2$ implies $S_2^\circ \subseteq S_1^\circ$.

2.11 (Lemma 2.5) Prove that if $S \subseteq \mathbb{R}^n$, then S° is a closed convex cone and $S \subseteq S^{\circ\circ}$.

2.12 Specify the sets conv S, ray S, cone S and S° when

(a) $S = \{(1, 1)\}$
(b) $S = \{(1, 1), (1, 2), (2, 1)\}$
(c) $S = \text{int } \mathbb{R}^2_+ \cup \{(0, 0)\}$.

2.13 Let $C \subseteq \mathbb{R}^n$ be a closed convex cone. Show that $K_C(\boldsymbol{0}) = C$.

2.14 (Theorem 2.16) Prove that the cone of global feasible directions $G_S(\boldsymbol{x})$ of the nonempty convex set S at $\boldsymbol{x} \in S$ is a convex cone.

2.15 Let $S \subseteq \mathbb{R}^n$ be convex. Show that $K_S(\boldsymbol{x}) = N_S(\boldsymbol{x})^\circ$.

2.16 Specify the sets $K_{\mathbb{R}^2_+}(\boldsymbol{0})$ and $N_{\mathbb{R}^2_+}(\boldsymbol{0})$.

2.17 Let $S \subseteq \mathbb{R}^n$ be convex and $\boldsymbol{x} \in \text{int } S$. Show that $K_S(\boldsymbol{x}) = \mathbb{R}^n$ and $N_S(\boldsymbol{x}) = \emptyset$.

2.18 Let $S_1, S_2 \subseteq \mathbb{R}^n$ be convex and $\boldsymbol{x} \in S_1 \cap S_2$. Show that

(a) $K_{S_1 \cap S_2}(\boldsymbol{x}) \subseteq K_{S_1}(\boldsymbol{x}) \cap K_{S_2}(\boldsymbol{x})$,

(b) $N_{S_1 \cap S_2}(\boldsymbol{x}) \supseteq N_{S_1}(\boldsymbol{x}) + N_{S_2}(\boldsymbol{x})$.

2.19 (Theorem 2.20) Prove that if S is a nonempty convex set such that $\mathbf{0} \in S$, then

(a) $G_S(\mathbf{0}) = \text{ray } S$,

(b) $K_S(\mathbf{0}) = \text{cl ray } S$,

(c) $N_S(\mathbf{0}) = S^\circ$.

2.20 Show that the function $f: \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) := e^{x^2}$$

is convex.

2.21 By exploiting Exercise 2.20 show that for all $x, y > 0$ we have

$$\frac{x}{4} + \frac{3y}{4} \leq \sqrt{\ln\left(\frac{e^{x^2}}{4} + \frac{3e^{y^2}}{4}\right)}.$$

2.22 (Theorem 2.24) Prove that the function $f: \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its epigraph epi f is a convex set.

2.23 How should the concept of a 'concave set' to be defined?

2.24 (Corollary 2.3) Prove that if $f: \mathbb{R}^n \to \mathbb{R}$ is a convex function, then the function $\boldsymbol{d} \mapsto f'(\boldsymbol{x}; \boldsymbol{d})$ is convex, its epigraph epi $f'(\boldsymbol{x}; \cdot)$ is a convex cone and we have

$$f'(\boldsymbol{x}; -\boldsymbol{d}) \geq -f'(\boldsymbol{x}; \boldsymbol{d}) \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^n.$$

2.25 Show that the function $f: \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) := \max\{|x|, x^2\}$$

is convex. Calculate $f'(1; \pm 1)$ and $\partial_c f(1)$.

2.26 Let $f: \mathbb{R}^2 \to \mathbb{R}$ be such that

$$f(x, y) := \max\{-\max\{-x, y\}, y - x\}.$$

Calculate $\partial_c f(0, 0)$.

2.27 Let $f: \mathbb{R}^n \to \mathbb{R}$ be such that $f(\boldsymbol{x}) := \|\boldsymbol{x}\|$ and $g: \mathbb{R} \to \mathbb{R}$ such that $g(x) := x^2$. Calculate $\partial_c f(0)$ and $\partial_c g(f(0))$.

2.28 Let $f: \mathbb{R}^n \to \mathbb{R}$ be convex. Show that the mapping $\boldsymbol{x} \mapsto \partial_c f(\boldsymbol{x})$ is monotonic, in other words for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ we have

$$(\boldsymbol{\xi_x} - \boldsymbol{\xi_y})^T(\boldsymbol{x} - \boldsymbol{y}) \geq 0 \quad \text{for all} \quad \boldsymbol{\xi_x} \in \partial_c f(\boldsymbol{x}),\ \boldsymbol{\xi_y} \in \partial_c f(\boldsymbol{y}).$$

2.29 Prove that if $f: \mathbb{R}^n \to \mathbb{R}$ is continuous, then $\operatorname{epi} f$ and $\operatorname{lev}_\alpha f$ are closed for all $\alpha \in \mathbb{R}$.

2.30 Let the functions $f_i: \mathbb{R}^n \to \mathbb{R}$ be convex for all $i = 1, \ldots, m$ and define $f: \mathbb{R}^n \to \mathbb{R}$ by

$$f(\boldsymbol{x}) := \max \{f_i(\boldsymbol{x}) \mid i = 1, \ldots, m\}.$$

Show that

(a) $\operatorname{lev} f = \bigcap_{i=1}^m \operatorname{lev} f_i$,

(b) $\operatorname{epi} f = \bigcap_{i=1}^m \operatorname{epi} f_i$.

2.31 Show that the equality does not hold in Theorem 2.36 without the extra assumption $\mathbf{0} \notin \partial_c f(\boldsymbol{x})$. In other words, if the function $f: \mathbb{R}^n \to \mathbb{R}$ is convex, then

$$K_{\operatorname{lev}_{f(\boldsymbol{x})} f}(\boldsymbol{x}) \not\supseteq \operatorname{lev}_0 f'(\boldsymbol{x};\ \cdot).$$

(Hint: Consider the function $f(\boldsymbol{x}) := \|\boldsymbol{x}\|^2$).

2.32 Let $S \subseteq \mathbb{R}^n$ convex and $x \in S$. Show that if $\mathbf{0} \notin \partial_c d_S(\boldsymbol{x})$, then

(a) $K_S(\boldsymbol{x}) = K_{\operatorname{lev}_{d_S(\boldsymbol{x})} d_S}(\boldsymbol{x}) \cap K_{\operatorname{lev}_{-d_S(\boldsymbol{x})} -d_S}(\boldsymbol{x})$,

(a) $N_S(\boldsymbol{x}) = N_{\operatorname{lev}_{d_S(\boldsymbol{x})} d_S}(\boldsymbol{x})$.

2.33 Let

$$S = \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1^2 \leq x_2 \text{ and } |x_1| \leq x_2\}.$$

Calculate $K_S((1, 1))$ and $N_S((1, 1))$.

2.34 Let

$$S = \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1 \leq 2,\ x_1 \geq -2x_2 \text{ and } x_1 \geq 2x_2\}.$$

Calculate $K_S((0, 0))$ and $N_S((0, 0))$.

http://www.springer.com/978-3-319-08113-7

Introduction to Nonsmooth Optimization
Theory, Practice and Software
Bagirov, A.M.; Karmitsa, N.; Mäkelä, M.M.
2014, XVIII, 372 p. 72 illus., Hardcover
ISBN: 978-3-319-08113-7