Chapter 2
Convex Functions

2.1 Definition and Basic Properties

Given a function $f : S \to [-\infty, +\infty]$ on a nonempty set $S \subset \mathbb{R}^n$, the sets

$$\begin{aligned} \mathrm{dom} f &= \{x \in S | \, f(x) < +\infty\} \\ \mathrm{epi} f &= \{(x, \alpha) \in S \times \mathbb{R} | \, f(x) \leq \alpha\} \end{aligned}$$

are called the *effective domain* and the *epigraph* of $f(x)$, respectively. If $\mathrm{dom} f \neq \emptyset$ and $f(x) > -\infty$ for all $x \in S$, then we say that the function $f(x)$ is *proper*.

A function $f : S \to [-\infty, +\infty]$ is called *convex* if its epigraph is a convex set in $\mathbb{R}^n \times R$. This is equivalent to saying that S is a convex set in $\mathbb{R}^n$ and for any $x^1, x^2 \in S$ and $\lambda \in [0, 1]$, we have

$$f((1-\lambda)x^1 + \lambda x^2) \leq (1-\lambda)f(x^1) + \lambda f(x^2) \tag{2.1}$$

whenever the right-hand side is defined. In other words (2.1) must always hold unless $f(x^1) = -f(x^2) = \pm\infty$. By induction it can be proved that if $f(x)$ is convex then for any finite set $x^1, \ldots, x^k \in S$ and any nonnegative numbers $\lambda_1, \ldots, \lambda_k$ summing up to 1, we have

$$f\left(\sum_{i=1}^{k} \lambda_i x^i\right) \leq \sum_{i=1}^{k} \lambda_i f(x^i)$$

whenever the right-hand side is defined. A function $f(x)$ is said to be *concave* on S if $-f(x)$ is convex; *affine* on S if $f(x)$ is finite and both convex and concave. An affine function on $\mathbb{R}^n$ has the form $f(x) = \langle a, x\rangle + \alpha$, with $a \in \mathbb{R}^n, \alpha \in \mathbb{R}$, because its epigraph is a halfspace in $\mathbb{R}^n \times \mathbb{R}$ containing no vertical line.

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6_2

For a given nonempty convex set $C \subset \mathbb{R}^n$ we can define the convex functions:

- the *indicator function* of C : $\delta_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases}$
- the *support function* (see Sect. 1.8) $s_C(x) = \sup_{y \in C} \langle y, x \rangle$
- the *distance function* $d_C(x) = \inf_{y \in C} \|x - y\|$.

The convexity of $\delta_C(x)$ is obvious; that of the two other functions can be verified directly or derived from Propositions 2.5 and 2.9 below.

Proposition 2.1 *If $f(x)$ is an improper convex function on $\mathbb{R}^n$ then $f(x) = -\infty$ at every relative interior point x of its effective domain.*

Proof By the definition of an improper convex function, $f(x^0) = -\infty$ for at least some $x^0 \in \mathrm{dom} f$ (unless $\mathrm{dom} f = \emptyset$). If $x \in \mathrm{ri}(\mathrm{dom} f)$ then there is a point $x' \in \mathrm{dom} f$ such that x is a relative interior point of the line segment $[x^0, x']$. Since $f(x') < +\infty$, it follows from $x = \lambda x^0 + (1 - \lambda)x'$ with $\lambda \in (0, 1)$, that $f(x) \leq \lambda f(x^0) + (1 - \lambda) f(x') = -\infty$. □

From the definition it is straightforward that a function $f(x)$ on $\mathbb{R}^n$ is convex if and only if its restriction to every straight line in $\mathbb{R}^n$ is convex. Therefore, convex functions on $\mathbb{R}^n$ can be characterized via properties of convex functions on the real line.

Theorem 2.1 *A real-valued function $f(x)$ on an open interval $(a, b) \subset \mathbb{R}$ is convex if and only if it is continuous and possesses at every $x \in (a, b)$ finite left and right derivatives*

$$f'_-(x) = \lim_{t \uparrow 0} \frac{f(x + t) - f(x)}{t}, \quad f'_+(x) = \lim_{t \downarrow 0} \frac{f(x + t) - f(x)}{t}$$

such that $f'_+(x)$ is nondecreasing and

$$f'_-(x) \leq f'_+(x), \qquad f'_+(x^1) \leq f'_-(x^2) \textit{ for } x^1 < x^2. \tag{2.2}$$

Proof

(i) Let $f(x)$ be convex. If $0 < s < t$ and $x + t < b$ then the point $(x + s, f(x + s))$ is below the segment joining $(x, f(x))$ and $(x + t, f(x + t))$, so

$$\frac{f(x + s) - f(x)}{s} \leq \frac{f(x + t) - f(x)}{t}. \tag{2.3}$$

This shows that the function $t \mapsto [f(x + t) - f(x)]/t$ is nonincreasing as $t \downarrow 0$. Hence it has a limit $f'_+(x)$ (finite or $= -\infty$). Analogously, $f'_-(x)$ exists (finite or $= +\infty$). Furthermore, setting $y = x + s$, $t = s + r$, we also have

$$\frac{f(x + s) - f(x)}{s} \leq \frac{f(y + r) - f(y)}{r}, \tag{2.4}$$

which implies $f'_+(x) \le f'_+(y)$ for $x < y$, i.e., $f'_+(x)$ is nondecreasing. Finally, writing (2.4) as

$$\frac{f(y-s)-f(y)}{-s} \le \frac{f(y+r)-f(y)}{r}.$$

and letting $-s \uparrow 0, r \downarrow 0$ yields $f'_-(y) \le f'_+(y)$, proving the left part of (2.2) and at the same time the finiteness of these derivatives. The continuity of $f(x)$ at every $x \in (a,b)$ then follows from the existence of finite $f'_-(x)$ and $f'_+(x)$. Furthermore, setting $x = x^1, y + r = x^2$ in (2.4) and letting $s, r \to 0$ yields the right part of (2.2).

(ii) Now suppose that $f(x)$ has all the properties mentioned in the Proposition and let $a < c < d < b$. Consider the function

$$g(x) = f(x) - f(c) - (x-c)\frac{f(d)-f(c)}{d-c}.$$

Since for any $x = (1-\lambda)c + \lambda d$, we have $g(x) = f(x) - f(c) - \lambda[f(d) - f(c)] = f(x) - [(1-\lambda)f(c) + \lambda f(d)]$, to prove the convexity of $f(x)$ it suffices to show that $g(x) \le 0$ for any $x \in [c,d]$. Suppose the contrary, that the maximum of $g(x)$ over the segment $[c,d]$ is positive (this maximum exists because $f(x)$ is continuous). Let $e \in [c,d]$ be the point where this maximum is attained. Note that $g(c) = g(d) = 0$, (hence $c < e < d$) and from its expression, $g(x)$ has the same properties as $f(x)$, namely: $g'_-(x), g'_+(x)$ exist at every $x \in (c,d)$, $g'_-(x) \le g'_+(x)$, $g'_+(x)$ is nondecreasing and $g'_+(x^1) \le g'_-(x^2)$ for $x^1 \le x^2$. Since $g(e) \ge g(x)\ \forall x \in [c,d]$, we must have $g'_-(e) \ge 0 \ge g'_+(e)$, consequently $g'_-(e) = g'_+(e) = 0$, and hence, since $g'_+(x)$ is nondecreasing, $g'_+(x) \ge 0\ \forall x \in [e,d]$. If $g'_-(y) \le 0$ for some $y \in (e,d]$ then $g'_+(x) \le g'_-(y) \le 0$ hence $g'(x) = 0$ for all $x \in [e,y)$, from which it follows that $g(y) = g(e) > 0$. Since $g(d) = 0$, there must exist $y \in (e,d)$ with $g'_-(y) > 0$. Let $x^1 \in [y,d)$ be the point where $g(x)$ attains its maximum over the segment $[y,d]$. Then $g'_+(x^1) \le 0$, contradicting $g'_+(y) \ge g'_-(y) > 0$. Therefore $g(x) \le 0$ for all $x \in [c,d]$, as was to be proved. □

Corollary 2.1 *A differentiable real-valued function $f(x)$ on an open interval is convex if and only if its derivative f' is a nondecreasing function. A twice differentiable real-valued function $f(x)$ on an open interval is convex if and only if its second derivative f'' is nonnegative throughout this interval.* □

Proposition 2.2 *A twice differentiable real-valued function $f(x)$ on an open convex set C in $\mathbb{R}^n$ is convex if and only if for every $x \in C$ its Hessian matrix*

$$Q_x = (q_{ij}(x)), \quad q_{ij}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \dots, x_n)$$

is positive semidefinite, i.e.,

$$\langle u, Q_x u\rangle \geq 0 \ \forall u \in \mathbb{R}^n.$$

Proof The function f is convex on C if and only if for each $a \in C$ and $u \in \mathbb{R}^n$ the function $\varphi_{a,u}(t) = f(a+tu)$ is convex on the open real interval $\{t|\ a+tu \in C\}$. The proposition then follows from the preceding corollary since an easy computation yields $\varphi''(t) = \langle u, Q_x u\rangle$ with $x = a + tu$. □

In particular, a *quadratic function*

$$f(x) = \frac{1}{2}\langle x, Qx\rangle + \langle x, a\rangle + \alpha,$$

where Q is a symmetric $n \times n$ matrix, is convex on $\mathbb{R}^n$ if and only if Q is positive semidefinite. It is concave on $\mathbb{R}^n$ if and only if its matrix Q is negative semidefinite.

Proposition 2.3 *A proper convex function f on $\mathbb{R}^n$ is continuous at every interior point of its effective domain.*

Proof Let $x^0 \in \mathrm{int}(\mathrm{dom}f)$. Without loss of generality one can assume $x^0 = 0$. By Theorem 2.1, for each $i = 1,\dots,n$ the restriction of f to the open interval $\{t|\ x^0 + te^i \in \mathrm{int}(\mathrm{dom}f)\}$ is continuous relative to this interval. Hence for any given $\varepsilon > 0$ and for each $i = 1,\dots,n$, we can select $\delta_i > 0$ so small that $|f(x) - f(x^0)| \leq \varepsilon$ for all $x \in [-\delta_i e^i, +\delta_i e^i]$. Let $\delta = \min\{\delta_i|\ i = 1,\dots,n\}$ and $B = \{x|\ \|x\|_1 \leq \delta\}$. Denote $u^i = \delta e^i, u^{i+n} = -\delta e^i,\ i = 1,\dots,n$. Then, as seen in the proof of Corollary 1.6, any $x \in B$ is of the form $x = \sum_{i=1}^{2n}\lambda_i u^i$, with $\sum_{i=1}^{2n}\lambda_i = 1, 0 \leq \lambda_i \leq 1$, hence $f(x) \leq \sum_{i=1}^{2n}\lambda_i f(u^i)$, and consequently, $f(x) - f(x^0) \leq \sum_{i=1}^{2n}\lambda_i[f(x^i) - f(x^0)]$. Therefore,

$$|f(x) - f(x^0)| \leq \sum_{i=1}^{2n}\lambda_i|f(x) - f(x^0)| \leq \varepsilon$$

for all $x \in B$, proving the continuity of $f(x)$ at x^0. □

Proposition 2.4 *Let f be a real-valued function on a convex set $C \subset \mathbb{R}^n$. If for every $x \in C$ there exists a convex open neighborhood U_x of x such that f is convex on $U_x \cap C$ then f is convex on C.*

Proof It suffices to show that for every $a \in C, u \in \mathbb{R}^n$, the function $\varphi(t) = f(a+tu)$ is convex on the interval $\Delta := \{t|\ a + tu \in C\}$. But from the hypothesis, this function is convex in a neighborhood of every $t \in \Delta$, hence is continuous and has left and right derivatives $\varphi'_-(t) \leq \varphi'_+(t)$ which are non decreasing in a neighborhood of every $t \in \Delta$. These derivatives thus exist and satisfy the conditions described in Theorem 2.1 on the whole interval Δ. Hence, $\varphi(t)$ is convex. □

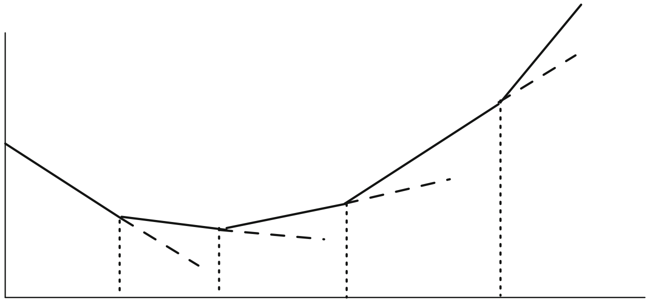

Fig. 2.1 Convex piecewise affine function

For example, let $f(x) : \mathbb{R} \to \mathbb{R}$ be a piecewise convex function on the real line, i.e., a function such that the real line can be partitioned into a finite number of intervals $\Delta_i, i = 1, \ldots, N$, in each of which $f(x)$ is convex. Then $f(x)$ is convex if and only if it is convex in the neighborhood of each breakpoint (endpoint of some interval Δ_i). In particular, a piecewise affine function on the real line is convex if and only if at each breakpoint the left slope is at most equal to the right slope (in other words, the sequence of slopes is nondecreasing) (see Fig. 2.1).

2.2 Operations That Preserve Convexity

A function with a complicated expression may be built up from a number of simpler ingredient functions via certain standard operations. The convexity of such a function can often be established indirectly, by proving that the ingredient functions are known convex functions, whereas the operations involved in the composition of the ingredient functions preserve convexity. It is therefore useful to be familiar with some of the most important operations which preserve convexity.

Proposition 2.5 *A positive combination of finitely many proper convex functions on $\mathbb{R}^n$ is convex. The upper envelope (pointwise supremum) of an arbitrary family of convex functions is convex.*

Proof If $f(x)$ is convex and $\alpha \geq 0$, then $\alpha f(x)$ is obviously convex. If f_1 and f_2 are proper convex functions on $\mathbb{R}^n$, then it is also evident that $f_1 + f_2$ is convex. This proves the first part of the proposition. The second part follows from the facts that if $f(x) = \sup\{f_i(x) |\ i \in I\}$, then $\operatorname{epi} f = \cap_{i \in I} \operatorname{epi} f_i$, and the intersection of a family of convex sets is a convex set. $\square$

Proposition 2.6 *Let Ω be a convex set in $\mathbb{R}^n$, G a convex set in $\mathbb{R}^m$, $\varphi(x, y)$ a real-valued convex function on $\Omega \times G$. Then the function*

$$f(x) = \inf_{y \in G} \varphi(x, y)$$

is convex on Ω.

Proof Let $x^1, x^2 \in \Omega$ and $x = \lambda x^1 + (1-\lambda)x^2$ with $\lambda \in [0,1]$. For each $i = 1,2$ select a sequence $\{y^{i,k}\} \subset G$ such that

$$\varphi(x^i, y^{i,k}) \to \inf_{y \in G} \varphi(x^i, y).$$

By convexity of φ,

$$f(x) \le \varphi(x, \lambda y^{1,k} + (1-\lambda)y^{2,k}) \le \lambda\varphi(x^1, y^{1,k}) + (1-\lambda)\varphi(x^2, y^{2,k}),$$

hence, letting $k \to \infty$ yields

$$f(x) \le \lambda f(x^1) + (1-\lambda) f(x^2).$$

□

Proposition 2.7 *If $g_i(x), i = 1,\dots,m$, are concave positive functions on a convex set $C \subset \mathbb{R}^n$ then their geometric mean*

$$f(x) = \left[\prod_{i=1}^m g_i(x)\right]^{1/m} \tag{2.5}$$

is a concave function (so $-f(x)$ is a convex function) on C.

Proof Let $T = \{t \in \mathbb{R}^m_+ |\ \prod_{i=1}^m t_i \ge 1\}$. We show that for any fixed $x \in C$:

$$\left[\prod_{i=1}^m g_i(x)\right]^{1/m} = \frac{1}{m}\min_{t\in T}\left\{\sum_{i=1}^m t_i g_i(x)\right\}. \tag{2.6}$$

Indeed, observe that the right-hand side of (2.6) is equal to

$$\frac{1}{m}\min\left\{\sum_{i=1}^m t_i g_i(x) |\ \prod_{i=1}^m t_i = 1,\ t_i > 0,\ i = 1,\dots,m\right\}$$

since if $\prod_{i=1}^m t_i > 1$ then by replacing t_i with $t'_i \le t_i$ such that $\prod_{i=1}^m t'_i = 1$, we can only decrease the value of $\sum_{i=1}^m t_i g_i(x)$. Therefore, it can be assumed that $\prod_{i=1}^m t_i = 1$ and hence

$$\prod_{i=1}^m t_i g_i(x) = \prod_{i=1}^m g_i(x). \tag{2.7}$$

Since $t_i g_i(x) > 0, i = 1,\dots,m$, and for fixed x the product of these positive numbers is constant [$= \prod_{i=1}^m g_i(x)$ by (2.7)] their sum is minimal when these numbers are equal (by theorem on arithmetic and geometric mean). That is, taking account of (2.7), the minimum of $\sum_{i=1}^m t_i g_i(x)$ is achieved when $t_i g_i(x) = [\prod_{i=1}^m g_i(x)]^{1/m}$ $\forall i = 1, \dots, m$, hence (2.6). Since for fixed $t \in T$ the function $x \mapsto \varphi_t(x) := \frac{1}{m}\sum_{i=1}^m t_i g_i(x)$ is concave by Proposition 2.5, their lower envelope $\inf_{t\in T}\varphi_t(x) = [\prod_{i=1}^m g_i(x)]^{1/m}$ is concave by the same proposition. □

An interesting concave function of the class (2.5) is the geometric mean:

$$f(x) = \begin{cases} (x_1 x_2 \cdots x_n)^{1/n} & \text{if } x_1 \geq 0, \dots, x_n \geq 0 \\ -\infty & \text{otherwise} \end{cases}$$

which corresponds to the case when $g_i(x) = x_i$.

Proposition 2.8 *Let $g(x) : \mathbb{R}^n \to (-\infty, +\infty)$ be a convex function and let $\varphi(t) : \mathbb{R} \to (-\infty, +\infty)$ be a nondecreasing convex function. Then $f(x) = \varphi(g(x))$ is convex on $\mathbb{R}^n$.*

Proof The proof is straightforward. For any $x^1, x^2 \in \mathbb{R}^n$ and $\lambda \in [0, 1]$ we have

$$g((1-\lambda)x^1 + \lambda x^2) \leq (1-\lambda)g(x^1) + \lambda g(x^2)$$

hence

$$\varphi(g((1-\lambda)x^1 + \lambda x^2)) \leq (1-\lambda)\varphi(g(x^1)) + \lambda\varphi(g(x^2)).$$

□

For example, by this proposition the function $f(x) = \sum_{i=1}^m c_i e^{g_i(x)}$ is convex if $c_i > 0$ and each $g_i(x)$ is convex proper.

Given the epigraph E of a convex function $f(x)$, one can restore $f(x)$ by the formula

$$f(x) = \inf\{t|\ (x,t) \in E\}. \tag{2.8}$$

Conversely, given a convex set $E \subset \mathbb{R}^{n+1}$ the function $f(x)$ defined by (2.8) is a convex function on $\mathbb{R}^n$ by Proposition 2.6. Therefore, if $f_1, \dots, f_m$ are m given convex functions, and $E \subset \mathbb{R}^{n+1}$ is a convex set resulting from some operation on their epigraphs $E_1, \dots, E_m$, then one can use (2.8) to define a corresponding new convex function $f(x)$.

Proposition 2.9 *Let $f_1, \dots, f_m$ be proper convex functions on $\mathbb{R}^n$. Then*

$$f(x) = \inf\left\{\sum_{i=1}^m f_i(x^i)|\ x^i \in \mathbb{R}^n,\ \sum_{i=1}^m x^i = x\right\}$$

is a convex function on $\mathbb{R}^n$.

Proof Indeed, $f(x)$ is defined by (2.8), where $E = E_1 + \cdots + E_m$ and $E_i = \text{epi}f_i$, $i = 1, \ldots, m$. □

The above constructed function $f(x)$ is called the *infimal convolution* of the functions $f_1, \ldots, f_m$. For example, the convexity of the distance function $d_C(x) = \inf\{\|x - y\| \mid y \in C\}$ associated with a convex set C follows from the above proposition because $d_C(x) = \inf_y\{\|x-y\|+\delta_C(y)\} = \inf\{\|x^1\|+\delta_C(x^2)|\, x^1+x^2 = x\}$.

Let $g(x)$ now be a nonconvex function, so that its epigraph is a nonconvex set. The relation (2.8) where $E = \text{conv(epi}g)$ defines a function $f(x)$ called the *convex envelope* or *convex hull* of $g(x)$ and denoted by conv$g$. Since E is the smallest convex set containing the epigraph of g it is easily seen that conv$g$ is the largest convex function majorized by g.

When C is a subset of $\mathbb{R}^n$, the convex envelope of the function

$$g|C = \begin{cases} g(x) & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases}$$

is called the *convex envelope of g over C*.

Proposition 2.10 *Let $g : \mathbb{R}^n \to \mathbb{R}$. The convex envelope of g over a set $C \subset \mathbb{R}^n$ such that dim(aff C) = k is given by*

$$f(x) = \inf\left\{\sum_{i=1}^{k+1}\lambda_i g(x^i)|\, x^i \in C, \lambda_i \geq 0, \sum_{i=1}^{k+1}\lambda_i = 1, \sum_{i=1}^{k+1}\lambda_i x^i = x\right\}.$$

Proof Let $X = C \times \{0\} \subset \mathbb{R}^n \times \mathbb{R},\ B = \{(0,t)|\ 0 \leq t \leq 1\} \subset \mathbb{R}^n \times \mathbb{R}$ and define $E = \{(x, g(x))|\ x \in C\} = \bigcup_{(x,0)\in X}((x,0) + g(x)B)$. Then X is a Caratheodory core of E (cf Sect. 1.4) and since dim(aff X) $= k$, by Proposition 1.14, we have

$$\text{conv}E = \left\{(x,t) = \sum_{i=1}^{k+1}\lambda_i(x^i, g(x^i))|\ x^i \in C,\ \lambda_i \geq 0,\ \sum_{i=1}^{k+1}\lambda_i = 1\right\}.$$

But clearly for every $(x,t) \in \text{epi}g$ there is $\theta \leq t$ such that $(x,\theta) \in E$, so for every $(x,t) \in \text{conv(epi}g)$ there is $\theta \leq t$ such that $(x,\theta) \in \text{conv}E$. Therefore, $(\text{conv}g)(x) = \inf\{t|\ (x,t) \in \text{conv(epi}g)\} = \inf\{t|\ (x,t) \in \text{conv}E\} = \inf\{\sum_{i=1}^{k+1}\lambda_i g(x^i)|\ x^i \in C$, $\lambda_i \geq 0, \sum_{i=1}^{k+1}\lambda_i x^i = x, \sum_{i=1}^{k+1}\lambda_i = 1\}$. □

Corollary 2.2 *The convex envelope of a concave function $g : \mathbb{R}^n \to \mathbb{R}$ over a polytope D in $\mathbb{R}^n$ with vertex set V is the function*

$$f(x) = \min\left\{\sum_{i=1}^{n+1}\lambda_i g(v^i)|\ v^i \in V, \lambda_i \geq 0, \sum_{i=1}^{n+1}\lambda_i = 1, \sum_{i=1}^{n+1}\lambda_i v^i = x\right\}.$$

Proof By the above proposition, $(\text{conv}g)(x) \leq f(x)$ but any $x \in D$ is of the form $x = \sum_{i=1}^{n+1} \lambda_i v^i$, with $v^i \in V, \lambda_i \geq 0, \sum_{i=1}^{n+1} \lambda_i = 1$, hence $f(x) \leq \sum_{i=1}^{n+1} \lambda_i g(v^i)$ (by definition of $f(x)$), while $\sum_{i=1}^{n+1} \lambda_i g(v^i) \leq g(x)$ by the concavity of g. Therefore, $f(x) \leq g(x)\ \forall x \in D$, and since $f(x)$ is convex, we must have $f(x) \leq (\text{conv}g)(x)$, and hence $f(x) = (\text{conv}g)(x)$. □

2.3 Lower Semi-Continuity

Given a function $f : \mathbb{R}^n \to [-\infty, +\infty]$ the sets

$$\{x | f(x) \leq \alpha\}, \quad \{x | f(x) \geq \alpha\}$$

where $\alpha \in [-\infty, +\infty]$ are called *lower* and *upper level sets*, respectively, of f.

Proposition 2.11 *The lower (upper, resp.) level sets of a convex (concave, resp.) function $f(x)$ are convex.*

Proof This property is equivalent to

$$f((1-\lambda)x^1 + \lambda x^2) \leq \max\{f(x^1), f(x^2)\} \quad \forall \lambda \in (0,1) \tag{2.9}$$

for all $x^1, x^2 \in \mathbb{R}^n$, which is an immediate consequence of the definition of convex functions. □

Note that the converse of this proposition is not true. For example, a real-valued function on the real line which is nondecreasing has all its lower level sets convex, but may not be convex. A function $f(x)$ whose every nonempty lower level set is convex (or, equivalently, which satisfies (2.9) for all $x^1, x^2 \in \mathbb{R}^n$), is said to be *quasiconvex*. If every nonempty upper level set is convex, $f(x)$ is said to be *quasiconcave*.

Proposition 2.12 *For any proper convex function f :*

(i) *The maximum of f over any line segment is attained at one endpoint.*
(ii) *If $f(x)$ is finite and bounded above on a halfline, then its maximum over the halfline is attained at the origin of the halfline.*
(iii) *If $f(x)$ is finite and bounded above on an affine set then it is constant on this set.*

Proof

(i) Immediate from Proposition 2.11.
(ii) If $f(b) > f(a)$ then for any $x = b + \lambda(b-a)$ with $\lambda \geq 0$, we have $b = \frac{1}{1+\lambda}x + \frac{\lambda}{1+\lambda}a$, hence $(1+\lambda)f(b) \leq f(x) + \lambda f(a)$, (whenever $f(x) < +\infty$),

i.e., $f(x) \geq \lambda[f(b) - f(a)] + f(b)$, which implies $f(x) \to +\infty$ as $\lambda \to +\infty$. Therefore, if $f(x)$ is finite and bounded above on a halfline of origin a, one must have $f(b) \leq f(a)$ for every b on this halfline.

(iii) Let M be an affine set on which $f(x)$ is finite. If $f(b) > f(a)$ for $a, b \in M$, then by (ii), $f(x)$ is unbounded above on the halfline in M from a through b. Therefore, if $f(x)$ is bounded above on M, it must be constant on M. □

A function f from a set $S \subset \mathbb{R}^n$ to $[-\infty, +\infty]$ is said to be *lower semi-continuous* (l.s.c.) at a point $x \in S$ if

$$\liminf_{\substack{y \in S \\ y \to x}} f(y) \geq f(x).$$

It is said to be *upper semi-continuous* (u.s.c.) at $x \in S$ if

$$\limsup_{\substack{y \in S \\ y \to x}} f(y) \leq f(x).$$

A function which is both lower and upper semi-continuous at x is continuous at x in the ordinary sense.

Proposition 2.13 *Let S be a closed set in $\mathbb{R}^n$. For an arbitrary function $f : S \to [-\infty, +\infty]$ the following conditions are equivalent:*

(i) *The epigraph of f is a closed set in $\mathbb{R}^{n+1}$;*
(ii) *For every $\alpha \in \mathbb{R}$ the set $\{x \in S | f(x) \leq \alpha\}$ is closed;*
(iii) *f is lower semi-continuous throughout S.*

Proof (i)⇒ (ii). Let $x^\nu \in S, x^\nu \to x, f(x^\nu) \leq \alpha$. Since $(x^\nu, \alpha) \in \text{epi} f$, it follows from the closedness of epi$f$ that $(x, \alpha) \in \text{epi} f$, i.e., $f(x) \leq \alpha$, proving (ii).

(ii)⇒(iii). Let $x^\nu \in S, x^\nu \to x$. If $\lim_{\nu \to \infty} f(x^\nu) < f(x)$ then there is $\alpha < f(x)$ such that $f(x^\nu) \leq \alpha$ for all sufficiently large ν. From (ii) it would then follow that $f(x) \leq \alpha$, a contradiction. Therefore $\lim_{\nu \to \infty} f(x^\nu) \geq f(x)$, proving (iii).

(iii)⇒ (i). Let $(x^\nu, t^\nu) \in \text{epi} f$, (i.e., $f(x^\nu) \leq t^\nu$) and $(x^\nu, t^\nu) \to (x, t)$. Then from (iii), we have $\liminf f(x^\nu) \geq f(x)$, hence $t \geq f(x)$, i.e., $(x, t) \in \text{epi} f$. □

Proposition 2.14 *Let f be a l.s.c. proper convex function. Then all the nonempty lower level sets $\{x | f(x) \leq \alpha\}$, $\alpha \in \mathbb{R}$, have the same recession cone and the same lineality space. The recession cone is made up of 0 and the directions of halflines over which f is bounded above, while the lineality space is the space parallel to the affine set on which f is constant.*

Proof By Proposition 2.13 every lower level set $C_\alpha := \{x | f(x) \leq \alpha\}$ is closed. Let $\Gamma = \{\lambda u | \lambda \geq 0\}$. If f is bounded above on a halfline $\Gamma_a = a + \Gamma$, then $f(a) \in \mathbb{R}$ (because f is proper) and by Proposition 2.12 $f(x) \leq f(a)\ \forall x \in \Gamma_a$. For any nonempty $C_\alpha, \alpha \in \mathbb{R}$, consider a point $b \in C_\alpha$ and let $\beta = \max\{f(a), \alpha\}$ so that

$\beta \in \mathbb{R}$ and $\Gamma_a \subset C_\beta$. Since C_β is closed and $b \in C_\beta$ it follows from Lemma 1.1 that $\Gamma_b \subset C_\beta$, i.e., $f(x)$ is finite and bounded above on Γ_b. Then by Proposition 2.12, $f(x) \leq f(b) \leq \alpha\ \forall x \in \Gamma_b$, hence $\Gamma_b \subset C_\alpha$. Thus, if f is bounded above on a halfline Γ_a then Γ is a direction of recession for every nonempty $C_\alpha, \alpha \in R$. The converse is obvious. Therefore, the recession cone of C_α is the same for all α and is made up of 0 and all directions of halflines over which f is bounded above. The rest of the proposition is straightforward. □

The recession cone and the lineality space common to each lower level set of f are also called the *recession cone* and the *constancy space* of f, respectively.

Corollary 2.3 *If the lower level set $\{x|\, f(x) \leq \alpha\}$ of a l.s.c. proper convex function f is nonempty and bounded for one α then it is bounded for every α.*

Proof Any lower level set of f is a closed convex set. Therefore, it is bounded if and only if its recession cone is the singleton $\{0\}$ (Corollary 1.8). □

Corollary 2.4 *If a l.s.c. proper convex function f is bounded above on a halfline then it is bounded above on every parallel halfline emanating from a point of dom$f$. If it is constant on a line then it is constant on every parallel line passing through a point of dom$f$.*

Proof Immediate. □

Proposition 2.15 *Let f be any proper convex function on $\mathbb{R}^n$. For any $y \in \mathbb{R}^n$, there exists $t \in \mathbb{R}$ such that (y, t) belongs to the lineality space of* epi$f$ *if and only if*

$$f(x + \lambda y) = f(x) + \lambda t \quad \forall x \in \mathrm{dom}f, \forall \lambda \in \mathbb{R}. \tag{2.10}$$

When f is l.s.c., this condition is satisfied provided for some $x \in$ dom$f$ *the function $\lambda \mapsto f(x + \lambda y)$ is affine.*

Proof (y, t) belongs to the lineality space of epi$f$ if and only if for any $x \in \mathrm{dom}f$: $(x, f(x)) + \lambda(y, t) \in \mathrm{epi}f\ \forall \lambda \in \mathbb{R}$, i.e., if and only if

$$f(x + \lambda y) - \lambda t \leq f(x) \quad \forall \lambda \in \mathbb{R}.$$

By Proposition 2.12 applied to the proper convex function $\varphi(\lambda) = f(x+\lambda t)-\lambda t$, this is equivalent to saying that $\varphi(\lambda) =$ constant, i.e., $f(x + \lambda y) - \lambda t = f(x)\ \forall \lambda \in \mathbb{R}$. This proves the first part of the proposition. If f is l.s.c., i.e., epi$f$ is closed, then (y, t) belongs to the lineality space of epi$f$ provided for some $x \in \mathrm{dom}f$ the line $\{(x, f(x)) + \lambda(y, t)|\ \lambda \in \mathbb{R}\}$ is contained in epi$f$ (Lemma 1.1). □

The projection of the lineality space of epi$f$ on $\mathbb{R}^n$, i.e., the set of vectors y for which there exists t such that (y, t) belongs to the lineality space of epi$f$, is called the *lineality space* of f. The directions of these vectors y are called *directions in which f is affine*. The dimension of the lineality space of f is called the *lineality* of f.

By definition, the *dimension of a convex function* f is the dimension of its domain. The number $\dim f - \text{lineality} f$ which is a measure of the nonlinearity of f is then called the *rank* of f :

$$\text{rank} f = \dim f - \text{lineality} f. \tag{2.11}$$

Corollary 2.5 *If a proper convex function f of full dimension on $\mathbb{R}^n$ has rank k then there exists a $k \times n$ matrix B with rank$B = k$ such that for any $b \in B(\text{dom} f)$ the restriction of f to the affine set $Bx = b$ is affine.*

Proof Let L be the lineality space of f. Since $\dim L = n - k$, there exists a $k \times n$ matrix B of rank k such that $L = \{u \in \mathbb{R}^n | \ Bu = 0\}$. If $b = Bx^0$ for some $x^0 \in \text{dom} f$ then for any x such that $Bx = b$, we have $B(x - x^0) = 0$, hence, denoting by $u^1, \ldots, u^h$ a basis of L, $x = x^0 + \sum_{i=1}^h \lambda_i u^i$. By Proposition 2.15 there exist $t_i \in R$ such that $f(x^0 + \sum_{i=1}^h \lambda_i u^i) = f(x^0) + \sum_{i=1}^h \lambda_i t_i$ for all $\lambda \in \mathbb{R}^h$. Therefore, $f(x)$ is affine on the affine set $Bx = b$. □

Given any proper function f on $\mathbb{R}^n$, the function whose epigraph is the closure of epi$f$ is the largest l.s.c. minorant of f. It is called the *l.s.c. hull* of f or the *closure* of f, and is denoted by cl$f$. Thus, for a proper function f,

$$\text{epi}(\text{cl} f) = \text{cl}(\text{epi} f). \tag{2.12}$$

A proper convex function f is said to be *closed* if cl$f = f$ (so for proper convex functions, closed is synonym of l.s.c.). An improper convex function is said to be closed only if $f \equiv +\infty$ or $f \equiv -\infty$ (so if $f(x) = -\infty$ for some x then its closure is the constant function $-\infty$).

Proposition 2.16 *The closure of a proper convex function f is a proper convex function which agrees with f except perhaps at the relative boundary points of dom$f$.*

Proof Since epi$f$ is convex, cl(epi$f$), i.e., epi(cl$f$), is also convex (Proposition 1.10). Hence by Proposition 2.13, cl$f$ is a closed convex function. Now the condition (2.12) is equivalent to

$$\text{cl} f(x) = \liminf_{y \to x} f(y) \quad \forall x \in \mathbb{R}^n. \tag{2.13}$$

If $x \in \text{ri}(\text{dom} f)$ then by Proposition 2.3 $f(x) = \lim_{y \to x} f(y)$, hence, by (2.13), $f(x) = \text{cl} f(x)$. Furthermore, if $x \notin \text{cl}(\text{dom} f)$ then $f(y) = +\infty$ for all y in a neighborhood of x and the same formula (2.13) shows that cl$f(x) = +\infty$. Thus the second half of the proposition is true. It remains to prove that cl$f$ is proper. For every $x \in \text{ri}(\text{dom} f)$, since f is proper, $-\infty < \text{cl} f(x) = f(x) < +\infty$. On the other hand, if cl$f(x) = -\infty$ at some relative boundary point x of dom$f$ = dom(cl$f$) then for an arbitrary $y \in \text{ri}(\text{dom} f)$, we have $\text{cl} f\left(\frac{x+y}{2}\right) \leq \frac{1}{2}\text{cl} f(x) + \frac{1}{2}\text{cl} f(y) = -\infty$. Noting that $\frac{x+y}{2} \in \text{ri}(\text{dom} f)$ this contradicts what has just been proved and thereby completes the proof. □

2.4 Lipschitz Continuity

By Proposition 2.3 a convex function is continuous relative to the relative interior of its domain. In this section we present further continuity properties of convex functions.

Proposition 2.17 *Let f be a convex function on $\mathbb{R}^n$ and D a polyhedron contained in dom$f$. Then f is u.s.c. relative to D, so that if f is l.s.c. f is actually continuous relative to D.*

Proof Consider any $x \in D$. By translating if necessary, we can assume that $x = 0$. Let $e^1, \ldots, e^n$ be the unit vectors of $\mathbb{R}^n$ and C the convex hull of the set $\{e^1, \ldots, e^n, -e^1, \ldots, -e^n\}$. This set C is partitioned by the coordinate hyperplanes into simplices $S_i, i = 1, \ldots, 2^n$, with a common vertex at 0. Since D is a polyhedron, each $D_i = S_i \cap D$ is a polytope and obviously $D \cap C = \cup_{i=1}^{2^n} D_i$. Now let $\{x^k\} \subset D \cap C$ be any sequence such that $x^k \to 0, f(x^k) \to \gamma$. Then at least one D_i, say D_1, contains an infinite subsequence of $\{x^k\}$. For convenience we also denote this subsequence by $\{x^k\}$. If V_1 is the set of vertices of D_1 other than 0, then each x^k is a convex combination of 0 and elements of V_1 : $x^k = (1 - \sum_{v \in V_1} \lambda_v^k)0 + \sum_{v \in V_1} \lambda_v^k v$, with $\lambda_v^k \geq 0, \sum_{v \in V_1} \lambda_v^k \leq 1$. By convexity of f we can write

$$f(x^k) \leq (1 - \sum_{v \in V_1} \lambda_v^k) f(0) + \sum_{v \in V_1} \lambda_v^k f(v).$$

As $k \to +\infty$, since $x^k \to 0$, it follows that $\lambda_v^k \to 0 \ \forall v$, hence $\gamma \leq f(0)$, i.e.,

$$\limsup f(x^k) \leq f(0).$$

This proves the upper semi-continuity of f at 0 relative to D. □

Theorem 2.2 *For a proper convex function f on $\mathbb{R}^n$ the following assertions are equivalent:*

(i) *f is continuous at some point;*
(ii) *f is bounded above on some open set;*
(iii) int(epi$f$) $\neq \emptyset$;
(iv) int(dom$f$) $\neq \emptyset$ *and f is Lipschitzian on every bounded set contained in* int(dom$f$);
(v) int(dom$f$) $\neq \emptyset$ *and f is continuous there.*

Proof (i) $\Rightarrow$ (ii) If f is continuous at a point x^0, then there exists an open neighborhood U of x^0 such that $f(x) < f(x^0) + 1$ for all $x \in U$.

(ii) $\Rightarrow$ (iii) If $f(x) \leq c$ for all x in an open set U, then $U \times [c, +\infty) \subset$ epi$f$, hence int(epi$f$) $\neq \emptyset$.

(iii) $\Rightarrow$ (iv) If int(epi$f$) $\neq \emptyset$, then there exists an open set U and an open interval $I \subset R$ such that $U \times I \subset$ epi$f$, hence $U \subset$ dom$f$, i.e., int(dom$f$) $\neq \emptyset$. Consider any compact set $C \subset \text{int}(\text{dom} f)$ and let B be the Euclidean unit ball. For every $r > 0$ the set $C + rB$ is compact, and the family of closed sets $\{(C + rB) \setminus \text{int}(\text{dom} f), r > 0\}$ has an empty intersection. In view of the compactness of $C + rB$ some finite subfamily of this family must have an empty intersection, hence for some $r > 0$, we must have $(C + rB) \setminus \text{int}(\text{dom} f) = \emptyset$, i.e., $C + rB \subset \text{int}(\text{dom} f)$. By Proposition 2.3 the function f is continuous on $\text{int}(\text{dom} f)$. Denote by μ_1 and μ_2 the maximum and the minimum of f over $C + rB$. Let x, x' be two distinct points in C and let $z = x + \frac{r(x-x')}{\|x-x'\|}$. Then $z \in C + rB \subset \text{int}(\text{dom} f)$. But

$$x = (1-\alpha)x' + \alpha z, \quad \alpha = \frac{\|x-x'\|}{r + \|x-x'\|}$$

and $z, x' \in \text{dom} f$, hence

$$f(x) \leq (1-\alpha)f(x') + \alpha f(z) = f(x') + \alpha(f(z) - f(x'))$$

and consequently

$$\begin{aligned} f(x) - f(x') &\leq \alpha(f(z) - f(x') \leq \alpha(\mu_1 - \mu_2) \\ &\leq \gamma\|x-x'\|, \quad \gamma = \frac{\mu_1 - \mu_2}{r}. \end{aligned}$$

By symmetry, we also have $f(x') - f(x) \leq \gamma\|x - x'\|$. Hence, for all x, x' such that $x \in C, x' \in C$:

$$|f(x) - f(x')| \leq \gamma\|x-x'\|,$$

proving the Lipschitz property of f over C.

(iv)$\Rightarrow$ (v) and (v)$\Rightarrow$ (i): obvious. □

2.5 Convex Inequalities

A *convex inequality* in x is an inequality of the form $f(x) \leq 0$ or $f(x) < 0$ where f is a convex function. Note that an inequality like $f(x) \geq 0$ or $f(x) > 0$, with $f(x)$ convex, is not convex but *reverse convex*, because it becomes convex only when reversed. A system of inequalities is said to be *consistent* if it has a solution, i.e., if there exists at least one value x satisfying all the inequalities; it is *inconsistent* otherwise. Many mathematical problems reduce to investigating the consistency (or inconsistency) of a system of inequalities.

Proposition 2.18 *Let $f_0, f_1, \ldots, f_m$ be convex functions, finite on some nonempty convex set $D \subset \mathbb{R}^n$. If the system*

$$x \in D,\ f_i(x) < 0 \ \ i = 0, 1, \ldots, m \tag{2.14}$$

is inconsistent, then there exist multipliers $\lambda_i \geq 0$, $i = 0, 1, \ldots, m$, *such that* $\sum_{i=0}^m \lambda_i > 0$ *and*

$$\lambda_0 f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \geq 0 \quad \forall x \in D. \tag{2.15}$$

If in addition

$$\exists x^0 \in D \quad f_i(x^0) < 0 \;\; i = 1, \ldots, m \tag{2.16}$$

then $\lambda_0 > 0$, *so that one can take* $\lambda_0 = 1$.

Proof Consider the set C of all vectors $y = (y_0, y_1, \ldots, y_m) \in \mathbb{R}^{m+1}$ for each of which there exists an $x \in D$ satisfying

$$f_i(x) < y_i, \quad i = 0, 1, \ldots, m. \tag{2.17}$$

As can readily be verified, C is a nonempty convex set and the inconsistency of the system (2.14) means that $0 \notin C$. By Lemma 1.2 there is a hyperplane in $\mathbb{R}^{m+1}$ properly separating 0 from C, i.e., a vector $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m) \neq 0$ such that

$$\sum_{i=0}^m \lambda_i y_i \geq 0 \quad \forall y = (y_0, y_1, \ldots, y_m) \in C. \tag{2.18}$$

If $x \in D$ then for every $\varepsilon > 0$, we have $f_i(x) < f_i(x) + \varepsilon$ for $i = 0, 1, \ldots, m$, so $(f_0(x) + \varepsilon, \ldots, f_m(x) + \varepsilon) \in C$ and hence,

$$\sum_{i=0}^m \lambda_i (f_i(x) + \varepsilon) \geq 0.$$

Since $\varepsilon > 0$ can be arbitrarily small, this implies (2.15). Furthermore, $\lambda_i \geq 0$, $i = 0, 1, \ldots, m$ because if $\lambda_j < 0$ for some j, then by fixing, for an arbitrary $x \in D$, all $y_i > f_i(x), i \neq j$, while letting $y_j \to +\infty$, we would have $\sum_{i=0}^m \lambda_i y_i \to -\infty$, contrary to (2.18). Finally, under (2.16), if $\lambda_0 = 0$ then $\sum_{i=1}^m \lambda_i > 0$, hence by (2.16) $\sum_{i=1}^m \lambda_i f_i(x^0) < 0$, contradicting the inequality $\sum_{i=1}^m \lambda_i f_i(x^0) \geq 0$ from (2.15). Therefore, $\lambda_0 > 0$ as was to be proved. □

Corollary 2.6 *Let D be a convex set in $\mathbb{R}^n$, g, f two convex functions finite on D. If $g(x^0) < 0$ for some $x^0 \in D$, while $f(x) \geq 0$ for all $x \in D$ satisfying $g(x) \geq 0$, then there exists a real number $\lambda \geq 0$ such that $f(x) + \lambda g(x) \geq 0$ $\forall x \in D$.*

Proof Apply the above Proposition for $f_0 = f, f_1 = g$. □

A more general result about inconsistent systems of convex inequalities is the following:

Theorem 2.3 (Generalized Farkas–Minkowski Theorem) *Let f_i, $i \in I_1 \subset \{1,\dots,m\}$, be affine functions on $\mathbb{R}^n$, and let f_0 and f_i, $i \in I_2 := \{1,\dots,m\} \setminus I_1$, be convex functions finite on some convex set $D \subset \mathbb{R}^n$. If there exists x^0 satisfying*

$$x^0 \in \mathrm{ri}D,\ f_i(x^0) \leq 0\ (i \in I_1),\ f_i(x^0) < 0\ (i \in I_2). \tag{2.19}$$

while the system

$$x \in D,\ f_i(x) \leq 0\ (i = 1,\dots,m),\ f_0(x) < 0, \tag{2.20}$$

is inconsistent, then there exist numbers $\lambda_i \geq 0, i = 1,\dots,m$, such that

$$f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \geq 0 \quad \forall x \in D. \tag{2.21}$$

Proof By replacing I_1 with $\{i|\ f_i(x^0) = 0\}$ we can assume $f_i(x^0) = 0$ for every $i \in I_1$. Arguing by induction on m, observe first that the theorem is true for $m = 1$. Indeed, in this case, if $I_1 = \emptyset$ the theorem follows from Proposition 2.18, so we can assume that $I_1 = \{1\}$, i.e., $f_1(x)$ is affine. In view of the fact $f_1(x^0) = 0$ for $x^0 \in \mathrm{ri}D$, if $f_1(x) \geq 0\ \forall x \in D$, then $f_1(x) = 0\ \forall x \in D$ and (2.21) holds with $\lambda = 0$. On the other hand, if there exists $x \in D$ satisfying $f_1(x) < 0$ then, since the system $x \in D, f_1(x) \leq 0, f_0(x) < 0$ is inconsistent, again the theorem follows from Proposition 2.18. Thus, in any event the theorem is true when $m = 1$. Assuming now that the theorem is true for $m = k - 1 \geq 1$, consider the case $m = k$. The hypotheses of the theorem imply that the system

$$x \in D,\ f_i(x) \leq 0\ (i = 1,\dots,k-1),\ \max\{f_k(x), f_0(x)\} < 0 \tag{2.22}$$

is inconsistent, and since the function $\max\{f_k(x), f_0(x)\}$ is convex, there exists, by the induction hypothesis, $t_i \geq 0, i = 1,\dots,k-1$, such that

$$\max\{f_k(x), f_0(x)\} + \sum_{i=1}^{k-1} t_i f_i(x) \geq 0 \quad \forall x \in D. \tag{2.23}$$

We show that this implies the inconsistency of the system

$$x \in D,\ \sum_{i=1}^{k-1} t_i f_i(x) \leq 0,\ f_k(x) \leq 0,\ f_0(x) < 0. \tag{2.24}$$

Indeed, from (2.23) it follows that no solution of (2.24) exists with $f_k(x) < 0$. Furthermore, if there exists $x \in D$ satisfying (2.24) with $f_k(x) = 0$ then, setting $x' = \alpha x^0 + (1-\alpha)x$ with $\alpha \in (0,1)$, we would have $x' \in D$, $\sum_{i=1}^{k-1} t_i f_i(x') \leq 0, f_k(x') < 0$ and $f_0(x') \leq \alpha f_0(x^0) + (1-\alpha) f_0(x) < 0$ for sufficiently small $\alpha > 0$.

Since this contradicts (2.23), the system (2.24) is inconsistent and, again by the induction hypothesis, there exist $\theta \geq 0$ and $\lambda_k \geq 0$ such that

$$f_0(x) + \theta \sum_{i=1}^{k-1} t_i f_i(x) + \lambda_k f_k(x) \geq 0 \quad \forall x \in D. \tag{2.25}$$

This is the desired conclusion with $\lambda_i = \theta t_i \geq 0$ for $i = 1, \ldots, k-1$. □

Corollary 2.7 (Farkas' Lemma) *Let A be an $m \times n$ matrix, and let $p \in \mathbb{R}^n$. If $\langle p, x\rangle \geq 0$ for all $x \in \mathbb{R}^n$ satisfying $Ax \geq 0$ then $p = A^T\lambda$ for some $\lambda = (\lambda_1, \ldots, \lambda_m) \geq 0$.*

Proof Apply the above theorem for $D = \mathbb{R}^n$, $f_0(x) = \langle p, x\rangle$, and $f_i(x) = -\langle a^i, x\rangle$, $i = 1, \ldots, m$, where a^i is the i-th row of A. Then there exist nonnegative $\lambda_1, \ldots, \lambda_m$ such that $\langle p, x\rangle - \sum_{i=1}^m \lambda_i \langle a^i, x\rangle \geq 0$ for all $x \in \mathbb{R}^n$, hence $\langle p, x\rangle - \sum_{i=1}^m \lambda_i \langle a^i, x\rangle = 0$ for all $x \in \mathbb{R}^n$, i.e., $p = \sum_{i=1}^m \lambda_i a^i$. □

Theorem 2.4 *Let $f_1, \ldots, f_m$ be convex functions finite on some convex set D in $\mathbb{R}^n$, and let A be a $k \times n$ matrix, $b \in \mathrm{ri}A(D)$. If the system*

$$x \in D,\ Ax = b,\ f_i(x) < 0,\ i = 1, \ldots, m \tag{2.26}$$

is inconsistent, then there exist a vector $t \in \mathbb{R}^m$ and nonnegative numbers $\lambda_1, \ldots, \lambda_m$ summing up to 1 such that

$$\langle t, Ax - b\rangle + \sum_{i=1}^m \lambda_i f_i(x) \geq 0 \quad \forall x \in D. \tag{2.27}$$

Proof Define $E = \{x \in D \mid Ax = b\}$. Since the system

$$x \in E,\ f_i(x) < 0,\ i = 1, \ldots, m$$

is inconsistent, there exist, by Proposition 2.18, nonnegative numbers $\lambda_1, \ldots, \lambda_m$, not all zero, such that

$$\sum_{i=1}^m \lambda_i f_i(x) \geq 0 \quad \forall x \in E. \tag{2.28}$$

By dividing $\sum_{i=1}^m \lambda_i$, we may assume $\sum_{i=1}^m \lambda_i = 1$. Obviously, the convex function $f(x) := \sum_{i=1}^m \lambda_i f_i(x)$ is finite on D. Consider the set C of all $(y, y_0) \in \mathbb{R}^k \times \mathbb{R}$ for which there exists an $x \in D$ satisfying

$$Ax - b = y,\ f(x) < y_0.$$

Since by (2.28) $0 \notin C$, and since C is convex there exists, again by Lemma 1.2, a vector $(t, t_0) \in \mathbb{R}^k \times R$ such that

$$\inf_{(y,y_0)\in C} [\langle t, y\rangle + t_0 y_0] \geq 0, \qquad \sup_{(y,y_0)\in C} [\langle t, y\rangle + t_0 y_0] > 0. \tag{2.29}$$

If $t_0 < 0$ then by fixing $x \in D, y = Ax - b$ and letting $y_0 \to +\infty$ we would have $\langle t, y\rangle + t_0 y_0 \to -\infty$, contradicting (2.29). Consequently $t_0 \geq 0$. We now contend that $t_0 > 0$. Suppose the contrary, that $t_0 = 0$, so that $\langle t, y\rangle \geq 0\ \forall (y, y_0) \in C$, and hence

$$\langle t, y\rangle \geq 0 \quad \forall y \in A(D) - b.$$

Since by hypothesis $b \in \mathrm{ri}A(D)$, i.e., $0 \in \mathrm{ri}(A(D) - b)$, this implies $\langle t, y\rangle = 0\ \forall y \in A(D)$, hence $\langle t, y\rangle + t_0 y_0 = 0$ for all $(y, y_0) \in C$, contradicting (2.29). Therefore, $t_0 > 0$ and we can take $t_0 = 1$. Then the left inequality (2.29), where $y = Ax - b$, $y_0 = f(x) + \varepsilon$ for $x \in D$, yields the desired relation (2.27) by making $\varepsilon \downarrow 0$. □

2.6 Approximation by Affine Functions

A general method of nonlinear analysis is linearization, i.e., the approximation of convex functions by affine functions. The basis for this approximation is provided by the next result on the structure of closed convex functions which is merely the analytical equivalent form of the corresponding theorem on the structure of closed convex sets (Theorem 1.5).

Theorem 2.5 *A proper closed convex function f on $\mathbb{R}^n$ is the upper envelope (pointwise supremum) of the family of all affine functions h on $\mathbb{R}^n$ minorizing f.*

Proof We first show that for any $(x^0, t^0) \notin \mathrm{epi}f$ there exists $(a, \alpha) \in \mathbb{R}^n \times \mathbb{R}$ such that

$$\langle a, x\rangle - t < \alpha < \langle a, x^0\rangle - t^0 \quad \forall (x, t) \in \mathrm{epi}f. \tag{2.30}$$

Indeed, since $\mathrm{epi}f$ is a closed convex set there exists by Theorem 1.3 a hyperplane strongly separating (x^0, t^0) from $\mathrm{epi}f$, i.e., an affine function $\langle a, x\rangle + \gamma t$ such that

$$\langle a, x\rangle + \gamma t < \alpha < \langle a, x^0\rangle + \gamma t^0 \quad \forall (x, t) \in \mathrm{epi}f.$$

It is easily seen that $\gamma \leq 0$ because if $\gamma > 0$ then by taking a point $\bar{x} \in \mathrm{dom}f$ and an arbitrary $t \geq f(\bar{x})$, we would have $(\bar{x}, t) \in \mathrm{epi}f$, hence $\langle a, \bar{x}\rangle + \gamma t < \alpha$ for all $t \geq f(\bar{x})$, which would lead to a contradiction as $t \to +\infty$. Furthermore, if $x^0 \in \mathrm{dom}f$ then $\gamma = 0$ would imply $\langle a, x^0\rangle < \langle a, x^0\rangle$, which is absurd. Hence, in this case, $\gamma < 0$, and by dividing a and α by $-\gamma$, we can assume $\gamma = -1$, so that (2.30) holds. On the other hand, if $x^0 \notin \text{dom}f$ and $\gamma = 0$, we can consider an $x^1 \in \text{dom}f$ and $t^1 < f(x^1)$, so that $(x^1, t^1) \notin \text{epi}f$ and by what has just been proved, there exists $(b, \beta) \in \mathbb{R}^n \times \mathbb{R}$ satisfying

$$\langle b, x\rangle - t < \beta < \langle b, x^1\rangle - t^1 \quad \forall (x,t) \in \text{epi}f.$$

For any $\theta > 0$ we then have for all $(x, t) \in \text{epi}f$:

$$\langle b + \theta a, x\rangle - t = (\langle b, x\rangle - t) + \theta\langle a, x\rangle < \beta + \theta\alpha,$$

while $\langle b + \theta a, x^0\rangle - t^0 = (\langle b, x^0\rangle - t^0) + \theta\langle a, x^0\rangle > \beta + \theta\alpha$ for sufficiently large θ because $\alpha < \langle a, x^0\rangle$. Thus, for $\theta > 0$ large enough, setting $a' = b+\theta a, \alpha' = \beta+\theta\alpha$, we have

$$\langle a', x\rangle - t < \alpha' < \langle a', x^0\rangle - t^0 \quad \forall (x,t) \in \text{epi}f,$$

i.e., (a', α') satisfies (2.30). Note that (2.30) implies $\langle a, x\rangle - \alpha \le f(x)$ $\forall x$, i.e., the affine function $h(x) = \langle a, x\rangle - \alpha$ minorizes $f(x)$. Now let $\mathbf{Q}$ be the family of all affine functions h minorizing f. We contend that

$$f(x) = \sup\{h(x)|\ h \in \mathbf{Q}\}. \tag{2.31}$$

Suppose the contrary, that $f(x^0) > \mu = \sup\{h(x)|\ h \in \mathbf{Q}\}$ for some x^0. Then $(x^0, \mu) \notin \text{epi}f$ and by the above there exists $(a, \alpha) \in \mathbb{R}^n \times R$ satisfying (2.30) for $t^0 = \mu$. Hence, $h(x) = \langle a, x\rangle - \alpha \in \mathbf{Q}$ and $\alpha < \langle a, x^0\rangle - \mu$, i.e., $h(x^0) = \langle a, x^0\rangle - \alpha > \mu$, a contradiction. Thus (2.31) holds, as was to be proved. □

Corollary 2.8 *For any function $f : \mathbb{R}^n \to [-\infty, +\infty]$ the closure of the convex hull of f is equal to the upper envelope of all affine functions minorizing f.*

Proof An affine function h minorizes f if and only if it minorizes cl(conv$f$), hence the conclusion. □

Proposition 2.19 *Any proper convex function f has an affine minorant. If $x^0 \in$ int(dom$f$) then an affine minorant h exists which is exact at x^0, i.e., such that $h(x^0) = f(x^0)$.*

Proof Indeed, the collection $\mathbf{Q}$ in Theorem 2.5 for cl(f) is nonempty. If $x^0 \in$ int(dom$f$) then $(x^0, f(x^0))$ is a boundary point of the convex set epi$f$. Hence by Theorem 1.5 there exists a supporting hyperplane to epi$f$ at this point, i.e., there exists $(a, \alpha) \in \mathbb{R}^n \times \mathbb{R}$ such that either $a \neq 0$ or $\alpha \neq 0$ and

$$\langle a, x\rangle - \alpha t \le \langle a, x^0\rangle - \alpha f(x^0) \quad \forall (x,t) \in \text{epi}f.$$

As in the proof of Theorem 2.5, it is readily seen that $\alpha \le 0$. Furthermore, if $\alpha = 0$ then the above relation implies that $\langle a, x\rangle \le \langle a, x^0\rangle$ for all x in some open neighborhood U of x^0 contained in dom$f$, and hence that $a = 0$, a contradiction.

Therefore, $\alpha < 0$, so we can take $\alpha = -1$. Then $\langle a, x\rangle - t \leq \langle a, x^0\rangle - f(x^0)$ for all $(x, t) \in \mathrm{epi} f$ and the affine function

$$h(x) = \langle a, x - x^0\rangle + f(x^0)$$

satisfies $h(x) \leq f(x)$ $\forall x$ and $h(x^0) = f(x^0)$. □

2.7 Subdifferential

Given a proper function f on $\mathbb{R}^n$, a vector $p \in \mathbb{R}^n$ is called a *subgradient* of f at a point x^0 if

$$\langle p, x - x^0\rangle + f(x^0) \leq f(x) \quad \forall x. \tag{2.32}$$

The set of all subgradients of f at x^0 is called the *subdifferential* of f at x^0 and is denoted by $\partial f(x^0)$ (Fig. 2.2). The function f is said to be *subdifferentiable* at x^0 if $\partial f(x^0) \neq \emptyset$.

Theorem 2.6 *Let f be a proper convex function on $\mathbb{R}^n$. For any bounded set $C \subset \mathrm{int}(\mathrm{dom} f)$ the set $\bigcup_{x\in C} \partial f(x)$ is nonempty and bounded. In particular, $\partial f(x^0)$ is nonempty and bounded at every $x^0 \in \mathrm{int}(\mathrm{dom} f)$.*

Proof By Proposition 2.19 if $x^0 \in \mathrm{int}(\mathrm{dom} f)$ then f has an affine minorant $h(x)$ such that $h(x^0) = f(x^0)$, i.e., $h(x) = \langle p, x - x^0\rangle + f(x^0)$ for some $p \in \partial f(x^0)$. Thus, $\partial f(x^0) \neq \emptyset$ for every $x^0 \in \mathrm{int}(\mathrm{dom} f)$. Consider now any bounded set $C \subset \mathrm{int}(\mathrm{dom} f)$. As we saw in the proof of Theorem 2.2, there is $r > 0$ such that $C + rB \subset \mathrm{int}(\mathrm{dom} f)$, where B denotes the Euclidean unit ball. By definition, for any

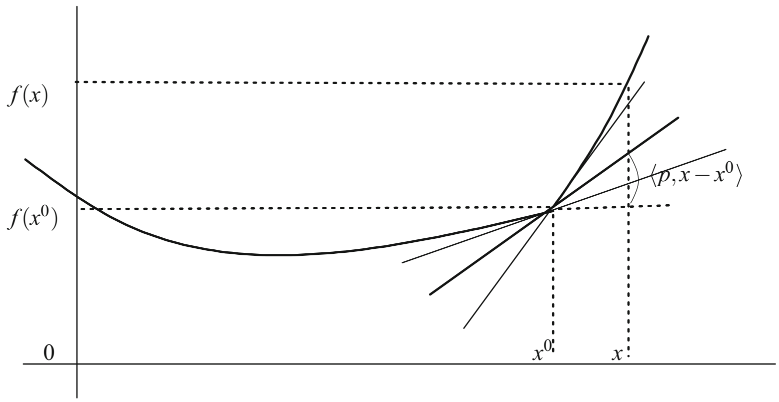

Fig. 2.2 The set $\partial f(x^0)$

$x \in C$ and $p \in \partial f(x)$, we have $\langle p, y - x\rangle + f(x) \le f(y)\ \forall y$, but by Theorem 2.2, there exists $\gamma > 0$ such that $|f(x) - f(y)| \le \gamma\|y - x\|$ for all $y \in C + rB$. Hence $|\langle p, y - x\rangle| \le \gamma\|y - x\|$ for all $y \in C + rB$, i.e., $|\langle p, u\rangle| \le \gamma\|u\|$ for all $u \in B$. By taking $u = p/\|p\|$ this implies $\|p\| \le \gamma$, so the set $\bigcup_{x\in C} \partial f(x)$ is bounded. □

Corollary 2.9 *Let f be a proper convex function on $\mathbb{R}^n$. For any bounded convex subset C of* $\mathrm{int}(\mathrm{dom}f)$ *there exists a positive constant γ such that*

$$f(x) = \sup\{h(x)|\ h \in \mathbf{Q}_0\} \quad \forall x \in C, \tag{2.33}$$

where every $h \in \mathbf{Q}_0$ has the form $h(x) = \langle a, x\rangle - \alpha$ with $\|a\| \le \gamma$.

Proof It suffices to take as $\mathbf{Q}_0$ the family of all affine functions $h(x) = \langle a, x - y\rangle + f(y)$, with $y \in C,\ a \in \partial f(y)$. □

Corollary 2.10 *Let $f : D \to \mathbb{R}$ be a convex function defined and continuous on a convex set D with nonempty interior. If the set $\cup\{\partial f(x)|\ x \in \mathrm{int}D\}$ is bounded, then f can be extended to a finite convex function on $\mathbb{R}^n$.*

Proof For each point $y \in \mathrm{int}D$ take a vector $p_y \in \partial f(y)$ and consider the affine function $h_y(x) = f(y) + \langle p_y, x - y\rangle$. The function $\tilde{f}(x) = \sup\{h_y(x)|\ y \in \mathrm{int}D\}$ is convex on $\mathbb{R}^n$ as the upper envelope of a family of affine functions. If a is any fixed point of D then $h_y(x) = f(y) + \langle p_y, a - y\rangle + \langle p_y, x - a\rangle \le f(a) + \langle p_y, x - a\rangle \le f(a) + \|p_y\|.\|x - a\|$. Since $\|p_y\|$ is bounded on $\mathrm{int}D$ the latter inequality shows that $-\infty < \tilde{f}(x) < +\infty\ \forall x \in \mathbb{R}^n$. Thus, $\tilde{f}(x)$ is a convex finite function on $\mathbb{R}^n$. Finally, since obviously $\tilde{f}(x) = f(x)$ for every $x \in \mathrm{int}D$ it follows from the continuity of both $f(x)$ and $\tilde{f}(x)$ on D that $\tilde{f}(x) = f(x)\ \forall x \in D$. □

Example 2.1 (*Positively Homogeneous Convex Function*) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a positively homogeneous convex function, i.e., a convex function $f : \mathbb{R}^n \to \mathbb{R}$ such that $f(\lambda x) = \lambda f(x)\ \forall\lambda > 0$. Then

$$\partial f(x^0) = \{p \in \mathbb{R}^n \mid \langle p, x^0\rangle = f(x^0),\ \langle p, x\rangle \le f(x)\ \forall x\} \tag{2.34}$$

Proof if $p \in \partial f(x^0)$ then $\langle p, x - x^0\rangle + f(x^0) \le f(x)\ \forall x$. Setting $x = 2x^0$ yields $\langle p, x^0\rangle + f(x^0) \le 2f(x^0)$, i.e., $\langle p, x^0\rangle \le f(x^0)$, then setting $x = 0$ yields $-\langle p, x^0\rangle \le -f(x^0)$, hence $\langle p, x^0\rangle = f(x^0)$. (Note that this condition is trivial and can be omitted if $x^0 = 0$). Furthermore, $\langle p, x - x^0\rangle = \langle p, x\rangle - \langle p, x^0\rangle = \langle p, x\rangle - f(x^0)$, hence $\langle p, x\rangle \le f(x)\ \forall x$. Conversely, if p belongs to the set on the right-hand side of (2.34) then obviously $\langle p, x - x^0\rangle \le f(x) - f(x^0)$, so $p \in \partial f(x^0)$. □

If, in addition, $f(-x) = f(x) \ge 0\ \forall x$ then the condition $\langle p, x\rangle \le f(x)\ \forall x$ is equivalent to $|\langle p, x\rangle| \le f(x)\ \forall x$. In particular:

1. If $f(x) = \|x\|$ (Euclidean norm) then

$$\partial f(x^0) = \begin{cases} \{p \mid \ \|p\| \le 1\} \text{ (unit ball)} & \text{if } x^0 = 0 \\ \{x^0/\|x^0\|\} & \text{if } x^0 \ne 0. \end{cases} \tag{2.35}$$

2. If $f(x) = \max\{|x_i| \mid i = 1, \ldots, n\}$ (Tchebycheff norm) then

$$\partial f(x^0) = \begin{cases} \text{conv}\{\pm e_1, \ldots, \pm e_n\} & \text{if } x^0 = 0 \\ \text{conv}\{(\text{sign} x_i^0) x_i^0 \mid i \in I_{x^0}\} & \text{if } x^0 \neq 0, \end{cases} \tag{2.36}$$

where $I_x = \{i \mid |x_i| = f(x)\}$.

3. If Q is a symmetric positive semidefinite matrix and $f(x) = \sqrt{\langle x, Qx\rangle}$ (elliptic norm) then

$$\partial f(x^0) = \begin{cases} \{p \mid \langle p, x\rangle \leq \sqrt{\langle x, Qx\rangle}\ \forall x\} & \text{if } x^0 \in \text{Ker}Q \\ \{(Qx^0)/\sqrt{\langle x^0, Qx^0\rangle}\} & \text{if } x^0 \notin \text{Ker}Q. \end{cases} \tag{2.37}$$

Example 2.2 (*Distance Function*) Let C be a closed convex set in $\mathbb{R}^n$, and $f(x) = \min\{\|y - x\| \mid y \in C\}$. Denote by $\pi_C(x)$ the projection of x on C, so that $\|\pi_C(x) - x\| = \min\{\|y - x\| \mid y \in C\}$ and $\langle x - \pi_C(x), y - \pi_C(x)\rangle \leq 0\ \forall y \in C$ (see Proposition 1.15). Then

$$\partial f(x^0) = \begin{cases} N_C(x^0) \cap B(0, 1) & \text{if } x^0 \in C \\ \left\{\frac{x^0 - \pi_C(x^0)}{\|x^0 - \pi_C(x^0)\|}\right\} & \text{if } x^0 \notin C, \end{cases} \tag{2.38}$$

where $N_C(x^0)$ denotes the outward normal cone of C at x^0 and $B(0, 1)$ the Euclidean unit ball.

Proof Let $x^0 \in C$, so that $f(x^0) = 0$. Then $p \in \partial f(x^0)$ implies $\langle p, x - x^0\rangle \leq f(x)\ \forall x$, hence, in particular, $\langle p, x - x^0\rangle \leq 0\ \forall x \in C$, i.e., $p \in N_C(x^0)$; furthermore, $\langle p, x - x^0\rangle \leq f(x) \leq \|x - x^0\|\ \forall x$, hence $\|p\| \leq 1$, i.e., $p \in B(0, 1)$. Conversely, if $p \in N_C(x^0) \cap B(0, 1)$ then $\langle p, x - \pi_C(x)\rangle \leq \|x - \pi_C(x)\| \leq f(x)$, and $\langle p, \pi_C(x) - x^0\rangle \leq 0$, consequently $\langle p, x - x^0\rangle = \langle p, x - \pi_C(x)\rangle + \langle p, \pi_C(x) - x^0\rangle \leq f(x) = f(x) - f(x^0)$ for all x, and so $p \in \partial f(x^0)$.

Turning to the case $x^0 \notin C$, observe that $p \in \partial f(x^0)$ implies $\langle p, x - x^0\rangle + f(x^0) \leq f(x)\ \forall x$, hence, setting $x = \pi_C(x^0)$ yields $\langle p, \pi_C(x^0) - x^0\rangle + \|\pi_C(x^0) - x^0\| \leq 0$, i.e., $\langle p, x^0 - \pi_C(x^0)\rangle \geq \|x^0 - \pi_C(x^0)\|$. On the other hand, setting $x = 2x^0 - \pi_C(x^0)$ yields $\langle p, x^0 - \pi_C(x^0)\rangle + \|\pi_C(x^0) - x^0\| \leq 2\|\pi_C(x^0) - x^0\|$, i.e., $\langle p, x^0 - \pi_C(x^0)\rangle \leq \|\pi_C(x^0) - x^0\|$. Thus, $\langle p, x^0 - \pi_C(x^0)\rangle = \|\pi_C(x^0) - x^0\|$ and consequently $p = \frac{x^0 - \pi_C(x^0)}{\|x^0 - \pi_C(x^0)\|}$. Conversely, the last equality implies $\langle p, x^0 - \pi_C(x^0)\rangle = \|x^0 - \pi_C(x^0)\| = f(x^0)$, $\langle p, x - \pi_C(x)\rangle \leq \|x - \pi_C(x)\| = f(x)$, hence $\langle p, x - x^0\rangle + f(x^0) = \langle p, x - x^0\rangle + \langle p, x^0 - \pi_C(x^0)\rangle = \langle p, x - \pi_C(x^0)\rangle = \langle p, x - \pi_C(x)\rangle + \langle p, \pi_C(x) - \pi_C(x^0)\rangle, \leq \|x - \pi_C(x)\| = f(x)$ for all x (note that $\langle p, \pi_C(x) - \pi_C(x^0)\rangle \leq 0$ because $p \in N_C(\pi_C(x^0))$). Therefore, $\langle p, x - x^0\rangle + f(x^0) \leq f(x)$ for all x, proving that $p \in \partial f(x^0)$. □

Observe from the above examples that there is a unique subgradient (which is just the gradient) at every point where f is differentiable. This is actually a general fact which we are now going to establish.

Let $f : \mathbb{R}^n \to [-\infty, +\infty]$ be any function and let x^0 be a point where f is finite. If for some $u \neq 0$ the limit (finite or infinite)

$$\lim_{\lambda \downarrow 0} \frac{f(x^0 + \lambda u) - f(x^0)}{\lambda}$$

exists, then it is called the *directional derivative* of f at x^0 in the direction u, and is denoted by $f'(x^0; u)$.

Proposition 2.20 *Let f be a proper convex function and $x^0 \in \mathrm{dom} f$. Then:*

(i) *$f'(x^0; u)$ exists for every direction u and satisfies*

$$f'(x^0; u) = \inf_{\lambda > 0} \frac{f(x^0 + \lambda u) - f(x^0)}{\lambda}; \tag{2.39}$$

(ii) *The function $u \mapsto f'(x^0; u)$ is convex and homogeneous and $p \in \partial f(x^0)$ if and only if*

$$\langle p, u \rangle \leq f'(x^0, u) \quad \forall u. \tag{2.40}$$

(iii) *If f is continuous at x^0 then $f'(x^0; u)$ is finite and continuous at every $u \in \mathbb{R}^n$, the subdifferential $\partial f(x^0)$ is compact and*

$$f'(x^0; u) = \max\{\langle p, u \rangle \mid p \in \partial f(x^0)\}. \tag{2.41}$$

Proof

(i) For any given $u \neq 0$ the function $\varphi(\lambda) = f(x^0 + \lambda u)$ is proper convex on the real line, and $0 \in \mathrm{dom}\varphi$. Therefore, its right derivative $\varphi'_+(0) = f'(x^0; u)$ exists (but may equal $+\infty$ if 0 is an endpoint of $\mathrm{dom}\varphi$). The relation (2.39) follows from the fact that $[\varphi(\lambda) - \varphi(0)]/\lambda$ is nonincreasing as $\lambda \downarrow 0$.

(ii) The homogeneity of $f'(x^0; u)$ is obvious. The convexity then follows from the relations

$$\begin{aligned} f'(x^0; u + v) &= \inf_{\lambda > 0} \frac{f(x^0 + \frac{\lambda}{2}(u + v)) - f(x^0)}{\frac{\lambda}{2}} \\ &\leq \inf_{\lambda > 0} \frac{f(x^0 + \lambda u) - f(x^0) + f(x^0 + \lambda v) - f(x^0)}{\lambda} \\ &= f'(x^0; u) + f'(x^0; v). \end{aligned}$$

Setting $x = x^0 + \lambda u$ we can turn the subgradient inequality (2.32) into the condition

$$\langle p, u \rangle \leq [f(x^0 + \lambda u) - f(x^0)]/\lambda \quad \forall u,\ \forall \lambda > 0,$$

which is equivalent to $\langle p, u \rangle \leq \inf_{\lambda > 0}[f(x^0 + \lambda u) - f(x^0)]/\lambda,\ \forall u$, i.e., by (i), $\langle p, u \rangle \leq f'(x^0; u)\ \forall u$.

(iii) If f is continuous at x^0 then there is a neighborhood U of 0 such that $f(x^0+u)$ is bounded above on U. Since by (i) $f'(x^0;u) \le f(x^0+u) - f(x^0)$, it follows that $f'(x^0;u)$ is also bounded above on U, and hence is finite and continuous on $\mathbb{R}^n$ (Theorem 2.2). The Condition (2.40) then implies that $\partial f(x^0)$ is closed and hence compact because it is bounded by Theorem 2.6. In view of the homogeneity of $f'(x^0;u)$, an affine minorant of it which is exact at some point must be of the form $\langle p,u\rangle$, with $\langle p,u\rangle \le f'(x^0;u)\ \forall u$, i.e., by (ii), $p \in \partial f(x^0)$. By Corollary 2.9, we then have $f'(x^0;u) = \max\{\langle p,u\rangle \mid p \in \partial f(x^0)\}$. □

According to the usual definition, a function f is *differentiable* at a point x^0 if there exists a vector $\nabla f(x^0)$ (the *gradient* of f at x^0) such that

$$f(x^0+u) = f(x^0) + \langle \nabla f(x^0), u\rangle + o(\|u\|).$$

This is equivalent to

$$\lim_{\lambda \downarrow 0} \frac{f(x^0+\lambda u) - f(x^0)}{\lambda} = \langle \nabla f(x^0), u\rangle, \quad \forall u \neq 0,$$

so the directional derivative $f'(x^0;u)$ exists, and is a linear function of u.

Proposition 2.21 *Let f be a proper convex function and $x^0 \in \mathrm{dom} f$. If f is differentiable at x^0 then $\nabla f(x^0)$ is its unique subgradient at x^0.*

Proof If f is differentiable at x^0 then $f'(x^0;u) = \langle \nabla f(x^0), u\rangle$, so by (ii) of Proposition 2.20, a vector p is a subgradient f at x^0 if and only if $\langle p,u\rangle \le \langle \nabla f(x^0), u\rangle\ \forall u$, i.e., if and only if $p = \nabla f(x^0)$. □

One can prove conversely that if f has a unique subgradient at x^0 then f is differentiable at x^0 (see, e.g., Rockafellar 1970).

2.8 Subdifferential Calculus

A convex function f may result from some operations on convex functions $f_i, i \in I$. (cf Sect. 2.2). It is important to know how the subdifferential of f can be computed in terms of the subdifferentials of the f_i' s.

Proposition 2.22 *Let $f_i,\ i = 1,\dots,m$, be proper convex functions on $\mathbb{R}^n$. Then for every $x \in \mathbb{R}^n$:*

$$\partial\left(\sum_{i=1}^m f_i(x)\right) \supset \sum_{i=1}^m \partial f_i(x).$$

If there exists a point $a \in \cap_{i=1}^m \mathrm{dom} f_i$, *where every function* f_i, *except perhaps one, is continuous, then the above inclusion is in fact an equality for every* $x \in \mathbb{R}^n$.

Proof It suffices to prove the proposition for $m = 2$ because the general case will follow by induction. Furthermore, the first part is straightforward, so we only need to prove the second part. If $p \in \partial(f_1 + f_2)(x^0)$, then the system

$$x - y = 0,\ f_1(x) + f_2(y) - f_1(x^0) - f_2(x^0) - \langle p, x - x^0\rangle < 0$$

is inconsistent. Define $D = \mathrm{dom} f_1 \times \mathrm{dom} f_2$ and $A(x, y) := x - y$. By hypothesis, f_1 is continuous at $a \in \mathrm{dom} f_1 \cap \mathrm{dom} f_2$, so there is a ball U around 0 such that $a + U \subset \mathrm{dom} f_1$, hence $U = (a + U) - a \subset \mathrm{dom} f_1 - \mathrm{dom} f_2 = A(D)$, i.e., $0 \in \mathrm{int} A(D)$. Therefore, by Theorem 2.4 there exists $t \in \mathbb{R}^n$ such that

$$\langle t, x - y\rangle + [f_1(x) + f_2(y) - f_1(x^0) - f_2(x^0) - \langle p, x - x^0\rangle] \geq 0$$

for all $x \in \mathbb{R}^n$ and all $y \in \mathbb{R}^n$. Setting $y = x^0$ yields $\langle p - t, x - x^0\rangle \leq f_1(x) - f_1(x^0)$ $\forall x \in \mathbb{R}^n$, i.e., $p - t \in \partial f_1(x^0)$. Then setting $x = x^0$ yields $\langle t, y - x^0\rangle \leq f_2(y) - f_2(x^0)$ $\forall y \in \mathbb{R}^n$, i.e., $t \in \partial f_2(x^0)$. Thus, $p = (p - t) + t \in \partial f_1(x^0) + \partial f_2(x^0)$, as was to be proved. □

Proposition 2.23 *Let* $A : \mathbb{R}^n \to \mathbb{R}^m$ *be a linear mapping and* g *be a proper convex function on* $\mathbb{R}^m$. *Then for every* $x \in \mathbb{R}^n$:

$$A^T \partial g(Ax) \subset \partial(g \circ A)(x).$$

If g *is continuous at some point in* $Im(A)$ (*the range of* A) *then the above inclusion is in fact an equality for every* $x \in \mathbb{R}^n$.

Proof The first part is straightforward. To prove the second part, consider any $p \in \partial(g \circ A)(x^0)$. Then the system

$$Ax - y = 0,\ g(y) - g(Ax^0) - \langle p, x - x^0\rangle < 0,\ x \in \mathbb{R}^n,\ y \in \mathbb{R}^m$$

is inconsistent. Define $D = \mathbb{R}^n \times \mathrm{dom} g$, $B(x, y) = Ax - y$. Since there is a point $b \in \mathrm{Im} A \cap \mathrm{int}(\mathrm{dom} g)$, we have $b \in \mathrm{int} B(D)$, so by Theorem 2.4 there exists $t \in \mathbb{R}^m$ such that

$$\langle t, Ax - y\rangle + g(y) - g(Ax^0) - \langle p, x - x^0\rangle \geq 0$$

for all $x \in \mathbb{R}^n$ and all $y \in \mathbb{R}^m$. Setting $y = 0$ then yields $\langle A^T t - p, x\rangle - g(Ax^0) + \langle p, x^0\rangle \geq 0$ $\forall x \in \mathbb{R}^n$, hence $p = A^T t$, while setting $x = x^0$ yields $\langle t, y - Ax^0\rangle \leq g(y) - g(Ax^0)$, i.e., $t \in \partial g(Ax^0)$. Therefore, $p \in A^T \partial g(Ax^0)$. □

Proposition 2.24 *Let* $g(x) = (g_1(x), \ldots, g_m(x))$, *where each* g_i *is a convex functions from* $\mathbb{R}^n$ *to* $\mathbb{R}$, *let* $\varphi : \mathbb{R}^m \to \mathbb{R}$ *be a convex function, component-wise* *increasing, i.e., such that* $\varphi(t) \geq \varphi(t')$ *whenever* $t_i \geq t'_i$, $i = 1, \ldots, m$. *Then the function* $f = \varphi \circ g$ *is convex and*

$$\partial f(x) = \left\{ \sum_{i=1}^{m} s_i p^i \mid p^i \in \partial g_i(x), (s_1, \ldots, s_m) \in \partial \varphi(g(x)) \right\}. \tag{2.42}$$

Proof The convexity of $f(x)$ follows from an obvious extension of Proposition 2.8 which corresponds to the case $m = 1$. To prove (2.42), let $p = \sum_{i=1}^{m} s_i p^i$ with $p^i \in \partial g_i(x^0), s \in \partial \varphi(g(x^0))$. First observe that $\langle s, y - g(x^0)\rangle \leq \varphi(y) - \varphi(g(x^0))\ \forall y \in \mathbb{R}^m$ implies, for all $y = g(x^0)+u$ with $u \leq 0 :\ \langle s, u\rangle \leq \varphi(g(x^0)+u) - \varphi(g(x^0)) \leq 0\ \forall u \leq 0$, hence $s \geq 0$. Now $\langle p, x - x^0\rangle\ = \sum_{i=1}^{m} s_i \langle p^i, x - x^0\rangle \leq \sum_{i=1}^{m} s_i [g_i(x) - g_i(x^0)]\ = \langle s, g(x) - g(x^0)\rangle\ \leq \varphi(g(x)) - \varphi(g(x^0))\ = f(x) - f(x^0)$ for all $x \in \mathbb{R}^n$. Therefore $p \in \partial f(x^0)$, i.e., the right-hand side of (2.42) is contained in the left-hand side. To prove the converse, let $p \in \partial f(x^0)$, so that the system

$$x \in \mathbb{R}^n,\ y \in \mathbb{R}^m,\ g_i(x) < y_i \quad i = 1, \ldots, m \tag{2.43}$$

$$\varphi(y) - \varphi(g(x^0)) - \langle p, x - x^0\rangle < 0 \tag{2.44}$$

is inconsistent, while the system (2.43) has a solution. By Proposition 2.18 there exists $s \in \mathbb{R}^m_+$ such that

$$\varphi(y) - \varphi(g(x^0)) - \langle p, x - x^0\rangle + \langle s, g(x) - y\rangle \geq 0$$

for all $x \in \mathbb{R}^n, y \in \mathbb{R}^m$. Setting $x = x^0$ yields $\varphi(y) - \varphi(g(x^0)) \geq \langle s, y - g(x^0)\rangle$ for all $y \in \mathbb{R}^m$, which means that $s \in \partial \varphi(g(x^0))$. On the other hand, setting $y = g(x^0)$ yields $\langle p, x - x^0\rangle \leq \sum_{i=1}^{m} s_i [g_i(x) - g_i(x^0)]$ for all $x \in \mathbb{R}^n$, which means that $p \in \partial(\sum_{i=1}^{m} s_i g_i(x^0))$, hence by Proposition 2.22, $p = \sum_{i=1}^{m} s_i p^i$ with $p^i \in \partial g_i(x^0)$. □

Note that when $\varphi(y)$ is differentiable at $g(x)$ the above formula (2.42) is similar to the classical chain rule, namely:

$$\partial(\varphi \circ g)(x) = \sum_{i=1}^{m} \frac{\partial \varphi}{\partial y_i}(g(x)) \partial g_i(x).$$

Proposition 2.25 *Let* $f(x) = \max\{g_1(x), \ldots, g_m(x)\}$, *where each* g_i *is a convex function from* $\mathbb{R}^n$ *to* $\mathbb{R}$. *Then*

$$\partial f(x) = \text{conv}\{\cup \partial g_i(x) |\ i \in I(x)\}, \tag{2.45}$$

where $I(x) = \{i \mid f(x) = g_i(x)\}$.

Proof If $p \in \partial f(x^0)$ then the system

$$g_i(x) - f(x^0) - \langle p, x - x^0\rangle < 0 \quad i = 1, \ldots, m$$

is inconsistent. By Proposition 2.18, there exist $\lambda_i \geq 0$ such that $\sum_{i=1}^m \lambda_i = 1$ and $\sum_{i=1}^m \lambda_i[g_i(x) - f(x^0) - \langle p, x - x^0\rangle] \geq 0$. Setting $x = x^0$, we have

$$\sum_{i \notin I(x^0)} \lambda_i[g_i(x^0) - f(x^0)] \geq 0,$$

with $g_i(x^0) - f(x^0) < 0$ for every $i \notin I(x^0)$. This implies that $\lambda_i = 0$ for all $i \notin I(x^0)$. Hence

$$\sum_{i \in I(x^0)} \lambda_i[g_i(x) - g_i(x^0) - \langle p, x - x^0\rangle] \geq 0$$

for all $x \in \mathbb{R}^n$ and so $p \in \partial(\sum_{i \in I(x^0)} \lambda_i g_i(x^0))$. By Proposition 2.22, $p = \sum_{i \in I(x^0)} p^i$, with $p^i \in \partial g_i(x^0)$. Thus $\partial f(x^0) \subset \mathrm{conv}\{\cup \partial g_i(x^0) | \ i \in I(x^0)\}$. The converse inclusion can be verified in a straightforward manner. □

2.9 Approximate Subdifferential

A proper convex function f on $\mathbb{R}^n$ may have an empty subdifferential at certain points. In practice, however, one often needs only a concept of approximate subdifferential .

Given a positive number $\varepsilon > 0$, a vector $p \in \mathbb{R}^n$ is called an *ε-subgradient* of f at point x^0 if

$$\langle p, x - x^0\rangle + f(x^0) \leq f(x) + \varepsilon \quad \forall x. \tag{2.46}$$

The set of all ε-subgradients of f at x^0 is called the *ε-subdifferential* of f at x^0, and is denoted by $\partial_\varepsilon f(x^0)$ (Fig. 2.3).

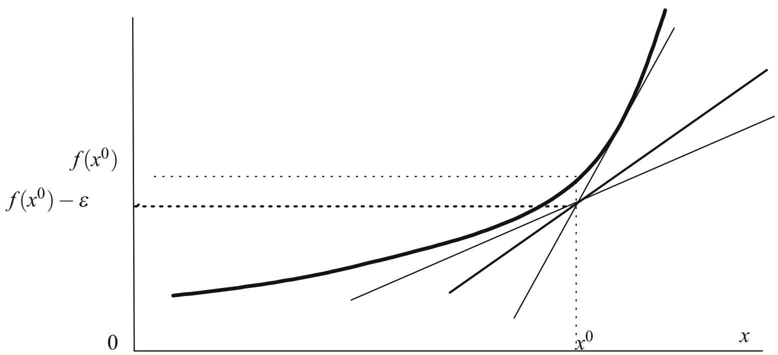

Fig. 2.3 The set $\partial_\varepsilon f(x^0)$

Proposition 2.26 *For any proper closed convex function f on* $\mathbb{R}^n$, *and any given* $\varepsilon > 0$ *the* ε*-subdifferential of f at any point* $x^0 \in \mathrm{dom} f$ *is nonempty. If C is a bounded subset of int(dom f) then the set* $\cup_{x\in C}\partial_\varepsilon f(x)$ *is bounded.*

Proof Since the point $(x^0, f(x^0) - \varepsilon) \notin \mathrm{epi} f$, there exists $p \in \mathbb{R}^n$ such that $\langle p, x - x^0\rangle < f(x) - (f(x^0) - \varepsilon) \quad \forall x$ (see (2.30) in the proof of Theorem 2.5). Thus, $\partial_\varepsilon f(x^0) \neq \emptyset$. The proof of the second part of the proposition is analogous to that of Theorem 2.6. □

Note that $\partial_\varepsilon f(x^0)$ is unbounded when x^0 is a boundary point of $\mathrm{dom} f$.

Proposition 2.27 *Let* $x^0, x^1 \in \mathrm{dom} f$. *If* $p \in \partial_\varepsilon f(x^0)$ $(\varepsilon \geq 0)$ *then* $p \in \partial_\eta f(x^1)$ *for* $\eta = f(x^1) - f(x^0) - \langle p, x^1 - x^0\rangle + \varepsilon \geq 0.$

Proof If $\langle p, x - x^0\rangle \leq f(x) - f(x^0) + \varepsilon$ for all x then $\langle p, x - x^1\rangle = \langle p, x - x^0\rangle + \langle p, x^0 - x^1\rangle \leq f(x) - f(x^0) + \varepsilon - \langle p, x^1 - x^0\rangle = f(x) - f(x^1) + \eta$ for all x. □

A function $f(x)$ is said to be *strongly convex* on a convex set C if there exists $r > 0$ such that

$$f((1-\lambda)x^1 + \lambda x^2) \leq (1-\lambda)f(x^1) + \lambda f(x^2) - (1-\lambda)\lambda r\|x^1 - x^2\|^2 \tag{2.47}$$

for all $x^1, x^2 \in C$, and all $\lambda \in [0, 1]$. The number $r > 0$ is then called the *modulus of strong convexity* of $f(x)$. Using the identity

$$(1-\lambda)\lambda\|x^1 - x^2\|^2 = (1-\lambda)\|x^1\|^2 + \lambda\|x^2\|^2 - \|(1-\lambda)x^1 + \lambda x^2\|^2$$

for all $x^1, x^2 \in \mathbb{R}^n$, and all $\lambda \in [0, 1]$, it is easily verified that a convex function $f(x)$ is strongly convex with modulus of strong convexity r if and only if the function $f(x) - r\|x\|^2$ is convex.

Proposition 2.28 *If* $f(x)$ *is a strongly convex function on* $\mathbb{R}^n$ *with modulus of strong convexity r then for any* $x^0 \in \mathbb{R}^n$ *and* $\varepsilon \geq 0$:

$$\partial f(x^0) + B(0, 2\sqrt{r\varepsilon}) \subset \partial_\varepsilon f(x^0), \tag{2.48}$$

where $B(0, \alpha)$ *denotes the ball of radius* α *around 0.*

Proof Let $p \in \partial f(x^0)$. Since $F(x) := f(x) - r\|x\|^2$ is convex and $p - 2rx^0 \in \partial F(x^0)$ we can write

$$f(x) - f(x^0) - r(\|x\|^2 - \|x^0\|^2) \geq \langle p - 2rx^0, x - x^0\rangle$$

for all x, hence

$$f(x) - f(x^0) \geq \langle p, x - x^0\rangle + r\|x - x^0\|^2.$$

Let us determine a vector u such that

$$r\|x - x^0\|^2 \geq \langle u, x - x^0 \rangle - \varepsilon \quad \forall x \in \mathbb{R}^n. \tag{2.49}$$

The convex quadratic function $r\|x - x^0\|^2 - \langle u, x - x^0 \rangle$ achieves its minimum at the point $\bar{x}$ such that $2r(\bar{x} - x^0) - u = 0$, i.e., $\bar{x} - x^0 = \frac{u}{2r}$. This minimum is equal to $r[\frac{u}{2r}]^2 - \frac{\|u\|^2}{2r} = \frac{-\|u\|^2}{4r}$. Thus, by choosing u such that $\|u\|^2 \leq 4r\varepsilon$, i.e., $u \in B(0, 2\sqrt{r\varepsilon})$ we will have (2.49), hence $p + u \in \partial_\varepsilon f(x^0)$. □

Corollary 2.11 *Let $f(x) = \frac{1}{2}\langle x, Qx \rangle + \langle x, a \rangle$, where $a \in \mathbb{R}^n$, and Q is an $n \times n$ symmetric positive definite matrix. Let $r > 0$ be the smallest eigenvalue of Q. Then*

$$Qx + a + u \in \partial_\varepsilon f(x)$$

for any $u \in \mathbb{R}^n$ such that $\|u\| \leq 2\sqrt{r\varepsilon}$.

Proof Clearly $f(x) - r\|x\|^2$ is convex (as its smallest eigenvalue is nonnegative), so $f(x)$ is a strongly convex function to which the above proposition applies. □

2.10 Conjugate Functions

Given an arbitrary function $f : \mathbb{R}^n \to [-\infty, +\infty]$, we consider the set of all affine functions h minorizing f. It is natural to restrict ourselves to proper functions, because an improper function either has no affine minorant (if $f(x) = -\infty$ for some x) or is minorized by every affine function (if $f(x)$ is identical to $+\infty$).

Observe that if $\langle p, x \rangle - \alpha \leq f(x)$ for all x then $\alpha \geq \langle p, x \rangle - f(x) \ \ \forall x$. The function

$$f^*(p) = \sup_{x \in \mathbb{R}^n} \{\langle p, x \rangle - f(x)\}. \tag{2.50}$$

which is clearly closed and convex is called the *conjugate* of f.

For instance, the conjugate of the function $f(x) = \delta_C(x)$ (indicator function of a set C, see Sect. 2.1) is the function $f^*(p) = \sup_{x \in C} \langle p, x \rangle = s_C(p)$ (support function of C). The conjugate of an affine function $f(x) = \langle c, x \rangle - \alpha$ is the function

$$f^*(p) = \sup_x \{\langle p, x \rangle - \langle c, x \rangle + \alpha\} = \begin{cases} +\infty, & p \neq c \\ \alpha, & p = c. \end{cases}$$

Two less trivial examples are the following:

Example 2.3 The conjugate of the proper convex function $f(x) = e^x, \ x \in \mathbb{R}$, is by definition $f^*(p) = \sup_x \{px - e^x\}$. Obviously, $f^*(p) = 0$ for $p = 0$ and $f^*(p) = +\infty$ for $p < 0$. For $p > 0$, the function $px - e^x$ achieves a maximum at $x = \xi$ satisfying $p = e^\xi$, so $f^*(p) = p \log p - p$. Thus,

$$f^*(p) = \begin{cases} 0, & p = 0 \\ +\infty, & p < 0 \\ p \log p - p, & p > 0. \end{cases}$$

The conjugate of $f^*(p)$ is in turn the function $f^{**}(x) = \sup_p \{px - f^*(p)\} = \sup\{px - p \log p + p \mid p > 0\} = e^x$.

Example 2.4 The conjugate of the function $f(x) = \frac{1}{\alpha} \sum_{i=1}^n |x_i|^\alpha,\ 1 < \alpha < +\infty$, is

$$f^*(p) = \frac{1}{\beta} \sum_{i=1}^n |p_i|^\beta,\ 1 < \beta < +\infty,$$

where $1/\alpha + 1/\beta = 1$. Indeed,

$$f^*(p) = \sup_x \left\{ \sum_{i=1}^n p_i x_i - \frac{1}{\alpha} \sum_{i=1}^n |x_i|^\alpha \right\}.$$

By differentiation, we find that the supremum on the right-hand side is achieved at $x = \xi$ satisfying $p_i = |\xi_i|^{\alpha-1} \mathrm{sign} \xi_i$, hence

$$f^*(p) = \sum_{i=1}^n |\xi_i|^\alpha \left(1 - \frac{1}{\alpha}\right) = \frac{1}{\beta} \sum_{i=1}^n |p_i|^\beta.$$

Proposition 2.29 *Let $f : \mathbb{R}^n \to [-\infty, +\infty]$ be an arbitrary proper function. Then:*

(i) $f(x) + f^*(p) \geq \langle p, x \rangle \quad \forall x \in \mathbb{R}^n,\ \forall p \in \mathbb{R}^n$;
(ii) $f^{**}(x) \leq f(x)\ \forall x$, *and* $f^{**} = f$ *if and only if* f *is convex and closed;*
(iii) $f^{**}(x) = \sup\{h(x) \mid h \text{ affine},\ h \leq f\}$, *i.e.,* $f^{**}(x)$ *is the largest closed convex function minorizing* $f(x) : f^{**} = \mathrm{cl(conv)}f$.

Proof (i) is obvious, and from (i) $f^{**}(x) = \sup_p \{\langle p, x \rangle - f^*(p)\} \leq f(x)$. Let Q be the set of all affine functions majorized by f. For every $h \in Q$, say $h(x) = \langle p, x \rangle - \alpha$, we have $\langle p, x \rangle - \alpha \leq f(x)\ \forall x$, hence $\alpha \geq \sup_x \{\langle p, x \rangle - f(x)\} = f^*(p)$, and consequently $h(x) \leq \langle p, x \rangle - f^*(p)\ \forall x$. Thus

$$\sup\{h \mid h \in Q\} \leq \sup_p \{\langle p, x \rangle - f^*(p)\} = f^{**}, \tag{2.51}$$

If f is convex and closed then, since it is proper, by Theorem 2.5 it is just equal to the function on the left-hand side of (2.51) and since $f \geq f^{**}$, it follows that $f = f^{**}$. Conversely, if $f = f^{**}$ then f is the conjugate of f^*, hence is convex and closed. Turning to (iii) observe that if $Q = \emptyset$ then $f^*(p) = \sup_x \{\langle p, x \rangle - f(x)\} = +\infty$ for all p and consequently, $f^{**} \equiv -\infty$. But in this case, $\sup\{h|\ h \in Q\} = -\infty$, too, hence $f^{**} = \sup\{h|\ h \in Q\}$. On the other hand, if there is $h \in Q$ then from (2.51) $h \leq f^{**}$; conversely, if $h \leq f^{**}$ then $h \leq f$ (because $f^{**} \leq f$). In this case, since $f^{**}(x) \leq f(x) < +\infty$ at least for some x, and $f^{**}(x) > -\infty\ \forall x$ (because there is an affine function minorizing f^{**}), it follows that f^{**} is proper. By Theorem 2.5, then $f^{**} = \sup\{h|\ h$ affine $h \leq f^{**}\} = \sup\{h|\ h \in Q\}$. This proves the equality in (iii), and so, by Corollary 2.8, $f^{**} = \mathrm{cl(conv)}f$. □

2.11 Extremal Values of Convex Functions

The smallest and the largest values of a convex function on a given convex set are often of particular interest.

Let $f : \mathbb{R}^n \to [-\infty, +\infty]$ be an arbitrary function, and C an arbitrary set in $\mathbb{R}^n$. A point $x^0 \in C \cap \mathrm{dom} f$ is called a *global minimizer* of $f(x)$ on C if $-\infty < f(x^0) \leq f(x)$ for all $x \in C$. It is called a *local minimizer* of $f(x)$ on C if there exists a neighborhood $U(x^0)$ of x^0 such that $-\infty < f(x^0) \leq f(x)$ for all $x \in C \cap U(x^0)$. The concepts of *global maximizer* and *local maximizer* are defined analogously. For an arbitrary function f on a set C we denote the set of all global minimizers (maximizers) of f on C by $\mathrm{argmin}_{x\in C} f(x)$ ($\mathrm{argmax}_{x\in C} f(x)$, resp.). Since $\min_{x\in C} f(x) = -\max_{x\in C}(-f(x))$ the theory of the minimum (maximum) of a convex function is the same as the theory of the maximum (minimum, resp.) of a concave function.

2.11.1 Minimum

Proposition 2.30 *Let C be a nonempty convex set in $\mathbb{R}^n$, and $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Any local minimizer of f on C is also global. The set* $\underset{x\in C}{\mathrm{argmin}} f(x)$ *is a convex subset of C.*

Proof Let $x^0 \in C$ be a local minimizer of f and $U(x^0)$ be a neighborhood such that $f(x^0) \leq f(x)\ \forall x \in C \cap U(x^0)$. For any $x \in C$ we have $x_\lambda := (1-\lambda)x^0 + \lambda x \in C \cap U(x^0)$ for sufficiently small $\lambda > 0$. Then $f(x^0) \leq f(x_\lambda) \leq (1-\lambda)f(x^0) + \lambda f(x)$, hence $f(x^0) \leq f(x)$, proving the first part of the proposition. If $\alpha = \min f(C)$ then $\mathrm{argmin}_{x\in C} f(x)$ coincides with the set $C \cap \{x|\ f(x) \leq \alpha\}$ which is a convex set by the convexity of $f(x)$ (Proposition 2.11). □

Remark 2.1 A real-valued function f on a convex set C is said to be *strictly convex* on C if

$$f((1-\lambda)x^1 + \lambda x^2) < (1-\lambda)f(x^1) + \lambda f(x^2)$$

for any two distinct points $x^1, x^2 \in C$ and $0 < \lambda < 1$. For such a function f the set $\mathrm{argmin}_{x\in C} f(x)$, if nonempty, is a singleton, i.e., a strictly convex function $f(x)$ on C has at most one minimizer over C. In fact, if there were two distinct minimizers x^1, x^2 then by strict convexity $f(\frac{x^1+x^2}{2}) < f(x^1)$, which is impossible.

Proposition 2.31 *Let C be a convex set in $\mathbb{R}^n$, and $f : \mathbb{R}^n \to [-\infty, +\infty]$ a convex function which is finite on C. For a point $x^0 \in C$ to be a minimizer of f on C it is necessary and sufficient that*

$$0 \in \partial f(x^0) + N_C(x^0), \tag{2.52}$$

*where $N_C(x^0)$ denotes the (outward) normal cone of C at x^0. (*cf Sect. 1.6*)*

Proof If (2.52) holds there is $p \in \partial f(x^0) \cap -N_C(x^0)$. For every $x \in C$, since $p \in \partial f(x^0)$, we have $\langle p, x - x^0\rangle \le f(x) - f(x^0)$, i.e., $f(x^0) \le f(x) - \langle p, x - x^0\rangle$; on the other hand, since $p \in -N_C(x^0)$, we have $\langle p, x - x^0\rangle \ge 0$, hence $f(x^0) \le f(x)$, i.e., x^0 is a minimizer. Conversely, if $x^0 \in \mathrm{argmin}_{x\in C} f(x)$ then the system

$$(x, y) \in C \times \mathbb{R}^n,\ x - y = 0,\ f(y) - f(x^0) < 0$$

is inconsistent. Define $D := C \times \mathbb{R}^n$ and $A(x, y) := x - y$, so that $A(D) = C - \mathbb{R}^n$. For any ball U around 0, $x^0 + U \subset \mathbb{R}^n$, hence $U = x^0 - (x^0 + U) \subset A(D)$, and so $0 \in \mathrm{int}A(D)$. Therefore, by Theorem 2.4, there exists a vector $p \in \mathbb{R}^n$ such that

$$\langle p, x - y\rangle + f(y) - f(x^0) \ge 0 \quad \forall (x, y) \in C \times \mathbb{R}^n.$$

Letting $y = x^0$ yields $\langle p, x - x^0\rangle \ge 0\ \forall x \in C$, i.e., $p \in -N_C(x^0)$, then letting $x = x^0$ yields $f(y) - f(x^0) \ge \langle p, y - x^0\rangle\ \forall y \in \mathbb{R}^n$, i.e., $p \in \partial f(x^0)$. Thus, $p \in -N_C(x^0) \cap \partial f(x^0)$, completing the proof. □

Corollary 2.12 *Under the assumptions of the above proposition, an interior point x^0 of C is a minimizer if and only if $0 \in \partial f(x^0)$.*

Proof Indeed, $N_C(x^0) = \{0\}$ if $x^0 \in \mathrm{int}C$. □

Proposition 2.32 *Let C be a nonempty compact set in $\mathbb{R}^n$, $f : C \to \mathbb{R}$ an arbitrary continuous function, f^c the convex envelope of f over C. Then any global minimizer of $f(x)$ on C is also a global minimizer of $f^c(x)$ on* conv$C$.

Proof Let $x^0 \in C$ be a global minimizer of $f(x)$ on C. Since f^c is a minorant of f, we have $f^c(x^0) \le f(x^0)$. If $f^c(x^0) < f(x^0)$ then the convex function $h(x) = \max\{f(x^0), f^c(x)\}$ would be a convex minorant of f larger than f^c, which is impossible. Thus, $f^c(x^0) = f(x^0)$ and $f^c(x) = h(x)\ \forall x \in \mathrm{conv}C$. Hence, $f^c(x^0) = f(x^0) \le f^c(x)\ \forall x \in \mathrm{conv}C$, i.e., x^0 is also a global minimizer of $f^c(x)$ on conv$C$. □

2.11.2 Maximum

In contrast with the minimum, a local maximum of a convex function may not be global. Generally speaking, local information is not sufficient to identify a global maximizer of a convex function.

Proposition 2.33 *Let C be a convex set in $\mathbb{R}^n$, and $f : C \to \mathbb{R}$ be a convex function. If $f(x)$ attains its maximum on C at a point $x^0 \in \mathrm{ri}C$ then $f(x)$ is constant on C. The set* $\operatorname{argmax}_{x\in C} f(x)$ *is a union of faces of C.*

Proof Suppose that f attains its maximum on C at a point $x^0 \in \mathrm{ri}C$ and let x be an arbitrary point of C. Since $x^0 \in \mathrm{ri}C$ there is $y \in C$ such that $x^0 = \lambda x + (1-\lambda)y$ for some $\lambda \in (0,1)$. Then $f(x^0) \le \lambda f(x) + (1-\lambda)f(y)$, hence $\lambda f(x) \ge f(x^0) - (1-\lambda)$ $f(y) \ge f(x^0) - (1-\lambda)f(x^0) = \lambda f(x^0)$. Thus $f(x) \ge f(x^0)$, hence $f(x) = f(x^0)$, proving the first part of the proposition. The second part follows, because for any maximizer x^0 there is a face F of C such that $x^0 \in \mathrm{ri}F$: then by the previous argument, any point of this face is a global minimizer. □

Proposition 2.34 *Let C be a closed convex set, and $f : C \to \mathbb{R}$ be a convex function. If C contains no line and $f(x)$ is bounded above on every halfline of C then*

$$\sup\{f(x)|\ x \in C\} = \sup\{f(x)|\ x \in V(C)\},$$

where $V(C)$ is the set of extreme points of C. If the maximum of $f(x)$ is attained at all on C, it is attained on $V(C)$.

Proof By Theorem 1.7, $C = \mathrm{conv}V(C) + K$, where K is the convex cone generated by the extreme directions of C. Any point of C which is actually not an extreme point belongs to a halfline emanating from some $v \in V(C)$ in the direction of a ray of K. Since $f(x)$ is finite and bounded above on this halfline, its maximum on the halfline is attained at v (Proposition 2.12, (ii)). Therefore, the supremum of $f(x)$ on C is reduced to the supremum on $\mathrm{conv}V(C)$. The conclusion then follows from the fact that any $x \in \mathrm{conv}V(C)$ is of the form $x = \sum_{i\in I}\lambda_i v^i$, with $|I| < +\infty, v^i \in V(C), \lambda_i \ge 0, \sum_{i\in I}\lambda_i = 1$, hence $f(x) \le \sum_{i\in I}\lambda_i f(v^i) \le \max_{i\in I} f(v^i)$. □

Corollary 2.13 *A real-valued convex function $f(x)$ on a polyhedron C containing no line is either unbounded above on some unbounded edge or attains its maximum at an extreme point of C.*

Corollary 2.14 *A real-valued convex function $f(x)$ on a compact convex set C attains its maximum at an extreme point of C.*

The latter result is in fact true for a wider class of functions, namely for *quasiconvex* functions. As was defined in Sect. 2.3, these are functions $f : \mathbb{R}^n \to [-\infty, +\infty]$ such that for any real number α, the level set $L_\alpha := \{x \in \mathbb{R}^n |\, f(x) \le \alpha\}$ is convex, or equivalently, such that

$$f((1-\lambda)x^1 + \lambda x^2) \le \max\{f(x^1), f(x^2)\} \tag{2.53}$$

for any $x^1, x^2 \in C$ and any $\lambda \in [0,1]$.

To see that Corollary 2.14 extends to quasiconvex functions, just note that a compact convex set C is the convex hull of its extreme points (Corollary 1.13), so any $x \in C$ can be represented as $x = \sum_{i \in I} \lambda_i v^i$, where v^i are extreme points, $\lambda_i \geq 0$ and $\sum_{i \in I} \lambda_i = 1$. If $f(x)$ is a quasiconvex function finite on C, and $\alpha = \max_{i \in I} f(v^i)$, then $v^i \in C \cap L_\alpha, \ \forall i \in I$, hence $x \in C \cap L_\alpha$, because of the convexity of the set $C \cap L_\alpha$. Therefore, $f(x) \leq \alpha = \max_{i \in I} f(v^i)$, i.e., the maximum of f on C is attained at some extreme point.

A function $f(x)$ is said to be *quasiconcave* if $-f(x)$ is quasiconvex. A convex (concave) function is of course quasiconvex (quasiconcave), but the converse may not be true, as can be demonstrated by a monotone nonconvex function of one variable. Also it is easily seen that an upper envelope of a family of quasiconvex functions is quasiconvex, but the sum of two quasiconvex functions may not be quasiconvex.

2.12 Minimax and Saddle Points

2.12.1 Minimax

Given a function $f(x, y) : C \times D \to \mathbb{R}$ we can compute $\inf_{x \in C} f(x, y)$ and $\sup_{y \in D} f(x, y)$. It is easy to see that there always holds

$$\sup_{y \in D} \inf_{x \in C} f(x, y) \leq \inf_{x \in C} \sup_{y \in D} f(x, y).$$

Indeed, $\inf_{x \in C} f(x, y) \leq f(z, y) \ \forall z \in C, y \in D$, so $\sup_{y \in D} \inf_{x \in C} f(x, y) \leq \sup_{y \in D} f(z, y) \ \forall z \in C$, hence $\sup_{y \in D} \inf_{x \in C} f(x, y) \leq \inf_{z \in C} \sup_{y \in D} f(z, y)$.

We would like to know when the reverse inequality is also true, i.e., when there holds the *minimax equality*

$$\gamma := \sup_{y \in D} \inf_{x \in C} f(x, y) = \inf_{x \in C} \sup_{y \in D} f(x, y) := \eta.$$

Investigations on this question date back to von Neumann (1928). A classical result of his states that if C, D are compact and $f(x, y)$ is convex in x, concave in y and continuous in each variable then the minimax equality holds. Since minimax theorems have found important applications, there has been a great deal of work on the extension of von Neumann's theorem. Almost all these extensions are based either on the separation theorem of convex sets or a fixed point argument. The most important result in this direction is due to Sion (1958). Later a more general minimax theorem was established by Tuy (1974) without any appeal to separation or fixed point argument. We present here a simplified version of the latter result which is also a refinement of Sion's theorem.

Theorem 2.7 *Let C, D be two closed convex sets in $\mathbb{R}^n, \mathbb{R}^m$, respectively, and let $f(x, y) : C \times D \to \mathbb{R}$ be a function quasiconvex, lower semi-continuous in x and quasiconcave, upper semi-continuous in y. Assume that*

(*) *There exists a finite set $N \subset D$ such that* $\sup_{y \in N} f(x, y) \to +\infty$ *as* $x \in C$, $\|x\| \to +\infty$.

Then there holds the minimax equality

$$\inf_{x\in C}\sup_{y\in D} f(x,y) = \sup_{y\in D}\inf_{x\in C} f(x,y). \tag{2.54}$$

Proof Since the inequality $\inf_{x\in C}\sup_{y\in D} f(x,y) \geq \sup_{y\in D}\inf_{x\in C} f(x,y)$ is trivial, it suffices to show the reverse inequality:

$$\inf_{x\in C}\sup_{y\in D} f(x,y) \leq \sup_{y\in D}\inf_{x\in C} f(x,y). \tag{2.55}$$

Let $\eta := \sup_{y\in D}\inf_{x\in C} f(x,y)$. If $\eta = +\infty$ then (2.55) is obvious, so we can assume $\eta < +\infty$. For an arbitrary $\alpha > \eta$ define

$$C_\alpha(y) = \{x \in C | f(x,y) \leq \alpha\}.$$

Since $\sup_{y\in D}\inf_{x\in C} < \alpha$, we have $C_\alpha(y) \neq \emptyset\ \forall y \in D$. If we can show that

$$\bigcap_{y\in D} C_\alpha(y) \neq \emptyset, \tag{2.56}$$

i.e., there is $x \in C$ satisfying $f(x,y) \leq \alpha$ for all $y \in D$, then $\inf_{x\in C}\sup_{y\in D} \leq \alpha$ and since this is true for every $\alpha > \eta$ it will follow that $\inf_{x\in C}\sup_{y\in D} \leq \eta$, proving (2.55). Thus, all is reduced to establishing (2.56). This will be done in three stages. To simplify the notation, from now on we shall omit the subscript α and write simply $C(a), C(b)$, etc…

I. Let us first show that for every pair $a, b \in D$ the two sets $C(a)$ and $C(b)$ intersect. Assume the contrary, that

$$C(a) \cap C(b) = \emptyset. \tag{2.57}$$

Consider an arbitrary $\lambda \in [0,1]$ and let $y_\lambda = (1-\lambda)a + \lambda b$. If $x \in C(y_\lambda)$, i.e., $f(x,y_\lambda) \leq \alpha$ then $\min\{f(x,a), f(x,b)\} \leq f(x,y_\lambda) \leq \alpha$ by quasiconcavity of $f(x,.)$, hence, either $f(x,a) \leq \alpha$ or $f(x,b) \leq \alpha$. Therefore,

$$C(y_\lambda) \subset C(a) \cup C(b). \tag{2.58}$$

Since $C(y_\lambda)$ is convex it follows from (2.58) that $C(y_\lambda)$ cannot meet both sets $C(a)$ and $C(b)$ which are disjoint by assumption (2.57). Consequently, for every $\lambda \in [0,1]$, one and only one of the following alternatives occurs:

$$\text{(a) } C(y_\lambda) \subset C(a); \qquad \text{(b) } C(y_\lambda) \subset C(b).$$

Denote by $M_a(M_b$, resp.) the set of all $\lambda \in [0, 1]$ satisfying (a) (satisfying (b), resp.). Clearly $0 \in M_a, 1 \in M_b, M_a \cup M_b = [0, 1]$ and, analogously to (2.58):

$$C(y_\lambda) \subset C(y_{\lambda_1}) \cup C(y_{\lambda_2}) \quad \forall \lambda \in [\lambda_1, \lambda_2]. \tag{2.59}$$

Therefore, $\lambda \in M_a$ implies $[0, \lambda] \subset M_a$, and $\lambda \in M_b$ implies $[\lambda, 1] \subset M_b$. Let $s = \sup M_a = \inf M_b$ and assume, for instance, that $s \in M_a$ (the argument is similar if $s \in M_b$). We show that (2.57) leads to a contradiction.

Since $\alpha > \eta \geq \inf_{x \in C} f(x, y_s)$, we have $f(\bar{x}, y_s) < \alpha$ for some $\bar{x} \in C$. By upper semi-continuity of $f(\bar{x}, .)$ there is $\varepsilon > 0$ such that $f(\bar{x}, y_{s+\varepsilon}) < \alpha$ and so $\bar{x} \in C(y_{s+\varepsilon})$. But $\bar{x} \in C(y_s) \subset C(a)$, hence $C(y_{s+\varepsilon}) \subset C(a)$, i.e., $s + \varepsilon \in M_a$, contradicting the definition of s. Thus (2.57) cannot occur and we must have $C(a) \cap C(b) \neq \emptyset$ for all $a, b \in C, \alpha < \gamma$.

II. We now show that any finite collection $C(y^1), \ldots, C(y^k)$ with $y^1, \ldots, y^k \in D$ has a nonempty intersection. By (I) this is true for $k = 2$. Assuming this is true for $k = h-1$ let us consider the case $k = h$. Set $C' = C(y^h), C'(y) = C' \cap C(y)$. From part I we know that $C'(y) \neq \emptyset$ for every $y \in D$. This means that for all $\alpha > \eta : \ \forall y \in D \quad \exists x \in C' \quad f(x, y) \leq \alpha$, so that $\sup_{y \in D} \inf_{x \in C'} f(x, y) \leq \alpha$. Since $\eta = \sup_{y \in D} \inf_{x \in C} f(x, y) \leq \sup_{y \in D} \inf_{x \in C'} f(x, y)$, it follows that $\eta = \sup_{y \in D} \inf_{x \in C'} f(x, y)$. So all the hypotheses of the theorem still hold when C is replaced by C'. It then follows from the induction assumption that the sets $C'(y^1), \ldots, C'(y^{h-1})$ have a nonempty intersection. Thus the family $\{C_\alpha(y), y \in D\}$ has the finite intersection property.

III. Finally, for every $y \in D$ let $C^*(y) = \{x \in C |\, f(x, y) \leq \alpha, \sup_{z \in N} f(x, z) \leq \alpha\}$. Then $C^*(y) \subset C^N := \{x \in C|\ \sup_{z \in N} f(x, z) \leq \alpha\}$ and the set C^N is compact because if it were not so there would exist a sequence $x^k \in C^N$ such that $\|x^k| \to +\infty$, contradicting assumption (*). On the other hand, for any finite set $E \subset D$ clearly $\cap_{y \in E} C^*(y) = \cap_{y \in E \cup N} C(y)$, so by part II $\cap_{y \in E} C^*(y) \neq \emptyset$, i.e., the family $\{C^*(y), y \in D\}$ has the finite intersection property. Since every $C^*(y)$ is a subset of the compact set C^N it follows that $\cap_{y \in D} C_\alpha(y) = \cap_{y \in D} C^*(y) \neq \emptyset$, i.e. (2.56) must hold. □

Remark 2.2 In view of the symmetry in the roles of x, y Theorem 2.7 still holds if instead of condition (*) one assumes that

(!) *There exists a finite set* $M \subset C$ *such that* $\inf_{x \in M} f(x, y) \to -\infty$ *as* $y \in D, \|y\| \to +\infty$.

The proof is analogous, using $D_\alpha(x) := \{y \in D |\, f(x, y) \geq \alpha\}$ with $\alpha > \gamma := \inf_{x \in C} \sup_{y \in D} f(x, y)$ (instead of $C_\alpha(y)$ with $\alpha < \eta := \sup_{y \in D} \inf_{x \in C} f(x, y)$) and proving that $\cap_{x \in C} D_\alpha(x) \neq \emptyset$.

2.12.2 Saddle Point of a Function

A pair $(\bar{x}, \bar{y}) \in C \times D$ is called a *saddle point* of the function $f(x, y) : C \times D \to \mathbb{R}$ if

$$f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \leq f(x, \bar{y}) \quad \forall x \in C, \ \forall y \in D. \tag{2.60}$$

This means

$$\min_{x \in C}(x, \bar{y}) = f(\bar{x}, \bar{y}) = \max_{y \in D} f(\bar{x}, y),$$

so in a neighborhood of the point $(\bar{x}, \bar{y}, f(\bar{x}, \bar{y}))$ the set epi$f := \{(x, y, t) | \ (x, y) \in C \times D, \ t \in \mathbb{R}, \ f(x, y) \leq t\}$ reminds the image of a saddle (or a mountain pass).

Proposition 2.35 *A point $(\bar{x}, \bar{y}) \in C \times D$ is a saddle point of $f(x, y) : C \times D \to \mathbb{R}$ if and only if*

$$\max_{y \in D} \inf_{x \in C} f(x, y) = \min_{x \in C} \sup_{y \in D} f(x, y). \tag{2.61}$$

Proof We have

$$\begin{aligned}
(2.60) &\Leftrightarrow \sup_{y \in D} f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \leq \inf_{y \in C} f(x, \bar{y}) \\
&\Leftrightarrow \inf_{x \in C} \sup_{y \in D} f(x, y) \leq \sup_{y \in D} f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \\
&\quad \leq \inf_{x \in C} f(x, \bar{y}) \leq \sup_{y \in D} \inf_{x \in C} f(x, y).
\end{aligned}$$

Since always $\inf_{x \in C} \sup_{y \in D} f(x, y) \geq \sup_{y \in D} \inf_{x \in C} f(x, y)$ we must have equality everywhere in the above sequence of inequalities. Therefore, (2.60) must be equivalent to (2.61). □

As a consequence of Theorem 2.7 and Remark 2.2 we can now state

Proposition 2.36 *Assume that $C \subset \mathbb{R}^n$, $D \subset \mathbb{R}^m$ are nonempty closed convex sets and the function $f(x, y) : C \times D \to \mathbb{R}$ is quasiconvex l.s.c in x and quasiconcave u.s.c. in y. If both the following conditions hold:*

(i) *There exists a finite set $N \subset D$ such that $\sup_{y \in N} f(x, y) \to +\infty$ as $x \in C, \ \|x\| \to +\infty$;*
(ii) *There exists a finite set $M \subset C$ such that $\inf_{x \in M} f(x, y) \to +\infty$ as $y \in D, \ \|y\| \to +\infty$;*

then there exists a saddle point $(\bar{x}, \bar{y})$ for the function $f(x, y)$.

Proof If (i) holds, we have $\min_{x \in C} \sup_{y \in D} f(x, y) = \sup_{y \in D} \inf_{x \in C} f(x, y)$ by Theorem 2.7. If (ii) holds, we have $\inf_{x \in C} \sup_{y \in D} f(x, y) = \max_{y \in D} \inf_{x \in C} f(x, y)$ by Remark 2.2. Hence the condition (2.61). □

2.13 Convex Optimization

We assume that the reader is familiar with convex optimization problems, i.e., optimization problems of the form

$$\min\{f(x)|\ g_i(x) \le 0\ (i = 1,\dots,m),\ h_j(x) = 0\ (j = 1,\dots,p),\ x \in \Omega\}, \qquad (2.62)$$

where $\Omega \subset \mathbb{R}^n$ is a closed convex set, $f, g_1,\dots,g_m$ are convex functions finite on an open domain containing Ω, $h_1,\dots,h_m$ are affine functions.

In this section we study a generalization of problem (2.62), where the convex inequality constraints are understood in a generalized sense.

2.13.1 *Generalized Inequalities*

a. Ordering Induced by a Cone

A cone $K \subset \mathbb{R}^n$ induces a partial ordering $\preceq_K$ on $\mathbb{R}^n$ such that

$$x \preceq_K y \quad \Leftrightarrow \quad y - x \in K.$$

We also write $x \succeq_K y$ to mean $y \preceq_K x$. In the case $K = \mathbb{R}^n_+$ this is the usual ordering $x \le y \ \Leftrightarrow x_i \le y_i,\ i = 1,\dots,m.$

The following properties of the ordering $\preceq_K$ are straightforward:

(i) transitivity: $x \preceq_K y,\ y \preceq_K z \Rightarrow x \preceq_K z$;
(ii) reflexivity: $x \preceq_K x\ \forall x \in \mathbb{R}^n$;
(iii) preservation under addition: $x \preceq_K y,\ x' \preceq_K y' \ \Rightarrow x + x' \preceq_K y + y'$;
(iv) preservation under nonnegative scaling: $x \preceq_K y,\ \alpha > 0 \ \Rightarrow \alpha x \preceq_K \alpha y$.

The ordering induced by a cone K is of particular interest when K is *closed, solid* (i.e., has nonempty interior), and *pointed* (i.e., contains no line: $x \in K \Rightarrow -x \notin K$). In that case a relation $x \preceq_K y$ is called a *generalized inequality*. We also write $x \prec_K y$ to mean that $y - x \in \text{int}K$ and call such a relation a *strict generalized inequality*.

Generalized inequalities enjoy the following important properties:

(i) $x \preceq_K y,\ y \preceq_K x \ \Rightarrow x = y$;
(ii) $x \not\prec_K x$;
(iii) $x^k \preceq_K y^k,\ x^k \to x,\ y^k \to y \Rightarrow x \preceq_K y$;
(iv) If $x \prec_K y$ then for u, v small enough $x + u \prec_K y + v$;
(v) $x \prec_K y,\ u \preceq_K \Rightarrow x + u \prec_K y + v$;
(vi) The set $\{z|\ x \preceq_K z \preceq_K y\}$ is bounded.

For instance, (ii) is true because $x \prec_K x$ implies $0 = x - x \in \text{int}K$, which is impossible as K is pointed; (vi) holds because the set $E = \{z|\ x \preceq_K z \preceq_K y\}$

is closed, so if there exists $\{z^k\} \subset E, \|z^k\| \to +\infty$ then, up to a subsequence, $z^k/\|z^k\| \to u$ with $\|u\| = 1$ and since $\frac{z^k}{\|z^k\|} \in \frac{x}{\|z^k\|} + K$, $\frac{z^k}{\|z^k\|} \in \frac{x}{\|z^k\|} - K$, letting $k \to +\infty$ yields $u \in K, u \in -K$, hence by (i) $u = 0$ conflicting with $\|u\| = 1$.

b. Dual Cone

The *dual cone* of a cone K is by definition the cone

$$K^* = \{y \in \mathbb{R}^n | \ \langle x, y\rangle \geq 0 \ \forall x \in K\}.$$

Note that $K^* = -K^o$, where K^o is the polar of K (Sect. 1.8). As can easily be seen:

(i) K^* is a closed convex cone;
(ii) K^* is also the dual cone of the closure of K;
(iii) $(K^*)^* = \text{cl}K$;
(iv) If $x \in \text{int}K$ then $\langle x, y\rangle > 0 \ \forall y \in K^* \setminus \{0\}$.

A cone K is said to be *self-dual* if $K^* = K$. Clearly the orthant $\mathbb{R}^n_+$ is a self-dual cone.

Lemma 2.1 *If K is a closed convex cone then*

$$y \notin K \ \Leftrightarrow \ \exists \lambda \in K^* \ \langle \lambda, y\rangle < 0.$$

Proof By (iii) above $K = (K^*)^*$ so $y \notin K$ if and only if $y \notin (K^*)^*$, hence if and only if there exists $\lambda \in K^*$ satisfying $\langle \lambda, y\rangle < 0$. □

c. *K*-Convex Functions

Given a convex cone $K \subset \mathbb{R}^m$ inducing an ordering $\preceq_K$ on $\mathbb{R}^m$ a map $g : \mathbb{R}^n \to \mathbb{R}^m$ is called K-convex if for every $x^1, x^2 \in \mathbb{R}^n$ and $0 \leq \alpha \leq 1$ we have the generalized inequality

$$g(\alpha x^1 + (1-\alpha)x^2) \preceq_K \alpha g(x^1) + (1-\alpha)g(x^2). \tag{2.63}$$

For instance, the map $g : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathbb{R}^m_+$-convex (or component-wise convex) if each function $g_i(x), i = 1 \ldots, m$, is convex.

Lemma 2.2 *If $g : \mathbb{R}^n \to \mathbb{R}^m$ is a K-convex map then for every $\lambda \in K^*$ the function $\langle \lambda, g(x)\rangle = \sum_{i=1}^m \lambda_i g_i(x)$ is convex in the usual sense and the set $\{x \in \mathbb{R}^n | \ g(x) \preceq_K 0\}$ is convex.*

Proof For every $x^1, x^2 \in \mathbb{R}^n$, we have by (2.63)

$$\alpha g(x^1) + (1-\alpha)g(x^2) - g(\alpha x^1 + (1-\alpha)x^2) \in K.$$

Since $\lambda \in K^*$ it follows that

$$\langle \lambda, \alpha g(x^1) + (1-\alpha)g(x^2) - g(\alpha x^1) + (1-\alpha)g(x^2)\rangle \geq 0,$$

hence $\alpha\langle \lambda, g(x^1)\rangle + (1-\alpha)\langle \lambda, g(x^1)\rangle \geq \langle \lambda, g(\alpha x^1) + (1-\alpha)g(x^2)\rangle$, proving that the function $\langle \lambda, g(x)\rangle$ is convex.

Further, by (2.2), if $g(x^1) \preceq_K 0,\ g(x^2) \preceq_K 0$ then

$$g(\alpha x^1) + (1-\alpha)x^2) \preceq_K \alpha g(x^1) + (1-\alpha)g(x^2) \preceq_K 0,$$

proving that the set $\{x|\ g(x) \preceq_K 0\}$ is convex. □

2.13.2 Generalized Convex Optimization

A *generalized convex optimization problem* is a problem of the form

$$\min\{f(x)|\ g_i(x) \preceq_{K_i} 0\ (i=1,\ldots,m),\ h(x)=0,\ x \in \Omega\}, \tag{2.64}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, K_i is a closed, solid, pointed convex cone in $\mathbb{R}^{s_i}$, $g_i : \mathbb{R}^n \to \mathbb{R}^{s_i}$, $i = 1,\ldots,m$, is K_i-convex, finite on the whole $\mathbb{R}^n$, $h : \mathbb{R}^n \to \mathbb{R}^p$ is an affine map, and $\Omega \subset \mathbb{R}^n$ is a closed convex set. By Lemma 2.2 the constraint set of this problem is convex, so this is also a problem of minimizing a convex function over a convex set.

Let $\lambda_i \in K_i^*$ be the Lagrange multiplier associated with the generalized inequality $g_i(x) \preceq_{K_i} 0$ and $\mu \in \mathbb{R}^p$ the Lagrange multiplier associated with the equality $h(x) = 0$. So the Lagrangian of the problem is the function

$$L(x,\lambda,\mu) := f(x) + \sum_{i=1}^{m}\langle \lambda_i, g_i(x)\rangle + \langle \mu, h(x)\rangle,$$

where $\lambda_i \in K_i^*, i = 1,\ldots,m$, and $\mu \in \mathbb{R}^p$. The dual Lagrange function is

$$\varphi(\lambda,\mu) = \inf_{x\in\Omega} L(x,\lambda,\mu) = \inf_{x\in\Omega}\{f(x) + \sum_{I=1}^{m}\langle \lambda_i, g_i(x)\rangle + \langle \mu, h(x)\rangle\}.$$

Since $\varphi(\lambda,\mu)$ is the lower envelope of a family of affine functions in (λ,μ), it is a concave function. Setting $K^* = K_1^* \times \cdots \times K_m^*,\ \lambda = (\lambda_1,\ldots,\lambda_m)$ we can show that

$$\sup_{\lambda\in K^*,\mu\in\mathbb{R}^p} L(x,\lambda,\mu)$$
$$= \begin{cases} f(x) & \text{if } g_i(x) \preceq_{K_i} 0\ (i=1,\ldots,m), h(x) = 0; \\ +\infty & \text{otherwise} \end{cases}$$

In fact, if $g_i(x) \preceq_{K_i} 0$ $(i = 1, \dots, m), h(x) = 0$ then the supremum is attained for $\lambda = 0, \mu = 0$ and equals $f(x)$. On the other hand, if $h(x) \neq 0$ there is $\mu \in \mathbb{R}^p$ satisfying $\langle \mu, h(x) \rangle > 0$ and for $\lambda = 0$ the supremum is equal to $\sup_{\theta > 0}\{f(x) + \langle \theta\mu, h(x)\rangle\} = +\infty$. If there is an $i = 1, \dots, m$ with $g_i(x) \not\preceq_{K_i} 0$ then by Lemma 2.1 there is $\lambda_i \in K_i^*$ satisfying $\langle \lambda_i, g_i(x)\rangle > 0$, hence for $\mu = 0, \lambda_j = 0\ \forall j \neq i$, the supremum equals $\sup_{\theta>0}\{f(x) + \theta\langle \lambda_i, g_i(x)\rangle\} = +\infty$.

Thus, the problem (2.64) can be written as

$$\inf_{x \in \Omega} \sup_{\lambda \in K^*, \mu \in \mathbb{R}^p} L(x, \lambda, \mu).$$

The dual problem is

$$\sup_{\lambda \in K^*, \mu \in \mathbb{R}^p} \inf_{x \in \Omega} L(x, \lambda, \mu),$$

that is,

$$\sup_{\lambda \in K^*, \mu \in \mathbb{R}^p} \varphi(\lambda, \mu). \tag{2.65}$$

Theorem 2.8

(i) (weak duality) *The optimal value in the dual problem never exceeds the optimal value in the primal problem (2.64).*

(ii) (strong duality) *The optimal values in the two problems are equal if the Slater condition holds, i.e., if*

$$\exists \bar{x} \in \mathrm{int}\Omega \quad h(\bar{x}) = 0,\ g_i(\bar{x}) \prec_{K_i} 0,\ i = 1, \dots, m.$$

Proof (i) is straightforward, we need only prove (ii). By Lemma 2.2 for every $\lambda_i \in K_i^*$ the function $\langle \lambda_i, g_i(x)\rangle$ is convex (in the usual sense), finite on $\mathbb{R}^n$ and hence continuous. So $L(x, \lambda, \mu)$ is a convex continuous function in $x \in \Omega$ for every fixed $(\lambda, \mu) \in D := K_1^* \times \cdots \times K_m^* \times \mathbb{R}^p$ and affine in $(\lambda, \mu) \in D$ for every fixed $x \in h(\Omega)$.

Since $h : \mathbb{R}^n \to \mathbb{R}^p$ is an affine map, without loss of generality it can be assumed that $h(\Omega) = \mathbb{R}^p$. Since $h(\bar{x}) = 0$ and $\bar{x} \in \mathrm{int}\Omega$, we have $0 \in \mathrm{int}\Omega$, so for every $j = 1, \dots, p$ there is an $a^j \in \Omega$ such that $h(a^j)$ has its j-th component equal to 1, and all other components equal to 0. Then for sufficiently small $\varepsilon > 0$, we have $x^i := \bar{x} + \varepsilon(a^j - \bar{x}) \in \Omega$, $\hat{x}^j := \bar{x} - \varepsilon(a^j - \bar{x}) \in \Omega$, and so

$$g_i(x^j) < 0\ (i = 1, \dots, m),\ h_j(x^j) > 0,\ h_j(x^j) = 0\ \forall i \neq j$$
$$g_i(\hat{x}^j) < 0\ (i = 1, \dots, m),\ h_j(\hat{x}^j) < 0,\ h_i(\hat{x}^j) = 0\ \forall i \neq j.$$

Let $M = \{x^j, \hat{x}^j, j = 1, \dots, p\}$. For every $(\lambda, \mu) \in D \setminus \{0\}$ we can write

$$\rho(\lambda, \mu) := \min_{x \in M}[\sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^p \mu_j h_j(x)] < 0$$

hence, $\theta := \max\{\rho(\lambda,\mu) |\ \lambda \in \mathbb{R}^m_+,\ \mu \in \mathbb{R}^p, \|\lambda\| + \|\mu\| = 1\} < 0$. Consequently,

$$\min_{x\in M} L(x,(\lambda,\mu)) \leq \max_{x\in M} f(x) + \min_{x\in M}\{\sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^p \mu_j h_j(x)\}$$
$$\leq \max_{x\in M} f(x) + (\|\lambda\| + \|\mu\|)\theta \to -\infty$$

as $(\lambda,\mu) \in D, \|\lambda\| + \|\mu\| \to +\infty$. So the function $L(x,(\lambda,\mu))$ satisfies condition (!) in Remark 2.2 with $C = \Omega, y = (\lambda,\mu)$. By virtue of this Remark,

$$\inf_{x\in\Omega} \sup_{(\lambda,\mu)\in D} L(x,(\lambda,\mu)) = \max_{(\lambda,\mu)\in D} \inf_{x\in\Omega} L(x,(\lambda,\mu)),$$

completing the proof of (ii). □

The difference between the optimal values in the primal and dual problems is called the *duality gap*. Theorem 2.8 says that for problem (2.64) the duality gap is never negative and is zero if Slater condition is satisfied.

2.13.3 Conic Programming

Let $K \subset \mathbb{R}^m$ be a closed, solid, and pointed convex cone, let $c \in \mathbb{R}^n$ and let A be an $m \times n$ matrix, and $b \in \mathbb{R}^m$. Then the following optimization problem:

$$\min\{\langle c,x\rangle |\ Ax \preceq_K b\} \tag{2.66}$$

is called a *conic programming problem*.

Since $A(\alpha x^1 + (1-\alpha)x^2) - [\alpha Ax^1 + (1-\alpha)Ax^2] = 0 \in K$, i.e.,

$$A(\alpha x^1 + (1-\alpha)x^2) \succeq_K \alpha Ax^1 + (1-\alpha)Ax^2,$$

for all $x^1, x^2 \in \mathbb{R}^n, 0 \leq \alpha \leq 1$, the map $x \mapsto b - Ax$ is K-convex. So a conic program is nothing but a special case of the above considered problem (2.64) when $m = 1, K_1 = K, f(x) = \langle c,x\rangle, g_1(x) = b - Ax, h \equiv 0, \Omega = \mathbb{R}^n$.

The Lagrangian of problem (2.66) is

$$L(x,y) = \langle c,x\rangle + \langle y, b - Ax\rangle = \langle c - A^T y, x\rangle + \langle b,y\rangle \quad (y \succeq_K 0).$$

But, as can easily be seen,

$$\inf_{x\in\mathbb{R}^n} L(x,y) = \begin{cases} \langle b,y\rangle \text{ if } A^T y = x \\ -\infty \quad \text{otherwise} \end{cases}$$

so the dual of the conic programming problem (2.66) is the problem

$$\max\{\langle b, y\rangle \mid A^T y = c,\ y \succeq_K 0\}. \tag{2.67}$$

Clearly linear programming is a special case of conic programming when $K = \mathbb{R}^n_+$. However, while strong duality always holds for linear programming (except only when both the primal and the dual problems are infeasible), it is not so for conic programming. By Theorem 2.8 a sufficient condition for strong duality in conic programming is

$$\exists \bar{x} \quad A\bar{x} \succ_K b.$$

2.14 Semidefinite Programming

2.14.1 The SDP Cone

Given a symmetric $n \times n$ matrix $A = [a_{ij}]$ (i.e., a matrix A such that $A^T = A$) the *trace* of A, written $\mathrm{Tr}(A)$, is the sum of its diagonal elements:

$$\mathrm{Tr}(A) := \sum_{i=1}^{n} a_{ii}.$$

From the definition it is readily seen that

$$\mathrm{Tr}(\alpha A + \beta B) = \alpha \mathrm{Tr}(A) + \beta \mathrm{Tr}(B), \quad \mathrm{Tr}(AB) = \mathrm{Tr}(BA)$$
$$\mathrm{Tr}(A) = \lambda_1 + \cdots + \lambda_n$$

where $\lambda_i, i = 1, \ldots, n$, are the eigenvalues of A, the latter equality being derived from the development of the characteristic polynomial $\det(\lambda I_n - A)$.

Consider now the space $\mathbf{S}^n$ of all $n \times n$ symmetric matrices. Using the concept of trace we can introduce an *interior product*[1] in $\mathbf{S}^n$ defined as follows:

$$\langle A, B\rangle = \mathrm{Tr}(AB) = \sum_{i,j} a_{ij} b_{ij} = \mathrm{vec}(A)^T \mathrm{vec}(B),$$

where $\mathrm{vec}(A)$ denotes the $n \times n$ column vector whose elements are elements of the matrix A in the order $a_{11}, \cdots, a_{1n}, a_{21}, \cdots, a_{2n}, \cdots, a_{n1}, a_{nn}$. The norm of a matrix A associated with this inner product is the *Frobenius norm* given by

[1] Sometimes also written as $A \bullet B$ and called *dot product*.

$$\|A\|_F = (\langle A, A\rangle)^{1/2} = (\mathrm{Tr}(A^T A)^{1/2} = \left(\sum_{i,j} a_{ij}^2\right)^{1/2}.$$

The set of semidefinite positive matrices $X \in \mathbf{S}^n$ forms a convex cone $\mathbf{S}^n_+$ called the *SDP cone* (semidefinite positive cone). For any two matrices $A, B \in \mathbf{S}^n$ we write $A \preceq B$ to mean that $B - A$ is semidefinite positive. In particular, $A \succeq 0$ means A is a semidefinite positive matrix.

Lemma 2.3 *The cone $\mathbf{S}^n_+$ is closed, convex, solid, pointed, and self-dual, i.e.,*

$$\mathbf{S}^n_+ = (\mathbf{S}^n_+)^*.$$

Proof We prove only the self-dual property. By definition

$$(\mathbf{S}^n_+)^* = \{Y \in \mathbf{S}^n_+ \mid \mathrm{Tr}(YX) \geq 0 \ \forall X \in \mathbf{S}^n_+\}.$$

Let $Y \in \mathbf{S}^n_+$. For every $X \in \mathbf{S}^n_+$, we have

$$\mathrm{Tr}(YX) = \mathrm{Tr}(YX^{1/2}X^{1/2}) = \mathrm{Tr}(X^{1/2}YX^{1/2}) \geq 0,$$

where the last inequality holds because the matrix $X^{1/2}YX^{1/2}$ is semidefinite positive. Therefore, $\mathbf{S}^n_+ \subset (\mathbf{S}^n_+)^*$. Conversely, let $Y \in (\mathbf{S}^n_+)^*$. For any $u \in \mathbb{R}^n$, we have

$$\mathrm{Tr}(Yuu^T) = \sum_{i=1}^n y_{1i}u_iu_1 + \cdots + \sum_{i=1}^n y_{ni}u_iu_n = \sum_{i,j=1}^n y_{ij}u_iu_j = u^TYu.$$

The matrix uu^T is obviously semidefinite positive, while the trace of Yuu^T is nonnegative by assumption, so the product u^TYu is nonnegative. Since u is arbitrary, this means $Y \in \mathbf{S}^n_+$. Hence, $(\mathbf{S}^n_+)^* \subset \mathbf{S}^n_+$. □

A map $F : \mathbb{R}^n \to \mathbf{S}^m$ is said to be *convex*, or more precisely, *convex with respect to matrix inequalities* if it is $\mathbf{S}^m_+$-convex, i.e., such that for every $X, Y \in \mathbf{S}^n$ and $0 \leq t \leq 1$

$$F(tX + (1-t)Y) \preceq tF(X) + (1-t)F(Y).$$

For instance, the map $X \mapsto X^2$ is convex because for every $u \in \mathbb{R}^m$ the function $u^TX^2u = \|Xu\|^2$ is convex quadratic with respect to the components of X and so

$$u^T(\lambda X + (1-\lambda)X)^2u \leq \lambda u^TX^2u + (1-\lambda)u^TX^2u,$$

which implies $(\lambda X + (1-\lambda)X)^2 \preceq \lambda X^2 + (1-\lambda)X^2$. Analogously, the function $X \mapsto XX^T$ is convex.

2.14.2 Linear Matrix Inequality

A linear matrix inequality (LMI) is a generalized inequality

$$A_0 + x_1A_1 + \cdots + x_nA_n \preceq 0$$

where $x \in \mathbb{R}^n$ is the variable and $A_i \in \mathbf{S}^p, i = 0, 1, \ldots, n$, are given $p \times p$ symmetric matrices. The inequality sign $\preceq$ is understood with respect to the cone $\mathbf{S}^p_+$: the notation $P \preceq 0$ means the matrix P is semidefinite negative. Obviously, $A(x) := A_0 + \sum_{k=1}^n x_kA_k \in \mathbf{S}^p_+$ and each element of this matrix is an affine function of x :

$$A(x) = [a_{ij}(x)],\ a_{ij}(x) = a^0_{ij} + \sum_{k=1}^n a^k_{ij}x_k.$$

Therefore an LMI can also be defined as an inequality of the form

$$A(x) \preceq 0,$$

where $A(x)$ is a square symmetric matrix whose every element is an affine function of x.

By definition

$$A(x) \preceq 0 \Leftrightarrow \langle y, A(x)y\rangle \leq 0\ \forall y \in \mathbb{R}^p,$$

so setting $C := \{x \in \mathbb{R}^n |\ A(x) \preceq 0\}$, we have

$$C = \cap_{y\in\mathbb{R}^p}\{x \in \mathbb{R}^n |\ \langle y, A(x)y\rangle \leq 0\}.$$

Since for every fixed y the set $\{x \in \mathbb{R}^n |\ \langle y, A(x)y\rangle \leq 0\}$ is a halfspace, we see that the solution set C of an LMI is a closed convex set. In other words, an LMI is nothing but a specific convex inequality which is equivalent to an infinite system of linear inequalities.

Obviously, the inequality $A(x) \succeq 0$ is also an LMI, determining a convex constraint for x. Furthermore, a finite system of LMI's of the form

$$A^{(1)}(x) \preceq 0, \ldots, A^{(m)}(x) \preceq 0$$

can be equivalently written as the single LMI

$$\mathrm{Diag}(A^{(1)}(x), \cdots, A^{(m)}(x)) \preceq 0.$$

2.14.3 SDP Programming

A *semidefinite program* (*SDP*) is a problem of minimizing a linear function under an LMI constraint, i.e., a problem of the form

$$(SDP) \qquad \min\{\langle c, x\rangle \mid A_0 + x_1A_1 + \ldots + x_nA_n \preceq 0\}.$$

Clearly this is a special case of generalized convex optimization. Specifically, (*SDP*) can be rewritten in the form (2.64), with $m = 1, f(x) = \langle c, x\rangle, g_1(x) = A_0 + x_1A_1 + \ldots + x_nA_n, K_1 = \mathbf{S}^p_+$.

To form the dual problem to (*SDP*) we associate with the LMI constraint a dual variable $Y \in (\mathbf{S}^p_+)^* = \mathbf{S}^p_+$ (see Lemma 2.3), so the Lagrangian is

$$L(x, Y) = c^T x + \mathrm{Tr}(Y(A_0 + x_1A_1 + \cdots + x_nA_n)).$$

The dual function is

$$\varphi(Y) = \inf\{L(x, Y) \mid x \in \mathbb{R}^n\}.$$

Since $L(x, Y)$ is affine in x it is unbounded below, except if it is identically zero, i.e., if $c_i + \mathrm{Tr}(YA_i) = 0, i = 1, \ldots, n$, in which case $L(x, Y) = \mathrm{Tr}(A_0Y)$. Therefore,

$$\varphi(Y) = \begin{cases} \mathrm{Tr}(A_0Y \text{ if } \mathrm{Tr}(A_iY) + c_i = 0, i = 1, \ldots, n \\ -\infty \qquad\qquad\qquad \text{otherwise} \end{cases}$$

Consequently, the dual of (*SDP*) is

$$(SDD) \qquad \max\{\mathrm{Tr}(A_0Y) \mid \mathrm{Tr}(A_iY) + c_i = 0, i = 1, \ldots, n, Y = Y^T \succeq 0\}.$$

Writing this problem in the form

$$\max\{\langle A_0, Y\rangle \mid \langle A_i, Y\rangle = -c_i, i = 1, \ldots, , Y \succeq 0\} \tag{2.68}$$

we see that (*SDP*) reminds a linear program in standard form.

By Theorem 2.8 strong duality holds for (*SDP*) if Slater condition is satisfied

$$\exists \bar{x} \quad A_0 + \bar{x}_1A_1 + \cdots + \bar{x}_nA_n. \tag{2.69}$$

2.15 Exercises

1 Let $f(x)$ be a convex function and $C = \mathrm{dom}f \subset \mathbb{R}^n$. Show that for all $x^1, x^2 \in C$ and $\lambda \notin [0, 1]$ such that $\lambda x^1 + (1 - \lambda)x^2 \in C$:

$$f(\lambda x^1 + (1 - \lambda)x^2) \geq \lambda f(x^1) + (1 - \lambda)f(x^2).$$

2 A real-valued function $f(t), -\infty < t < +\infty$, is strictly convex (cf Remark 2.1) if it has a strictly monotonically increasing derivative $f'(t)$. Apply this result to $f(t) = e^t$ and show that for any positive numbers $\alpha_1, \dots \alpha_k$;

$$\left(\prod_{i=1}^{k} \alpha_i\right)^{1/k} \le \frac{1}{k}\sum_{i=1}^{k} \alpha_i$$

with equality holding only when $\alpha_1 = \ldots = \alpha_k$.

3 A function $f(x)$ is strongly convex (cf Sect. 2.9) on a convex set C if and only if there exists $r > 0$ (modulus of strong convexity) such that $f(x) - r\|x\|^2$ is convex. Show that:

$$f((1-\lambda)x^1 + \lambda x^2) \ge (1-\lambda)f(x^1) + \lambda f(x^2) - (1-\lambda)\lambda r\|x^1 - x^2\|^2$$

for all $x^1, x^2 \in C$ and $\lambda \notin [0,1]$ such that $\lambda x^1 + (1-\lambda)x^2 \in C$.

4 Show that if $f(x)$ is strongly convex on $\mathbb{R}^n$ (with modulus of strong convexity r) then for any $x^0 \in \mathbb{R}^n$ and $p \in \partial f(x^0)$:

$$f(x) - f(x^0) \ge \langle p, x - x^0\rangle + r\|x - x^0\|^2 \ \forall x \in \mathbb{R}^n.$$

5 Notations being the same as in Exercise 4, show that for any $x^1, x^2 \in \mathbb{R}^n$ and $p^1 \in \partial f(x^1), p^2 \in \partial f(x^2)$:

$$\langle p^1 - p^2, x^1 - x^2\rangle \ge r\|x^1 - x^2\|^2.$$

6 If $f(x)$ is strongly convex on a convex set C then for any $x^0 \in C$ the level set $\{x \in C | f(x) \le f(x^0)\}$ is bounded.

7 Let C be a nonempty convex set in $\mathbb{R}^n$, $f : C \to \mathbb{R}$ a convex function, Lipschitzian with constant L on C. The function

$$F(x) = \inf\{f(y) + L\|x - y\| \,|\, y \in C\}$$

is convex, Lipschitzian with constant L on the whole space and satisfies $F(x) = f(x) \ \forall x \in C$.

8 Show that if $\partial_\varepsilon f(x^0)$ is a singleton for some $x^0 \in \mathrm{dom} f$ and $\varepsilon > 0$, then $f(x)$ is an affine function.

9 For a proper convex function f : $p \in \partial f(x^0)$ if and only if $(p, -1)$ is an outward normal to the set $\mathrm{epi} f$ at point $(x^0, f(x^0))$.

10 Let M be a nonempty set in $\mathbb{R}^n$, $h : M \to \mathbb{R}$ an arbitrary function, E an $n \times n$ matrix. The function

$$\varphi(x) = \max\{\langle x, Ey\rangle - h(y)|\ y \in M\}$$

is convex and for every $x^0 \in \mathrm{dom}\varphi$, if $y^0 \in \mathrm{argmax}_{y\in M}\{\langle x^0, Ey\rangle - h(y)\}$ then $Ey^0 \in \partial\varphi(x^0)$.

11 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, X a closed convex subset of $\mathbb{R}^n, A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{m\times p}, c \in \mathbb{R}^m$. Show that the function $\varphi(y) = \min\{f(x)|\ Ax + By \leq c, x \in X\}$ is convex and for every $y^0 \in \mathrm{dom}\varphi$, if λ is a Kuhn–Tucker vector for the problem $\min\{f(x)|\ Ax + By^0 \leq c, x \in X\}$ then the vector $B^T\lambda$ is a subgradient of φ at y^0.

http://www.springer.com/978-3-319-31482-2

Convex Analysis and Global Optimization
Tuy, H.
2016, XVI, 505 p. 36 illus., Hardcover
ISBN: 978-3-319-31482-2