1. Convex Set and Function
2. Separation Theorem
3. Short Fenchel Conjugation
4. Subgradient
5. Fenchel-conjugate
6. Directional Derivative
7. Kuhn-Tucker Theorem

CHAPTER 1

Convex Sets and Functions

This chapter presents definitions, examples, and basic properties of convex sets and functions in the Euclidean space $\mathbb{R}^n$ and also contains some related material.

1.1 PRELIMINARIES

We start with reviewing classical notions and properties of the Euclidean space $\mathbb{R}^n$. The proofs of the results presented in this section can be found in standard books on advanced calculus and linear algebra.

Let us denote by $\mathbb{R}^n$ the set of all $n-$tuples of real numbers $x = (x_1, \ldots, x_n)$. Then $\mathbb{R}^n$ is a linear space with the following operations:

$$\begin{aligned}&(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n),\\&\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n),\end{aligned}$$

where $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. The zero element of $\mathbb{R}^n$ and the number zero of $\mathbb{R}$ are often denoted by the same notation 0 if no confusion arises.

For any $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we identify it with the column vector $x = [x_1, \ldots, x_n]^T$, where the symbol "T" stands for *vector transposition*. Given $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, the *inner product* of x and y is defined by

$$\langle x, y\rangle := \sum_{i=1}^{n} x_i y_i.$$

The following proposition lists some important properties of the inner product in $\mathbb{R}^n$.

Proposition 1.1 *For $x, y, z \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, we have:*

(i) $\langle x, x\rangle \geq 0$, *and* $\langle x, x\rangle = 0$ *if and only if* $x = 0$.
(ii) $\langle x, y\rangle = \langle y, x\rangle$.
(iii) $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$.
(iv) $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$.

The Euclidean *norm* of $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is defined by

$$\|x\| := \sqrt{x_1^2 + \ldots + x_n^2}.$$

It follows directly from the definition that $\|x\| = \sqrt{\langle x, x\rangle}$.

Proposition 1.2 *For any $x, y \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, we have:*

(i) $\|x\| \geq 0$, *and* $\|x\| = 0$ *if and only if* $x = 0$.
(ii) $\|\lambda x\| = |\lambda| \cdot \|x\|$.
(iii) $\|x + y\| \leq \|x\| + \|y\|$ (*the triangle inequality*).
(iv) $|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$ (*the Cauchy-Schwarz inequality*).

Using the Euclidean norm allows us to introduce the balls in $\mathbb{R}^n$, which can be used to define other topological notions in $\mathbb{R}^n$.

Definition 1.3 *The* CLOSED BALL *centered at $\bar{x}$ with radius $r \geq 0$ and the* CLOSED UNIT BALL *of $\mathbb{R}^n$ are defined, respectively, by*

$$I\!B(\bar{x}; r) := \big\{x \in \mathbb{R}^n \;\big|\; \|x - \bar{x}\| \leq r\big\} \;\; \text{and} \;\; I\!B := \big\{x \in \mathbb{R}^n \;\big|\; \|x\| \leq 1\big\}.$$

It is easy to see that $I\!B = I\!B(0; 1)$ and $I\!B(\bar{x}; r) = \bar{x} + rI\!B$.

Definition 1.4 *Let $\Omega \subset \mathbb{R}^n$. Then $\bar{x}$ is an* INTERIOR POINT *of Ω if there is $\delta > 0$ such that*

$$I\!B(\bar{x}; \delta) \subset \Omega.$$

The set of all interior points of Ω is denoted by $\operatorname{int} \Omega$. *Moreover, Ω is said to be* OPEN *if every point of Ω is its interior point.*

We get that Ω is open if and only if for every $\bar{x} \in \Omega$ there is $\delta > 0$ such that $I\!B(\bar{x}; \delta) \subset \Omega$. It is obvious that the empty set $\emptyset$ and the whole space $\mathbb{R}^n$ are open. Furthermore, any *open ball* $B(\bar{x}; r) := \{x \in \mathbb{R}^n \mid \|x - \bar{x}\| < r\}$ centered at $\bar{x}$ with radius r is open.

Definition 1.5 *A set $\Omega \subset \mathbb{R}^n$ is* CLOSED *if its complement $\Omega^c = \mathbb{R}^n \setminus \Omega$ is open in $\mathbb{R}^n$.*

It follows that the empty set, the whole space, and any ball $I\!B(\bar{x}; r)$ are closed in $\mathbb{R}^n$.

Proposition 1.6 **(i)** *The union of any collection of open sets in $\mathbb{R}^n$ is open.*

(ii) *The intersection of any finite collection of open sets in $\mathbb{R}^n$ is open.*
(iii) *The intersection of any collection of closed sets in $\mathbb{R}^n$ is closed.*
(iv) *The union of any finite collection of closed sets in $\mathbb{R}^n$ is closed.*

Definition 1.7 *Let $\{x_k\}$ be a sequence in $\mathbb{R}^n$. We say that $\{x_k\}$* CONVERGES *to $\bar{x}$ if $\|x_k - \bar{x}\| \to 0$ as $k \to \infty$. In this case we write*

$$\lim_{k \to \infty} x_k = \bar{x}.$$

This notion allows us to define the following important topological concepts for sets.

Definition 1.8 *Let Ω be a nonempty subset of $\mathbb{R}^n$. Then:*

(i) *The* CLOSURE *of Ω, denoted by $\overline{\Omega}$ or $\text{cl}\,\Omega$, is the collection of limits of all convergent sequences belonging to Ω.*
(ii) *The* BOUNDARY *of Ω, denoted by $\text{bd}\,\Omega$, is the set $\overline{\Omega} \setminus \text{int}\,\Omega$.*

We can see that the closure of Ω is the intersection of all closed sets containing Ω and that the interior of Ω is the union of all open sets contained in Ω. It follows from the definition that $\bar{x} \in \overline{\Omega}$ if and only if for any $\delta > 0$ we have $I\!B(\bar{x};\delta) \cap \Omega \neq \emptyset$. Furthermore, $\bar{x} \in \text{bd}\,\Omega$ if and only if for any $\delta > 0$ the closed ball $I\!B(\bar{x};\Omega)$ intersects both sets Ω and its complement Ω^c.

Definition 1.9 *Let $\{x_k\}$ be a sequence in $\mathbb{R}^n$ and let $\{k_\ell\}$ be a strictly increasing sequence of positive integers. Then the new sequence $\{x_{k_\ell}\}$ is called a* SUBSEQUENCE *of $\{x_k\}$.*

We say that a set Ω is *bounded* if it is contained in a ball centered at the origin with some radius $r > 0$, i.e., $\Omega \subset I\!B(0;r)$. Thus a sequence $\{x_k\}$ is bounded if there is $r > 0$ with

$$\|x_k\| \le r \quad \text{for all} \quad k \in \mathbb{N}\,.$$

The following important result is known as the *Bolzano-Weierstrass theorem*.

Theorem 1.10 Any bounded sequence in $\mathbb{R}^n$ contains a convergent subsequence.

The next concept plays a very significant role in analysis and optimization.

Definition 1.11 *We say that a set Ω is* COMPACT *in $\mathbb{R}^n$ if every sequence in Ω contains a subsequence converging to some point in Ω.*

The following result is a consequence of the Bolzano-Weierstrass theorem.

Theorem 1.12 A subset Ω of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

For subsets Ω, Ω_1, and Ω_2 of $\mathbb{R}^n$ and for $\lambda \in \mathbb{R}$, we define the operations:

$$\Omega_1 + \Omega_2 := \big\{x + y \;\big|\; x \in \Omega_1,\; y \in \Omega_2\big\}, \quad \lambda\Omega := \big\{\lambda x \;\big|\; x \in \Omega\big\}.$$

The next proposition can be proved easily.

Proposition 1.13 *Let Ω_1 and Ω_2 be two subsets of $\mathbb{R}^n$.*

(i) *If Ω_1 is open or Ω_2 is open, then $\Omega_1 + \Omega_2$ is open.*
(ii) *If Ω_1 is closed and Ω_2 is compact, then $\Omega_2 + \Omega_2$ is closed.*

Recall now the notions of *bounds* for subsets of the real line.

Definition 1.14 *Let D be a subset of the real line. A number $m \in \mathbb{R}$ is a* LOWER BOUND *of D if we have*

$$x \geq m \quad \text{for all} \quad x \in D.$$

If the set D has a lower bound, then it is BOUNDED BELOW. *Similarly, a number $M \in \mathbb{R}$ is an* UPPER BOUND *of D if*

$$x \leq M \quad \text{for all} \quad x \in D,$$

and D is BOUNDED ABOVE *if it has an upper bound. Furthermore, we say that the set D is* BOUNDED *if it is simultaneously bounded below and above.*

Now we are ready to define the concepts of *infimum* and *supremum* of sets.

Definition 1.15 *Let $D \subset \mathbb{R}$ be nonempty and bounded below. The infimum of D, denoted by* $\inf D$, *is the greatest lower bound of D. When D is nonempty and bounded above, its supremum, denoted by* $\sup D$, *is the least upper bound of D. If D is not bounded below (resp. above), then we set* $\inf D := -\infty$ (*resp.* $\sup D := \infty$). *We also use the convention that* $\inf \emptyset := \infty$ *and* $\sup \emptyset := -\infty$.

The following fundamental axiom ensures that these notions are well-defined.

Completeness Axiom. *For every nonempty subset D of $\mathbb{R}$ that is bounded above, the least upper bound of D exists as a real number.*

Using the Completeness Axiom, it is easy to see that if a nonempty set is bounded below, then its greatest lower bound exists as a real number.

Throughout the book we consider for convenience *extended-real-valued* functions, which take values in $\overline{\mathbb{R}} := (-\infty, \infty]$. The usual conventions of extended arithmetics are that $a + \infty = \infty$ for any $a \in \mathbb{R}$, $\infty + \infty = \infty$, and $t \cdot \infty = \infty$ for $t > 0$.

Definition 1.16 *Let $f : \Omega \to \overline{\mathbb{R}}$ be an extended-real-valued function and let $\bar{x} \in \Omega$ with $f(\bar{x}) < \infty$. Then f is* CONTINUOUS *at $\bar{x}$ if for any $\epsilon > 0$ there is $\delta > 0$ such that*

$$|f(x) - f(\bar{x})| < \epsilon \quad \text{whenever} \quad \|x - \bar{x}\| < \delta, \; x \in \Omega.$$

We say that f is continuous on Ω if it is continuous at every point of Ω.

It is obvious from the definition that if $f : \Omega \to \overline{\mathbb{R}}$ is continuous at $\bar{x}$ (with $f(\bar{x}) < \infty$), then it is finite on the intersection of Ω and a ball centered at $\bar{x}$ with some radius $r > 0$. Furthermore, $f : \Omega \to \overline{\mathbb{R}}$ is continuous at $\bar{x}$ (with $f(\bar{x}) < \infty$) if and only if for every sequence $\{x_k\}$ in Ω converging to $\bar{x}$ the sequence $\{f(x_k)\}$ converges to $f(\bar{x})$.

Definition 1.17 *Let $f : \Omega \to \overline{\mathbb{R}}$ and let $\bar{x} \in \Omega$ with $f(\bar{x}) < \infty$. We say that f has a* LOCAL MINIMUM *at $\bar{x}$ relative to Ω if there is $\delta > 0$ such that*

$$f(x) \geq f(\bar{x}) \text{ for all } x \in I\!B(\bar{x}; \delta) \cap \Omega.$$

We also say that f has a GLOBAL/ABSOLUTE MINIMUM *at $\bar{x}$ relative to Ω if*

$$f(x) \geq f(\bar{x}) \text{ for all } x \in \Omega.$$

The notions of *local and global maxima* can be defined similarly.

Finally in this section, we formulate a fundamental result of mathematical analysis and optimization known as the *Weierstrass existence theorem*.

Theorem 1.18 Let $f : \Omega \to \mathbb{R}$ be a continuous function, where Ω is a nonempty, compact subset of $\mathbb{R}^n$. Then there exist $\bar{x} \in \Omega$ and $\bar{u} \in \Omega$ such that

$$f(\bar{x}) = \inf\{f(x) \mid x \in \Omega\} \text{ and } f(\bar{u}) = \sup\{f(x) \mid x \in \Omega\}.$$

In Section 4.1 we present some "unilateral" versions of Theorem 1.18.

1.2 CONVEX SETS

We start the study of convexity with sets and then proceed to functions. Geometric ideas play an underlying role in convex analysis, its extensions, and applications. Thus we implement the geometric approach in this book.

Given two elements a and b in $\mathbb{R}^n$, define the *interval/line segment*

$$[a, b] := \{\lambda a + (1 - \lambda)b \mid \lambda \in [0, 1]\}.$$

Note that if $a = b$, then this interval reduces to a singleton $[a, b] = \{a\}$.

Definition 1.19 *A subset Ω of $\mathbb{R}^n$ is* CONVEX *if $[a, b] \subset \Omega$ whenever $a, b \in \Omega$. Equivalently, Ω is convex if $\lambda a + (1 - \lambda)b \in \Omega$ for all $a, b \in \Omega$ and $\lambda \in [0, 1]$.*

Given $\omega_1, \ldots, \omega_m \in \mathbb{R}^n$, the element $x = \sum_{i=1}^m \lambda_i \omega_i$, where $\sum_{i=1}^m \lambda_i = 1$ and $\lambda_i \geq 0$ for some $m \in \mathbb{N}$, is called a *convex combination* of $\omega_1, \ldots, \omega_m$.

Proposition 1.20 *A subset Ω of $\mathbb{R}^n$ is convex if and only if it contains all convex combinations of its elements.*

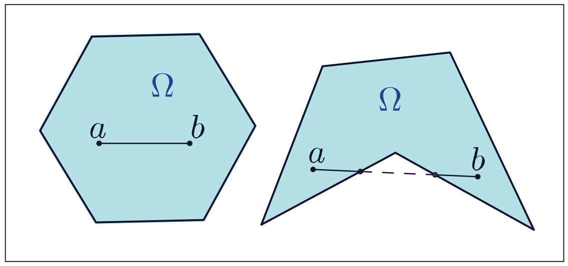


Figure 1.1: Convex set and nonconvex set.

Proof. The sufficient condition is trivial. To justify the necessity, we show by induction that any convex combination $x = \sum_{i=1}^{m} \lambda_i \omega_i$ of elements in Ω is an element of Ω. This conclusion follows directly from the definition for $m = 1, 2$. Fix now a positive integer $m \geq 2$ and suppose that every convex combination of $k \in \mathbb{N}$ elements from Ω, where $k \leq m$, belongs to Ω. Form the convex combination

$$y := \sum_{i=1}^{m+1} \lambda_i \omega_i, \ \sum_{i=1}^{m+1} \lambda_i = 1, \ \lambda_i \geq 0$$

and observe that if $\lambda_{m+1} = 1$, then $\lambda_1 = \lambda_2 = \ldots = \lambda_m = 0$, so $y = \omega_{m+1} \in \Omega$. In the case where $\lambda_{m+1} < 1$ we get the representations

$$\sum_{i=1}^{m} \lambda_i = 1 - \lambda_{m+1} \ \text{ and } \ \sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} = 1,$$

which imply in turn the inclusion

$$z := \sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} \omega_i \in \Omega.$$

It yields therefore the relationships

$$y = (1 - \lambda_{m+1}) \sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} \omega_i + \lambda_{m+1} \omega_{m+1} = (1 - \lambda_{m+1}) z + \lambda_{m+1} \omega_{m+1} \in \Omega$$

and thus completes the proof of the proposition. □

Proposition 1.21 *Let Ω_1 be a convex subset of $\mathbb{R}^n$ and let Ω_2 be a convex subset of $\mathbb{R}^p$. Then the Cartesian product $\Omega_1 \times \Omega_2$ is a convex subset of $\mathbb{R}^n \times \mathbb{R}^p$.*

Proof. Fix $a = (a_1, a_2), b = (b_1, b_2) \in \Omega_1 \times \Omega_2$, and $\lambda \in (0,1)$. Then we have $a_1, b_1 \in \Omega_1$ and $a_2, b_2 \in \Omega_2$. The convexity of Ω_1 and Ω_2 gives us

$$\lambda a_i + (1-\lambda)b_i \in \Omega_i \ \text{ for } \ i = 1, 2,$$

which implies therefore that

$$\lambda a + (1-\lambda)b = \big(\lambda a_1 + (1-\lambda)b_1, \lambda a_2 + (1-\lambda)b_2\big) \in \Omega_1 \times \Omega_2.$$

Thus the Cartesian product $\Omega_1 \times \Omega_2$ is convex. □

Let us continue with the definition of affine mappings.

Definition 1.22 *A mapping $B : \mathbb{R}^n \to \mathbb{R}^p$ is* AFFINE *if there exist a linear mapping $A : \mathbb{R}^n \to \mathbb{R}^p$ and an element $b \in \mathbb{R}^p$ such that $B(x) = A(x) + b$ for all $x \in \mathbb{R}^n$.*

Every linear mapping is affine. Moreover, $B : \mathbb{R}^n \to \mathbb{R}^p$ is affine if and only if

$$B(\lambda x + (1-\lambda)y) = \lambda B(x) + (1-\lambda)B(y) \text{ for all } x, y \in \mathbb{R}^n \text{ and } \lambda \in \mathbb{R}\,.$$

Now we show that set convexity is preserved under *affine operations*.

Proposition 1.23 *Let $B : \mathbb{R}^n \to \mathbb{R}^p$ be an affine mapping. Suppose that Ω is a convex subset of $\mathbb{R}^n$ and Θ is a convex subset of $\mathbb{R}^p$. Then $B(\Omega)$ is a convex subset of $\mathbb{R}^p$ and $B^{-1}(\Theta)$ is a convex subset of $\mathbb{R}^n$.*

Proof. Fix any $a, b \in B(\Omega)$ and $\lambda \in (0,1)$. Then $a = B(x)$ and $b = B(y)$ for $x, y \in \Omega$. Since Ω is convex, we have $\lambda x + (1-\lambda)y \in \Omega$. Then

$$\lambda a + (1-\lambda)b = \lambda B(x) + (1-\lambda)B(y) = B(\lambda x + (1-\lambda)y) \in B(\Omega),$$

which justifies the convexity of the image $B(\Omega)$.

Taking now any $x, y \in B^{-1}(\Theta)$ and $\lambda \in (0,1)$, we get $B(x), B(y) \in \Theta$. This gives us

$$\lambda B(x) + (1-\lambda)B(y) = B\big(\lambda x + (1-\lambda)y\big) \in \Theta$$

by the convexity of Θ. Thus we have $\lambda x + (1-\lambda)y \in B^{-1}(\Theta)$, which verifies the convexity of the inverse image $B^{-1}(\Theta)$. □

Proposition 1.24 *Let $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ be convex and let $\lambda \in \mathbb{R}$. Then both sets $\Omega_1 + \Omega_2$ and $\lambda\Omega_1$ are also convex in $\mathbb{R}^n$.*

Proof. It follows directly from the definitions. □

Next we proceed with *intersections* of convex sets.

Proposition 1.25 *Let $\{\Omega_\alpha\}_{\alpha \in I}$ be a collection of convex subsets of $\mathbb{R}^n$. Then $\bigcap_{\alpha \in I} \Omega_\alpha$ is also a convex subset of $\mathbb{R}^n$.*

Proof. Taking any $a, b \in \bigcap_{\alpha \in I} \Omega_\alpha$, we get that $a, b \in \Omega_\alpha$ for all $\alpha \in I$. The convexity of each Ω_α ensures that $\lambda a + (1-\lambda)b \in \Omega_\alpha$ for any $\lambda \in (0,1)$. Thus $\lambda a + (1-\lambda)b \in \bigcap_{\alpha \in I} \Omega_\alpha$ and the intersection $\bigcap_{\alpha \in I} \Omega_\alpha$ is convex. □

Definition 1.26 *Let Ω be a subset of $\mathbb{R}^n$. The* CONVEX HULL *of Ω is defined by*

$$\operatorname{co}\Omega := \bigcap \Big\{ C \;\Big|\; C \text{ is convex and } \Omega \subset C \Big\}.$$

The next proposition follows directly from the definition and Proposition 1.25.

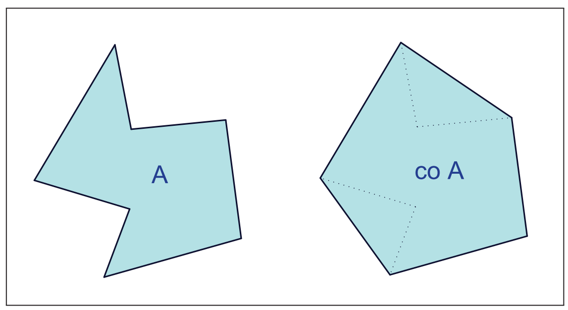


Figure 1.2: Nonconvex set and its convex hull.

Proposition 1.27 *The convex hull* $\operatorname{co}\Omega$ *is the smallest convex set containing Ω.*

Proof. The convexity of the set $\operatorname{co}\Omega \supset \Omega$ follows from Proposition 1.25. On the other hand, for any convex set C that contains Ω we clearly have $\operatorname{co}\Omega \subset C$, which verifies the conclusion. □

Proposition 1.28 *For any subset Ω of $\mathbb{R}^n$, its convex hull admits the representation*

$$\operatorname{co}\Omega = \Big\{ \sum_{i=1}^{m} \lambda_i a_i \;\Big|\; \sum_{i=1}^{m} \lambda_i = 1,\ \lambda_i \geq 0,\ a_i \in \Omega,\ m \in \mathbb{N} \Big\}.$$

Proof. Denoting by C the right-hand side of the representation to prove, we obviously have $\Omega \subset C$. Let us check that the set C is convex. Take any $a, b \in C$ and get

$$a := \sum_{i=1}^{p} \alpha_i a_i, \quad b := \sum_{j=1}^{q} \beta_j b_j,$$

where $a_i, b_j \in \Omega, \alpha_i, \beta_j \geq 0$ with $\sum_{i=1}^{p} \alpha_i = \sum_{j=1}^{q} \beta_j = 1$, and $p, q \in \mathbb{N}$. It follows easily that for every number $\lambda \in (0, 1)$, we have

$$\lambda a + (1-\lambda)b = \sum_{i=1}^{p} \lambda \alpha_i a_i + \sum_{j=1}^{q} (1-\lambda)\beta_j b_j.$$

Then the resulting equality

$$\sum_{i=1}^{p} \lambda \alpha_i + \sum_{j=1}^{q} (1-\lambda)\beta_j = \lambda \sum_{i=1}^{p} \alpha_i + (1-\lambda) \sum_{j=1}^{q} \beta_j = 1$$

ensures that $\lambda a + (1-\lambda)b \in C$, and thus $\operatorname{co} \Omega \subset C$ by the definition of $\operatorname{co} \Omega$. Fix now any $a = \sum_{i=1}^{m} \lambda_i a_i \in C$ with $a_i \in \Omega \subset \operatorname{co} \Omega$ for $i = 1, \ldots, m$. Since the set $\operatorname{co} \Omega$ is convex, we conclude by Proposition 1.20 that $a \in \operatorname{co} \Omega$ and thus $\operatorname{co} \Omega = C$. □

Proposition 1.29 *The interior* $\operatorname{int} \Omega$ *and closure* $\overline{\Omega}$ *of a convex set* $\Omega \subset \mathbb{R}^n$ *are also convex.*

Proof. Fix $a, b \in \operatorname{int} \Omega$ and $\lambda \in (0, 1)$. Then find an open set V such that

$$a \in V \subset \Omega \text{ and so } \lambda a + (1-\lambda)b \in \lambda V + (1-\lambda)b \subset \Omega.$$

Since $\lambda V + (1-\lambda)b$ is open, we get $\lambda a + (1-\lambda)b \in \operatorname{int} \Omega$, and thus the set $\operatorname{int} \Omega$ is convex.

To verify the convexity of $\overline{\Omega}$, we fix $a, b \in \overline{\Omega}$ and $\lambda \in (0, 1)$ and then find sequences $\{a_k\}$ and $\{b_k\}$ in Ω converging to a and b, respectively. By the convexity of Ω, the sequence $\{\lambda a_k + (1-\lambda)b_k\}$ lies entirely in Ω and converges to $\lambda a + (1-\lambda)b$. This ensures the inclusion $\lambda a + (1-\lambda)b \in \overline{\Omega}$ and thus justifies the convexity of the closure $\overline{\Omega}$. □

To proceed further, for any $a, b \in \mathbb{R}^n$, define the interval

$$[a, b) := \big\{\lambda a + (1-\lambda)b \;\big|\; \lambda \in (0, 1]\big\}.$$

We can also define the intervals $(a, b]$ and (a, b) in a similar way.

Lemma 1.30 *For a convex set* $\Omega \subset \mathbb{R}^n$ *with nonempty interior, take any* $a \in \operatorname{int} \Omega$ *and* $b \in \overline{\Omega}$. *Then* $[a, b) \subset \operatorname{int} \Omega$.

Proof. Since $b \in \overline{\Omega}$, for any $\epsilon > 0$, we have $b \in \Omega + \epsilon I\!B$. Take now $\lambda \in (0, 1]$ and let $x_\lambda := \lambda a + (1-\lambda)b$. Choosing $\epsilon > 0$ such that $a + \epsilon \frac{2-\lambda}{\lambda} I\!B \subset \Omega$ gives us

$$\begin{aligned} x_\lambda + \epsilon I\!B &= \lambda a + (1-\lambda)b + \epsilon I\!B \\ &\subset \lambda a + (1-\lambda)\big[\Omega + \epsilon I\!B\big] + \epsilon I\!B \\ &= \lambda a + (1-\lambda)\Omega + (1-\lambda)\epsilon I\!B + \epsilon I\!B \\ &\subset \lambda\Big[a + \epsilon\frac{2-\lambda}{\lambda} I\!B\Big] + (1-\lambda)\Omega \\ &\subset \lambda\Omega + (1-\lambda)\Omega \subset \Omega. \end{aligned}$$

This shows that $x_\lambda \in \operatorname{int}\Omega$ and thus verifies the inclusion $[a, b) \subset \operatorname{int}\Omega$. □

Now we establish relationships between taking the interior and closure of convex sets.

Proposition 1.31 *Let $\Omega \subset \mathbb{R}^n$ be a convex set with nonempty interior. Then we have:*
(i) $\overline{\operatorname{int}\Omega} = \overline{\Omega}$ *and* **(ii)** $\operatorname{int}\Omega = \operatorname{int}\overline{\Omega}$.

Proof. **(i)** Obviously, $\overline{\operatorname{int}\Omega} \subset \overline{\Omega}$. For any $b \in \overline{\Omega}$ and $a \in \operatorname{int}\Omega$, define the sequence $\{x_k\}$ by

$$x_k := \frac{1}{k}a + \Big(1 - \frac{1}{k}\Big)b,\ k \in N.$$

Lemma 1.30 ensures that $x_k \in \operatorname{int}\Omega$. Since $x_k \to b$, we have $b \in \overline{\operatorname{int}\Omega}$ and thus verify **(i)**.
(ii) Since the inclusion $\operatorname{int}\Omega \subset \operatorname{int}\overline{\Omega}$ is obvious, it remains to prove the opposite inclusion $\operatorname{int}\overline{\Omega} \subset \operatorname{int}\Omega$. To proceed, fix any $b \in \operatorname{int}\overline{\Omega}$ and $a \in \operatorname{int}\Omega$. If $\epsilon > 0$ is sufficiently small, then

$$c := b + \epsilon(b-a) \in \overline{\Omega}, \text{ and hence } b = \frac{\epsilon}{1+\epsilon}a + \frac{1}{1+\epsilon}c \in (a, c) \subset \operatorname{int}\Omega,$$

which verifies that $\operatorname{int}\overline{\Omega} \subset \operatorname{int}\Omega$ and thus completes the proof. □

1.3 CONVEX FUNCTIONS

This section collects basic facts about general (extended-real-valued) *convex functions* including their analytic and geometric characterizations, important properties as well as their specifications for particular subclasses. We also define *convex set-valued mappings* and use them to study a remarkable class of *optimal value functions* employed in what follows.

Definition 1.32 *Let $f : \Omega \to \overline{\mathbb{R}}$ be an extended-real-valued function define on a convex set $\Omega \subset \mathbb{R}^n$. Then the function f is* CONVEX *on Ω if*

$$f\big(\lambda x + (1-\lambda)y\big) \le \lambda f(x) + (1-\lambda)f(y) \ \text{ for all } x, y \in \Omega \text{ and } \lambda \in (0,1). \tag{1.1}$$

If the inequality in (1.1) *is strict for* $x \neq y$*, then* f *is* STRICTLY CONVEX *on* Ω.

Given a function $f : \Omega \to \overline{\mathbb{R}}$, the *extension* of f to $\mathbb{R}^n$ is defined by

$$\widetilde{f}(x) := \begin{cases} f(x) & \text{if } x \in \Omega, \\ \infty & \text{otherwise.} \end{cases}$$

Obviously, if f is convex on Ω, where Ω is a convex set, then $\widetilde{f}$ is convex on $\mathbb{R}^n$. Furthermore, if $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a convex function, then it is also convex on every convex subset of $\mathbb{R}^n$. This allows to consider without loss of generality extended-real-valued convex functions on the whole space $\mathbb{R}^n$.

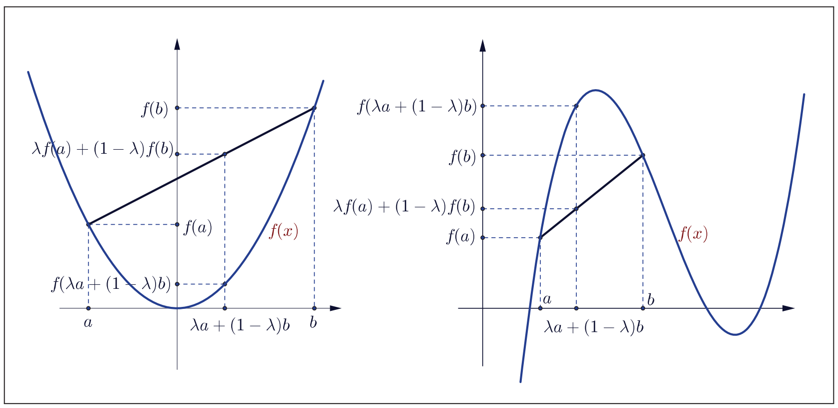


Figure 1.3: Convex function and nonconvex function.

Definition 1.33 *The* DOMAIN *and* EPIGRAPH *of* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *are defined, respectively, by*

$$\text{dom } f := \{x \in \mathbb{R}^n \mid f(x) < \infty\} \quad \textit{and}$$

$$\text{epi } f := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid x \in \mathbb{R}^n,\ t \geq f(x)\} = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid x \in \text{dom } f,\ t \geq f(x)\}.$$

Let us illustrate the convexity of functions by examples.

Example 1.34 The following functions are convex:

(i) $f(x) := \langle a, x\rangle + b$ for $x \in \mathbb{R}^n$, where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$.
(ii) $g(x) := \|x\|$ for $x \in \mathbb{R}^n$.
(iii) $h(x) := x^2$ for $x \in \mathbb{R}$.

Indeed, the function f in **(i)** is convex since

$$\begin{aligned} f\big(\lambda x + (1-\lambda)y\big) &= \langle a, \lambda x + (1-\lambda)y\rangle + b = \lambda\langle a, x\rangle + (1-\lambda)\langle a, y\rangle + b \\ &= \lambda(\langle a, x\rangle + b) + (1-\lambda)(\langle a, y\rangle + b) \\ &= \lambda f(x) + (1-\lambda)f(y) \text{ for all } x, y \in \mathbb{R}^n \text{ and } \lambda \in (0,1). \end{aligned}$$

The function g in **(ii)** is convex since for $x, y \in \mathbb{R}^n$ and $\lambda \in (0,1)$, we have

$$g\big(\lambda x + (1-\lambda)y\big) = \|\lambda x + (1-\lambda)y\| \le \lambda\|x\| + (1-\lambda)\|y\| = \lambda g(x) + (1-\lambda)g(y),$$

which follows from the triangle inequality and the fact that $\|\alpha u\| = |\alpha| \cdot \|u\|$ whenever $\alpha \in \mathbb{R}$ and $u \in \mathbb{R}^n$. The convexity of the simplest quadratic function h in **(iii)** follows from a more general result for the quadratic function on $\mathbb{R}^n$ given in the next example.

Example 1.35 Let A be an $n \times n$ symmetric matrix. It is called *positive semidefinite* if $\langle Au, u\rangle \ge 0$ for all $u \in \mathbb{R}^n$. Let us check that A is positive semidefinite if and only if the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f(x) := \frac{1}{2}\langle Ax, x\rangle, \quad x \in \mathbb{R}^n,$$

is convex. Indeed, a direct calculation shows that for any $x, y \in \mathbb{R}^n$ and $\lambda \in (0,1)$ we have

$$\lambda f(x) + (1-\lambda)f(y) - f\big(\lambda x + (1-\lambda)y\big) = \frac{1}{2}\lambda(1-\lambda)\langle A(x-y), x-y\rangle. \tag{1.2}$$

If the matrix A is positive semidefinite, then $\langle A(x-y), x-y\rangle \ge 0$, so the function f is convex by (1.2). Conversely, assuming the convexity of f and using equality (1.2) for $x = u$ and $y = 0$ verify that A is positive semidefinite.

The following characterization of convexity is known as the *Jensen inequality*.

Theorem 1.36 A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if and only if for any numbers $\lambda_i \ge 0$ as $i = 1, \ldots, m$ with $\sum_{i=1}^m \lambda_i = 1$ and for any elements $x_i \in \mathbb{R}^n$, $i = 1, \ldots, m$, it holds that

$$f\Big(\sum_{i=1}^m \lambda_i x_i\Big) \le \sum_{i=1}^m \lambda_i f(x_i). \tag{1.3}$$

Proof. Since (1.3) immediately implies the convexity of f, we only need to prove that any convex function f satisfies the Jensen inequality (1.3). Arguing by induction and taking into account that for $m = 1$ inequality (1.3) holds trivially and for $m = 2$ inequality (1.3) holds by the definition of convexity, we suppose that it holds for an integer $m := k$ with $k \ge 2$. Fix numbers $\lambda_i \ge 0$, $i = 1, \ldots, k+1$, with $\sum_{i=1}^{k+1} \lambda_i = 1$ and elements $x_i \in \mathbb{R}^n$, $i = 1, \ldots, k+1$. We obviously have the relationship

$$\sum_{i=1}^k \lambda_i = 1 - \lambda_{k+1}.$$

If $\lambda_{k+1} = 1$, then $\lambda_i = 0$ for all $i = 1, \ldots, k$ and (1.3) obviously holds for $m := k + 1$ in this case. Supposing now that $0 \le \lambda_{k+1} < 1$, we get

$$\sum_{i=1}^{k} \frac{\lambda_i}{1 - \lambda_{k+1}} = 1$$

and by direct calculations based on convexity arrive at

$$\begin{aligned} f\Big(\sum_{i=1}^{k+1} \lambda_i x_i\Big) &= f\Big[(1 - \lambda_{k+1}) \frac{\sum_{i=1}^{k} \lambda_i x_i}{1 - \lambda_{k+1}} + \lambda_{k+1} x_{k+1}\Big] \\ &\le (1 - \lambda_{k+1}) f\Big(\frac{\sum_{i=1}^{k} \lambda_i x_i}{1 - \lambda_{k+1}}\Big) + \lambda_{k+1} f(x_{k+1}) \\ &= (1 - \lambda_{k+1}) f\Big(\sum_{i=1}^{k} \frac{\lambda_i}{1 - \lambda_{k+1}} x_i\Big) + \lambda_{k+1} f(x_{k+1}) \\ &\le (1 - \lambda_{k+1}) \sum_{i=1}^{k} \frac{\lambda_i}{1 - \lambda_{k+1}} f(x_i) + \lambda_{k+1} f(x_{k+1}) \\ &= \sum_{i=1}^{k+1} \lambda_i f(x_i). \end{aligned}$$

This justifies inequality (1.3) and completes the proof of the theorem. □

The next theorem gives a *geometric characterization* of the function convexity via the convexity of the associated epigraphical set.

Theorem 1.37 A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if and only if its epigraph epi f is a convex subset of the product space $\mathbb{R}^n \times \mathbb{R}$.

Proof. Assuming that f is convex, fix pairs $(x_1, t_1), (x_2, t_2) \in \text{epi } f$ and a number $\lambda \in (0, 1)$. Then we have $f(x_i) \le t_i$ for $i = 1, 2$. Thus the convexity of f ensures that

$$f\big(\lambda x_1 + (1 - \lambda)x_2\big) \le \lambda f(x_1) + (1 - \lambda) f(x_2) \le \lambda t_1 + (1 - \lambda)t_2.$$

This implies therefore that

$$\lambda(x_1, t_1) + (1 - \lambda)(x_2, t_2) = (\lambda x_1 + (1 - \lambda)x_2, \lambda t_1 + (1 - \lambda)t_2) \in \text{epi } f,$$

and thus the epigraph epi f is a convex subset of $\mathbb{R}^n \times \mathbb{R}$.

Conversely, suppose that the set epi f is convex and fix $x_1, x_2 \in \text{dom } f$ and a number $\lambda \in (0, 1)$. Then $(x_1, f(x_1)), (x_2, f(x_2)) \in \text{epi } f$. This tells us that

$$\big(\lambda x_1 + (1 - \lambda)x_2, \lambda f(x_1) + (1 - \lambda) f(x_2)\big) = \lambda\big(x_1, f(x_1)\big) + (1 - \lambda)\big(x_2, f(x_2)\big) \in \text{epi } f$$

and thus implies the inequality

$$f\big(\lambda x_1 + (1-\lambda)x_2\big) \le \lambda f(x_1) + (1-\lambda) f(x_2),$$

which justifies the convexity of the function f. □

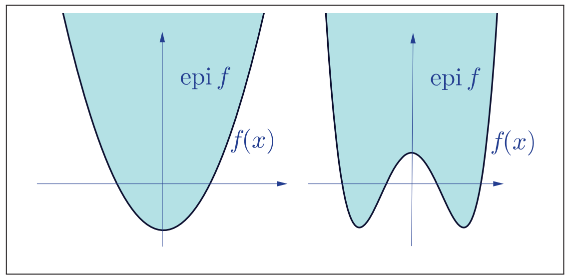


Figure 1.4: Epigraphs of convex function and nonconvex function.

Now we show that convexity is preserved under some important operations.

Proposition 1.38 *Let $f_i\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex functions for all $i = 1, \ldots, m$. Then the following functions are convex as well:*

(i) *The multiplication by scalars λf for any $\lambda > 0$.*
(ii) *The sum function $\sum_{i=1}^{m} f_i$.*
(iii) *The maximum function $\max_{1 \le i \le m} f_i$.*

Proof. The convexity of λf in **(i)** follows directly from the definition. It is sufficient to prove **(ii)** and **(iii)** for $m = 2$, since the general cases immediately follow by induction.

(ii) Fix any $x, y \in \mathbb{R}^n$ and $\lambda \in (0,1)$. Then we have

$$\begin{aligned}\big(f_1 + f_2\big)\big(\lambda x + (1-\lambda)y\big) &= f_1\big(\lambda x + (1-\lambda)y\big) + f_2\big(\lambda x + (1-\lambda)y\big)\\ &\le \lambda f_1(x) + (1-\lambda) f_1(y) + \lambda f_2(x) + (1-\lambda) f_2(y)\\ &= \lambda(f_1 + f_2)(x) + (1-\lambda)(f_1 + f_2)(y),\end{aligned}$$

which verifies the convexity of the sum function $f_1 + f_2$.

(iii) Denote $g := \max\{f_1, f_2\}$ and get for any $x, y \in \mathbb{R}^n$ and $\lambda \in (0,1)$ that

$$f_i\big(\lambda x + (1-\lambda)y\big) \le \lambda f_i(x) + (1-\lambda) f_i(y) \le \lambda g(x) + (1-\lambda) g(y)$$

for $i = 1, 2$. This directly implies that

$$g\big(\lambda x + (1-\lambda)y\big) = \max\big\{f_1\big(\lambda x + (1-\lambda)y\big), f_2\big(\lambda x + (1-\lambda)y\big)\big\} \le \lambda g(x) + (1-\lambda) g(y),$$

which shows that the maximum function $g(x) = \max\{f_1(x), f_2(x)\}$ is convex. □

The next result concerns the preservation of convexity under function compositions.

Proposition 1.39 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and let $\varphi : \mathbb{R} \to \overline{\mathbb{R}}$ be nondecreasing and convex on a convex set containing the range of the function f. Then the composition $\varphi \circ f$ is convex.*

Proof. Take any $x_1, x_2 \in \mathbb{R}^n$ and $\lambda \in (0, 1)$. Then we have by the convexity of f that

$$f\big(\lambda x_1 + (1-\lambda)x_2\big) \le \lambda f(x_1) + (1-\lambda) f(x_2).$$

Since φ is nondecreasing and it is also convex, it follows that

$$\begin{aligned}(\varphi \circ f)\big(\lambda x_1 + (1-\lambda)x_2\big) &= \varphi\big(f\big(\lambda x_1 + (1-\lambda)x_2\big)\big)\\ &\le \varphi\big(\lambda f(x_1) + (1-\lambda) f(x_2)\big)\\ &\le \lambda\varphi\big(f(x_1)\big) + (1-\lambda)\varphi\big(f(x_2)\big)\\ &= \lambda(\varphi \circ f)(x_1) + (1-\lambda)(\varphi \circ f)(x_2),\end{aligned}$$

which verifies the convexity of the composition $\varphi \circ f$. □

Now we consider the composition of a convex function and an affine mapping.

Proposition 1.40 *Let $B : \mathbb{R}^n \to \mathbb{R}^p$ be an affine mapping and let $f : \mathbb{R}^p \to \overline{\mathbb{R}}$ be a convex function. Then the composition $f \circ B$ is convex.*

Proof. Taking any $x, y \in \mathbb{R}^n$ and $\lambda \in (0, 1)$, we have

$$\begin{aligned}(f \circ B)\big(\lambda x + (1-\lambda)y\big) &= f(B(\lambda x + (1-\lambda)y)) = f\big(\lambda B(x) + (1-\lambda)B(y)\big)\\ &\le \lambda f\big(B(x)\big) + (1-\lambda) f\big(B(y)\big) = \lambda(f \circ B)(x) + (1-\lambda)(f \circ B)(y)\end{aligned}$$

and thus justify the convexity of the composition $f \circ B$. □

The following simple consequence of Proposition 1.40 is useful in applications.

Corollary 1.41 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function. For any $\bar{x}, d \in \mathbb{R}^n$, the function $\varphi : \mathbb{R} \to \overline{\mathbb{R}}$ defined by $\varphi(t) := f(\bar{x} + td)$ is convex as well. Conversely, if for every $\bar{x}, d \in \mathbb{R}^n$ the function φ defined above is convex, then f is also convex.*

Proof. Since $B(t) = \bar{x} + td$ is an affine mapping, the convexity of φ immediately follows from Proposition 1.40. To prove the converse implication, take any $x_1, x_2 \in \mathbb{R}^n$, $\lambda \in (0, 1)$ and let $\bar{x} := x_2$, $d := x_1 - x_2$. Since $\varphi(t) = f(\bar{x} + td)$ is convex, we have

$$\begin{aligned}f\big(\lambda x_1 + (1-\lambda)x_2\big) &= f\big(x_2 + \lambda(x_1 - x_2)\big) = \varphi(\lambda) = \varphi\big(\lambda(1) + (1-\lambda)(0)\big)\\ &\le \lambda\varphi(1) + (1-\lambda)\varphi(0) = \lambda f(x_1) + (1-\lambda) f(x_2),\end{aligned}$$

which verifies the convexity of the function f. □

The next proposition is trivial while useful in what follows.

Proposition 1.42 *Let* $f : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$ *be convex. For* $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^p$*, the functions* $\varphi(y) := f(\bar{x}, y)$ *and* $\psi(x) := f(x, \bar{y})$ *are also convex.*

Now we present an important extension of Proposition 1.38(iii).

Proposition 1.43 *Let* $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ *for* $i \in I$ *be a collection of convex functions with a nonempty index set* I*. Then the supremum function* $f(x) := \sup_{i \in I} f_i(x)$ *is convex.*

Proof. Fix $x_1, x_2 \in \mathbb{R}^n$ and $\lambda \in (0, 1)$. For every $i \in I$, we have

$$f_i\big(\lambda x_1 + (1-\lambda)x_2\big) \le \lambda f_i(x_1) + (1-\lambda) f_i(x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2),$$

which implies in turn that

$$f\big(\lambda x_1 + (1-\lambda)x_2\big) = \sup_{i \in I} f_i\big(\lambda x_1 + (1-\lambda)x_2\big) \le \lambda f(x_1) + (1-\lambda) f(x_2).$$

This justifies the convexity of the supremum function. □

Our next intention is to characterize convexity of *smooth* functions of one variable. To proceed, we begin with the following lemma.

Lemma 1.44 *Given a convex function* $f : \mathbb{R} \to \overline{\mathbb{R}}$*, assume that its domain is an open interval* I*. For any* $a, b \in I$ *and* $a < x < b$*, we have the inequalities*

$$\frac{f(x) - f(a)}{x - a} \le \frac{f(b) - f(a)}{b - a} \le \frac{f(b) - f(x)}{b - x}.$$

Proof. Fix a, b, x as above and form the numbers $t := \dfrac{x - a}{b - a} \in (0, 1)$. Then

$$f(x) = f\big(a + (x - a)\big) = f\Big(a + \frac{x - a}{b - a}(b - a)\Big) = f\big(a + t(b - a)\big) = f\big(tb + (1 - t)a\big).$$

This gives us the inequalities $f(x) \le t f(b) + (1 - t) f(a)$ and

$$f(x) - f(a) \le t f(b) + (1 - t) f(a) - f(a) = t\big[f(b) - f(a)\big] = \frac{x - a}{b - a}\Big(f(b) - f(a)\Big),$$

which can be equivalently written as

$$\frac{f(x) - f(a)}{x - a} \le \frac{f(b) - f(a)}{b - a}.$$

Similarly, we have the estimate

$$f(x) - f(b) \le tf(b) + (1-t)f(a) - f(b) = (t-1)\big[f(b) - f(a)\big] = \frac{x-b}{b-a}\Big(f(b) - f(a)\Big),$$

which finally implies that

$$\frac{f(b)-f(a)}{b-a} \le \frac{f(b)-f(x)}{b-x}$$

and thus completes the proof of the lemma. □

Theorem 1.45 Suppose that $f : \mathbb{R} \to \overline{\mathbb{R}}$ is differentiable on its domain, which is an open interval I. Then f is convex if and only if the derivative f' is nondecreasing on I.

Proof. Suppose that f is convex and fix $a < b$ with $a, b \in I$. By Lemma 1.44, we have

$$\frac{f(x)-f(a)}{x-a} \le \frac{f(b)-f(a)}{b-a}$$

for any $x \in (a,b)$. This implies by the derivative definition that

$$f'(a) \le \frac{f(b)-f(a)}{b-a}.$$

Similarly, we arrive at the estimate

$$\frac{f(b)-f(a)}{b-a} \le f'(b)$$

and conclude that $f'(a) \le f'(b)$, i.e., f' is a nondecreasing function.

To prove the converse, suppose that f' is nondecreasing on I and fix $x_1 < x_2$ with $x_1, x_2 \in I$ and $t \in (0,1)$. Then

$$x_1 < x_t < x_2 \ \text{ for } \ x_t := tx_1 + (1-t)x_2.$$

By the mean value theorem, we find c_1, c_2 such that $x_1 < c_1 < x_t < c_2 < x_2$ and

$$\begin{aligned} f(x_t) - f(x_1) &= f'(c_1)(x_t - x_1) = f'(c_1)(1-t)(x_2 - x_1), \\ f(x_t) - f(x_2) &= f'(c_2)(x_t - x_2) = f'(c_2)t(x_1 - x_2). \end{aligned}$$

This can be equivalently rewritten as

$$\begin{aligned} &tf(x_t) - tf(x_1) = f'(c_1)t(1-t)(x_2 - x_1), \\ &(1-t)f(x_t) - (1-t)f(x_2) = f'(c_2)t(1-t)(x_1 - x_2). \end{aligned}$$

Since $f'(c_1) \le f'(c_2)$, adding these equalities yields

$$f(x_t) \le tf(x_1) + (1-t)f(x_2),$$

which justifies the convexity of the function f. □

Corollary 1.46 *Let $f : \mathbb{R} \to \overline{\mathbb{R}}$ be twice differentiable on its domain, which is an open interval I. Then f is convex if and only if $f''(x) \geq 0$ for all $x \in I$.*

Proof. Since $f''(x) \geq 0$ for all $x \in I$ if and only if the derivative function f' is nondecreasing on this interval. Then the conclusion follows directly from Theorem 1.45. □

Example 1.47 Consider the function

$$f(x) := \begin{cases} \dfrac{1}{x} & \text{if } x > 0, \\ \infty & \text{otherwise.} \end{cases}$$

To verify its convexity, we get that $f''(x) = \dfrac{2}{x^3} > 0$ for all x belonging to the domain of f, which is $I = (0, \infty)$. Thus this function is convex on $\mathbb{R}$ by Corollary 1.46.

Next we define the notion of *set-valued mappings* (or *multifunctions*), which plays an important role in modern convex analysis, its extensions, and applications.

Definition 1.48 *We say that F is a* SET-VALUED MAPPING *between $\mathbb{R}^n$ and $\mathbb{R}^p$ and denote it by $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ if $F(x)$ is a subset of $\mathbb{R}^p$ for every $x \in \mathbb{R}^n$. The* DOMAIN *and* GRAPH *of F are defined, respectively, by*

$$\operatorname{dom} F := \big\{x \in \mathbb{R}^n \;\big|\; F(x) \neq \emptyset\big\} \;\; \textit{and} \;\; \operatorname{gph} F := \big\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^p \;\big|\; y \in F(x)\big\}.$$

Any single-valued mapping $F : \mathbb{R}^n \to \mathbb{R}^p$ is a particular set-valued mapping where the set $F(x)$ is a singleton for every $x \in \mathbb{R}^n$. It is essential in the following definition that the mapping F is set-valued.

Definition 1.49 *Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ and let $\varphi : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$. The* OPTIMAL VALUE *or* MARGINAL FUNCTION *associated with F and φ is defined by*

$$\mu(x) := \inf\big\{\varphi(x, y) \;\big|\; y \in F(x)\big\}, \;\; x \in \mathbb{R}^n. \tag{1.4}$$

Throughout this section we assume that $\mu(x) > -\infty$ for every $x \in \mathbb{R}^n$.

Proposition 1.50 *Assume that $\varphi : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$ is a convex function and that $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ is of convex graph. Then the optimal value function μ in (1.4) is convex.*

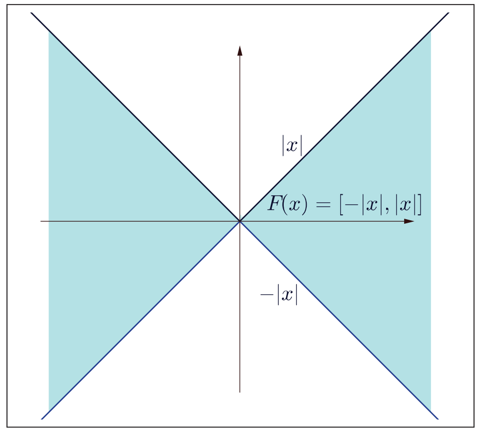


Figure 1.5: Graph of set-valued mapping.

Proof. Take $x_1, x_2 \in \operatorname{dom} \mu$, $\lambda \in (0, 1)$. For any $\epsilon > 0$, find $y_i \in F(x_i)$ such that

$$\varphi(x_i, y_i) < \mu(x_i) + \epsilon \quad \text{for } i = 1, 2.$$

It directly implies the inequalities

$$\lambda\varphi(x_1, y_1) < \lambda\mu(x_1) + \lambda\epsilon, \quad (1 - \lambda)\varphi(x_2, y_2) < (1 - \lambda)\mu(x_2) + (1 - \lambda)\epsilon.$$

Summing up these inequalities and employing the convexity of φ yield

$$\begin{aligned} \varphi\big(\lambda x_1 + (1 - \lambda)x_2, \lambda y_1 + (1 - \lambda)y_2\big) &\le \lambda\varphi(x_1, y_1) + (1 - \lambda)\varphi(x_2, y_2) \\ &< \lambda\mu(x_1) + (1 - \lambda)\mu(x_2) + \epsilon. \end{aligned}$$

Furthermore, the convexity of $\operatorname{gph} F$ gives us

$$\big(\lambda x_1 + (1 - \lambda)x_2, \lambda y_1 + (1 - \lambda)y_2\big) = \lambda(x_1, y_1) + (1 - \lambda)(x_2, y_2) \in \operatorname{gph} F,$$

and therefore $\lambda y_1 + (1 - \lambda)y_2 \in F(\lambda x_1 + (1 - \lambda)x_2)$. This implies that

$$\mu\big(\lambda x_1 + (1 - \lambda)x_2\big) \le \varphi\big(\lambda x_1 + (1 - \lambda)x_2, \lambda y_1 + (1 - \lambda)y_2\big) < \lambda\mu(x_1) + (1 - \lambda)\mu(x_2) + \epsilon.$$

Letting finally $\epsilon \to 0$ ensures the convexity of the optimal value function μ. □

Using Proposition 1.50, we can verify convexity in many situations. For instance, given two convex functions $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 1, 2$, let $\varphi(x, y) := f_1(x) + y$ and $F(x) := [f_2(x), \infty)$. Then the function φ is convex and set $\operatorname{gph} F = \operatorname{epi} f_2$ is convex as well, and hence we justify the convexity of the sum

$$\mu(x) = \inf_{y \in F(x)} \varphi(x, y) = f_1(x) + f_2(x).$$

Another example concerns compositions. Let $f : \mathbb{R}^p \to \overline{\mathbb{R}}$ be convex and let $B : \mathbb{R}^n \to \mathbb{R}^p$ be affine. Define $\varphi(x, y) := f(y)$ and $F(x) := \{B(x)\}$. Observe that φ is convex while F is of convex graph. Thus we have the convex composition

$$\mu(x) = \inf_{y \in F(x)} \varphi(x, y) = f\big(B(x)\big), \ x \in \mathbb{R}^n .$$

The examples presented above recover the results obtained previously by direct proofs. Now we establish via Proposition 1.50 the convexity of three new classes of functions.

Proposition 1.51 *Let $\varphi : \mathbb{R}^p \to \overline{\mathbb{R}}$ be convex and let $B : \mathbb{R}^p \to \mathbb{R}^n$ be affine. Consider the set-valued inverse image mapping $B^{-1} : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$, define*

$$f(x) := \inf\big\{\varphi(y) \bigm| y \in B^{-1}(x)\big\}, \quad x \in \mathbb{R}^n,$$

and suppose that $f(x) > -\infty$ for all $x \in \mathbb{R}^n$. Then f is a convex function.

Proof. Let $\varphi(x, y) \equiv \varphi(y)$ and $F(x) := B^{-1}(x)$. Then the set

$$\text{gph}\, F = \big\{(u, v) \in \mathbb{R}^n \times \mathbb{R}^p \bigm| B(v) = u\big\}$$

is obviously convex. Since we have the representation

$$f(x) = \inf_{y \in F(x)} \varphi(y), \quad x \in \mathbb{R}^n,$$

the convexity of f follows directly from Proposition 1.50. □

Proposition 1.52 *For convex functions $f_1, f_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$, define the* INFIMAL CONVOLUTION

$$(f_1 \oplus f_2)(x) := \inf\big\{f_1(x_1) + f_2(x_2) \bigm| x_1 + x_2 = x\big\}$$

and suppose that $(f_1 \oplus f_2)(x) > -\infty$ for all $x \in \mathbb{R}^n$. Then $f_1 \oplus f_2$ is also convex.

Proof. Define $\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ by $\varphi(x_1, x_2) := f_1(x_1) + f_2(x_2)$ and $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ by $B(x_1, x_2) := x_1 + x_2$. We have

$$\inf\big\{\varphi(x_1, x_2) \bigm| (x_1, x_2) \in B^{-1}(x)\big\} = (f_1 \oplus f_2)(x) \ \text{ for all } \ x \in \mathbb{R}^n,$$

which implies the convexity of $(f_1 \oplus f_2)$ by Proposition 1.51. □

Definition 1.53 *A function* $g : \mathbb{R}^p \to \overline{\mathbb{R}}$ *is called* NONDECREASING COMPONENTWISE *if*

$$\big[x_i \le y_i \ \text{ for all } \ i = 1, \ldots, p\big] \Longrightarrow \big[g(x_1, \ldots, x_p) \le g(y_1, \ldots, y_p)\big].$$

Now we are ready to present the final consequence of Proposition 1.50 in this section that involves the composition.

Proposition 1.54 *Define* $h : \mathbb{R}^n \to \mathbb{R}^p$ *by* $h(x) := (f_1(x), \ldots, f_p(x))$, *where* $f_i : \mathbb{R}^n \to \mathbb{R}$ *for* $i = 1, \ldots, p$ *are convex functions. Suppose that* $g : \mathbb{R}^p \to \overline{\mathbb{R}}$ *is convex and nondecreasing componentwise. Then the composition* $g \circ h : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is a convex function.*

Proof. Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ be a set-valued mapping defined by

$$F(x) := [f_1(x), \infty) \times [f_2(x), \infty) \times \ldots \times [f_p(x), \infty).$$

Then the graph of F is represented by

$$\text{gph}\, F = \big\{(x, t_1, \ldots, t_p) \in \mathbb{R}^n \times \mathbb{R}^p \ \big|\ t_i \ge f_i(x)\big\}.$$

Since all f_i are convex, the set $\text{gph}\, F$ is convex as well. Define further $\varphi : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$ by $\varphi(x, y) := g(y)$ and observe, since g is increasing componentwise, that

$$\inf\big\{\varphi(x, y) \ \big|\ y \in F(x)\big\} = g\big(f_1(x), \ldots, f_p(x)\big) = (g \circ h)(x),$$

which ensures the convexity of the composition $g \circ h$ by Proposition 1.50. □

1.4 RELATIVE INTERIORS OF CONVEX SETS

We begin this section with the definition and properties of *affine sets*. Given two elements a and b in $\mathbb{R}^n$, the line connecting them is

$$\mathcal{L}[a, b] := \big\{\lambda a + (1 - \lambda) b \ \big|\ \lambda \in \mathbb{R}\big\}.$$

Note that if $a = b$, then $\mathcal{L}[a, b] = \{a\}$.

Definition 1.55 *A subset* Ω *of* $\mathbb{R}^n$ *is* AFFINE *if for any* $a, b \in \Omega$ *we have* $\mathcal{L}[a, b] \subset \Omega$.

For instance, any point, line, and plane in $\mathbb{R}^3$ are affine sets. The empty set and the whole space are always affine. It follows from the definition that the intersection of any collection of affine sets is affine. This leads us to the construction of the *affine hull* of a set.

Definition 1.56 *The* AFFINE HULL *of a set* $\Omega \subset \mathbb{R}^n$ *is*

$$\text{aff}\, \Omega := \bigcap \big\{C \ \big|\ C \text{ is affine and } \ \Omega \subset C\big\}.$$

An element x in $\mathbb{R}^n$ of the form

$$x = \sum_{i=1}^{m} \lambda_i \omega_i \text{ with } \sum_{i=1}^{m} \lambda_i = 1, \quad m \in \mathbb{N},$$

is called an *affine combination* of $\omega_1, \ldots, \omega_m$. The proof of the next proposition is straightforward and thus is omitted.

Proposition 1.57 *The following assertions hold:*

(i) *A set Ω is affine if and only if Ω contains all affine combinations of its elements.*
(ii) *Let Ω, Ω_1, and Ω_2 be affine subsets of $\mathbb{R}^n$. Then the sum $\Omega_1 + \Omega_2$ and the scalar product $\lambda\Omega$ for any $\lambda \in \mathbb{R}$ are also affine subsets of $\mathbb{R}^n$.*
(iii) *Let $B : \mathbb{R}^n \to \mathbb{R}^p$ be an affine mapping. If Ω is an affine subset of $\mathbb{R}^n$ and Θ is an affine subset of $\mathbb{R}^p$, then the image $B(\Omega)$ is an affine subset of $\mathbb{R}^p$ and the inverse image $B^{-1}(\Theta)$ is an affine subset of $\mathbb{R}^n$.*
(iv) *Given $\Omega \subset \mathbb{R}^n$, its affine hull is the smallest affine set containing Ω. We have*

$$\operatorname{aff} \Omega = \Big\{ \sum_{i=1}^{m} \lambda_i \omega_i \ \Big|\ \sum_{i=1}^{m} \lambda_i = 1,\ \omega_i \in \Omega,\ m \in \mathbb{N} \Big\}.$$

(v) *A set Ω is a (linear) subspace if and only if Ω is an affine set containing the origin.*

Next we consider relationships between affine sets and (linear) subspaces.

Lemma 1.58 *A nonempty subset Ω of $\mathbb{R}^n$ is affine if and only if $\Omega - \omega$ is a subspace of $\mathbb{R}^n$ for any $\omega \in \Omega$.*

Proof. Suppose that a nonempty set $\Omega \subset \mathbb{R}^n$ is affine. It follows from Proposition 1.57(v) that $\Omega - \omega$ is a subspace for any $\omega \in \Omega$. Conversely, fix $\omega \in \Omega$ and suppose that $\Omega - \omega$ is a subspace denoted by L. Then the set $\Omega = \omega + L$ is obviously affine. □

The preceding lemma leads to the following notion.

Definition 1.59 *An affine set Ω is* PARALLEL *to a subspace L if $\Omega = \omega + L$ for some $\omega \in \Omega$.*

The next proposition justifies the form of the parallel subspace.

Proposition 1.60 *Let Ω be a nonempty, affine subset of $\mathbb{R}^n$. Then it is parallel to the unique subspace L of $\mathbb{R}^n$ given by $L = \Omega - \Omega$.*

Proof. Given a nonempty, affine set Ω, fix $\omega \in \Omega$ and come up to the linear subspace $L := \Omega - \omega$ parallel to Ω. To justify the uniqueness of such L, take any $\omega_1, \omega_2 \in \Omega$ and any subspaces $L_1, L_2 \subset \mathbb{R}^n$ such that $\Omega = \omega_1 + L_1 = \omega_2 + L_2$. Then $L_1 = \omega_2 - \omega_1 + L_2$. Since $0 \in L_1$, we have $\omega_1 - \omega_2 \in L_2$. This implies that $\omega_2 - \omega_1 \in L_2$ and thus $L_1 = \omega_2 - \omega_1 + L_2 \subset L_2$. Similarly, we get $L_2 \subset L_1$, which justifies that $L_1 = L_2$.

It remains to verify the representation $L = \Omega - \Omega$. Let $\Omega = \omega + L$ with the unique subspace L and some $\omega \in \Omega$. Then $L = \Omega - \omega \subset \Omega - \Omega$. Fix any $x = u - \omega$ with $u, \omega \in \Omega$ and observe that $\Omega - \omega$ is a subspace parallel to Ω. Hence $\Omega - \omega = L$ by the uniqueness of L proved above. This ensures that $x \in \Omega - \omega = L$ and thus $\Omega - \Omega \subset L$. □

The uniqueness of the parallel subspace shows that the next notion is well defined.

Definition 1.61 *The* DIMENSION OF AN AFFINE SET $\emptyset \neq \Omega \subset \mathbb{R}^n$ *is the dimension of the linear subspace parallel to* Ω*. Furthermore, the* DIMENSION OF A CONVEX SET $\emptyset \neq \Omega \subset \mathbb{R}^n$ *is the dimension of its affine hull* aff Ω.

To proceed further, we need yet another definition important in what follows.

Definition 1.62 *The elements* $v_0, \ldots, v_m$ *in* $\mathbb{R}^n$, $m \geq 1$, *are* AFFINELY INDEPENDENT *if*

$$\Big[\sum_{i=0}^{m} \lambda_i v_i = 0,\ \sum_{i=0}^{m} \lambda_i = 0\Big] \Longrightarrow \big[\lambda_i = 0 \ \textit{for all}\ i = 0, \ldots, m\big].$$

It is easy to observe the following relationship with the linear independence.

Proposition 1.63 *The elements* $v_0, \ldots, v_m$ *in* $\mathbb{R}^n$ *are affinely independent if and only if the elements* $v_1 - v_0, \ldots, v_m - v_0$ *are linearly independent.*

Proof. Suppose that $v_0, \ldots, v_m$ are affinely independent and consider the system

$$\sum_{i=1}^{m} \lambda_i (v_i - v_0) = 0, \ \text{ i.e., } \ \lambda_0 v_0 + \sum_{i=1}^{m} \lambda_i v_i = 0,$$

where $\lambda_0 := -\sum_{i=1}^{m} \lambda_i$. Since the elements $v_0, \ldots, v_m$ are affinely independent and $\sum_{i=0}^{m} \lambda_i = 0$, we have that $\lambda_i = 0$ for all $i = 1, \ldots, m$. Thus $v_1 - v_0, \ldots, v_m - v_0$ are linearly independent. The proof of the converse statement is straightforward. □

Recall that the *span* of some set C, span C, is the linear subspace generated by C.

Lemma 1.64 *Let* $\Omega :=$ aff$\{v_0, \ldots, v_m\}$, *where* $v_i \in \mathbb{R}^n$ *for all* $i = 0, \ldots, m$. *Then the span of the set* $\{v_1 - v_0, \ldots, v_m - v_0\}$ *is the subspace parallel to* Ω.

Proof. Denote by L the subspace parallel to Ω. Then $\Omega - v_0 = L$ and therefore $v_i - v_0 \in L$ for all $i = 1, \ldots, m$. This gives

$$\operatorname{span}\{v_i - v_0 \mid i = 1, \ldots, m\} \subset L.$$

To prove the converse inclusion, fix any $v \in L$ and get $v + v_0 \in \Omega$. Thus we have

$$v + v_0 = \sum_{i=0}^{m} \lambda_i v_i, \quad \sum_{i=0}^{m} \lambda_i = 1.$$

This implies the relationship

$$v = \sum_{i=1}^{m} \lambda_i (v_i - v_0) \in \operatorname{span}\{v_i - v_0 \mid i = 1, \ldots, m\},$$

which justifies the converse inclusion and hence completes the proof. □

The proof of the next proposition is rather straightforward.

Proposition 1.65 *The elements $v_0, \ldots, v_m$ are affinely independent in $\mathbb{R}^n$ if and only if its affine hull $\Omega := \operatorname{aff}\{v_0, \ldots, v_m\}$ is m-dimensional.*

Proof. Suppose that $v_0, \ldots, v_m$ are affinely independent. Then Lemma 1.64 tells us that the subspace $L := \operatorname{span}\{v_i - v_0 \mid i = 1, \ldots, m\}$ is parallel to Ω. The linear independence of $v_1 - v_0, \ldots, v_m - v_0$ by Proposition 1.63 means that the subspace L is m-dimensional and so is Ω. The proof of the converse statement is also straightforward. □

Affinely independent systems lead us to the construction of *simplices*.

Definition 1.66 *Let $v_0, \ldots, v_m$ be affinely independent in $\mathbb{R}^n$. Then the set*

$$\Delta_m := \operatorname{co}\{v_i \mid i = 0, \ldots, m\}$$

is called an m-SIMPLEX in $\mathbb{R}^n$ with the vertices v_i, $i = 0, \ldots, m$.

An important role of simplex vertices is revealed by the following proposition.

Proposition 1.67 *Consider an m-simplex Δ_m with vertices v_i for $i = 0, \ldots, m$. For every $v \in \Delta_m$, there is a unique element $(\lambda_0, \ldots, \lambda_m) \in \mathbb{R}^{m+1}_+$ such that*

$$v = \sum_{i=0}^{m} \lambda_i v_i, \quad \sum_{i=0}^{m} \lambda_i = 1.$$

Proof. Let $(\lambda_0, \ldots, \lambda_m) \in \mathbb{R}_+^{m+1}$ and $(\mu_0, \ldots, \mu_m) \in \mathbb{R}_+^{m+1}$ satisfy

$$v = \sum_{i=0}^{m} \lambda_i v_i = \sum_{i=0}^{m} \mu_i v_i, \quad \sum_{i=0}^{m} \lambda_i = \sum_{i=0}^{m} \mu_i = 1.$$

This immediately implies the equalities

$$\sum_{i=0}^{m} (\lambda_i - \mu_i) v_i = 0, \quad \sum_{i=0}^{m} (\lambda_i - \mu_i) = 0.$$

Since $v_0, \ldots, v_m$ are affinely independent, we have $\lambda_i = \mu_i$ for $i = 0, \ldots, m$. □

Now we are ready to define a major notion of *relative interiors* of convex sets.

Definition 1.68 *Let Ω be a convex set. We say that an element $v \in \Omega$ belongs to the* RELATIVE INTERIOR ri Ω *of Ω if there exists $\epsilon > 0$ such that $I\!B(v;\epsilon) \cap \operatorname{aff} \Omega \subset \Omega$.*

We begin the study of relative interiors with the following lemma.

Lemma 1.69 *Any linear mapping $A : \mathbb{R}^n \to \mathbb{R}^p$ is continuous.*

Proof. Let $\{e_1, \ldots, e_n\}$ be the standard orthonormal basis of $\mathbb{R}^n$ and let $v_i := A(e_i), i = 1, \ldots, n$. For any $x \in \mathbb{R}^n$ with $x = (x_1, \ldots, x_n)$, we have

$$A(x) = A\Big(\sum_{i=1}^{n} x_i e_i\Big) = \sum_{i=1}^{n} x_i A(e_i) = \sum_{i=1}^{n} x_i v_i.$$

Then the triangle inequality and the Cauchy-Schwarz inequality give us

$$\|A(x)\| \le \sum_{i=1}^{n} |x_i| \|v_i\| \le \sqrt{\sum_{i=1}^{n} |x_i|^2} \sqrt{\sum_{i=1}^{n} \|v_i\|^2} = M\|x\|,$$

where $M := \sqrt{\sum_{i=1}^{n} \|v_i\|^2}$. It follows furthermore that

$$\|A(x) - A(y)\| = \|A(x-y)\| \le M\|x-y\| \text{ for all } x, y \in \mathbb{R}^n,$$

which implies the continuity of the mapping A. □

The next proposition plays an essential role in what follows.

Proposition 1.70 *Let Δ_m be an m-simplex in $\mathbb{R}^n$ with some $m \ge 1$. Then* ri $\Delta_m \neq \emptyset$.

Proof. Consider the vertices $v_0, \ldots, v_m$ of the simplex Δ_m and denote

$$v := \frac{1}{m+1} \sum_{i=0}^{m} v_i.$$

We prove the proposition by showing that $v \in \operatorname{ri} \Delta_m$. Define

$$L := \operatorname{span}\{v_i - v_0 \mid i = 1, \ldots, m\}$$

and observe that L is the m-dimensional subspace of $\mathbb{R}^n$ parallel to $\operatorname{aff} \Delta_m = \operatorname{aff}\{v_0, \ldots, v_m\}$. It is easy to see that for every $x \in L$ there is a unique collection $(\lambda_0, \ldots, \lambda_m) \in \mathbb{R}^{m+1}$ with

$$x = \sum_{i=0}^{m} \lambda_i v_i, \quad \sum_{i=0}^{m} \lambda_i = 0.$$

Consider the mapping $A : L \to \mathbb{R}^{m+1}$, which maps x to the corresponding coefficients $(\lambda_0, \ldots, \lambda_m) \in \mathbb{R}^{m+1}$ as above. Then A is linear, and hence it is continuous by Lemma 1.69. Since $A(0) = 0$, we can choose $\delta > 0$ such that

$$\|A(u)\| < \frac{1}{m+1} \text{ whenever } \|u\| \le \delta.$$

Let us now show that $(v + \delta I\!B) \cap \operatorname{aff} \Delta_m \subset \Delta_m$, which means that $v \in \operatorname{ri} \Delta_m$. To proceed, fix any $x \in (v + \delta I\!B) \cap \operatorname{aff} \Delta_m$ and get that $x = v + u$ for some $u \in \delta I\!B$. Since $v, x \in \operatorname{aff} \Delta_m$ and $u = x - v$, we have $u \in L$. Denoting $A(u) := (\alpha_0, \ldots, \alpha_m)$ gives us the representation $u = \sum_{i=0}^{m} \alpha_i v_i$ with $\sum_{i=0}^{m} \alpha_i = 0$ and the estimate

$$|\alpha_i| \le \|A(u)\| < \frac{1}{m+1} \text{ for all } i = 0, \ldots, m.$$

Then implies in turn that

$$v + u = \sum_{i=0}^{m} \Big(\frac{1}{m+1} + \alpha_i\Big) v_i = \sum_{i=0}^{m} \mu_i v_i,$$

where $\mu_i := \dfrac{1}{m+1} + \alpha_i \ge 0$ for $i = 0, \ldots, m$. Since $\sum_{i=0}^{m} \mu_i = 1$, this ensures that $x \in \Delta_m$. Thus $(v + \delta I\!B) \cap \operatorname{aff} \Delta_m \subset \Delta_m$ and therefore $v \in \operatorname{ri} \Delta_m$. □

Lemma 1.71 *Let Ω be a nonempty, convex set in $\mathbb{R}^n$ of dimension $m \ge 1$. Then there exist $m + 1$ affinely independent elements $v_0, \ldots, v_m$ in Ω.*

Proof. Let $\Delta_k := \{v_0, \ldots, v_k\}$ be a $k-$simplex of maximal dimension contained in Ω. Then $v_0, \ldots, v_k$ are affinely independent. To verify now that $k = m$, form $K := \text{aff}\{v_0, \ldots, v_k\}$ and observe that $K \subset \text{aff}\,\Omega$ since $\{v_0, \ldots, v_k\} \subset \Omega$. The opposite inclusion also holds since we have $\Omega \subset K$. Justifying it, we argue by contradiction and suppose that there exists $w \in \Omega$ such that $w \notin K$. Then a direct application of the definition of affine independence shows that $v_0, \ldots, v_k, w$ are affinely independent being a subset of Ω, which is a contradiction. Thus $K = \text{aff}\,\Omega$, and hence we get $k = \dim K = \dim \text{aff}\ \Omega = \dim \Omega = m$. □

The next is one of the most fundamental results of convex finite-dimensional geometry.

Theorem 1.72 Let Ω be a nonempty, convex set in $\mathbb{R}^n$. The following assertions hold:

(i) We always have $\text{ri}\,\Omega \neq \emptyset$.
(ii) We have $[a, b) \subset \text{ri}\,\Omega$ for any $a \in \text{ri}\,\Omega$ and $b \in \overline{\Omega}$.

Proof. **(i)** Let m be the dimension of Ω. Observe first that the case where $m = 0$ is trivial since in this case Ω is a singleton and $\text{ri}\,\Omega = \Omega$. Suppose that $m \geq 1$ and find $m + 1$ affinely independent elements $v_0, \ldots, v_m$ in Ω as in Lemma 1.71. Consider further the m-simplex

$$\Delta_m := \text{co}\big\{v_0, \ldots, v_m\big\}.$$

We can show that $\text{aff}\,\Delta_m = \text{aff}\,\Omega$. To complete the proof, take $v \in \text{ri}\,\Delta_m$, which exists by Proposition 1.70, and get for any small $\epsilon > 0$ that

$$I\!B(v, \epsilon) \cap \text{aff}\,\Omega = I\!B(v, \epsilon) \cap \text{aff}\,\Delta_m \subset \Delta_m \subset \Omega.$$

This verifies that $v \in \text{ri}\,\Omega$ by the definition of relative interior.

(ii) Let L be the subspace of $\mathbb{R}^n$ parallel to $\text{aff}\,\Omega$ and let $m := \dim L$. Then there is a bijective linear mapping $A : L \to \mathbb{R}^m$ such that both A and A^{-1} are continuous. Fix $x_0 \in \text{aff}\,\Omega$ and define the mapping $f : \text{aff}\,\Omega \to \mathbb{R}^m$ by $f(x) := A(x - x_0)$. It is easy to check that f is a bijective affine mapping and that both f and f^{-1} are continuous. We also see that $a \in \text{ri}\,\Omega$ if and only if $f(a) \in \text{int}\, f(\Omega)$, and that $b \in \overline{\Omega}$ if and only if $f(b) \in \overline{f(\Omega)}$. Then $[f(a), f(b)) \subset \text{int}\, f(\Omega)$ by Lemma 1.30. This shows that $[a, b) \subset \text{ri}\,\Omega$. □

We conclude this section by the following properties of the relative interior.

Proposition 1.73 *Let Ω be a nonempty, convex subset of $\mathbb{R}^n$. For the convex sets* $\text{ri}\,\Omega$ *and* $\overline{\Omega}$, *we have that* **(i)** $\overline{\text{ri}\,\Omega} = \overline{\Omega}$ *and* **(ii)** $\text{ri}\,\Omega = \text{ri}\,\overline{\Omega}$.

Proof. Note that the convexity of $\text{ri}\,\Omega$ follows from Theorem 1.72(ii) while the convexity of $\overline{\Omega}$ was proved in Proposition 1.29. To justify assertion **(i)** of this proposition, observe that the inclusion $\overline{\text{ri}\,\Omega} \subset \overline{\Omega}$ is obvious. Fix $b \in \overline{\Omega}$, choose $a \in \text{ri}\,\Omega$, and form the sequence

$$x_k := \frac{1}{k}a + \Big(1 - \frac{1}{k}\Big)b, \quad k \in \mathbb{N},$$

which converges to b as $k \to \infty$. Since $x_k \in \text{ri}\,\Omega$ by Theorem 1.72(ii), we have $b \in \overline{\text{ri}\,\Omega}$. Thus $\overline{\Omega} \subset \overline{\text{ri}\,\Omega}$, which verifies **(i)**. The proof **(ii)** is similar to that of Proposition 1.31(ii). □

1.5 THE DISTANCE FUNCTION

The last section of this chapter is devoted to the study of the class of distance functions for convex sets, which belongs to the most interesting and important subjects of convex analysis and its extensions. Functions of this type are intrinsically *nondifferentiable* while they naturally and frequently appear in analysis and applications.

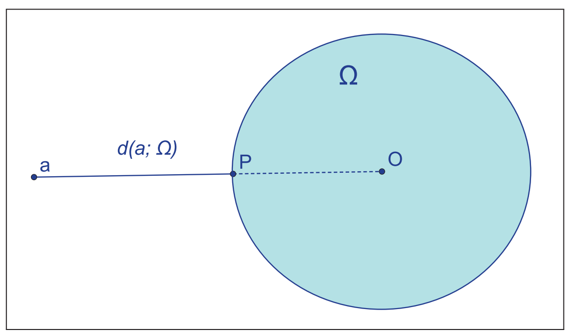


Figure 1.6: Distance function.

Given a set $\Omega \subset \mathbb{R}^n$, the *distance function* associated with Ω is defined by

$$d(x;\Omega) := \inf\big\{\|x-\omega\| \;\big|\; \omega \in \Omega\big\}. \tag{1.5}$$

Recall further that a mapping $f\colon \mathbb{R}^n \to \mathbb{R}^p$ is *Lipschitz continuous* with constant $\ell \geq 0$ on some set $C \subset \mathbb{R}^n$ if we have the estimate

$$\|f(x) - f(y)\| \leq \ell\|x-y\| \text{ for all } x, y \in C. \tag{1.6}$$

Note that the Lipschitz continuity of f in (1.6) specifies its continuity with a *linear rate*.

Proposition 1.74 *Let Ω be a nonempty subset of $\mathbb{R}^n$. The following hold:*

(i) *$d(x;\Omega) = 0$ if and only if $x \in \overline{\Omega}$.*

(ii) *The function $d(x;\Omega)$ is Lipschitz continuous with constant $\ell = 1$ on $\mathbb{R}^n$.*

Proof. **(i)** Suppose that $d(x;\Omega) = 0$. For each $k \in \mathbb{N}$, find $\omega_k \in \Omega$ such that

$$0 = d(x;\Omega) \le \|x - \omega_k\| < d(x;\Omega) + \frac{1}{k} = \frac{1}{k},$$

which ensures that the sequence $\{\omega_k\}$ converges to x, and hence $x \in \overline{\Omega}$.

Conversely, let $x \in \overline{\Omega}$ and find a sequence $\{\omega_k\} \subset \Omega$ converging to x. Then

$$0 \le d(x;\Omega) \le \|x - \omega_k\| \ \text{ for all } \ k \in \mathbb{N},$$

which implies that $d(x;\Omega) = 0$ since $\|x - \omega_k\| \to 0$ as $k \to \infty$.

(ii) For any $\omega \in \Omega$, we have the estimate

$$d(x;\Omega) \le \|x - \omega\| \le \|x - y\| + \|y - \omega\|,$$

which implies in turn that

$$d(x;\Omega) \le \|x - y\| + \inf\big\{\|y - \omega\| \ \big|\ \omega \in \Omega\big\} = \|x - y\| + d(y;\Omega).$$

Similarly, we get $d(y;\Omega) \le \|y - x\| + d(x;\Omega)$ and thus $|d(x;\Omega) - d(y;\Omega)| \le \|x - y\|$, which justifies by (1.6) the Lipschitz continuity of $d(x;\Omega)$ on $\mathbb{R}^n$ with constant $\ell = 1$. □

For each $x \in \mathbb{R}^n$, the *Euclidean projection* from x to Ω is defined by

$$\Pi(x;\Omega) := \big\{\omega \in \Omega \ \big|\ \|x - \omega\| = d(x;\Omega)\big\}. \tag{1.7}$$

Proposition 1.75 *Let Ω be a nonempty, closed subset of $\mathbb{R}^n$. Then for any $x \in \mathbb{R}^n$ the Euclidean projection $\Pi(x;\Omega)$ is nonempty.*

Proof. By definition (1.7), for each $k \in \mathbb{N}$ there exists $\omega_k \in \Omega$ such that

$$d(x;\Omega) \le \|x - \omega_k\| < d(x;\Omega) + \frac{1}{k}.$$

It is clear that $\{\omega_k\}$ is a bounded sequence. Thus it has a subsequence $\{\omega_{k_\ell}\}$ that converges to ω. Since Ω is closed, $\omega \in \Omega$. Letting $\ell \to \infty$ in the inequality

$$d(x;\Omega) \le \|x - \omega_{k_\ell}\| < d(x;\Omega) + \frac{1}{k_\ell},$$

we have $d(x;\Omega) = \|x - \omega\|$, which ensures that $\omega \in \Pi(x;\Omega)$. □

An interesting consequence of convexity is the following unique projection property.

Corollary 1.76 *If Ω is a nonempty, closed, convex subset of $\mathbb{R}^n$, then for each $x \in \mathbb{R}^n$ the Euclidean projection $\Pi(x;\Omega)$ is a singleton.*

Proof. The nonemptiness of the projection $\Pi(x;\Omega)$ follows from Proposition 1.75. To prove the uniqueness, suppose that $\omega_1, \omega_2 \in \Pi(x;\Omega)$ with $\omega_1 \neq \omega_2$. Then

$$\|x-\omega_1\| = \|x-\omega_2\| = d(x;\Omega).$$

By the classical parallelogram equality, we have that

$$2\|x-\omega_1\|^2 = \|x-\omega_1\|^2 + \|x-\omega_2\|^2 = 2\Big\|x - \frac{\omega_1+\omega_2}{2}\Big\|^2 + \frac{\|\omega_1-\omega_2\|^2}{2}.$$

This directly implies that

$$\Big\|x - \frac{\omega_1+\omega_2}{2}\Big\|^2 = \|x-\omega_1\|^2 - \frac{\|\omega_1-\omega_2\|^2}{4} < \|x-\omega_1\|^2 = \big[d(x;\Omega)\big]^2,$$

which is a contradiction due to the inclusion $\dfrac{\omega_1+\omega_2}{2} \in \Omega$. □

Now we show that the convexity of a nonempty, closed set and its distance function are equivalent. It is an easy exercise to show that the convexity of an arbitrary set Ω implies the convexity of its distance function.

Proposition 1.77 *Let Ω be a nonempty, closed subset of $\mathbb{R}^n$. Then the function $d(\cdot;\Omega)$ is convex if and only if the set Ω is convex.*

Proof. Suppose that Ω is convex. Taking $x_1, x_2 \in \mathbb{R}^n$ and $\omega_i := \Pi(x_i;\Omega)$, we have

$$\|x_i - \omega_i\| = d(x_i;\Omega) \ \text{ for } \ i = 1,2.$$

The convexity of Ω ensures that $\lambda\omega_1 + (1-\lambda)\omega_2 \in \Omega$ for any $\lambda \in (0,1)$. It yields

$$\begin{aligned} d\big(\lambda x_1 + (1-\lambda)x_2;\Omega\big) &\le \|\lambda x_1 + (1-\lambda)x_2 - [\lambda\omega_1 + (1-\lambda)\omega_2]\| \\ &\le \lambda\|x_1-\omega_1\| + (1-\lambda)\|x_2-\omega_2\| \\ &= \lambda d(x_1;\Omega_1) + (1-\lambda)d(x_2;\Omega_2), \end{aligned}$$

which implies therefore the convexity of the distance function $d(\cdot;\Omega)$ by

$$d\big(\lambda x_1 + (1-\lambda)x_2;\Omega\big) \le \lambda d(x_1;\Omega) + (1-\lambda)d(x_2;\Omega).$$

To prove the converse implication, suppose that $d(\cdot;\Omega)$ is convex and fix any $\omega_i \in \Omega$ for $i = 1, 2$ and $\lambda \in (0, 1)$. Then we have

$$d\big(\lambda\omega_1 + (1-\lambda)\omega_2;\Omega\big) \le \lambda d(\omega_1;\Omega) + (1-\lambda)d(\omega_2;\Omega) = 0.$$

Since Ω is closed, this yields $\lambda\omega_1 + (1-\lambda)\omega_2 \in \Omega$ and so justifies the convexity of Ω. □

Next we characterize the Euclidean projection to convex sets in $\mathbb{R}^n$. In the proposition below and in what follows we often identify the projection $\Pi(x;\Omega)$ with its unique element if Ω is a nonempty, closed, convex set.

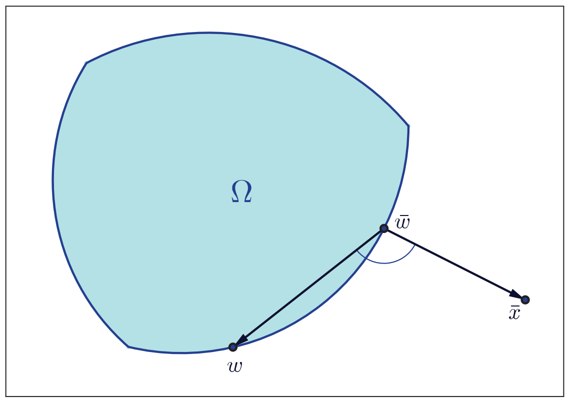


Figure 1.7: Euclidean projection.

Proposition 1.78 *Let Ω be a nonempty, convex subset of $\mathbb{R}^n$ and let $\bar{\omega} \in \Omega$. Then we have $\bar{\omega} \in \Pi(\bar{x};\Omega)$ if and only if*

$$\langle \bar{x} - \bar{\omega}, \omega - \bar{\omega}\rangle \le 0 \ \textit{for all}\ \ \omega \in \Omega. \tag{1.8}$$

Proof. Take $\bar{\omega} \in \Pi(\bar{x};\Omega)$ and get for any $\omega \in \Omega$, $\lambda \in (0,1)$ that $\bar{\omega} + \lambda(\omega - \bar{\omega}) \in \Omega$. Thus

$$\begin{aligned}\|\bar{x} - \bar{\omega}\|^2 &= \big[d(\bar{x};\Omega)\big]^2 \le \|\bar{x} - \big[\bar{\omega} + \lambda(\omega - \bar{\omega})\|^2\\ &= \|\bar{x} - \bar{\omega}\|^2 - 2\lambda\langle \bar{x} - \bar{\omega}, \omega - \bar{\omega}\rangle + \lambda^2\|\omega - \bar{\omega}\|^2.\end{aligned}$$

This readily implies that

$$2\langle \bar{x} - \bar{\omega}, \omega - \bar{\omega}\rangle \le \lambda\|\omega - \bar{\omega}\|^2.$$

Letting $\lambda \to 0^+$, we arrive at (1.8).

To verify the converse, suppose that (1.8) holds. Then for any $\omega \in \Omega$ we get

$$\begin{aligned}\|\bar{x}-\omega\|^2 &= \|\bar{x}-\bar{\omega}+\bar{\omega}-\omega\|^2\\ &= \|\bar{x}-\bar{\omega}\|^2+\|\bar{\omega}-\omega\|^2+2\langle \bar{x}-\bar{\omega},\bar{\omega}-\omega\rangle\\ &= \|\bar{x}-\bar{\omega}\|^2+\|\bar{\omega}-\omega\|^2-2\langle \bar{x}-\bar{\omega},\omega-\bar{\omega}\rangle \geq \|\bar{x}-\bar{\omega}\|^2.\end{aligned}$$

Thus $\|\bar{x}-\bar{\omega}\| \leq \|\bar{x}-\omega\|$ for all $\omega \in \Omega$, which implies $\bar{\omega} \in \Pi(\bar{x};\Omega)$ and completes the proof. □

We know from the above that for any nonempty, closed, convex set Ω in $\mathbb{R}^n$ the Euclidean projection $\Pi(x;\Omega)$ is a singleton. Now we show that the projection mapping is in fact *nonexpansive*, i.e., it satisfies the following Lipschitz property.

Proposition 1.79 *Let Ω be a nonempty, closed, convex subset of $\mathbb{R}^n$. Then for any elements $x_1, x_2 \in \mathbb{R}^n$ we have the estimate*

$$\big\|\Pi(x_1;\Omega)-\Pi(x_2;\Omega)\big\|^2 \leq \big\langle \Pi(x_1;\Omega)-\Pi(x_2;\Omega), x_1-x_2\big\rangle.$$

In particular, it implies the Lipschitz continuity of the projection with constant $\ell = 1$:

$$\big\|\Pi(x_1;\Omega)-\Pi(x_2;\Omega)\big\| \leq \|x_1-x_2\| \text{ for all } x_1,x_2 \in \mathbb{R}^n.$$

Proof. It follows from the preceding proposition that

$$\big\langle \Pi(x_2;\Omega)-\Pi(x_1;\Omega), x_1-\Pi(x_1;\Omega)\big\rangle \leq 0 \text{ for all } x_1,x_2 \in \mathbb{R}^n.$$

Changing the roles of x_1, x_2 in the inequality above and summing them up give us

$$\big\langle \Pi(x_1;\Omega)-\Pi(x_2;\Omega), x_2-x_1+\Pi(x_1;\Omega)-\Pi(x_2;\Omega)\big\rangle \leq 0.$$

This implies the first estimate in the proposition. Finally, the nonexpansive property of the Euclidean projection follows directly from

$$\begin{aligned}\big\|\Pi(x_1;\Omega)-\Pi(x_2;\Omega)\big\|^2 &\leq \big\langle \Pi(x_1;\Omega)-\Pi(x_2;\Omega), x_1-x_2\big\rangle\\ &\leq \|\Pi(x_1;\Omega)-\Pi(x_2;\Omega)\|\cdot\|x_1-x_2\|\end{aligned}$$

for all $x_1, x_2 \in \mathbb{R}^n$, which completes the proof of the proposition. □

1.6 EXERCISES FOR CHAPTER 1

Exercise 1.1 Let Ω_1 and Ω_2 be nonempty, closed, convex subsets of $\mathbb{R}^n$ such that Ω_1 is bounded or Ω_2 is bounded. Show that $\Omega_1 - \Omega_2$ is a nonempty, closed, convex set. Give an example of two nonempty, closed, convex sets Ω_1, Ω_2 for which $\Omega_1 - \Omega_2$ is not closed.

Exercise 1.2 Let Ω be a subset of $\mathbb{R}^n$. We say that Ω is a *cone* if $\lambda x \in \Omega$ whenever $\lambda \geq 0$ and $x \in \Omega$. Show that the following are equivalent:

(i) Ω is a convex cone.
(ii) $x + y \in \Omega$ whenever $x, y \in \Omega$, and $\lambda x \in \Omega$ whenever $x \in \Omega$ and $\lambda \geq 0$.

Exercise 1.3 **(i)** Let Ω be a nonempty, convex set that contains 0 and let $0 \leq \lambda_1 \leq \lambda_2$. Show that $\lambda_1 \Omega \subset \lambda_2 \Omega$.
(ii) Let Ω be a nonempty, convex set and let $\alpha, \beta \geq 0$. Show that $\alpha\Omega + \beta\Omega \subset (\alpha + \beta)\Omega$.

Exercise 1.4 **(i)** Let Ω_i for $i = 1, \ldots, m$ be nonempty, convex sets in $\mathbb{R}^n$. Show that $x \in \text{co} \bigcup_{i=1}^m \Omega_i$ if and only if there exist elements $\omega_i \in \Omega_i$ and $\lambda_i \geq 0$ for $i = 1, \ldots, m$ with $\sum_{i=1}^m \lambda_i = 1$ such that $x = \sum_{i=1}^m \lambda_i \omega_i$.
(ii) Let Ω_i for $i = 1, \ldots, m$ be nonempty, convex cones in $\mathbb{R}^n$. Show that

$$\sum_{i=1}^m \Omega_i = \text{co}\,\{\bigcup_{i=1}^m \Omega_i\}.$$

Exercise 1.5 Let Ω be a nonempty, convex cone in $\mathbb{R}^n$. Show that Ω is a linear subspace of $\mathbb{R}^n$ if and only if $\Omega = -\Omega$.

Exercise 1.6 Show that the following functions are convex on $\mathbb{R}^n$:

(i) $f(x) = \alpha\|x\|$, where $\alpha \geq 0$.
(ii) $f(x) = \|x - a\|^2$, where $a \in \mathbb{R}^n$.
(iii) $f(x) = \|Ax - b\|$, where A is an $p \times n$ matrix and $b \in \mathbb{R}^p$.
(iv) $f(x) = \|x\|^q$, where $q \geq 1$.

Exercise 1.7 Show that the following functions are convex on the given domains:

(i) $f(x) = e^{ax}$, $x \in \mathbb{R}$, where a is a constant.
(ii) $f(x) = x^q$, $x \in [0, \infty)$, where $q \geq 1$ is a constant.
(iii) $f(x) = -\ln(x)$, $x \in (0, \infty)$.
(iv) $f(x) = x\ln(x)$, $x \in (0, \infty)$.

Exercise 1.8 **(i)** Give an example of a function $f : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$, which is convex with respect to each variable but not convex with respect to both.
(ii) Let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ for $i = 1, 2$ be convex functions. Can we conclude that the *minimum function* $\min\{f_1, f_2\}(x) := \min\{f_1(x), f_2(x)\}$ is convex?

Exercise 1.9 Give an example showing that the product of two real-valued convex functions is not necessarily convex.

Exercise 1.10 Verify that the set $\{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid t \geq \|x\|\}$ is closed and convex.

Exercise 1.11 The *indicator function* associated with a set Ω is defined by

$$\delta(x;\Omega) := \begin{cases} 0 & \text{if } x \in \Omega, \\ \infty & \text{otherwise.} \end{cases}$$

(i) Calculate $\delta(\cdot;\Omega)$ for $\Omega = \emptyset$, $\Omega = \mathbb{R}^n$, and $\Omega = [-1,1]$.
(ii) Show that the set Ω is convex set if and only if its indicator function $\delta(\cdot;\Omega)$ is convex.

Exercise 1.12 Show that if $f : \mathbb{R} \to [0,\infty)$ is a convex function, then its q-power $f^q(x) := (f(x))^q$ is also convex for any $q \geq 1$.

Exercise 1.13 Let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ for $i = 1,2$ be convex functions. Define

$$\begin{aligned} \Omega_1 &:= \{(x,\lambda_1,\lambda_2) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \mid \lambda_1 \geq f_1(x)\}, \\ \Omega_2 &:= \{(x,\lambda_1,\lambda_2) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \mid \lambda_2 \geq f_2(x)\}. \end{aligned}$$

(i) Show that the sets Ω_1, Ω_2 are convex.
(ii) Define the set-valued mapping $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^2$ by $F(x) := [f_1(x),\infty) \times [f_2(x),\infty)$ and verify that the graph of F is $\Omega_1 \cap \Omega_2$.

Exercise 1.14 We say that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *positively homogeneous* if $f(\alpha x) = \alpha f(x)$ for all $\alpha > 0$, and that f is *subadditive* if $f(x+y) \leq f(x) + f(y)$ for all $x, y \in \mathbb{R}^n$. Show that a positively homogeneous function is convex if and only if it is subadditive.

Exercise 1.15 Given a nonempty set $\Omega \subset \mathbb{R}^n$, define

$$K_\Omega := \{\lambda x \mid \lambda \geq 0,\ x \in \Omega\} = \bigcup_{\lambda \geq 0} \lambda\Omega.$$

(i) Show that K_Ω is a cone.
(ii) Show that K_Ω is the smallest cone containing Ω.
(iii) Show that if Ω is convex, then the cone K_Ω is convex as well.

Exercise 1.16 Let $\varphi : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$ be a convex function and let K be a nonempty, convex subset of $\mathbb{R}^p$. Suppose that for each $x \in \mathbb{R}^n$ the function $\varphi(x,\cdot)$ is bounded below on K. Verify that the function on $\mathbb{R}^n$ defined by $f(x) := \inf\{\varphi(x,y) \mid y \in K\}$ is convex.

Exercise 1.17 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function.

(i) Show that for every $\alpha \in \mathbb{R}$ the *level set* $\{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$ is convex.
(ii) Let $\Omega \subset \overline{\mathbb{R}}$ be a convex set. Is it true in general that the *inverse image* $f^{-1}(\Omega)$ is a convex subset of $\mathbb{R}^n$?

Exercise 1.18 Let C be a convex subset of $\mathbb{R}^{n+1}$.

(i) For $x \in \mathbb{R}^n$, define $F(x) := \{\lambda \in \mathbb{R} \mid (x, \lambda) \in C\}$. Show that F is a set-valued mapping with convex graph and give an explicit formula for its graph.
(ii) For $x \in \mathbb{R}^n$, define the function

$$f_C(x) := \inf\big\{\lambda \in \mathbb{R} \,\big|\, (x, \lambda) \in C\big\}. \tag{1.9}$$

Find an explicit formula for f_C when C is the closed unit ball of $\mathbb{R}^2$.
(iii) For the function f_C defined in **(ii)**, show that if $f_C(x) > -\infty$ for all $x \in \mathbb{R}^n$, then f_C is a convex function.

Exercise 1.19 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be bounded from below. Define its *convexification*

$$(\text{co}\, f)(x) := \inf\Big\{\sum_{i=1}^{m} \lambda_i f(x_i) \,\Big|\, \lambda_i \geq 0,\ \sum_{i=1}^{m} \lambda_i = 1,\ \sum_{i=1}^{m} \lambda_i x_i = x,\ m \in \mathbb{N}\Big\}. \tag{1.10}$$

Verify that (1.10) is convex with $\text{co}\, f = f_C$, $C := \text{co}\,(\text{epi}\, f)$, and f_C defined in (1.9).

Exercise 1.20 We say that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *quasiconvex* if

$$f\big(\lambda x + (1-\lambda)y\big) \leq \max\big\{f(x), f(y)\big\} \text{ for all } x, y \in \mathbb{R}^n,\ \lambda \in (0, 1).$$

(i) Show that a function f is quasiconvex if and only if for any $\alpha \in \mathbb{R}$ the level set $\{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$ is a convex set.
(ii) Show that any convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is quasiconvex. Give an example demonstrating that the converse is not true.

Exercise 1.21 Let a be a nonzero element in $\mathbb{R}^n$ and let $b \in \mathbb{R}$. Show that

$$\Omega := \big\{x \in \mathbb{R}^n \,\big|\, \langle a, x\rangle = b\big\}$$

is an affine set with $\dim \Omega = 1$.

Exercise 1.22 Let the set $\{v_1, \ldots, v_m\}$ consist of affinely independent elements and let $v \notin \text{aff}\{v_1, \ldots, v_m\}$. Show that $v_1, \ldots, v_m, v$ are affinely independent.

Exercise 1.23 Suppose that Ω is a convex subset of $\mathbb{R}^n$ with $\dim \Omega = m$, $m \geq 1$. Let the set $\{v_1, \ldots, v_m\} \subset \Omega$ consist of affinely independent elements and let

$$\Delta_m := \text{co}\,\big\{v_1, \ldots, v_m\big\}.$$

Show that $\text{aff}\, \Omega = \text{aff}\, \Delta_m = \text{aff}\,\big\{v_1, \ldots, v_m\big\}$.

Exercise 1.24 Let Ω be a nonempty, convex subset of $\mathbb{R}^n$.

(i) Show that $\operatorname{aff}\Omega = \operatorname{aff}\overline{\Omega}$.
(ii) Show that for any $\bar{x} \in \operatorname{ri}\Omega$ and $x \in \overline{\Omega}$ there exists $t > 0$ such that $\bar{x} + t(\bar{x} - x) \in \Omega$.
(iii) Prove Proposition 1.73(ii).

Exercise 1.25 Let $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ be convex sets with $\operatorname{ri}\Omega_1 \cap \operatorname{ri}\Omega_2 \neq \emptyset$. Show that:

(i) $\overline{\Omega_1 \cap \Omega_2} = \overline{\Omega}_1 \cap \overline{\Omega}_2$.
(ii) $\operatorname{ri}(\Omega_1 \cap \Omega_2) = \operatorname{ri}\Omega_1 \cap \operatorname{ri}\Omega_2$.

Exercise 1.26 Let $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ be convex sets with $\overline{\Omega}_1 = \overline{\Omega}_2$. Show that $\operatorname{ri}\Omega_1 = \operatorname{ri}\Omega_2$.

Exercise 1.27 **(i)** Let $B : \mathbb{R}^n \to \mathbb{R}^p$ be an affine mapping and let Ω be a convex subset of $\mathbb{R}^n$. Prove the equality

$$B(\operatorname{ri}\Omega) = \operatorname{ri} B(\Omega).$$

(ii) Let Ω_1 and Ω_2 be convex subsets of $\mathbb{R}^n$. Show that $\operatorname{ri}(\Omega_1 - \Omega_2) = \operatorname{ri}\Omega_1 - \operatorname{ri}\Omega_2$.

Exercise 1.28 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function. Show that:

(i) $\operatorname{aff}(\operatorname{epi} f) = \operatorname{aff}(\operatorname{dom} f) \times \mathbb{R}$.
(ii) $\operatorname{ri}(\operatorname{epi} f) = \big\{(x, \lambda) \;\big|\; x \in \operatorname{ri}(\operatorname{dom} f),\ \lambda > f(x)\big\}$.

Exercise 1.29 Find the explicit formulas for the distance function $d(x; \Omega)$ and the Euclidean projection $\Pi(x; \Omega)$ in the following cases:

(i) Ω is the closed unit ball of $\mathbb{R}^n$.
(ii) $\Omega := \big\{(x_1, x_2) \in \mathbb{R}^2 \;\big|\; |x_1| + |x_2| \le 1\big\}$.
(iii) $\Omega := [-1, 1] \times [-1, 1]$.

Exercise 1.30 Find the formulas for the projection $\Pi(x; \Omega)$ in the following cases:

(i) $\Omega := \big\{x \in \mathbb{R}^n \;\big|\; \langle a, x\rangle = b\big\}$, where $0 \neq a \in \mathbb{R}^n$ and $b \in \mathbb{R}$.
(ii) $\Omega := \big\{x \in \mathbb{R}^n \;\big|\; Ax = b\big\}$, where A is an $m \times n$ matrix with rank $A = m$ and $b \in \mathbb{R}^m$.
(iii) Ω is the nonnegative orthant $\Omega := \mathbb{R}^n_+$.

Exercise 1.31 For $a_i \le b_i$ with $i = 1, \ldots, n$, define

$$\Omega := \big\{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \;\big|\; a_i \le x \le b_i \ \text{ for all } \ i = 1, \ldots, n\big\}.$$

Show that the projection $\Pi(x; \Omega)$ has the following representation:

$$\Pi(x; \Omega) = \big\{u = (u_1, \ldots, u_n) \in \mathbb{R}^n \;\big|\; u_i = \max\{a_i, \min\{b, x_i\}\} \ \text{ for } \ i \in \{1, \ldots, n\}\big\}.$$

A Standard material on convexity

Definition A.1 A set S in $\mathbb{R}^n$ is said to be *convex* if for every $x_1, x_2 \in S$ the line segment $\{\lambda x_1 + (1-\lambda)x_2 : 0 \le \lambda \le 1\}$ belongs to S.

For instance, a hyperplane $S = \{x \in \mathbb{R}^n : p^t x = \alpha\}$ or a ball $S = \{x \in \mathbb{R}^n : |x - x_0| \le \beta\}$ are examples of convex sets. However, the sphere $S = \{x \in \mathbb{R}^n : |x - x_0| = \beta\}$ provides an example of a set that is not convex ($\beta > 0$). It is easy to see that arbitrary intersections of convex sets are again convex; also finite sums of convex sets are convex again.

Theorem A.2 (strict point-set separation [1, Thm. 2.4.4]) *Let S be a nonempty closed convex subset of $\mathbb{R}^n$ and let $y \in \mathbb{R}^n \backslash S$. Then there exists $p \in \mathbb{R}^n$, $p \ne 0$, such that*

$$\sup_{x \in S} p^t x < p^t y.$$

PROOF. It is a standard result that there exists $\hat{x} \in S$ such that $\sup_{s \in S} |y - s| = |y - \hat{x}|$ (consider a suitable closed ball around y and apply the theorem of Weierstrass [1, Thm. 2.3.1]). By convexity of S, this means that for every $x \in S$ and every $\lambda \in (0, 1]$

$$|y - (\lambda x + (1-\lambda)\hat{x})|^2 \ge |y - \hat{x}|^2.$$

Obviously, the expression on the left equals

$$|y - \hat{x} - \lambda(x - \hat{x})|^2 = |y - \hat{x}|^2 - 2\lambda(y - \hat{x})^t(x - \hat{x}) + \lambda^2|x - \hat{x}|^2,$$

so the above inequality amounts to

$$2\lambda(y - \hat{x})^t(x - \hat{x}) \le \lambda^2|x - \hat{x}|^2$$

for every $x \in S$ and every $\lambda \in (0, 1]$. Dividing by $\lambda > 0$ and letting λ go to zero then gives

$$(y - \hat{x}) \cdot (x - \hat{x}) \le 0 \text{ for all } x \in S.$$

Set $p := y - \hat{x}$; then $p \ne 0$ (note that $p = 0$ would imply $y \in S$). We clearly have $p^t x \le p^t \hat{x}$. Also, we have now $p^t \hat{x} > p^t y$, for otherwise $(y - \hat{x})^t(\hat{x} - y) \ge 0$ would imply $y = \hat{x} \in S$, which is impossible. QED

For our next result, recall that $\partial S := \text{cl}S \cap \text{cl}(\mathbb{R}^n \backslash S) = \text{cl}S \backslash \text{int S}$ denotes the boundary of a set $S \subset \mathbb{R}^n$.

Theorem A.3 (supporting hyperplane [1, Thm. 2.4.7]) *Let S be a nonempty convex subset of $\mathbb{R}^n$ and let $y \in \partial S$. Then there exists $q \in \mathbb{R}^n$, $q \ne 0$, such that*

$$\sup_{x \in cl\ S} q^t x \le q^t y.$$

In geometric terms, $H := \{x \in \mathbb{R}^n : q^t x = q^t y\}$ is said to be a supporting hyperplane *for S at y: the hyperplane H contains the point y and the set S (as well as* cl S*) is contained the halfspace $\{x \in \mathbb{R}^n : p^t x \le p^t y\}$.*

Proof. Let $Z := \text{cl } S$; then $\partial S \subset \partial Z$ (exercise). Of course, Z is closed and it is easy to show that Z is convex (use limit arguments). So there exists a sequence (y_k) in $\mathbb{R}^n \backslash Z$ such that $y_k \to y$. By Theorem A.2 there exists for every k a nonzero vector $p_k \in \mathbb{R}^n$ such that

$$\sup_{x \in Z} p_k^t x < p_k^t y_k.$$

Division by $|p_k|$ turns this into

$$\sup_{x \in Z} q_k^t x < q_k^t y_k,$$

where $q_k := p_k/|p_k|$ belongs to the unit sphere of $\mathbb{R}^n$. This sphere is compact (Bolzano-Weierstrass theorem), so we can suppose without loss of generality that (q_k) converges to some q, $|q| = 1$ (so q is nonzero). Now for every $x \in Z$ the inequality $q_k^t x < q_k^t y_k$, which holds for all k, implies

$$q^t x = \lim_k q_k^t x \leq \lim_k q_k^t y_k = q^t y,$$

and the proof is finished. QED

Theorem A.4 (set-set separation [1, Thm. 2.4.8]) *Let S_1, S_2 be two nonempty convex sets in $\mathbb{R}^n$ such that $S_1 \cap S_2 = \emptyset$. Then there exist $p \in \mathbb{R}^n$, $p \neq 0$, and $\alpha \in \mathbb{R}$ such that*

$$\sup_{x \in S_1} p^t x \leq \alpha \leq \inf_{y \in S_2} p^t y.$$

In geometric terms, $H := \{x \in \mathbb{R}^n : p^t x = \alpha\}$ is said to be a separating hyperplane *for S_1 and S_2: each of the two convex sets is contained in precisely one of the two halfspaces $\{x \in \mathbb{R}^n : p^t x \leq \alpha\}$ and $\{x \in \mathbb{R}^n : p^t x \geq \alpha\}$.*

Proof. It is easy to see that $S := S_1 - S_2$ is convex. Now $0 \notin S$, for otherwise we get an immediate contradiction to $S_1 \cap S_2 = \emptyset$. W distinguish now two cases: (i) $0 \in \text{cl } S$ and (ii) $0 \notin \text{cl } S$.

In case (i) we have $0 \in \partial S$, so by Theorem A.3 we then have the existence of a nonzero $p \in \mathbb{R}^n$ such that

$$p^t z \leq 0 \text{ for every } z \in S = S_1 - S_2, \tag{2}$$

i.e., for every $z = x - y$, with $x \in S_1$ and $y \in S_2$. This gives $p^t x \leq p^t y$ for all $x \in S_1$ and $y \in S_2$, whence the result.

In case (ii) we apply Theorem A.2 to get immediately (2) as well. The result follows just as in case (i). QED

Theorem A.5 (strong set-set separation [1, Thm. 2.4.10]) *Let S_1, S_2 be two nonempty closed convex sets in $\mathbb{R}^n$ such that $S_1 \cap S_2 = \emptyset$ and such that S_1 is bounded. Then there exist $p \in \mathbb{R}^n$, $p \neq 0$, and $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$ such that*

$$\sup_{x \in S_1} p^t x \leq \alpha < \beta \leq \inf_{y \in S_2} p^t y.$$

Proof. As in the previous proof, it is easy to see that $S := S_1 - S_2$ is convex. Now S is also seen to be closed (exercise). As in the previous proof, we have $0 \notin S$. We can now apply Theorem A.2 to get the desired result, just as in case (ii) of the previous proof. QED

B Fenchel conjugation

Definition B.1 For a function $f : \mathbb{R}^n \to (-\infty, +\infty]$ the *(Fenchel) conjugate* function of f is $f^* : \mathbb{R}^n \to [-\infty, +\infty]$, given by

$$f^*(\xi) := \sup_{x \in \mathbb{R}^n} [\xi^t x - f(x)].$$

By repeating the conjugation operation one also defines the *(Fenchel) biconjugate* of f, which is simply given by $f^{**} := (f^*)^*$.

Example B.2 Consider $f : \mathbb{R} \to \mathbb{R}$, given by

$$f(x) := \begin{cases} x \log x & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ +\infty & \text{if } x < 0. \end{cases}$$

Observe that this function is convex. Then (counting $0 \log 0$ as 0) we clearly have $f^*(\xi) = \sup_{x \geq 0} \xi x - x \log x$ for the conjugate. For an *interior* maximum in $\mathbb{R}_+$ (by concavity of the function to be maximized) the necessary and sufficient condition is $\xi - \log x - 1 = 0$, i.e., $x = \exp(\xi - 1)$, which gives the value $\xi x - x \log x = \exp(\xi - 1)$. Since this value is positive, we conclude that the point $x = 0$ stands no chance for the maximum, i.e., the maximum is always interior, as calculated above, giving $f^*(\xi) = \exp(\xi - 1)$ for the conjugate function. We can also determine the biconjugate function: by definition, $f^{**}(x) = \sup_{\xi \in \mathbb{R}} x\xi - \exp(\xi - 1)$. If $x < 0$, then, by $\exp(\xi - 1) \to 0$ as $\xi \to -\infty$, the supremum value is clearly $+\infty$. Hence, $f^{**}(x) = +\infty$ for $x < 0$. If $x > 0$, then setting the derivative of the concave function $\xi \mapsto x\xi - \exp(\xi - 1)$ equal to zero gives a solution (whence a global maximum) for $\xi = \log x + 1$. Hence $f^{**}(x) = x \log x$ for $x > 0$. Finally, if $x = 0$, then the supremum of $-\exp(\xi - 1)$ is clearly the limit value 0. So $f^{**}(0) = 0$. We conclude that $f^{**} = f$ in this example. The Fenchel-Moreau theorem below will support this observation.

Exercise B.1 Determine for each of the following functions f the conjugate function f^* and verify also explicitly if $f = f^{**}$ holds.
a. $f(x) = ax^2 + bx + c$, $a \geq 0$,
b. $f(x) = |x| + |x - 1|$,
c. $f(x) = x^a/a$ for $x \geq 0$ and $f(x) = +\infty$ for $x < 0$ (here $a \geq 1$).
d. $f = \chi_B$, where B is the closed unit ball in $\mathbb{R}^n$.

Example B.3 Let K be a nonempty convex cone in $\mathbb{R}^n$ (recall that a *cone* (at zero) is a set such that $\alpha x \in K$ for every $\alpha > 0$ and $x \in K$; cf. Definition 2.5.1 in [1]). Let $f := \chi_K$. Then

$$f^*(\xi) = \sup_{x \in K} \xi^t x = \begin{cases} 0 & \text{if } \xi \in K^*, \\ +\infty & \text{otherwise.} \end{cases}$$

Recall here that K^*, the *polar cone* of K, is defined by $K^* := \{\xi \in \mathbb{R}^n : \xi^t x \leq 0 \text{ for all } x \in K\}$. Hence, we conclude that $(\chi_K)^* = \chi_{K^*}$.

Denote the closure of K by $\bar{K}$. We also observe that $\xi \in \partial \chi_{\bar{K}}(0)$ is equivalent to $\xi^t x \leq 0$ for all $x \in \bar{K}$, i.e., to $\xi^t x \leq 0$ for all $x \in K$, i.e., to $\xi \in K^*$.

Proposition B.4 *Let* $f,\ g : \mathbb{R}^n \to (-\infty, +\infty]$.

(*i*) *If* $f \geq g$ *then* $f^* \leq g^*$.

(*ii*) *If* $f^*(x) = -\infty$ *for some* $x \in \mathbb{R}^n$, *then* $f \equiv +\infty$.

(*iii*) *For every* $x_0,\ \xi \in \mathbb{R}^n$

$$f^*(\xi) \geq \xi^t x_0 - f(x_0) \text{ (Young's inequality).}$$

(*iv*) $f \geq f^{**}$.

(*v*) *For every* $x_0,\ \xi \in \mathbb{R}^n$

$$f^*(\xi) = \xi^t x_0 - f(x_0) \text{ if and only if } \xi \in \partial f(x_0).$$

Exercise B.2 Give a proof of Proposition B.4.

Theorem B.5 (Fenchel-Moreau) *Let* $f : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex. Then*

$$f(x_0) = f^{**}(x_0) \text{ if and only if } f \text{ is lower semicontinuous at } x_0.$$

Proof. One implication is very simple: if $f(x_0) = f^{**}(x_0)$, and if $x_n \to x_0$ then $\liminf_n f(x_n) \geq \liminf_n f^{**}(x_n)$ by Proposition B.4(*iv*). Also, $\liminf_n f^{**}(x_n) \geq f^{**}(x_0)$ because every conjugate, being the supremum of a collection of continuous functions, is automatically lower semicontinuous. So we conclude that $\liminf_n f(x_n) \geq f^{**}(x_0) = f(x_0)$, i.e., f is lower semicontinuous at x_0.

In the converse direction, by Proposition B.4(*iv*) it is enough to prove $f^{**}(x_0) \geq r$ for an arbitrary $r < f(x_0)$, both when $f(x_0) < +\infty$ and when $f(x_0) = +\infty$.

Case 1: $f(x_0) < +\infty$. It is easy to check that $C := \text{epi } f := \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : r \geq f(x)\}$, the epigraph of f, is a convex set in $\mathbb{R}^{n+1}$ (this is Theorem 3.2.2 in [1] – as can be seen immediately from its proof, it continues to hold for functions with values in $(-\infty, +\infty]$ and we know already that this theorem also holds for sets with empty interior). Hence, the closure cl C is also convex. We claim now that $(x_0, r) \notin \text{cl } C$. For suppose (x_0, r) would be the limit of a sequence of points $(x_n, y_n) \in C$. Then $y_n \geq f(x_n)$ for each n, and in the limit this would give $r \geq \liminf_n f(x_n) \geq f(x_0)$ by lower semicontinuity of f at x_0. This contradiction proves that the claim holds. We may now apply separation [1, Theorem 2.4.10]: there exist $\alpha \in \mathbb{R}$ and $p =: (\xi_0, \mu) \neq (0, 0)$, with $\xi_0 \in \mathbb{R}^n$ and $\mu \in \mathbb{R}$, such that

$$\xi_0^t x + \mu y \leq \alpha < \xi_0^t x_0 + \mu r \text{ for all } (x, y) \in C. \tag{3}$$

It is clear that $\mu \leq 0$ by the definition of C. Also, it is obvious that $\mu \neq 0$ (just consider what happens if we take $(x, y) = (x_0, f(x_0))$ in (3) – and we may do this by virtue of $f(x_0) \in \mathbb{R}$). Hence, we can divide by $-\mu$ in (3) and get

$$\xi_1^t x - f(x) \leq \xi_1^t x_0 - r \text{ for all } x \in \text{dom } f.$$

Notice that this inequality continues to hold outside dom f as well; thus, $f^*(\xi_1) \leq \xi_1^t x_0 - r$, which implies the desired inequality $f^{**}(x_0) \geq r$.

Case 2a: $f \equiv +\infty$. In this case, the desired result is trivial, for $f^* \equiv -\infty$, so $f^{**} \equiv +\infty$.

Case 2b: $f(x_1) < +\infty$ for some $x_1 \in \mathbb{R}^n$. We can repeat the proof of Case 1 until (3). If μ happens to be nonzero, then of course we finish as in Case 1. However, if $\mu = 0$ we only get

$$\xi_0^t x \leq \alpha < \xi_0^t x_0 \text{ for all } x \in \text{dom } f$$

from (3). We then repeat the full proof of Case 1, but with x_0 replaced by x_1 and r by $f(x_1) - 1$. This gives the existence of $\xi \in \mathbb{R}^n$ such that

$$\xi^t x - f(x) \leq \xi^t x_1 - f(x_1) + 1 \text{ for all } x \in \text{dom } f.$$

Now for any $\lambda > 0$, observe that by the two previous inequalities

$$f(x) \geq (\xi + \lambda\xi_0)^t x - \xi^t x_1 + f(x_1) - 1 - \alpha\lambda \text{ for all } x \in \mathbb{R}^n,$$

which implies $f^*(\xi + \lambda\xi_0) \leq \xi^t x_1 - f(x_1) + 1 + \lambda\alpha$. By definition of $f^{**}(x_0)$, this gives

$$f^{**}(x_0) \geq \lambda(\xi_0^t x_0 - \alpha) + \xi^t x_0 - \xi^t x_1 + f(x_1) - 1,$$

which implies $f^{**}(x_0) = +\infty$, by letting λ go to infinity (note that $\xi_0^t x_0 - \alpha > 0$ by the above). QED

Corollary B.6 (bipolar theorem for cones) *Let K be a closed convex cone in $\mathbb{R}^n$. Then $K = K^{**} := (K^*)^*$.*

Proof. Observe that $f := \chi_K$ is a lower semicontinuous convex function. Hence, $f^{**} = f$ by Theorem B.5. By Example B.3 we know that $f^* = \chi_{K^*}$, so $f^{**} = \chi_{K^{**}}$ follows by another application of this fact. Hence $\chi_K = \chi_{K^{**}}$. QED

Exercise B.3 Prove Farkas' theorem (see Exercise 3.5) by means of Corollary B.6.

Exercise B.4 Redo Exercise 3.3 by making it a special case of Corollary B.6.

References

[1] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms.* Wiley, New York, 1993.

[2] J. van Tiel, *Convex Analysis: An Introductory Text.* Wiley, 1984.

On subdifferential calculus *

Tieyong Zeng
zeng@math.cuhk.edu.hk
Feb, 2020

Definition 2.30 *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be a convex function and let* $\bar{x} \in \operatorname{dom} f$. *An element* $v \in \mathbb{R}^n$ *is called a* SUBGRADIENT *of* f *at* $\bar{x}$ *if*

$$\langle v, x - \bar{x} \rangle \leq f(x) - f(\bar{x}) \ \text{ for all } \ x \in \mathbb{R}^n . \tag{2.13}$$

The collection of all the subgradients of f *at* $\bar{x}$ *is called the* SUBDIFFERENTIAL *of the function at this point and is denoted by* $\partial f(\bar{x})$.

Subdifferential

the *subdifferential* $\partial f(x)$ of f at x is the set of all subgradients:

$$\partial f(x) = \{g \mid g^T(y - x) \leq f(y) - f(x), \ \forall y \in \mathbf{dom}\, f\}$$

Properties

- $\partial f(x)$ is a closed convex set (possibly empty)

  this follows from the definition: $\partial f(x)$ is an intersection of halfspaces

- if $x \in \mathbf{int}\,\mathbf{dom}\, f$ then $\partial f(x)$ is nonempty and bounded

  proof on next two pages

Proof: we show that $\partial f(x)$ is nonempty when $x \in \mathbf{int}\,\mathbf{dom}\, f$

- $(x, f(x))$ is in the boundary of the convex set $\mathbf{epi}\, f$

- therefore there exists a supporting hyperplane to $\mathbf{epi}\, f$ at $(x, f(x))$:

$$\exists (a, b) \neq 0, \qquad \begin{bmatrix} a \\ b \end{bmatrix}^T \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0 \qquad \forall (y, t) \in \mathbf{epi}\, f$$

- $b > 0$ gives a contradiction as $t \to \infty$

- $b = 0$ gives a contradiction for $y = x + \epsilon a$ with small $\epsilon > 0$

- therefore $b < 0$ and $g = \dfrac{1}{|b|} a$ is a subgradient of f at x

Proof: $\partial f(x)$ is bounded when $x \in \operatorname{int} \operatorname{dom} f$

- for small $r > 0$, define a set of $2n$ points

$$B = \{x \pm r e_k \mid k = 1, \ldots, n\} \subset \operatorname{dom} f$$

  and define $M = \max_{y \in B} f(y) < \infty$

- for every $g \in \partial f(x)$, there is a point $y \in B$ with

$$r \|g\|_\infty = g^T (y - x)$$

  (choose an index k with $|g_k| = \|g\|_\infty$, and take $y = x + r \operatorname{sign}(g_k) e_k$)

- since g is a subgradient, this implies that

$$f(x) + r\|g\|_\infty = f(x) + g^T (y - x) \leq f(y) \leq M$$

- we conclude that $\partial f(x)$ is bounded:

$$\|g\|_\infty \leq \frac{M - f(x)}{r} \quad \text{for all } g \in \partial f(x)$$

Definition 2.34 *We say that* $f : \mathbb{R}^n \to \mathbb{R}$ *is (Fréchet)* DIFFERENTIABLE *at* $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$ *if there exists an element* $v \in \mathbb{R}^n$ *such that*

$$\lim_{x \to \bar{x}} \frac{f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} = 0.$$

In this case the element v *is uniquely defined and is denoted by* $\nabla f(\bar{x}) := v$.

Proposition 2.35 *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be convex and let* $\bar{x} \in \operatorname{dom} f$*. Then* f *attains its local/global minimum at* $\bar{x}$ *if and only if* $0 \in \partial f(\bar{x})$.

Proof. Suppose that f attains its global minimum at $\bar{x}$. Then

$$f(\bar{x}) \le f(x) \quad \text{for all} \quad x \in \mathbb{R}^n,$$

which can be rewritten as

$$0 = \langle 0, x - \bar{x} \rangle \le f(x) - f(\bar{x}) \quad \text{for all} \quad x \in \mathbb{R}^n.$$

The definition of the subdifferential shows that this is equivalent to $0 \in \partial f(\bar{x})$. □

Now we show that the subdifferential (2.13) is indeed a singleton for differentiable functions reducing to the classical derivative/gradient at the reference point and clarifying the notion of differentiability in the case of convex functions.

Proposition 2.36 *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be convex and differentiable at* $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$. *Then we have* $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$ *and*

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \le f(x) - f(\bar{x}) \ \text{ for all } \ x \in \mathbb{R}^n . \tag{2.17}$$

Proposition 2.36 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex and differentiable at $\bar{x} \in \text{int}(\text{dom } f)$. Then we have $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$ and*

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \le f(x) - f(\bar{x}) \ \textit{for all} \ \ x \in \mathbb{R}^n . \tag{2.17}$$

Proof. It follows from the differentiability of f at $\bar{x}$ that for any $\epsilon > 0$ there is $\delta > 0$ with

$$-\epsilon\|x - \bar{x}\| \le f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle \le \epsilon\|x - \bar{x}\| \ \text{ whenever } \ \|x - \bar{x}\| < \delta. \tag{2.18}$$

Consider further the convex function

$$\varphi(x) := f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \epsilon\|x - \bar{x}\|, \quad x \in \mathbb{R}^n,$$

and observe that $\varphi(x) \ge \varphi(\bar{x}) = 0$ for all $x \in I\!B(\bar{x}; \delta)$. The convexity of φ ensures that $\varphi(x) \ge \varphi(\bar{x})$ for all $x \in \mathbb{R}^n$. Thus

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \le f(x) - f(\bar{x}) + \epsilon\|x - \bar{x}\| \ \text{ whenever } \ x \in \mathbb{R}^n,$$

which yields (2.17) by letting $\epsilon \downarrow 0$.

It follows from (2.17) that $\nabla f(\bar{x}) \in \partial f(\bar{x})$. Picking now $v \in \partial f(\bar{x})$, we get

$$\langle v, x - \bar{x} \rangle \le f(x) - f(\bar{x}).$$

Then the second part of (2.18) gives us that

$$\langle v - \nabla f(\bar{x}), x - \bar{x} \rangle \le \epsilon\|x - \bar{x}\| \ \text{ whenever } \ \|x - \bar{x}\| < \delta.$$

Finally, we observe that $\|v - \nabla f(\bar{x})\| \le \epsilon$, which yields $v = \nabla f(\bar{x})$ since $\epsilon > 0$ was chosen arbitrarily. Thus $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$. □

Example 2.38 Let $p(x) := \|x\|$ be the Euclidean norm function on $\mathbb{R}^n$. Then we have

$$\partial p(x) = \begin{cases} I\!B & \text{if } x = 0, \\ \left\{ \dfrac{x}{\|x\|} \right\} & \text{otherwise.} \end{cases}$$

To verify this, observe first that the Euclidean norm function p is differentiable at any nonzero point with $\nabla p(x) = \dfrac{x}{\|x\|}$ as $x \neq 0$. It remains to calculate its subdifferential at $x = 0$. To proceed by definition (2.13), we have that $v \in \partial p(0)$ if and only if

$$\langle v, x \rangle = \langle v, x - 0 \rangle \le p(x) - p(0) = \|x\| \text{ for all } x \in \mathbb{R}^n .$$

Letting $x = v$ gives us $\langle v, v \rangle \le \|v\|$, which implies that $\|v\| \le 1$, i.e., $v \in I\!B$. Now take $v \in I\!B$ and deduce from the Cauchy-Schwarz inequality that

$$\langle v, x - 0 \rangle = \langle v, x \rangle \le \|v\| \cdot \|x\| \le \|x\| = p(x) - p(0) \text{ for all } x \in \mathbb{R}^n$$

and thus $v \in \partial p(0)$, which shows that $\partial p(0) = I\!B$.

Theorem 2.40 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a differentiable function on its domain D, which is an open convex set. Then f is convex if and only if

$$\langle \nabla f(u), x - u \rangle \le f(x) - f(u) \quad \text{for all } \; x, u \in D. \tag{2.21}$$

Proof. The "only if" part follows from Proposition 2.36. To justify the converse, suppose that (2.21) holds and then fix any $x_1, x_2 \in D$ and $t \in (0, 1)$. Denoting $x_t := tx_1 + (1 - t)x_2$, we have $x_t \in D$ by the convexity of D. Then

$$\langle \nabla f(x_t), x_1 - x_t \rangle \le f(x_1) - f(x_t), \quad \langle \nabla f(x_t), x_2 - x_t \rangle \le f(x_2) - f(x_t).$$

It follows furthermore that

$$\begin{aligned} &t\langle \nabla f(x_t), x_1 - x_t \rangle \le tf(x_1) - tf(x_t) \;\text{ and} \\ &(1-t)\langle \nabla f(x_t), x_2 - x_t \rangle \le (1-t)f(x_2) - (1-t)f(x_t). \end{aligned}$$

Summing up these inequalities, we arrive at

$$0 \le tf(x_1) + (1-t)f(x_2) - f(x_t),$$

which ensures that $f(x_t) \le tf(x_1) + (1-t)f(x_2)$, and so verifies the convexity of f. □

Moreau-Rockafellar theorem

Corollary 2.45 *Let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ for $i = 1, 2$ be convex functions such that there exists $u \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$ for which f_1 is continuous at u or f_2 is continuous at u. Then*

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x) \tag{2.28}$$

whenever $x \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$. Consequently, if both functions f_i are finite-valued on $\mathbb{R}^n$, then the sum rule (2.28) holds for all $x \in \mathbb{R}^n$.

Theorem 2.9 (Moreau-Rockafellar) *Let* $f, g : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions. Then for every* $x_0 \in \mathbb{R}^n$

$$\partial f(x_0) + \partial g(x_0) \subset \partial(f+g)(x_0).$$

Moreover, suppose that $\operatorname{int}\operatorname{dom} f \cap \operatorname{dom} g \neq \emptyset$. *Then for every* $x_0 \in \mathbb{R}^n$ *also*

$$\partial(f+g)(x_0) \subset \partial f(x_0) + \partial g(x_0).$$

PROOF. The proof of the first part is elementary: Let $\xi_1 \in \partial f(x_0)$ and $\xi_2 \in \partial g(x_0)$. Then for all $x \in \mathbb{R}^n$

$$f(x) \geq f(x_0) + \xi_1^t(x - x_0), \; g(x) \geq g(x_0) + \xi_2^t(x - x_0),$$

so addition gives $f(x) + g(x) \geq f(x_0) + g(x_0) + (\xi_1 + \xi_2)^t(x - x_0)$. Hence $\xi_1 + \xi_2 \in \partial(f + g)(x_0)$.

To prove the second part, let $\xi \in \partial(f + g)(x_0)$. First, observe that $f(x_0) = +\infty$ implies $(f + g)(x_0) = +\infty$, whence $f + g \equiv +\infty$, which is impossible by $\xi \in \partial(f + g)(x_0)$. Likewise, $g(x_0) = +\infty$ is impossible. Hence, from now on we know that both $f(x_0)$ and $g(x_0)$ belong to $\mathbb{R}$. We form the following two sets in $\mathbb{R}^{n+1}$.

$$\Lambda_f := \{(x - x_0, y) \in \mathbb{R}^n \times \mathbb{R} : y > f(x) - f(x_0) - \xi^t(x - x_0)\}$$

$$\Lambda_g := \{(x - x_0, y) : -y \geq g(x) - g(x_0)\}.$$

$$\Lambda_f := \{(x - x_0, y) \in \mathbb{R}^n \times \mathbb{R} : y > f(x) - f(x_0) - \xi^t(x - x_0)\}$$

$$\Lambda_g := \{(x - x_0, y) : -y \geq g(x) - g(x_0)\}.$$

Observe that both sets are nonempty and convex (see Exercise 2.8), and that $\Lambda_f \cap \Lambda_g = \emptyset$ (the latter follows from $\xi \in \partial(f + g)(x_0)$). Hence, by the set-set-separation Theorem A.4, there exists $(\xi_0, \mu) \in \mathbb{R}^{n+1}$ and $\alpha \in \mathbb{R}$, $(\xi_0, \mu) \neq (0, 0)$, such that

$$\xi_0^t(x - x_0) + \mu y \leq \alpha \text{ for all } (x, y) \text{ with } y > f(x) - f(x_0) - \xi^t(x - x_0),$$

$$\xi_0^t(x - x_0) + \mu y \geq \alpha \text{ for all } (x, y) \text{ with } -y \geq g(x) - g(x_0).$$

By $(0, 0) \in \Lambda_g$ we get $\alpha \leq 0$. But also $(0, \epsilon) \in \Lambda_f$ for every $\epsilon > 0$, and this gives $\mu\epsilon \leq \alpha$, so $\mu \leq 0$ (take $\epsilon = 1$). In the limit, for $\epsilon \to 0$, we find $\alpha \geq 0$. Hence $\alpha = 0$ and $\mu \leq 0$. We now claim that $\mu = 0$ is impossible. Indeed, if one had $\mu = 0$, then the first of the above two inequalities would give

$$\xi_0^t(x - x_0) \leq 0 \text{ for all } (x, y) \text{ with } y > f(x) - f(x_0) - \xi^t(x - x_0),$$

which is equivalent to

$$\xi_0^t(x - x_0) \leq 0 \text{ for all } x \in \text{dom } f$$

(simply note that when $f(x) < +\infty$ one can always achieve $y > f(x) - f(x_0) - \xi^t(x - x_0)$ by choosing y sufficiently large). Likewise, the second inequality would give

$$\xi_0^t(x - x_0) \geq 0 \text{ for all } x \in \text{dom } g.$$

In particular, for $\tilde{x}$ as above this would imply $\xi_0^t(\tilde{x} - x_0) = 0$. But since $\tilde{x}$ lies in the interior of dom f (so for some $\delta > 0$ the ball $N_\delta(\tilde{x})$ belongs to dom f), the preceding would imply

$$\xi_0^t u = \xi_0^t(\tilde{x} + u - x_0) \leq 0 \text{ for all } u \in N_\delta(0).$$

Clearly, this would give $\xi_0 = 0$ (take $u := \delta\xi_0/2$), which would be in contradiction to $(\xi_0, \mu) \neq (0, 0)$. Hence, we conclude $\mu < 0$. Dividing the separation inequalities by $-\mu$ and setting $\bar{\xi}_0 := -\xi_0/\mu$, this results in

$$\bar{\xi}_0^t(x - x_0) \leq y \text{ for all } (x, y) \text{ with } y > f(x) - f(x_0) - \xi^t(x - x_0),$$

$$\bar{\xi}_0^t(x - x_0) \geq y \text{ for all } (x, y) \text{ with } -y \geq g(x) - g(x_0).$$

The last inequality gives $-\bar{\xi}_0 \in \partial g(x_0)$ (set $y := g(x_0) - g(x)$) and the one but last inequality gives $\xi + \bar{\xi}_0 \in \partial f(x_0)$ (take $y := f(x) - f(x_0) - \xi^t(x - x_0) + \epsilon$ and let $\epsilon \downarrow 0$). Since $\xi = (\xi + \bar{\xi}_0) - \bar{\xi}_0$, this finishes the proof. QED

As a precursor to the Karush-Kuhn-Tucker theorem, we have now the following application of the Moreau-Rockafellar theorem.

Theorem 2.10 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $S \subset \mathbb{R}^n$ be a nonempty convex set. Consider the optimization problem*

$$(P) \quad \inf_{x \in S} f(x).$$

Then $\bar{x} \in S$ is an optimal solution of (P) if and only if there exists a subgradient $\bar{\xi} \in \partial f(\bar{x})$ such that

$$\bar{\xi}^t (x - \bar{x}) \geq 0 \text{ for all } x \in S. \tag{1}$$

Proof. Recall from Definition 2.3 that χ_S is the indicator function of S. Now let $\bar{x} \in S$ be arbitrary. Then the following is trivial: $\bar{x}$ is an optimal solution of (P) if and only if

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

By the Moreau-Rockafellar Theorem 2.9, we have

$$\partial(f + \chi_S)(\bar{x}) = \partial f(\bar{x}) + \partial \chi_S(\bar{x}).$$

To see that its conditions hold, observe that $\operatorname{dom} f = \mathbb{R}^n$ and $\operatorname{dom} \chi_S = S$. So it follows that $\bar{x}$ is an optimal solution of (P) if and only if $0 \in \partial f(\bar{x}) + \partial \chi_S(\bar{x})$. By the definition of the sum of two sets this means that $\bar{x}$ is an optimal solution of (P) if and only if $0 = \bar{\xi} + \bar{\xi}'$ for some $\bar{\xi} \in \partial f(\bar{x})$ and $\bar{\xi}' \in \partial \chi_S(\bar{x})$. Of course, the former means $\bar{\xi}' = -\bar{\xi}$, so $-\bar{\xi} \in \partial \chi_S(\bar{x})$, which is equivalent to

$$\chi_S(x) \geq \chi_S(\bar{x}) + (-\bar{\xi})^t (x - \bar{x}) \text{ for all } x \in \mathbb{R}^n,$$

i.e., to (1). QED

Definition 2.13 The *directional derivative* of a convex function $f : \mathbb{R}^n \to (-\infty, +\infty]$ at the point $x_0 \in \text{dom} f$ in the direction $d \in \mathbb{R}^n$ is defined as

$$f'(x_0; d) := \lim_{\lambda \downarrow 0} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}.$$

The above limit is a well-defined number in $[-\infty, +\infty]$. This follows from the following proposition (why?), which shows that the difference quotients of a convex functions possess a monotonicity property:

Proposition 2.14 *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex function and let x_0 be a point in* $\mathrm{dom} f$*. Then for every direction $d \in \mathbb{R}^n$ and every $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $\lambda_2 > \lambda_1 > 0$ we have*

$$\frac{f(x_0 + \lambda_1 d) - f(x_0)}{\lambda_1} \leq \frac{f(x_0 + \lambda_2 d) - f(x_0)}{\lambda_2}$$

PROOF. Note that

$$x_0 + \lambda_1 d = \frac{\lambda_1}{\lambda_2}(x_0 + \lambda_2 d) + (1 - \frac{\lambda_1}{\lambda_2})x_0.$$

So by convexity of f

$$f(x_0 + \lambda_1 d) \leq \frac{\lambda_1}{\lambda_2} f(x_0 + \lambda_2 d) + (1 - \frac{\lambda_1}{\lambda_2}) f(x_0).$$

Simple algebra shows that this is equivalent to the desired inequality. QED

Theorem 2.15 *Let* $f : \mathbb{R}^n \to (-\infty, +\infty]$ *be a convex function and let* x_0 *be a point in* int dom f. *Then*

$$f'(x_0; d) = \sup_{\xi \in \partial f(x_0)} \xi^t d \text{ for every } d \in \mathbb{R}^n.$$

PROOF OF THEOREM 2.15. By Proposition 2.14

$$q(d) := f'(x_0; d) := \lim_{\lambda \downarrow 0} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda} = \inf_{\lambda > 0} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}.$$

Since the pointwise limit of a sequence of convex functions is convex, it follows that $q : \mathbb{R}^n \to \mathbb{R}$ is convex (by the infimum expression for $q(d)$ the fact that $x_0 \in \text{int dom } f$ implies automatically $q(d) < +\infty$ for every d; also, $q(d) > -\infty$ for every d, because of the nonemptiness part of Lemma 2.16). Hence, q is continuous at every point $d \in \mathbb{R}^n$ (apply the continuity part of Lemma 2.16). So by the Fenchel-Moreau theorem (Theorem B.5 in the Appendix) we have for every d

$$q(d) = q^{**}(d) := \sup_{\xi \in \mathbb{R}^n} [d^t \xi - q^*(\xi)].$$

Let us calculate q^*. For any $\xi \in \mathbb{R}^n$ we have

$$q^*(\xi) := \sup_{d \in \mathbb{R}^n} [\xi^t d - q(d)] = \sup_{d, \lambda > 0} [\xi^t d - \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}] = \sup_{\lambda > 0} \sup_{d} [\xi^t d - \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}]$$

by the above infimum expression for $q(d)$. Fix $\lambda > 0$; then $z := x_0 + \lambda d$ runs through all of $\mathbb{R}^n$ as d runs through $\mathbb{R}^n$. Hence

$$\sup_d [\xi^t d - \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}] = \frac{f(x_0) - \xi^t x_0 + \sup_z [\xi^t z - f(z)]}{\lambda}.$$

Clearly, this gives

$$q^*(\xi) = \sup_{\lambda > 0} \frac{f(x_0) - \xi^t x_0 + f^*(\xi)}{\lambda} = \begin{cases} 0 & \text{if } \xi \in \partial f(x_0) \\ +\infty & \text{otherwise} \end{cases}$$

where we use Proposition B.4(v). Observe that in terms of the indicator function of the subdifferential this can be rewritten as $q^* = \chi_{\partial f(x_0)}$. Now that q^* has been calculated, we conclude from the above that for every $d \in \mathbb{R}^n$

$$f'(x_0; d) = q(d) = q^{**}(d) = \chi^*_{\partial f(x_0)}(d) = \sup_{\xi \in \partial f(x_0)} \xi^t d,$$

which proves the result. QED

Proposition 2.54 *Let* $f_i\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 1, \ldots, m$, *be convex functions. Take any point* $\bar{x} \in \bigcap_{i=1}^m \operatorname{dom} f_i$ *and assume that each* f_i *is continuous at* $\bar{x}$*. Then we have the maximum rule*

$$\partial\big(\max f_i\big)(\bar{x}) = \operatorname{co} \bigcup_{i \in I(\bar{x})} \partial f_i(\bar{x}).$$

Theorem 2.17 (Dubovitskii-Milyutin) *Let* $f_1, \cdots, f_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions and let* x_0 *be a point in* $\cap_{i=1}^m \text{int dom } f_i$*. Let* $f : \mathbb{R}^n \to (-\infty, +\infty]$ *be given by*

$$f(x) := \max_{1 \le i \le m} f_i(x)$$

and let $I(x_0)$ *be the (nonempty) set of all* $i \in \{1, \cdots, m\}$ *for which* $f_i(x_0) = f(x_0)$*. Then*

$$\partial f(x_0) = \text{co } \cup_{i \in I(x_0)} \partial f_i(x_0).$$

PROOF. For our convenience we write $I := I(x_0)$. To begin with, observe that $\xi \in \partial f_i(x_0)$ easily implies $\xi \in \partial f(x_0)$ for each $i \in I$. Since $\partial f(x_0)$ is evidently convex, the inclusion "$\supset$" follows with ease. To prove the opposite inclusion, let ξ_0 be arbitrary in $\partial f(x_0)$. If ξ_0 were not to belong to the compact set co $\cup_{i \in I} \partial f_i(x_0)$, then we could separate strictly (note that each set $\partial f_i(x_0)$ is both closed and compact (exercise)): by Theorem A.2 there would exist $d \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ such that

$$\xi_0^t d > \alpha \geq \max_{i \in I} \sup_{\xi \in \partial f_i(x_0)} \xi^t d = \max_{i \in I} f_i'(x_0; d),$$

where the final identity follows from Theorem 2.15. But now observe that

$$f'(x_0; d) := \lim_{\lambda \downarrow 0} \max_{i \in I} \frac{f_i(x_0 + \lambda d) - f_i(x_0)}{\lambda} = \max_{i \in I} \lim_{\lambda \downarrow 0} \frac{f_i(x_0 + \lambda d) - f_i(x_0)}{\lambda} = \max_{i \in I} f_i'(x_0; d),$$

so the above gives $\xi_0^t d > f'(x_0; d)$. On the other hand, by $\xi_0 \in \partial f(x_0)$ it follows that $f(x_0 + \lambda d) \geq f(x_0) + \lambda \xi_0^t d$ for every $\lambda > 0$, whence $f'(x_0; d) \geq \xi_0^t d$. We thus have arrived at a contradiction. So the inclusion "$\subset$" must hold as well. QED

Directional derivative

Definition (for general f): the *directional derivative* of f at x in the direction y is

$$
\begin{aligned}
f'(x;y) &= \lim_{\alpha \searrow 0} \frac{f(x+\alpha y) - f(x)}{\alpha} \\
&= \lim_{t\to\infty} \left(t(f(x + \frac{1}{t}y) - t f(x) \right)
\end{aligned}
$$

(if the limit exists)

- $f'(x;y)$ is the right derivative of $g(\alpha) = f(x+\alpha y)$ at $\alpha = 0$
- $f'(x;y)$ is homogeneous in y:

$$f'(x;\lambda y) = \lambda f'(x;y) \quad \text{for } \lambda \geq 0$$

Directional derivative of a convex function

Equivalent definition (for convex f): replace $\lim$ with $\inf$

$$f'(x;y) = \inf_{\alpha>0} \frac{f(x+\alpha y) - f(x)}{\alpha}$$
$$= \inf_{t>0} \left(t f(x + \frac{1}{t} y) - t f(x) \right)$$

Proof

- the function $h(y) = f(x+y) - f(x)$ is convex in y, with $h(0) = 0$
- its perspective $th(y/t)$ is nonincreasing in t (ECE236B ex. A2.5); hence

$$f'(x;y) = \lim_{t\to\infty} th(y/t) = \inf_{t>0} th(y/t)$$

Properties

consequences of the expressions (for convex f)

$$
\begin{aligned}
f'(x; y) &= \inf_{\alpha>0} \frac{f(x+\alpha y) - f(x)}{\alpha} \\
&= \inf_{t>0} \left(t f(x + \frac{1}{t} y) - t f(x) \right)
\end{aligned}
$$

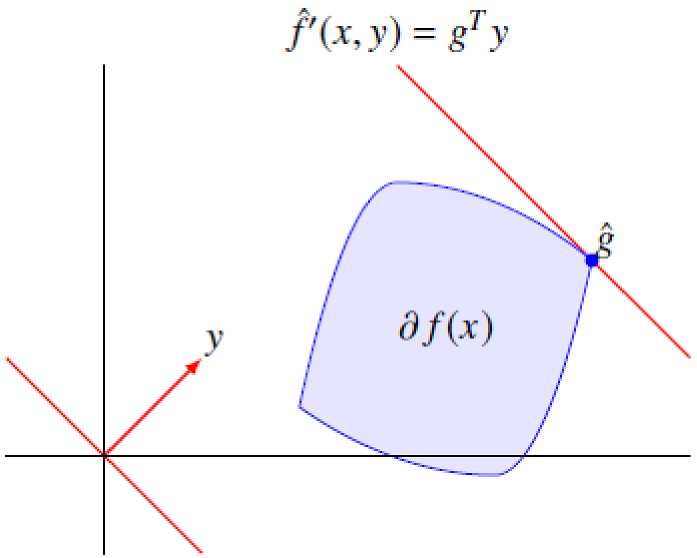
- $f'(x; y)$ is convex in y (partial minimization of a convex function in y, t)
- $f'(x; y)$ defines a lower bound on f in the direction y:

$$f(x+\alpha y) \geq f(x) + \alpha f'(x; y) \quad \text{for all } \alpha \geq 0$$

Directional derivative and subgradients

for convex f and $x \in \operatorname{int} \operatorname{dom} f$

$$f'(x; y) = \sup_{g \in \partial f(x)} g^T y$$



$f'(x; y)$ is *support function* of $\partial f(x)$

- generalizes $f'(x; y) = \nabla f(x)^T y$ for differentiable functions
- implies that $f'(x; y)$ exists for all $x \in \operatorname{int} \operatorname{dom} f$, all y (see page 2.4)

Proof: if $g \in \partial f(x)$ then from page 2.29

$$f'(x; y) \geq \inf_{\alpha>0} \frac{f(x) + \alpha g^T y - f(x)}{\alpha} = g^T y$$

it remains to show that $f'(x; y) = \hat{g}^T y$ for at least one $\hat{g} \in \partial f(x)$

- $f'(x; y)$ is convex in y with domain $\mathbf{R}^n$, hence subdifferentiable at all y
- let $\hat{g}$ be a subgradient of $f'(x; y)$ at y: then for all v, $\lambda \geq 0$,

$$\lambda f'(x; v) = f'(x; \lambda v) \geq f'(x; y) + \hat{g}^T(\lambda v - y)$$

- taking $\lambda \to \infty$ shows that $f'(x; v) \geq \hat{g}^T v$; from the lower bound on page 2.30,

$$f(x + v) \geq f(x) + f'(x; v) \geq f(x) + \hat{g}^T v \quad \text{for all } v$$

  hence $\hat{g} \in \partial f(x)$

- taking $\lambda = 0$ we see that $f'(x; y) \leq \hat{g}^T y$

Proof. Taking $\Omega_1 \subset \Omega_2$ as in the proposition, for any $v \in \mathbb{R}^n$ we have

$$\sigma_{\Omega_1}(v) = \sup\{\langle v, x\rangle \mid x \in \Omega_1\} \le \sup\{\langle v, x\rangle \mid x \in \Omega_2\} = \sigma_{\Omega_2}(v).$$

Conversely, suppose that $\sigma_{\Omega_1}(v) \le \sigma_{\Omega_2}(v)$ whenever $v \in \mathbb{R}^n$. Since $\sigma_{\Omega_1}(0) = \sigma_{\Omega_2}(0) = 0$, it follows from definition (2.13) of the subdifferential that

$$\partial\sigma_{\Omega_1}(0) \subset \partial\sigma_{\Omega_2}(0),$$

which yields $\Omega_1 \subset \Omega_2$ by formula (2.44). □

2.8 FENCHEL CONJUGATES

Many important issues of convex analysis and its applications (in particular, to optimization) are based on *duality*. The following notion plays a crucial role in duality considerations.

Definition 2.70 *Given a function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ (*not necessarily convex*), *its* FENCHEL CONJUGATE $f^* : \mathbb{R}^n \to [-\infty, \infty]$ *is*

$$f^*(v) := \sup\{\langle v, x\rangle - f(x) \mid x \in \mathbb{R}^n\} = \sup\{\langle v, x\rangle - f(x) \mid x \in \operatorname{dom} f\}. \tag{2.45}$$

Note that $f^*(v) = -\infty$ is allowed in (2.45) while $f^*(v) > -\infty$ for all $v \in \mathbb{R}^n$ if $\operatorname{dom} f \neq \emptyset$. It follows directly from the definitions that for any nonempty set $\Omega \subset \mathbb{R}^n$ the conjugate to its indicator function is the support function of Ω:

$$\delta^*_\Omega(v) = \sup\{\langle v, x\rangle \mid x \in \Omega\} = \sigma_\Omega(v), \quad v \in \Omega. \tag{2.46}$$

The next two propositions can be easily verified.

Proposition 2.71 *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be a function, not necessarily convex, with* $\operatorname{dom} f \neq \emptyset$. *Then its Fenchel conjugate* f^* *is convex on* $\mathbb{R}^n$.

Proof. Function (2.45) is convex as the supremum of a family of affine functions. □

Proposition 2.72 *Let* $f, g : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be such that* $f(x) \le g(x)$ *for all* $x \in \mathbb{R}^n$. *Then we have* $f^*(v) \ge g^*(v)$ *for all* $v \in \mathbb{R}^n$.

Proof. For any fixed $v \in \mathbb{R}^n$, it follows from (2.45) that

$$\langle v, x\rangle - f(x) \ge \langle v, x\rangle - g(x), \quad x \in \mathbb{R}^n.$$

This readily implies the relationships

$$f^*(v) = \sup\{\langle v, x\rangle - f(x) \mid x \in \mathbb{R}^n\} \ge \sup\{\langle v, x\rangle - g(x) \mid x \in \mathbb{R}^n\} = g^*(v)$$

for all $v \in \mathbb{R}^n$, and therefore $f^* \ge g^*$ on $\mathbb{R}^n$. □

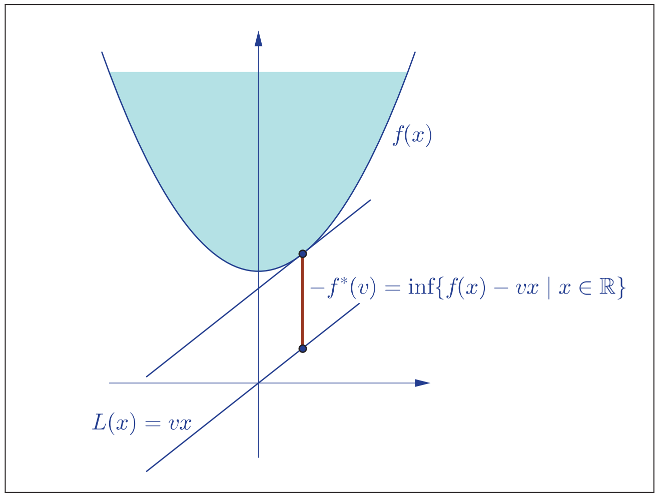


Figure 2.8: Fenchel conjugate.

The following two examples illustrate the calculation of conjugate functions.

Example 2.73 **(i)** Given $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$, consider the affine function

$$f(x) := \langle a, x \rangle + b, \quad x \in \mathbb{R}^n .$$

Then it can be seen directly from the definition that

$$f^*(v) = \begin{cases} -b & \text{if } v = a, \\ \infty & \text{otherwise.} \end{cases}$$

(ii) Given any $p > 1$, consider the power function

$$f(x) := \begin{cases} \dfrac{x^p}{p} & \text{if } x \geq 0, \\ \infty & \text{otherwise.} \end{cases}$$

For any $v \in \mathbb{R}$, the conjugate of this function is given by

$$f^*(v) = \sup\Big\{vx - \frac{x^p}{p} \;\Big|\; x \geq 0\Big\} = -\inf\Big\{\frac{x^p}{p} - vx \;\Big|\; x \geq 0\Big\}.$$

It is clear that $f^*(v) = 0$ if $v \leq 0$ since in this case $vx - p^{-1}x^p \leq 0$ when $x \geq 0$. Considering the case of $v > 0$, we see that function $\psi_v(x) := p^{-1}x^p - vx$ is convex and differentiable on

$(0, \infty)$ with $\psi_v'(x) = x^{p-1} - v$. Thus $\psi_v'(x) = 0$ if and only if $x = v^{1/(p-1)}$, and so ψ_v attains its minimum at $x = v^{1/(p-1)}$. Therefore, the conjugate function is calculated by

$$f^*(v) = \Big(1 - \frac{1}{p}\Big) v^{p/(p-1)}, \quad v \in \mathbb{R}^n.$$

Taking q from $q^{-1} = 1 - p^{-1}$, we express the conjugate function as

$$f^*(v) = \begin{cases} 0 & \text{if } v \le 0, \\ \dfrac{v^q}{q} & \text{otherwise.} \end{cases}$$

Note that the calculation in Example 2.73(ii) shows that

$$vx \le \frac{x^p}{p} + \frac{v^q}{q} \quad \text{for any } \ x, v \ge 0.$$

The first assertion of the next proposition demonstrates that such a relationship in a more general setting. To formulate the second assertion below, we define the *biconjugate* of f as the conjugate of f^*, i.e., $f^{**}(x) := (f^*)^*(x)$.

Proposition 2.74 *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be a function with* $\operatorname{dom} f \neq \emptyset$. *Then we have:*

(i) $\langle v, x\rangle \le f(x) + f^*(v)$ *for all* $x, v \in \mathbb{R}^n$.
(ii) $f^{**}(x) \le f(x)$ *for all* $x \in \mathbb{R}^n$.

Proof. Observe first that **(i)** is obvious if $f(x) = \infty$. If $x \in \operatorname{dom} f$, we get from (2.45) that $f^*(v) \ge \langle v, x\rangle - f(x)$, which verifies **(i)**. It implies in turn that

$$\sup\big\{\langle v, x\rangle - f^*(v) \ \big|\ v \in \mathbb{R}^n\big\} \le f(x) \ \text{ for all } \ x, v \in \mathbb{R}^n,$$

which thus verifies **(ii)** and completes the proof. □

The following important result reveals a close relationship between subgradients and Fenchel conjugates of convex functions.

Theorem 2.75 For any convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and any $\bar{x} \in \operatorname{dom} f$, we have that $v \in \partial f(\bar{x})$ if and only if

$$f(\bar{x}) + f^*(v) = \langle v, \bar{x}\rangle. \tag{2.47}$$

Proof. Taking any $v \in \partial f(\bar{x})$ and using definition (2.13) gives us

$$f(\bar{x}) + \langle v, x \rangle - f(x) \leq \langle v, \bar{x} \rangle \quad \text{for all} \quad x \in \mathbb{R}^n .$$

This readily implies the inequality

$$f(\bar{x}) + f^*(v) = f(\bar{x}) + \sup \big\{ \langle v, x \rangle - f(x) \; \big| \; x \in \mathbb{R}^n \big\} \leq \langle v, \bar{x} \rangle.$$

Since the opposite inequality holds by Proposition 2.74(i), we arrive at (2.47).

Conversely, suppose that $f(\bar{x}) + f^*(v) = \langle v, \bar{x} \rangle$. Applying Proposition 2.74(i), we get the estimate $f^*(v) \geq \langle v, x \rangle - f(x)$ for every $x \in \mathbb{R}^n$. This shows that $v \in \partial f(\bar{x})$. □

The result obtained allows us to find conditions ensuring that the biconjugate f^{**} of a convex function agrees with the function itself.

Proposition 2.76 *Let $\bar{x} \in \text{dom}\, f$ for a convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. Suppose that $\partial f(\bar{x}) \neq \emptyset$. Then we have the equality $f^{**}(\bar{x}) = f(\bar{x})$.*

Proof. By Proposition 2.74(ii) it suffices to verify the opposite inequality therein. Fix $v \in \partial f(\bar{x})$ and get $\langle v, \bar{x} \rangle = f(\bar{x}) + f^*(v)$ by the preceding theorem. This shows that

$$f(\bar{x}) = \langle v, \bar{x} \rangle - f^*(v) \leq \sup \big\{ \langle \bar{x}, v \rangle - f^*(v) \; \big| \; v \in \mathbb{R}^n \big\} = f^{**}(\bar{x}),$$

which completes the proof of this proposition. □

Taking into account Proposition 2.47, we arrive at the following corollary.

Corollary 2.77 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function and let $\bar{x} \in \text{int}(\text{dom}\, f)$. Then we have the equality $f^{**}(\bar{x}) = f(\bar{x})$.*

Finally in this section, we prove a necessary and sufficient condition for the validity of the biconjugacy equality $f = f^{**}$ known as the *Fenchel-Moreau theorem*.

Theorem 2.78 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function with $\text{dom}\, f \neq \emptyset$ and let $\mathcal{A}$ be the set of all affine functions of the form $\varphi(x) = \langle a, x \rangle + b$ for $x \in \mathbb{R}^n$, where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Denote

$$\mathcal{A}(f) := \big\{ \varphi \in \mathcal{A} \; \big| \; \varphi(x) \leq f(x) \quad \text{for all} \quad x \in \mathbb{R}^n \big\}.$$

Then $\mathcal{A}(f) \neq \emptyset$ whenever $\text{epi}\, f$ is closed and convex. Moreover, the following are equivalent:

(i) $\text{epi}\, f$ is closed and convex.
(ii) $f(x) = \sup_{\varphi \in \mathcal{A}(f)} \varphi(x)$ for all $x \in \mathbb{R}^n$.
(iii) $f^{**}(x) = f(x)$ for all $x \in \mathbb{R}^n$.

Proof. Let us first show that $\mathcal{A}(f) \neq \emptyset$. Fix any $x_0 \in \text{dom}\, f$ and choose $\lambda_0 < f(x_0)$. Then $(x_0, \lambda_0) \notin \text{epi}\, f$. By Proposition 2.1, there exist $(\bar{v}, \bar{\gamma}) \in \mathbb{R}^n \times \mathbb{R}$ and $\epsilon > 0$ such that

$$\langle \bar{v}, x \rangle + \bar{\gamma}\lambda < \langle \bar{v}, x_0 \rangle + \bar{\gamma}\lambda_0 - \epsilon \ \text{ whenever } \ (x, \lambda) \in \text{epi}\, f. \tag{2.48}$$

Since $(x_0, f(x_0) + \alpha) \in \text{epi}\, f$ for all $\alpha \geq 0$, we get

$$\bar{\gamma}(f(\bar{x}) + \alpha) < \bar{\gamma}\lambda_0 - \epsilon \ \text{ whenever } \ \alpha \geq 0.$$

This implies $\bar{\gamma} < 0$ since if not, we can let $\alpha \to \infty$ and arrive at a contradiction. For any $x \in \text{dom}\, f$, it follows from (2.48) as $(x, f(x)) \in \text{epi}\, f$ that

$$\langle \bar{v}, x \rangle + \bar{\gamma} f(x) < \langle \bar{v}, x_0 \rangle + \bar{\gamma}\lambda_0 - \epsilon \ \text{ for all } \ x \in \text{dom}\, f.$$

This allows us to conclude that

$$f(x) > \langle \frac{\bar{v}}{\bar{\gamma}}, x_0 - x \rangle + \lambda_0 - \frac{\epsilon}{\bar{\gamma}} \ \text{ if } \ x \in \text{dom}\, f.$$

Define now $\varphi(x) := \langle \frac{\bar{v}}{\bar{\gamma}}, x_0 - x \rangle + \lambda_0 - \frac{\epsilon}{\bar{\gamma}}$. Then $\varphi \in \mathcal{A}(f)$, and so $\mathcal{A}(f) \neq \emptyset$.

Let us next prove that **(i)**$\Longrightarrow$**(ii)**. By definition we need to show that for any $\lambda_0 < f(x_0)$ there is $\varphi \in \mathcal{A}(f)$ such that $\lambda_0 < \varphi(x_0)$. Since $(x_0, \lambda_0) \notin \text{epi}\, f$, we apply again Proposition 2.1 to obtain (2.48). In the case where $x_0 \in \text{dom}\, f$, it was proved above that $\varphi \in \mathcal{A}(f)$. Moreover, we have $\varphi(x_0) = \lambda_0 - \frac{\epsilon}{\bar{\gamma}} < \lambda_0$ since $\bar{\gamma} < 0$. Consider now the case where $x_0 \notin \text{dom}\, f$. It follows from (2.48) by taking any $x \in \text{dom}\, f$ and letting $\lambda \to \infty$ that $\bar{\gamma} \leq 0$. If $\bar{\gamma} < 0$, we can apply the same procedure and arrive at the conclusion. Hence we only need to consider the case where $\bar{\gamma} = 0$. In this case

$$\langle \bar{v}, x - x_0 \rangle + \epsilon < 0 \ \text{ whenever } \ x \in \text{dom}\, f.$$

Since $\mathcal{A}(f) \neq \emptyset$, choose $\varphi_0 \in \mathcal{A}(f)$ and define

$$\varphi_k(x) := \varphi_0(x) + k(\langle \bar{v}, x - x_0 \rangle + \epsilon), \ k \in \mathbb{N}\,.$$

It is obvious that $\varphi_k \in \mathcal{A}(f)$ and $\varphi_k(x_0) = \varphi_0(x_0) + k\epsilon > \lambda_0$ for large k. This justifies **(ii)**.

Let us now verify implication **(ii)**$\Longrightarrow$**(iii)**. Fix any $\varphi \in \mathcal{A}(f)$. Then $\varphi \leq f$, and hence $\varphi^{**} \leq f^{**}$. Applying Proposition 2.76 ensures that $\varphi = \varphi^{**} \leq f^{**}$. It follows therefore that

$$f(x) = \sup\left\{\varphi(x) \;\middle|\; \varphi \in \mathcal{A}(f)\right\} \leq f^{**} \ \text{ for every } \ x \in \mathbb{R}^n\,.$$

The opposite inequality $f^{**} \leq f$ holds by Proposition 2.74(ii), and thus $f^{**} = f$.

The last implication **(iii)**$\Longrightarrow$**(i)** is obvious because the set epi g^* is always closed and convex for any function $g : \mathbb{R}^n \to \overline{\mathbb{R}}$. $\square$

2.9 DIRECTIONAL DERIVATIVES

Our next topic is *directional differentiability* of convex functions and its relationships with subdifferentiation. In contrast to classical analysis, directional derivative constructions in convex analysis are *one-sided* and related to directions with no classical plus-minus symmetry.

Definition 2.79 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be an extended-real-valued function and let $\bar{x} \in \operatorname{dom} f$. The* DIRECTIONAL DERIVATIVE *of the function f at the point $\bar{x}$ in the direction $d \in \mathbb{R}^n$ is the following limit—if it exists as either a real number or $\pm\infty$:*

$$f'(\bar{x}; d) := \lim_{t \to 0^+} \frac{f(\bar{x} + td) - f(\bar{x})}{t}. \tag{2.49}$$

Note that construction (2.49) is sometimes called the *right* directional derivative f at $\bar{x}$ in the direction d. Its *left* counterpart is defined by

$$f'_-(\bar{x}; d) := \lim_{t \to 0^-} \frac{f(\bar{x} + td) - f(\bar{x})}{t}.$$

It is easy to see from the definitions that

$$f'_-(\bar{x}; d) = -f'(\bar{x}; -d) \text{ for all } d \in \mathbb{R}^n,$$

and thus properties of the left directional derivative $f'_-(\bar{x}; d)$ reduce to those of the right one (2.49), which we study in what follows.

Lemma 2.80 *Given a convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ with $\bar{x} \in \operatorname{dom} f$ and given $d \in \mathbb{R}^n$, define*

$$\varphi(t) := \frac{f(\bar{x} + td) - f(\bar{x})}{t}, \quad t > 0.$$

Then the function φ is nondecreasing on $(0, \infty)$.

Proof. Fix any numbers $0 < t_1 < t_2$ and get the representation

$$\bar{x} + t_1 d = \frac{t_1}{t_2}\Big(\bar{x} + t_2 d\Big) + \Big(1 - \frac{t_1}{t_2}\Big)\bar{x}.$$

It follows from the convexity of f that

$$f(\bar{x} + t_1 d) \le \frac{t_1}{t_2} f(\bar{x} + t_2 d) + \Big(1 - \frac{t_1}{t_2}\Big) f(\bar{x}),$$

which implies in turn the inequality

$$\varphi(t_1) = \frac{f(\bar{x} + t_1 d) - f(\bar{x})}{t_1} \le \frac{f(\bar{x} + t_2 d) - f(\bar{x})}{t_2} = \varphi(t_2).$$

This verifies that φ is nondecreasing on $(0, \infty)$. □

The next proposition establishes the directional differentiability of convex functions.

Proposition 2.81 *For any convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and any $\bar{x} \in \text{dom}\, f$, the directional derivative $f'(\bar{x}; d)$ (and hence its left counterpart) exists in every direction $d \in \mathbb{R}^n$. Furthermore, it admits the representation via the function φ is defined in Lemma 2.80:*

$$f'(\bar{x}; d) = \inf_{t>0} \varphi(t), \quad d \in \mathbb{R}^n$$

Proof. Lemma 2.80 tells us that the function φ is nondecreasing. Thus we have

$$f'(\bar{x}; d) = \lim_{t \to 0+} \frac{f(\bar{x} + td) - f(\bar{x})}{t} = \lim_{t \to 0^+} \varphi(t) = \inf_{t>0} \varphi(t),$$

which verifies the results claimed in the proposition. □

Corollary 2.82 *If $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a convex function, then $f'(\bar{x}; d)$ is a real number for any $\bar{x} \in$ int(dom f) and $d \in \mathbb{R}^n$.*

Proof. It follows from Theorem 2.29 that f is locally Lipschitz continuous around $\bar{x}$, i.e., there is $\ell \geq 0$ such that

$$\left| \frac{f(\bar{x} + td) - f(\bar{x})}{t} \right| \leq \frac{\ell t \|d\|}{t} = \ell \|d\| \text{ for all small } t > 0,$$

which shows that $|f'(\bar{x}; d)| \leq \ell \|d\| < \infty$. □

To establish relationships between directional derivatives and subgradients of general convex functions, we need the following useful observation.

Lemma 2.83 *For any convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\bar{x} \in \text{dom}\, f$, we have*

$$f'(\bar{x}; d) \leq f(\bar{x} + d) - f(\bar{x}) \ \textit{whenever} \ d \in \mathbb{R}^n .$$

Proof. Using Lemma 2.80, we have for the function φ therein that

$$\varphi(t) \leq \varphi(1) = f(\bar{x} + d) - f(\bar{x}) \text{ for all } t \in (0, 1),$$

which justifies the claimed property due to $f'(\bar{x}; d) = \inf_{t>0} \varphi(t) \leq \varphi(1)$. □

Theorem 2.84 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex with $\bar{x} \in \text{dom}\, f$. The following are equivalent:

(i) $v \in \partial f(\bar{x})$.
(ii) $\langle v, d \rangle \leq f'(\bar{x}; d)$ for all $d \in \mathbb{R}^n$.
(iii) $f'_-(\bar{x}; d) \leq \langle v, d \rangle \leq f'(\bar{x}; d)$ for all $d \in \mathbb{R}^n$.

Proof. Picking any $v \in \partial f(\bar{x})$ and $t > 0$, we get

$$\langle v, td \rangle \le f(\bar{x} + td) - f(\bar{x}) \;\text{ whenever }\; d \in \mathbb{R}^n,$$

which verifies the implication **(i)**$\Longrightarrow$**(ii)** by taking the limit as $t \to 0^+$. Assuming now that assertion **(ii)** holds, we get by Lemma 2.83 that

$$\langle v, d \rangle \le f'(\bar{x}; d) \le f(\bar{x} + d) - f(\bar{x}) \;\text{ for all }\; d \in \mathbb{R}^n\,.$$

It ensures by definition (2.13) that $v \in \partial f(\bar{x})$, and thus assertions **(i)** and **(ii)** are equivalent.

It is obvious that **(iii)** yields **(ii)**. Conversely, if **(ii)** is satisfied, then for $d \in \mathbb{R}^n$ we have $\langle v, -d \rangle \le f'(\bar{x}; -d)$, and thus

$$f'_{-}(\bar{x}; d) = -f'(\bar{x}; -d) \le \langle v, d \rangle \;\text{ for any }\; d \in \mathbb{R}^n\,.$$

This justifies the validity of **(iii)** and completes the proof of the theorem. □

Let us next list some properties of (2.49) as a function of the direction.

Proposition 2.85 *For any convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ with $\bar{x} \in \operatorname{dom} f$, we define the directional function $\psi(d) := f'(\bar{x}; d)$, which satisfies the following properties:*

(i) $\psi(0) = 0$.
(ii) $\psi(d_1 + d_2) \le \psi(d_1) + \psi(d_2)$ *for all* $d_1, d_2 \in \mathbb{R}^n$.
(iii) $\psi(\alpha d) = \alpha \psi(d)$ *whenever* $d \in \mathbb{R}^n$ *and* $\alpha > 0$.
(iv) *If furthermore* $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$, *then* ψ *is finite on* $\mathbb{R}^n$.

Proof. It is straightforward to deduce properties **(i)–(iii)** directly from definition (2.49). For instance, **(ii)** is satisfied due to the relationships

$$\begin{aligned}
\psi(d_1 + d_2) &= \lim_{t \to 0^+} \frac{f(\bar{x} + t(d_1 + d_2)) - f(\bar{x})}{t} \\
&= \lim_{t \to 0^+} \frac{f\big(\frac{\bar{x} + 2td_1 + \bar{x} + 2td_2}{2}\big) - f(\bar{x})}{t} \\
&\le \lim_{t \to 0^+} \frac{f(\bar{x} + 2td_1) - f(\bar{x})}{2t} + \lim_{t \to 0^+} \frac{f(\bar{x} + 2td_2) - f(\bar{x})}{2t} = \psi(d_1) + \psi(d_2).
\end{aligned}$$

Thus it remains to check that $\psi(d)$ is finite for every $d \in \mathbb{R}^n$ when $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$. To proceed, choose $\alpha > 0$ so small that $\bar{x} + \alpha d \in \operatorname{dom} f$. It follows from Lemma 2.83 that

$$\psi(\alpha d) = f'(\bar{x}; \alpha d) \le f(\bar{x} + \alpha d) - f(\bar{x}) < \infty.$$

Employing **(iii)** gives us $\psi(d) < \infty$. Further, we have from **(i)** and **(ii)** that

$$0 = \psi(0) = \psi\big(d + (-d)\big) \le \psi(d) + \psi(-d), \quad d \in \mathbb{R}^n,$$

which implies that $\psi(d) \ge -\psi(-d)$. This yields $\psi(d) > -\infty$ and so verifies **(iv)**. □

To derive the second major result of this section, yet another lemma is needed.

Lemma 2.86 *We have the following relationships between a convex function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *and the directional function* ψ *defined in Proposition 2.85:*

(i) $\partial f(\bar{x}) = \partial \psi(0)$.
(ii) $\psi^*(v) = \delta_\Omega(v)$ *for all* $v \in \mathbb{R}^n$, *where* $\Omega := \partial \psi(0)$.

Proof. It follows from Theorem 2.84 that $v \in \partial f(\bar{x})$ if and only if

$$\langle v, d - 0 \rangle = \langle v, d \rangle \le f'(\bar{x}; d) = \psi(d) = \psi(d) - \psi(0), \quad d \in \mathbb{R}^n .$$

This is equivalent to $v \in \partial \psi(0)$, and hence **(i)** holds.

To justify **(ii)**, let us first show that $\psi^*(v) = 0$ for all $v \in \Omega = \partial \psi(0)$. Indeed, we have

$$\psi^*(v) = \sup \left\{ \langle v, d \rangle - \psi(d) \;\middle|\; d \in \mathbb{R}^n \right\} \ge \langle v, 0 \rangle - \psi(0) = 0.$$

Picking now any $v \in \partial \psi(0)$ gives us

$$\langle v, d \rangle = \langle v, d - 0 \rangle \le \psi(d) - \psi(0) = \psi(d), \quad d \in \mathbb{R}^n,$$

which implies therefore that

$$\psi^*(v) = \sup \left\{ \langle v, d \rangle - \psi(d) \;\middle|\; d \in \mathbb{R}^n \right\} \le 0$$

and so ensures the validity of $\psi^*(v) = 0$ for any $v \in \partial \psi(0)$.

It remains to verify that $\psi^*(v) = \infty$ if $v \notin \partial \psi(0)$. For such an element v, find $d_0 \in \mathbb{R}^n$ with $\langle v, d_0 \rangle > \psi(d_0)$. Since ψ is positively homogeneous by Proposition 2.85, it follows that

$$\begin{aligned} \psi^*(v) &= \sup \left\{ \langle v, d \rangle - \psi(d) \;\middle|\; d \in \mathbb{R}^n \right\} \ge \sup_{t>0} (\langle v, t d_0 \rangle - \psi(t d_0)) \\ &= \sup_{t>0} t(\langle v, d_0 \rangle - \psi(d_0)) = \infty, \end{aligned}$$

which completes the proof of the lemma. □

Now we are ready establish a major relationship between the directional derivative and the subdifferential of an arbitrary convex function.

Theorem 2.87 Given a convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$, we have

$$f'(\bar{x}; d) = \max \left\{ \langle v, d \rangle \;\middle|\; v \in \partial f(\bar{x}) \right\} \text{ for any } d \in \mathbb{R}^n .$$

Proof. It follows from Proposition 2.76 that

$$f'(\bar{x};d) = \psi(d) = \psi^{**}(d), \quad d \in \mathbb{R}^n,$$

by the properties of the function $\psi(d) = f'(\bar{x};d)$ from Proposition 2.85. Note that ψ is a finite convex function in this setting. Employing now Lemma 2.86 tells us that $\psi^*(v) = \delta_\Omega(v)$, where $\Omega = \partial\psi(0) = \partial f(\bar{x})$. Hence we have by (2.46) that

$$\psi^{**}(d) = \delta^*_\Omega(d) = \sup\big\{\langle v, d\rangle \;\big|\; v \in \Omega\big\}.$$

Since the set $\Omega = \partial f(\bar{x})$ is compact by Proposition 2.47, we complete the proof. □

2.10 SUBGRADIENTS OF SUPREMUM FUNCTIONS

Let T be a nonempty subset of $\mathbb{R}^p$ and let $g : T \times \mathbb{R}^n \to \mathbb{R}$. For convenience, we also use the notation $g_t(x) := g(t,x)$. The *supremum function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ for g_t over T is

$$f(x) := \sup_{t\in T} g(t,x) = \sup_{t\in T} g_t(x), \quad x \in \mathbb{R}^n. \tag{2.50}$$

If the supremum in (2.50) is attained (this happens, in particular, when the index set T is compact and $g(\cdot,x)$ is continuous), then (2.50) reduces to the *maximum function*, which can be written in form (2.35) when T is a finite set.

The main goal of this section is to calculate the subdifferential (2.13) of (2.50) when the functions g_t are convex. Note that in this case the supremum function (2.50) is also convex by Proposition 1.43. In what follows we assume without mentioning it that *the functions* $g_t\colon \mathbb{R}^n \to \mathbb{R}$ *are convex for all* $t \in T$.

For any point $\bar{x} \in \mathbb{R}^n$, define the *active index set*

$$S(\bar{x}) := \big\{t \in T \;\big|\; g_t(\bar{x}) = f(\bar{x})\big\}, \tag{2.51}$$

which may be empty if the supremum in (2.50) is not attained for $x = \bar{x}$. We first present a simple lower subdifferential estimate for (2.66).

Proposition 2.88 *Let* dom $f \neq \emptyset$ *for the supremum function* (2.50) *over an arbitrary index set* T. *For any* $\bar{x} \in$ dom f, *we have the inclusion*

$$\operatorname{clco} \bigcup_{t\in S(\bar{x})} \partial g_t(\bar{x}) \subset \partial f(\bar{x}). \tag{2.52}$$

Proof. Inclusion (2.52) obviously holds if $S(\bar{x}) = \emptyset$. Supposing now that $S(\bar{x}) \neq \emptyset$, fix $t \in S(\bar{x})$ and $v \in \partial g_t(\bar{x})$. Then we get $g_t(\bar{x}) = f(\bar{x})$ and therefore

$$\langle v, x - \bar{x}\rangle \le g_t(x) - g_t(\bar{x}) = g_t(x) - f(\bar{x}) \le f(x) - f(\bar{x}),$$

which shows that $v \in \partial f(\bar{x})$. Since the subgradient set $\partial f(\bar{x})$ is a closed and convex, we conclude that inclusion (2.52) holds in this case. □

On subdifferential calculus *

Erik J. Balder

1 Introduction

The main purpose of these lectures is to familiarize the student with the basic ingredients of convex analysis, especially its subdifferential calculus. This is done while moving to a clearly discernible end-goal, the Karush-Kuhn-Tucker theorem, which is one of the main results of nonlinear programming. Of course, in the present lectures we have to limit ourselves most of the time to the Karush-Kuhn-Tucker theorem for *convex* nonlinear programming. While this is on the one hand restrictive, it is somewhat compensated for by extra structure that the Karush-Kuhn-Tucker theory gains in the presence of convexity.

The material is presented in the following way. It is assumed that several – but perhaps not all – students have already been exposed to some standard material on convex sets. *This material has been collected in the appendix*; it will be referred to during the lectures whenever the need arises. Sometimes further references will be given; as a rule these concern results that can be found in the textbooks [1] or [2]. The less standard part of the material, notably subdifferential calculus, is treated in the main part of the text.

2 Fundamental results on subdifferentials

The introduction of $+\infty$ and $-\infty$ as *extended* real numbers is an essential, simplifying ingredient of convex analysis, as we shall see below. The additional arithmetic is simple, but needs some care. Of course, one has $\alpha + (+\infty) = (+\infty) + \alpha = +\infty$ for every $\alpha \in (-\infty, +\infty]$; also, $\alpha - (+\infty) = -\infty$ for every $\alpha \in [-\infty, +\infty)$. Similar rules for adding/subtracting $-\infty$ can easily be gathered. However, neither $(+\infty) - (+\infty)$ nor $(+\infty) + (-\infty)$ is defined. This requires constant vigilance on the part of the reader: for instance, the identity $\alpha + \beta = \gamma + \beta$ can only be used to conclude that $\alpha = \gamma$ for $\alpha, \gamma \in [-\infty, +\infty]$ if $\beta \in \mathbb{R}$. For multiplication the additional rules apply: $\alpha \cdot (+\infty) = +\infty$ for every $\alpha \in (0, +\infty]$ and $\alpha \cdot (+\infty) = -\infty$ for every $\alpha \in [-\infty, 0)$. By definition, one also sets $0 \cdot (+\infty) = 0 \cdot (-\infty) = 0$. As for division, it is consistent with

*LNMB Ph.D. course "Convex Analysis for Optimization", September, 2010. These notes grew from earlier Ph.D. courses in Utrecht and Naples; the author is indebted to participants in those courses for helpful comments.

the above to have $\alpha/(+\infty) = \alpha/(-\infty) = 0$ for every $\alpha \in \mathbb{R}$, but of course fractions like $(+\infty)/(+\infty)$, etc. are undefined. Similar warnings hold: for instance, $\alpha/\beta = \gamma/\beta$ can only be used to conclude that $\alpha = \gamma$ for $\alpha, \gamma \in [-\infty, +\infty]$ if $\beta \in \mathbb{R}\backslash\{0\}$. Recall that the definition of a convex set can be found in Appendix A (Definition A.1). We now introduce a fundamental concept of this course.

Definition 2.1 A function $f : S \to (-\infty, +\infty]$, defined on a convex set $S \subset \mathbb{R}^n$, is said to be *convex on S* if for every $x_1, x_2 \in S$ and every $\lambda \in [0, 1]$

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).$$

The same function is said to be *strictly convex* if for every $x_1, x_2 \in S$, $x_1 \neq x_2$, and for every $\lambda \in (0, 1)$

$$f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2).$$

This definition does not take into consideration functions that can take the value $-\infty$, even though it could be expanded to include these.[1] By a "sign-mirror treatment" the above definition can be turned into the following: a function $f : S \to [-\infty, +\infty)$, defined on a convex set $S \subset \mathbb{R}^n$, is said to be *[strictly] concave on S* if the function $-f$ is [strictly] convex, as defined above. Because concave functions can always be turned into convex ones by changing the signs, this course will not consider concave functions explicitly.

Exercise 2.1 Prove the following:
a. Every linear[2] function $f(x) := a^t x + \alpha$, with $a \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, is a convex function on $\mathbb{R}^n$ (and note that it is also concave).
b. The function $f(x) := \beta|x|^2$ is strictly convex on $\mathbb{R}^n$ if $\beta > 0$ (note that f is strictly concave if $\beta < 0$).
c. The function f defined on $\mathbb{R}_+$ by $f(x) := 1/x$ if $x > 0$ and by $f(0) := \gamma$ can only be made convex by choosing $\gamma = +\infty$.
d. The function f defined on $\mathbb{R}$ by $f(x) := 1/x$ if $x > 0$ and $f(x) := +\infty$ if $x \leq 0$ is convex.
e. The function $f(x) := -\sqrt{x}$ is convex on $\mathbb{R}_+$.
f. The function f defined on $\mathbb{R}$ by $f(x) := -\sqrt{x}$ if $x > 0$ and by defining $f(x) \in (-\infty, +\infty]$ for $x \leq 0$ can only be a convex function if one sets $f(x) := +\infty$ for every $x < 0$ and $f(0) := \gamma$ with $\gamma \in [0, +\infty]$.

Exercise 2.2 Let $S \subset \mathbb{R}^n$ be a convex set and let $f : S \to (-\infty, +\infty]$. Then f is said to be *quasiconvex* on S if for every $\alpha \in \mathbb{R}$ the so-called *lower level* set

$$S_\alpha := \{x \in S : f(x) \leq \alpha\}$$

[1] Functions that can take the value $-\infty$ are called *improper* in convex analysis. It can be shown that improper convex functions have a certain "pathological" structure, which is never encountered in realistic convex optimization problems.

[2] More accurately, such a function is called *affine*.

is convex.
a. Prove that if f is convex on S, then it is also quasiconvex on S.
b. Prove that f is quasiconvex on S if and only if for every $\alpha \in \mathbb{R}$ the set $\{x \in S : f(x) < \alpha\}$ is convex.
c. Let $g : D \to \mathbb{R}$ be a nondecreasing function on an interval $D \subset \mathbb{R}$ with $D \supset f(S)$ (note that this forces f to have values in $\mathbb{R}$). Prove that the composed function $h(x) := g(f(x))$ is also quasiconvex on S. *Hint:* Be careful: the function g is allowed to have discontinuities.

Exercise 2.3 Prove that the function $f(x) := -\exp(-x^2)$ is quasiconvex on $\mathbb{R}$, but not convex on $\mathbb{R}$. *Hint:* Prove monotonicity properties of f on respectively $\mathbb{R}_+$ and $\mathbb{R}_-$.

Exercise 2.4 For a function $f : S \to (-\infty, +\infty]$ one denotes by $\text{argmin}_{x \in S} f(x)$ the set (possibly empty) of all minimizers of f on S. That is to say

$$\text{argmin}_{x \in S} f(x) := \{z \in S : f(z) = \inf_{x \in S} f(x)\}.$$

a. Prove that the set $\text{argmin}_{x \in S} f(x)$ is convex if the function f is quasiconvex on S.
b. Prove that the set $\text{argmin}_{x \in S} f(x)$ contains at most one element if the function f is strictly convex on S.

Exercise 2.5 Prove the following *automatic* extension result for the domain of a convex function: if $f : S \to (-\infty, +\infty]$, defined on the convex set $S \subset \mathbb{R}^n$, is convex on S, then $\hat{f} : \mathbb{R}^n \to (-\infty, +\infty]$ is convex on $\mathbb{R}^n$, where $\hat{f}(x) := f(x)$ if $x \in S$ and $\hat{f}(x) := +\infty$ if $x \notin S$.

Note that this kind of extension has been practiced already in Exercise 2.1d, f above. As an important consequence of Exercise 2.5, we can often limit ourselves to the study of convex functions on the full space $\mathbb{R}^n$. This standardization can be very convenient. In the converse direction, we distinguish the subset of $\mathbb{R}^n$ on which a convex function $f : \mathbb{R}^n \to (-\infty, +\infty]$ "really matters" in the following way:

Definition 2.2 The *essential domain* of a function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is the set dom f, given by

$$\text{dom } f := \{x \in \mathbb{R}^n : f(x) < +\infty\}.$$

It is clear that for every $x_0 \in \mathbb{R}^n$ the following equivalence holds: $x_0 \in \text{dom} f$ if and only if $f(x_0) \in \mathbb{R}$. Note also that if $f : \mathbb{R}^n \to (-\infty, +\infty]$ is a convex function (see Definition 2.1), then dom f is a convex set (see Definition A.1).

Next, we discuss some methods to create new convex functions from known convex functions. To begin with, it is easy to see that if $f_1, \ldots, f_m : \mathbb{R}^n \to (-\infty, +\infty]$ are convex functions, then so are their pointwise sum $f(x) := \sum_{i=1}^m f_i(x)$ and pointwise maximum $\max_{1 \le i \le m} f_i(x)$. More generally, if $\alpha_1, \ldots, \alpha_m$ are in $\mathbb{R}_+$, then the pointwise sum $f(x) := \sum_{i=1}^m \alpha_i f_i(x)$ is also a convex function (on $\mathbb{R}^n$). Another, more powerful device to create new convex functions out of known convex functions is composition; this is the subject of the following two exercises:

Exercise 2.6 Let $f : S \to \mathbb{R}$ be a convex function on the convex set S and let $g : D \to \mathbb{R}$ be a convex function on the convex set $D \subset \mathbb{R}$, with $D \supset f(S)$. Suppose in addition that the function g is also *nondecreasing* on D (i.e., $\xi_1 \leq \xi_2$ implies $g(\xi_1) \leq g(\xi_2)$ for all $\xi_1, \xi_2 \in D$). Demonstrate that the composed function $h(x) := g(f(x))$ is also convex on S. Prove also that if g is merely nondecreasing (but perhaps not convex), then h is a quasiconvex function on S.

Exercise 2.7 a. Let $f : \mathbb{R}^n \to [0, +\infty]$ be convex on $\mathbb{R}^n$. Prove that f^2 is also a convex function on $\mathbb{R}^n$.
b. Prove that the function $f(x) := 1 - \sqrt{1 - x^2}$ is convex on $[-1, +1]$.
c. Prove that the function $f(x) := \exp(x^2)$ is convex on $\mathbb{R}$.

Below, in Proposition 2.7, the reader will find another important tool to determine whether a given function is convex.

Definition 2.3 Given $S \subset \mathbb{R}^n$, consider the following function $\chi_S : \mathbb{R}^n \to \{0, +\infty\}$

$$\chi_S(x) := \begin{cases} 0 & \text{if } x \in S \\ +\infty & \text{if } x \notin S. \end{cases}$$

This function is called the *indicator function* of the set S.

This definition turns sets into closely related functions. It is easy to see that $S \subset \mathbb{R}^n$ is a convex set if and only if its indicator function χ_S is a convex function. In a converse direction, convex functions can also be turned into closely related convex sets:

Definition 2.4 The *epigraph* of a function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is the subset epi f of $\mathbb{R}^n \times \mathbb{R}$ defined by

$$\text{epi } f := \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq y\}$$

Exercise 2.8 Let $f : \mathbb{R}^n \to (-\infty, +\infty]$. Prove the following: the function f is convex if and only if its epigraph epi f is a convex subset of $\mathbb{R}^n \times \mathbb{R}$.

Definition 2.5 a. A *subgradient* of a function $f : \mathbb{R}^n \to (-\infty, +\infty]$, $f \not\equiv +\infty$, at the point $x_0 \in \mathbb{R}^n$ is a vector $\xi \in \mathbb{R}^n$ such that

$$f(x) \geq f(x_0) + \xi^t(x - x_0) \text{ for all } x \in \mathbb{R}^n.$$

The set $\partial f(x_0)$ (possibly empty) of all such subgradients is called the *subdifferential* of f at the point x_0. Observe that this definition is only nontrivial if $x_0 \in \text{dom} f$: if $x_0 \in \mathbb{R}^n \backslash \text{dom } f$, then $f(x_0) = +\infty$, so $\partial f(x_0) = \emptyset$.

From now on, the trivial function $f \equiv +\infty$ is excluded from our considerations. For convex functions, subgradients form a generalization of the classical notion of gradient:

Proposition 2.6 *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex function that is differentiable at the point $x_0 \in \operatorname{int} \operatorname{dom} f$. Then $\partial f(x_0) = \{\nabla f(x_0)\}$.*

The proof of this proposition will be given later, because its proof uses Theorem 2.15. Observe that below this proposition applies to some points in Example 2.9(a) and also to Example 2.9(b).

Exercise 2.9 a. Consider the function $f : \mathbb{R} \to (-\infty, +\infty]$, defined by

$$f(x) := \begin{cases} 0 & \text{if } x \in [-1, +1] \\ |x| - 1 & \text{if } x \in [-2, -1) \cup (1, 2] \\ +\infty & \text{if } x \in (-\infty, -2) \cup (2, +\infty). \end{cases}$$

Demonstrate that

$$\partial f(x) = \begin{cases} \{0\} & \text{if } x \in (-1, 1) \\ [-1, 0] & \text{if } x = -1 \\ [0, 1] & \text{if } x = 1 \\ \{-1\} & \text{if } x \in (-2, -1) \\ \{1\} & \text{if } x \in (1, 2) \\ (-\infty, -1] & \text{if } x = -2 \\ [1, +\infty) & \text{if } x = 2 \\ \text{undefined} & \text{if } x \in (-\infty, -2) \cup (2, +\infty). \end{cases}$$

b. Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be given by $f(x) := 1 - \sqrt{1 - x^2}$ if $x \in [-1, +1]$ and by $f(x) := +\infty$ if $x < -1$ or $x > 1$. Demonstrate that

$$\partial f(x) = \begin{cases} \{x/\sqrt{1 - x^2}\} & \text{if } x \in (-1, +1) \\ \emptyset & \text{if } x \leq -1 \text{ or } x \geq 1. \end{cases}$$

Proposition 2.6 can be used to provide a very useful characterization of convexity for *differentiable* functions on $\mathbb{R}$:

Proposition 2.7 (i) *Let $f : S \to \mathbb{R}$ be a differentiable function on the open, convex set $S \subset \mathbb{R}^n$. Then f is convex on S if and only if the following monotonicity property holds*

$$(\nabla f(x_1) - \nabla f(x_2))^t (x_1 - x_2) \geq 0 \textit{ for every } x_1, x_2 \in S.$$

(i') *Let $f : S \to \mathbb{R}$ be a differentiable function on the open, convex set $S \subset \mathbb{R}^n$. Then f is strictly convex on S if and only if the following monotonicity property holds*

$$(\nabla f(x_1) - \nabla f(x_2))^t (x_1 - x_2) > 0 \textit{ for every } x_1, x_2 \in S,\ x_1 \neq x_2.$$

(ii) *Let $f : S \to \mathbb{R}$ be a second order continuously differentiable function on the open, convex set $S \subset \mathbb{R}^n$. Then f is convex on S if and only if its Hessian matrix*

$$H_f(x) := \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right)_{i,j}$$

is positive semidefinite at every point x of S.

(ii') *Let $f : S \to \mathbb{R}$ be a second order continuously differentiable function on the open, convex set $S \subset \mathbb{R}^n$. Then f is strictly convex on S if its Hessian matrix*

$$H_f(x) := \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right)_{i,j}$$

is positive definite at every point x of S.

Recall here that an $n \times n$ matrix M is *positive semidefinite* if $d^t M d \geq 0$ for all $d \in \mathbb{R}^n$. It is *positive definite* if $d^t M d > 0$ for all $d \in \mathbb{R}^n$.

PROOF. (i) If f is convex on S, then Proposition 2.6, together with the definition of subdifferential, implies

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^t (x_2 - x_1) \text{ and } f(x_1) \geq f(x_2) + \nabla f(x_2)^t (x_1 - x_2).$$

This immediately gives the desired monotonicity.

Conversely, given monotonicity, fix x, x' in S and let $\phi(t) := f(tx' + (1-t)x'')$, $t \in [0,1]$. By the mean value theorem there exists $\theta \in (0,1)$ such that $\phi(1) - \phi(0) = \phi'(\theta)$, i.e., $f(x') - f(x'') = \nabla f(\tilde{x})^t (x' - x'')$, where $\tilde{x} := \theta x' + (1-\theta)x''$. Monotonicity implies $(\nabla f(\tilde{x}) - \nabla f(x''))^t (\tilde{x} - x'') \geq 0$, i.e., $\theta(\nabla f(\tilde{x}) - \nabla f(x''))^t (x' - x'') \geq 0$. Hence, $\nabla f(\tilde{x})^t (x' - x'') \geq \nabla f(x''))^t (x' - x'')$. Thus, it follows that

$$f(x') \geq f(x'') + \nabla f(x'')^t (x' - x'') \text{ for every pair } x', x'' \in S.$$

To prove that this property implies the convexity of f, let $x_1, x_2 \in S$, let $\lambda \in [0,1]$ and set $x_3 := \lambda x_1 + (1-\lambda) x_2$. By applying the previous property to $x'' := x_3$ and successively to $x' = x_1$ and $x' = x_2$, we obtain

$$f(x_1) \geq f(x_3) + \nabla f(x_3)^t (x_1 - x_3) \text{ and } f(x_2) \geq f(x_3) + \nabla f(x_3)^t (x_2 - x_3).$$

Multiplying the left hand sides by λ and $1 - \lambda$ respectively, this easily leads to $\lambda f(x_1) + (1-\lambda) f(x_2) \geq f(x_3)$.

(ii) The underlying idea is that monotonicity (as in part (i)) of the first order derivative of f can, in turn, be characterized by "nonnegativity" (i.e., positive semidefiniteness) of the second order derivative. We refer to [2] for the details. Parts (i') and (ii') go analogously (exercise). QED

Specialized to $n = 1$, Proposition 2.7 is as follows:

Corollary 2.8 (i) *Let $f : S \to \mathbb{R}$ be a differentiable function on the open, convex set $S \subset \mathbb{R}$. Then f is convex [strictly convex] on S if and only if its derivative is nondecreasing [increasing].*

(ii) *Let $f : S \to \mathbb{R}$ be a second order continuously differentiable function on the open, convex set $S \subset \mathbb{R}$. Then f is convex [strictly convex] on S if and only if [if] its second derivative is nonnegative [positive].*

Exercise 2.10 Find the smallest $\alpha \in \mathbb{R}$ for which $f(x) := x \exp(-x)$ is convex on the set $[\alpha, +\infty)$.

Exercise 2.11 Consider for $\alpha, \beta > 0$ the function $f(x_1, x_2) := -x_1^\alpha x_2^\beta$ on $\mathbb{R}^2_+$. Prove the following:
a. If $\alpha + \beta \leq 1$, then f is convex on $\mathbb{R}^2_+$.
b. If $\alpha + \beta > 1$, then f is not convex on $\mathbb{R}^2_+$, but it is still quasiconvex. *Hint:* use Exercise 2.6.

Theorem 2.9 (Moreau-Rockafellar) *Let $f, g : \mathbb{R}^n \to (-\infty, +\infty]$ be convex functions. Then for every $x_0 \in \mathbb{R}^n$*

$$\partial f(x_0) + \partial g(x_0) \subset \partial(f + g)(x_0).$$

Moreover, suppose that $\text{int dom } f \cap \text{dom } g \neq \emptyset$. *Then for every $x_0 \in \mathbb{R}^n$ also*

$$\partial(f + g)(x_0) \subset \partial f(x_0) + \partial g(x_0).$$

Proof. The proof of the first part is elementary: Let $\xi_1 \in \partial f(x_0)$ and $\xi_2 \in \partial g(x_0)$. Then for all $x \in \mathbb{R}^n$

$$f(x) \geq f(x_0) + \xi_1^t(x - x_0), \ g(x) \geq g(x_0) + \xi_2^t(x - x_0),$$

so addition gives $f(x) + g(x) \geq f(x_0) + g(x_0) + (\xi_1 + \xi_2)^t(x - x_0)$. Hence $\xi_1 + \xi_2 \in \partial(f + g)(x_0)$.

To prove the second part, let $\xi \in \partial(f + g)(x_0)$. First, observe that $f(x_0) = +\infty$ implies $(f + g)(x_0) = +\infty$, whence $f + g \equiv +\infty$, which is impossible by $\xi \in \partial(f + g)(x_0)$. Likewise, $g(x_0) = +\infty$ is impossible. Hence, from now on we know that both $f(x_0)$ and $g(x_0)$ belong to $\mathbb{R}$. We form the following two sets in $\mathbb{R}^{n+1}$.

$$\Lambda_f := \{(x - x_0, y) \in \mathbb{R}^n \times \mathbb{R} : y > f(x) - f(x_0) - \xi^t(x - x_0)\}$$

$$\Lambda_g := \{(x - x_0, y) : -y \geq g(x) - g(x_0)\}.$$

Observe that both sets are nonempty and convex (see Exercise 2.8), and that $\Lambda_f \cap \Lambda_g = \emptyset$ (the latter follows from $\xi \in \partial(f + g)(x_0)$). Hence, by the set-set-separation Theorem A.4, there exists $(\xi_0, \mu) \in \mathbb{R}^{n+1}$ and $\alpha \in \mathbb{R}$, $(\xi_0, \mu) \neq (0, 0)$, such that

$$\xi_0^t(x - x_0) + \mu y \leq \alpha \text{ for all } (x, y) \text{ with } y > f(x) - f(x_0) - \xi^t(x - x_0),$$

$$\xi_0^t(x - x_0) + \mu y \geq \alpha \text{ for all } (x, y) \text{ with } -y \geq g(x) - g(x_0).$$

By $(0, 0) \in \Lambda_g$ we get $\alpha \leq 0$. But also $(0, \epsilon) \in \Lambda_f$ for every $\epsilon > 0$, and this gives $\mu\epsilon \leq \alpha$, so $\mu \leq 0$ (take $\epsilon = 1$). In the limit, for $\epsilon \to 0$, we find $\alpha \geq 0$. Hence $\alpha = 0$ and $\mu \leq 0$. We now claim that $\mu = 0$ is impossible. Indeed, if one had $\mu = 0$, then the first of the above two inequalities would give

$$\xi_0^t(x - x_0) \leq 0 \text{ for all } (x, y) \text{ with } y > f(x) - f(x_0) - \xi^t(x - x_0),$$

which is equivalent to

$$\xi_0^t(x - x_0) \leq 0 \text{ for all } x \in \text{dom } f$$

(simply note that when $f(x) < +\infty$ one can always achieve $y > f(x) - f(x_0) - \xi^t(x - x_0)$ by choosing y sufficiently large). Likewise, the second inequality would give

$$\xi_0^t(x - x_0) \geq 0 \text{ for all } x \in \text{dom } g.$$

In particular, for $\tilde{x}$ as above this would imply $\xi_0^t(\tilde{x} - x_0) = 0$. But since $\tilde{x}$ lies in the interior of dom f (so for some $\delta > 0$ the ball $N_\delta(\tilde{x})$ belongs to dom f), the preceding would imply

$$\xi_0^t u = \xi_0^t(\tilde{x} + u - x_0) \leq 0 \text{ for all } u \in N_\delta(0).$$

Clearly, this would give $\xi_0 = 0$ (take $u := \delta\xi_0/2$), which would be in contradiction to $(\xi_0, \mu) \neq (0, 0)$. Hence, we conclude $\mu < 0$. Dividing the separation inequalities by $-\mu$ and setting $\bar{\xi}_0 := -\xi_0/\mu$, this results in

$$\bar{\xi}_0^t(x - x_0) \leq y \text{ for all } (x, y) \text{ with } y > f(x) - f(x_0) - \xi^t(x - x_0),$$

$$\bar{\xi}_0^t(x - x_0) \geq y \text{ for all } (x, y) \text{ with } -y \geq g(x) - g(x_0).$$

The last inequality gives $-\bar{\xi}_0 \in \partial g(x_0)$ (set $y := g(x_0) - g(x)$) and the one but last inequality gives $\xi + \bar{\xi}_0 \in \partial f(x_0)$ (take $y := f(x) - f(x_0) - \xi^t(x - x_0) + \epsilon$ and let $\epsilon \downarrow 0$). Since $\xi = (\xi + \bar{\xi}_0) - \bar{\xi}_0$, this finishes the proof. QED

Exercise 2.12 Show by means of an example that the condition int dom $f \cap$dom $g \neq \emptyset$ in Theorem 2.9 cannot be omitted.

Exercise 2.13 Find and prove an version of the Moreau-Rockafellar theorem that applies to the subdifferentials of a finite sum of convex functions.

As a precursor to the Karush-Kuhn-Tucker theorem, we have now the following application of the Moreau-Rockafellar theorem.

Theorem 2.10 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $S \subset \mathbb{R}^n$ be a nonempty convex set. Consider the optimization problem*

$$(P) \ \inf_{x \in S} f(x).$$

Then $\bar{x} \in S$ is an optimal solution of (P) if and only if there exists a subgradient $\bar{\xi} \in \partial f(\bar{x})$ such that

$$\bar{\xi}^t(x - \bar{x}) \geq 0 \textit{ for all } x \in S. \tag{1}$$

PROOF. Recall from Definition 2.3 that χ_S is the indicator function of S. Now let $\bar{x} \in S$ be arbitrary. Then the following is trivial: $\bar{x}$ is an optimal solution of (P) if and only if

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

By the Moreau-Rockafellar Theorem 2.9, we have

$$\partial(f + \chi_S)(\bar{x}) = \partial f(\bar{x}) + \partial \chi_S(\bar{x}).$$

To see that its conditions hold, observe that dom $f = \mathbb{R}^n$ and dom $\chi_S = S$. So it follows that $\bar{x}$ is an optimal solution of (P) if and only if $0 \in \partial f(\bar{x}) + \partial \chi_S(\bar{x})$. By the definition of the sum of two sets this means that $\bar{x}$ is an optimal solution of (P) if and only if $0 = \bar{\xi} + \bar{\xi}'$ for some $\bar{\xi} \in \partial f(\bar{x})$ and $\bar{\xi}' \in \partial \chi_S(\bar{x})$. Of course, the former means $\bar{\xi}' = -\bar{\xi}$, so $-\bar{\xi} \in \partial \chi_S(\bar{x})$, which is equivalent to

$$\chi_S(x) \geq \chi_S(\bar{x}) + (-\bar{\xi})^t(x - \bar{x}) \text{ for all } x \in \mathbb{R}^n,$$

i.e., to (1). QED

Remark 2.11 *As the application of the Moreau-Rockafellar theorem in the above proof shows, the sufficiency part of Theorem 2.10 remains valid for a convex function $f : \mathbb{R}^n \to (-\infty, +\infty]$, i.e., a function that can attain the value $+\infty$. In that same situation the necessity also remains valid, provided that we suppose either* int dom $f \cap S \neq \emptyset$ *or* dom $f \cap$int $S \neq \emptyset$. *In particular, this remark applies to automatic extensions of the type introduced in Exercise 2.5.*

Exercise 2.14 Show by means of an example that, without the additional condition suggested in Remark 2.11, it is essential in Theorem 2.10 to have a function f with values in $\mathbb{R}$. [*Hint:* In boundary points of dom f the subdifferential of f can be empty, as shown in certain examples above.]

Exercise 2.15 Let $S \subset \mathbb{R}^2$ be given by the following system of inequalities: $\xi_1 \geq 0, \xi_2 \geq 0$, $-\xi_1 + \xi_2 \leq 2$, $2\xi_1 + 3\xi_2 \leq 11$. Let $f(\xi_1, \xi_2) := \xi_1^2 + \xi_2^2 - 8\xi_1 - 20\xi_2 + 89$.
a. Prove that S is a convex set and that $f : S \to \mathbb{R}$ is convex.
b. Use Theorem 2.10 to show that $\xi_1 = 1$, $\xi_2 = 3$ is an optimal solution for minimizing f over S.
c. Prove that, actually, f is *strictly* convex, i.e., prove that $f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda) f(x_2)$ for every $x_1, x_2 \in S$, $x_1 \neq x_2$, and every $\lambda \in (0, 1)$.
d. Use part c to prove that $(1, 3)$ in part b is the only optimal solution.

Example 2.12 Let the convex set $S \subset \mathbb{R}^2$ be given by the following four inequalities: $\xi_1 \geq 0, \xi_2 \geq 0, \xi_2 \geq \xi_1^2$ and $\xi_2 \leq 4$. Let $f(\xi_1, \xi_2) := (\xi_1 - 10)^2 + (\xi_2 - 5)^2$; this measures the squared distance from (ξ_1, ξ_2) to the point $(10, 5)$. From a picture of S it would seem that $\bar{x} = (2, 4)$ is the point in S that is closest to $(10, 5)$. To check that $\bar{x} = (2, 4)$ is indeed the optimal solution of $\min_{x \in S} f(x)$, we apply Theorem 2.10: it is enough to verify that $\nabla f(2, 4)^t(\xi_1 - 2, \xi_2 - 4) \geq 0$ for every $(\xi_1, \xi_2) \in S$. Now $\nabla f(2, 4) = (-16, -2)$, so it must be verified that $-16(\xi_1 - 2) - 2(\xi_2 - 4) \geq 0$, i.e., that $8\xi_1 + \xi_2 \leq 20$ for every $(\xi_1, \xi_2) \in S$. This holds, because $(\xi_1, \xi_2) \in S$ implies directly $\xi_1 \leq 2$ and $\xi_2 \leq 4$. Since the function f is strictly convex, we conclude, moreover, from Exercise 2.4 that $(2, 4)$ is the unique point in S that is closest to $(10, 5)$.

Definition 2.13 The *directional derivative* of a convex function $f : \mathbb{R}^n \to (-\infty, +\infty]$ at the point $x_0 \in \text{dom} f$ in the direction $d \in \mathbb{R}^n$ is defined as

$$f'(x_0; d) := \lim_{\lambda \downarrow 0} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}.$$

The above limit is a well-defined number in $[-\infty, +\infty]$. This follows from the following proposition (why?), which shows that the difference quotients of a convex functions possess a monotonicity property:

Proposition 2.14 *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex function and let x_0 be a point in* dom$f$. *Then for every direction $d \in \mathbb{R}^n$ and every $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $\lambda_2 > \lambda_1 > 0$ we have*

$$\frac{f(x_0 + \lambda_1 d) - f(x_0)}{\lambda_1} \leq \frac{f(x_0 + \lambda_2 d) - f(x_0)}{\lambda_2}$$

PROOF. Note that

$$x_0 + \lambda_1 d = \frac{\lambda_1}{\lambda_2}(x_0 + \lambda_2 d) + (1 - \frac{\lambda_1}{\lambda_2})x_0.$$

So by convexity of f

$$f(x_0 + \lambda_1 d) \leq \frac{\lambda_1}{\lambda_2} f(x_0 + \lambda_2 d) + (1 - \frac{\lambda_1}{\lambda_2}) f(x_0).$$

Simple algebra shows that this is equivalent to the desired inequality. QED

In Appendix B the Fenchel conjugation of convex functions is studied; this tool plays a major role in the proof of the next theorem:

Theorem 2.15 *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex function and let x_0 be a point in* int dom f. *Then*

$$f'(x_0; d) = \sup_{\xi \in \partial f(x_0)} \xi^t d \text{ for every } d \in \mathbb{R}^n.$$

Exercise 2.16 You are asked to verify the identity of Theorem 2.15 explicitly in each of the following cases (so in each case you are asked to determine both the left and right hand sides independently, and then to show that the identity holds).
a. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function which is differentiable at the point $x_0 \in \text{int dom } f$.
b. Let $x_0 := 0$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be the convex function given by $f(x) := |x| := (\sum_i x_i^2)^{1/2}$ (Euclidean norm). *Hint:* here you must show, among other things, that $\partial f(0) = \{x \in \mathbb{R}^n : |x| \leq 1\}$.
c. Let $x_0 := 1$ and let $f : \mathbb{R} \to \mathbb{R}$ be the convex function $f(x) := \max(1, x)$.

The proof of Theorem 2.15 uses the following lemma:

Lemma 2.16 *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex function. Then f is continuous at any point $x_0 \in \text{int dom } f$; moreover, then $\partial f(x_0)$ is nonempty and compact.*

PROOF. Continuity: Consider $g(x) := f(x_0 + x) - f(x_0)$. Then g is convex and $g(0) = 0$. Let $e_1, \ldots, e_n$ be the unit vectors in $\mathbb{R}^n$. Denote the set $\{e_1, \ldots, e_n, -e_1, \ldots, -e_n\}$ by $\{y_1, \ldots, y_{2n}\}$. Let $\alpha \in (0, 1]$ be so small that $x_0 + \alpha y_i \in \text{dom } f$ for all i. Now for every $x \in \mathbb{R}^n$ such that $|x_i| \leq \alpha/n$ one has

$$x = \sum_{i, x_i > 0} \frac{x_i}{\alpha} \alpha e_i + \sum_{i, x_i < 0} \frac{-x_i}{\alpha} \alpha(-e_i) + (1 - \sum_i \frac{|x_i|}{\alpha})0$$

so that

$$g(x) \leq \sum_{i, x_i > 0}^{n} \frac{|x_i|}{\alpha} g(\alpha e_i) + \sum_{i, x_i < 0}^{n} \frac{|x_i|}{\alpha} g(-\alpha e_i) \leq \beta \sum_i |x_i|,$$

where $\beta := \alpha^{-1} \max_{1 \leq i \leq 2n}(f(x_0 + \alpha y_i) - f(x_0)) < +\infty$. Also, for the same x one has $0 = \frac{1}{2}x + \frac{1}{2}(-x)$, so

$$0 \leq \frac{1}{2} g(x) + \frac{1}{2} g(-x),$$

Hence $g(x) \geq -g(-x) \geq -\beta \sum_i |x_i|$ holds as well. We conclude therefore that g is continuous (and even Lipschitz-continuous) at 0, i.e., f is continuous at the original point x_0.

Nonemptiness: Let $g := \chi_{x_0}$. Then by the Moreau-Rockafellar theorem $\partial(f + g)(x_0) = \partial f(x_0) + \partial g(x_0)$. But both $\partial(f + g)(x_0)$ and $\partial g(x_0)$ are equal to $\mathbb{R}^n$ in this case, so $\partial f(x_0)$ cannot be empty (because of $\emptyset + \mathbb{R}^n = \emptyset$).

Compactness: Exercise 2.17. QED

Exercise 2.17 Prove the compactness part of Lemma 2.16. *Hint:* Use the continuity part and mimic certain components of the proof of that part.

PROOF OF THEOREM 2.15. By Proposition 2.14

$$q(d) := f'(x_0; d) := \lim_{\lambda \downarrow 0} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda} = \inf_{\lambda > 0} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}.$$

Since the pointwise limit of a sequence of convex functions is convex, it follows that $q : \mathbb{R}^n \to \mathbb{R}$ is convex (by the infimum expression for $q(d)$ the fact that $x_0 \in \text{int dom } f$ implies automatically $q(d) < +\infty$ for every d; also, $q(d) > -\infty$ for every d, because of the nonemptiness part of Lemma 2.16). Hence, q is continuous at every point $d \in \mathbb{R}^n$ (apply the continuity part of Lemma 2.16). So by the Fenchel-Moreau theorem (Theorem B.5 in the Appendix) we have for every d

$$q(d) = q^{**}(d) := \sup_{\xi \in \mathbb{R}^n} [d^t \xi - q^*(\xi)].$$

Let us calculate q^*. For any $\xi \in \mathbb{R}^n$ we have

$$q^*(\xi) := \sup_{d \in \mathbb{R}^n} [\xi^t d - q(d)] = \sup_{d, \lambda > 0} [\xi^t d - \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}] = \sup_{\lambda > 0} \sup_{d} [\xi^t d - \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}]$$

by the above infimum expression for $q(d)$. Fix $\lambda > 0$; then $z := x_0 + \lambda d$ runs through all of $\mathbb{R}^n$ as d runs through $\mathbb{R}^n$. Hence

$$\sup_d [\xi^t d - \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}] = \frac{f(x_0) - \xi^t x_0 + \sup_z [\xi^t z - f(z)]}{\lambda}.$$

Clearly, this gives

$$q^*(\xi) = \sup_{\lambda > 0} \frac{f(x_0) - \xi^t x_0 + f^*(\xi)}{\lambda} = \begin{cases} 0 & \text{if } \xi \in \partial f(x_0) \\ +\infty & \text{otherwise} \end{cases}$$

where we use Proposition B.4(v). Observe that in terms of the indicator function of the subdifferential this can be rewritten as $q^* = \chi_{\partial f(x_0)}$. Now that q^* has been calculated, we conclude from the above that for every $d \in \mathbb{R}^n$

$$f'(x_0; d) = q(d) = q^{**}(d) = \chi^*_{\partial f(x_0)}(d) = \sup_{\xi \in \partial f(x_0)} \xi^t d,$$

which proves the result. QED

PROOF OF PROPOSITION 2.6. By Theorem 2.15 we get

$$\nabla f(x_0)^t d = \sup_{\xi \in \partial f(x_0)} \xi^t d.$$

The remainder of the proof is left as an exercise.

Theorem 2.17 (Dubovitskii-Milyutin) *Let $f_1, \cdots, f_m : \mathbb{R}^n \to (-\infty, +\infty]$ be convex functions and let x_0 be a point in $\cap_{i=1}^m \text{int dom } f_i$. Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be given by*

$$f(x) := \max_{1 \le i \le m} f_i(x)$$

and let $I(x_0)$ be the (nonempty) set of all $i \in \{1, \cdots, m\}$ for which $f_i(x_0) = f(x_0)$. Then

$$\partial f(x_0) = \text{co } \cup_{i \in I(x_0)} \partial f_i(x_0).$$

PROOF. For our convenience we write $I := I(x_0)$. To begin with, observe that $\xi \in \partial f_i(x_0)$ easily implies $\xi \in \partial f(x_0)$ for each $i \in I$. Since $\partial f(x_0)$ is evidently convex, the inclusion "$\supset$" follows with ease. To prove the opposite inclusion, let ξ_0 be arbitrary in $\partial f(x_0)$. If ξ_0 were not to belong to the compact set co $\cup_{i \in I} \partial f_i(x_0)$, then we could separate strictly (note that each set $\partial f_i(x_0)$ is both closed and compact (exercise)): by Theorem A.2 there would exist $d \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ such that

$$\xi_0^t d > \alpha \ge \max_{i \in I} \sup_{\xi \in \partial f_i(x_0)} \xi^t d = \max_{i \in I} f_i'(x_0; d),$$

where the final identity follows from Theorem 2.15. But now observe that

$$f'(x_0; d) := \lim_{\lambda \downarrow 0} \max_{i \in I} \frac{f_i(x_0 + \lambda d) - f_i(x_0)}{\lambda} = \max_{i \in I} \lim_{\lambda \downarrow 0} \frac{f_i(x_0 + \lambda d) - f_i(x_0)}{\lambda} = \max_{i \in I} f_i'(x_0; d),$$

so the above gives $\xi_0^t d > f'(x_0; d)$. On the other hand, by $\xi_0 \in \partial f(x_0)$ it follows that $f(x_0 + \lambda d) \ge f(x_0) + \lambda \xi_0^t d$ for every $\lambda > 0$, whence $f'(x_0; d) \ge \xi_0^t d$. We thus have arrived at a contradiction. So the inclusion "$\subset$" must hold as well. QED

Exercise 2.18 a. In the above proof the following property is used: if $S \subset \mathbb{R}^n$ is compact, then its convex hull co S is compact. Prove this, using the following result of Carathéodory: in $\mathbb{R}^n$ every convex combination x of $p \geq n+1$ points $x_1, \ldots, x_p$ (i.e., $x = \sum_1^p \alpha_i x_i$ for $\alpha_i \geq 0$ and $\sum_1^p \alpha_i = 1$) can also be written as a convex combination of at most $n+1$ points $x_{i_1}, \ldots, x_{i_{n+1}} \subset \{x_1, \ldots, x_p\}$.
b. Give an example of a closed set $S \subset \mathbb{R}^n$ for which co S is *not* closed (conclusion: in the above proof it is essential to work with compactness).

Exercise 2.19 Let $f(x) := |x|$ on $S := \mathbb{R}$. Then $\partial f(0) = [-1, 1]$ (by Exercise 2.16(b) for $n = 1$). Demonstrate how this result can also be derived from Theorem 2.17.

Exercise 2.20 Show by means of an example that in Theorem 2.17 it is essential to have $x_0 \in \cap_i \text{int dom } f_i$.

3 The Kuhn-Tucker theorem for convex programming

We use the results of the previous section to derive the celebrated Kuhn-Tucker theorem for convex programming. Unlike its counterparts in section 4 of [1], this theorem gives necessary *and* sufficient conditions for optimality for the standard convex programming problem. First we discuss the situation with inequality constraints only.

Theorem 3.1 (Kuhn-Tucker – no equality constraints) *Let $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ be convex functions and let $S \subset \mathbb{R}^n$ be a convex set. Consider the* convex programming problem

$$(P) \ \inf_{x \in S} \{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0\}.$$

Let $\bar{x}$ be a feasible point of (P); denote by $I(\bar{x})$ the set of all $i \in \{1, \cdots, m\}$ for which $g_i(\bar{x}) = 0$.

(i) $\bar{x}$ is an optimal solution of (P) if there exist vectors of multipliers $\bar{u} := (\bar{u}_1, \cdots, \bar{u}_m) \in \mathbb{R}^m_+$ *and $\bar{\eta} \in \mathbb{R}^n$ such that the following three relationships hold:*

$$\bar{u}_i g_i(\bar{x}) = 0 \text{ for } i = 1, \cdots, m \quad \textbf{(complementary slackness)},$$

$$0 \in \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + \bar{\eta} \quad \textbf{(normal Lagrange inclusion)},$$

$$\bar{\eta}^t (x - \bar{x}) \leq 0 \text{ for all } x \in S \quad \textbf{(obtuse angle property)}.$$

(ii) Conversely, if $\bar{x}$ is an optimal solution of (P) and if $\bar{x} \in \text{int dom } f \cap \cap_{i \in I(\bar{x})} \text{int dom } g_i$, then there exist multipliers $\bar{u}_0 \in \{0, 1\}$, $\bar{u} \in \mathbb{R}^m_+$, $(\bar{u}_0, \bar{u}) \neq (0, 0)$, and $\bar{\eta} \in \mathbb{R}^n$ such that the complementary slackness relationship and obtuse angle property of part (i) hold, as well as the following:

$$0 \in \bar{u}_0 \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + \bar{\eta} \quad \textbf{(Lagrange inclusion)}.$$

Here the normal *case is said to occur when* $\bar{u}_0 = 1$ *and the* abnormal *case when* $\bar{u}_0 = 0$.

Remark 3.2 (minimum principle) *By Theorem 2.9, the normal Lagrange inclusion in Theorem 3.1 implies*

$$-\bar{\eta} \in \partial(f + \sum_{i \in I(\bar{x})} \bar{u}_i g_i)(\bar{x}).$$

So by Theorem 2.10 and Remark 2.11 it follows that

$$\bar{x} \in \mathrm{argmin}_{x \in S}[f(x) + \sum_{i \in I(\bar{x})} \bar{u}_i g_i(x)]\textbf{(minimum principle)}.$$

Likewise, under the additional condition $\mathrm{dom}\, f \cap \cap_{i \in I(\bar{x})} \mathrm{int}\ \mathrm{dom}\, g_i \neq \emptyset$*, this minimum principle implies the normal Lagrange inclusion by the converse parts of Theorem 2.10/Remark 2.11 and Theorem 2.9.*

Remark 3.3 (Slater's constraint qualification) *The following* Slater constraint qualification *guarantees normality: Suppose that there exists* $\tilde{x} \in S$ *such that* $g_i(\tilde{x}) < 0$ *for* $i = 1, \cdots, m$*. Then in part (ii) of Theorem 3.1 we have the normal case* $\bar{u}_0 = 1$*.*

Indeed, suppose we had $\bar{u}_0 = 0$*. For* $\bar{u}_0 = 0$ *instead of* $\bar{u}_0 = 1$ *the proof of the minimum principle in Remark 3.2 can be mimicked and gives*

$$\sum_{i=1}^{m} \bar{u}_i g_i(\bar{x}) \leq \sum_{i=1}^{m} \bar{u}_i g_i(\tilde{x}).$$

Since $(\bar{u}_1, \cdots, \bar{u}_m) \neq (0, \cdots, 0)$*, this gives* $\sum_{i=1}^{m} \bar{u}_i g_i(\bar{x}) < 0$*, in contradiction to complementary slackness.*

Proof of Theorem 3.1. Let us write $I := I(\bar{x})$. (*i*) By Remark 3.2 the minimum principle holds, i.e., for any $x \in S$ we have

$$f(x) + \sum_{i \in I} \bar{u}_i g_i(x) \geq f(\bar{x})$$

(observe that $\sum_{i \in I} \bar{u}_i g_i(\bar{x}) = 0$ by complementary slackness). Hence, for any *feasible* $x \in S$ we have

$$f(x) \geq f(x) + \sum_{i \in I} \bar{u}_i g_i(x) \geq f(\bar{x}),$$

by nonnegativity of the multipliers. Clearly, this proves optimality of $\bar{x}$.

(*ii*) Consider the auxiliary optimization problem

$$(P') \ \inf_{x \in S} \phi(x),$$

where $\phi(x) := \max[f(x) - f(\bar{x}), \max_{1 \leq i \leq m} g_i(x)]$. Since $\bar{x}$ is an optimal solution of (P), it is not hard to see that $\bar{x}$ is also an optimal solution of (P') (observe that $\phi(\bar{x}) = 0$

and that $x \in S$ is feasible if and only if $\max_{1 \leq i \leq m} g_i(x) \leq 0$). By Theorem 2.10 and Remark 2.11 there exists $\bar{\eta}$ in $\mathbb{R}^n$ such that $\bar{\eta}$ has the obtuse angle property and $-\bar{\eta} \in \partial\phi(\bar{x})$. By Theorem 2.17 this gives

$$-\bar{\eta} \in \partial\phi(\bar{x}) = \text{co}(\partial f(\bar{x}) \cup \cup_{i \in I} \partial g_i(\bar{x})).$$

Since subdifferentials are convex, we get the existence of $(u_0, \xi_0) \in \mathbb{R}_+ \times \partial f(\bar{x})$ and $(u_i, \xi_i) \in \mathbb{R}_+ \times \partial g_i(\bar{x})$, $i \in I$, such that $\sum_{i \in \{0\} \cup I} u_i = 1$ and

$$-\bar{\eta} = \sum_{i \in \{0\} \cup I} u_i \xi_i.$$

In case $u_0 = 0$, we are done by setting $\bar{u}_i := u_i$ for $i \in \{0\} \cup I$ and $\bar{u}_i := 0$ otherwise. Observe that in this case $(\bar{u}_1, \cdots, \bar{u}_m) \neq (0, \cdots, 0)$ by $\sum_{i \in I} u_i = 1$. In case $u_0 \neq 0$, we know that $u_0 > 0$, so we can set $\bar{u}_i := u_i / u_0$ for $i \in \{0\} \cup I$ and $\bar{u}_i := 0$ otherwise. QED

Example 3.4 Consider the following optimization problem:

$$(P) \text{ minimize } (x_1 - \frac{9}{4})^2 + (x_2 - 2)^2$$

over all $(x_1, x_2) \in \mathbb{R}^2_+$ such that

$$\begin{aligned} x_1^2 - x_2 &\leq 0 \\ x_1 + x_2 - 6 &\leq 0 \\ -x_1 + 1 &\leq 0 \end{aligned}$$

Since Slater's constraint qualification clearly holds, we get that a feasible point $(\bar{x}_1, \bar{x}_2)$ is optimal if and only if there exists $(\bar{u}_1, \bar{u}_2, \bar{u}_3) \in \mathbb{R}^3_+$ such that

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2(\bar{x}_1 - \frac{9}{4}) \\ 2(\bar{x}_2 - 2) \end{pmatrix} + \bar{u}_1 \begin{pmatrix} 2\bar{x}_1 \\ -1 \end{pmatrix} + \bar{u}_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \bar{u}_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \begin{pmatrix} \bar{\eta}_1 \\ \bar{\eta}_2 \end{pmatrix}$$

for some $\bar{\eta} := (\bar{\eta}_1, \bar{\eta}_2)^t$ with

$$\bar{\eta}^t (x - \bar{x}) \leq 0 \text{ for all } x \in \mathbb{R}^2_+$$

and such that

$$\begin{aligned} \bar{u}_1(\bar{x}_1^2 - \bar{x}_2) &= 0 \\ \bar{u}_2(\bar{x}_1 + \bar{x}_2 - 6) &= 0 \\ \bar{u}_3(-\bar{x}_1 + 1) &= 0 \end{aligned}$$

Let us first deal with $\bar{\eta}$: observe that the above obtuse angle property forces $\bar{\eta}_1$ and $\bar{\eta}_2$ to be nonpositive, and $\bar{x}_i > 0$ even implies $\bar{\eta}_i = 0$ for $i = 1, 2$ (this can be seen as a form of complementarity). Since $\bar{x}_1 \geq 1$, this means $\bar{\eta}_1 = 0$. Also, $\bar{x}_2 = 0$ stands no chance, because it would mean $\bar{x}_1^2 \leq 0$. Hence, $\bar{\eta} = 0$. We now distinguish the following possibilities for the set $I := I(\bar{x})$:

Case 1 ($I = \emptyset$): By complementary slackness, $\bar{u}_1 = \bar{u}_2 = \bar{u}_3 = 0$, so the Lagrange inclusion gives $\bar{x}_1 = 9/4$, $\bar{x}_2 = 2$, which violates the first constraint ($(9/4)^2 \not\leq 2$).

Case 2 ($I = \{1\}$): By complementary slackness, $\bar{u}_2 = \bar{u}_3 = 0$. The Lagrange inclusion gives $\bar{x}_1 = \frac{9}{4}(1 + \bar{u}_1)^{-1}$, $\bar{x}_2 = \bar{u}_1/2 + 2$, so, since $\bar{x}_1^2 = \bar{x}_2$, by definition of I, we obtain the equation $\bar{u}_1^3 + 6\bar{u}_1^2 + 9\bar{u}_1 = 49/8$, which has $\bar{u}_1 = 1/2$ as its only solution. It follows then that $\bar{x} = (3/2, 9/4)^t$.

At this stage we can already stop: Theorem 3.1(i) guarantees that, in fact, $\bar{x} = (3/2, 9/4)^t$ is an optimal solution of (P). Moreover, since the objective function $(x_1, x_2) \mapsto (x_1 - \frac{9}{4})^2 + (x_2 - 2)^2$ is *strictly* convex, it follows that any optimal solution of (P) must be unique. So $\bar{x} = (3/2, 9/4)^t$ is the unique optimal solution of (P).

Exercise 3.1 Consider the optimization problem

$$(P) \quad \sup_{(\xi_1,\xi_2)\in\mathbb{R}^2_+} \{\xi_1\xi_2 : 2\xi_1 + 3\xi_2 \leq 5\}.$$

Solve this problem using Theorem 3.1. *Hint:* The set of optimal solutions does not change if we apply a monotone transformation to the objective function. So one can use $f(\xi_1, \xi_2) := \sqrt{\xi_1\xi_2}$ to ensure convexity (see Exercise 2.11).

Exercise 3.2 Let $a_i > 0$, $i = 1, \ldots, n$ and let $p \geq 1$. Consider the optimization problem

$$(P) \quad \text{maximize} \sum_{i=1}^{n} a_i\xi_i \text{ over } (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$$

subject to $g(\xi) := \sum_{i=1}^n |\xi_i|^p = 1$.

a. Show that if the constraint $\sum_{i=1}^n |\xi_i|^p = 1$ is replaced by $\sum_{i=1}^n |\xi_i|^p \leq 1$, then this results in exactly the same optimal solutions.

b. Prove that $g : \mathbb{R}^n \to \mathbb{R}$, as defined above, is convex. Prove also that g is in fact strictly convex if $p > 1$.

c. Apply Theorem 3.1 to determine the optimal solutions of (P). *Hint:* Treat the cases $p = 1$ and $p > 1$ separately.

d. Derive from the result obtained in part (c) for $p > 1$ the following famous *Hölder inequality,* which is an extension of the Cauchy-Schwarz inequality: $|\sum_i a_i\xi_i| \leq (\sum_i a_i^q)^{1/q}(\sum_i |\xi_i|^p)^{1/p}$ for all $(\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$. Here q is defined by $q := p/(p-1)$.

Corollary 3.5 (Kuhn-Tucker – general case) *Let $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ be convex functions, let $S \subset \mathbb{R}^n$ be a convex set. Also, let A be a $p \times n$-matrix and let $b \in \mathbb{R}^p$. Define $L := \{x : Ax = b\}$. Consider the* convex programming problem

$$(P) \quad \inf_{x\in S}\{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0, Ax - b = 0\}.$$

Let $\bar{x}$ be a feasible point of (P); denote by $I(\bar{x})$ the set of all $i \in \{1, \cdots, m\}$ for which $g_i(\bar{x}) = 0$.

(i) $\bar{x}$ is an optimal solution of (P) if there exist vectors of multipliers $\bar{u} \in \mathbb{R}^m_+$, $\bar{v} \in \mathbb{R}^p$ and $\bar{\eta} \in \mathbb{R}^n$ such that the complementary slackness relationship and the obtuse angle property hold just as in Theorem 3.1(i), as well as the following version of the normal Lagrange inclusion:

$$0 \in \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + A^t \bar{v} + \bar{\eta}.$$

(ii) Conversely, if $\bar{x}$ is an optimal solution of (P) and if both $\bar{x} \in \text{int dom } f \cap \cap_{i \in I(\bar{x})} \text{int dom } g_i$ and $\text{int } S \cap L \neq \emptyset$, then there exist multipliers $\bar{u}_0 \in \{0, 1\}$, $\bar{u} \in \mathbb{R}^m_+$, $(\bar{u}_0, \bar{u}) \neq (0, 0)$, and $\bar{v} \in \mathbb{R}^p$, $\bar{\eta} \in \mathbb{R}^n$ such that the complementary slackness relationship and obtuse angle property of part (i) hold, as well as the following Lagrange inclusion:

$$0 \in \bar{u}_0 \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + A^t \bar{v} + \bar{\eta}.$$

PROOF. Observe that $\partial \chi_L(\bar{x}) = \text{im } A^t$. Indeed, $\eta \in \partial \chi_L(\bar{x})$ is equivalent to $\eta^t(x - \bar{x}) \leq 0$ for all $x \in L$, i.e., to $\eta^t(x - \bar{x}) = 0$ for all $x \in \mathbb{R}^n$ with $A(x - \bar{x}) = 0$. But the latter states that η belongs to the bi-orthoplement of the linear subspace im A^t, so it belongs to im A^t itself. This proves the observation. Let us note that the above problem (P) is precisely the same problem as the one of Theorem 3.1, but with S replaced by $S' := S \cap L$. Thus, parts (i) and (ii) follow directly from Theorem 3.1, but now $\bar{\eta}$ as in Theorem 3.1 has to be replaced by an element (say η') in $\partial \chi_{S'}$. From Theorem 2.9 we know that

$$\partial \chi_{S'}(\bar{x}) = \partial \chi_S(\bar{x}) + \partial \chi_L(\bar{x}),$$

in view of the condition int $S \cap L \neq \emptyset$. Therefore, η' can be decomposed as $\eta' = \bar{\eta} + \eta$, with $\bar{\eta} \in \partial \chi_S(\bar{x})$ (this amounts to the obtuse angle property, of course), and with $\eta \in \partial \chi_L(\bar{x})$. By the above there exists $\bar{v} \in \mathbb{R}^m$ with $\eta = A^t \bar{v}$ and this finishes the proof. QED

Example 3.6 Let $c_1, \cdots, c_n$, $a_1, \cdots, a_n$ and b be positive real numbers. Consider the following optimization problem:

$$(P) \text{ minimize } \sum_{i=1}^{n} \frac{c_i}{x_i}$$

over all $x = (x_1, \cdots, x_n)^t \in \mathbb{R}^n_{++}$ (the strictly positive orthant) such that

$$\sum_{i=1}^{n} a_i x_i = b.$$

Let us try to meet the sufficient conditions of Corollary 3.5(i). Thus, we must find a feasible $\bar{x} \in \mathbb{R}^n$ and multipliers $\bar{v} \in \mathbb{R}$, $\bar{\eta} \in \mathbb{R}^n$ such that

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{c_1}{\bar{x}_1^2} \\ \vdots \\ -\frac{c_n}{\bar{x}_n^2} \end{pmatrix} + \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \bar{v} + \bar{\eta}.$$

and such that the obtuse angle property holds for $\bar{\eta}$. To begin with the latter, since we seek $\bar{x}$ in the open set $S := \mathbb{R}^n_{++}$, the only $\bar{\eta}$ with the obtuse angle property is $\bar{\eta} = 0$. The above Lagrange inclusion gives $\bar{x}_i = (c_i/(\bar{v}a_i))^{1/2}$ for all i. To determine $\bar{v}$, which must certainly be positive, we use the constraint: $b = \sum_i a_i\bar{x}_i = \sum_i(a_ic_i/\bar{v})^{1/2}$, which gives $\bar{v} = (\sum_i(a_ic_i)^{1/2}/b)^2$. Thus, all conditions of Corollary 3.5(i) are seen to hold: an optimal solution of (P) is $\bar{x}$, given by

$$\bar{x}_i = \sqrt{\frac{c_i}{a_i}}\frac{b}{\sum_{j=1}^n \sqrt{a_jc_j}},$$

and it is implicit in our derivation that this solution is unique (exercise).

Remark 3.7 *By using the* relative interior *(denoted as "ri") of a convex set, i.e., the interior relative to the linear variety spanned by that set, one can obtain the following improvement of the nonempty intersection condition in Theorem 2.9: it is already enough that* ri dom $f \cap$ dom g *is nonempty. Since one can also prove that* $A(\text{ri } S) = \text{ri } A(S)$ *for any convex set* $S \subset \mathbb{R}^n$ *and any linear mapping* $A : \mathbb{R}^n \to \mathbb{R}^p$ *[2, Theorem 4.9], it follows that the nonempty intersection condition in Corollary 3.5 can be improved considerably into* ri $S \cap L \neq \emptyset$ *or,* equivalently*, into* $b \in A(\text{ri } S)$.

Exercise 3.3 In the above proof of Corollary 3.5 the fact was used that for a linear subspace M of $\mathbb{R}^n$ the following holds: let

$$M^\perp := \{x \in \mathbb{R}^n : x^t\xi = 0 \text{ for all } \xi \in M\},$$

This is a linear subspace itself (prove this), so $M^{\perp\perp} := (M^\perp)^\perp$ is well-defined. Prove that $M = M^{\perp\perp}$. *Hint:* This identity can be established by proving two inclusions; one of these is elementary and the other requires the use of projections.

Exercise 3.4 What becomes of Corollary 3.5 in the situation where there are no inequality constraints (i.e., just equality constraints)? Derive this version.

Exercise 3.5 Use Corollary 3.5 to prove the following famous theorem of Farkas. Let A be a $p \times n$-matrix and let $c \in \mathbb{R}^n$. Then precisely one of the following is true:

$$(1)\ \exists_{x\in\mathbb{R}^n} Ax \leq 0 \text{ (componentwise) and } c^tx > 0,\ \ (2)\ \exists_{y\in\mathbb{R}^p_+} A^ty = c.$$

Hint: Show first, by elementary means, that validity of (2) implies that (1) cannot hold. Next, apply Corollary 3.5 to a suitably chosen optimization problem in order to prove that if (1) does not hold, then (2) must be true.

A Standard material on convexity

Definition A.1 A set S in $\mathbb{R}^n$ is said to be *convex* if for every $x_1, x_2 \in S$ the line segment $\{\lambda x_1 + (1-\lambda)x_2 : 0 \leq \lambda \leq 1\}$ belongs to S.

For instance, a hyperplane $S = \{x \in \mathbb{R}^n : p^t x = \alpha\}$ or a ball $S = \{x \in \mathbb{R}^n : |x - x_0| \leq \beta\}$ are examples of convex sets. However, the sphere $S = \{x \in \mathbb{R}^n : |x - x_0| = \beta\}$ provides an example of a set that is not convex ($\beta > 0$). It is easy to see that arbitrary intersections of convex sets are again convex; also finite sums of convex sets are convex again.

Theorem A.2 (strict point-set separation [1, Thm. 2.4.4]) *Let S be a nonempty closed convex subset of $\mathbb{R}^n$ and let $y \in \mathbb{R}^n \backslash S$. Then there exists $p \in \mathbb{R}^n$, $p \neq 0$, such that*

$$\sup_{x \in S} p^t x < p^t y.$$

PROOF. It is a standard result that there exists $\hat{x} \in S$ such that $\sup_{s \in S} |y - s| = |y - \hat{x}|$ (consider a suitable closed ball around y and apply the theorem of Weierstrass [1, Thm. 2.3.1]). By convexity of S, this means that for every $x \in S$ and every $\lambda \in (0, 1]$

$$|y - (\lambda x + (1-\lambda)\hat{x})|^2 \geq |y - \hat{x}|^2.$$

Obviously, the expression on the left equals

$$|y - \hat{x} - \lambda(x - \hat{x})|^2 = |y - \hat{x}|^2 - 2\lambda(y - \hat{x})^t(x - \hat{x}) + \lambda^2|x - \hat{x}|^2,$$

so the above inequality amounts to

$$2\lambda(y - \hat{x})^t(x - \hat{x}) \leq \lambda^2|x - \hat{x}|^2$$

for every $x \in S$ and every $\lambda \in (0, 1]$. Dividing by $\lambda > 0$ and letting λ go to zero then gives

$$(y - \hat{x}) \cdot (x - \hat{x}) \leq 0 \text{ for all } x \in S.$$

Set $p := y - \hat{x}$; then $p \neq 0$ (note that $p = 0$ would imply $y \in S$). We clearly have $p^t x \leq p^t \hat{x}$. Also, we have now $p^t \hat{x} > p^t y$, for otherwise $(y - \hat{x})^t(\hat{x} - y) \geq 0$ would imply $y = \hat{x} \in S$, which is impossible. QED

For our next result, recall that $\partial S := \text{cl}S \cap \text{cl}(\mathbb{R}^n \backslash S) = \text{cl}S \backslash \text{int S}$ denotes the boundary of a set $S \subset \mathbb{R}^n$.

Theorem A.3 (supporting hyperplane [1, Thm. 2.4.7]) *Let S be a nonempty convex subset of $\mathbb{R}^n$ and let $y \in \partial S$. Then there exists $q \in \mathbb{R}^n$, $q \neq 0$, such that*

$$\sup_{x \in cl\ S} q^t x \leq q^t y.$$

In geometric terms, $H := \{x \in \mathbb{R}^n : q^t x = q^t y\}$ is said to be a supporting hyperplane *for S at y: the hyperplane H contains the point y and the set S (as well as* cl S*) is contained the halfspace $\{x \in \mathbb{R}^n : p^t x \leq p^t y\}$.*

PROOF. Let $Z := \text{cl } S$; then $\partial S \subset \partial Z$ (exercise). Of course, Z is closed and it is easy to show that Z is convex (use limit arguments). So there exists a sequence (y_k) in $\mathbb{R}^n \backslash Z$ such that $y_k \to y$. By Theorem A.2 there exists for every k a nonzero vector $p_k \in \mathbb{R}^n$ such that

$$\sup_{x \in Z} p_k^t x < p_k^t y_k.$$

Division by $|p_k|$ turns this into

$$\sup_{x \in Z} q_k^t x < q_k^t y_k,$$

where $q_k := p_k/|p_k|$ belongs to the unit sphere of $\mathbb{R}^n$. This sphere is compact (Bolzano-Weierstrass theorem), so we can suppose without loss of generality that (q_k) converges to some q, $|q| = 1$ (so q is nonzero). Now for every $x \in Z$ the inequality $q_k^t x < q_k^t y_k$, which holds for all k, implies

$$q^t x = \lim_k q_k^t x \leq \lim_k q_k^t y_k = q^t y,$$

and the proof is finished. QED

Theorem A.4 (set-set separation [1, Thm. 2.4.8]) *Let S_1, S_2 be two nonempty convex sets in $\mathbb{R}^n$ such that $S_1 \cap S_2 = \emptyset$. Then there exist $p \in \mathbb{R}^n$, $p \neq 0$, and $\alpha \in \mathbb{R}$ such that*

$$\sup_{x \in S_1} p^t x \leq \alpha \leq \inf_{y \in S_2} p^t y.$$

In geometric terms, $H := \{x \in \mathbb{R}^n : p^t x = \alpha\}$ is said to be a separating hyperplane *for S_1 and S_2: each of the two convex sets is contained in precisely one of the two halfspaces $\{x \in \mathbb{R}^n : p^t x \leq \alpha\}$ and $\{x \in \mathbb{R}^n : p^t x \geq \alpha\}$.*

PROOF. It is easy to see that $S := S_1 - S_2$ is convex. Now $0 \notin S$, for otherwise we get an immediate contradiction to $S_1 \cap S_2 = \emptyset$. W distinguish now two cases: (i) $0 \in \text{cl } S$ and (ii) $0 \notin \text{cl } S$.

In case (i) we have $0 \in \partial S$, so by Theorem A.3 we then have the existence of a nonzero $p \in \mathbb{R}^n$ such that

$$p^t z \leq 0 \text{ for every } z \in S = S_1 - S_2, \tag{2}$$

i.e., for every $z = x - y$, with $x \in S_1$ and $y \in S_2$. This gives $p^t x \leq p^t y$ for all $x \in S_1$ and $y \in S_2$, whence the result.

In case (ii) we apply Theorem A.2 to get immediately (2) as well. The result follows just as in case (i). QED

Theorem A.5 (strong set-set separation [1, Thm. 2.4.10]) *Let S_1, S_2 be two nonempty closed convex sets in $\mathbb{R}^n$ such that $S_1 \cap S_2 = \emptyset$ and such that S_1 is bounded. Then there exist $p \in \mathbb{R}^n$, $p \neq 0$, and $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$ such that*

$$\sup_{x \in S_1} p^t x \leq \alpha < \beta \leq \inf_{y \in S_2} p^t y.$$

PROOF. As in the previous proof, it is easy to see that $S := S_1 - S_2$ is convex. Now S is also seen to be closed (exercise). As in the previous proof, we have $0 \notin S$. We can now apply Theorem A.2 to get the desired result, just as in case (ii) of the previous proof. QED

B Fenchel conjugation

Definition B.1 For a function $f : \mathbb{R}^n \to (-\infty, +\infty]$ the *(Fenchel) conjugate* function of f is $f^* : \mathbb{R}^n \to [-\infty, +\infty]$, given by

$$f^*(\xi) := \sup_{x \in \mathbb{R}^n} [\xi^t x - f(x)].$$

By repeating the conjugation operation one also defines the *(Fenchel) biconjugate* of f, which is simply given by $f^{**} := (f^*)^*$.

Example B.2 Consider $f : \mathbb{R} \to \mathbb{R}$, given by

$$f(x) := \begin{cases} x \log x & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ +\infty & \text{if } x < 0. \end{cases}$$

Observe that this function is convex. Then (counting $0 \log 0$ as 0) we clearly have $f^*(\xi) = \sup_{x \geq 0} \xi x - x \log x$ for the conjugate. For an *interior* maximum in $\mathbb{R}_+$ (by concavity of the function to be maximized) the necessary and sufficient condition is $\xi - \log x - 1 = 0$, i.e., $x = \exp(\xi - 1)$, which gives the value $\xi x - x \log x = \exp(\xi - 1)$. Since this value is positive, we conclude that the point $x = 0$ stands no chance for the maximum, i.e., the maximum is always interior, as calculated above, giving $f^*(\xi) = \exp(\xi - 1)$ for the conjugate function. We can also determine the biconjugate function: by definition, $f^{**}(x) = \sup_{\xi \in \mathbb{R}} x\xi - \exp(\xi - 1)$. If $x < 0$, then, by $\exp(\xi - 1) \to 0$ as $\xi \to -\infty$, the supremum value is clearly $+\infty$. Hence, $f^{**}(x) = +\infty$ for $x < 0$. If $x > 0$, then setting the derivative of the concave function $\xi \mapsto x\xi - \exp(\xi - 1)$ equal to zero gives a solution (whence a global maximum) for $\xi = \log x + 1$. Hence $f^{**}(x) = x \log x$ for $x > 0$. Finally, if $x = 0$, then the supremum of $-\exp(\xi - 1)$ is clearly the limit value 0. So $f^{**}(0) = 0$. We conclude that $f^{**} = f$ in this example. The Fenchel-Moreau theorem below will support this observation.

Exercise B.1 Determine for each of the following functions f the conjugate function f^* and verify also explicitly if $f = f^{**}$ holds.
a. $f(x) = ax^2 + bx + c$, $a \geq 0$,
b. $f(x) = |x| + |x - 1|$,
c. $f(x) = x^a/a$ for $x \geq 0$ and $f(x) = +\infty$ for $x < 0$ (here $a \geq 1$).
d. $f = \chi_B$, where B is the closed unit ball in $\mathbb{R}^n$.

Example B.3 Let K be a nonempty convex cone in $\mathbb{R}^n$ (recall that a *cone* (at zero) is a set such that $\alpha x \in K$ for every $\alpha > 0$ and $x \in K$; cf. Definition 2.5.1 in [1]). Let $f := \chi_K$. Then

$$f^*(\xi) = \sup_{x \in K} \xi^t x = \begin{cases} 0 & \text{if } \xi \in K^*, \\ +\infty & \text{otherwise.} \end{cases}$$

Recall here that K^*, the *polar cone* of K, is defined by $K^* := \{\xi \in \mathbb{R}^n : \xi^t x \leq 0 \text{ for all } x \in K\}$. Hence, we conclude that $(\chi_K)^* = \chi_{K^*}$.

Denote the closure of K by $\bar{K}$. We also observe that $\xi \in \partial \chi_{\bar{K}}(0)$ is equivalent to $\xi^t x \leq 0$ for all $x \in \bar{K}$, i.e., to $\xi^t x \leq 0$ for all $x \in K$, i.e., to $\xi \in K^*$.

Proposition B.4 *Let* f, $g : \mathbb{R}^n \to (-\infty, +\infty]$.

(*i*) *If* $f \geq g$ *then* $f^* \leq g^*$.

(*ii*) *If* $f^*(x) = -\infty$ *for some* $x \in \mathbb{R}^n$, *then* $f \equiv +\infty$.

(*iii*) *For every* x_0, $\xi \in \mathbb{R}^n$

$$f^*(\xi) \geq \xi^t x_0 - f(x_0) \text{ (Young's inequality).}$$

(*iv*) $f \geq f^{**}$.

(*v*) *For every* x_0, $\xi \in \mathbb{R}^n$

$$f^*(\xi) = \xi^t x_0 - f(x_0) \text{ if and only if } \xi \in \partial f(x_0).$$

Exercise B.2 Give a proof of Proposition B.4.

Theorem B.5 (Fenchel-Moreau) *Let* $f : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex. Then*

$$f(x_0) = f^{**}(x_0) \text{ if and only if } f \text{ is lower semicontinuous at } x_0.$$

Proof. One implication is very simple: if $f(x_0) = f^{**}(x_0)$, and if $x_n \to x_0$ then $\liminf_n f(x_n) \geq \liminf_n f^{**}(x_n)$ by Proposition B.4(*iv*). Also, $\liminf_n f^{**}(x_n) \geq f^{**}(x_0)$ because every conjugate, being the supremum of a collection of continuous functions, is automatically lower semicontinuous. So we conclude that $\liminf_n f(x_n) \geq f^{**}(x_0) = f(x_0)$, i.e., f is lower semicontinuous at x_0.

In the converse direction, by Proposition B.4(*iv*) it is enough to prove $f^{**}(x_0) \geq r$ for an arbitrary $r < f(x_0)$, both when $f(x_0) < +\infty$ and when $f(x_0) = +\infty$.

Case 1: $f(x_0) < +\infty$. It is easy to check that $C := \text{epi } f := \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : r \geq f(x)\}$, the epigraph of f, is a convex set in $\mathbb{R}^{n+1}$ (this is Theorem 3.2.2 in [1] – as can be seen immediately from its proof, it continues to hold for functions with values in $(-\infty, +\infty]$ and we know already that this theorem also holds for sets with empty interior). Hence, the closure cl C is also convex. We claim now that $(x_0, r) \notin \text{cl } C$. For suppose (x_0, r) would be the limit of a sequence of points $(x_n, y_n) \in C$. Then $y_n \geq f(x_n)$ for each n, and in the limit this would give $r \geq \liminf_n f(x_n) \geq f(x_0)$ by lower semicontinuity of f at x_0. This contradiction proves that the claim holds. We may now apply separation [1, Theorem 2.4.10]: there exist $\alpha \in \mathbb{R}$ and $p =: (\xi_0, \mu) \neq (0, 0)$, with $\xi_0 \in \mathbb{R}^n$ and $\mu \in \mathbb{R}$, such that

$$\xi_0^t x + \mu y \leq \alpha < \xi_0^t x_0 + \mu r \text{ for all } (x, y) \in C. \tag{3}$$

It is clear that $\mu \leq 0$ by the definition of C. Also, it is obvious that $\mu \neq 0$ (just consider what happens if we take $(x, y) = (x_0, f(x_0))$ in (3) – and we may do this by virtue of $f(x_0) \in \mathbb{R}$). Hence, we can divide by $-\mu$ in (3) and get

$$\xi_1^t x - f(x) \leq \xi_1^t x_0 - r \text{ for all } x \in \text{dom } f.$$

Notice that this inequality continues to hold outside dom f as well; thus, $f^*(\xi_1) \leq \xi_1^t x_0 - r$, which implies the desired inequality $f^{**}(x_0) \geq r$.

Case 2a: $f \equiv +\infty$. In this case, the desired result is trivial, for $f^* \equiv -\infty$, so $f^{**} \equiv +\infty$.

Case 2b: $f(x_1) < +\infty$ for some $x_1 \in \mathbb{R}^n$. We can repeat the proof of Case 1 until (3). If μ happens to be nonzero, then of course we finish as in Case 1. However, if $\mu = 0$ we only get

$$\xi_0^t x \leq \alpha < \xi_0^t x_0 \text{ for all } x \in \text{dom } f$$

from (3). We then repeat the full proof of Case 1, but with x_0 replaced by x_1 and r by $f(x_1) - 1$. This gives the existence of $\xi \in \mathbb{R}^n$ such that

$$\xi^t x - f(x) \leq \xi^t x_1 - f(x_1) + 1 \text{ for all } x \in \text{dom } f.$$

Now for any $\lambda > 0$, observe that by the two previous inequalities

$$f(x) \geq (\xi + \lambda \xi_0)^t x - \xi^t x_1 + f(x_1) - 1 - \alpha\lambda \text{ for all } x \in \mathbb{R}^n,$$

which implies $f^*(\xi + \lambda\xi_0) \leq \xi^t x_1 - f(x_1) + 1 + \lambda\alpha$. By definition of $f^{**}(x_0)$, this gives

$$f^{**}(x_0) \geq \lambda(\xi_0^t x_0 - \alpha) + \xi^t x_0 - \xi^t x_1 + f(x_1) - 1,$$

which implies $f^{**}(x_0) = +\infty$, by letting λ go to infinity (note that $\xi_0^t x_0 - \alpha > 0$ by the above). QED

Corollary B.6 (bipolar theorem for cones) *Let K be a closed convex cone in $\mathbb{R}^n$. Then $K = K^{**} := (K^*)^*$.*

PROOF. Observe that $f := \chi_K$ is a lower semicontinuous convex function. Hence, $f^{**} = f$ by Theorem B.5. By Example B.3 we know that $f^* = \chi_{K^*}$, so $f^{**} = \chi_{K^{**}}$ follows by another application of this fact. Hence $\chi_K = \chi_{K^{**}}$. QED

Exercise B.3 Prove Farkas' theorem (see Exercise 3.5) by means of Corollary B.6.

Exercise B.4 Redo Exercise 3.3 by making it a special case of Corollary B.6.

References

[1] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms.* Wiley, New York, 1993.

[2] J. van Tiel, *Convex Analysis: An Introductory Text.* Wiley, 1984.