Convex Analysis[1]

Patrick Cheridito

Princeton University, Spring 2013

[1]Many thanks to Andreas Hamel for providing his lecture notes. Large parts of chapters 2-4 are based on them.

Contents

Chapter 1

Convex Analysis in $\mathbb{R}^d$

The following notation is used:

- $d \in \mathbb{N} := \{1, 2, \ldots\}$
- e_i is the i-th unit vector in $\mathbb{R}^d$
- $\langle x, y \rangle := \sum_{i=1}^d x_i y_i$ for $x, y \in \mathbb{R}^d$
- $||x|| := \sqrt{\langle x, x \rangle}$ for $x \in \mathbb{R}^d$
- $B_\varepsilon(x) := \left\{ y \in \mathbb{R}^d : ||x - y|| \leq \varepsilon \right\}$
- $\mathbb{R}_+ := \{x \in \mathbb{R} : x \geq 0\}$, $\mathbb{R}_{++} := \{x \in \mathbb{R} : x > 0\}$
- $x \vee y := \max\{x, y\}$ and $x \wedge y := \min\{x, y\}$ for $x, y \in \mathbb{R}$

1.1 Subspaces, affine sets, convex sets, cones and half-spaces

Definition 1.1.1 *Let C be a subset of $\mathbb{R}^d$. C is a subspace of $\mathbb{R}^d$ if*

$$\lambda x + y \in C \quad \text{for all } x, y \in C \text{ and } \lambda \in \mathbb{R}.$$

C is an affine set if

$$\lambda x + (1 - \lambda) y \in C \quad \text{for all } x, y \in C \text{ and } \lambda \in \mathbb{R}.$$

C is a convex set if

$$\lambda x + (1 - \lambda) y \in C \quad \text{for all } x, y \in C \text{ and } \lambda \in [0, 1].$$

C is a cone if

$$\lambda x \in C \quad \text{for all } x \in C \text{ and } \lambda \in \mathbb{R}_{++}.$$

Exercise 1.1.2 Let C, D be non-empty subsets of $\mathbb{R}^d$.

1. Show that if C, D are subspaces, then so is

$$C - D := \{x - y : x \in C, y \in D\},$$

and the same is true for affine sets, convex sets and cones.

2. Show that if C is affine, then $C + v$ is affine for every $v \in \mathbb{R}^k$.

3. Show that if C is affine and contains 0, it is a subspace.

4. Show that if C is affine and $v \in C$, then $C - v = C - C$ is a subspace.

5. Show that the intersection of arbitrarily many subspaces is a subspace, and that the same is true for affine subsets, convex subsets and cones.

6. Show that there exists a smallest subspace containing C, and that the same is true for affine sets, convex sets and cones.

Definition 1.1.3 *If C is a non-empty subset of $\mathbb{R}^d$, we denote by* $\operatorname{lin} C$*,* $\operatorname{aff} C$*,* $\operatorname{conv} C$*,* $\operatorname{cone} C$ *the smallest subspace, affine set, convex set, cone containing C, respectively.*

Exercise 1.1.4 Let C be a non-empty subset of $\mathbb{R}^d$. Show that

$$\begin{aligned}
\operatorname{lin} C &= \left\{ \sum_{i=1}^{n} \lambda_i x_i : n \in \mathbb{N},\ \lambda_i \in \mathbb{R},\ x_i \in C \right\} \\
\operatorname{aff} C &= \left\{ \sum_{i=1}^{n} \lambda_i x_i : n \in \mathbb{N},\ \lambda_i \in \mathbb{R},\ x_i \in C,\ \sum_{i=1}^{n} \lambda_i = 1 \right\} \\
\operatorname{conv} C &= \left\{ \sum_{i=1}^{n} \lambda_i x_i : n \in \mathbb{N},\ \lambda_i \in \mathbb{R}_+,\ x_i \in C,\ \sum_{i=1}^{n} \lambda_i = 1 \right\} \\
\operatorname{cone} C &= \{\lambda x :\ \lambda \in \mathbb{R}_{++},\ x \in C\}
\end{aligned}$$

Definition 1.1.5 *The dimension of an affine subset M of $\mathbb{R}^d$ is the dimension of the subspace $M - M$. The dimension of an arbitrary subset C is the dimension of* $\operatorname{aff} C$*.*

Definition 1.1.6 *Let C be a non-empty subset of $\mathbb{R}^d$. The dual cone of C is the set*

$$C^* := \left\{ z \in \mathbb{R}^d : \langle x, z \rangle \geq 0 \text{ for all } x \in C \right\}.$$

Exercise 1.1.7 Show that the dual cone C^* of a non-empty subset $C \subseteq \mathbb{R}^d$ is a closed convex cone and C is contained in C^{**}.

Definition 1.1.8 *The recession cone 0^+C of a subset C of $\mathbb{R}^d$ consists of all $y \in \mathbb{R}$ satisfying*

$$x + \lambda y \in C \quad \text{for all } x \in C \text{ and } \lambda \in \mathbb{R}_{++}.$$

Every $y \in 0^+C \setminus \{0\}$ is called a direction of recession for C.

Definition 1.1.9 *Let C be a subset of $\mathbb{R}^d$. The closure* $\mathrm{cl}\, C$ *of C is the smallest closed subset of $\mathbb{R}^d$ containing C. The interior* $\mathrm{int}\, C$ *consists of all $x \in C$ such that $B_\varepsilon(x) \subseteq C$ for some $\varepsilon \in \mathbb{R}_{++}$. The relative interior* $\mathrm{ri}\, C$ *is the set of all $x \in C$ such that $B_\varepsilon(x) \cap \mathrm{aff}\, C \subseteq C$ for some $\varepsilon \in \mathbb{R}_{++}$. The boundary of C is the set* $\mathrm{bd}\, C := \mathrm{cl}\, C \setminus \mathrm{int}\, C$. *The relative boundary is* $\mathrm{rbd}\, C := \mathrm{cl}\, C \setminus \mathrm{ri}\, C$

Exercise 1.1.10

1. Show that an affine subset of $\mathbb{R}^d$ is closed.

2. Show that the closure of a cone is a cone.

3. Show that the closure of a convex set is convex.

Lemma 1.1.11 *Let C be a non-empty convex subset of $\mathbb{R}^d$ and $\lambda \in (0,1]$. If $\mathrm{int}\, C \neq \emptyset$, then*

$$\lambda \mathrm{int}\, C + (1-\lambda)\mathrm{cl}\, C \subseteq \mathrm{int}\, C. \tag{1.1.1}$$

If $\mathrm{ri}\, C \neq \emptyset$, then

$$\lambda \mathrm{ri}\, C + (1-\lambda)\mathrm{cl}\, C \subseteq \mathrm{ri}\, C \tag{1.1.2}$$

In particular, $\mathrm{int}\, C$ *and* $\mathrm{ri}\, C$ *are convex.*

Proof. Let $x \in \mathrm{int}\, C$, $y \in \mathrm{cl}\, C$ and $\lambda \in (0,1]$. There exists $\varepsilon > 0$ such that $B_{2\varepsilon}(x) \subseteq C$ and $z \in C$ such that $(1-\lambda)||y - z|| \leq \lambda\varepsilon$. Choose $v \in B_{\lambda\varepsilon}(0)$. Then

$$w = \frac{v}{\lambda} + \frac{1-\lambda}{\lambda}(y - z) \in B_{2\varepsilon}(0),$$

and therefore,

$$\lambda x + (1-\lambda)y + v = \lambda(x + w) + (1-\lambda)z \in C.$$

This shows (1.1.1). (1.1.2) follows by working in $\mathrm{aff}\, C$ instead of $\mathbb{R}^d$. □

Lemma 1.1.12 *Let C be a convex subset of $\mathbb{R}^d$. Then $\mathrm{int}\, C \neq \emptyset$ if and only if* $\mathrm{aff}\ C = \mathbb{R}^d$.

Proof. If $x \in \mathrm{int}\, C$, then $0 \in \mathrm{int}\, C - x$, and it follows that

$$\mathrm{aff}\,(C) - x = \mathrm{aff}\,(C - x) = \mathrm{lin}\,(C - x) = \mathbb{R}^d.$$

On the other hand, if $\mathrm{aff}\ C = \mathbb{R}^d$, choose $x \in C$. Then

$$\mathrm{lin}\,(C - x) = \mathrm{aff}\,(C - x) = \mathrm{aff}\,(C) - x = \mathbb{R}^d.$$

So there there exist d vectors $x_1, \ldots, x_d$ in C such that $v_i := x_i - x$ are linearly independent. Since C is convex, one has

$$\frac{1}{d+1}(x + x_1 + \cdots + x_d) + \lambda v_i \in C \quad \text{for } |\lambda| \leq \frac{1}{d+1} \quad \text{and } i = 1, \ldots, d,$$

and therefore,

$$\frac{1}{d+1}(x + x_1 + \cdots + x_d) + V \subseteq C,$$

where $V := \left\{\sum_{i=1}^d \lambda_i v_i : \sum_{i=1}^d |\lambda_i| \leq \frac{1}{d+1}\right\}$. So since $n(x) := \sum_{i=1}^d |\lambda_i|$ for $x = \sum_{i=1}^d \lambda_i v_i$ defines a norm and all norms on $\mathbb{R}^d$ are equivalent, there exists an $\varepsilon > 0$ such that

$$\frac{1}{d+1}(x + x_1 + \cdots + x_d) + B_\varepsilon(0) \subseteq C.$$

□

Corollary 1.1.13 *Let C be a non-empty convex subset of $\mathbb{R}^d$. Then $\operatorname{ri} C$ is dense in C. In particular, $\operatorname{ri} C$ is non-empty.*

Proof. If C consists of only one point x_0, then $\operatorname{ri} C = C = \{x_0\}$. If C contains at least two different points, one can, by shifting, assume that one of them is 0. Then $\operatorname{lin} C = \operatorname{aff} C$ is at least one-dimensional. So by restricting to $\operatorname{lin} C$, one can assume that $\operatorname{lin} C = \mathbb{R}^d$. It follows from Lemma 1.1.12 that $\operatorname{ri} C \neq \emptyset$. Now the corollary follows from Lemma 1.1.11. □

Definition 1.1.14 *A half-space in $\mathbb{R}^d$ is a set of the form*

$$\left\{x \in \mathbb{R}^d : \langle x, z \rangle \geq c\right\} \quad \text{for some } z \in \mathbb{R}^d \setminus \{0\} \text{ and } c \in \mathbb{R}.$$

We say a subset C of $\mathbb{R}^d$ is supported at $x_0 \in C$ by $z \in \mathbb{R}^d \setminus \{0\}$ if $\langle x_0, z \rangle = \inf_{x \in C} \langle x, z \rangle$.

Note that if a subset C of $\mathbb{R}^d$ is supported at $x_0 \in C$ by some $z \in \mathbb{R}^d \setminus \{0\}$, then x_0 is in the boundary of C and C is contained in the half-space

$$\left\{x \in \mathbb{R}^d : \langle x, z \rangle \geq \langle x_0, z \rangle\right\}.$$

1.2 Separation results in finite dimensions

Lemma 1.2.1 *Let C be a non-empty closed subset of $\mathbb{R}^d$. Then there exists $x_0 \in C$ such that*

$$||x_0|| = \inf_{x \in C} ||x||.$$

If in addition, C is convex, then x_0 is unique.

Proof. For fixed $y \in C$, the set $\{x \in C : ||x|| \leq ||y||\}$ is closed and bounded. So the existence of x_0 follows because the norm is continuous. If C is convex and x_0, x_1 are two different norm minimizers, one has

$$||\frac{x_0 + x_1}{2}|| < ||x_0|| = ||x_1||,$$

a contradiction. □

Theorem 1.2.2 (Strong separation)
Let C, D be non-empty convex subsets of $\mathbb{R}^d$. Then there exists $z \in \mathbb{R}^d$ satisfying

$$\inf_{x \in C} \langle x, z \rangle > \sup_{y \in D} \langle y, z \rangle \tag{1.2.3}$$

if and only if $0 \notin \mathrm{cl}\,(C - D)$.

Proof. The "only if" direction is clear. On the other hand, if $0 \notin \mathrm{cl}\,(C - D)$, the unique norm minimizer $z \in \mathrm{cl}\,(C - D)$ is different from zero. For all $w \in C - D$ and $\lambda \in (0, 1]$, one has

$$||z||^2 \leq ||(1-\lambda)z + \lambda w||^2 = ||z||^2 + 2\lambda \langle w - z, z \rangle + \lambda^2 ||w - z||^2.$$

By dividing by λ and sending λ to 0, one obtains

$$\langle w, z \rangle \geq ||z||^2 > 0 \quad \text{for all } w \in C - D.$$

This proves (1.2.3). □

Lemma 1.2.3 *Let C and D be two non-empty closed convex sets with no common direction of recession. Then $C - D$ is closed.*

Proof. Let (x_n) be a sequence in C and (y_n) a sequence in D such that $x_n - y_n \to w \in \mathbb{R}^d$. If (x_n) is unbounded, one can pass to a subsequence such that $||x_n|| \to \infty$ and

$$\frac{x_n}{||x_n||} \to \bar{x} \quad \text{for some } \bar{x} \in S^{d-1} := \left\{ x \in \mathbb{R}^d : ||x|| = 1 \right\}.$$

But then one has for all $x_0 \in C$ and $\lambda \in \mathbb{R}_{++}$,

$$x_0 + \frac{\lambda}{||x_n||}(x_n - x_0) \to x_0 + \lambda \bar{x} \in C$$

since C is closed. This shows that $\bar{x} \in 0^+ C$. However,

$$\lim_n \frac{y_n}{||y_n||} = \lim_n \frac{x_n - w}{||x_n|| + (||y_n|| - ||x_n||)} = \bar{x},$$

and it follows as above that $\bar{x} \in 0^+ D$, a contradiction. So (x_n) and (y_n) must both be bounded. After passing to subsequences, one has $x_n \to x \in C$ and $y_n \to y \in D$. So $w = x - y \in C - D$. □

Corollary 1.2.4 *If C, D are non-empty closed convex disjoint subsets of $\mathbb{R}^d$ with no common direction of recession, there exists $z \in \mathbb{R}^d$ such that*

$$\inf_{x \in C} \langle x, z \rangle > \sup_{y \in D} \langle y, z \rangle.$$

Proof. By Lemma 1.2.3, $C - D$ is closed and does not contain 0. So the corollary follows from Theorem 1.2.2. □

Corollary 1.2.5 *If C, D are non-empty closed convex disjoint subsets of $\mathbb{R}^d$ such that D is bounded, there exists $z \in \mathbb{R}^d$ such that*

$$\inf_{x \in C} \langle x, z \rangle > \sup_{y \in D} \langle y, z \rangle .$$

Proof. D has no direction of recession. So the corollary follows from Corollary 1.2.4. □

Corollary 1.2.6 *Every proper closed convex subset of $\mathbb{R}^d$ is equal to the intersection of all half-spaces containing it.*

Proof. Consider a closed convex subset $C \subsetneq \mathbb{R}^d$. It is clear that C is contained in the intersection of all half-spaces enveloping it. On the other hand, if $x_0 \in \mathbb{R}^d \setminus C$, it follows from Corollary 1.2.5 that there exists a half-space containing C but not x_0. This proves the corollary. □

Corollary 1.2.7 *Let C be a non-empty subset of $\mathbb{R}^d$. Then C^{**} is equal to the smallest closed convex cone containing C.*

Proof. Since C^{**} contains C, it also contains the smallest closed convex cone D enveloping C. To show $C^{**} = D$, assume that there exists $x_0 \in C^{**} \setminus D$. But then it follows from Corollary 1.2.5 that there exists a $z \in \mathbb{R}^d$ such that

$$\inf_{x \in D} \langle x, z \rangle > \langle x_0, z \rangle .$$

Since D is a cone, this implies

$$\inf_{x \in D} \langle x, z \rangle = 0 > \langle x_0, z \rangle ,$$

from which one obtains that $z \in C^*$ and $x_0 \notin C^{**}$, a contradiction. □

Lemma 1.2.8 *Let C be a non-empty convex cone in $\mathbb{R}^d$ such that $C \neq \mathbb{R}^d$. Then there exists $z \in \mathbb{R}^d \setminus \{0\}$ such that*

$$\inf_{x \in C} \langle x, z \rangle = 0. \qquad (1.2.4)$$

Proof. If $\operatorname{int} C = \emptyset$, it follows from Lemma 1.1.12, that $M = \operatorname{aff} C$ is different from $\mathbb{R}^d$. Since C is a cone, M contains 0. Therefore, it is a proper subspace of $\mathbb{R}^d$, and one can choose $z \in M^{\perp}$.

If there exists $x_0 \in \operatorname{int} C$, $-x_0$ cannot be in $\operatorname{cl} C$. Otherwise, it would follow from Corollary 1.1.11 that $0 \in \operatorname{int} C$, implying $C = \mathbb{R}^d$. So one obtains from Corollary 1.2.5 that there exists $z \in \mathbb{R}^d$ such that

$$\inf_{x \in C} \langle x, z \rangle \geq \inf_{x \in \operatorname{cl} C} \langle x, z \rangle > \langle -x_0, z \rangle .$$

This implies (1.2.4) and $z \neq 0$. □

Theorem 1.2.9 (Weak separation)
Let C, D be non-empty convex subsets of $\mathbb{R}^d$. Then there exists $z \in \mathbb{R}^d \setminus \{0\}$ such that

$$\inf_{x \in C} \langle x, z \rangle \geq \sup_{y \in D} \langle y, z \rangle \tag{1.2.5}$$

if and only if $0 \notin \operatorname{int}(C - D)$.

Proof. The "only if" direction is clear. To show the other direction, let us assume $0 \notin \operatorname{int}(C - D)$. If we can show that

$$\operatorname{cone}(C - D) \neq \mathbb{R}^d, \tag{1.2.6}$$

we obtain from Lemma 1.2.8 the existence of a $z \in \mathbb{R}^d \setminus \{0\}$ such that

$$\inf_{x \in \operatorname{cone}(C-D)} \langle x, z \rangle \geq 0,$$

which implies (1.2.5). To prove (1.2.6), we assume by way of contradiction that $\operatorname{cone}(C - D) = \mathbb{R}^d$. But then there exists $\varepsilon > 0$ such that all the vectors $\pm\varepsilon e_i$, $i = 1, \ldots, d$, are in $C - D$. This implies $0 \in \operatorname{int}(C - D)$, contradicting the assumption. So (1.2.6) must hold. □

Corollary 1.2.10 *Let C, D be non-empty convex disjoint subsets of $\mathbb{R}^d$ such that D is open. Then there exists $z \in \mathbb{R}^d$ such that*

$$\inf_{x \in C} \langle x, z \rangle > \langle y, z \rangle \quad \textit{for every } y \in D.$$

Proof. By Theorem 1.2.9, there exists $z \in \mathbb{R}^d \setminus \{0\}$ such that

$$\inf_{x \in C} \langle x, z \rangle \geq \sup_{y \in D} \langle y, z \rangle .$$

Since D is open, the sup is not attained in D, and the corollary follows. □

Corollary 1.2.11 *A convex subset C of $\mathbb{R}^d$ is supported at every point $x_0 \in C \setminus \operatorname{int} C$ by at least one vector $z \in \mathbb{R}^d \setminus \{0\}$.*

Proof. If $x_0 \in C \setminus \operatorname{int} C$, then $0 \notin \operatorname{int}(C - x_0)$. So it follows from Theorem 1.2.9 that there exists $z \in \mathbb{R} \setminus \{0\}$ such that $\inf_{x \in C} \langle x, z \rangle \geq \langle x_0, z \rangle$, proving the corollary. □

Corollary 1.2.12 *Let C be a non-empty convex subset of $\mathbb{R}^d$. Then $\operatorname{int} C = \operatorname{int} \operatorname{cl} C$.*

Proof. It is enough to show that $\operatorname{int} \operatorname{cl} C \subseteq \operatorname{int} C$. To do that we assume $x_0 \notin \operatorname{int} C$. Then it follows from Theorem 1.2.9 that there exists $z \in \mathbb{R}^d \setminus \{0\}$ such that

$$\inf_{x \in C} \langle x, z \rangle \geq \langle x_0, z \rangle .$$

It follows that

$$\inf_{x \in \operatorname{cl} C} \langle x, z \rangle \geq \langle x_0, z \rangle ,$$

which implies $x_0 \notin \operatorname{int} \operatorname{cl} C$. This proves the corollary. □

Corollary 1.2.13 *Let C be a dense convex subset of $\mathbb{R}^d$. Then $C = \mathbb{R}^d$.*

Proof. By Corollary 1.2.12, one has $\operatorname{int} C = \operatorname{int}\operatorname{cl} C = \mathbb{R}^d$. □

Theorem 1.2.14 (Proper separation)
Let C, D be non-empty convex subsets of $\mathbb{R}^d$. Then there exists $z \in \mathbb{R}^d$ satisfying

$$\inf_{x \in C} \langle x, z\rangle \geq \sup_{y \in D} \langle y, z\rangle \quad \textit{and} \quad \sup_{x \in C} \langle x, z\rangle > \inf_{y \in D} \langle y, z\rangle \tag{1.2.7}$$

if and only if $0 \notin \operatorname{ri}(C - D)$.

Proof. To show the "only if" direction, let us assume there exists a $z \in \mathbb{R}^d$ satisfying (1.2.7) and $0 \in \operatorname{ri}(C-D)$. Then the affine hull M of $C-D$ is a subspace. Decompose $z = z_1 + z_2$ such that $z_1 \in M$ and $z_2 \in M^\perp$. Then

$$\inf_{x \in C-D} \langle x, z_1\rangle \geq 0 \quad \text{and} \quad \sup_{x \in C-D} \langle x, z_1\rangle > 0.$$

But this contradicts $0 \in \operatorname{ri}(C - D)$.

To show the "if" direction, assume $0 \notin \operatorname{ri}(C-D)$. If $0 \notin M$, then $0 \notin \operatorname{cl}(C-D)$, and (1.2.7) follows from Theorem 1.2.2. If $0 \in M$, one can without loss of generality assume that $M = \mathbb{R}^d$. But then $0 \notin \operatorname{int}(C - D)$, and one obtains from Theorem 1.2.9 that there exists $z \in \mathbb{R}^d \setminus \{0\}$ such that

$$\inf_{x \in C-D} \langle x, z\rangle \geq 0.$$

Moreover, there must exist an $x \in C - D$ satisfying $\langle x, z\rangle > 0$. Otherwise, one would have $\langle x, z\rangle = 0$ for all $x \in C - D$, contradicting $M = \mathbb{R}^d$. □

1.3 Linear, affine and convex functions

Definition 1.3.1 *A function $f : \mathbb{R}^d \to \mathbb{R}^k$ is linear if*

$$f(\lambda x + y) = \lambda f(x) + f(y) \quad \textit{for all } x, y \in \mathbb{R}^d \textit{ and } \lambda \in \mathbb{R}.$$

f is affine if

$$f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y) \quad \textit{for all } x, y \in \mathbb{R}^d \textit{ and } \lambda \in \mathbb{R}.$$

Exercise 1.3.2 Let $f : \mathbb{R}^d \to \mathbb{R}^k$ be an affine function and $v \in \mathbb{R}^k$.

1. Show that $f + v$ is affine.

2. Show that $f - f(0)$ is linear.

3. Show that $f(x) = Ax + f(0)$ for some $k \times d$-matrix A.

Proposition 1.3.3 *Every affine function $f : \mathbb{R}^d \to \mathbb{R}^k$ is Lipschitz-continuous.*

Proof. It is enough to show that $f - f(0)$ is Lipschitz-continuous. So one can assume that f is linear. But then there exists a $k \times d$-matrix A such that $f(x) = Ax$, and one has

$$||f|| := \sup_{||x|| \le 1} ||f(x)|| \le \sup_{||x|| \le 1} \left(\sum_{i=1}^k \left(\sum_{j=1}^d A_{ij} x_j \right)^2 \right)^{1/2} \le \left(\sum_{ij} A_{ij}^2 \right)^{1/2} .$$

So

$$||f(x) - f(y)|| \le ||f|| \, ||x - y|| \quad \text{for all } x, y \in \mathbb{R}^d.$$

□

Definition 1.3.4 *A function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex if*

$$f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y) \quad \textit{for all } x, y \in \mathbb{R} \textit{ and } \lambda \in (0,1)$$

and quasi-convex if

$$f(\lambda x + (1-\lambda) y) \le f(x) \vee f(y) \quad \textit{for all } x, y \in \mathbb{R} \textit{ and } \lambda \in (0,1).$$

A function $f : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ is (quasi-) concave if $-f$ is (quasi-) convex.

The effective domain of a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ or $f : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ is the set

$$\operatorname{dom} f := \left\{ x \in \mathbb{R}^d : f(x) \in \mathbb{R} \right\}.$$

Exercise 1.3.5 Show that a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is quasi-convex if and only if all the sublevel sets

$$\left\{ x \in \mathbb{R}^d : f(x) \le y \right\}, \quad y \in \mathbb{R},$$

are convex.

Exercise 1.3.6 Let $f, g : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be convex functions and $\lambda > 0$. Show that $\lambda f + g$ is convex.

Definition 1.3.7 *We say a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ is quasi-convex if all sub-level sets $\left\{ x \in \mathbb{R}^d : f(x) \le y \right\}$, $y \in \mathbb{R}$, are convex. We say f is quasi-concave if $-f$ is quasi-convex.*

Exercise 1.3.8
1. Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ be quasi-convex and $h : \mathbb{R} \cup \{\pm\infty\} \to \mathbb{R} \cup \{\pm\infty\}$ non-decreasing. Show that $h \circ f$ is quasi-convex.

2. Give an example of a convex function $f : \mathbb{R}^d \to \mathbb{R}$ and a non-decreasing function $h : \mathbb{R} \to \mathbb{R}$ such that $h \circ f$ is not convex.

3. Let $f_i : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$, $i \in I$, be a family of quasi-convex functions. Show that $\sup_{i \in I} f_i$ is quasi-convex.

Definition 1.3.9 *The epigraph of a function* $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ *is the set*

$$\operatorname{epi} f := \left\{(x, y) \in \mathbb{R}^d \times \mathbb{R} : f(x) \leq y\right\}.$$

The hypograph of f *is given by*

$$\operatorname{hypo} f := \left\{(x, y) \in \mathbb{R}^d \times \mathbb{R} : f(x) \geq y\right\}.$$

Exercise 1.3.10 Show that a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex if and only if $\operatorname{epi} f$ is convex.

Definition 1.3.11 *We say a function* $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ *is convex if* $\operatorname{epi} f$ *is a convex subset of* $\mathbb{R}^{d+1}$. *A convex function* $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ *is said to be proper convex if* $f(x) > -\infty$ *for all* $x \in \mathbb{R}^d$ *and* $f(x) < +\infty$ *for at least one* $x \in \mathbb{R}^d$. *We say* f *is concave if* $-f$ *is convex and proper concave if* $-f$ *is proper convex.*

Exercise 1.3.12

1. Show that for a convex function $f : \mathbb{R} \to \mathbb{R} \cup \{\pm\infty\}$ and $x_0 \in \mathbb{R}$ such that $f(x_0) \in \mathbb{R}$,

$$\frac{f(x_0 + \varepsilon) - f(x_0)}{\varepsilon}$$

is non-decreasing in $\varepsilon \in \mathbb{R} \setminus \{0\}$

2. Show that for a convex function $f : \mathbb{R} \to \mathbb{R} \cup \{\pm\infty\}$ and $x_0 \in \mathbb{R}$ such that $f(x_0) \in \mathbb{R}$,

$$f'_+(x_0) := \lim_{\varepsilon \downarrow 0} \frac{f(x_0 + \varepsilon) - f(x_0)}{\varepsilon} \quad \text{and} \quad f'_-(x_0) := \lim_{\varepsilon \downarrow 0} \frac{f(x_0) - f(x_0 - \varepsilon)}{\varepsilon}$$

exist and $f'_-(x) \leq f'_+(x)$.

3. Let $f_i : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$, $i \in I$, be a family of convex functions. Show that $\sup_{i \in I} f_i$ is convex.

4. Show that every function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ has a greatest convex minorant.

Definition 1.3.13 *We denote the greatest convex minorant of a function* $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ *by* $\operatorname{conv} f$ *and call it the convex hull of* f.

Theorem 1.3.14 *Let* $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ *be a convex function and* $x_0 \in \mathbb{R}^d$ *such that* $f(x_0) \in \mathbb{R}$. *Assume there exists a neighborhood* U *of* x_0 *such that* $f(x) < +\infty$ *for all* $x \in U$. *Then* f *is proper convex and continuous at* x_0.

Proof. There is an $\varepsilon > 0$ such that $m := \max_i f(x_0 \pm \varepsilon e_i) < +\infty$. By convexity, one has $f(x) \leq m$ for all $x \in x_0 + V$, where $V := \left\{x \in \mathbb{R}^d : \sum_{i=1}^d |x_i| \leq \varepsilon\right\}$. Since

$f(x_0) \in \mathbb{R}$ and $x_0 + V$ is a neighborhood of x_0, one obtains $f(x) > -\infty$ for all $x \in \mathbb{R}^d$. In particular, f is proper convex. Now choose $x \in V$ and $0 < \lambda \le 1$. Then

$$f(x_0 + \lambda x) = f(\lambda(x_0 + x) + (1 - \lambda)x_0) \le \lambda f(x_0 + x) + (1 - \lambda)f(x_0),$$

and therefore,

$$f(x_0 + \lambda x) - f(x_0) \le \lambda[f(x_0 + x) - f(x_0)] \le \lambda(m - f(x_0)).$$

On the other hand,

$$x_0 = \frac{1}{1+\lambda}(x_0 + \lambda x) + \frac{\lambda}{1+\lambda}(x_0 - x).$$

So

$$f(x_0) \le \frac{1}{1+\lambda} f(x_0 + \lambda x) + \frac{\lambda}{1+\lambda} f(x_0 - x),$$

from which one obtains

$$f(x_0) - f(x_0 + \lambda x) \le \lambda[f(x_0 - x) - f(x_0)] \le \lambda(m - f(x_0)).$$

Hence, we have shown that

$$|f(x) - f(x_0)| \le \lambda(m - f(x_0)) \quad \text{for all } x \in x_0 + \lambda V,$$

which proves the theorem. □

Corollary 1.3.15 *A convex function* $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *is continuous on* $\operatorname{int} \operatorname{dom} f$.

Proof. If $x_0 \in \operatorname{int} \operatorname{dom} f$, there exists a neighborhood U of x_0 such that $f(x) < +\infty$ for all $x \in U$. Now the corollary follows from Theorem 1.3.14. □

Definition 1.3.16 *A function* $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ *is said to be positively homogeneous if* $f(\lambda x) = \lambda f(x)$ *for all* $x \in \mathbb{R}^d$ *and* $\lambda \in \mathbb{R}_{++}$. *If* f *is convex and positively homogeneous, it is called sub-linear.*

Exercise 1.3.17

1. Show that a positively homogeneous function $f : \mathbb{R}^d \to \mathbb{R}$ satisfies $f(0) = 0$.

2. Show that a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ is positively homogeneous if and only if $\operatorname{epi} f$ is a cone in $\mathbb{R}^{d+1}$.

3. Show that a positively homogeneous function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex if and only if $f(x + y) \le f(x) + f(y)$, $x, y \in \mathbb{R}^d$.

Corollary 1.3.18 (Hahn–Banach extension theorem in finite dimensions)
Let $g : \mathbb{R}^d \to \mathbb{R}$ *be a sub-linear function and* $f : M \to \mathbb{R}$ *a linear function on a subspace* M *of* $\mathbb{R}^d$ *such that* $f(x) \le g(x)$ *for all* $x \in M$. *Then there exists a linear extension* $F : \mathbb{R}^d \to \mathbb{R}$ *of* f *such that* $F(x) \le g(x)$ *for all* $x \in \mathbb{R}^d$.

Proof. $\operatorname{epi} g = \left\{(x,y) \in \mathbb{R}^d \times \mathbb{R} : g(x) \le y\right\}$ is a non-empty convex cone in $\mathbb{R}^{d+1}$ and $\operatorname{graph} f := \{(x, f(x)) : x \in M\}$ a subspace. Since $\operatorname{epi} g - \operatorname{graph} f$ is a cone that does not contain $(0,-1)$, the point $(0,0)$ cannot be in the interior of $\operatorname{epi} g - \operatorname{graph} f$. By Theorem 1.2.9, there exists $(z,v) \in \mathbb{R}^d \times \mathbb{R} \setminus \{0\}$ such that

$$\inf_{(x,y)\in \operatorname{epi} g} (\langle x, z\rangle + yv) \ge \sup_{x \in M} (\langle x, z\rangle + f(x)v).$$

It follows that $v > 0$, and by rescaling, one can assume $v = 1$. Since M is a subspace, one must have $f(x) = \langle x, -z\rangle$, $x \in M$, and therefore, $\langle x, z\rangle + g(x) \ge 0$, $x \in \mathbb{R}^d$. This shows that $F(x) = \langle x, -z\rangle$ has the desired properties. □

1.4 Derivatives, directional derivatives and sub-gradients

Definition 1.4.1 *Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ and $x_0 \in \mathbb{R}^d$ such that $f(x_0) \in \mathbb{R}$. If there exists $z \in \mathbb{R}^d$ such that*

$$\lim_{x \neq 0,\, x \to 0} \frac{f(x_0 + x) - f(x_0) - \langle x, z\rangle}{||x||} = 0,$$

then f is said to be differentiable at x_0 with gradient $\nabla f(x_0) = z$ (or derivative $Df(x_0) = z$).

Definition 1.4.2 *Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ and $x_0 \in \mathbb{R}^d$ such that $f(x_0) \in \mathbb{R}$. If the limit*

$$f'(x_0; x) := \lim_{\varepsilon \downarrow 0} \frac{f(x_0 + \varepsilon x) - f(x_0)}{\varepsilon}$$

exists (it is allowed to be $+\infty$ or $-\infty$), we call it the directional derivative of f at x_0 in the direction x.

Note that if f is differentiable at x_0, then

$$f'(x_0; x) = \langle x, \nabla f(x_0)\rangle .$$

is linear in x.

Definition 1.4.3 *Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ and $x_0 \in \mathbb{R}^d$ such that $f(x_0) \in \mathbb{R}$. $z \in \mathbb{R}^d$ is a sub-gradient of f at x_0 if*

$$f(x_0 + x) - f(x_0) \ge \langle x, z\rangle \quad \textit{for all } x \in \mathbb{R}^d.$$

The set of all sub-gradients of f at x_0 is denoted by $\partial f(x_0)$ and called sub-differential of f at x_0.

Exercise 1.4.4 Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ be a convex function and $x_0 \in \mathbb{R}^d$ such that $f(x_0) \in \mathbb{R}$. Show the following:

1.

$$f'(x_0; x) = \inf_{\varepsilon > 0} \frac{f(x_0 + \varepsilon x) - f(x_0)}{\varepsilon}.$$

In particular, $f'(x_0; x)$ exists for all $x \in \mathbb{R}^d$.

2. $f'(x_0, .)$ is sub-linear.

3. If $x_0 \in \operatorname{int} \left\{x \in \mathbb{R}^d : f(x) \in \mathbb{R}\right\}$, then $f'(x_0; x) \in \mathbb{R}$ for all $x \in \mathbb{R}^d$.

4. The following are equivalent:

(i) $f(x_0) = \min_x f(x)$

(ii) $0 \in \partial f(x_0)$

(iii) $f'(x_0; x) \geq 0$ for all $x \in \mathbb{R}^d$.

5. The sub-differential $\partial f(x_0)$ is a closed convex subset of $\mathbb{R}^d$.

6. $\partial f(x_0) = \partial g(0)$, where $g(x) := f'(x_0; x)$.

7. If f is differentiable at x_0, then $\partial f(x_0) = \{\nabla f(x_0)\}$.

8. The following are equivalent:

(i) $z \in \partial f(x_0)$

(ii) $(-z, 1)$ supports $\operatorname{epi} f$ at $(x_0, f(x_0))$.

Theorem 1.4.5 *Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a convex function and $x_0 \in \operatorname{ri} \operatorname{dom} f$. Then $\partial f(x_0) \neq \emptyset$.*

Proof. Since $(x_0, f(x_0) + 1) \in \operatorname{epi} f$, the point $(x_0, f(x_0))$ is not in $\operatorname{ri} \operatorname{epi} f$. By Theorem 1.2.14, there exists $(z, v) \in \mathbb{R}^d \times \mathbb{R}$ such that

$$\inf_{(x,y) \in \operatorname{epi} f} (\langle x, -z \rangle + vy) \geq \langle x_0, -z \rangle + v f(x_0) \tag{1.4.8}$$

and

$$\sup_{(x,y) \in \operatorname{epi} f} (\langle x, -z \rangle + vy) > \langle x_0, -z \rangle + v f(x_0) \tag{1.4.9}$$

It follows from (1.4.8) that $v \geq 0$. Now assume that $v = 0$. Then, since $x_0 \in \operatorname{ri} \operatorname{dom} f$, (1.4.9) contradicts (1.4.8). So $v > 0$, and by scaling, one can assume $v = 1$. Then $(-z, 1)$ supports $\operatorname{epi} f$ at $(x_0, f(x_0))$, which by Exercise 1.4.4.8, proves that $z \in \partial f(x_0)$. $\square$

Definition 1.4.6 *A function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ is lower semi-continuous (lsc) at $x_0 \in \mathbb{R}^d$ if $f(x_0) \leq \liminf_{x \to x_0} f(x)$. f is lsc if it is lsc everywhere. f is upper semi-continuous (usc) at x_0 if $f(x_0) \geq \limsup_{x \to x_0} f(x)$. f is usc if it is usc everywhere. By $\underline{f}$, we denote the function given by*

$$\underline{f}(x) := \liminf_{y \to x} f(y)$$

and call it lsc hull of f. By $\underline{\mathrm{conv}}\, f$ we denote the lsc hull of $\mathrm{conv}\, f$ and call it lsc convex hull of f.

Exercise 1.4.7
Consider a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$.

1. Show that the following are equivalent:

(i) f is lsc

(ii) All sub-level sets $\left\{x \in \mathbb{R}^d : f(x) \leq y\right\}$, $y \in \mathbb{R}$, are closed

(iii) $\mathrm{epi}\, f$ is closed

2. Show that the epigraph of $\underline{f}$ is the closure of $\mathrm{epi}\, f$ and $\underline{f}$ is the greatest lsc minorant of f.

3. Show that if f is convex, then so is $\underline{f}$.

4. Show that $\underline{\mathrm{conv}}\, f$ is the greatest lsc convex minorant of f.

5. Let $f_i : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$, $i \in I$, be a family of lsc functions. Show that $\sup_{i \in I} f_i$ is lsc.

Lemma 1.4.8 *Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ be a lsc convex function and $x_0 \in \mathbb{R}^d$ such that $f(x_0) \in \mathbb{R}$. Then f is proper convex.*

Proof. Assume there exists $x_1 \in \mathbb{R}^d$ such that $f(x_1) = -\infty$. Then $f(\lambda x_0 + (1 - \lambda)x_1) = -\infty$ for all $\lambda \in [0, 1)$. But since f is lsc, one must have $f(x_0) = -\infty$, a contradiction. □

Lemma 1.4.9 *Let f be a proper convex function on $\mathbb{R}^d$ and $x_0 \in \mathrm{dom}\, f$ such that $\partial f(x_0) \neq \emptyset$. Then $f(x_0) = \underline{f}(x_0)$ and $\partial f(x_0) = \partial \underline{f}(x_0)$.*

Proof. Choose $z \in \partial f(x_0)$. The affine function $g(x) = f(x_0) + \langle x - x_0, z\rangle$ minorizes f and equals f at x_0. So g also minorizes $\underline{f}$ and equals $\underline{f}$ at x_0. This shows $f(x_0) = g(x_0) = \underline{f}(x_0)$ and $\partial f(x_0) \subseteq \partial \underline{f}(x_0)$. $\partial f(x_0) \supseteq \partial \underline{f}(x_0)$ follows since $f(x_0) = \underline{f}(x_0)$ and $f \geq \underline{f}$. □

Corollary 1.4.10 *Let f be a proper convex function on $\mathbb{R}^d$. Then so is $\underline{f}$. Moreover, $f(x) = \underline{f}(x)$ for all $x \in \mathrm{ri}\,\mathrm{dom}\, f \cup (\mathrm{cl}\,\mathrm{dom}\, f)^c$ and $\partial f(x) = \partial \underline{f}(x)$ for all $x \in \mathrm{ri}\,\mathrm{dom}\, f$*

Proof. We already know that $\underline{f}$ is convex, and it is clear that it cannot be identically equal to $+\infty$. By Corollary 1.1.13, $\operatorname{ri}\operatorname{dom} f$ is not empty. Choose $x \in \operatorname{ri}\operatorname{dom} f$. By Theorem 1.4.5, there exists $z \in \partial f(x)$. So one obtains from Lemma 1.4.9 that $f(x) = \underline{f}(x)$ and $\partial f(x) = \partial \underline{f}(x)$, which implies that $\underline{f}$ is proper. Finally, note that $\operatorname{dom} \underline{f} \subseteq \operatorname{cl}\operatorname{dom} f$. So if $x \notin \operatorname{cl}\operatorname{dom} f$, then $f(x) = \underline{f}(x) = +\infty$. □

Theorem 1.4.11 *A lsc convex function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ equals the point-wise supremum of all its affine minorants.*

Proof. If f is constantly equal to $+\infty$, the theorem is clear. So we can assume $\operatorname{dom} f \neq \emptyset$. Choose a pair $(x_0, w) \in \mathbb{R}^d \times \mathbb{R}$ that does not belong to $\operatorname{epi} f$. By Corollary 1.2.5, there exists $(z, v) \in \mathbb{R}^d \times \mathbb{R}$ such that

$$m := \inf_{(x,y) \in \operatorname{epi} f} (\langle x, z \rangle + yv) > \langle x_0, z \rangle + wv.$$

It follows that $v \geq 0$. If $v > 0$, one can scale and assume $v = 1$. Then $m - \langle x, z \rangle$ is an affine minorant of f whose epigraph does not contain (x_0, w). If $v = 0$, set $\lambda := m - \langle x_0, z \rangle > 0$ and choose $x_1 \in \operatorname{dom} f$. Since $(x_1, f(x_1) - 1)$ is not in $\operatorname{epi} f$, there exists $(z', v') \in \mathbb{R}^d \times \mathbb{R}$ such that

$$m' := \inf_{(x,y) \in \operatorname{epi} f} (\langle x, z' \rangle + yv') > \langle x_1, z' \rangle + (f(x_1) - 1)v'.$$

Since $x_1 \in \operatorname{dom} f$, one must have $v' > 0$. So by scaling, one can assume $v' = 1$. Now choose

$$\delta > \frac{1}{\lambda}(w + \langle x_0, z' \rangle - m')^+$$

and set $z'' := \delta z + z'$. Then

$$\begin{aligned} m'' &:= \inf_{(x,y) \in \operatorname{epi} f} (\langle x, z'' \rangle + y) \geq \delta m + m' \\ &= \delta \lambda + \delta \langle x_0, z \rangle + m' > \langle x_0, z'' \rangle + w. \end{aligned}$$

So $m'' - \langle x, z'' \rangle$ is an affine minorant of f whose epigraph does not contain (x_0, w). This completes the proof of the theorem. □

1.5 Convex conjugates

Definition 1.5.1 *The convex conjugate of a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ is the function $f^* : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ given by*

$$f^*(z) := \sup_{x \in \mathbb{R}^d} \{\langle x, z \rangle - f(x)\}.$$

Exercise 1.5.2
Consider functions $f, g : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$. Show that ...

1. f^* is convex and lsc.

2. $f \geq f^{**}$

3. $f \leq g$ implies $f^* \geq g^*$

4. $f^{***} = f^*$.

Exercise 1.5.3 Calculate f^* in the cases

1. $f(x) = \sum_{i=1}^d |x_i|^p$ for $p \geq 1$

2. $f(x) = \exp(\lambda x)$ for $\lambda \in \mathbb{R}$

Definition 1.5.4 *Let C be a subset of $\mathbb{R}^d$. The indicator function $\delta_C : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is defined to be 0 on C and $+\infty$ outside of C. The convex conjugate δ_C^* is called support function of C.*

Exercise 1.5.5 Let $f : \mathbb{R}^d \to \mathbb{R}$ be an affine function of the form $f(x) = \langle x, z \rangle - v$ for a pair $(z, v) \in \mathbb{R}^d \times \mathbb{R}$. Show that $f^* = v + \delta_z$ and $f^{**} = f$.

Exercise 1.5.6 Consider a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$.

1. Show that the Young–Fenchel inequality holds:

$$f^*(z) \geq \langle x, z \rangle - f(x) \quad \text{for all } x, z \in \mathbb{R}^d.$$

2. Show that if $f(x_0) \in \mathbb{R}$, the following are equivalent

(i) $z \in \partial f(x_0)$

(ii) $\langle x, z \rangle - f(x)$ achieves its supremum in x at $x = x_0$

(iii) $f(x_0) + f^*(z) = \langle x_0, z \rangle$

3. Show that if $f(x_0) = f^{**}(x_0) \in \mathbb{R}$, the following conditions are equivalent to (i)–(iii)

(iv) $x_0 \in \partial f^*(z)$

(v) $\langle x_0, v \rangle - f^*(v)$ achieves its supremum in v at $v = z$

(vi) $z \in \partial f^{**}(x_0)$

Theorem 1.5.7 (Fenchel–Moreau Theorem)
Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a function whose lsc convex hull $\underline{\text{conv}}\, f$ does not take the value $-\infty$. Then $\underline{\text{conv}}\, f = f^{**}$. In particular, if f is lsc and convex, then $f = f^{**}$.

Proof. We know that $f \geq f^{**}$. Since f^{**} is lsc and convex, one has $\underline{\text{conv}}\, f \geq f^{**}$. Now let h be an affine minorant of $\underline{\text{conv}}\, f$. Then it is also an affine minorant of f. So one has $h = h^{**} \leq f^{**}$. Since by Theorem 1.4.11, $\underline{\text{conv}}\, f$ is the point-wise supremum of its affine minorants, it follows that $\underline{\text{conv}}\, f \leq f^{**}$. □

Corollary 1.5.8 *If f is a proper convex function on $\mathbb{R}^d$, then f^* is lsc proper convex.*

Proof. f^* is lsc convex for every function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$. If f is proper convex, one obtains from Corollary 1.4.10 that so is $\underline{f}$, and it follows from Theorem 1.5.7 that $\underline{f} = f^{**}$. This implies that f^* is proper convex. □

Corollary 1.5.9 *Let C be a non-empty subset of $\mathbb{R}^d$ with closed convex hull D. Then $\delta^*_C(z) = \sup_{x \in D} \langle x, z \rangle$ and $\delta^{**}_C = \delta_D$.*

Proof. $\delta^{**}_C = \delta_D$ follows from Theorem 1.5.7 since δ_D is the lsc convex hull of δ_C. Now one obtains $\delta^*_C = \delta^{***}_C = \delta^*_D$, and the proof is complete. □

Corollary 1.5.10 *Let f be a lsc proper sub-linear function on $\mathbb{R}^d$. Then $f = \delta^*_{\partial f(0)}$ and $f^* = \delta_{\partial f(0)}$. In particular, $f(0) = 0$ and $\partial f(0) \neq \emptyset$.*

Proof. It can easily be checked that $f^* = \delta_C$ for the set

$$C = \left\{ z \in \mathbb{R}^d : \langle x, z \rangle \leq f(x) \text{ for all } x \in \mathbb{R}^d \right\}.$$

By Theorem 1.5.7, one has $f = \delta^*_C$. It follows that C is non-empty, which implies $f(0) = 0$ and $\partial f(0) = C$. □

Exercise 1.5.11 Calculate f^* for

$$f(x) = ||x||_p := \left(\sum_{i=1}^d |x_i|^p \right)^{1/p} \quad \text{for } p \geq 1.$$

Corollary 1.5.12 *Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ be a convex function and $x_0 \in \mathbb{R}^d$ such that $f(x_0) \in \mathbb{R}$. Assume there exists a neighborhood U of x_0 and a constant $M \in \mathbb{R}_+$ such that*

$$f(x) - f(x_0) \geq -M||x - x_0|| \quad \textit{for all } x \in U. \tag{1.5.10}$$

Then f has a sub-gradient z at x_0 such that $||z|| \leq M$.

Proof. Denote by $h : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ the lsc hull of the directional derivative $g(x) = f'(x_0; x)$. It follows from (1.5.10) that $h(x) \geq -M||x||$. In particular, $h(0) = 0$. h is a lsc sublinear function satisfying $\partial h(0) \subseteq \partial f(x_0)$. So it is enough to show that h has a sub-gradient z at x_0 such that $||z|| \le M$. It follows from Corollary 1.5.10 that $\partial h(0)$ is non-empty and

$$h(x) = \sup_{z \in \partial h(0)} \langle x, z \rangle .$$

Now assume that $\partial h(0) \cap B_M(0) = \emptyset$. Since $\partial h(0)$ is closed and convex, there exists an x such that

$$h(x) = \sup_{z \in \partial h(0)} \langle x, z \rangle < \inf_{z \in B_M(0)} \langle x, z \rangle = -M||x||,$$

a contradiction. □

Theorem 1.5.13 *Let f be a proper convex function on $\mathbb{R}^d$ and $x_0 \in \operatorname{ri} \operatorname{dom} f$. Then*

$$f'(x_0; x) = \sup_{z \in \partial f(x_0)} \langle x, z \rangle , \quad x \in \mathbb{R}^d. \tag{1.5.11}$$

Proof. Consider the sub-linear function $g(x) = f'(x_0; x)$. It follows from Theorem 1.4.5 that $\partial g(0) = \partial f(x_0) \neq \emptyset$. So g is proper convex with $\operatorname{dom} g = \operatorname{aff} \operatorname{dom} f - x_0$. In particular, $\operatorname{dom} g$ is closed, and g restricted to $\operatorname{dom} g$ is a real-valued convex function. It follows from Corollary 1.3.15 that g is continuous on $\operatorname{dom} g$, and therefore lsc on $\mathbb{R}^d$. So one obtains from Corollary 1.5.10 that $g = \delta^*_C$ for $C = \partial g(0) = \partial f(x_0)$. This proves the theorem. □

Theorem 1.5.14 *Let f be a proper convex function on $\mathbb{R}^d$ and $x_0 \in \operatorname{dom} f$. Then $\partial f(x_0)$ is non-empty and bounded if and only if $x_0 \in \operatorname{int} \operatorname{dom} f$.*

Proof. Let us first assume that $x_0 \in \operatorname{int} \operatorname{dom} f$. Then it follows from Theorem 1.4.5 that $\partial f(x_0) \neq \emptyset$. If there exists a sequence (z_n) in $\partial f(x_0)$ such that $||z_n|| \ge n$, then one has for every ε,

$$f(x_0 + \varepsilon z_n/||z_n||) \ge f(x_0) + \varepsilon \langle z_n/||z_n||, z_n \rangle = f(x_0) + \varepsilon ||z_n|| \ge f(x_0) + \varepsilon n.$$

That is, f is unbounded from above on every neighborhood of x_0, and it follows from Corollary 1.3.15 that $x_0 \notin \operatorname{int} \operatorname{dom} f$, a contradiction. So $\partial f(x_0)$ must be bounded.

Now we assume that $\partial f(x_0)$ is non-empty and bounded but $x_0 \notin \operatorname{int} \operatorname{dom} f$. Define $g(x) := f'(x_0; x)$. By Corollary 1.2.11, there exists a $z \in \mathbb{R}^d \setminus \{0\}$ such that

$$\operatorname{dom} f \subseteq \{x \in \mathbb{R} : \langle x, z \rangle \ge \langle x_0, z \rangle\} .$$

It follows that

$$\underline{g} = +\infty \quad \text{on the set } \left\{x \in \mathbb{R}^d : \langle x, z \rangle < 0\right\}. \tag{1.5.12}$$

Since $\partial f(x_0) = \partial g(0)$ is not empty, it follows from Lemma 1.4.9 that $\underline{g}(0) = g(0) = 0$ and $\partial \underline{g}(0) = \partial g(0) = \partial f(x_0)$. In particular, $\underline{g}$ is a lsc proper sub-linear function, and one obtains from Corollary 1.5.10 that

$$\underline{g}(x) = \sup_{z \in \partial f(x_0)} \langle x, z \rangle ,$$

contradicting (1.5.12). This shows that $x_0 \in \operatorname{int} \operatorname{dom} f$. □

Theorem 1.5.15 *Let f be a proper convex function on $\mathbb{R}^d$ and $x_0 \in \operatorname{dom} f$ such that $\partial f(x_0) = \{z\}$ for some $z \in \mathbb{R}^d$. Then f is differentiable at x_0 with $\nabla f(x_0) = z$.*

Proof. It follows from Theorems 1.5.14 and 1.5.13 that $x_0 \in \operatorname{int} \operatorname{dom} f$ and $f'(x_0; x) = \langle x, z\rangle$, $x \in \mathbb{R}^d$. So for given $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$f(x_0 + \lambda e_i) - f(x_0) - \langle \lambda e_i, z\rangle \le \varepsilon|\lambda|, \tag{1.5.13}$$

for all $i = 1, \dots, d$ and $\lambda \in [-\delta, \delta]$. Now choose $x \in \mathbb{R}^d$ such that

$$||x||_1 := \sum_{i=1}^d |x_i| \in (0, \delta].$$

By convexity of the function $g(x) := f(x_0 + x) - f(x_0) - \langle x, z\rangle$, one obtains from (1.5.13) that

$$g(x) = \sum_{i=1}^d g\left(||x||_1 \frac{\sum_i |x_i| \operatorname{sign}(x_i) e_i}{||x||_1}\right) \le \sum_{i=1}^d \frac{|x_i|}{||x||_1} g(||x||_1 \operatorname{sign}(x_i) e_i) \le \varepsilon ||x||_1.$$

Since

$$f(x_0 + x) - f(x_0) \ge \langle x, z\rangle \quad \text{for all } x \in \mathbb{R}^d,$$

and all norms on $\mathbb{R}^d$ are equivalent, one obtains

$$\lim_{x \neq 0,\, x \to 0} \frac{f(x_0 + x) - f(x_0) - \langle x, z\rangle}{||x||} = 0.$$

□

The following example shows that Theorem 1.5.15 does not hold for non-convex functions.

Example 1.5.16 The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) := \begin{cases} e^{x+1} & \text{for } x < -1 \\ |x| & \text{for } -1 \le x \le 1 \\ e^{1-x} & \text{for } 1 \le x \end{cases}$$

is not differentiable at 0. But $\partial f(0) = \{0\}$.

1.6 Inf-convolution

Definition 1.6.1 *Consider functions $f_j : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, $j = 1, \dots, n$. The inf-convolution of f_1 and f_2 is given by*

$$f_1 \Box f_2(x) := \inf_{y \in \mathbb{R}^d} (f_1(x - y) + f_2(y)) = \inf_{x_1 + x_2 = x} (f(x_1) + f(x_2)).$$

The inf-convolution of f_j, $j = 1, \dots, n$, is the function

$$\square_{j=1}^n f_j(x) := \inf_{x_1+\cdots+x_n=x} \sum_{j=1}^n f_j(x_j).$$

The inf-convolution is said to be exact if the infimum is attained.

Lemma 1.6.2 *Let $f_j : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, $j = 1, \dots, n$, be convex functions. Then $f = \square_{j=1}^n f_j$ is convex.*

Proof. If $f \equiv +\infty$, the lemma is clear. Otherwise, let $(x, v), (y, w) \in \operatorname{epi} f$, $\lambda \in (0, 1)$ and $\varepsilon > 0$. There exist x_j and y_j, $j = 1, \dots, n$, such that $\sum_{j=1}^n x_j = x$, $\sum_{j=1}^n f(x_j) \le v + \varepsilon$, $\sum_{j=1}^n y_j = y$ and $\sum_{j=1}^n f(y_j) \le w + \varepsilon$. Set $z_j = \lambda x_j + (1 - \lambda) y_j$. Then $z := \sum_{j=1}^n z_j = \lambda x + (1 - \lambda) y$ and

$$f(z) \le \sum_{j=1}^n f_j(z_j) \le \sum_{j=1}^n \lambda f_j(x_j) + (1 - \lambda) f(y_j) \le \lambda v + (1 - \lambda) w + \varepsilon.$$

It follows that $f(z) \le \lambda v + (1 - \lambda) w$, which shows that $\operatorname{epi} f$ and f are convex. □

Lemma 1.6.3 *Let f_j, $j = 1, \dots, n$, be proper convex functions on $\mathbb{R}^d$ and denote $f = \square_{j=1}^n f_j$. Assume $f(x_0) = \sum_j f_j(x_j) < +\infty$ for some x_j summing up to x_0 and $f_1(x) < +\infty$ for all x in some neighborhood of x_1. Then f is a proper convex function, $x_0 \in \operatorname{int} \operatorname{dom} f$ and f is continuous on $\operatorname{int} \operatorname{dom} f$.*

Proof. By definition of f, one has

$$f(x_0 + x) - f(x_0) \le f_1(x_1 + x) + \sum_{j=2}^n f_j(x_j) - \sum_{j=1}^n f_j(x_j) = f_1(x_1 + x) - f_1(x_1)$$

for all $x \in \mathbb{R}^d$. Therefore, $f(x) < +\infty$ for all x in some neighborhood of x_0. Since by Lemma 1.6.2, f is convex, the result follows from Theorem 1.3.14. □

Lemma 1.6.4 *Consider functions $f_j : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, $j = 1, \dots, n$, and denote $f = \square_{j=1}^n f_j$. Assume $f(x_0) = \sum_{j=1}^n f_j(x_j) < +\infty$ for some x_j summing up to x_0. Then $\partial f(x_0) = \bigcap_{j=1}^n \partial f_j(x_j)$.*

Proof. Assume $z \in \partial f(x_0)$ and $x \in \mathbb{R}^d$. Then

$$f_1(x_1+x) - f_1(x_1) = f_1(x_1+x) + \sum_{j=2}^n f_j(x_j) - \sum_{j=1}^n f_j(x_j) \ge f(x_0+x) - f(x_0) \ge \langle x, z \rangle .$$

Hence $z \in \partial f_1(x_1)$, and it follows by symmetry that $\partial f(x_0) \subseteq \bigcap_{j=1}^n \partial f_j(x_j)$. On the other hand, if $z \in \bigcap_{j=1}^n \partial f_j(x_j)$ and $x \in \mathbb{R}^d$, choose y_j such that $\sum_{j=1}^n y_j = x_0 + x$. Then

$$\sum_{j=1}^n f_j(y_j) \ge \sum_{j=1}^n f_j(x_j) + \langle y_j - x_j, z \rangle = \sum_{j=1}^n f_j(x_j) + \langle x, z \rangle .$$

So $f(x_0 + x) - f(x_0) \ge \langle x, z \rangle$, and the lemma follows. □

Lemma 1.6.5 *Let f_j, $j = 1, \dots, n$, be proper convex functions on $\mathbb{R}^d$ and denote $f = \square_{j=1}^n f_j$. Assume $f(x_0) = \sum_j f_j(x_j) < +\infty$ for some x_j summing up to and f_1 is differentiable at x_1. Then f is differentiable at x_0 with $\nabla f(x_0) = \nabla f_1(x_1)$.*

Proof. One has

$$f(x_0 + x) - f(x_0) \le f_1(x_1 + x) + \sum_{j=2}^n f_j(x_j) - \sum_{j=1}^n f_j(x_j) = f_1(x_1 + x) - f_1(x_1)$$

for all $x \in \mathbb{R}^d$. It follows that the directional derivative $g(x) := f'(x_0; x)$ satisfies

$$g(x) \le f_1'(x_1; x) = \langle x, \nabla f_1(x_1) \rangle$$

for all $x \in \mathbb{R}^d$. But by Lemma 1.6.2, f is convex. So g is sub-linear, and it follows that $g(x) = \langle x, \nabla f_1(x_1) \rangle$. This implies that $\partial f(x_0) = \partial g(0) = \{\nabla f_1(x_1)\}$, and the lemma follows from Theorem 1.5.15. □

Lemma 1.6.6 *Consider functions $f_j : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, $j = 1, \dots, n$, none of which is identically equal to $+\infty$. Then $\left(\square_{j=1}^n f_j\right)^* = \sum_{j=1}^n f_j^*$.*

Proof.

$$\left(\square_{j=1}^n f_j\right)^*(z) = \sup_x (\langle x, z \rangle - \square_{j=1}^n f_j(x)) = \sup_{x_1, \dots, x_n} \sum_{j=1}^n (\langle x_j, z \rangle - f_j(x_j)) = \sum_{j=1}^n f_j^*(z).$$

□

Corollary 1.6.7 *Let f_j, $j = 1, \dots, n$, be lsc proper convex functions on $\mathbb{R}^d$. Then $(\sum_{j=1}^n f_j)^* = \square_{j=1}^n f_j^*$.*

Proof. We know from Corollary 1.5.8 that f_j^*, $i = 1, \dots, n$, are lsc proper convex. So one obtains from Theorem 1.5.7 and Lemma 1.6.6 that

$$\sum_{j=1}^n f_j = \sum_{j=1}^n f_j^{**} = \left(\square_{j=1}^n f_j^*\right)^*,$$

and therefore, $(\sum_{j=1}^n f_j)^* = \square_{j=1}^n f_j^*$. □

Chapter 2

General Vector Spaces

2.1 Definitions

A general vector space is a set whose elements can be added and multiplied with scalars. It can be defined over a general field of scalars. But here we just consider vector spaces over $\mathbb{R}$. The precise definition is as follows:

Definition 2.1.1 *A vector space is a non-empty set X with an addition*

$$(x, y) \in X \times X \mapsto x + y \in X$$

and a scalar multiplication

$$(\lambda, x) \in \mathbb{R} \times X \mapsto \lambda x \in X$$

satisfying the following properties:

1. $(x + y) + z = x + (y + z)$ *for all* $x, y, z \in X$
2. $x + y = y + x$ *for all* $x, y \in X$
3. *There exists an element* $0 \in X$ *such that* $x + 0 = x$ *for all* $x \in X$.
4. *For every* $x \in X$ *there exists* $-x \in X$ *such that* $x + (-x) = 0$
5. $\lambda(x + y) = \lambda x + \lambda y$ *for all* $\lambda \in \mathbb{R}$ *and* $x, y \in X$
6. $(\lambda + \mu)x = \lambda x + \mu x$ *for all* $\lambda, \mu \in \mathbb{R}$ *and* $x \in X$
7. $\lambda(\mu x) = (\lambda\mu)x$
8. $1x = x$

Exercise 2.1.2
1. Show that there exists only one element $0 \in X$ satisfying 3. It is called zero-element or neutral element of the addition.

2. Show that $0x = 0$ for all $x \in X$.

3. Show that for given $x \in X$, there exists only one $-x \in X$ satisfying 4. It is called the negative or additive inverse of x.

4. Show that $(-1)x = -x$.

Examples 2.1.3 The following are vector spaces:

1. $\{0\}$

2. $\mathbb{R}^d$

3. The set of all linear functions $f : \mathbb{R}^d \to \mathbb{R}^k$.

4. The set of all functions $f : X \to Y$, where X is an arbitrary set and Y a vector space.

5. All polynomials on $\mathbb{R}^d$.

6. All real sequences.

7. All real sequences that converge.

8. $L^p(\Omega, \mathcal{F}, \mu)$, where $(\Omega, \mathcal{F}, \mu)$ is a measure space.

9. The product $X \times Y$ of two vector spaces X and Y.

10. The quotient X/Y if Y is a subspace of X. (In X/Y, x and x' are identified if $x - x' \in Y$.)

Definition 2.1.4 *Let Y be a subset of a vector space X.*

- *Y is said to be linearly independent if for every non-empty finite subset $\{y_1, \dots, y_k\}$ of Y, $(0, \dots, 0)$ is the only vector λ in $\mathbb{R}^k$ such that $\lambda_1 y_1 + \dots + \lambda_k y_k = 0$.*

- *If Y is linearly independent and for every $x \in X$, there exists a finite subset $\{y_1, \dots, y_k\}$ of Y and $\lambda \in \mathbb{R}^k$ such that $x = \lambda_1 x_1 + \dots + \lambda_k x_k$, then Y is called a Hamel basis of X.*

Exercise 2.1.5
1. Let Y be a Hamel basis of a vector space X. Show that the representation of points $x \in X$ as linear combinations of elements in Y is unique.

2. Show that $1, \cos(2\pi nx), \sin(2\pi nx)$, $n = 1, 2, \dots$ are linearly independent in $L^2[0, 1]$.

Definition 2.1.6 *Let C be a subset of $\mathbb{R}^d$. C is a subspace of $\mathbb{R}^d$ if*

$$\lambda x + y \in C \quad \textit{for all } x, y \in C \textit{ and } \lambda \in \mathbb{R}.$$

C is an affine set if

$$\lambda x + (1-\lambda)y \in C \quad \text{for all } x, y \in C \text{ and } \lambda \in \mathbb{R}.$$

C is a convex set if

$$\lambda x + (1-\lambda)y \in C \quad \text{for all } x, y \in C \text{ and } \lambda \in [0,1].$$

C is a cone if

$$\lambda x \in C \quad \text{for all } x \in C \text{ and } \lambda \in \mathbb{R}_{++}.$$

Exercise 2.1.7 Show that the statements of Exercise 1.1.2 hold for non-empty subsets C, D of a vector space.

Definition 2.1.8 *If C is a non-empty subset of a vector space, we denote by $\operatorname{lin} C$, $\operatorname{aff} C$, $\operatorname{conv} C$, $\operatorname{cone} C$ the smallest subspace, affine set, convex set, cone containing C, respectively.*

Definition 2.1.9 *A function $f : X \to Y$ between vector spaces is linear if*

$$f(\lambda x + y) = \lambda f(x) + f(y) \quad \text{for all } \lambda \in \mathbb{R} \text{ and } x, y \in X,$$

and affine if

$$f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda) f(y) \quad \text{for all } \lambda \in \mathbb{R} \text{ and } x, y \in X.$$

Definition 2.1.10 *The algebraic dual X' of a vector space X is the vector space of all linear functions $f : X \to \mathbb{R}$. Elements of X' are usually called linear functionals.*

Definition 2.1.11 *A function $f : X \to \mathbb{R} \cup \{\pm\infty\}$ on a vector space X is ...*

- *quasi-convex if all sub-level sets $\{x \in X : f(x) \leq y\}$, $y \in \mathbb{R}$, are convex.*
- *quasi-concave if all super-level sets $\{x \in X : f(x) \geq y\}$, $y \in \mathbb{R}$, are convex.*
- *convex if $\operatorname{epi} f := \{(x,y) \in X \times \mathbb{R} : f(x) \leq y\}$ is convex.*
- *proper convex if it is convex, $f(x) > -\infty$ for all $x \in X$ and $f(x) < +\infty$ for at least one $x \in X$.*
- *concave if $-f$ is convex.*
- *proper concave if $-f$ is proper concave.*
- *positively homogeneous if $\operatorname{epi} f$ is a cone.*
- *sub-linear if $\operatorname{epi} f$ is a convex cone.*

Exercise 2.1.12 *Let X be a vector space. Show the following:*

1. *The pointwise supremum of quasi-convex functions on X is quasi-convex.*

2. *The point wise supremum of convex functions on X is convex.*

3. *A positively homogeneous function $f : X \to \mathbb{R}$ satisfies $f(0) = 0$.*

4. *A positively homogeneous function $f : X \to \mathbb{R} \cup \{+\infty\}$ is convex if and only if $f(x + y) \leq f(x) + f(y)$ for all $x, y \in X$.*

2.2 Zorn's lemma and extension results

Definition 2.2.1 *A binary relation on a non-empty set X is a subset R of $X \times X$. One usually writes xRy instead of $(x, y) \in R$. R is said to be ...*

- *reflexive if xRx for all $x \in X$.*
- *symmetric if xRy implies yRx.*
- *antisymmetric if xRy and yRx imply $x = y$.*
- *transitive if xRy and yRz imply xRz.*
- *total if for all $x, y \in X$, one has xRy, yRx or both.*
- *an equivalence relation if it is reflexive, symmetric and transitive.*
- *a preorder if it is reflexive and transitive.*
- *a partial order if it is an antisymmetric preorder.*
- *a total order (or linear order) if it is a total partial order.*

Definition 2.2.2 *Let V be a subset of a partially ordered set $(X, \geq)$.*

- *V is called a chain if $(V, \geq)$ is totally ordered.*
- *An upper (lower) bound of V is an element $x \in X$ such that $x \geq v$ ($x \leq v$) for all $v \in V$.*
- *If $x \in V$ is an upper (lower) bound of V, it is called largest (smallest) element of V.*
- *An element $x \in V$ is called maximal (minimal) if there is no element $y \in V \setminus \{x\}$ such that $x \leq y$ ($x \geq y$).*

Zorn's lemma is equivalent to the axiom of choice. We use it as an axiom.

Zorn's lemma
Let X be a partially ordered set in which every chain has an upper bound. Then X has a maximal element.

Theorem 2.2.3 *Every vector space has a Hamel basis.*

Proof. Let X be a vector space and denote by W be the set of all linearly independent subsets Y of X. $Y_1 \geq Y_2 :\Leftrightarrow Y_1 \supseteq Y_2$ defines a partial order on W. If V is a chain in W, then $\bigcup_{Y \in V} Y$ is an upper bound of V. So it follows from Zorn's lemma that there exists a maximal element $Y \in W$. Y is a Hamel Basis of X. □

Exercise 2.2.4 Let Y be subspace of a vector space X.

1. Show that there exist subsets $V \subseteq Y$ and $W \subseteq X$ such that V is a Hamel Basis of Y and $V \cup W$ is a Hamel basis of X.

2. Show that every linear function $f : Y \to \mathbb{R}$ has a linear extension $F : X \to \mathbb{R}$.

Theorem 2.2.5 (Hahn–Banach extension theorem)
Let $g : X \to \mathbb{R}$ be a sub-linear function on a vector space X and $f : Y \to \mathbb{R}$ a linear function on a subspace Y of X such that $f(x) \leq g(x)$ for all $x \in Y$. Then there exists a linear extension $F : X \to \mathbb{R}$ of f such that $F(x) \leq g(x)$ for all $x \in X$.

Proof. If $Y \neq X$, choose $z \in X \setminus Y$ and set $\hat{Y} := \{y + \lambda z : y \in Y, \lambda \in \mathbb{R}\}$. For all $x, y \in Y$, one has

$$f(x) + f(y) = f(x + y) \leq g(x + y) \leq g(x - z) + g(y + z).$$

So there exists a number $\beta \in \mathbb{R}$ such that

$$\sup_{x \in Y} \{f(x) - g(x - z)\} \leq \beta \leq \inf_{y \in Y} \{-f(y) + g(y + z)\}.$$

Hence, if f is extended to $\hat{Y}$ by setting

$$f(y + \lambda z) = f(y) + \lambda \beta,$$

it stays dominated by g.

Now let W be the set of all pairs (V, F), where V is a subspace of X containing Y and $F : V \to \mathbb{R}$ a linear extension of f that is dominated by g on V. Write $(V_1, F_1) \geq (V_2, F_2)$ if $V_1 \supseteq V_2$ and $F_1 = F_2$ on V_2. If U is a chain in W, $\hat{V} = \bigcup_{V \in U} V$ is a vector space and $\hat{F}(x) := F(x)$ if $x \in V$ for some $(V, F) \in U$, defines a linear function $\hat{F} : \hat{V} \to \mathbb{R}$ such that $(\hat{V}, \hat{F})$ is an upper bound of W. So it follows from Zorn's lemma that W has a maximal element (V, F). But this means $V = X$. Otherwise, there would exist a $z \in X \setminus V$ and F could be extended to $\operatorname{lin}(V \cup \{z\})$ while staying dominated by g, a contradiction to the maximality of (V, F). □

Remark 2.2.6 If $g : X \to \mathbb{R}$ is a sub-linear function on a vector space X, then $g(0) = 0$. Since $\{0\}$ is a subspace of X, and $f(0) = 0$ is a linear function on $\{0\}$, one obtains from the Hahn–Banach extension theorem that there exists a linear function $F : X \to \mathbb{R}$ dominated by g.

Theorem 2.2.7 (Mazur–Orlicz)
Let $g : X \to \mathbb{R}$ be a sub-linear function on a vector space X and C a non-empty convex subset of X. Then there exists a linear function $f : X \to \mathbb{R}$ that is dominated by g and satisfies

$$\inf_{x \in C} f(x) = \inf_{x \in C} g(x).$$

Proof. If $\alpha := \inf_{x \in C} g(x) = -\infty$, choose any $f \in X'$ that is dominated by g (such an f exists by Hahn–Banach). Then $\inf_{x \in C} f(x) = \inf_{x \in C} g(x) = -\infty$. If $\alpha \in \mathbb{R}$, define

$$h(x) := \inf_{y \in C,\, \lambda > 0} \left\{ g(x + \lambda y) - \lambda \alpha \right\}.$$

Since $\alpha \le g(y)$, one has

$$g(x + \lambda y) - \lambda \alpha \ge g(x + \lambda y) - \lambda g(y) = g(x + \lambda y) - g(\lambda y) \ge -g(-x),$$

which shows that $h(x)$ is real-valued on $\mathbb{R}$. It is clear that h is positively homogeneous. Moreover, if $x_1, x_2 \in \mathbb{R}$, one has for all $y_1, y_2 \in C$ and $\lambda_1, \lambda_2 > 0$,

$$\begin{aligned} & g\left(x_1 + x_2 + (\lambda_1 + \lambda_2)\frac{\lambda_1 y_1 + \lambda_2 y_2}{\lambda_1 + \lambda_2}\right) - (\lambda_1 + \lambda_2)\alpha \\ = \; & g(x_1 + x_2 + \lambda_1 y_1 + \lambda_2 y_2) - (\lambda_1 + \lambda_2)\alpha \\ \le \; & g(x_1 + \lambda_1 y_1) - \lambda_1 \alpha + g(x_2 + \lambda_2 y_2) - \lambda_2 \alpha, \end{aligned}$$

which shows that $h(x_1 + x_2) \le h(x_1) + h(x_2)$. From the Hahn-Banach extension theorem one obtains an $f \in X'$ that is dominated by h. Note that

$$f(x) \le h(x) \le \inf_{y \in C} \{g(x + y) - \alpha\} \le \inf_{y \in C} \{g(x) + g(y) - \alpha\} = g(x) \quad \text{for all } x \in X.$$

In particular, $\inf_{x \in C} f(x) \le \inf_{x \in C} g(x)$. On the other hand,

$$-f(y) = f(-y) \le h(-y) \le g(-y + y) - \alpha = -\alpha \quad \text{for all } y \in C,$$

and it follows that $\inf_{x \in C} f(x) \ge \alpha = \inf_{x \in C} g(x)$. □

Corollary 2.2.8 *Let $g : X \to \mathbb{R}$ be a sub-linear function on a vector space X and $x_0 \in X$. Then there exists an $f \in X'$ that is dominated by g such that $f(x_0) = g(x_0)$.*

Proof. Apply Mazur–Orlicz with $C = \{x_0\}$. □

2.3 Algebraic interior and separation results

Definition 2.3.1 *Let C be a subset of a vector space X.*

- *The algebraic interior,* $\operatorname{core} C$*, of C consists of all points $x_0 \in C$ with the property that for every $x \in X$, there exists $\lambda_x > 0$ such that*

$$x_0 + \lambda x \in C \quad \textit{for all } \lambda \in [0, \lambda_x].$$

- *If $x_0 \in \operatorname{core} C$, we call C an algebraic neighborhood of x_0.*
- *If $0 \in \operatorname{core} C$, we call C absorbing.*

Lemma 2.3.2 *Let C be a convex subset of a vector space X such that* $\operatorname{core} C \neq \emptyset$*. Then*

$$\lambda \operatorname{core} C + (1 - \lambda) C \subseteq \operatorname{core} C \tag{2.3.1}$$

for all $\lambda \in (0, 1]$. In particular, $\operatorname{core} C$ *is convex.*

Proof. Let $x \in \operatorname{core} C$, $y \in C$, $\lambda \in (0, 1]$ and $z \in X$. There exists $\mu_z > 0$ such that $x + \mu z \in C$ for all $\mu \in [0, \mu_z]$. So one has

$$\lambda x + (1 - \lambda) y + \lambda \mu z = \lambda(x + \mu z) + (1 - \lambda) y \in C$$

for all $\mu \in [0, \mu_z]$. □

Definition 2.3.3 *Let C be a non-empty subset of a vector space X. The Minkowski functional $\mu_C : X \to [0, +\infty]$ is given by*

$$\mu_C(x) := \inf \{\lambda > 0 : x \in \lambda C\},$$

where $\inf \emptyset$ *is understood as $+\infty$.*

Lemma 2.3.4 *Let C be an absorbing convex subset of a vector space X. Then the Minkowski functional μ_C has the following properties:*

(i) *μ_C is real-valued and sub-linear*

(ii) *$\mu_C(x) < 1$ if $x \in \operatorname{core} C$, $\mu_C(x) \leq 1$ if $x \in C$ and $\mu_C(x) \geq 1$ if $x \notin \operatorname{core} C$.*

Proof. It is clear that μ_C is real-valued and positively homogeneous. Moreover, if $x, y \in X$ and $\lambda, \mu > 0$ are such that $x \in \lambda C$ and $y \in \mu C$. Then $x + y \in \lambda C + \mu C = (\lambda + \mu) C$ (the inclusion $\subseteq$ holds because C is convex). This shows that $\mu_C(x + y) \leq \mu_C(x) + \mu_C(y)$. So μ_C is sub-linear. The first two statements of (ii) are obvious. To show that last one, assume $\mu_C(x) < 1$. Then there exists a $\mu > 1$ such that $\mu x \in C$. Since $0 \in \operatorname{core} C$, it follows from Lemma 2.3.2 that $x \in \operatorname{core} C$. So if $x \notin \operatorname{core} C$, then $\mu_C(x) \geq 1$. □

Theorem 2.3.5 (Algebraic weak separation)
Let C and D be non-empty convex subsets of a vector space X such that $\operatorname{core} D \neq \emptyset$. *Then there exists $f \in X' \setminus \{0\}$ such that*

$$\inf_{x \in C} f(x) \geq \sup_{y \in D} f(y)$$

if and only if $C \cap \operatorname{core} D = \emptyset$.

Proof. The "only if" direction is clear. To show the "if" direction, we assume that $C \cap \operatorname{core} D = \emptyset$. Choose $x_0 \in \operatorname{core} D$. Then $A = C - x_0$ and $B = D - x_0$ are non-empty convex sets such that $A \cap \operatorname{core} B = \emptyset$ and B is absorbing. Therefore, the Minkowski functional μ_B is real-valued and sub-linear. It follows from Mazur–Orlicz that there exists an $f \in X'$ satisfying

$$f \leq \mu_B \text{ on } X \quad \text{and} \quad \inf_{x \in A} f(x) = \inf_{x \in A} \mu_B(x).$$

By Lemma 2.3.2, one has $\mu_B \leq 1$ on B. On the other hand, $\mu_B \geq 1$ on $X \setminus \operatorname{core} B$, and therefore, $\inf_{x \in A} \mu_B(x) \geq 1$. So one obtains

$$f(x) \geq 1 \geq f(y) \quad \text{for all } x \in A \text{ and } y \in B.$$

In particular, $f \in X' \setminus \{0\}$ and

$$f(x) \geq 1 + f(x_0) \geq f(y) \quad \text{for all } x \in C \text{ and } y \in D.$$

□

Theorem 2.3.6 (Algebraic strong separation)
Let C and D be non-empty convex subsets of a vector space X. Then there exists $f \in X'$ such that

$$\inf_{x \in C} f(x) > \sup_{y \in D} f(y) \tag{2.3.2}$$

if and only if there exists a convex absorbing set U such that $C \cap (D + U) = \emptyset$.

Proof. If there exists $f \in X'$ such that (2.3.2) holds, set

$$\beta := \inf_{x \in C} f(x) - \sup_{y \in D} f(y) > 0.$$

The set $U := \{x \in X : f(x) < \beta\}$ is convex absorbing, and C does not intersect $D + U$. This shows the "only if" direction.

For the "if" direction, assume there exists a convex absorbing set U such that $C \cap (D + U) = \emptyset$. Then $0 \notin D + U - C$. Since $\operatorname{core}(D + U - C) \neq \emptyset$, one obtains from Theorem 2.3.5 an $f \in X' \setminus \{0\}$ such that

$$0 \geq \sup_{x \in D + U - C} f(x),$$

or equivalently,

$$\inf_{x\in C} f(x) \geq \sup_{y\in D} f(y) + \sup_{u\in U} f(u).$$

Since U is absorbing, there exists $u \in U$ such that $f(u) > 0$, and it follows that

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

□

2.4 Directional derivatives and sub-gradients

Definition 2.4.1 *Let $f : X \to \mathbb{R} \cup \{\pm\infty\}$ be a function on a vector space X and $x_0 \in X$ such that $f(x_0) \in \mathbb{R}$. If the limit*

$$f'(x_0; x) := \lim_{\varepsilon \downarrow 0} \frac{f(x_0 + \varepsilon x) - f(x_0)}{\varepsilon}$$

exists (it is allowed to be $+\infty$ or $-\infty$), we call it the directional derivative of f at x_0 in the direction x.

If there exists $x' \in X'$ such that $f'(x_0; x) = x'(x)$ for all $x \in X$, x' is called algebraic Gâteaux derivative of f at x_0.

Definition 2.4.2 *Let $f : X \to \mathbb{R} \cup \{\pm\infty\}$ be a function on a vector space X and $x_0 \in X$ such that $f(x_0) \in \mathbb{R}$. $x' \in X'$ is an algebraic sub-gradient of f at x_0 if*

$$f(x_0 + x) - f(x_0) \geq x'(x) \quad \text{for all } x \in X.$$

We denote the set of all algebraic sub-gradients of f at x_0 by $\partial_a f(x_0)$ and call it algebraic sub-differential of f at x_0.

Definition 2.4.3 *The effective domain of a function $f : X \to \mathbb{R} \cup \{+\infty\}$ or $f : X \to \mathbb{R} \cup \{-\infty\}$ on a set X is*

$$\operatorname{dom} f := \{x \in X : f(x) \in \mathbb{R}\}.$$

Exercise 2.4.4 Let $f : X \to \mathbb{R} \cup \{\pm\infty\}$ be a convex function on a vector space and $x_0 \in X$ such that $f(x_0) \in \mathbb{R}$. Show the following:

1.

$$f'(x_0; x) = \inf_{\varepsilon > 0} \frac{f(x_0 + \varepsilon x) - f(x_0)}{\varepsilon}.$$

In particular, $f'(x_0; x)$ exists for all $x \in \mathbb{R}^d$.

2. $f'(x_0, .)$ is sub-linear.

3. If $x_0 \in \operatorname{core} \{x \in X : f(x) \in \mathbb{R}\}$, then $f'(x_0; x) \in \mathbb{R}$ for all $x \in X$.

4. The following are equivalent:

(i) $f(x_0) = \min_x f(x)$

(ii) $0 \in \partial_a f(x_0)$

(iii) $f'(x_0; x) \geq 0$ for all $x \in X$.

5. $\partial_a f(x_0)$ is a convex subset of X'.

6. $\partial_a f(x_0) = \partial_a g(0)$, where $g(x) := f'(x_0; x)$.

7. The following are equivalent:

(i) $z \in \partial f_a(x_0)$

(ii) $(-z, 1)$ supports $\operatorname{epi} f$ at $(x_0, f(x_0))$.

Lemma 2.4.5 *Let $f : X \to \mathbb{R} \cup \{\pm\infty\}$ be a convex function on a vector space X such that $f(x_0) \in \mathbb{R}$. Assume there exists an algebraic neighborhood U of x_0 such that $f(x) < +\infty$ for all $x \in U$. Then $f(x) > -\infty$ for all $x \in X$.*

Proof. Assume there exists $x_1 \in X$ such that $f(x_1) = -\infty$. Then there exists $x_2 \in U$ and $\lambda \in (0, 1)$ such that $x_0 = \lambda x_1 + (1 - \lambda)x_2$. It follows that $f(x_0) = -\infty$, a contradiction. □

Theorem 2.4.6 *Let f be a proper convex function on a vector space X and $x_0 \in X$. Assume there exists an algebraic neighborhood U of x_0 such that $f(x) < +\infty$ for all $x \in U \cap \operatorname{aff} \operatorname{dom} f$. Then $\partial_a f(x_0) \neq \emptyset$.*

Proof. The restriction of the directional derivative $g(x) := f'(x_0, x)$ to the subspace $Y = \operatorname{aff} \operatorname{dom} f - x_0$ is sub-linear and real-valued because $f(x) < +\infty$ for all $x \in U \cap \operatorname{aff} \operatorname{dom} f$. So it follows from the Hahn–Banach extension theorem that there exists a $y' \in Y'$ such that $y'(y) \leq g(y)$, $y \in Y$. By Exercise 2.2.4, y' has a linear extension $x' \in X'$, and since $g(x) = +\infty$ for $x \in X \setminus Y$, one has $x'(x) \leq g(x)$, $x \in X$. This shows that $x' \in \partial_a g(0) = \partial f_a(x_0)$. □

Chapter 3

Topological Vector Spaces

3.1 Topological spaces

Definition 3.1.1 *A topological space is a non-empty set X with a family τ of subsets of X satisfying:*

(i) $\emptyset, X \in \tau$

(ii) $\bigcup_{V \in \eta} V \in \tau$ *for every non-empty subset* $\eta \subseteq \tau$

(iii) $\bigcap_{i=1}^{k} V_i \in \tau$ *for every finite subset* $\{V_1, \dots, V_k\}$ *of* τ.

τ is called a topology and the members of τ open sets. A set $V \subseteq X$ is called closed if $X \setminus V$ is open. The interior $\operatorname{int} C$ *of a set $C \subseteq X$ is the largest open set contained in C. The closure* $\operatorname{cl} C$ *is the smallest closed set containing C. The boundary* $\operatorname{bd} C$ *of C is the set* $\operatorname{cl} C \setminus \operatorname{int} C$. *$C$ is dense in X if* $\operatorname{cl} C = X$. *(X, τ) is separable if it contains a countable dense subset.*

Definition 3.1.2 *A filter on a non-empty set X is a family $\mathcal{V}$ of subsets satisfying*

(i) *$\emptyset \notin \mathcal{V}$ and $X \in \mathcal{V}$.*

(ii) *If $U, V \in \mathcal{V}$, then $U \cap V \in \mathcal{V}$.*

(iii) *If $U \in \mathcal{V}$ and $U \subseteq V$, then $V \in \mathcal{V}$.*

Definition 3.1.3 *A subset U of a topological space (X, τ) is a neighborhood of a point $x \in X$ if $x \in$* int *U. The neighborhood filter τ_x of x is the family of all neighborhoods of x. A subset $\mathcal{B}_x$ of τ_x is called a neighborhood base of x if for every $U \in \tau_x$ there exists a $V \in \mathcal{B}_x$ such that $V \subseteq U$. (X, τ) is called first countable if every $x \in X$ has a countable neighborhood base. The neighborhood system of the topology τ consists of all neighborhood filters τ_x, $x \in X$. (X, τ) is said to be Hausdorff (or separated) if any two different points have disjoint neighborhoods.*

Exercise 3.1.4 Show that every point in a Hausdorff topological space is closed.

Exercise 3.1.5 Let (X, τ) be a topological space and $x \in X$. Show the following:

1. τ_x is a filter on X such that each $U \in \tau_x$ contains x.

2. Each $U \in \tau_x$ contains a $V \in \tau_x$ such that $U \in \tau_y$ for all $y \in V$.

Exercise 3.1.6 Let X be a non-empty set and $\mathcal{N}_x$, $x \in X$, a collection of filters on X satisfying 1. and 2. of Exercise 3.1.5. Show that the collection of all sets $V \subseteq X$ satisfying $V \in \mathcal{N}_x$ for every $x \in V$, forms a topology τ on X such that $\tau_x = \mathcal{N}_x$ for all $x \in X$.

Hint: The proof of the inclusion $\tau_x \subseteq \mathcal{N}_x$ is straight-forward. To show the other inclusion, let $U \in \mathcal{N}_x$ and note that $x \in V := \{y \in U : U \in \mathcal{N}_y\}$. If it can be shown that V belongs to τ, it follows that $U \in \tau_x$.

Definition 3.1.7 *A directed set is a non-empty set A with a preorder $\geq$ such that for every pair $(a, b) \in A^2$ there exists a $c \in A$ such that $c \geq a$ and $c \geq b$.*

A net in a set X is a family $(x_a)_{a \in A}$ of elements in X indexed by a directed set A.

A net $(x_a)_{a \in A}$ in a topological space (X, τ) is said to converge to a point $x \in X$ if for every neighborhood U of x there exists an $a_0 \in A$ such that $x_a \in U$ for all $a \geq a_0$.

Exercise 3.1.8 Let C be a non-empty subset of a topological space X and $x \in X$. Show that the following are equivalent:

(i) $x \in \operatorname{cl} C$;

(ii) $C \cap U \neq \emptyset$ for every neighborhood U of x;

(iii) There exists a net $(x_a)_{a \in A}$ in C converging to x.

Definition 3.1.9 *Let (X, τ) be a topological space. A subset Y of X is compact if for every subset η of τ satisfying $\bigcup_{V \in \eta} V \supseteq Y$ there exists a finite subset $\{V_1, \dots, V_k\}$ of η such that $\bigcup_{i=1}^k V_i \supseteq Y$.*

Exercise 3.1.10 Let (X, τ) be a topological space. Show the following:

(i) Single points in X are compact but not necessarily closed.

(ii) If (X, τ) is Hausdorff, then compact sets in X are closed.

Definition 3.1.11 *Let (X, τ) be a topological space and Y a subset of X. The topology induced by τ on Y is*

$$\tau_Y := \{V \cap Y : V \in \tau\} .$$

Members of τ_Y are called relatively open in Y.

Definition 3.1.12 *A function $f : (X, \tau) \to (Y, \eta)$ between topological spaces is continuous at a point $x_0 \in X$ if $f^{-1}(U)$ is a neighborhood of x_0 for every neighborhood U of $f(x_0)$. f is said to be continuous if it is continuous at every $x \in X$.*

Exercise 3.1.13 Let $f : (X, \tau) \to (Y, \eta)$ be a function between topological spaces. Show the following:

1. f is continuous if and only if $f^{-1}(V) \in \tau$ for every $V \in \eta$.

2. f is continuous at a point $x \in X$ if and only if $f(x_a)$ converges to $f(x)$ for every net $(x_a)_{a \in A}$ in X that converges to x.

3. If (X, τ) is first countable, then f is continuous at a point $x \in X$ if and only if $f(x_n)$ converges to $f(x)$ for every sequence $(x_n)_{n \in \mathbb{N}}$ in X that converges to x.

4. If (X, τ) is not first countable, it is possible that f is not continuous at some $x \in X$ but $f(x_n)$ converges to $f(x)$ for every sequence $(x_n)_{n\mathbb{N}}$ that converges to x.

Definition 3.1.14 *A function $f : (X, \tau) \to \mathbb{R} \cup \{\pm\infty\}$ on a topological space is lsc at a point $x_0 \in X$ if for every $\varepsilon > 0$ there exists a neighborhood U of x_0 such that $f(x) \geq f(x_0) - \varepsilon$ for all $x \in U$. It is said to be lsc if it is lsc everywhere on X. f is usc at x_0 if $-f$ is lsc at x_0 and usc if $-f$ is lsc. By $\underline{f}$, we denote the function given by*

$$\underline{f}(x_0) := \sup_{U \in \tau_x} \inf_{x \in U} f(x)$$

and call it lsc hull of f.

Exercise 3.1.15
Consider a function $f : X \to \mathbb{R} \cup \{\pm\infty\}$ on a topological vector space.

1. Show that the following are equivalent:

(i) f is lsc

(ii) All sub-level sets $\left\{x \in \mathbb{R}^d : f(x) \leq c\right\}$, $c \in \mathbb{R}$, are closed

(iii) $\operatorname{epi} f$ is closed

2. Show that the epigraph of $\underline{f}$ is the closure of $\operatorname{epi} f$ and $\underline{f}$ is the greatest lsc minorant of f.

3. Let $f_i : X \to \mathbb{R} \cup \{\pm\infty\}$, $i \in I$, be a family of lsc functions. Show that $\sup_{i \in I} f_i$ is lsc.

Definition 3.1.16 *Let (X_i, τ_i), $i \in I$, be a family of topological spaces. The product topology on $\prod_{i \in I} X_i$ is the coarsest topology that makes all the projections continuous.*

Definition 3.1.17 *A pseudo-metric on a non-empty set X is a function $d : X \times X \to \mathbb{R}_+$ with the following three properties:*

(i) $d(x,x) = 0$ *for all* $x \in X$.

(ii) $d(x,y) = d(y,x)$ *for all* $x, y \in X$.

(iii) $d(x,y) + d(y,z) \geq d(x,z)$ *for all* $x, y, z \in X$.

If in addition to (i)–(iii), d *satisfies*

(iv) $d(x,y) = 0$ *implies* $x = y$,

then d *is a metric.*

Exercise 3.1.18 Let d be a pseudo-metric on a non-empty set X and define

$$B_n(x) := \{y \in X : d(x,y) \leq 1/n\}, \quad x \in X, n \in \mathbb{N}.$$

Show that

$$\mathcal{B}_x := \{B_n(x) : n \in \mathbb{N}\}, \quad x \in X,$$

define neighborhood bases inducing a first countable topology τ on X, which is separable if and only if d is a metric.

Definition 3.1.19 *A semi-norm on a vector space* X *is a sub-linear function* $p : X \to \mathbb{R}_+$ *such that*

$$p(\lambda x) = |\lambda| p(x) \quad \text{for all } x \in X \text{ and } \lambda \in \mathbb{R}.$$

If in addition, $p(x) = 0$ *implies* $x = 0$, p *is a norm.*

Exercise 3.1.20 Let p be a semi-norm on a vector space X. Show that ...

1. $d(x,y) := p(x - y)$ defines a pseudo-metric.

2. if p is a norm, then d is a metric.

Definition 3.1.21 *An inner product (or scalar product) on a vector space is a mapping* $(x,y) \in X \times X \mapsto \langle x,y \rangle \in \mathbb{R}$ *with the properties:*

(i) $\langle \lambda x + y, z \rangle = \lambda \langle x, z \rangle + \langle y, z \rangle$ *for all* $\lambda \in \mathbb{R}$ *and* $x, y, z \in X$.

(ii) $\langle x, y \rangle = \langle y, x \rangle$ *for all* $x, y \in X$.

(iii) $\langle x, x \rangle > 0$ *for all* $x \in X \setminus \{0\}$.

Exercise 3.1.22 Let $\langle x, x \rangle$ be an inner product on a vector space X. Show that $||x|| := \langle x, x \rangle^{1/2}$ defines a norm.

Definition 3.1.23 *A topological vector space is a vector space X with a topology τ such that the operations*

$$(x, y) \in X \times X \mapsto x + y \in X \quad \text{and} \quad (\lambda, x) \in \mathbb{R} \times X \mapsto \lambda x \in X$$

are continuous with respect to the product topologies on $X \times X$ and $\mathbb{R} \times X$, respectively, where $\mathbb{R}$ is endowed with the usual topology induced by $d(x, y) = |x - y|$.

X is said to be locally convex if 0 has a neighborhood base consisting of convex sets.

Exercise 3.1.24 Show that for a vector space X the following hold:

1. A norm on X induces a topology under which X is a locally convex topological vector space.

2. For every $x' \in X'$, $|x'(x)|$ defines a semi-norm on X.

3. Let D be a non-empty subset of X'. Write neighborhood bases of the coarsest topology on X making every $x' \in D$ continuous.

Remark 3.1.25

1. Let X be a topological vector space. Since the addition is continuous, the translation $x \mapsto x + x_0$ is a homeomorphism for each x_0 with inverse $x \mapsto x - x_0$. Therefore, a subset $V \subseteq X$ is open/closed/a neighborhood of 0 if and only if $V + x_0$ is open/closed/a neighborhood of x_0, respectively.

2. The multiplication with real numbers is also continuous. Therefore, for every $\lambda \in \mathbb{R} \setminus \{0\}$, the mapping $x \mapsto \lambda x$ is a homeormorphism with inverse $x \mapsto x/\lambda$. So a subset $V \subseteq X$ is open/closed/a neighborhood of 0 if and only if λV is open/closed/a neighborhood of 0, respectively.

Lemma 3.1.26 *Let C be subset of a topological vector space X. Then $\operatorname{int} C \subseteq \operatorname{core} C$. In particular, every 0-neighborhood in X is absorbing.*

Proof. Let $x \in \operatorname{int} C$ and $y \in X$. Since the vector space operations are continuous, there exists a $\varepsilon > 0$ such that $x + \lambda y \in C$ for all $0 \leq \lambda \leq \varepsilon$. Hence, $x \in \operatorname{core} C$. If U is a 0-neighborhood in X, then $0 \in \operatorname{int} U$, and therefore, U is absorbing. □

Lemma 3.1.27 *Let C be a convex subset of a topological vector space X. Then the following hold:*

(i) *If $\operatorname{int} C \neq \emptyset$, then $\lambda \operatorname{int} C + (1 - \lambda) \operatorname{cl} C \subseteq \operatorname{int} C$ for all $\lambda \in (0, 1]$.*

(ii) *$\operatorname{int} C$ and $\operatorname{cl} C$ are convex.*

(iii) *If $\operatorname{int} C \neq \emptyset$, then $\operatorname{int} C = \operatorname{core} C$.*

Proof. (i) Let $x \in \operatorname{int} C$, $y \in \operatorname{cl} C$ and $\lambda \in (0,1]$. There exists a neighborhood U of 0 in X such that $x + U \subseteq C$. Since the vector space operations are continuous, there exist neighborhoods V and W of 0 in X such that

$$\frac{(1-\lambda)}{\lambda} V + \frac{1}{\lambda} W \subseteq U.$$

Moreover, there is a $z \in C$ such that $y - z \in V$. So one has

$$\lambda x + (1-\lambda) y + w = \lambda \left(x + \frac{(1-\lambda)}{\lambda}(y-z) + \frac{1}{\lambda} w \right) + (1-\lambda) z \in C$$

for all $w \in W$. This proves (i).

(ii) That $\operatorname{int} C$ is convex is a consequence of (i). If $x, y \in \operatorname{cl} C$, there exist nets $(x_a)_{a \in A}$ and $(y_a)_{a \in A}$ converging to x and y, respectively. But then $\lambda x_a + (1-\lambda) y_a \to \lambda x + (1-\lambda) y$ for every $0 < \lambda < 1$, and it follows that $\operatorname{cl} C$ is convex.

(iii) We know from Lemma 3.1.26 that $\operatorname{int} C \subseteq \operatorname{core} C$. On the other hand if $x \in \operatorname{core} C$ and there exist a $y \in \operatorname{int} C$, there is a $z \in C$ such that $x = \lambda y + (1-\lambda) z$ for some $\lambda \in (0,1]$. So it follows from (i) that $x \in \operatorname{int} C$. □

Exercise 3.1.28 Let $f : X \to \mathbb{R} \cup \{\pm\infty\}$ be a convex function on a topological vector space. Show that $\underline{f}$ is still convex.

Definition 3.1.29 *We call a subset C of a vector space X balanced if $\lambda C \subseteq C$ for all $\lambda \in [-1,1]$.*

Lemma 3.1.30 *Let X be a topological vector space. Then 0 has a neighborhood base consisting of open balanced sets. If X is locally convex, 0 has a neighborhood base consisting of convex open balanced sets.*

Proof. Let U be a 0-neighborhood in X. Then there exists an open 0-neighborhood V in X and $\varepsilon > 0$ such that $\lambda x \in U$ for all $\lambda \in [-\varepsilon, \varepsilon]$ and $x \in V$. $W = \varepsilon V$ is still an open 0-neighborhood in X and $\bigcup_{-1 \le \lambda \le 1} \lambda W$ is an open balanced 0-neighborhood contained in U.

If X is locally convex, there exists a convex 0-neighborhood V contained in U. $W = \operatorname{int} V$ is a convex open neighborhood of 0 contained in U and $W \cap (-W)$ a convex open balanced neighborhood of 0 contained in U. □

3.2 Continuous linear functionals and extension results

Theorem 3.2.1 *Let X be a topological vector space and $f \in X' \setminus \{0\}$. Then the following are equivalent:*

(i) *f is continuous;*

(ii) *f is continuous at 0;*

(iii) *$f^{-1}(0)$ is closed;*

(iv) *$f^{-1}(0)$ is not dense in X;*

(v) *f is bounded on some 0-neighborhood U in X;*

(vi) *There exists a non-empty open subset V of X such that $f(V) \neq \mathbb{R}$.*

Proof. It is clear that (i) implies (ii) and (iii). (ii) $\Rightarrow$ (i) follows since for every $x \in X$, U is a 0-neighborhood if and only if $x + U$ is an x-neighborhood. (iii) $\Rightarrow$ (iv) follows since $f^{-1}(0) \neq X$. (iv) $\Rightarrow$ (v): If $f^{-1}(0)$ is not dense in X, it follows from Lemma 3.1.30 that there exist $x \in X$ and a balanced 0-neighborhood U such that $(x + U) \cap f^{-1}(0) = \emptyset$. This implies that f is bounded on U. (v) $\Rightarrow$ (ii): If U is a 0-neighborhood on which f is bounded by $m > 0$, then $|f(x)| \leq m/n$ for all all $x \in U/n$, which shows (ii). (v) $\Rightarrow$ (vi): If f is bounded on a 0-neighborhood U in X, then $V = \operatorname{int} U$ is a non-empty open set such that $f(V) \neq \mathbb{R}$. (vi) $\Rightarrow$ (iv): If $V \subseteq X$ satisfies (vi), there exists $a \in \mathbb{R}$ such that $V \cap f^{-1}(a) = \emptyset$. Since f is non-trivial, there exists a $x \in f^{-1}(a)$. Then $V - x$ is a non-empty open set that does not intersect $f^{-1}(0)$. It follows that $f^{-1}(0)$ is not dense in X. □

Remark 3.2.2 Theorem 3.2.1 shows that for a non-zero linear functional $f : X \to \mathbb{R}$ on a topological vector space one of the following holds:

(i) $f^{-1}(0)$ is a proper closed subspace of X and f is continuous.

(ii) $f^{-1}(0)$ is dense in X and f is not continuous.

Corollary 3.2.3 *Let $f : X \to \mathbb{R}$ be a linear function on a topological vector space X that is dominated by a sub-linear function $g : X \to \mathbb{R}$ which is continuous at 0. Then f is continuous.*

Proof. It follows from Lemma 3.1.30 that for given $\varepsilon > 0$, there exists a balanced 0-neighborhood U in X such that $|g(x)| \leq \varepsilon$ for all $x \in U$. Hence, $f(x) \leq g(x) \leq \varepsilon$ and $-f(x) = f(-x) \leq g(-x) \leq \varepsilon$ for all $x \in U$. This shows that f is continuous at 0, which by Theorem 3.2.1 implies that it is continuous everywhere. □

Theorem 3.2.4 (Hahn–Banach topological extension theorem)
Let $g : X \to \mathbb{R}$ be a sub-linear function on a topological vector space that is continuous at 0 and $f : Y \to \mathbb{R}$ a linear function on a subspace Y of X such that $f(x) \leq g(x)$ for all $x \in Y$. Then there exists a continuous linear extension $F : X \to \mathbb{R}$ of f such that $F(x) \leq g(x)$ for all $x \in X$.

Proof. We know from the algebraic version of Hahn–Banach that there exists a linear extension $F : X \to \mathbb{R}$ of f that is dominated by g. By Corollary 3.2.3, F is continuous. □

Theorem 3.2.5 (Topological version of Mazur–Orlicz)
Let $g : X \to \mathbb{R}$ be a sub-linear function on a topological vector space X that is continuous at 0 and C a non-empty convex subset of X. Then there exists a continuous linear function $f : X \to \mathbb{R}$ that is dominated by g and satisfies

$$\inf_{x \in C} f(x) = \inf_{x \in C} g(x). \tag{3.2.1}$$

Proof. From the algebraic version of Mazur–Orlicz we know that there exist a linear function $f : X \to \mathbb{R}$ that is dominated by g and satisfies (3.2.1). By Corollary 3.2.3, f is continuous. □

Definition 3.2.6 *The topological dual of a topological vector space X consists of the vector space*

$$X^* := \{x' \in X' : x' \textit{ is continuous}\}.$$

Remark 3.2.7 Every linear functional on $\mathbb{R}^d$ is continuous and can be represented by a vector $z \in \mathbb{R}^d$. Hence, $(\mathbb{R}^d)^* = (\mathbb{R}^d)' = \mathbb{R}^d$.

Remark 3.2.8 For a general topological vector space X, the topological dual X^* depends on the topology. But it is possible that there exist different topologies inducing the same space X^* of continuous linear functionals.

3.3 Separation with continuous linear functionals

Theorem 3.3.1 (Topological weak separation)
Let C and D be non-empty convex subsets of a topological vector space X such that $\operatorname{int} D \neq \emptyset$. Then there exists an $f \in X^* \setminus \{0\}$ such that

$$\inf_{x \in C} f(x) \geq \sup_{y \in D} f(y) \tag{3.3.2}$$

if and only if $C \cap \operatorname{int} D = \emptyset$.

Proof. We know from Lemma 3.1.27 that $\operatorname{int} C = \operatorname{core} C$. So the "only if" direction is clear. On the other hand, if $C \cap \operatorname{int} D = \emptyset$, it follows from algebraic weak separation that there exists an $f \in X' \setminus \{0\}$ satisfying (3.3.2). But then $\operatorname{int} D$ is a non-empty open subset of X such that $f(\operatorname{int} D) \neq \mathbb{R}$. Thus one obtains from Theorem 3.2.1 that f is continuous. □

The following is an immediate consequence of Theorem 3.3.1:

Corollary 3.3.2 *Let C be a closed convex subset of a topological vector space. If C has non-empty interior, then it is supported at every boundary point by a non-trivial continuous linear functional.*

Another consequence of Theorem 3.3.1 is:

Corollary 3.3.3 *Let X be a topological vector space. Then $X^* \neq \{0\}$ if and only if 0 has a convex neighborhood different from X.*

Proof. If there exists $f \in X^* \backslash \{0\}$, then $\{x \in X : f(x) < 1\}$ is a convex 0-neighborhood different from X. On the other hand, if U is such a neighborhood, there exists $x \in X \setminus U$. Since $\operatorname{int} U \neq \emptyset$ and $\{x\} \cap \operatorname{int} U = \emptyset$, the existence of an $f \in X^* \setminus \{0\}$ follows from Theorem 3.3.1. □

Lemma 3.3.4 *Let C and D be non-empty disjoint subsets of a topological vector space X such that C is closed and D compact. Then there exists a neighborhood U of 0 in X such that $C \cap (D + U) = \emptyset$.*

Proof. For every $x \in D$ there exists a neighborhood V_x of 0 in X such that $C \cap (x + V_x) = \emptyset$. Since the vector space operations are continuous, there is an open neighborhood U_x of 0 in X satisfying $U_x + U_x \subseteq V_x$. Due to compactness, there are finitely many $x_1, \dots, x_n \in D$ such that $D \subseteq \bigcup_{i=1}^n (x_i + U_{x_i})$. $U = \bigcap_{i=1}^n U_{x_i}$ is again a 0-neighborhood, and for every $x \in D$ there exists an i such that $x = x_i + u_i$ for some $u_i \in U_{x_i}$. So for all $u \in U$, one has

$$x + u = x_i + u_i + u \subseteq x_i + U_{x_i} + U_{x_i} \subseteq x_i + V_{x_i},$$

and therefore, $C \cap (x + U) = \emptyset$. □

Theorem 3.3.5 (Topological strong separation)
Let C and D be non-empty disjoint convex subsets of a locally convex topological vector space X such that C is closed and D is compact. Then there exists an $f \in X^* \setminus \{0\}$ such that

$$\inf_{x \in C} f(x) > \sup_{y \in D} f(y). \tag{3.3.3}$$

Proof. We know from Lemma 3.3.4 that there exists a neighborhood U of 0 in X such that $C \cap (D+U) = \emptyset$. Since X is locally convex, there exists a convex neighborhood V of 0 with the same property. $D + V$ is a convex set satisfying $\operatorname{int}(D + V) \neq \emptyset$ and $C \cap \operatorname{int}(D + V) = \emptyset$. So it follows from Theorem 3.3.1 that there exists an $f \in X^* \setminus \{0\}$ such that

$$\inf_{x \in C} f(x) \geq \sup_{y \in D+V} f(y).$$

But, by By Lemma 3.1.26, V is absorbing, and one obtains (3.3.3). □

Remark 3.3.6 Note that in Theorem 3.3.5 we did not assume X to by Hausdorff or D to be closed.

Corollary 3.3.7 *Let C be a non-empty closed convex subset of a locally convex topological vector space X and $x_0 \in X \setminus C$. Then there exists an $f \in X^* \setminus \{0\}$ such that*

$$\inf_{x \in C} f(x) > f(x_0).$$

Proof. The corollary is a consequence of Theorem 3.3.5 since $\{x_0\}$ is compact. □

As an immediate consequence one obtains the following:

Corollary 3.3.8 *Let C be a proper non-empty closed convex subset of a locally convex topological vector space X. Then*

$$C = \bigcap \{H(x^*, c) : x^* \in X^*,\, c \in \mathbb{R},\, C \subseteq H(x^*, c)\},$$

where

$$H(x^*, c) := \{x \in X : x^*(x) \geq c\}.$$

Corollary 3.3.9 *Let X be a locally convex topological vector space. Then the following two are equivalent:*

(i) *X is Hausdorff.*

(ii) *For any two different points $x, y \in X$, there exists an $f \in X^*$ such that $f(x) \neq f(y)$.*

Proof. If X is Hausdorff, then single points are closed. So one obtains from Corollary 3.3.7 that different points can be separated with continuous linear functionals.

On the other hand, if there exists an $f \in X^*$ such that $f(x) < f(y)$, set $m := (f(x) + f(y))/2$. Then $\{z \in X : f(z) < m\}$ is an x-neighborhood that does not intersect the y-neighborhood $\{z \in X : f(z) > m\}$. □

Definition 3.3.10 *The topological dual cone of a non-empty subset C of a topological vector space X is given by*

$$C^* := \{x^* \in X^* : x^*(x) \geq 0 \text{ for all } x \in C\}.$$

Exercise 3.3.11 Let C be a non-empty subset of a locally convex topological vector space X. Show the following:

1. C^* is a convex cone in X^* that is closed with respect to $\sigma(X^*, X)$ (the coarsest topology on X^* such that all x, viewed as linear functionals on X^*, are continuous).

2. The set

$$\{x \in X : x^*(x) \geq 0 \text{ for all } x^* \in C^*\}$$

is the smallest closed convex cone in X that contains C.

3.4 Continuity of convex functions

Theorem 3.4.1 *Let $f : X \to \mathbb{R} \cup \{\pm\infty\}$ be a convex function on a topological vector space X and $x_0 \in X$ such that $f(x_0) \in \mathbb{R}$. Assume there exists a neighborhood U of 0 such that $\sup_{x \in U} f(x_0 + x) < +\infty$. Then f is proper convex, $x_0 \in \operatorname{int} \operatorname{dom} f$ and f is continuous on $\operatorname{int} \operatorname{dom} f$.*

Proof. Since $x_0 \in \operatorname{core}(x_0 + U)$, it follows from the convexity of f that $f(x) > -\infty$ for all $x \in X$. Hence f is proper convex, and $x_0 \in \operatorname{int} \operatorname{dom} f$.

Now choose a balanced 0-neighborhood V contained in U and set

$$m := \sup_{x \in V} f(x) \in \mathbb{R}.$$

Then for $x \in V$ and $0 < \lambda \leq 1$, one has

$$f(x_0 + \lambda x) = f(\lambda(x_0 + x) + (1 - \lambda)x_0) \leq \lambda f(x_0 + x) + (1 - \lambda) f(x_0),$$

and therefore,

$$f(x_0 + \lambda x) - f(x_0) \leq \lambda[f(x_0 + x) - f(x_0)] \leq \lambda(m - f(x_0)).$$

On the other hand,

$$x_0 = \frac{1}{1 + \lambda}(x_0 + \lambda x) + \frac{\lambda}{1 + \lambda}(x_0 - x).$$

So

$$f(x_0) \leq \frac{1}{1 + \lambda} f(x_0 + \lambda x) + \frac{\lambda}{1 + \lambda} f(x_0 - x),$$

from which one obtains

$$f(x_0) - f(x_0 + \lambda x) \leq \lambda[f(x_0 - x) - f(x_0)] \leq \lambda(m - f(x_0)).$$

Hence, we have proved that

$$|f(x) - f(x_0)| \leq \lambda(m - f(x_0)) \quad \text{for all } x \in x_0 + \lambda V,$$

showing that f is continuous at x_0.

Finally, let $x_1 \in \operatorname{int} \operatorname{dom} f$. Then there exists a $\mu > 1$ such that

$$x_0 + \mu(x_1 - x_0) \in \operatorname{dom} f.$$

So one has for all $x \in V$

$$\begin{aligned} & f(x_1 + (1 - 1/\mu)x) = f(x_1 - (1 - 1/\mu)x_0 + (1 - 1/\mu)(x_0 + x)) \\ \leq \; & \frac{1}{\mu} f(x_0 + \mu(x_1 - x_0)) + \left(1 - \frac{1}{\mu}\right) f(x_0 + x) \\ \leq \; & \frac{1}{\mu} f(x_0 + \mu(x_1 - x_0)) + \left(1 - \frac{1}{\mu}\right) m. \end{aligned}$$

This shows that f is bounded above on $x_1 + (1 - 1/\mu)V$, and it follows as above that f is continuous at x_1. □

Corollary 3.4.2 *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a convex function on a topological vector space. Then the following are equivalent:*

(i) $\operatorname{int}\operatorname{dom} f$ *is not empty, and f is continuous on* $\operatorname{int}\operatorname{dom} f$.

(ii) $\operatorname{int}\operatorname{epi} f$ *is not empty.*

Proof. (i) $\Rightarrow$ (ii): If (i) holds, there exists a neighborhood U of some $x_0 \in X$ and a $y \in \mathbb{R}$ such that $f(x) \leq y$ for all $x \in U$. It follows that $U \times [b, +\infty) \subseteq \operatorname{epi} f$, which implies (ii).

(ii) $\Rightarrow$ (i): If $(x_0, y_0) \in \operatorname{int}\operatorname{epi} f$, there exists a neighborhood U of x_0 in X and an $\varepsilon > 0$ such that $U \times [y_0 - \varepsilon, x_0 + \varepsilon] \subseteq \operatorname{epi} f$. In particular, $f(x_0) \in \mathbb{R}$ and $\sup_{x \in U} f(x) < +\infty$. So (ii) follows from Theorem 3.4.1. □

Definition 3.4.3 *Let C be a subset of a topological vector space X.*

- *C is called a barrel if it is closed, convex, balanced and absorbing.*
- *X is called a barreled space if it is locally convex and every barrel is a neighborhood of* 0.

Remark 3.4.4 It can be shown that every Banach space is barreled. But there exist normed vector spaces that are not barreled.

Corollary 3.4.5 *Let f be a lsc proper convex function on a barreled space X. Then f is continuous on* $\operatorname{int}\operatorname{dom} f$.

Proof. Let us suppose that $\operatorname{int}\operatorname{dom} f$ is not empty. Then we can assume without loss of generality that $0 \in \operatorname{int}\operatorname{dom} f$. Choose a number $m > f(0)$. Then

$$U := \{x \in X : f(x) \leq m \text{ and } f(-x) \leq m\}$$

is closed, convex and balanced. Next, note that for every $x \in X$, the function $f^x(\lambda) := f(\lambda x)$ is a proper convex function on $\mathbb{R}$ with $0 \in \operatorname{int}\operatorname{dom} f^x$. It follows that f^x is continuous at 0. So there exists an $\varepsilon > 0$ such that $f(\lambda x) \leq m$ for all $\lambda \in [-\varepsilon, \varepsilon]$. This shows that U is absorbing and therefore, a barrel. Since X is barreled, U is a 0-neighborhood. Now the corollary follows from Theorem 3.4.1. □

3.5 Derivatives and sub-gradients

Definition 3.5.1 *Let $f : X \to \mathbb{R} \cup \{\pm\infty\}$ be a function on a normed vector space and $x_0 \in X$ such that $f(x_0) \in \mathbb{R}$. A Fréchet derivative of f at x_0 is a continuous linear functional $x^* \in X^*$ satisfying*

$$\lim_{x \neq 0,\, ||x|| \to 0} \frac{f(x_0 + x) - f(x_0) - x^*(x)}{||x||} = 0 \quad \textit{for all } x \in X.$$

Definition 3.5.2 *Let $f : X \to \mathbb{R} \cup \{\pm\infty\}$ be a function on a topological vector space and $x_0 \in X$ such that $f(x_0) \in \mathbb{R}$. A Gâteaux-derivative of f at x_0 is a continuous linear functional $x^* \in X^*$ satisfying $x^*(x) = f'(x_0; x)$ for all $x \in X$.*

Definition 3.5.3 *Let $f : X \to \mathbb{R} \cup \{\pm\infty\}$ be a function on a topological vector space and $x_0 \in X$ such that $f(x_0) \in \mathbb{R}$. The sub-differential of f at x_0 is the set $\partial f(x_0) := \partial_a f(x_0) \cap X^*$. Elements of $\partial f(x_0)$ are called sub-gradients of f at x_0.*

Exercise 3.5.4 Let $f : X \to \cup \{\pm\infty\}$ be a convex function on a topological vector space and $x_0 \in X$ such that $f(x_0) \in \mathbb{R}$. Show the following:

1. $\partial f(x_0)$ is a $\sigma(X^*, X)$-closed convex subset of X^*.

2. If the function $g(x) := f'(x_0; x)$ is continuous at $x = 0$, then

$$\partial f(x_0) = \partial_a f(x_0) = \partial g(0) = \partial_a g(0).$$

Theorem 3.5.5 *Let $f : X \to \mathbb{R} \{\pm\infty\}$ be a convex function on a topological vector space and $x_0 \in X$ such that $f(x_0) \in \mathbb{R}$. If f is continuous at x_0, then $\partial f(x_0) \neq \emptyset$.*

Proof. It follows from Theorem 3.4.1 that f is proper convex, and x_0 has a neighborhood U on which f is bounded from above. So one obtains from Theorem 2.32 that there exists $x' \in \partial_a f(x_0)$. It follows that x' is bounded from above on $U - x_0$, which by Theorem 3.2.1, impies that it is continuous. □

Lemma 3.5.6 *Let $f : X \to \mathbb{R} \cup \{\pm\infty\}$ be a lsc convex function on a topological vector space and $x_0 \in X$ such that $f(x_0) \in \mathbb{R}$. Then f is proper convex.*

Proof. Assume there exists $x_1 \in X$ such that $f(x_1) = -\infty$. Then $f(\lambda x_0 + (1 - \lambda)x_1) = -\infty$ for all $\lambda \in [0, 1)$. Since $\lambda x_0 + (1 - \lambda)x_1)$ converges to x_0 for $\lambda \to 1$, one obtains $f(x_0) = -\infty$, which contradicts the assumption. □

Lemma 3.5.7 *Let f be a proper convex function on X and $x_0 \in \operatorname{dom} f$ such that $\partial f(x_0) \neq \emptyset$. Then $f(x_0) = \underline{f}(x_0)$ and $\partial f(x_0) = \partial \underline{f}(x_0)$. In particular, $\underline{f}$ is proper convex.*

Proof. Choose $x^* \in \partial f(x_0)$. The affine function $g(x) = f(x_0) + x^*(x - x_0)$ minorizes f and equals f at x_0. So g also minorizes $\underline{f}$ and equals $\underline{f}$ at x_0. This shows $f(x_0) = g(x_0) = \underline{f}(x_0)$ and $\partial f(x_0) \subseteq \partial \underline{f}(x_0)$. $\partial f(x_0) \supseteq \partial \underline{f}(x_0)$ follows since $f(x_0) = \underline{f}(x_0)$ and $f \geq \underline{f}$. □

Theorem 3.5.8 *A lsc convex function $f : X \to \mathbb{R} \cup \{+\infty\}$ on a locally convex topological vector space equals the point-wise supremum of all its continuous affine minorants.*

Proof. If f is constantly equal to $+\infty$, the theorem is clear. So we can assume $\operatorname{dom} f \neq \emptyset$. Choose a pair $(x_0, w) \in X \times \mathbb{R}$ that does not belong to $\operatorname{epi} f$. By Corollary 3.3.9, there exists $(x^*, v) \in X^* \times \mathbb{R}$ such that

$$m := \inf_{(x,y) \in \operatorname{epi} f} (x^*(x) + yv) > x^*(x_0) + wv.$$

It follows that $v \geq 0$. If $v > 0$, one can scale and assume $v = 1$. Then $m - x^*(x)$ is an affine minorant of f whose epigraph does not contain (x_0, w). If $v = 0$, set $\lambda := m - x^*(x_0) > 0$ and choose $x_1 \in \operatorname{dom} f$. Since $(x_1, f(x_1) - 1)$ is not in $\operatorname{epi} f$, there exists $(y^*, v') \in X^* \times \mathbb{R}$ such that

$$m' := \inf_{(x,y) \in \operatorname{epi} f} (y^*(x) + yv') > y^*(x_1) + (f(x_1) - 1)v'.$$

Since $x_1 \in \operatorname{dom} f$, one must have $v' > 0$. So by scaling, one can assume $v' = 1$. Now choose

$$\delta > \frac{1}{\lambda}(w + y^*(x_0) - m')^+$$

and set $z^* := \delta x^* + y^*$. Then

$$\begin{aligned} m'' &:= \inf_{(x,y) \in \operatorname{epi} f} (z^*(x) + y) \geq \delta m + m' \\ &= \delta\lambda + \delta x^*(x) + m' > z^*(x_0) + w. \end{aligned}$$

So $m'' - z^*(x)$ is an affine minorant of f whose epigraph does not contain (x_0, w). This completes the proof of the theorem. □

3.6 Dual pairs

Definition 3.6.1 *Two vector spaces X and Y together with a bilinear function $\langle ., . \rangle : X \times Y \to \mathbb{R}$ form a dual pair if the following hold:*

(i) *For every $x \in X \setminus \{0\}$ there exists a $y \in Y$ such that $\langle x, y \rangle \neq 0$;*

(ii) *For every $y \in Y \setminus \{0\}$ there exists a $x \in X$ $\langle x, y \rangle \neq 0$.*

$\sigma(X, Y)$ is the coarsest topology on X making all $y \in Y$ continuous. It is called weak topology induced by Y. A locally convex topology τ on X is said to be consistent with Y if $(X, \tau)^* = Y$. Analogously, the weak topology $\sigma(Y, X)$ is the coarsest topology on Y such that all $x \in X$ are continuous. A locally convex topology τ on Y is consistent with X if $(Y, \tau)^* = X$.

Exercise 3.6.2 Show that the following are dual pairs:

1. $X = Y = \mathbb{R}^d$, $\langle x, y \rangle = \sum_{i=1}^d x_i y_i$;

2. $X = Y = H$ if H is a vector space with an inner product $\langle ., . \rangle$;

3. $Y = X'$ for a vector space X with $\langle x, y\rangle = y(x)$;

4. $Y = X^*$ for a Hausdorff locally convex topological vector space X with $\langle x, y\rangle = y(x)$; e.g., X could be a normed vector space;

5. $X = L^p(\Omega, \mathcal{F}, \mu)$, $Y = L^q(\Omega, \mathcal{F}, \mu)$ with $\langle x, y\rangle = \int xy d\mu$, where $(\Omega, \mathcal{F}, \mu)$ is a measure space and $1/p + 1/q = 1$.

Exercise 3.6.3 Let (X, Y) be a dual pair. Show the following:

1. For each $y \in Y$,

$$U(y) := \{x \in X : |\langle x, y\rangle| \leq 1\}$$

is a convex balanced neighborhood of 0 in X with respect to $\sigma(X, Y)$.

2.

$$\mathcal{U} := \{U(y_1) \cap \cdots \cap U(y_n) : n \in \mathbb{R},\, y_1, \ldots, y_n \in Y\}$$

is a neighborhood base of 0 in X with respect to $\sigma(X, Y)$.

3. X with the topology $\sigma(X, Y)$ is a Hausdorff locally convex topological vector space.

Exercise 3.6.4 Let H be a Hilbert space. Show the following:

(i) $||x + y||^2 + ||x - y||^2 = 2||x||^2 + 2||y||^2$ for all $x, y \in H$.

(ii) If C is a non-empty closed convex subset of H, there exists a unique $x_0 \in C$ such that

$$||x_0|| = \inf_{x \in C} ||x||.$$

(iii) If D is a non-empty closed subspace of H and $x \in H$, there exists a unique $y \in D$ such that

$$||x - y|| = \inf_{v \in D} ||x - v||.$$

This $y \in D$ satisfies

$$\langle x - y, v\rangle = 0 \quad \text{for all } v \in D.$$

In particular, $H = D + D^{\perp}$ and $D \cap D^{\perp} = \{0\}$.

(iv) If $f : H \to \mathbb{R}$ is a continuous linear functional, $f^{-1}(0)$ is a closed linear subspace of H. Show that there exists a $z \in H$ such that $f^{-1}(0)^{\perp} = \{\lambda z : \lambda \in \mathbb{R}\}$. It follows that there exists a $y \in H$ such that $f(x) = \langle x, y\rangle$ for all $x \in X$. This shows that H^* can be identified with H.

Theorem 3.6.5 (Fundamental theorem of duality)
Let X be a vector space and $x'_0, \ldots, x'_n \in X'$. Then the following are equivalent:

(i) $x'_0 = \sum_{i=1}^n \lambda_i x'_i$ *for some* $\lambda \in \mathbb{R}^n$

(ii) $\bigcap_{i=1}^n x_i'^{-1}(0) \subseteq x_0'^{-1}(0)$.

Proof. (i) $\Rightarrow$ (ii) is clear. To show (ii) $\Rightarrow$ (i), define a linear function $f : X \to \mathbb{R}^n$ by $f(x) := (x_1'(x), \dots, x_n'(x))$. Due to (ii), there exists a linear function $g : f(X) \to \mathbb{R}$ such that $x_0'(x) = g \circ f(x)$ for all $x \in X$. g can be extended to a linear function $G : \mathbb{R}^n \to \mathbb{R}$, and G has a representation of the form $G(x) = \lambda^T x$ for some $\lambda \in \mathbb{R}^n$. This shows (i). □

Theorem 3.6.6 (Duality theorem for dual pairs)
Let (X, Y) be a dual pair of vector spaces. Then $(X, \sigma(X,Y))^* = Y$ and $(Y, \sigma(Y,X))^* = X$.

Proof. First note that it follows from Definition 3.6.1 that two different elements $y_1, y_2 \in Y$ induce different continuous linear functionals on $(X, \sigma(X,Y))$.

Now pick a $x' \in X'$ that is continuous with respect to $\sigma(X,Y)$. Then there exist $y_1, \dots, y_n \in Y$ such that

$$\{x \in X : |\langle x, y_i \rangle| \le 1 \text{ for all } i = 1, \dots, n\} \subseteq \{x \in X : |x'(x)| \le 1\},$$

implying that

$$\bigcap_{i=1}^n y_i^{-1}(0) \subseteq x'^{-1}(0).$$

By Theorem 3.6.5, there exists $\lambda \in \mathbb{R}^n$ such that $x' = \sum_{i=1}^n \lambda_i y_i$, implying that $x' \in Y$. This shows $(X, \sigma(X,Y))^* = Y$. $(Y, \sigma(Y,X))^* = X$ follows by symmetry. □

Remark 3.6.7 Let X be a Hausdorff locally convex topological vector space. It follows from Theorem 3.6.6 that $(X, \sigma(X, X^*))^* = X^*$ and $(X^*, \sigma(X^*, X))^* = X$. $\sigma(X, X^*)$ is called the weak topology on X and $\sigma(X^*, X)$ the weak* topology on X^*.

For $1 < p, q < \infty$ such that $1/p + 1/q$ one has $(L^p, ||.||_p)^* = L^q$ and $L^p = (L^q, ||.||_q)^*$. But $(L^1, ||.||_1)^* = L^\infty$ and $(L^\infty, ||.||_\infty)^* = ba$, which is strictly larger than L^1.

Theorem 3.6.8 (Closed convex sets in dual pairs)
Let (X, Y) be a dual pair of vector spaces. Then all locally convex vector space topologies on X consistent with Y have the same collection of closed convex sets in X.

Proof. By Corollary 3.3.8, every proper closed convex subset C of X equals the intersection of all closed half-spaces containing C. But this intersection depends only on Y. □

Corollary 3.6.9 *Let (X, Y) be a dual pair of vector spaces. Then all locally convex vector space topologies on X consistent with Y have the same collections of lsc convex functions $f : X \to \mathbb{R} \cup \{\pm\infty\}$ and lsc quasi-convex functions $f : X \to \mathbb{R} \cup \{\pm\infty\}$.*

Proof. A function $f : X \to \mathbb{R} \cup \{\pm\infty\}$ is lsc if and only if all sub-level sets $\{x \in X : f(x) \le c\}$, $c \in \mathbb{R}$, are closed. If f is (quasi-)convex, its sub-level sets are convex. So the corollary follows from Theorem 3.6.8. □

3.7 Convex conjugates

In this whole subsection, (X, Y) is dual pair of vector spaces. X is endowed with the topology $\sigma(X, Y)$ and Y with $\sigma(Y, X)$. For instance, X could be a normed vector space and $Y = X^*$, or more generally, X could be a Hausdorff locally convex topological vector space and $Y = X^*$.

Definition 3.7.1 *The convex conjugate of a function $f : X \to \mathbb{R} \cup \{\pm\infty\}$ is the function $f^* : Y \to \mathbb{R} \cup \{\pm\infty\}$ given by*

$$f^*(y) := \sup_{x \in X} \{\langle x, y\rangle - f(x)\}.$$

The convex conjugate of a function $h : Y \to \mathbb{R} \cup \{\pm\infty\}$ is the function $h^* : X \to \mathbb{R} \cup \{\pm\infty\}$ given by

$$h^*(x) := \sup_{y \in Y} \{\langle x, y\rangle - h(y)\}.$$

Exercise 3.7.2
Consider functions $f, g : X \to \mathbb{R} \cup \{\pm\infty\}$. Show that ...

1. f^* is convex and lsc.

2. $f \geq f^{**}$

3. $f \leq g$ implies $f^* \geq g^*$

4. $f^{***} = f^*$.

Definition 3.7.3 *Let C be a subset of X. The indicator function $\delta_C : X \to \mathbb{R} \cup \{+\infty\}$ is defined to be 0 on C and $+\infty$ outside of C. The convex conjugate δ_C^* is called support function of C.*

Exercise 3.7.4 Let $f : X \to \mathbb{R}$ be a continuous affine function of the form $f(x) = \langle x, y\rangle - v$ for a pair $(y, v) \in Y \times \mathbb{R}$. Show that $f^* = v + \delta_y$ and $f^{**} = f$.

Exercise 3.7.5 Consider a function $f : X \to \mathbb{R} \cup \{\pm\infty\}$.

1. Show that the Young–Fenchel inequality holds:

$$f^*(y) \geq \langle x, y\rangle - f(x) \quad \text{for all } (x, y) \in X \times Y.$$

2. Show that if $f(x_0) \in \mathbb{R}$, the following are equivalent

(i) $y \in \partial f(x_0)$

(ii) $\langle x, y\rangle - f(x)$ achieves its supremum in x at $x = x_0$

(iii) $f(x_0) + f^*(y) = \langle x_0, y\rangle$

3. Show that if $f(x_0) = f^{**}(x_0) \in \mathbb{R}$, the following conditions are equivalent to (i)–(iii)

(iv) $x_0 \in \partial f^*(y)$

(v) $\langle x_0, v\rangle - f^*(v)$ achieves its supremum in v at $v = y$

(vi) $y \in \partial f^{**}(x_0)$

Theorem 3.7.6 (Fenchel–Moreau Theorem)
Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a function whose lsc convex hull $\underline{\text{conv}}\, f$ does not take the value $-\infty$. Then $\underline{\text{conv}}\, f = f^{**}$. In particular, if f is lsc and convex, then $f = f^{**}$.

Proof. We know that $f \geq f^{**}$. Since f^{**} is lsc and convex, one obtains $\underline{\text{conv}}\, f \geq f^{**}$. Now let h be a continuous affine minorant of $\underline{\text{conv}}\, f$. Then it also minorizes f. So one has $h = h^{**} \leq f^{**}$. But by Theorem 3.5.8, $\underline{\text{conv}}\, f$ is the point-wise supremum of its continuous affine minorants. So one gets $\underline{\text{conv}}\, f \leq f^{**}$. □

Corollary 3.7.7 *If f is a lsc proper convex function on X, then f^* is lsc proper convex.*

Proof. f^* is lsc convex for every function $f : X \to \mathbb{R} \cup \{\pm\infty\}$. If f is lsc proper convex, one obtains from Theorem 3.7.6 that $f = f^{**}$, and it follows that f^* is proper convex. □

Corollary 3.7.8 *Let C be a non-empty subset of X with closed convex hull D. Then $\delta^*_C(y) = \sup_{x \in D} \langle x, y\rangle$ and $\delta^{**}_C = \delta_D$.*

Proof. $\delta^{**}_C = \delta_D$ follows from Theorem 3.7.6 since δ_D is the lsc convex hull of δ_C. Now one obtains $\delta^*_C = \delta^{***}_C = \delta^*_D$, and the corollary follows. □

Corollary 3.7.9 *Let f be a lsc proper sub-linear function on X. Then $f = \delta^*_{\partial f(0)}$ and $f^* = \delta_{\partial f(0)}$. In particular, $f(0) = 0$ and $\partial f(0) \neq \emptyset$.*

Proof. It can easily be checked that $f^* = \delta_C$ for the set

$$C = \{y \in Y : \langle x, y\rangle \leq f(x) \text{ for all } x \in X\}.$$

By Theorem 3.7.6, one has $f = \delta^*_C$. In particular, C is non-empty, $f(0) = 0$ and $\partial f(0) = C$. □

Corollary 3.7.10 *Let $f : X \to \mathbb{R} \cup \{\pm\infty\}$ be a convex function on a normed vector space and $x_0 \in \mathbb{R}^d$ such that $f(x_0) \in \mathbb{R}$. Assume there exists a neighborhood U of x_0 and a constant $M \in \mathbb{R}_+$ such that*

$$f(x) - f(x_0) \geq -M||x - x_0|| \quad \text{for all } x \in U. \tag{3.7.4}$$

Then $\partial f(x_0) \neq \emptyset$.

Proof. It follows from condition (3.7.4) that $g(x) := f'(x_0;x) \geq -M||x||$, and therefore, $\underline{g}(x) \geq -M||x||$ for all $x \in X$. So one obtains from Corollary 3.7.9 that $\underline{g}(0) = 0 = g(0)$ and $\partial \underline{g}(0) \neq \emptyset$, which implies that $\partial f(x_0) = \partial g(0) \neq \emptyset$. □

Theorem 3.7.11 *Let f be a proper convex function on X and $x_0 \in \operatorname{dom} f$. If f is continuous at x_0, then*

$$f'(x_0;x) = \sup_{y \in \partial f(x_0)} \langle x, y \rangle, \quad x \in X. \tag{3.7.5}$$

Proof. Consider the sub-linear function $g(x) = f'(x_0;x)$. It follows from Theorem 3.5.5 that $\partial g(0) = \partial f(x_0) \neq \emptyset$. Since g is bounded above on a neighborhood of 0, one obtains from Theorem 3.4.1 that g is continuous on X. So it follows from Corollary 3.7.9 that $g = \delta^*_C$ for $C = \partial g(0) = \partial f(x_0)$, which proves the theorem. □

3.8 Inf-convolution

Definition 3.8.1 *Let $f_j : X \to \mathbb{R} \cup \{+\infty\}$, $j = 1, \dots, n$, be functions on a vector space. The inf-convolution of f_j, $j = 1, \dots, n$, is the function*

$$\square_{j=1}^n f_j(x) := \inf_{x_1 + \dots + x_n = x} \sum_{j=1}^n f_j(x_j).$$

The inf-convolution is said to be exact if the infimum is attained.

Lemma 3.8.2 *Let $f_j : X \to \mathbb{R} \cup \{+\infty\}$, $j = 1, \dots, n$, be convex functions on a vector space X. Then $f = \square_{j=1}^n f_j$ is convex.*

Proof. If $f \equiv +\infty$, it is convex. If not, let $(x,v), (y,w) \in \operatorname{epi} f$, $\lambda \in (0,1)$ and $\varepsilon > 0$. There exist x_j and y_j, $j = 1, \dots, n$, such that $\sum_{j=1}^n x_j = x$, $\sum_{j=1}^n f(x_j) \leq v + \varepsilon$, $\sum_{j=1}^n y_j = y$ and $\sum_{j=1}^n f(y_j) \leq w + \varepsilon$. Set $z_j = \lambda x_j + (1-\lambda) y_j$. Then $z := \sum_{j=1}^n z_j = \lambda x + (1-\lambda) y$ and

$$f(z) \leq \sum_{j=1}^n f_j(z_j) \leq \sum_{j=1}^n \lambda f_j(x_j) + (1-\lambda) f(y_j) \leq \lambda v + (1-\lambda) w + \varepsilon.$$

It follows that $f(z) \leq \lambda v + (1-\lambda) w$, which shows that $\operatorname{epi} f$ and f are convex. □

Lemma 3.8.3 *Let f_j, $j = 1, \dots, n$, be proper convex functions on a topological vector space X and denote $f = \square_{j=1}^n f_j$. Assume $f(x_0) = \sum_j f_j(x_j) < +\infty$ for some x_j summing up to x_0 and f_1 is bounded from above on a neighborhood of x_1. Then f is a proper convex function, $x_0 \in \operatorname{int} \operatorname{dom} f$ and f is continuous on $\operatorname{int} \operatorname{dom} f$.*

Proof. By definition of f, one has

$$f(x_0+x)-f(x_0) \leq f_1(x_1+x)+\sum_{j=2}^{n} f_j(x_j)-\sum_{j=1}^{n} f_j(x_j)=f_1(x_1+x)-f_1(x_1)$$

for all $x \in X$. It follows that f is bounded from above on a neighborhood of x_0. Now the lemma is a consequence of Theorem 3.4.1. □

Lemma 3.8.4 *Consider functions $f_j : X \to \mathbb{R} \cup \{+\infty\}$, $j = 1, \ldots, n$, on a topological vector space and denote $f = \Box_{j=1}^n f_j$. Assume $f(x_0) = \sum_{j=1}^n f_j(x_j) < +\infty$ for some x_j summing up to x_0. Then $\partial f(x_0) = \bigcap_{j=1}^n \partial f_j(x_j)$.*

Proof. Assume $x^* \in \partial f(x_0)$ and $x \in X$. Then

$$f_1(x_1+x)-f_1(x_1)=f_1(x_1+x)+\sum_{j=2}^{n} f_j(x_j)-\sum_{j=1}^{n} f_j(x_j) \geq f(x_0+x)-f(x_0) \geq x^*(x).$$

Hence $x^* \in \partial f_1(x_1)$, and it follows by symmetry that $\partial f(x_0) \subseteq \bigcap_{j=1}^n \partial f_j(x_j)$. On the other hand, if $x^* \in \bigcap_{j=1}^n \partial f_j(x_j)$ and $x \in X$, choose y_j such that $\sum_{j=1}^n y_j = x_0 + x$. Then

$$\sum_{j=1}^{n} f_j(y_j) \geq \sum_{j=1}^{n} f_j(x_j)+x^*(y_j-x_j)=\sum_{j=1}^{n} f_j(x_j)+x^*(x).$$

So $f(x_0+x)-f(x_0) \geq x^*(x)$, and the lemma follows. □

Lemma 3.8.5 *Let f_j, $j = 1, \ldots, n$, be proper convex functions on a topological vector space X and denote $f = \Box_{j=1}^n f_j$. Assume $f(x_0) = \sum_j f_j(x_j) < +\infty$ for some x_j summing up to x_0 and f_1 is Gâteaux-differentiable at x_1 with $f_1'(x_1; x) = x^*(x)$ for some $x^* \in X^*$. Then f is Gâteaux-differentiable at x_0 with $f'(x_0; x) = x^*(x)$. In particular, $\partial f(x_0) = \{x^*\}$.*

Proof. One has

$$f(x_0+x)-f(x_0) \leq f_1(x_1+x)+\sum_{j=2}^{n} f_j(x_j)-\sum_{j=1}^{n} f_j(x_j)=f_1(x_1+x)-f_1(x_1)$$

for all $x \in X$. It follows that the directional derivative $g(x) := f'(x_0; x)$ satisfies

$$g(x) \leq f_1'(x_1; x) = x^*(x)$$

for all $x \in X$. But by Lemma 3.8.2, f is convex. So g is sub-linear, and it follows that $g(x) = x^*(x)$. □

Lemma 3.8.6 *Let (X, Y) be a dual pair of vector spaces and $f_j : X \to \mathbb{R} \cup \{+\infty\}$, $j = 1, \ldots, n$, functions none of which is identically equal to $+\infty$. Then $\left(\Box_{j=1}^n f_j\right)^* = \sum_{j=1}^n f_j^*$.*

Proof.

$$\left(\square_{j=1}^{n} f_j\right)^* (y) = \sup_x (\langle x, y\rangle - \square_{j=1}^{n} f_j(x)) = \sup_{x_1,\ldots,x_n} \sum_{j=1}^{n} (\langle x_j, y\rangle - f_j(x_j)) = \sum_{j=1}^{n} f_j^*(y).$$

□

Chapter 4

Convex Optimization

In this chapter we study the minimization problem

$$\inf_{x \in X} f(x) \qquad \text{(P)}$$

for a function $f : X \to \mathbb{R} \cup \{\pm\infty\}$ on a vector space. If one wants to constrain x to be in a subset $C \subseteq X$, one can replace f with $f + \delta_C$.

4.1 Perturbation and the dual problem

We assume that there exist vector spaces Y, W, Z such that (X, W) and (Y, Z) are dual pairs. A perturbation of f is a function $F : X \times Y \to \mathbb{R} \cup \{\pm\infty\}$ such that $f(x) = F(x, 0)$. Note that $((X, Y), (W, Z))$ is again a dual pair with pairing $\langle (x, y), (w, z) \rangle := \langle x, w \rangle + \langle y, z \rangle$. The value function associated with F is the function $u : Y \to \mathbb{R} \cup \{\pm\infty\}$ given by

$$u(y) := \inf_{x \in X} F(x, y).$$

In particular, $u(0) = \inf_x f(x)$.

The dual problem of (P) is

$$\sup_{z \in Z} -F^*(0, z) = -\inf_{z \in Z} F^*(0, z), \qquad \text{(D)}$$

where F^* is the convex conjugate

$$F^*(w, z) := \sup_{(x,y) \in X \times Y} \left(\langle x, w \rangle + \langle y, z \rangle - F(x, y) \right).$$

The dual value function is the function $v : W \to \mathbb{R} \cup \{\pm\infty\}$, given by

$$v(w) := \sup_{z \in Z} -F^*(w, z) = -\inf_{z \in Z} F^*(w, z).$$

Proposition 4.1.1 (Weak Duality)
One always has $u(0) \geq v(0)$.

Proof. By the Young–Fenchel inequality, one has

$$F^*(w,z) \geq \langle x, w \rangle + \langle y, z \rangle - F(x,y) \quad \text{for all } x, y, w, z.$$

In particular,

$$F(x,0) \geq -F^*(0,z) \quad \text{for all } x, z,$$

and the proposition follows. □

The dual problem of (D) is

$$\sup_{x \in X} -F^{**}(x,0) = -\inf_{x \in X} F^{**}(x,0), \qquad \text{(BD)}.$$

If $F = F^{**}$, then (BD) is equivalent to (P). In the general case, one obtains from Proposition 4.1.1 applied to (D) and (BD) that

$$\sup_z -F^*(0,z) = -\inf_z F^*(0,z) \leq \inf_x F^{**}(x,0) \leq \inf_x F(x,0),$$

and both inequalities can be strict. Note that the first term is a "concave max", the third term a "convex min", and the last term a "min" of a general function.

Lemma 4.1.2 *If* F *is convex, then* $u : Y \to \mathbb{R} \cup \{\pm\infty\}$ *is convex too.*

Proof. Assume there exist $(y_1, r_1), (y_2, r_2) \in \operatorname{epi} u$. Choose $\lambda \in (0,1)$ and $\varepsilon > 0$. There are $x_1, x_2 \in X$ such that

$$F(x_i, y_i) \leq r_i + \varepsilon, \quad i = 1, 2.$$

So

$$\begin{aligned} & u(\lambda y_1 + (1-\lambda) y_2) \leq F(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda) y_2) \\ \leq \; & \lambda F(x_1, y_1) + (1-\lambda) F(x_2, y_2) \leq \lambda r_1 + (1-\lambda) r_2 + \varepsilon, \end{aligned}$$

which shows that $\operatorname{epi} u$ and u are convex. □

Exercise 4.1.3 Show that $u^*(z) = F^*(0,z)$ and $v(0) = u^{**}(0)$. In particular, strong duality $u(0) = v(0)$ is equivalent to $u(0) = u^{**}(0)$.

Definition 4.1.4 *Problem* (P) *is called normal if* $u(0) = v(0) \in \mathbb{R}$. *It is called stable if it is normal and problem* (D) *has a solution.*

Lemma 4.1.5 *Assume that* F *is convex. Then* (P) *is normal if and only if* $u(0) = \underline{u}(0) \in \mathbb{R}$.

Proof. If (P) is normal, then $u(0) = v(0) = u^{**}(0) \in \mathbb{R}$, which implies $u(0) = \underline{u}(0) \in \mathbb{R}$. On the other hand, we know from Lemma 4.1.2 that u is convex. So if $u(0) = \underline{u}(0) \in \mathbb{R}$, one obtains from Lemma 3.5.6 that $\underline{u}$ is a lsc proper convex function, and it follows from Theorem 3.7.6 that $u(0) = \underline{u}(0) = u^{**}(0) = v(0) \in \mathbb{R}$.□

Proposition 4.1.6 (P) *is stable if and only if* $u(0) \in \mathbb{R}$ *and* $\partial u(0) \neq \emptyset$.

Proof. If (P) is stable, then there exists z such that $u(0) = v(0) = -F^*(0, z) \in \mathbb{R}$. So one has

$$u(0) = v(0) = \langle 0, z\rangle - u^*(z) \in \mathbb{R},$$

and it follows that $z \in \partial u(0)$. On the other hand, if $u(0) \in \mathbb{R}$ and $z \in \partial u(0)$, then

$$u(0) = \langle 0, z\rangle - u^*(z) = -F^*(0, z),$$

which by weak duality, implies that z is a solution of (D). □

Theorem 4.1.7 (Fundamental duality formula of convex analysis)
Assume F *is convex and* $u(0) \in \mathbb{R}$. *Then* (P) *is stable if one of the following conditions holds:*

(i) *There exists a neighborhood* U *of* 0 *in* Y *such that* $\sup_{y \in U} u(y) < +\infty$.

(ii) Y *is barreled,* u *is lsc and* $0 \in \operatorname{int} \operatorname{dom} u$;

(iii) Y *is a normed vector space and there exists a constant* $M \in \mathbb{R}_+$ *such that*

$$u(y) - u(0) \geq -M||y||$$

for all y *in a neighborhood of* 0 *in* Y;

(iv) $Y = \mathbb{R}^d$, u *does not take the value* $-\infty$ *and* $0 \in \operatorname{ri} \operatorname{dom} u$;

(v) $Y = \mathbb{R}^d$, $u(y) < +\infty$ *for* y *in a neighborhood of* 0 *in* Y.

Proof. By Proposition 4.1.6, it is enough to show that $\partial u(0) \neq \emptyset$. We know from Lemma 4.1.2 that u is convex. So $\partial u(0) \neq \emptyset$ follows from each of the conditions (i)–(v). □

In the following, consider functions $f : X \to \mathbb{R} \cup \{+\infty\}$ and $g : Y \to \mathbb{R} \cup \{+\infty\}$. Moreover, let $A : X \to Y$ be a continuous linear function and define the adjoint $A^* : Z \to W$ by $\langle x, A^*z\rangle := \langle Ax, z\rangle$. Denote

$$p := \inf_{x \in X} \{f(x) + g(Ax)\} \qquad (\mathrm{P-FR})$$

$$d := \sup_{z \in Z} \{-f^*(-A^*z) - g^*(z)\} \qquad (\mathrm{D-FR})$$

As a consequence of Proposition 4.1.1 and Theorem 4.1.7, one obtains the following

Corollary 4.1.8 (Fenchel–Rockafellar duality theorem)
One always has $p \geq d$. Moreover, $p = d$ and (D-FR) *has a solution if f and g are convex, $p \in \mathbb{R}$ and one of the following conditions holds:*

(i) *The function $h(y) := \inf_x \{f(x) + g(Ax + y)\}$ satisfies $\sup_{y \in U} h(y) < +\infty$ for some neighborhood U of 0 in Y;*

(ii) *Y is barreled, h is lsc and $0 \in \operatorname{int} \operatorname{dom} h$;*

(iii) *Y is a normed vector space and there exists a constant $M \in \mathbb{R}_+$ such that*

$$h(y) - h(0) \geq -M||y||$$

for all y in a neighborhood of 0 in Y;

(iv) *$Y = \mathbb{R}^d$, h does not take the value $-\infty$ and $0 \in \operatorname{ri} \operatorname{dom} u$;*

(v) *$Y = \mathbb{R}^d$, $h(y) < +\infty$ for y in a neighborhood of 0 in Y.*

Proof. Define the function $F : X \times Y \to \mathbb{R} \cup \{+\infty\}$ by

$$F(x, y) := f(x) + g(Ax + y).$$

Then

$$\begin{aligned} F^*(w, z) &= \sup_{x,y} \{\langle x, w\rangle + \langle y, z\rangle - f(x) - g(Ax + y)\} \\ &= \sup_{x,y} \{\langle x, w\rangle + \langle y - Ax, z\rangle - f(x) - g(y)\} \\ &= \sup_{x,y} \{\langle x, w - A^*z\rangle + \langle y, z\rangle - f(x) - g(y)\} \\ &= f^*(w - A^*z) + g^*(z). \end{aligned}$$

So $u(0) = p$ and $v(0) = d$, and it follows from Proposition 4.1.1 that $p \geq d$. The rest of the corollary follows from Theorem 4.1.7. □

Example 4.1.9 Let A be an $m \times n$-matrix, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. Denote by $p \in [-\infty, \infty]$ the value of the primal problem

(P) minimize $c^T x$ subject to $Ax = b$ and $x \geq 0$

and by $d \in [-\infty, \infty]$ the value of the dual problem

(D) maximize $b^T y$ subject to $A^T y \leq c$.

If one sets

$$f(x) = c^T x + \delta_{\mathbb{R}^n_+}(x) \quad \text{and} \quad g(y) = \delta_b(y),$$

then (P) corresponds to the problem (P-FR) and (D) to (D-FR). So one obtains from Proposition 4.1.1 that $p \geq d$.

Corollary 4.1.10 (Sandwich Theorem)
Let $f : X \to \mathbb{R} \cup \{+\infty\}$ and $g : Y \to \mathbb{R} \cup \{+\infty\}$ be convex functions and $A : X \to Y$ a continuous linear function. Assume $f(x) \geq -g(Ax)$ for all $x \in X$ and one of the conditions (i)–(v) *of Corollary 4.1.8 holds. Then there exist $z \in Z$ and $r \in \mathbb{R}$ such that*

$$f(x) \geq \langle x, A^* z \rangle - r \geq -g(Ax) \quad \textit{for all } x \in X.$$

Proof. It follows from Corollary 4.1.8 that there exists a $z \in Z$ such that

$$0 \leq \inf_{x \in X} \{f(x) + g(Ax)\} = -f^*(A^* z) - g^*(-z).$$

Choose $r \in \mathbb{R}$ such that $g^*(-z) \leq -r \leq -f^*(A^* z)$. Then

$$f(x) - \langle x, A^* z \rangle \geq -f^*(A^* z) \geq -r \quad \text{for all } x \in X,$$

and

$$\langle y, -z \rangle - g(y) \leq g^*(-z) \leq -r \quad \text{for all } y \in Y. \tag{4.1.1}$$

Choosing $y = Ax$ in (4.1.1) gives

$$\langle Ax, -z \rangle - g(Ax) \leq -r,$$

which is equivalent to

$$\langle x, A^* z \rangle - r \geq -g(Ax) \quad \text{for all } x \in X.$$

□

Corollary 4.1.11 (Subdifferential Calculus)
Let $f : X \to \mathbb{R} \cup \{+\infty\}$ and $g : Y \to \mathbb{R} \cup \{+\infty\}$ be convex functions and $A : X \to Y$ a continuous linear function. Then

$$\partial f(x) + A^* \partial g(Ax) \subseteq \partial(f + g \circ A)(x) \quad \textit{for all } x \in X.$$

Moreover, if $x \in \operatorname{dom} f$ and $\sup_{y \in U} g(y) < +\infty$ for some neighborhood U of Ax, then the inclusion is an equality.

Proof. That the inclusion holds for all $x \in X$ is straightforward to check. Now assume that $x \in \operatorname{dom} f$ and $\sup_{y \in U} g(y) < +\infty$ for some neighborhood U of Ax. If there exists a $w \in \partial(f + g \circ A)(x)$, then the mapping

$$x' \mapsto f(x') + g(Ax') - \langle x', w \rangle$$

takes its minimum at $x' = x$, and by shifting f, one can assume that this minimum is 0. Then it follows from the sandwich theorem that there exist $z \in Z$ and $r \in \mathbb{R}$ such that

$$f(x') - \langle x', w \rangle \geq \langle x', A^* z \rangle - r \geq -g(Ax') \quad \text{for all } x' \in X. \tag{4.1.2}$$

In particular,

$$f(x) - \langle x, w \rangle = \langle x, A^* z \rangle - r = -g(Ax). \tag{4.1.3}$$

By subtracting (4.1.3) from (4.1.2), one obtains that $w + A^* z \in \partial f(x)$ and

$$g(Ax') - g(Ax) \geq \langle Ax' - Ax, -z \rangle \quad \text{for all } x \in X.$$

Moreover, it follows from the assumptions that g is proper convex and continuous at Ax. So $g'(Ax; y)$ is a real-valued continuous sub-linear function on Y that dominates $\langle ., -z \rangle$ on the subspace $\{Ax' : x' \in X\}$. By Hahn–Banach, there exists $\tilde{z} \in Z$ such that $\langle Ax', \tilde{z} \rangle = \langle Ax', z \rangle$ for all $x' \in X$ and $g'(Ax; y) \geq \langle y, -\tilde{z} \rangle$ for all $y \in Y$. It follows that $-\tilde{z} \in \partial g(Ax)$ and $A^* \tilde{z} = A^* z$. So $w = w + A^* z - A^* \tilde{z} \in \partial f(x) + A^* \partial g(Ax)$. □

Corollary 4.1.12 (Sum Rule)
Let $f, g : X \to \mathbb{R} \cup \{+\infty\}$ be convex functions. Then

$$\partial f(x) + \partial g(x) \subseteq \partial (f + g)(x) \quad \textit{for all } x \in X.$$

Moreover, if $x \in \operatorname{dom} f$ and $\sup_{y \in U} g(y) < +\infty$ for some neighborhood U of x, then the inclusion is an equality.

Proof. Choose $X = Y$ and $A = \mathrm{id}$ in Corollary 4.1.11. □

Corollary 4.1.13 (Chain Rule)
Let $g : Y \to \mathbb{R} \cup \{+\infty\}$ be a convex function and $A : X \to Y$ a continuos linear function. Then

$$A^* \partial g(Ax) \subseteq \partial (g \circ A)(x) \quad \textit{for all } x \in X.$$

Moreover, if $\sup_{y \in U} g(y) < +\infty$ for some neighborhood U of Ax, then the inclusion is an equality.

Proof. Choose $f \equiv 0$ in Corollary 4.1.11. □

Corollary 4.1.14 *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a convex function and C a non-empty convex subset of X. If $0 \in \partial f(x_0) + \partial \delta_C(x_0)$ for some $x_0 \in C$, then x_0 solves the optimization problem*

$$\min_{x \in C} f(x). \tag{4.1.4}$$

On the other hand, if $x_0 \in C$ solves (4.1.4) and $\sup_{x \in U} f(x) < +\infty$ for a neighborhood U of x_0, then $0 \in \partial f(x_0) + \partial \delta_C(x_0)$.

Proof. The minimization problem (4.1.4) is equivalent to

$$\min_{x \in X} \{f(x) + \delta_C(x)\}, \tag{4.1.5}$$

and $x_0 \in C$ solves (4.1.5) if and only if $0 \in \partial (f + \delta_C)(x_0)$, which by Corollary 4.1.12 follows if $0 \in \partial f(x_0) + \partial \delta_C(x_0)$. Moreover, if $\sup_{x \in U} f(x) < +\infty$ for a neighborhood U of x_0, one obtains from Corollary 4.1.12 that $\partial f(x_0) + \partial \delta_C(x_0) = \partial (f + \delta_C)(x_0)$. □

4.2 Lagrangians and saddle points

Definition 4.2.1 *A saddle point of a function* $L : X \times Z \to \mathbb{R} \cup \{\pm\infty\}$ *is a pair* $(\bar{x}, \bar{z}) \in X \times Z$ *satisfying*

$$\sup_z L(\bar{x}, z) \leq L(\bar{x}, \bar{z}) \leq \inf_x L(x, \bar{z}).$$

Lemma 4.2.2 *For every function* $L : X \times Z \to \mathbb{R} \cup \{\pm\infty\}$*, one has*

$$\sup_z \inf_x L(x, z) \leq \inf_x \sup_z L(x, z), \tag{4.2.6}$$

and if L *has a saddle point* $(\bar{x}, \bar{z})$*, then*

$$\sup_z \inf_x L(x, z) = L(\bar{x}, \bar{z}) = \inf_x \sup_z L(x, z).$$

Proof. For every x', one has

$$\sup_z \inf_x L(x, z) \leq \sup_z L(x', z),$$

and one obtains (4.2.6). If $(\bar{x}, \bar{z})$ is a saddle point of L, then

$$\inf_x \sup_z L(x, z) \leq \sup_z L(\bar{x}, z) \leq L(\bar{x}, \bar{z}) \leq \inf_x L(x, \bar{z}) \leq \sup_z \inf_x L(x, z),$$

and the lemma follows. □

Now we assume that $-L$ is the y-conjugate of a function $F : X \times Y \to \mathbb{R} \cup \{\pm\infty\}$:

$$L(x, z) = \inf_{y \in Y} \{F(x, y) - \langle y, z \rangle\}. \tag{4.2.7}$$

Then L is called the Lagrangian of the problem (P) related to the perturbation F.

Lemma 4.2.3 *If* L *is of the form* (4.2.7)*, then it is concave and usc in* z*. If moreover,* F *is convex, then* L *is convex in* x*.*

Proof. That L is concave and usc in z is clear. That L is convex in x if F is convex, follows as in the proof of Lemma 4.1.2. □

Lemma 4.2.4 *Assume* L *is of the form* (4.2.7)*. Then*

$$F^*(w, z) = \sup_x \{\langle x, w \rangle - L(x, z)\}.$$

In particular,

$$\sup_z -F^*(0, z) = \sup_z \inf_x L(x, z).$$

Proof.

$$\begin{aligned} F^*(w,z) &= \sup_{x,y} \{\langle x,w\rangle + \langle y,z\rangle - F(x,y)\} \\ &= \sup_{x} \{\langle x,w\rangle - L(x,z)\}. \end{aligned}$$

□

Lemma 4.2.5 *If L is of the form (4.2.7) for a lsc convex function $F : X \times Y \to \mathbb{R} \cup \{+\infty\}$, then*

$$F(x,y) = \sup_z \{\langle y,z\rangle + L(x,z)\}.$$

In particular,

$$\inf_x F(x,0) = \inf_x \sup_z L(x,z).$$

Proof. For fixed x, $F(x,.)$ is identically equal to $+\infty$ or lsc proper convex. So one obtains from Theorem 3.7.6 that

$$F(x,y) = \sup_z \{\langle y,z\rangle + L(x,z)\}.$$

□

Lemma 4.2.6 *Let L be of the form (4.2.7) for a lsc convex F and $(\bar{x},\bar{z}) \in X \times Z$. Then the following two are equivalent:*

(i) *$(\bar{x},\bar{z})$ is a saddle point of L*

(ii) *$\bar{x}$ is a solution of the primal problem (P), $\bar{z}$ is a solution of the dual problem (D), and both problems have the the same value.*

If (i)–(ii) hold, then the value of (P) and (D) is equal to $L(\bar{x},\bar{z})$.

Proof. By Lemmas 4.2.2, 4.2.4 and 4.2.5, one has

$$\inf_x F(x,0) = \inf_x \sup_z L(x,z) \geq \sup_z \inf_x L(x,z) = \sup_z -F^*(0,z). \qquad (4.2.8)$$

If $(\bar{x},\bar{z})$ is a saddle point of L, one obtains from Lemmas 4.2.2, 4.2.4 and 4.2.5 that

$$F(\bar{x},0) = L(\bar{x},\bar{z}) = -F^*(0,\bar{z}).$$

On the other hand, if (ii) holds, one obtains from (4.2.8) that

$$\sup_z L(\bar{x},z) = \inf_x L(x,\bar{z}),$$

which implies that $(\bar{x},\bar{z})$ is a saddle point of L. □

Proposition 4.2.7 *Let L be of the form (4.2.7) for a lsc convex F and assume the primal problem (P) is stable. Then for fixed $\bar{x} \in X$, the following two are equivalent:*

(i) *$\bar{x}$ is a solution of the primal problem (P);*

(ii) *There exists a $\bar{z} \in Z$ such that $(\bar{x},\bar{z})$ is a saddle point of L.*

Proof. (i) ⇒ (ii) follows from stability and Lemma 4.2.6. (ii) ⇒ (i) is a consequence of Lemma 4.2.6. □

4.3 Karush–Kuhn–Tucker-type conditions

Let $f, g_1, \dots, g_m : X \to \mathbb{R} \cup \{\pm\infty\}$ be functions and C a non-empty subset of X such that

$$f(x), g_1(x), \dots, g_m(x) \in \mathbb{R} \quad \text{for all } x \in C.$$

We consider the constraint minimization problem:

$$\inf f(x) \quad \text{subject to } x \in C \text{ and } g_i(x) \le 0 \text{ for all } i = 1, \dots, m. \tag{CP}$$

Let us define the Lagrange functions

$$L : C \times \mathbb{R}^m_+ \to \mathbb{R} \quad \text{and} \quad M : C \times \mathbb{R}^{m+1}_+ \to \mathbb{R}$$

by

$$L(x, z) = f(x) + z^T g(x) \quad \text{and} \quad M(x, z_0, z) = z_0 f(x) + z^T g(x),$$

where $z = (z_1, \dots, z_m) \in \mathbb{R}^m_+$ and $z_0 \in \mathbb{R}_+$.

We call $(\bar{x}, \bar{z}) \in C \times \mathbb{R}^m_+$ a saddle point of L on $C \times \mathbb{R}^m_+$ if

$$L(\bar{x}, z) \le L(\bar{x}, \bar{z}) \le L(x, \bar{z}) \quad \text{for all } (x, z) \in C \times \mathbb{R}^m_+.$$

The following is called Slater condition:

(SC) There exists $x_0 \in C$ such that $g_i(x_0) < 0$ for all $i = 1, \dots, m$.

For given $\bar{x} \in C$ we consider the following conditions:

(S) $\bar{x}$ is a solution of (CP);

(SP) There exists $\bar{z} \in \mathbb{R}^m_+$ such that $(\bar{x}, \bar{z})$ is a saddle point of L on $C \times \mathbb{R}^m_+$;

(L) There exists $\bar{z} \in \mathbb{R}^m_+$ such that the following hold:

- (i) $L(\bar{x}, \bar{z}) = \min_{x \in C} L(x, \bar{z})$
- (ii) $g_i(\bar{x}) \le 0$ and $\bar{z}_i g_i(\bar{x}) = 0$ for all $i = 1, \dots, m$;

(M) There exists $(\bar{z}_0, \bar{z}) \in \mathbb{R}^{m+1}_+ \setminus \{0\}$ such that the following hold:

- (i) $M(\bar{x}, \bar{z}_0, \bar{z}) = \min_{x \in C} M(x, \bar{z}_0, \bar{z})$
- (ii) $g_i(\bar{x}) \le 0$ and $\bar{z}_i g_i(\bar{x}) = 0$ for all $i = 1, \dots, m$.

Theorem 4.3.1 *Let $\bar{x} \in C$. Then one has*

(i) (SP) $\Leftrightarrow$ (L) $\Rightarrow$ (S);

(ii) *If $C, f, g_1, \dots, g_m$ are convex, then* (S) $\Rightarrow$ (M);

(iii) *If $C, f, g_1, \dots, g_m$ are convex and* (SC) *holds, then* (SP) $\Leftrightarrow$ (L) $\Leftrightarrow$ (S) $\Leftrightarrow$ (M).

Proof. (i) First, assume that $(\bar{x}, \bar{z})$ is a saddle point of L on $C \times \mathbb{R}^m_+$. Then $L(\bar{x}, \bar{z}) = \min_{x \in C} L(x, \bar{z}) \in \mathbb{R}$. Therefore, one obtains from $\max_{z \in \mathbb{R}^m_+} L(\bar{x}, z) = L(\bar{x}, \bar{z})$ that $g_i(\bar{x}) \le 0$ and $\bar{z}_i g_i(\bar{x}) = 0$ for all $i = 1, \dots, m$.

On the other hand, if (L) holds, then $L(\bar{x}, \bar{z}) \le L(x, \bar{z})$ for all $x \in C$, and $L(\bar{x}, z) = f(\bar{x}) + z^T g(\bar{x}) \le f(\bar{x}) + \bar{z}^T g(\bar{x}) = L(\bar{x}, \bar{z})$. This shows that $(\bar{x}, \bar{z})$ is a saddle point. Moreover, it follows from (L) that $f(\bar{x}) = L(\bar{x}, \bar{z}) \le L(x, \bar{z}) \le f(x)$ for all $x \in C$ satisfying $g_i(x) \le 0$ for all $i = 1, \dots, m$.

To show (ii), assume that $C, f, g_1, \dots, g_m$ are convex. Denote

$$K := \operatorname{conv} \{(f(x) - f(\bar{x}), g_1(x), \dots, g_m(x)) : x \in D\} \subseteq \mathbb{R}^{m+1}.$$

Condition (S) implies $K \cap \operatorname{int} \mathbb{R}^{m+1}_- = \emptyset$. Indeed, otherwise there would exist $x_1, \dots, x_n \in C$ and $\lambda_1, \dots, \lambda_n \ge 0$ such that $\sum_j \lambda_j = 1$ and

$$\sum_{j=1}^n \lambda_j (f(x_j) - f(\bar{x}), g_1(x_j), \dots, g_m(x_j)) \in \operatorname{int} \mathbb{R}^{m+1}_-.$$

But this would imply $\sum_j \lambda_j x_j \in C$, $f(\sum_j \lambda_j x_j) \le \sum_j \lambda_j f(x_j) < f(\bar{x})$ and $g_i(\sum_j \lambda_j x_j) \le \sum_j \lambda_j g_i(x_j) \le 0$, a contradiction to (S). Therefore there exists $(\bar{z}_0, \bar{z}) \in \mathbb{R}^{m+1} \setminus \{0\}$ such that

$$\inf_{v \in K} \langle v, (\bar{z}_0, \bar{z}) \rangle \ge \sup_{w \in \mathbb{R}^{m+1}_-} \langle w, \bar{z} \rangle .$$

It follows that $(\bar{z}_0, \bar{z}) \in \mathbb{R}^{m+1}_+ \setminus \{0\}$ and

$$\bar{z}_0 f(x) + \bar{z}^T g(x) \ge \bar{z}_0 f(\bar{x}) \quad \text{for all } x \in C.$$

In particular,

$$\bar{z}_0 f(\bar{x}) + \bar{z}^T g(\bar{x}) \ge \bar{z}_0 f(\bar{x}) \ge \bar{z}_0 f(\bar{x}) + \bar{z}^T g(\bar{x}).$$

So $\bar{z}_i g_i(\bar{x}) = 0$ for all i and $M(\bar{x}, \bar{z}_0, \bar{z}) \le M(x, \bar{z}_0, \bar{z})$ for all $x \in C$.

(iii) We show that if $C, f, g_1, \dots, g_m$ are convex and (SC) holds, then (M) $\Rightarrow$ (L). So assume (M) holds for some $(\bar{z}_0, \bar{z}) \in \mathbb{R}^{m+1}_+ \setminus \{0\}$. If $\bar{z}_0 = 0$, one has

$$0 > \bar{z}^T g(x_0) = M(x_0, \bar{z}_0, \bar{z}) \ge M(\bar{x}, \bar{z}_0, \bar{z}) = \bar{z}^T g(\bar{x}) = 0,$$

a contradiction. So $\bar{z}_0 > 0$. By rescaling, one can assume $\bar{z}_0 = 1$. Then (L) holds. □

Now for given $\bar{x} \in C$, consider the Karush–Kuhn–Tucker condition:

(KKT) There exists $\bar{z} \in \mathbb{R}^m_+$ such that the following hold:

(i) $0 \in \partial f(\bar{x}) + \sum_{i=1}^m \bar{z}_i \partial g_i(\bar{x}) + \partial \delta_C(\bar{x})$

(ii) $g_i(\bar{x}) \le 0$ and $\bar{z}_i g_i(\bar{x}) = 0$ for all $i = 1, \dots, m$;

Theorem 4.3.2 *Assume $C, f, g_1, \dots, g_m$ are convex and let $\bar{x} \in C$. Then the following hold:*

(i) (KKT) $\Rightarrow$ (S);

(ii) *If* $f, g_1, \ldots, g_m$ *are continuous at* $\bar{x}$ *and* (SC) *is satisfied, then* (KKT) $\Leftrightarrow$ (S).

Proof. (i) If (KKT) holds, it follows from Corollary 4.1.14 that $\bar{z}$ satisfies (L), which by Theorem 4.3.1, implies (S).

(ii) We know that under (SC), (S) implies (L). So there exists $\bar{z} \in \mathbb{R}^m_+$ such that $0 \in \partial(f + \bar{z}^T g + \delta_C)(\bar{x})$. But if $f, g_1, \ldots, g_m$ are continuous at $\bar{x}$, one obtains from Corollary 4.1.12 that $\partial(f + \bar{z}^T g + \delta_C)(\bar{x}) = \partial f(\bar{x}) + \sum_{i=1}^m \bar{z}_i \partial g_i(\bar{x}) + \partial \delta_C(\bar{x})$. So (KKT) holds. $\square$