Chapter 1

Convexity, optimization, and convex duality

The purpose of this chapter is to cover some background theory in convexity, optimization, and convex duality. The structure of this chapter is as follows: Section 1.1 recalls some basic notions of convexity theory, such as convex sets, convex functions and properties of these. In Section 1.2 we consider a weaker requirement than convexity, namely quasiconvexity. Section 1.3 covers some of the most central theorems and ideas of optimization theory. In Section 1.4 we consider a method for soving constrained optimization problems, called Lagrange duality. Section 1.5 introduces the convex (conjugate) duality framework of Rockafellar [18] which can be applied to rephrase and solve a large variety of optimization problems, due to its generality. The convex duality framework is a generalized version of the Lagrange duality in Section 1.4. Some examples of optimization using convex duality is given in Section 1.6. Section 1.7 introduces conjugate functions. In Section 1.8, we introduce the Lagrange function of convex duality theory.

[1]

1.1 Basic convexity

This section summarizes some of the most important definitions and properties of convexity theory. The material of this section is mainly based on the presentation of convexity in Rockafellar [18], Hiriart-Urruty and Lemarèchal [11] and Dahl [4]. The last two consider $X = \mathbb{R}^n$, but the extension to a general inner product space is straightforward. Therefore, in the following, let X be a

[1] These notes are an adaptation of parts of Dahl [5]: Exercises and solutions have been added, some new material has been added and other things have been removed. Some material has been rewritten and new figures have been added.

real inner product space, i.e. a vector space X equipped with an inner product $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R}$ (so the function $\langle \cdot, \cdot \rangle$ is symmetric, linear in the first component and positive definite in the sense that $\langle x, x \rangle \geq 0$ for all $x \in X$, with equality if and only if $x = 0$). For instance, $X = \mathbb{R}^n$, $n \in \mathbb{N}$ is such a space.

We begin with some core definitions.

Definition 1.1.1 (*i*) *(Convex set) A set $C \subseteq X$ is called convex if $\lambda x_1 + (1-\lambda)x_2 \in C$ for all $x_1, x_2 \in C$ and $0 \leq \lambda \leq 1$.*

(*ii*) *(Convex combination) A convex combination of elements $x_1, x_2, \ldots, x_k$ in X is an element of the form $\sum_{i=1}^k \lambda_i x_i$ where $\sum_{i=1}^k \lambda_i = 1$ and $\lambda_i \geq 0$ for all $i = 1, \ldots, k$.*

(*iii*) *(Convex hull, $\mathrm{conv}(\cdot)$) Let $A \subseteq X$ be a set. The convex hull of A, denoted $\mathrm{conv}(A)$ is the set of all convex combinations of elements of A.*

(*iv*) *(Extreme points) Let $C \subseteq X$ be a convex set. An extreme point of C is a point that cannot be written as a convex combination of any other points than itself. That is: $e \in C$ is an extreme point for C if $\lambda x + (1-\lambda)y = e$ for some $x, y \in C$ implies $x = y = e$.*

(*v*) *(Hyperplane) $H \subset X$ is called a hyperplane if it is of the form $H = \{x \in X : \langle a, x \rangle = \alpha\}$ for some nonzero vector $a \in X$ and some real number α.*

(*vi*) *(Halfspace) A hyperplane H divides X into two sets $H^+ = \{x \in X : \langle a, x \rangle \geq \alpha\}$ and $H^- = \{x \in X : \langle a, x \rangle \leq \alpha\}$, these sets intersect in H. These sets are called halfspaces.*

We will now look at some hyperplane theorems in $\mathbb{R}^n$. These will be used in connection to environmental contours later in the course. Note that most of these theorems generalise to an arbitrary real inner product space X. However, the proofs are more complicated in the general case. Since the $\mathbb{R}^n$ versions are sufficient for our purposes in this course, we restrict ourselves to this.

Any hyperplane in $\mathbb{R}^n$ can be written in the form $\Pi = \{\boldsymbol{x} : \boldsymbol{c}'\boldsymbol{x} = d\}$, where $\boldsymbol{c} \in \mathbb{R}^n$ is a normal vector to Π and $d \in \mathbb{R}$. Let $\Pi^- = \{\boldsymbol{x} : \boldsymbol{c}'\boldsymbol{x} \leq d\}$ and $\Pi^+ = \{\boldsymbol{x} : \boldsymbol{c}'\boldsymbol{x} \geq d\}$ denote the two half-spaces bounded by Π. Let $\mathcal{S} \subseteq \mathbb{R}^n$. A *supporting hyperplane* of $\mathcal{S}$, is a hyperplane Π such that we either have $\mathcal{S} \subseteq \Pi^-$ or $\mathcal{S} \subseteq \Pi^+$, and such that $\Pi \cap \partial\mathcal{S} \neq \emptyset$.

If Π is a supporting hyperplane of the set $\mathcal{S}$, and $\mathcal{S} \subseteq \Pi^-$, we say that Π^+ is a *supporting half-space* of $\mathcal{S}$. We observe that if Π^+ is a supporting half-space of $\mathcal{S}$, we also have:

$$\Pi^+ \cap \mathcal{S} \subseteq \partial\mathcal{S}.$$

Moreover, we introduce the notation:

$$\mathcal{P}(\mathcal{S}) = \text{The family of supporting half-spaces of } \mathcal{S}.$$

For a given nonempty set $\mathcal{S} \subseteq \mathbb{R}^n$ and a vector $\boldsymbol{x}_0 \notin \mathcal{S}$, the vector $\boldsymbol{x}^* \in \mathcal{S}$ is said to be the *projection* of $\boldsymbol{x}_0$ onto $\mathcal{S}$ if $\boldsymbol{x}^*$ is the point in $\mathcal{S}$ which is closest to $\boldsymbol{x}_0$. In general the projection $\boldsymbol{x}^*$ may neither exist nor be unique. However, if $\mathcal{S}$ is a closed convex set, $\boldsymbol{x}^*$ is well-defined, and we have:

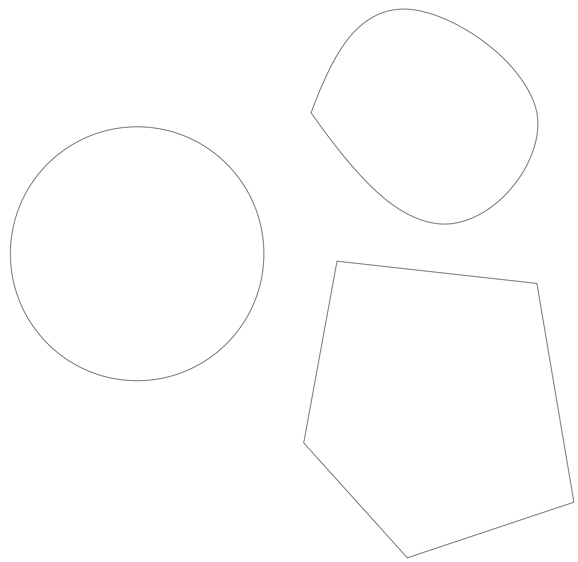

Figure 1.1: Some convex sets in the plane.

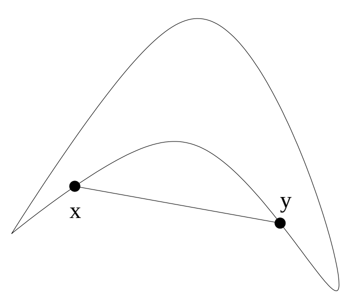

Figure 1.2: A non-convex set.

Theorem 1.1.2 (Projection) *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a closed convex set, and let $\boldsymbol{x}_0 \notin \mathcal{S}$. Then the following holds true:*

- *There exists a unique solution to the projection problem*
- *A vector $\boldsymbol{x}^* \in \mathcal{S}$ is the projection of $\boldsymbol{x}_0$ onto $\mathcal{S}$ if and only if:*

$$(\boldsymbol{x}^* - \boldsymbol{x}_0)'(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0 \textit{ for all } \boldsymbol{x} \in \mathcal{S}.$$

See Figure 1.3 for an illustration of the projection in $\mathbb{R}^2$.

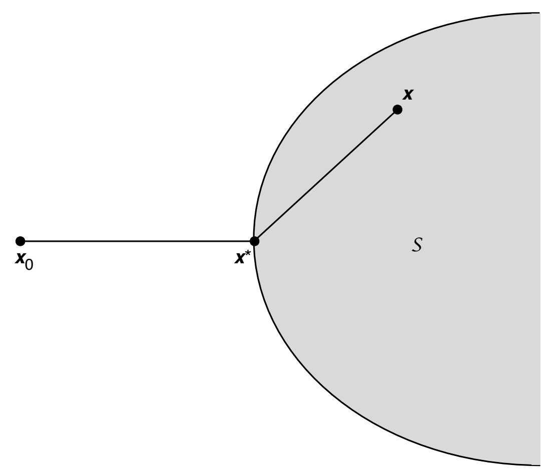

Figure 1.3: The point $\boldsymbol{x}^*$ is the projection of $\boldsymbol{x}_0$ onto the closed convex set $\mathcal{S}$.

Remark 1.1.3 *If $\boldsymbol{x} \in \mathcal{S}$, and θ is the angle between $(\boldsymbol{x}^* - \boldsymbol{x}_0)$ and $(\boldsymbol{x} - \boldsymbol{x}^*)$, then we must have $\theta \in [-\pi/2, \pi/2]$. This holds if and only if:*

$$(\boldsymbol{x}^* - \boldsymbol{x}_0)'(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0 \textit{ for all } \boldsymbol{x} \in \mathcal{S}.$$

Theorem 1.1.4 (Projection hyperplane) *Let $\mathcal{S} \in \mathbb{R}^n$ be a closed convex set, and assume that $\boldsymbol{x}_0 \notin \mathcal{S}$. Then there exists a supporting hyperplane $\Pi = \{\boldsymbol{x} : \boldsymbol{c}'\boldsymbol{x} = d\}$ of $\mathcal{S}$ such that:*

$$\boldsymbol{c}'\boldsymbol{x} \leq d \textit{ for all } \boldsymbol{x} \in \mathcal{S}, \quad \textit{and} \quad \boldsymbol{c}'\boldsymbol{x}_0 > d.$$

Proof: Since $\mathcal{S}$ is a closed convex set, it follows by the projection theorem that the projection of $\boldsymbol{x}_0$ onto $\mathcal{S}$, denoted $\boldsymbol{x}^*$ exists and satisfies:

$$(\boldsymbol{x}^* - \boldsymbol{x}_0)'(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0 \quad \text{for all } \boldsymbol{x} \in \mathcal{S}. \tag{1.1}$$

Now, we let $\boldsymbol{c} = (\boldsymbol{x}_0 - \boldsymbol{x}^*)$, and $d = \boldsymbol{c}'\boldsymbol{x}^*$. Then (1.1) can be written as:

$$\boldsymbol{c}'(\boldsymbol{x} - \boldsymbol{x}^*) \leq 0 \quad \text{for all } \boldsymbol{x} \in \mathcal{S}. \tag{1.2}$$

Hence, by (1.2) we have:

$$\boldsymbol{c}'\boldsymbol{x} \le \boldsymbol{c}'\boldsymbol{x}^* = d \quad \text{for all } \boldsymbol{x} \in \mathcal{S}.$$

Thus, $\mathcal{S} \subseteq \Pi^- = \{\boldsymbol{x} : \boldsymbol{c}'\boldsymbol{x} \le d\}$, and since $\boldsymbol{x}^* \in \mathcal{S} \cap \Pi$, Π is a supporting hyperplane of $\mathcal{S}$. Furthermore, we have:

$$\boldsymbol{c}'(\boldsymbol{x}_0 - \boldsymbol{x}^*) = (\boldsymbol{x}_0 - \boldsymbol{x}^*)'(\boldsymbol{x}_0 - \boldsymbol{x}^*) > 0.$$

Hence, it follows that:

$$\boldsymbol{c}'\boldsymbol{x}_0 > \boldsymbol{c}'\boldsymbol{x}^* = d.$$

□

Theorem 1.1.5 (Supporting hyperplane) *Let $\mathcal{S} \in \mathbb{R}^n$ be a convex set, and assume that either $\boldsymbol{x}_0 \notin \mathcal{S}$ or $\boldsymbol{x}_0 \in \partial\mathcal{S}$. Then there exists a hyperplane Π such that $\mathcal{S} \subseteq \Pi^-$ and such that $\boldsymbol{x}_0 \in \Pi$. If $\boldsymbol{x}_0 \in \partial\mathcal{S}$, Π is a supporting hyperplane of $\mathcal{S}$.*

Proof: The result follows by a similar argument as for the projection hyperplane theorem and is left as an exercise to the reader. □

Let $\mathcal{S}, \mathcal{T} \subseteq \mathbb{R}^n$. A hyperplane Π separates $\mathcal{S}$ and $\mathcal{T}$ if either $\mathcal{S} \subseteq \Pi^-$ and $\mathcal{T} \subseteq \Pi^+$ or $\mathcal{S} \subseteq \Pi^+$ and $\mathcal{T} \subseteq \Pi^-$.

Theorem 1.1.6 (Separating hyperplane) *Assume that $\mathcal{S}, \mathcal{T} \subseteq \mathbb{R}^n$ are convex, and that $\mathcal{S} \cap \mathcal{T} \subseteq \partial\mathcal{S}$. Then there exists a hyperplane Π separating $\mathcal{S}$ and $\mathcal{T}$ such that $\mathcal{S} \subseteq \Pi^-$ and $\mathcal{T} \subseteq \Pi^+$.*

Proof: We let $\boldsymbol{u}_0 = \boldsymbol{0} \in \mathbb{R}^n$ and introduce the set:

$$\mathcal{U} = \{\boldsymbol{x} - \boldsymbol{y} : \boldsymbol{x} \in \mathcal{S}_o, \boldsymbol{y} \in \mathcal{T}\},$$

where $\mathcal{S}_o = \mathcal{S} \setminus \partial\mathcal{S}$ is the (convex) set of inner points in $\mathcal{S}$.

We first argue that $\mathcal{U}$ is *convex*. To show this we must show that if $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathcal{U}$, then $\alpha\boldsymbol{u}_1 + (1-\alpha)\boldsymbol{u}_2 \in \mathcal{U}$ for all $\alpha \in [0,1]$. Since $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathcal{U}$ there exists $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{S}_o$ and $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathcal{T}$ such that:

$$\boldsymbol{u}_1 = \boldsymbol{x}_1 - \boldsymbol{y}_1 \quad \text{and} \quad \boldsymbol{u}_2 = \boldsymbol{x}_2 - \boldsymbol{y}_2.$$

Since $\mathcal{S}_o$ and $\mathcal{T}$ are convex, it follows that for any $\alpha \in [0,1]$, we have:

$$\alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2 \in \mathcal{S}_o \quad \text{and} \quad \alpha\boldsymbol{y}_1 + (1-\alpha)\boldsymbol{y}_2 \in \mathcal{T}$$

Hence, we have:

$$\begin{aligned}\alpha\boldsymbol{u}_1 + (1-\alpha)\boldsymbol{u}_2 &= \alpha(\boldsymbol{x}_1 - \boldsymbol{y}_1) + (1-\alpha)(\boldsymbol{x}_2 - \boldsymbol{y}_2) \\ &= (\alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) - (\alpha\boldsymbol{y}_1 + (1-\alpha)\boldsymbol{y}_2) \in \mathcal{U}.\end{aligned}$$

By the assumption that $\mathcal{S} \cap \mathcal{T} \subseteq \partial \mathcal{S}$ it follows that $\mathcal{S}_o$ and $\mathcal{T}$ do not have any element in common.

Hence, it follows that:

$$\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{y} \neq \mathbf{0}, \text{ for all } \boldsymbol{x} \in \mathcal{S}_o \text{ and } \boldsymbol{y} \in \mathcal{T}.$$

Thus, we conclude that:

$$\boldsymbol{u}_0 = \mathbf{0} \notin \mathcal{U}.$$

Then, by the supporting hyperplane theorem there exists a hyperplane $\Pi_0 = \{\boldsymbol{x} : \boldsymbol{c}'\boldsymbol{x} = d_0\}$ such that $\mathcal{U} \subseteq \Pi_0^-$ and such that $\boldsymbol{u}_0 \in \Pi_0$.

In fact, $\boldsymbol{u}_0 \in \Pi_0$ implies that $\boldsymbol{c}'\boldsymbol{u}_0 = \boldsymbol{c}'\mathbf{0} = d_0$. Thus, $d_0 = 0$.

Since $\mathcal{U} \subseteq \Pi_0^-$, we have $\boldsymbol{c}'\boldsymbol{u} \leq d_0 = 0$ for all $\boldsymbol{u} \in \mathcal{U}$, implying that:

$$\boldsymbol{c}'(\boldsymbol{x} - \boldsymbol{y}) \leq 0 \text{ for all } \boldsymbol{x} \in \mathcal{S}_o \text{ and } \boldsymbol{y} \in \mathcal{T},$$

or equivalently:

$$\boldsymbol{c}'\boldsymbol{x} \leq \boldsymbol{c}'\boldsymbol{y} \text{ for all } \boldsymbol{x} \in \mathcal{S}_0 \text{ and } \boldsymbol{y} \in \mathcal{T}. \tag{1.3}$$

We then let $d = \sup_{\boldsymbol{x} \in \mathcal{S}_o} \boldsymbol{c}'\boldsymbol{x}$. By the definition of d we have:

$$\boldsymbol{c}'\boldsymbol{x} \leq d, \text{ for all } \boldsymbol{x} \in \mathcal{S}_o. \tag{1.4}$$

If $\boldsymbol{x}_0 \in \partial \mathcal{S}$, there exists $\{\boldsymbol{x}_k\} \subseteq \mathcal{S}_o$ such that $\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}_0$, and so:

$$\boldsymbol{c}'\boldsymbol{x}_0 = \lim_{k\to\infty} \boldsymbol{c}'\boldsymbol{x}_k \leq d. \tag{1.5}$$

By combining (1.4) and (1.5) it follows that we have:

$$\boldsymbol{c}'\boldsymbol{x} \leq d, \text{ for all } \boldsymbol{x} \in \mathcal{S}. \tag{1.6}$$

Moreover, by (1.3) and the definition of d it follows that we have:

$$\boldsymbol{c}'\boldsymbol{y} \geq d, \text{ for all } \boldsymbol{y} \in \mathcal{T}. \tag{1.7}$$

Hence, by letting $\Pi = \{\boldsymbol{x} : \boldsymbol{c}'\boldsymbol{x} = d\}$, we conclude that:

$$\mathcal{S} \subseteq \Pi^- \text{ and } \mathcal{T} \subseteq \Pi^+ \tag{1.8}$$

□

In the proof of the separating hyperplane theorem, we defined $\Pi = \{\boldsymbol{x} : \boldsymbol{c}'\boldsymbol{x} = d\}$, where:

$$d = \sup_{\boldsymbol{x} \in \mathcal{S}_o} \boldsymbol{c}'\boldsymbol{x}.$$

This implies that there exists a sequence $\{\boldsymbol{x}_k\} \subseteq \mathcal{S}$ with limit $\boldsymbol{x}_0 = \lim_{k\to\infty} \boldsymbol{x}_k \in \partial \mathcal{S}$, such that $\boldsymbol{c}'\boldsymbol{x}_0 = d$.

Thus, $\boldsymbol{x}_0 \in \Pi \cap \partial \mathcal{S}$, implying that Π^+ is a *supporting halfspace* of $\mathcal{S}$.

Theorem 1.1.7 (Separating hyperplane, supporting halfspace) *Assume that $\mathcal{S}, \mathcal{T} \subseteq \mathbb{R}^n$ are convex, and that $\mathcal{S} \cap \mathcal{T} \subseteq \partial\mathcal{S}$. Then there exists a hyperplane Π separating $\mathcal{S}$ and $\mathcal{T}$ such that $\mathcal{S} \subseteq \Pi^-$, $\mathcal{T} \subseteq \Pi^+$ and where $\Pi^+ \in \mathcal{P}(\mathcal{S})$.*

We now define polyhedrons and polytopes. These subsets of $\mathbb{R}^n$ are useful because they are descriptions of the solution set of systems of linear inequalities. The following definitions are from Rockafellar [17].

Definition 1.1.8 (Polyhedron) *A set $K \subseteq \mathbb{R}^n$ is called a polyhedron if it can be described as the intersection of finitely many closed half-spaces.*

Hence, a polyhedron can be described as the solution set of a system of finitely many (non-strict) linear inequalities. It is straightforward to show that a polyhedron is a convex set.

A *(convex) polytope* is a set of the following form:

Definition 1.1.9 (Polytope) *A set $K \subseteq \mathbb{R}^n$ is called a (convex) polytope if it is the convex hull of finitely many points.*

Clearly, all polytopes are convex since a convex hull is always convex. Examples of (convex) polytopes in $\mathbb{R}^2$ are triangles, squares and hexagons.

Actually, all polytopes in $\mathbb{R}^n$ are compact sets.

Lemma 1.1.10 *Let $K \subseteq \mathbb{R}^n$ be a polytope. Then K is a compact set.*

Proof: Since K is a polytope, it is the convex hull of finitely many points, say $K = \mathrm{conv}(\{k_1\}, ..., \{k_m\})$, so

$$K = \{\sum_{i=1}^{m} \lambda_i k_i : \sum_{i=1}^{m} \lambda_i = 1, \lambda_i \geq 0 \text{ for all } i = 1, ..., m\}.$$

Consider the continuous function $f : \mathbb{R}^m \to \mathbb{R}^n$, $f(x_1, ..., x_m) = \sum_{i=1}^m x_i k_i$, and the compact set

$$S = \{(\lambda_1, ..., \lambda_m) : \sum_{i=1}^{m} \lambda_i = 1, \lambda_i \geq 0 \text{ for all } i = 1, ..., m\} \subseteq \mathbb{R}^m$$

(S is closed and bounded, hence compact in $\mathbb{R}^m$)

Then, since f is continuous and S is compact, $f(S) := \{x : x = f(s) \text{ for some } s \in S\} \subseteq \mathbb{R}^n$ is a compact set (see for example Munkres [15]). But $f(S) = K$ from the definitions, and hence K is compact. $\square$

From Lemma 1.1.10, any polytope is a closed and bounded set, since compactness is equivalent to being closed and bounded in $\mathbb{R}^n$.

The following theorem connects the notion of polytope and polyhedron.

Theorem 1.1.11 *A set $K \subseteq \mathbb{R}^n$ is a polytope if and only if it is a bounded polyhedron.*

For a proof of this, see Ziegler [25].

Sometimes, one needs to consider what is called the relative interior of a set.

Definition 1.1.12 (Relative interior, $\mathrm{rint}(\cdot)$) *Let $S \subseteq X$. $x \in S$ is a relative interior point of S if it is contained in some open set whose intersection with $\mathrm{aff}(S)$ is contained in S. $\mathrm{rint}(S)$ is the set of all relative interior points of S. Here, $\mathrm{aff}(S)$ is the smallest affine set that contains S (where a set is affine if it contains any affine combination of its points; an affine combination is like a convex combination except the coefficients are allowed to be negative).*

Another useful notion is that of a convex cone.

Definition 1.1.13 (Convex cone) *$C \subseteq X$ is called a convex cone if for all $x, y \in C$ and all $\alpha, \beta \geq 0$:*

$$\alpha x + \beta y \in C$$

From these definitions, one can derive some properties of convex sets.

Theorem 1.1.14 (Properties of convex sets) *(i) If $\{C_j\}_{j \in J} \subseteq X$ is an arbitrary family of convex sets, then the intersection $\cap_{j \in J} C_j$ is also a convex set.*

(ii) $\mathrm{conv}(A)$ is a convex set, and it is the smallest (set inclusion-wise) convex set containing A.

(iii) If $C_1, C_2, \ldots, C_m \subseteq X$ are convex sets, then the Cartesian product $C_1 \times C_2 \times \ldots \times C_m$ is also a convex set.

(iv) If $C \subseteq X$ is a convex set, then the interior of C, $\mathrm{int}(C)$, the relative interior $\mathrm{rint}(C)$ and the closure of C, $\mathrm{cl}(C)$, are convex sets as well.

The proof is left as an exercise.

Sometimes, one considers not just $\mathbb{R}$, but $\bar{\mathbb{R}}$, the extended real numbers.

Definition 1.1.15 (The extended real numbers, $\bar{\mathbb{R}}$) *Let $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ denote the extended real numbers.*

When working with the extended real numbers the following computational rules apply: $a - \infty = -\infty$, $a + \infty = \infty$, $\infty + \infty = \infty$, $-\infty - \infty = -\infty$ and $\infty - \infty$ is not defined.

The following function is often useful, in particular in optimization.

Definition 1.1.16 (The indicator function for a set M, δ_M) *Let $M \subseteq X$ be a set. The indicator function for the set M, $\delta_M : X \to \bar{\mathbb{R}}$ is defined as*

$$\delta_M(x) = \begin{cases} 0 & \text{if } x \in M \\ +\infty & \text{if } x \notin M. \end{cases}$$

The following example shows why this function is useful in optimization. Consider the constrained minimization problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in M \end{array}$$

for some function $f : X \to \bar{\mathbb{R}}$ and some set $M \subseteq X$. This can be transformed into an unconstrained minimization problem by altering the objective function as follows

$$\min f(x) + \delta_M(x).$$

This is the same problem as before because the minimum above cannot be achieved for $x \notin M$, because then $\delta_M = +\infty$, so the objective function is infinitely large as well.

The next definition is very important.

Definition 1.1.17 (Convex function) *Let $C \subseteq X$ be a convex set. A function $f : C \to \mathbb{R}$ is called convex if the inequality*

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) \tag{1.9}$$

holds for all $x, y \in C$ and every $0 \le \lambda \le 1$.

There is an alternative way of defining convex functions, which is based on the notion of epigraph.

Definition 1.1.18 (Epigraph, $\operatorname{epi}(\cdot)$**)** *Let $f : X \to \bar{\mathbb{R}}$ be a function. Then the epigraph of f is defined as $\operatorname{epi}(f) = \{(x, \alpha) : x \in X,\ \alpha \in \mathbb{R},\ \alpha \ge f(x)\}$.*

Definition 1.1.19 (Convex function) *Let $A \subseteq X$. A function $f : A \to \bar{\mathbb{R}}$ is called convex if the epigraph of f is convex (as a subset of the vector space $X \times \mathbb{R}$).*

Of course, these definitions are actually equivalent.

Theorem 1.1.20 *Definitions 1.1.17 and 1.1.19 are equivalent if the set A in Definition 1.1.19 is convex (A must be convex in order for Definition 1.1.17 to make sense).*

Proof:

1.1.17 $\Rightarrow$ 1.1.19: Assume that f is a convex function according to Definition 1.1.17. Let $(x, a), (y, b) \in \operatorname{epi}(f)$ and let $\lambda \in [0, 1]$. Then

$$\lambda(x, a) + (1-\lambda)(y, b) = (\lambda x + (1-\lambda)y, \lambda a + (1-\lambda)b).$$

But $f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y)$ from Definition 1.1.17, so

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

So $(\lambda x + (1-\lambda)y, \lambda a + (1-\lambda)b) \in \operatorname{epi}(f)$.

1.1.19 $\Rightarrow$ 1.1.17 uses the same type of arguments, thus it is omitted. $\square$

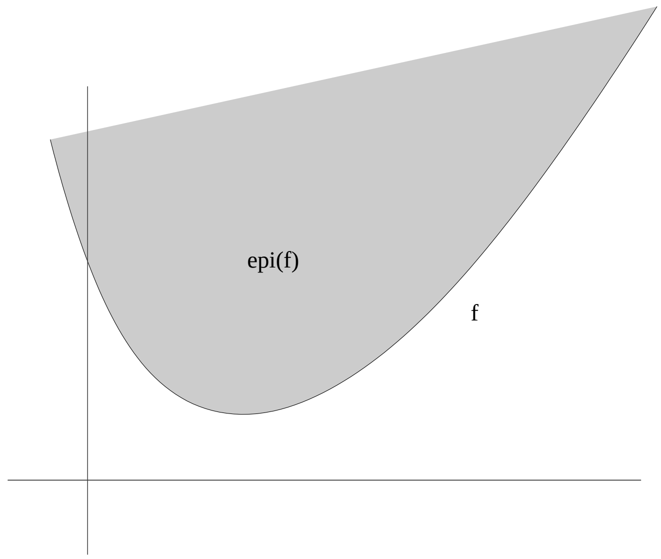

Figure 1.4: The epigraph of a function f.

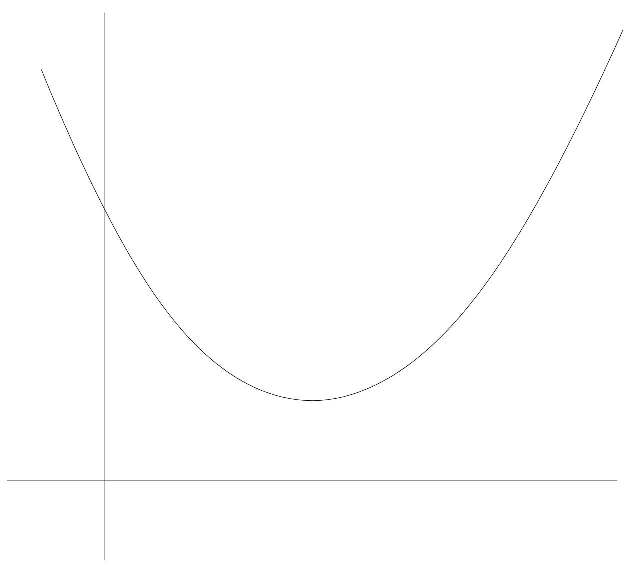

Figure 1.5: A convex function.

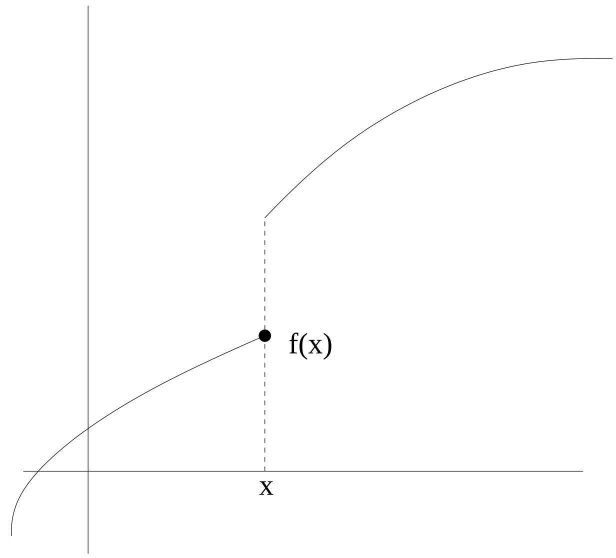

Figure 1.6: A lower semi-continuous function f.

Definition 1.1.21 (Concave function) *A function g is concave if the function $f := -g$ is convex.*

When minimizing a function, the points where it is infinitely large are uninteresting, this motivates the following definitions.

Definition 1.1.22 (Effective domain, $\mathrm{dom}(\cdot)$**)** *Let $A \subseteq X$ and let $f : A \to \bar{\mathbb{R}}$ be a function. The effective domain of f is defined as* $\mathrm{dom}(f) = \{x \in A : f(x) < +\infty\}$.

Definition 1.1.23 (Proper function) *Let $A \subseteq X$ and let $f : A \to \bar{\mathbb{R}}$ be a function. f is called proper if* $\mathrm{dom}(f) \neq \emptyset$ *and $f(x) > -\infty$ for all $x \in A$.*

For definitions of general topological terms, such as convergence, continuity and neighborhood, see any basic topology book, for instance *Topology* by James Munkres [15].

Definition 1.1.24 (Lower semi-continuity, lsc**)** *Let $A \subseteq X$ be a set, and let $f : A \to \bar{\mathbb{R}}$ be a function. f is called lower semi-continuous,* lsc, *at a point $x_0 \in A$ if for each $k \in \mathbb{R}$ such that $k < f(x_0)$ there exists a neighborhood U of x_0 such that $f(u) > k$ for all $u \in U$. Equivalently: f is lower semi-continuous at x_0 if and only if* $\liminf_{x \to x_0} f(x) \geq f(x_0)$.

Definition 1.1.25 (α-sublevel set of a function, $S_\alpha(f)$) *Let $f : X \to \bar{\mathbb{R}}$ be a function and let $\alpha \in \mathbb{R}$. The α-sublevel set of f, $S_\alpha(f)$, is defined as*

$$S_\alpha(f) = \{x \in X : f(x) \leq \alpha\}.$$

Theorem 1.1.26 *Let $f : X \to \bar{\mathbb{R}}$ be a function. Then, f is lower semi-continuous if and only if the sublevel sets $S_\alpha(f)$ are closed for all $\alpha \in \bar{\mathbb{R}}$.*

Proof: The sublevel sets $S_\alpha(f) := \{x \in X : f(x) \leq \alpha\}$ are closed for all $\alpha \in \mathbb{R}$ iff. the complement sets $Y = X - S_\alpha(f) = \{x \in X : f(x) > \alpha\}$ are open for all α. But this happens iff. all $y \in Y$ are interior points, which is equivalent with that for each $y \in Y$ there is a neighborhood U such that $U \subseteq Y$, i.e. $f(U) > \alpha$. But this is the definition of f being lower semi-continuous at the point y. Since this argument holds for all $y \in X$ (by choosing different α), f is lower semi-continuous. $\square$

Definition 1.1.27 (Convex hull of a function, $\mathrm{co}(f)$**)** *Let $A \subseteq X$ be a set, and let $f : A \to \bar{\mathbb{R}}$ be a function. Then the convex hull of f is the (pointwise) largest convex function h such that $h(x) \leq f(x)$ for all $x \in A$.*

Clearly, if f is a convex function $\mathrm{co}(f) = f$. One can define *the lower semi-continuous hull*, $\mathrm{lsc}(f)$ of a function f in a similar way.

Definition 1.1.28 (Closure of a function, $\mathrm{cl}f$**)** *Let $A \subseteq X$ be a set, and let $f : A \to \bar{\mathbb{R}}$ be a function. We define: $\mathrm{cl}(f)(x) = \mathrm{lsc}(f(x))$ for all $x \in A$ if $\mathrm{lsc}(f(x)) > -\infty \ \forall \ x \in X$ and $\mathrm{cl}(f)(x) = -\infty$ for all $x \in A$ if $\mathrm{lsc}(f(x')) = -\infty$ for some $x' \in A$.*

We say that a function f is *closed* if $\mathrm{cl}(f) = f$. Hence, f is closed if it is lower semi-continuous and $f(x) > -\infty$ for all x or if $f(x) = -\infty$ for all x.

Theorem 1.1.29 *Let $M \subseteq X$, and consider the indicator function for the set M, δ_M, as defined in Definition 1.1.16. Then, the following properties hold:*

- *If $N \subseteq X$, then $M \subseteq N \iff \delta_N \leq \delta_M$.*
- *M is a convex set $\iff$ δ_M is a convex function.*
- *δ_M is lower semi-continuous $\iff$ M is a closed set.*

Proof:

- From Definition 1.1.16: $\delta_N \leq \delta_M$ iff. (If $\delta_M(x) < +\infty$ then $\delta_N(x) < +\infty$) iff. ($x \in M \Rightarrow x \in N$) iff. $M \subseteq N$.
- δ_M is convex if and only if $\delta_M(\lambda x + (1-\lambda)y) \leq \lambda\delta_M(x) + (1-\lambda)\delta_M(y)$ holds for all $0 \leq \lambda \leq 1$ and all $x, y \in X$ such that $\delta_M(x), \delta_M(y) < +\infty$, that is, for all $x, y \in M$. But this means that $\lambda x + (1-\lambda)y \in M$, equivalently, M is convex.
- Assume δ_M is lower semi-continuous. Then it follows from Theorem 1.1.26 that $S_\alpha(\delta_M)$ is closed for all $\alpha \in \mathbb{R}$. But, for any $\alpha \in \mathbb{R}$, $S_\alpha(\delta_M) = \{x \in X : \delta_M(x) \leq \alpha\} = M$ (from the definition of δ_M), so M is closed. Conversely, assume that M is closed. Then, for any $\alpha \in \mathbb{R}$, $S_\alpha(\delta_M) = M$, hence δ_M is lower semi-continuous from Theorem 1.1.26.

□

A *global minimum* for a function $f : A \to \bar{\mathbb{R}}$, where $A \subset X$, is an $x' \in A$ such that $f(x') \leq f(x)$ for all $x \in A$. A *local minimum* for f is an $x' \in A$ such that there exists a neighborhood U of x' such that $x \in U \Rightarrow f(x') \leq f(x)$.

Based on all these definitions, one can derive the following properties of convex functions.

Theorem 1.1.30 (Properties of convex functions) *Let $C \subseteq X$ be a convex set, $f : C \to \mathbb{R}$ be a convex function. Then the following properties hold:*

1. *If f has a local minimum x', then x' is also a global minimum for f.*
2. *If $C = \mathbb{R}$, so that $f : \mathbb{R} \to \mathbb{R}$ and f is differentiable, then f' is monotonically increasing.*
3. *If a function $g : \mathbb{R} \to \mathbb{R}$ is twice differentiable and $g''(x) > 0$, then g is convex.*
4. Jensen's inequality*: For $x_1, \ldots, x_n \in X, \lambda_1, \ldots, \lambda_n \in \mathbb{R}, \lambda_k \geq 0$, for $k = 1, \ldots, n$, $\sum_{k=1}^n \lambda_k = 1$, the following inequality holds*
$$f(\sum_{k=1}^n \lambda_k x_k) \leq \sum_{k=1}^n \lambda_k f(x_k).$$
5. *The sum of convex functions is convex.*
6. *αf is convex if $\alpha \in \mathbb{R}, \alpha \geq 0$.*
7. *If $(f_n)_{n \in \mathbb{N}}$ is a sequence of convex functions, $f_n : X \to \mathbb{R}$, and $f_n \to f$ pointwise as $n \to \infty$, then f is convex.*
8. $\mathrm{dom}(f)$ *is a convex set*
9. *If $\alpha \in \bar{\mathbb{R}}$, then the sublevel set for f, $S_\alpha(f)$ is a convex set. Similarly, $\{x \in C : f(x) < \alpha\}$ is a convex set.*
10. Maximization*: Let $\{f_\lambda\}$ be an arbitrary family of convex functions, then $g(x) = \sup_\lambda f_\lambda(x)$ is convex. Also, $g(x) = \sup_y f(x, y)$ is convex if $f(x, y)$ is convex in x for all y.*
11. Minimization*: Let $f : X \times X \to \bar{\mathbb{R}}$ be a convex function. Then $g(x) = \inf_y f(x, y)$ is convex.*

Proof:

1. Suppose x' is a local minimum for f, that is: There exists a neighborhood $U \subseteq C$ of x' such that $f(x') \leq f(x)$ for all $x \in U$. We want to show that $f(x') \leq f(x)$ for all $x \in C$. Let $x \in C$. Consider the convex combination $\lambda x + (1 - \lambda)x'$. This convex combination converges towards x' as $\lambda \to 0$.

Therefore, for a sufficiently small λ^*, $\lambda^* x + (1-\lambda^*)x' \in U$, so since f is convex

$$\begin{aligned} f(x') &\leq f(\lambda^* x + (1-\lambda^*)x') \\ &\leq \lambda^* f(x) + (1-\lambda^*) f(x') \end{aligned}$$

which, by rearranging the terms, shows that $f(x') \leq f(x)$. Therefore, x' is a global minimum as well.

2. Follows from Definition 1.1.17 and the definition of the derivative.

3. Left as an exercise.

4. Left as an exercise.

5. Left as an exercise.

6. Follows from Definition 1.1.17.

7. Use Definition 1.1.17 and the homogeneity and additivity of limits.

8. Follows from the definitions.

9. Follows from the definitions, but is included here as an example of a typical basic proof. Let $x, y \in S_\alpha(f)$. Then $f(x), f(y) \leq \alpha$. Then $\lambda x + (1-\lambda)y \in S_\alpha(f)$ because

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \leq \lambda\alpha + (1-\lambda)\alpha = \alpha$$

where the first inequality follows from the convexity of f, and the second inequality follows from that $x, y \in S_\alpha(f)$.

10. sup is a limit, so the result is a consequence of property 7.

11. Same as property 10.

□

1.2 Quasiconvex functions

In this section we introduce quasiconvex functions. Quasi-convexity is a weaker requirement than convexity, but still strong enough that quasiconvex functions have applications in optimization, game theory and economics.

Definition 1.2.1 (Quasiconvex function) *Let $S \subseteq X$ be convex. A function $f : S \to \mathbb{R}$ is quasiconvex if for all $x, y \in S$ and $\lambda \in [0,1]$, we have*

$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}.$$

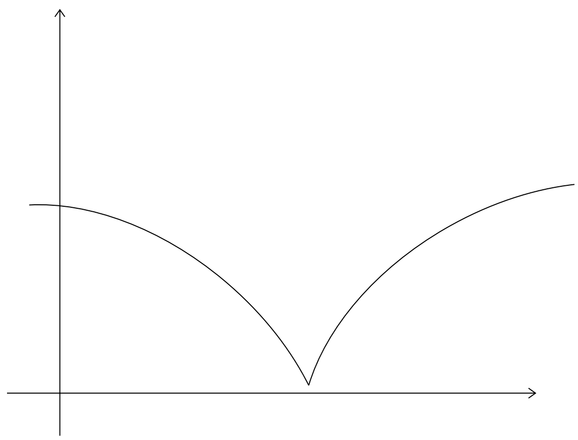

Figure 1.7: A function which is quasiconvex, but not convex.

An equivalent way to define quasiconvexity is via convexity of the sublevel sets $S_\alpha := \{x \in S : f(x) \leq \alpha\}$ for all α.

Proposition 1.2.2 *Let $S \subseteq X$ be convex and let $f : S \to \mathbb{R}$. Then, f is quasiconvex if and only if the α-sublevel sets*

$$S_\alpha = \{x \in S : f(x) \leq \alpha\}$$

are convex for all $\alpha \in \mathbb{R}$.

Proof: Left as an exercise to the reader. □

All convex functions are quasiconvex. However, the opposite implication is not true: There exists quasiconvex functions that are not convex, see Figure 1.7. Furthermore, concave functions can be quasiconvex. An example of this is $f(x) = log(x)$, defined on the positive real numbers.

Not all functions are quasiconvex, and an example of a function which is not quasiconvex is illustrated in Figure 1.7. Note that this fuction is not quasiconvex because the set of points in the domain where the function values are below the horizontal red line is the union of the two bold, red intervals, which is not a convex set. Hence, the sublevel set S_α for this particular α is not convex, and therefore the function does not satisfy the condition in Proposition 1.2.2

1.3 Optimization

Optimization is the mathematical theory of maximization and minimization problems. It is useful in many applications, for example in logistic problems, finding the best spot to drill for oil, and in mathematical finance. In finance,

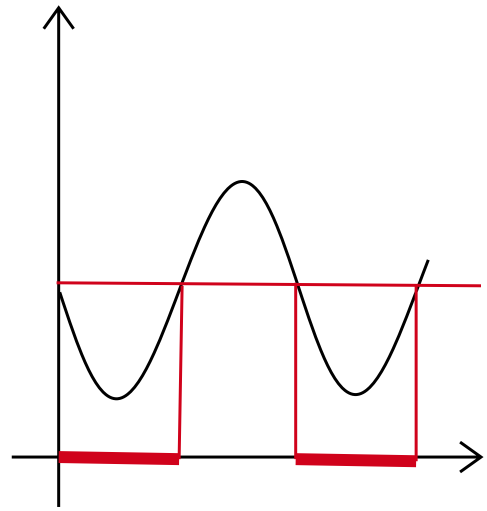

Figure 1.8: A function which is not quasiconvex.

one often considers an investor who wishes to maximize her utility, given various constraints (for instance her salary). The question is how one can solve such problems. This section gives a short summary of some of some background theory on optimization.

Let X be a vector space, $f : X \to \bar{\mathbb{R}}$, $g : X \to \mathbb{R}^n$ and $S \subseteq X$. Consider an optimization problem of the form

$$\begin{array}{ll} \min & f(x) \\ \text{subject to} & \\ & g(x) \leq 0 \text{ (componentwise)} \\ & x \in S. \end{array} \tag{1.10}$$

In problem (1.10), f is called the *objective function*, while $g(x) \leq 0$, $x \in S$ are called the *constraints* of the problem.

A useful technique when dealing with optimization problems is transforming the problem. For example, a constraint of the form $h(x) \geq y$ (for $h : X \to \mathbb{R}^n$, $y \in \mathbb{R}^n$) is equivalent to $y - h(x) \leq 0$, which is of the form $g(x) \leq 0$ with $g(x) = y - h(x)$. Similarly, any maximization problem can be turned into a minimization problem (and visa versa) by using that $\inf f(x) = -\sup(-f(x))$. Also, any equality constraint can be transformed into two inequality constraints: $h(x) = 0$ is equivalent to $h(x) \leq 0$ *and* $h(x) \geq 0$.

One of the most important theorems of optimization is the extreme value theorem (see Munkres [15]).

Theorem 1.3.1 (The extreme value theorem) *If $f : X \to \mathbb{R}$ is a continuous function from a compact set into the real numbers, then there exist points*

$a, b \in X$ such that $f(a) \geq f(x) \geq f(b)$ for all $x \in X$. That is, f attains *a maximum and a minimum.*

The importance of the extreme value theorem is that it gives the existence of a maximum and a minimum in a fairly general situation. However, these may not be unique. But, for convex (or concave) functions, Theorem 1.1.30 implies that any local minimum (maximum) is a global minimum (maximum). This makes convex functions useful in optimization.

For a function $f : \mathbb{R}^n \to \mathbb{R}$, the maximum and minimum are attained in *critical points*. Critical points are points x such that

- $f'(x) = 0$, where f is differentiable at x,
- the function f is not differentiable at x or
- x is on the boundary of the set one is optimizing over.

Hence, for a differentiable function which is optimized without extra constraints, one can find maximum and minimum points by solving $f'(x) = 0$ and comparing the objective value in these points to those of the points on the boundary.

Constrained optimization can be tricky to handle. An example of constrained optimization is linear programming (LP); maximization of linear functions under linear constraints. In this situation, strong theorems regarding the solution has been derived. It turns out that corresponding to each LP problem, there is a "dual" problem, and these two problems have the same optimal value. This dual problem is introduced in order to get a second chance at solving an otherwise difficult problem. There is also an effective numerical method for solving LP problems, called the *simplex algorithm*. See Vanderbei [24] for more about linear programming.

The concept of deriving a "dual" problem to handle constraints is the idea of Lagrange duality as well. Lagrange duality begins with a problem of the form (1.10) (or the corresponding maximization problem), and derives a dual problem which gives lower (upper) bounds on the optimal value of the problem. Actually, linear programming duality is a special case of Lagrange duality, but since Lagrange duality is more general, one cannot get the strong theorems of linear programming. The duality concept is generalized even more in convex duality theory, which is the topic of Section 1.5.

1.4 Lagrange duality

The method of Lagrange duality can be described as follows: Let X be a general inner product space with inner product $\langle \cdot, \cdot \rangle$. Assume there is a function $f : X \to \mathbb{R}$ to be maximized under certain constraints.

Consider a problem of the following, very general, form

$$\text{maximize } f(x) \text{ subject to } g(x) \leq 0, \; x \in S \tag{1.11}$$

where g is a function such that $g : X \to \mathbb{R}^N$ and $S \neq \emptyset$ (to exclude a trivial case). This will be called *the primal problem.* Note that if one has a problem with equality constraints, one can rewrite this in the form of problem (1.11) by writing each equality as two inequalities. Also, $\geq$ can be turned into $\leq$ by multiplying with -1, and by basic algebra, one can always make sure there is 0 on one side of the inequality. Note that there are no constraints on f or S and only one (weak) constraint on g. Hence, many problems can be written in the form (1.11).

Let $\lambda \in \mathbb{R}^N$ be such that $\lambda \geq 0$ (componentwise), and assume that $g(x) \leq 0$ (componentwise) for all $x \in S$. Then:

$$f(x) \leq f(x) - \lambda \cdot g(x) \tag{1.12}$$

because $\lambda \cdot g(x) \leq 0$ (where $\cdot$ denotes the Euclidean inner product). This motivates the definition of the *Lagrange function*, $L(x, \lambda)$

$$L(x, \lambda) = f(x) - \lambda \cdot g(x).$$

Hence, $L(x, \lambda)$ is an upper bound on the objective function for each $\lambda \in \mathbb{R}^N$, $\lambda \geq 0$ and $x \in X$ such that $g(x) \leq 0$. By taking supremum on each side of the inequality in (1.12), for each $\lambda \geq 0$,

$$\begin{aligned} \sup\{f(x) : g(x) \leq 0,\ x \in S\} &\leq \sup\{f(x) - \lambda \cdot g(x) : g(x) \leq 0, x \in S\} \\ &= \sup\{L(x, \lambda) : x \in S, g(x) \leq 0\} \\ &\leq \sup_{x \in S} L(x, \lambda) \\ &:= L(\lambda) \end{aligned} \tag{1.13}$$

where the second inequality follows because we are maximizing over a larger set, hence the optimal value cannot decrease.

This implies that for all $\lambda \geq 0$, $L(\lambda)$ is an upper bound for the optimal value function. We want to find the smallest upper bound. This motivates the definition of the *Lagrangian dual problem*

$$\inf_{\lambda \geq 0} L(\lambda). \tag{1.14}$$

Therefore, the following theorem is proven (by taking the infimum on the right hand side of equation (1.13)).

Theorem 1.4.1 (Weak Lagrange duality) *In the setting above, the following inequality holds*

$$\sup\{f(x) : g(x) \leq 0,\ x \in S\} \leq \inf\{L(\lambda) : \lambda \geq 0\}.$$

This theorem shows that the Lagrangian dual problem gives the smallest upper bound on the optimal value of problem (1.11) generated by the Lagrange

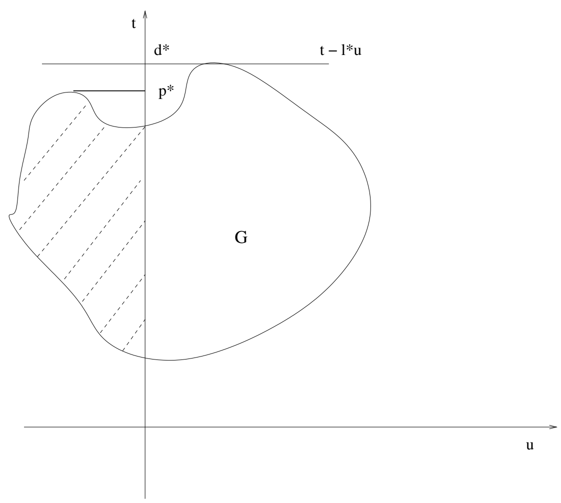

Figure 1.9: Illustration of Lagrange duality with duality gap.

function. The Lagrangian dual problem has only one, quite simple, constraint, namely $\lambda \geq 0$, and this may mean that the dual problem is easier to solve than the original problem.

In some special cases, one can proceed to show duality theorems, proving that $\sup\{f(x) : g(x) \leq 0,\ x \in S\} = \inf_{\lambda \geq 0} L(\lambda)$. If this is the case, one says that there is no *duality gap*. This typically happens in convex optimization problems under certain assumptions. However, often there actually is a duality gap, but the Lagrangian dual problem still gives us an upper bound, and hence some idea of the optimal value of our problem.

An example where Lagrangian duality is applied, and a duality theorem is derived, is linear programming (LP) duality. The calculation of the LP dual from the primal is omitted here, but it is fairly straight-forward.

The definition of conjugate functions (see Section 1.7) also has its roots in Lagrangian duality. The conjugate function shows up naturally when finding the Lagrangean dual of a minimization problem with linear inequalities as constrains.

One can illustrate Lagrange duality such that it is simple to see graphically whether there is a duality gap. Consider problem (1.11) where $S = X$, and define the set $\mathcal{G} = \{(g(x), f(x)) \in \mathbb{R}^{N+1} : x \in X\}$. The optimal value of problem (1.11), denoted p^*, can then be written as $p^* = \sup\{t : (u,t) \in \mathcal{G}, u \leq 0\}$ (from the definitions). This can be illustrated for $g : X \to \mathbb{R}$ (i.e. for only one inequality) as in Figures 1.9 and 1.10.

Figure 1.10 shows the set $\mathcal{G}$, the optimal primal value p^* and the Lagrange-function for two different Lagrange multipliers. The value of the function $L(l) = \sup_{x \in X}\{f(x) - lg(x)\}$ is given by the intersection of the line $t - lu$ and the t-axis.

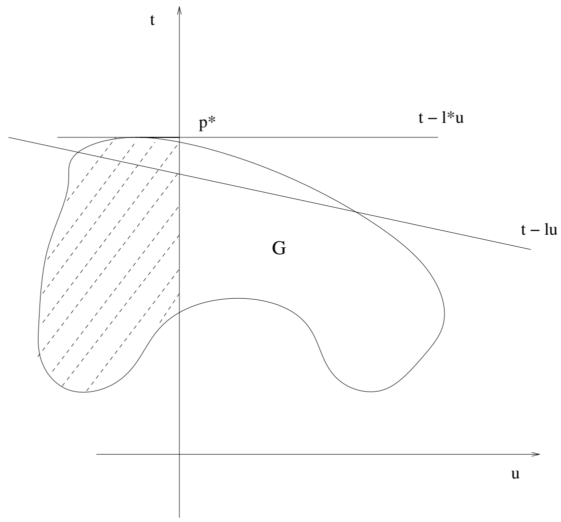

Figure 1.10: Illustration of Lagrange duality with no duality gap.

Note that the shaded part of $\mathcal{G}$ corresponds to the feasible solutions of problem (1.11). Hence, to find the optimal primal solution p^* in the figure, find the point (u^*, t^*) in the shaded area of $\mathcal{G}$ such that t^* is as large as possible.

How can one find the optimal dual solution in Figure 1.10? Fix an $l \geq 0$, and draw the line $t - lu$ into the figure. Now, find the function $L(l)$ by parallel-adjusting the line so that the intersection of $t - lu$ is as large as possible, while making sure that the line still intersects $\mathcal{G}$. Having done this, tilt the line such that l is still greater than or equal 0, but such that the intersection of the line and the t-axis becomes as big as possible. The final intersection is the optimal dual solution.

Actually, there is no duality gap in the problem of Figure 1.10, since the optimal primal value corresponds to the optimal dual value, given by the intersection of the line $t - l^*u$ and the t-axis.

In Figure 1.9 there is a duality gap, since the optimal dual value, denoted d^* is greater than the optimal primal value, denoted p^*. What goes wrong? By examining the two figures above, one sees that the absence of a duality gap has something to do with the set $\mathcal{G}$ being "locally convex" near the t-axis. Bertsekas [2] formalizes this idea, and shows a condition for the absence of a duality gap (in the Lagrange duality case), called the Slater condition.

The Slater condition, in the case where $X = \mathbb{R}^n$ (see Boyd and Vandenberghe [3]), states the following: Assume there is a problem of the form (1.11). If f is concave, $S = X$, each component function of g is convex and there exists $x \in \text{rint}(D)$ (see Definition 1.1.12), where D is defined as the set of $x \in X$ where both f and g are defined, such that $g(x) < 0$, then there is no duality gap.

Actually, (from Boyd and Vandenberghe [3]) this condition can be weakened in the case where the component-functions g are actually affine (and f is still concave) and $\mathrm{dom}(f)$ is open. In this case it is sufficient that there exists a feasible solution for the absence of a duality gap. Note that for a minimization problem, the same condition holds as long as f is convex (since a maximization problem can be turned into a minimization problem by using that $\sup f = -\inf(-f)$).

There is also an alternative version of the Slater condition, where $X = \mathbb{R}^n$. This is from Bertsekas et. al [2, p.371]: If the optimal value of the primal problem (1.11) is finite, S is a convex set, f and g are convex functions and there exists $x' \in S$ such that $g(x) < 0$, then there is no duality gap.

There is also a generalized version of the Lagrange duality method. The previous Lagrange duality argument can be done for $g : X \to Z$, where Z is some normed space (see Rynne and Youngston [22] for more on normed spaces) with an ordering that defines a non-negative orthant. From this, one can derive a slightly more general version of the Slater condition (using the separating hyperplane theorem). This version of the Slater condition is Theorem 5 in Luenberger [13] (adapted to the notation of this section): Let X be a normed space and let f be a concave function, defined on a convex subset C of X. Also, let g be a convex function which maps into a normed space Z (with some ordering). Assume there exists some $x' \in C$ such that $g(x') < 0$. Then the optimal value of the Lagrange primal problem equals the optimal value of the Lagrange dual problem, i.e. there is no duality gap.

In particular, since $\mathbb{R}^m$ is a normed space with an ordering that defines a non-negative orthant (componentwise ordering), this generalized Slater condition applies to the Lagrange problem at the beginning of this section.

Finally, note that the Lagrange duality method is quite general, since it holds for an arbitrary vector space X.

1.5 Convex duality and optimization

This section is based on *Conjugate Duality and Optimization* by Rockafellar [18]. As mentioned, convex functions are very handy in optimization problems because of property 1 of Theorem 1.1.30: For any convex function, a local minimum is also a global minimum.

Another advantage with convex functions in optimization is that one can exploit *duality properties* in order to solve problems. In the following, let X be a linear space, and let $f : X \to \mathbb{R}$ be a function. The main idea of convex duality is to view a given minimization problem $\min_{x \in X} f(x)$ (note that it is common to write min instead of inf when introducing a minimization problem even though one does not know that the minimum is attained) as one half of a minimax problem where a saddle value exists. Very roughly, one does this by looking at an abstract optimization problem

$$\min_{x \in X} F(x, u) \tag{1.15}$$

where $F : X \times U \to \mathbb{R}$ is a function such that $F(x, 0) = f(x)$, U is a linear space and $u \in U$ is a parameter one chooses depending on the particular problem at hand. For example, u can represent time or some stochastic vector expressing uncertainty in the problem data. Corresponding to this problem, one defines an optimal value function

$$\varphi(u) = \inf_{x \in X} F(x, u)\,, \quad u \in U. \tag{1.16}$$

We then have the following theorem:

Theorem 1.5.1 *Let X, U be real vector spaces, and let $F : X \times U \to \overline{\mathbb{R}}$ be a convex function. Then φ is convex as well.*

Proof: This follows from property 10 of Theorem 1.1.30. □

The following is a more detailed illustration of the dual optimization method: Let X and Y be general linear spaces, and let $K : X \times Y \to \overline{\mathbb{R}}$ be a function. Define

$$f(x) = \sup_{y \in Y} K(x, y) \tag{1.17}$$

and

$$g(y) = \inf_{x \in X} K(x, y). \tag{1.18}$$

Then, consider two optimization problems

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

and

$$(D) \qquad \max_{y \in Y} g(y).$$

From the definitions

$$g(y) \leq K(x, y) \leq f(x), \quad \forall\, x \in X, \forall\, y \in Y. \tag{1.19}$$

By taking the infimum over x and then the supremum over y in equation (1.19)

$$\inf_{x \in X} \sup_{y \in Y} K(x, y) = \inf_{x \in X} f(x) \geq \sup_{y \in Y} g(y) = \sup_{y \in Y} \inf_{x \in X} K(x, y). \tag{1.20}$$

If there is equality in equation (1.20), then the common value is called the *saddle value* of K.

The saddle value exists if K has a *saddle point*, i.e. there exists a point (x', y') such that

$$K(x', y) \leq K(x', y') \leq K(x, y') \tag{1.21}$$

for all $x \in X$ and for all $y \in Y$. If such a point exists, the saddle value of K is $K(x', y')$.

From these definitions, one can prove the following theorem.

Theorem 1.5.2 *A point (x', y') is a saddle point for K if and only if x' solves (P), y' solves (D), and the saddle value of K exists, i.e.*

$$\inf_{x \in X} f(x) = \sup_{y \in Y} g(y)$$

The proof is left as an exercise.

Because of this theorem, (P) and (D) are called *dual problems*, since they can be considered as half of the problem of finding a saddle point for K.

Hence, in order to prove that (P) and (D) have a solution, and actually find this solution one can instead attempt to find a saddle point for the function K.

In convex optimization, one is interested in what has been done above in the opposite order: If one starts with (P), where $f : X \to \mathbb{R}$, how can one choose a space Y and a function K on $X \times Y$ such that $f(x) = \sup_{y \in Y} K(x, y)$ holds? This approach gives freedom to choose Y and K in different ways, so that one can (hopefully) achieve the properties one would like Y and K to have. This idea is called *the duality approach.*

1.6 Examples of convex optimization via duality

Example 1.6.1 (Nonlinear programming) *Let $f_0, f_1, \dots, f_m$ be real valued, convex functions on a nonempty, convex set C in the vector space X. The duality approach consists of the following steps:*

1. *The given problem:* $\min f_0(x)$ *over* $\{x \in C\ :\ f_i(x) \leq 0\ \forall\ i = 1, \dots, m\}$.
2. *Abstract representation:* $\min f$ *over X, where*

$$f(x) = \begin{cases} f_0(x) & x \in C,\ f_i(x) \leq 0 \text{ for } i = 1, \dots, m \\ +\infty & \text{for all other } x \in X. \end{cases}$$

3. *Parametrization: Define (for example) $F(x, u)$ for $u = (u_1, \dots, u_m) \in \mathbb{R}^m$ by $F(x, u) = f_0(x)$ if $x \in C, f_i(x) \leq u_i$ for $i = 1, \dots, m$, and $F(x, u) = +\infty$ for all other x. Then, $F : X \times \mathbb{R}^m \to [-\infty, +\infty]$ is convex and $F(x, 0) = f(x)$*

Example 1.6.2 (Nonlinear programming with infinitely many constraints) *Let $f_0 : C \to \mathbb{R}$ where $C \subset X$ is convex, and let $h : X \times S \to \bar{\mathbb{R}}$ be convex in the x-argument, where S is an arbitrary set*

1. *The problem:* $\min f_0(x)$ *over* $K = \{x \in C : h(x, s) \leq 0\ \forall\ s \in S\}$.
2. *Abstract representation:* $\min f(x)$ *over X, where $f(x) = f_0(x)$ if $x \in K$, and $f(x) = +\infty$ for all other x.*
3. *Parametrization: Choose u analoglously with Example 1.6.1: Let U be the linear space of functions $u : S \to \mathbb{R}$ and let $F(x, u) = f_0(x)$ if $x \in C$, $h(x, s) \leq u(s)\ \forall\ s \in S$ and $F(x, u) = +\infty$ for all other x. As in the previous example, this makes F convex and satisfies $F(x, 0) = f(x)$.*

Example 1.6.3 (Stochastic optimization) *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $h : X \times \Omega \to \bar{\mathbb{R}}$ be convex in the x-argument, where X is a linear, topological space. Let C be a closed, convex subset of X.*

1. *The general problem:* $\min h(x, \omega)$ *over all $x \in C$, where ω is a stochastic element with a known distribution. The difficulty here is that x must be chosen before ω has been observed.*
2. *We therefore solve the following problem: Minimize the expectation $f(x) = \int_\Omega h(x, \omega) dP(\omega)$ over all $x \in X$. Here, it is assumed that h is measurable, so that f is well defined. Rockafellar then shows in [18], Theorem 3, that f actually is convex.*
3. *Parametrization: Let $F(x, u) = \int_\Omega h(x - u(\omega), \omega) dP(\omega) + \delta_C(u)$ for $u \in U$, where U is a linear space of measurable functions and δ_C is the indicator function of C, as defined in Definition 1.1.16. Then F is (by the same argument as for f) well defined and convex, with $F(x, 0) = f(x)$.*

1.7 Conjugate functions in paired spaces

The material in this section is based on Rockafellar [18] and Rockafellar and Wets [20].

Definition 1.7.1 (Pairing of spaces) *A pairing of two linear spaces X and V is a real valued bilinear form $\langle \cdot, \cdot \rangle$ on $X \times V$.*

The pairing associates for each $v \in V$ a linear function $\langle \cdot, v \rangle : x \mapsto \langle x, v \rangle$ on X, and similarly for X.

Definition 1.7.2 (Compatible topology) *Assume there is a pairing between the spaces X and V. A topology on X is compatible with the pairing if it is a locally convex topology such that the linear function $\langle \cdot, v \rangle$ is continuous, and any continuous linear function on X can be written in this form for some $v \in V$. A compatible topology on V is defined similarly.*

Definition 1.7.3 (Paired spaces) *X and V are paired spaces if one has chosen a pairing between X and V, and the two spaces have compatible topologies with respect to the pairing.*

Example 1.7.4 *Let $X = \mathbb{R}^n$ and $V = \mathbb{R}^n$. Then, the standard Euclidean inner product is a bilinear form, so X and V become paired spaces.*

Example 1.7.5 *Let $X = L^1(\Omega, \mathcal{F}, P)$ and $V = L^\infty(\Omega, \mathcal{F}, P)$. Then X and V are paired via the bilinear form $\langle x, v \rangle = \int_\Omega x(s) v(s) dP(s)$. Similarly, the spaces $X = L^p(\Omega, F, P)$ and $V = L^q(\Omega, F, P)$, where $\frac{1}{p} + \frac{1}{q} = 1$, are paired.*

We now come to a central notion of convex duality, the conjugate of a function.

Definition 1.7.6 (Convex conjugate of a function, f^*) *Let X and V be paired spaces. For a function $f : X \to \bar{\mathbb{R}}$, define the conjugate of f, denoted by $f^* : V \to \bar{\mathbb{R}}$, by*

$$f^*(v) = \sup\{\langle x, v\rangle - f(x) : x \in X\}. \tag{1.22}$$

Finding f^* is called taking the conjugate of f in the convex sense. One may also define the conjugate g^* of a function $g : V \to \bar{\mathbb{R}}$ similarly.

Similarly, define

Definition 1.7.7 (Biconjugate of a function, f^{**}) *Let X and V be paired spaces. For a function $f : X \to \bar{\mathbb{R}}$, define* the biconjugate *of f, f^{**}, to be the conjugate of f^*, so $f^{**}(x) = \sup\{\langle x, v\rangle - f^*(v) : v \in V\}$.*

Definition 1.7.8 (The Fenchel transform) *The operation $f \mapsto f^*$ is called* the Fenchel transform*.*

If $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, then the operation $f \mapsto f^*$ is sometimes called the *Legendre-Fenchel transform*.

To understand why the conjugate function f^* is important, it is useful to consider it via the epigraph. This is most easily done in $\mathbb{R}^n$, so let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ and consider $X = \mathbb{R}^n = V$. From equation (1.22), it is not difficult to show that

$$(v, b) \in \mathrm{epi}(f^*) \iff b \geq \langle v, x\rangle - a \text{ for all } (x, a) \in \mathrm{epi}(f). \tag{1.23}$$

This can also be expressed as

$$(v, b) \in \mathrm{epi}(f^*) \iff l_{v,b} \leq f \tag{1.24}$$

where $l_{v,b}(x) := \langle v, x\rangle - b$. So, since specifying a function on $\mathbb{R}^n$ is equivalent to specifying its epigraph, equation (1.24) shows that f^* describes the family of all affine functions that are majorized by f (since all affine functions on $\mathbb{R}^n$ are of the form $\langle v, x\rangle - b$ for fixed v, b).

But from equation (1.23)

$$b \geq f^*(v) \iff b \geq l_{x,a}(v) \text{ for all } (x, a) \in \mathrm{epi}(f).$$

This means that f^* is the pointwise supremum of all affine functions $l_{x,a}$ for $(x, a) \in \mathrm{epi}(f)$.

This is illustrated in Figures 1.11 and 1.12.

We then have the following very central theorem on duality, which is Theorem 5 in Rockafellar [18]:

Theorem 1.7.9 *Let $f : X \to \bar{\mathbb{R}}$ be arbitrary. Then the conjugate f^* is a closed (as defined in Section 1.1), convex function on V and $f^{**} = \mathrm{cl}(\mathrm{co}(f))$. Similarly if one starts with a function in V. In particular, the Fenchel transform induces a one-to-one correspondence $f \mapsto h, h = f^*$ between the closed, convex functions on X and the closed, convex functions on V.*

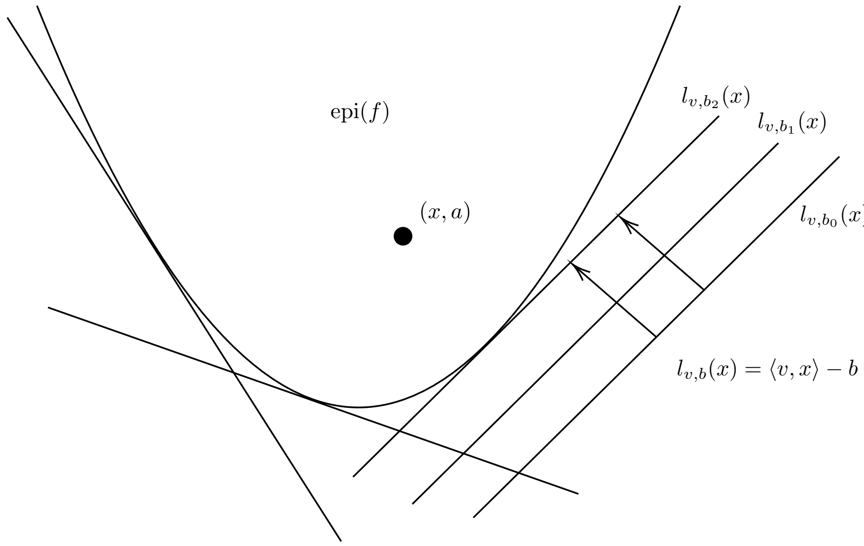

Figure 1.11: Affine functions majorized by f.

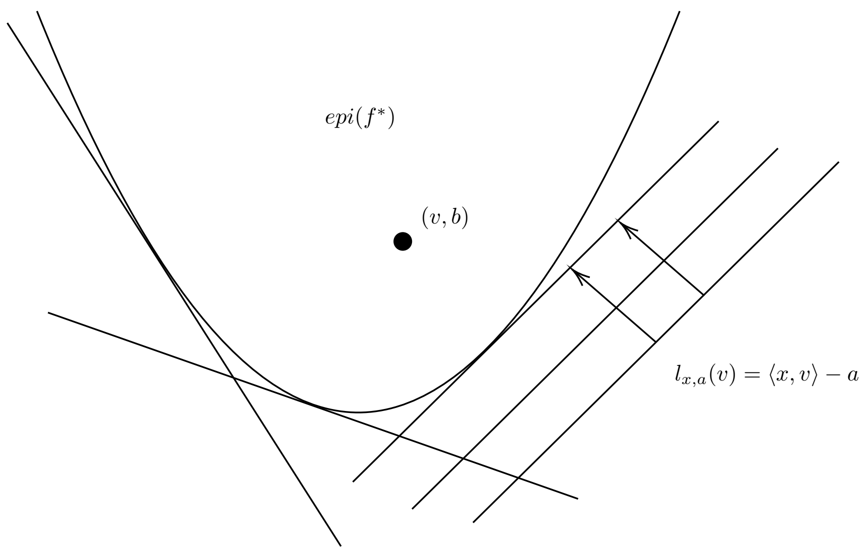

Figure 1.12: Affine functions majorized by f^*.

Proof: By definition f^* is the pointwise supremum of the continuous, affine functions $V \mapsto \langle x, v\rangle - \alpha$, where $(x, \alpha) \in \mathrm{epi}(f)$. Therefore, f^* is convex and lsc, hence it is closed. $(v, \beta) \in \mathrm{epi}(f^*)$ if and only if the continuous affine function $x \mapsto \langle x, v\rangle - \beta$ satisfies $f(x) \geq \langle x, v\rangle - \beta$ for all $x \in X$, that is if the epigraph of this affine function contains the epigraph of f. Thus, $\mathrm{epi}(f^{**})$ is the intersection of all the nonvertical, closed halfspaces in $X \times \mathbb{R}$ containing epi(f). This implies, using what a closed, convex set is, that $f^{**} = \mathrm{cl}(\mathrm{co}(f))$. $\square$

Theorem 1.7.9 implies that if f is convex and closed, then $f = f^{**}$. This gives a one-to-one correspondence between the closed convex functions on X, and the same type of functions on V. Hence, all properties and operations on such functions must have conjugate counterparts (see [20]).

Example 1.7.10 *Let X and V be paired spaces, and let $f = \delta_L$ where $L \subseteq X$ is a subspace (so in particular, L is convex) and δ_L is the indicator function of L, as in Definition 1.1.16. It follows from Example 1.1.29 that $f = \delta_L$ is convex. Then*

$$\begin{aligned} \delta_L^*(v) &= \sup\{\langle x, v\rangle - \delta_L(x) : x \in X\} \\ &= \sup\{\langle x, v\rangle ; x \in L\} \end{aligned}$$

since $\langle x, v\rangle - \delta_L(x) = -\infty$ if $x \notin L$. This function δ_L^* is called the support function *of L (and is often denoted by ψ_L). Note also that*

$$\delta_L^*(v) = \delta_{L^\perp}(v)$$

because if $v \in L^\perp$, then $\langle x, v\rangle = 0$ for all $x \in L$, but if $v \notin L^\perp$ then $\langle x', v\rangle \neq 0$ for some $x' \in L$. Hence, since L is a subspace, $\langle x', v\rangle$ can be made arbitrarily large by multiplying x' by either $+t$ or $-t$ (in order to make $\langle x', v\rangle$ positive), and letting $t \to +\infty$.

By a similar argument

$$\delta_L^{**} = \delta_{(L^\perp)^\perp}. \tag{1.25}$$

We will now use conjugate duality to prove a central result in functional analysis, namely that for any subspace $L \subseteq X$, $(L^\perp)^\perp = \bar{L}$ (see for instance *Linear Functional Analysis* by Rynne and Youngston [22]).

Theorem 1.7.11 *Let $L \subseteq X$ be a subspace. Then $(L^\perp)^\perp = \bar{L}$.*

Proof:

From Example 1.7.10

$$\delta_L^{**} = \delta_{(L^\perp)^\perp} \tag{1.26}$$

But then, Theorem 1.7.9 implies that $\delta_{(L^\perp)^\perp} = \mathrm{cl}(\mathrm{co}(\delta_L))$. δ_L is convex, so $\mathrm{co}(\delta_L) = \delta_L$. To proceed, we make the following claim:

Claim: $\mathrm{cl}(\delta_L) = \delta_{\bar{L}}$.

Proof of Claim: From Definition 1.1.28, $\mathrm{cl}(\delta_L) = \mathrm{lsc}(\delta_L) =$ the largest lower semi-continuous function that is less than or equal to δ_L. $\delta_{\bar{L}}$ is lower semi-continuous since $\bar{L}$ is closed (from Theorem 1.1.26). Also, from Example 1.7.10,

$\delta_{\bar{L}} \leq \delta_L$ since $L \subseteq \bar{L}$. All that remains to be proved is that if f is lower semi-continuous and $f \leq \delta_L$, then $f \leq \delta_{\bar{L}}$.

So assume that f is lower semi-continuous and $f \leq \delta_L$. We know that $\delta_L(L) = \delta_{\bar{L}}(L)$, so $f(L) \leq \delta_L(L) \leq \delta_{\bar{L}}(L)$, from the assumption that $f \leq \delta_L$.

If $x \in (\bar{L})^\perp$, then $\delta_{\bar{L}}(x) = +\infty$, so $f(x) \leq \delta_{\bar{L}}(x)$.

Finally, if $x \in \bar{L} \setminus L$, then $\delta_L(x) = +\infty$, but $\delta_{\bar{L}}(x) = 0$. Hence, we must show that $f(x) \leq 0$. Since f is lower semi-continuous, Theorem 1.1.26 implies that the sublevel set $S_0(f) = \{x \in X : f(x) \leq 0\}$ is closed. Because $f \leq \delta_L$, $L \subseteq S_0(f)$, hence (since $S_0(f)$ is closed) $\bar{L} \subseteq S_0(f)$, so $f(x) \leq 0$ for all $x \in \bar{L}$.

So the claim is proved.

The arguments above imply that

$$\delta_{(L^\perp)^\perp} = \delta_L^{**} = \mathrm{cl}(\mathrm{co}(\delta_L)) = \mathrm{cl}(\delta_L) = \delta_{\bar{L}}$$

where the final equality uses the claim. But this again implies that $(L^\perp)^\perp = \bar{L}$. □

For a concave function $g : X \to \bar{\mathbb{R}}$ one can define the conjugate as:

$$g^*(v) = \inf\{\langle x, v\rangle - g(x) : x \in X\} \tag{1.27}$$

This is called taking the *conjugate of g in the concave sense*.

1.8 Dual problems and Lagrangians

This is our situation as of now. We have an abstract minimization problem:

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

which is assumed to have the representation:

$$f(x) = F(x, 0), \quad F : X \times U \to \bar{\mathbb{R}} \tag{1.28}$$

(where U is some linear space). Everything now depends on the choice of U and F. We want to exploit duality, so let X be paired with V, and U paired with Y, where U and Y are linear spaces (the choice of pairings may also be important in applications). Preferably, we want to choose (F, U) such that F is a closed, jointly convex function of x and u.

Definition 1.8.1 (The Lagrange function, $K(x, y)$) *Define the Lagrange function $K : X \times Y \to \bar{\mathbb{R}}$ to be*

$$K(x, y) = \inf\{F(x, u) + \langle u, y\rangle : u \in U\}. \tag{1.29}$$

The following theorem is Theorem 6 in Rockafellar [18]. It says that K is a closed convex function which satisfies a certain inequality, and that all functions of this form actually are the Lagrange function associated with some function f.

Theorem 1.8.2 *The Lagrange function K is closed, concave in $y \in Y$ for each $x \in X$, and if $F(x,u)$ is closed and convex in u*

$$f(x) = \sup_{y \in Y} K(x,y). \tag{1.30}$$

Conversely, if K is an arbitrary extended-real valued function on $X \times Y$ such that (1.30) holds, and if K is closed and concave in y, then K is the Lagrange function associated with a unique representation $f(x) = F(x,0), F : X \times U \to \bar{\mathbb{R}}$ where F is closed and convex in u. This means that

$$F(x,u) = \sup\{K(x,y) - \langle u, y\rangle : y \in Y\}.$$

Further, if F is closed and convex in u, K is convex in x if and only if $F(x,u)$ is jointly convex in x and u.

Proof: Everything in the theorem, apart from the last statement, follows from Theorem 1.7.9. For the last statement, assume that F and K respectively are convex, use the definitions of F and K and that the supremum and infimum of convex functions are convex (see Theorem 1.1.30). □

We now define, motivated by equation (1.30), the dual problem of (P),

$$(D) \qquad \max_{y \in Y} g(y)$$

where $g(y) = \inf_{x \in X} K(x,y)$.

Note that this dual problem gives a lower bound on the primal problem, from (1.30) since

$$K(x,y) \geq \inf_{x \in X} K(x,y) = g(y).$$

But then

$$\sup_{y \in Y} K(x,y) \geq \sup_{y \in Y} g(y).$$

So from equation (1.30), $f(x) \geq \sup_{y \in Y} g(y)$. Therefore, taking the infimum with respect to $x \in X$ on the left hand side implies $(D) \leq (P)$. This is called *weak duality*. Sometimes, one can prove that the dual and primal problems have the same optimal value. If this is the case, there is *no duality gap* and *strong duality holds*.

The next theorem (Theorem 7 in Rockafellar [18]) is important:

Theorem 1.8.3 *The function g in (D) is closed and concave. By taking the conjugate in concave sense, $g = -\varphi^*$, hence $-g^* = \mathrm{cl}(\mathrm{co}(\varphi))$, so*

$$\sup_{y \in Y} g(y) = \mathrm{cl}(\mathrm{co}(\varphi))(0)$$

while

$$\inf_{x \in X} f(x) = \varphi(0)$$

In particular, if $F(x,u)$ *is convex in* (x,u)*, then* $-g^* = \text{cl}(\varphi)$ *and* $\sup_{y \in Y} g(y) = \liminf_{u \to 0} \varphi(u)$ *(except if* $0 \notin \text{cl}(\text{dom}(\varphi)) \neq \emptyset$*, and* $\text{lsc}(\varphi)$ *is nowhere finite valued).*

For the proof, see Rockafellar [18].

What makes this theorem important is that it converts the question of whether $\inf_{x \in X} f(x) = \sup_{y \in Y} g(y)$ and the question of whether the saddle value of the Lagrange function K exists, to a question of whether the optimal value function φ satisfies $\varphi(0) = (\text{cl}(\text{co}(\varphi)))(0)$. Hence, *if the value function* φ *is convex, the lower semi-continuity of* φ *is a sufficient condition for the absence of a duality gap.*

By combining the results of the previous sections, we get the following rough summary of the duality method, based on conjugate duality:

- To begin, there is a minimization problem $\min_{x \in X} f(x)$ which cannot be solved directly.
- Find a function $F : X \times U \to \bar{\mathbb{R}}$, where U is a vector space, such that $f(x) = F(x, 0)$.
- Introduce the linear space Y, paired with U, and define the Lagrange function $K : X \times Y \to \bar{\mathbb{R}}$ by $K(x,y) = \inf_{u \in U}\{F(x,u) + \langle u, y \rangle\}$.
- Try to find a saddle point for K. If this succeeds, Theorem 1.5.2 tells us that this gives the solution of (P) and (D).
- Theorem 1.8.3 tells us that K has a saddle point if and only if $\varphi(0) = (\text{cl}(\text{co}(\varphi)))(0)$. Hence, if the value function φ is convex, the lower semi-continuity of φ is a sufficient condition for the absence of a duality gap.

We can look at an example illustrating these definitions, based on Example 1.6.1.

Example 1.8.4 (Nonlinear programming) *The Lagrange function takes the form*

$$\begin{aligned}
K(x,y) &= \inf\{F(x,u) + \langle u, y \rangle : u \in U\} \\
&= \inf\{\begin{cases} f_0(x) + \langle u, y \rangle; x \in C, f_i(x) \leq u_i \\ +\infty + \langle u, y \rangle; \forall \text{ other } x \end{cases} : u \in U\} \\
&= \begin{cases} f_0(x) + \inf\{\langle u, y \rangle : u \in U, f_i(x) \leq u_i\}, x \in C \\ +\infty, \text{ otherwise.} \end{cases} \\
&= \begin{cases} \inf\{f_0(x) + f_1(x)y_1 + \ldots + f_m(x)y_m\}, u \in U, x \in C, y \in \mathbb{R}^m_+ \\ -\infty, x \in C, y \notin \mathbb{R}^m_+ \\ +\infty, \text{ otherwise.} \end{cases}
\end{aligned}$$

where the last equality follows because if there is at least one negative y_j, one can choose u_j arbitrarily large and make the above expression arbitrarily small. Therefore, the dual function is

$$\begin{aligned} g(y) &= \inf_{x\in X} K(x,y) \\ &= \inf_{x\in X} \begin{cases} f_0(x)+f_1(x)y_1+\ldots+f_m(x)y_m \text{ if } x\in C, y\in\mathbb{R}^m_+ \\ -\infty, x\in C, y\notin\mathbb{R}^m_+ \\ +\infty, \text{ otherwise.} \end{cases} \\ &= \begin{cases} \inf_{x\in C}\{f_0(x)+f_1(x)y_1+\ldots+f_m(x)y_m\} \text{ if } y\in\mathbb{R}^m_+ \\ -\infty, y\notin\mathbb{R}^m_+. \end{cases} \end{aligned}$$

By making some small alterations to the approach above, Rockafellar [18] shows that by beginning with the standard primal linear programming problem (abbreviated LP-problem)

$$\max\{\langle c,x\rangle : Ax\leq b,\ x\geq 0\}$$

where c and b are given vectors and A is a given matrix, and finding its dual problem (in the above sense), one gets the standard dual LP-problem back. That is

$$\min\{\langle b,y\rangle : A^T y\geq c,\ y\geq 0\}$$

(see Vanderbei [24]).

1.9 Exercises to Chapter 1

Exercise 1.1: Prove Theorem 1.1.14.

Exercise 1.2: Prove items 3.-5. of Theorem 1.1.30.

Exercise 1.3: Prove Proposition 1.2.2.

Exercise 1.4: Prove that all convex functions are quasiconvex.

Exercise 1.5: Prove that $f(x)=\log(x)$ defined on the positive real numbers is concave *and* quasiconvex.

Exercise 1.6: Prove Theorem 1.5.2.

Exercise 1.7: Prove that Definition 1.1.19 implies Definition 1.1.17. Note that this completes the proof of Theorem 1.1.20.

Chapter 2

Convex risk measures and duality

The purpose of this chapter is to present a possible way to quantify monetary risk. In practice, several different risk measures are used for this purpose. However, we will focus on a class of such measures, called convex risk measures, which satisfy certain economically reasonable properties. In Section 2.1, we define convex risk measures and some corresponding notions, such as coherent risk measures and acceptance sets. Then, we derive some properties of convex risk measures. In Section 1.5 we present a brief introduction to convex duality theory. This theory is needed in Section 2.2, where we prove dual representation theorems for convex risk measures. These theorems provide an alternative characterisation of these measures. In Section 2.3, we give some examples of measures of risk commonly used in finance.

2.1 Convex and coherent risk measures

In the literature, there are many different methods for quantifying risk depending on the context and purpose. In this section, we focus on monetary measures of financial risk. We will introduce measures for the risk of a financial position, X, which takes a random value at some set terminal time. This value depends on the current world scenario. An intuitive approach for quantifying risk is the variance. However, the variance does not separate between negative and positive deviations. Hence, it is only suitable as a measure of risk in cases where any kind of deviation from the target is a problem. In situations where deviations in one direction is OK, or even good, while deviations is the other direction is bad, the variance is unsuitable. Clearly, the variance is an unsuitable measure of financial risk. In finance, positive deviations are good (earning more money), but negative deviations are bad (earning less money). In order to resolve this, Artzner et al. [1] set up some economically reasonable axioms that a measure of risk should satisfy and thereby introduced coherent risk measures. This notion has later been modified to socalled convex risk measures.

In order to understand convex risk measures properly, we need some essential concepts from measure theory. We recall these definitions here for completeness, and refer the reader to Shilling [23] for a detailed introduction to measure- and integration theory. Consider a given scenario space Ω. This may be a finite set $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ or an infinite set. On this space, we can define a *σ-algebra* $\mathcal{F}$, i.e. a family of subsets of Ω that contains the empty set $\emptyset$ and is closed under complements and countable unions. The elements in the σ-algebra $\mathcal{F}$ are called *measurable sets*. $(\Omega, \mathcal{F})$ is then called a *measurable space*. A *measurable function* is a function $f : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ (where $(\Omega', \mathcal{F}')$ is another measurable space) such that for any measurable set, $E \in \mathcal{F}'$, the inverse image (preimage), $f^{-1}(E)$, is a measurable set, i.e.,

$$f^{-1}(E) := \{\omega \in \Omega : f(\omega) \in E\} \in \mathcal{F}.$$

A *random variable* is a real-valued measurable function. On a measurable space $(\Omega, \mathcal{F})$ one can define a *measure*, i.e. a non-negative countably additive function $\mu : \Omega \to \mathbb{R}$ such that $\mu(\emptyset) = 0$. Then, $(\Omega, \mathcal{F}, \mu)$ is called a measure space. A *signed measure* is the same as a measure, but without the non-negativity requirement. A *probability measure* is a measure P such that $P(\Omega) = 1$. Let $\mathcal{P}$ denote the set of all probability measures on $(\Omega, \mathcal{F})$, and $\mathbb{V}$ the set of all measures on $(\Omega, \mathcal{F})$. Then $\mathbb{V}$ is a vector space (also called linear space), and $\mathcal{P} \subseteq \mathbb{V}$ is a convex set (check this yourself as an exercise!).

In the following, let Ω be a fixed set of scenarios, or possible states of the world. Note that we make no further assumptions on Ω, so in particular, it may be infinite. Consider the measure space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F}$ is a given σ-algebra on Ω and P is a given probability measure on $(\Omega, \mathcal{F})$. A financial position (such as a portfolio of stocks) can be described by a mapping $X : \Omega \to \mathbb{R}$, where $X(\omega)$ is the value of the position at the end of the trading period if the state ω occurs. More formally, X is a random variable. Hence, the dependency of X on ω describes the uncertainty of the value of the portfolio. Let $\mathbb{X}$ be a given vector space of such random variables $X : \Omega \to \mathbb{R}$, which contains the constant functions. For $c \in \mathbb{R}$, let $c\mathbf{1} \in \mathbb{X}$ denote the constant function $c\mathbf{1}(\omega) = c$ for all $\omega \in \Omega$. An example of such a space is

$$L^p(\Omega, \mathcal{F}, P) := \{f : f \text{ is measurable and } \big(\int_\Omega |f(\omega)|^p dP(\omega)\big)^{1/p} < \infty\},\ 1 \le p \le \infty.$$

A convex risk measure is defined as follows:

Definition 2.1.1 (Convex risk measure) *A convex risk measure is a function $\rho : \mathbb{X} \to \mathbb{R}$ which satisfies the following for each $X, Y \in \mathbb{X}$:*

(i) (Convexity) $\rho(\lambda X + (1-\lambda)Y) \le \lambda\rho(X) + (1-\lambda)\rho(Y)$ *for* $0 \le \lambda \le 1$.

(ii) (Monotonicity) If $X \le Y$, *then* $\rho(X) \ge \rho(Y)$.

(iii) (Translation invariance) If $m \in \mathbb{R}$, *then* $\rho(X + m\mathbf{1}) = \rho(X) - m$.

If $\rho(X) \leq 0$, X is *acceptable* since it does not have a positive risk. On the other hand, if $\rho(X) > 0$, X is *unacceptable*.

If a convex risk measure also satisfies *positive homogeneity*, that is if

$$\lambda \geq 0 \Rightarrow \rho(\lambda X) = \lambda \rho(X)$$

then ρ is called a *coherent risk measure*. The original definition of a coherent risk measure, did not involve convexity directly, but instead required subadditivity:

Definition 2.1.2 (Coherent risk measure) *A coherent risk measure is a function $\pi : \mathbb{X} \to \mathbb{R}$ which satisfies the following for each $X, Y \in \mathbb{X}$:*

(*i*) *(Positive homogeneity) $\pi(\lambda X) = \lambda \pi(X)$ for $\lambda \geq 0$.*

(*ii*) *(Subadditivity) $\pi(X + Y) \leq \pi(X) + \pi(Y)$.*

(*iii*) *(Monotonicity) If $X \leq Y$, then $\pi(X) \geq \pi(Y)$.*

(*iv*) *(Translation invariance) If $m \in \mathbb{R}$, then $\pi(X + m\mathbf{1}) = \pi(X) - m$.*

We can interpret ρ as a *capital requirement*, that is: $\rho(X)$ is the extra amount of money which should be added to the portfolio in a risk free way to make the position acceptable for an agent.

The conditions in Definition 2.1.1 are quite natural. The convexity reflects that diversification reduces risk. The total risk of loss in two portfolios should be reduced when the two are weighed into a mixed portfolio. Roughly speaking, spreading your eggs in several baskets should reduce the risk of broken eggs.

Monotonicity says that the downside risk, the risk of loss, is reduced by choosing a portfolio that has a higher value in every possible state of the world.

Finally, translation invariance can be interpreted in the following way: ρ is the amount of money one needs to add to the portfolio in order to make it acceptable for an agent. Hence, if one adds a risk free amount m to the portfolio, the capital requirement should be reduced by the same amount.

As mentioned, Artzner et al. [1] originally defined coherent risk measures, that is, they required positive homogeneity. The reason for skipping this requirement in the definition of a convex risk measure is that positive homogeneity means that risk grows linearly with X, and this may not always be the case. In the following, we consider convex risk measures. However, the results can be proved for coherent risk measures as well.

Starting with n convex risk measures, one can derive more convex risk measures, as in the following theorem. This was proven by Rockafellar in [19], Theorem 3, for coherent risk measures.

Theorem 2.1.3 *Let $\rho_1, \rho_2, \ldots, \rho_n$ be convex risk measures.*

1. *If $\lambda_1, \lambda_2, \ldots, \lambda_n \geq 0$ and $\sum_{i=1}^n \lambda_i = 1$, then $\rho = \sum_{i=1}^n \lambda_i \rho_i$ is a convex risk measure as well.*
2. *$\rho = \max\{\rho_1, \rho_2, \ldots, \rho_n\}$ is a convex risk measure.*

Proof:

1. Let's check the conditions of Definition 2.1.1. Obviously, $\rho : \mathbb{X} \to \mathbb{R}$, so we check for any $X, Y \in \mathbb{X}$, $0 \leq \lambda \leq 1$:

   (i) : This follows from that a sum of convex functions is also a convex function, and that a positive constant times a convex function is still convex.

   (ii) : If $X \leq Y$, then $\rho(X) = \sum_{i=1}^{n} \lambda_i \rho_i(X) \geq \sum_{i=1}^{n} \lambda_i \rho_i(Y) = \rho(Y)$.

   (iii) : If $m \in \mathbb{R}$,

$$\begin{aligned} \rho(X+m) &= \sum_{i=1}^{n} \lambda_i \rho_i(X+m) \\ &= \sum_{i=1}^{n} \lambda_i (\rho_i(X) - m) \\ &= \sum_{i=1}^{n} \lambda_i \rho_i(X) - m \sum_{i=1}^{n} \lambda_i \\ &= \rho(X) - m. \end{aligned}$$

2. The proof is left as an exercise.

□

Associated with every convex risk measure ρ, there is a natural set of all acceptable portfolios, called the *acceptance set*, $\mathcal{A}_\rho$, of ρ.

Definition 2.1.4 (The acceptance set of a convex risk measure, $\mathcal{A}_\rho$) *A convex risk measure ρ induces a set*

$$\mathcal{A}_\rho = \{X \in \mathbb{X} : \rho(X) \leq 0\}$$

The set $\mathcal{A}_\rho$ is called the acceptance set *of ρ.*

Conversely, given a class $\mathcal{A} \subseteq \mathbb{X}$, one can associate a quantitative risk measure $\rho_{\mathcal{A}}$ to it.

Definition 2.1.5 (Associated measure of risk) *Let $\mathcal{A} \subseteq \mathbb{X}$ be a set of "acceptable" random variables. This set has an associated measure of risk $\rho_{\mathcal{A}}$ defined as follows: For $X \in \mathbb{X}$, let*

$$\rho_{\mathcal{A}}(X) = \inf\{m \in \mathbb{R} : X + m \in \mathcal{A}\}. \tag{2.1}$$

This means that $\rho_{\mathcal{A}}(X)$ measures how much one must add to the portfolio X, in a risk free way, to get the portfolio into the set $\mathcal{A}$ of acceptable portfolios. This is the same interpretation as for a convex risk measure.

The previous definitions show that one can either start with a risk measure, and derive an acceptance set, or one can start with a set of acceptable random variables, and derive a risk measure.

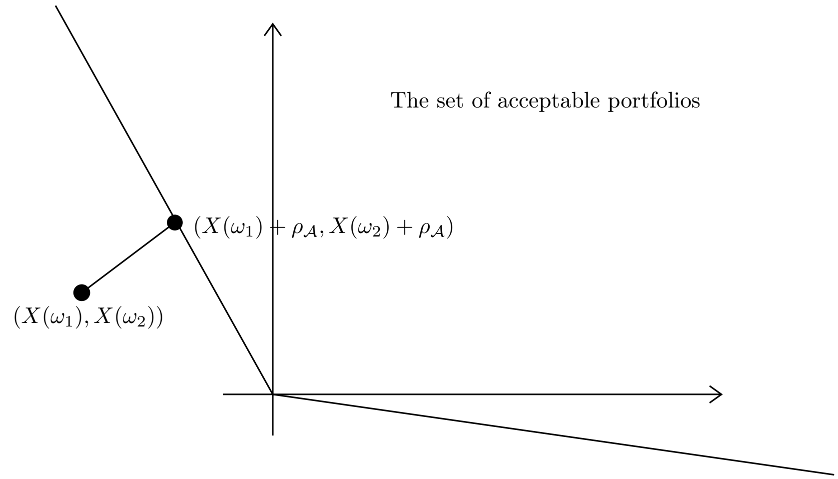

Figure 2.1: Illustration of the risk measure $\rho_{\mathcal{A}}$ associated with a set $\mathcal{A}$ of acceptable portfolios.

Example 2.1.6 (Illustration of the risk measure $\rho_{\mathcal{A}}$ associated with a set $\mathcal{A}$ of acceptable portfolios) *Let $\Omega = \{\omega_1, \omega_2\}$, and let $X : \Omega \to \mathbb{R}$ be a portfolio. Let $x = (X(\omega_1), X(\omega_2))$. If the set of acceptable portfolios is as in Figure 2.1 , the risk measure $\rho_{\mathcal{A}}$ associated with the set $\mathcal{A}$ can be illustrated as in the figure.*

Based on this theory, a theorem on the relationship between risk measures and acceptable sets can be derived. The following theorem is a version of Proposition 2.2 in Föllmer and Schied [8].

Theorem 2.1.7 *Let ρ be a convex risk measure with acceptance set $\mathcal{A}_\rho$. Then:*

(i) $\rho_{\mathcal{A}_\rho} = \rho$

(ii) *$\mathcal{A}_\rho$ is a nonempty, convex set.*

(iii) *If $X \in \mathcal{A}_\rho$ and $Y \in \mathbb{X}$ are such that $X \leq Y$, then $Y \in \mathcal{A}_\rho$*

(iv) *If ρ is a coherent risk measure, then $\mathcal{A}_\rho$ is a convex cone.*

Conversely, let $\mathcal{A}$ be a nonempty, convex subset of $\mathbb{X}$. Let $\mathcal{A}$ be such that if $X \in \mathcal{A}$ and $Y \in \mathbb{X}$ satisfy $X \leq Y$, then $Y \in \mathcal{A}$. Then, the following holds:

(v) *$\rho_{\mathcal{A}}$ is a convex risk measure.*

(vi) *If $\mathcal{A}$ is a convex cone, then $\rho_{\mathcal{A}}$ is a coherent risk measure.*

(vii) $\mathcal{A} \subseteq \mathcal{A}_{\rho_{\mathcal{A}}}$.

Proof:

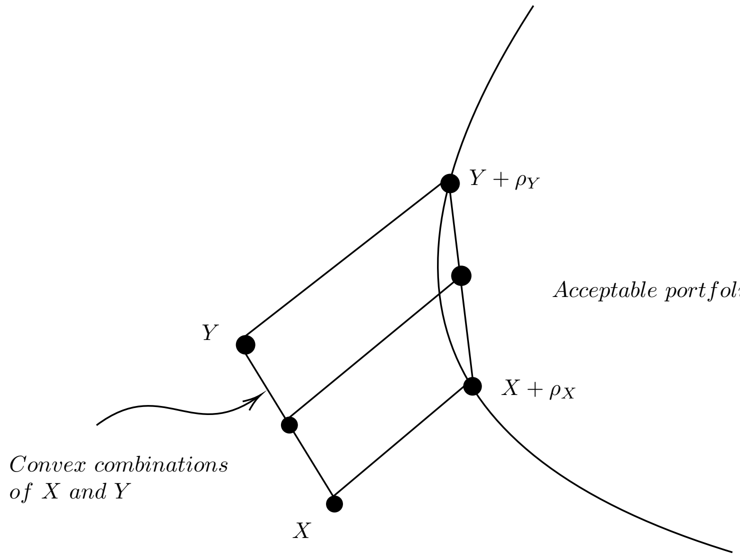

Figure 2.2: Illustration of proof of Theorem 2.1.7 part (v).

(i) For any $X \in \mathcal{A}_\rho$

$$\begin{aligned}
\rho_{\mathcal{A}_\rho}(X) &= \inf\{m \in \mathbb{R} : m + X \in \mathcal{A}_\rho\} \\
&= \inf\{m \in \mathbb{R} : m + X \in \{Y \in \mathbb{X} : \rho(Y) \leq 0\}\} \\
&= \inf\{m \in \mathbb{R} : \rho(m + X) \leq 0\} \\
&= \inf\{m \in \mathbb{R} : \rho(X) - m \leq 0\} \\
&= \inf\{m \in \mathbb{R} : \rho(X) \leq m\} \\
&= \rho(X)
\end{aligned}$$

where we have used the definition of a convex risk measure (Definition 2.1.1) and an acceptance set (Definition 2.1.4).

(ii) $\mathcal{A}_\rho \neq \emptyset$ because $X = 0 \in \mathcal{A}_\rho$. Since ρ is a convex function, $\mathcal{A}_\rho$ is a convex set.

(iii) The proof is left as an exercise.

(iv) The proof is left as an exercise.

(v) We check Definition 2.1.1: $\rho_{\mathcal{A}} : \mathbb{X} \to \mathbb{R}$. Also, for $0 \leq \lambda \leq 1$, $X, Y \in \mathbb{X}$

$$\begin{aligned}
\rho_{\mathcal{A}}(\lambda X + (1-\lambda)Y) &= \inf\{m \in \mathbb{R} : m + \lambda X + (1-\lambda)Y \in \mathcal{A}\} \\
&\leq \lambda \inf\{m \in \mathbb{R} : m + X \in \mathcal{A}\} \qquad (2.2) \\
&\quad +(1-\lambda) \inf\{m \in \mathbb{R} : m + Y \in \mathcal{A}\} \\
&= \lambda \rho_{\mathcal{A}}(X) + (1-\lambda)\rho_{\mathcal{A}}(Y)
\end{aligned}$$

where the inequality follows because $\lambda\rho_{\mathcal{A}}(X) + (1-\lambda)\rho_{\mathcal{A}}(Y) = K + L$, is a real number which will make the portfolio become acceptable since

$$\begin{aligned}(K+L) + (\lambda X + (1-\lambda)Y) &= (K+\lambda X) + (L + (1-\lambda)Y) \\ &= \lambda(\inf\{m \in \mathbb{R} : m + X \in \mathcal{A}\} + X) + \\ &\quad (1-\lambda)(\inf\{m \in \mathbb{R} : m + Y \in \mathcal{A}\} + Y) \in \mathcal{A}\end{aligned}$$

since $\mathcal{A}$ is convex (see Figure 2.2). In addition, if $X, Y \in \mathbb{X}, X \leq Y$

$$\begin{aligned}\rho_{\mathcal{A}}(X) &= \inf\{m \in \mathbb{R} : m + X \in \mathcal{A}\} \\ &\geq \inf\{m \in \mathbb{R} : m + Y \in \mathcal{A}\} \\ &= \rho_{\mathcal{A}}(Y)\end{aligned}$$

since $X \leq Y$. Finally, for $k \in \mathbb{R}$ and $X \in \mathbb{X}$

$$\begin{aligned}\rho_{\mathcal{A}}(X+k) &= \inf\{m \in \mathbb{R} : m + X + k \in \mathcal{A}\} \\ &= \inf\{s - k \in \mathbb{R} : s + X \in \mathcal{A}\} \\ &= \inf\{s \in \mathbb{R} : s + X \in \mathcal{A}\} - k \\ &= \rho_{\mathcal{A}}(X) - k.\end{aligned}$$

Hence, $\rho_{\mathcal{A}}$ is a convex risk measure.

(vi) From (v), all that remains to show is positive homogeneity. For $\alpha > 0$

$$\begin{aligned}\rho_{\mathcal{A}}(\alpha X) &= \inf\{m \in \mathbb{R} : m + \alpha X \in \mathcal{A}\} \\ &= \inf\{m \in \mathbb{R} : \alpha(\frac{m}{\alpha} + X) \in \mathcal{A}\} \\ &= \inf\{m \in \mathbb{R} : \frac{m}{\alpha} + X \in \mathcal{A}\} \\ &= \inf\{\alpha k \in \mathbb{R} : k + X \in \mathcal{A}\} \\ &= \alpha \inf\{k \in \mathbb{R} : k + X \in \mathcal{A}\} \\ &= \alpha\rho_{\mathcal{A}}(X)\end{aligned}$$

where we have used that $\mathcal{A}$ is a convex cone in equality number three. Hence, $\rho_{\mathcal{A}}$ is a coherent risk measure.

(vii) Note that $\mathcal{A}_{\rho_{\mathcal{A}}} = \{X \in \mathbb{X} : \rho_{\mathcal{A}}(X) \leq 0\} = \{X \in \mathbb{X} : \inf\{m \in \mathbb{R} : m + X \in \mathcal{A}\} \leq 0\}$.

Let $X \in \mathcal{A}$, then $\inf\{m \in \mathbb{R} : m + X \in \mathcal{A}\} \leq 0$, since $m = 0$ will suffice (because $X \in \mathcal{A}$). Hence $X \in \mathcal{A}_{\rho_{\mathcal{A}}}$.

□

We would like to derive a alternative, dual characterisation of convex risk measures. However, in order to do so, we need convex duality theory (also called conjugate duality theory). This theory was first introduced by Rockafellar [18]:

2.2 A dual characterisation of convex risk measures

Now, we are ready to derive the dual characterisation of a convex risk measure ρ. Therefore, let V be a vector space that is paired with the vector space $\mathbb{X}$ of financial positions. For instance, if $\mathbb{X}$ is given a Hausdorff topology, so that it becomes a topological vector space (for definitions of these terms, see Pedersen [16]), V can be the set of all continuous linear functionals from $\mathbb{X}$ into $\mathbb{R}$, as in Frittelli and Gianin [10]. Using the theory presented in Section 1.5, a dual characterisation of a convex risk measure ρ can be derived. The following Theorem 2.2.1 was originally proved by Frittelli and Gianin [10]. In the following, ρ^* denotes the conjugate of ρ in the sense of Definition 1.7.6, ρ^{**} is the biconjugate of ρ as in Definition 1.7.7, and $\langle \cdot, \cdot \rangle$ is a pairing.

Theorem 2.2.1 *Let $\rho : \mathbb{X} \to \mathbb{R}$ be a convex risk measure. Assume in addition that ρ is lower semi-continuous. Then $\rho = \rho^{**}$. Hence for each $X \in \mathbb{X}$*

$$\begin{aligned} \rho(X) &= \sup\{\langle X, v \rangle - \rho^*(v) : v \in V\} \\ &= \sup\{\langle X, v \rangle - \rho^*(v) : v \in \mathrm{dom}(\rho^*)\} \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is a pairing between $\mathbb{X}$ and V.

Proof: Since ρ is a convex risk measure, it is a convex function (see Definition 2.1.1). Hence, the convex hull of ρ is equal to ρ, i.e., $\mathrm{co}(\rho) = \rho$ (see Definition 1.1.27). In addition, since ρ is lower semi-continuous and always greater than $-\infty$, ρ is closed (see comment after Definition 1.1.28), so $\mathrm{cl}(\rho) = \rho$. Therefore

$$\mathrm{cl}(\mathrm{co}(\rho)) = \mathrm{cl}(\rho) = \rho.$$

But Theorem 1.7.9 says that $\rho^{**} = \mathrm{cl}(\mathrm{co}(\rho))$, hence $\rho = \rho^{**}$.

The second to last equation in the theorem follows directly from the definition of ρ^{**} (Definition 1.7.7), while the last equation follows because the supremum cannot be achieved when $\rho^* = +\infty$. □

2.2.1 The finite dimensional case

Theorem 2.2.1 is quite abstract, but by choosing a specific set of paired spaces, $\mathbb{X}$ and V, some nice results can be derived. The next theorem is due to Föllmer and Schied [7]. Consider the paired spaces $\mathbb{X} = \mathbb{R}^n, V = \mathbb{R}^n$ with the standard Euclidean inner product, denoted $\cdot$, as pairing. In the following, let $(\Omega, \mathcal{F})$ be a measurable space and let $\mathcal{P}$ denote the set of all probability measures over Ω.

Theorem 2.2.2 *Assume that Ω is finite. Then, any convex risk measure $\rho : \mathbb{X} \to \mathbb{R}$ can be represented in the form*

$$\rho(X) = \sup_{Q \in \mathcal{P}} \{E_Q[-X] - \alpha(Q)\} \tag{2.3}$$

where $E_Q[\cdot]$ denotes the expectation with respect to Q and $\alpha : \mathcal{P} \to (-\infty, \infty]$ is a "penalty function" which is convex and closed. Actually, $\alpha(Q) = \rho^*(-Q)$ for all $Q \in \mathcal{P}$.

Proof: (Luthi and Doege [14]) To show that $\rho : \mathbb{X} \to \mathbb{R}$ (as in Theorem 2.2.2) is a convex risk measure we check Definition 2.1.1: Let $\lambda \in [0,1], m \in \mathbb{R}, X, Y \in \mathbb{X}$.

(i) :

$$\begin{aligned}
\rho(\lambda X + (1-\lambda)Y) &= \sup_{Q \in \mathcal{P}} \{E_Q[-(\lambda X + (1-\lambda)Y)] - \alpha(Q)\} \\
&= \sup_{Q \in \mathcal{P}} \{\lambda E_Q[-X] + (1-\lambda)E_Q[-Y] - \alpha(Q)\} \\
&\leq \lambda \sup_{Q \in \mathcal{P}} \{E_Q[-X] - \alpha(Q)\} \\
&\quad + (1-\lambda) \sup_{Q \in \mathcal{P}} \{E_Q[-Y] - \alpha(Q)\} \\
&= \lambda\rho(X) + (1-\lambda)\rho(Y).
\end{aligned}$$

(ii) : Assume $X \leq Y$. Then $-X \geq -Y$, so

$$\begin{aligned}
\rho(X) &= \sup_{Q \in \mathcal{P}} \{E_Q[-X] - \alpha(Q)\} \\
&\geq \sup_{Q \in \mathcal{P}} \{E_Q[-Y] - \alpha(Q)\} \\
&= \rho(Y).
\end{aligned}$$

(iii) :

$$\begin{aligned}
\rho(X + m\mathbf{1}) &= \sup_{Q \in \mathcal{P}} \{E_Q[-(X + m\mathbf{1})] - \alpha(Q)\} \\
&= \sup_{Q \in \mathcal{P}} \{E_Q[-X] - mE_Q[\mathbf{1}] - \alpha(Q)\} \\
&= \sup_{Q \in \mathcal{P}} \{E_Q[-X] - m - \alpha(Q)\} \\
&= \sup_{Q \in \mathcal{P}} \{E_Q[-X] - \alpha(Q)\} - m \\
&= \rho(X) - m.
\end{aligned}$$

Hence, ρ is a convex risk measure.

So, assume that ρ is a convex risk measure. The conjugate function of ρ, denoted ρ^*, is then defined as $\rho^*(v) = \sup_{X \in \mathbb{X}}\{v \cdot X - \rho(X)\}$ (where $\cdot$ denotes Euclidean inner product) for all $v \in V = \mathbb{R}^n$. Fix an $X \in \mathbb{X}$ and consider $Y_m := X + m\mathbf{1} \in \mathbb{X}$ for an arbitrary $m \in \mathbb{R}$. Then

$$\rho^*(v) \geq \sup_{m \in \mathbb{R}} \{v \cdot Y_m - \rho(Y_m)\}$$

because $\{Y_m\}_{m\in\mathbb{R}} \subset \mathbb{X}$. This means that

$$\begin{aligned} \rho^*(v) &\geq \sup_{m\in\mathbb{R}}\{v\cdot(X+m\mathbf{1})-\rho(X+m\mathbf{1})\} \\ &= \sup_{m\in\mathbb{R}}\{m(v\cdot\mathbf{1}+1)\}+v\cdot X-\rho(X) \end{aligned}$$

where the equality follows from the translation invariance of ρ (see Definition 2.1.1). The first term in the last expression is only finite if $v\cdot\mathbf{1}+1=0$, (where $\mathbf{1}=(1,1,\ldots,1)\in\mathbb{R}^n$) i.e. if $\sum_{i=1}^n v_i=-1$ (if not, one can make the first term go towards $+\infty$ by letting m go towards either $+\infty$ or $-\infty$). It is now proved that in order for $\rho^*(v)<+\infty$, $\sum_{i=1}^n v_i=-1$ must hold.

Again, consider an arbitrary, but fixed $X\in\mathbb{X}, X\geq 0$ (here, $X\geq 0$ means component-wise). Then, for all $\lambda\geq 0$, we have $\lambda X\geq 0$, and $\lambda X\in\mathbb{X}$, and hence $\rho(\lambda X)\leq\rho(0)$, from the monotonicity of ρ (again, see Definition 2.1.1). Therefore, by the same type of arguments as above

$$\rho^*(v)\geq\sup_{\lambda\geq 0}\{v\cdot\lambda X-\rho(\lambda X)\}\geq\sup_{\lambda\geq 0}\{v\cdot(\lambda X)\}-\rho(0).$$

Here, $\rho^*(v)$ is only finite if $v\cdot X\leq 0$ for all $X\geq 0$, hence $v\leq 0$.

We then get that the conjugate ρ^* is reduced to

$$\rho^*(v)=\begin{cases}\sup_{X\in\mathbb{X}}\{v\cdot X-\rho(X)\} \text{ where } v\cdot\mathbf{1}=-1 \text{ and } v\leq 0\\ +\infty \text{ otherwise .}\end{cases}$$

Now, define $\alpha(Q)=\rho^*(-Q)$ for all $Q\in\mathcal{P}$. From Theorem 2.2.1, $\rho=\rho^{**}$. But

$$\begin{aligned} \rho^{**}(X) &= \sup_{v\in V}\{v\cdot X-\rho^*(v)\} \\ &= \sup_{Q\in\mathcal{P}}\{(-Q)\cdot X-\alpha(Q)\} \\ &= \sup_{Q\in\mathcal{P}}\{\sum_{i=1}^n Q_i(-X_i)-\alpha(Q)\} \\ &= \sup_{Q\in\mathcal{P}}\{E_Q[-X]-\alpha(Q)\} \end{aligned}$$

where Q_i, X_i denote the i'th components of the vectors Q, X respectively. Hence $\rho(X)=\rho^{**}(X)=\sup_{Q\in\mathcal{P}}\{E_Q[-X]-\alpha(Q)\}$. □

Theorem 2.2.2 says that any convex risk measure $\rho:\mathbb{R}^n\to\mathbb{R}$ is the expected value of the negative of a contingent claim, $-X$, minus a penalty function, $\alpha(\cdot)$, under the worst case probability. Note that we consider the expectation of $-X$, not X, since losses are negative in our context.

We already know that the penalty function α in Theorem 2.2.2 is of the form $\alpha(Q)=\rho^*(-Q)$. Actually, Luthi and Doege [14] proved that it is possible to derive a more intuitive representation of α (see Corollary 2.5 in [14]).

Theorem 2.2.3 *Let $\rho : \mathbb{R}^n \to \mathbb{R}$ be a convex risk measure, and let $\mathcal{A}_\rho$ be its acceptance set (in the sense of Definition 2.1.4). Then, Theorem 2.2.2 implies that $\rho(X) = \sup_{Q \in \mathcal{P}}\{E_Q[-X] - \alpha(Q)\}$, where $\alpha : \mathcal{P} \to \mathbb{R}$ is a penalty function. Then, α is of the form*

$$\alpha(Q) = \sup_{X \in \mathcal{A}_\rho} \{E_Q[-X]\}.$$

Proof: It suffices to prove that for all $Q \in \mathcal{P}$,

$$\rho^*(-Q) = \sup_{X \in \mathbb{X}}\{E_Q[-X] - \rho(X)\} = \sup_{X \in \mathcal{A}_\rho} \{E_Q[-X]\} \tag{2.4}$$

since we know that $\alpha(Q) = \rho^*(-Q)$. For all $X \in \mathcal{A}_\rho$, $\rho(X) \leq 0$ (see Definition 2.1.4), so $E_Q[-X] - \rho(X) \geq E_Q[-X]$. Hence, since $\mathcal{A}_\rho \subseteq \mathbb{X}$

$$\rho^*(-Q) \geq \sup_{X \in \mathcal{A}_\rho} \{E_Q[X] - \rho(X)\} \geq \sup_{X \in \mathcal{A}_\rho} \{E_Q[-X]\}.$$

To prove the opposite inequality, and hence to prove equation (2.4), assume for contradiction that there exists $Q \in \mathcal{P}$ such that $\rho^*(-Q) > \sup_{X \in \mathcal{A}_\rho}\{E_Q[-X]\}$. From the definition of supremum, there exists a $Y \in \mathbb{X}$ such that

$$E_Q[-Y] - \rho(Y) > E_Q[-X] \text{ for all } X \in \mathcal{A}_\rho.$$

Note that $Y + \rho(Y)\mathbf{1} \in \mathcal{A}_\rho$ since $\rho(Y + \rho(Y)\mathbf{1}) = \rho(Y) - \rho(Y) = 0$. Therefore $E_Q[-Y] - \rho(Y) > E_Q[-(Y + \rho(Y)\mathbf{1})] = E_Q(-Y) + \rho(Y)E_Q[-\mathbf{1}] = E_Q(-Y) - \rho(Y)$, which is a contradiction. Hence, the result is proved. $\square$

Together, Theorem 2.2.2 and Theorem 2.2.3 provide a good understanding of convex risk measures in $\mathbb{R}^n$: Any convex risk measure $\rho : \mathbb{R}^n \to \mathbb{R}$ can be written in the form $\rho(X) = \sup_{Q \in \mathcal{P}}\{E_Q[-X] - \alpha(Q)\}$, where $\alpha(Q) = \sup_{X \in \mathcal{A}_\rho}\{E_Q[-X]\}$ and $\mathcal{A}_\rho$ is the acceptance set of ρ.

2.2.2 A little measure theory

We would like to prove a dual representation theorem for convex risk measures in the case where Ω is infinite as well. In order to do so, we need some concepts from measure theory. For completeness, we include the required definitions here. However, we refer the reader to Shilling [23] for a thorough introduction.

First, we consider the Jordan decomposition theorem, which says that any signed measure can be uniquely decomposed into a positive and a negative part:

Theorem 2.2.4 (Jordan decomposition theorem) *Every signed measure μ has a unique decomposition into a difference:*

$$\mu = \mu_+ - \mu_-$$

of two positive measures, μ_+ and μ_- where at least one of these two measures is finite. We say that μ_+ is the positive part, and μ_- is the negative part, of μ, respectively.

We also need to define absolutely continuous measures.

Definition 2.2.5 (Absolutely continuous measure) *Let μ, ν be two measures on the measurable space $(\Omega, \mathcal{F})$. If*

$$F \in \mathcal{F}, \mu(F) = 0 \implies \nu(F) = 0,$$

we say that ν is absolutely continuous *w.r.t. μ and write $\nu << \mu$.*

We also say that μ *dominates* ν. There is a very useful way of representing an absolutely continuous measure via its dominating measure. This representation is called the Radon-Nikodým theorem. In order to state this theorem, we need the concept of σ-finite measures.

Definition 2.2.6 (σ-finite measure) *Let μ be a measure on the measurable space $(\Omega, \mathcal{F})$. Then, μ is called* σ-finite *if the set Ω can be covered with at most countably many measurable sets with finite measure, i.e. there exists $F_1, F_2, \ldots \in \mathcal{F}$ with $\mu(A_i) < \infty$ for all $i \in \mathbb{N}$ such that*

$$\bigcup_{i \in \mathbb{N}} F_i = \Omega.$$

Now, we are ready to state the important Radon-Nikodým theorem. For a proof of this theorem, we refer to Shilling [23].

Theorem 2.2.7 (Radon-Nikodým theorem) *Let μ, ν be two measures on the measurable space $(\Omega, \mathcal{F})$. If μ is σ-finite, then the following are equivalent:*

- *$\nu(F) = \int_F f(x)\mu(dx)$ for some almost everywhere unique, non-negative measurable function f.*
- *$\nu << \mu$.*

The unique function f is called the Radon-Nikodým derivative *and is often denoted by $f = d\nu/d\mu$*

With these results from measure theory at hand, we are ready to generalize the dual representation of convex risk measures to the infinite dimensional case.

2.2.3 The infinite dimensional case

How about infinite-dimensional spaces? Can a similar representation of ρ be derived? This is partially answered in the following Theorem 2.2.8, which is Theorem 2.2 in Ruszczynski and Shapiro [21], modified slightly to our setting.

First, let's introduce the setting. Let $(\Omega, \mathcal{F})$ be a measurable space and let $\bar{\mathbb{V}}$ be the vector space of all finite signed measures on $(\Omega, \mathcal{F})$. For each $v \in \bar{\mathbb{V}}$ we know that there exists a Jordan decomposition of v, so $v = v^+ - v^-$ (see Shilling [23]). Let $|v| = v^+ + v^-$. Let $\mathbb{X}$ be a vector space of measurable functions

$X : \Omega \to \mathbb{R}$. Also, let $\mathbb{X}_+ = \{X \in \mathbb{X} : X(\omega) \geq 0 \; \forall \; \omega \in \Omega\}$. This gives a partial order relation on X, so for $X, Y \in \mathbb{X}$, $X \leq Y$ means that $Y - X \in \mathbb{X}_+$.

Let $V \subseteq \bar{\mathbb{V}}$ be the measures $v \in \mathbb{V}$ such that $\int_\Omega |X(\omega)| d|v| < +\infty$ for all $X \in \mathbb{X}$. V is a vector space because of uniqueness of the Jordan decomposition and linearity of integrals. For example: If $v, w \in V$ then $|v+w| = (v+w)^+ + (v+w)^- = (v^+ + w^+) + (v^- + w^-) = |v| + |w|$ by uniqueness of the decomposition, hence $\int_\Omega |X| d|v+w| = \int_\Omega |X| d|v| + \int_\Omega |X| d|w| < +\infty$. Define the pairing $\langle X, v \rangle = \int_\Omega X(\omega) dv(\omega)$. Let $V_- \subseteq V$ be the non-positive measures in V and let $\mathcal{P}$ be the set of probability measures in V.

Assume the following:

$$\text{(A): If } v \notin V_- = \{v \in V : v \leq 0\}, \text{ then there exists an } X' \in \mathbb{X}_+ \text{ such that } \langle X', v \rangle > 0.$$

Now, let $\mathbb{X}$ and V have topologies so that they become paired spaces under the pairing $\langle \cdot, \cdot \rangle$.

For example, let $\mathbb{X} = L^p(\Omega, \mathcal{F}, P)$ where P is a measure, and let V be as above. Each signed measure $v \in V$ can be decomposed so that $v = v_P + v'$, where v_P is absolutely continuous with respect to P (i.e. $P(A) = 0 \Rightarrow v_P(A) = 0$). Then, $dv_P = M dP$, where $M : \Omega \to \mathbb{R}$ is the Radon-Nikodym density of v w.r.t. P. Look at $V' := \{v \in V : \int_\Omega |X(\omega)| d|v_P| < +\infty\} \subseteq V$. This is a vector space for the same reasons that V is a vector space. Then, any signed measure $v \in V'$ can be identified by the Radon-Nikodym derivative of v_P w.r.t. P, that is by M. Actually, $M \in L^q(\Omega, \mathcal{F}, P)$, where $\frac{1}{p} + \frac{1}{q} = 1$, because $\int_\Omega |M|^q dP = \int_\Omega |M|^{q-1} d|v_P| < +\infty$. Hence, each signed measure $v \in V'$ is identified in L^q by its Radon-Nikodym density with respect to P.

Note that the pairing defined above actually is the usual bilinear form between L^p and L^q since for $\bar{p} \in L^p, \bar{q} \in L^q$

$$\begin{aligned} \langle \bar{p}, \bar{q} \rangle &= \int_\Omega \bar{p}(\omega) \bar{q}(\omega) dP(\omega) \\ &= \int_\Omega \bar{p}(\omega) M(\omega) dP(\omega) \\ &= \int_\Omega \bar{p}(\omega) dv(\omega) \end{aligned} \tag{2.5}$$

where the second equality follows from that any $\bar{q} \in L^q$ can be viewed as a Radon-Nikodym derivative w.r.t. P for some signed measure $v \in V'$ and the third equality from the definition of a Radon-Nikodym derivative.

In the following theorem monotonicity and translation invariance mean the same as in Definition 2.1.1.

Theorem 2.2.8 *Let $\mathbb{X}$ be a vector space paired with the space V, both of the form above. Let $\rho : \mathbb{X} \to \mathbb{R}$ be a proper, lower semi-continuous, convex function. From Theorem 1.7.9 the following holds: $\rho(X) = \sup\{\langle X, v \rangle - \rho^*(v) : v \in \mathrm{dom}(\rho^*)\}$. Then,* (*i*) *$\rho$ is monotone* $\iff$ *All $v \in \mathrm{dom}(\rho^*)$ are such that $v \leq 0$*

(*ii*) *ρ is translation invariant* $\iff$ *$v(\Omega) = -1$ for all $v \in \mathrm{dom}(\rho^*)$.*

Hence, if ρ is a convex risk measure (so monotonicity and translation invariance hold), then $v \in \mathrm{dom}(\rho^*)$ implies that $Q := -v \in \mathcal{P}$ and

$$\begin{aligned}\rho(X) &= \sup_{Q\in\mathcal{P}}\{\langle X, -Q\rangle - \rho^*(-Q)\} \\ &= \sup_{Q\in\mathcal{P}}\{\langle -X, Q\rangle - \rho^*(-Q)\} \\ &= \sup_{Q\in\mathcal{P}}\{E_Q[-X] - \alpha(Q)\}\end{aligned}$$

where $\alpha(Q) := \rho^*(-Q)$ is a penalty function and the pairing, i.e. the integral, is viewed as an expectation.

Proof:

(*i*) : Assume monotonicity of ρ. We want to show that $\rho^*(v) = +\infty$ for all $v \notin V_-$. From assumption (A), $v \notin V_- \Rightarrow$ there exists $X' \in X_+$ such that $\langle X', v\rangle > 0$. Take $X \in \mathrm{dom}(\rho)$, so that $\rho(X) < +\infty$ and consider $Y_m := X + mX'$. For $m \geq 0$, monotonicity implies that $\rho(X) \geq \rho(Y_m)$ (since $Y_m = X + mX' \geq X$ because $X' \geq 0$). Hence

$$\begin{aligned}\rho^*(v) &\geq \sup_{m\in\mathbb{R}_+}\{\langle Y_m, v\rangle - \rho(Y_m)\} \\ &= \sup_{m\in\mathbb{R}_+}\{\langle X, v\rangle + m\langle X', v\rangle - \rho(X + mX')\} \\ &\geq \sup_{m\in\mathbb{R}_+}\{\langle X, v\rangle + m\langle X', v\rangle - \rho(X)\}\end{aligned}$$

where the last inequality uses the monotonicity. But since $\langle X', v\rangle > 0$, by letting $m \to +\infty$, one gets $\rho^*(v) = +\infty$ (since $X \in \mathrm{dom}(\rho)$, so $\rho(X) < +\infty$, and $\langle X, v\rangle$ is bounded since $\langle X, \cdot\rangle$ and $\langle \cdot, v\rangle$ are bounded linear functionals).

Hence, monotonicity implies that $\rho^*(v) = +\infty$, unless $v \leq 0$, so all $v \in \mathrm{dom}(\rho^*)$ are such that $v \leq 0$.

Conversely, assume that all $v \in \mathrm{dom}(\rho^*)$ are such that $v \leq 0$. Take $X, Y \in \mathbb{X}$ such that $Y \leq X$ (i.e. $X - Y \geq 0$). Then $\langle Y, v\rangle \geq \langle X, v\rangle$ (from the linearity of the pairing). Since $\rho(X) = \sup_{v\in\mathrm{dom}(\rho^*)}\{\langle X, v\rangle - \rho^*(v)\}$, it follows that $\rho(X) \leq \rho(Y)$. Hence (i) is proved.

(*ii*) : Assume translation invariance. Let $\mathbf{1} : \Omega \to \mathbb{R}$ denote the random variable constantly equal to 1, so $\mathbf{1}(\omega) = 1\ \forall\ \omega \in \Omega$. This random variable is clearly measurable, so $\mathbf{1} \in \mathbb{X}$. For $X \in \mathrm{dom}(\rho)$

$$\begin{aligned} \rho^*(v) &\geq \sup_{m \in \mathbb{R}} \{\langle X + m\mathbf{1}, v\rangle - \rho(X + m\mathbf{1})\} \\ &= \sup_{m \in \mathbb{R}} \{m\langle \mathbf{1}, v\rangle + \langle X, v\rangle - \rho(X) + m\} \\ &= \sup_{m \in \mathbb{R}} \{m(v(\Omega) + 1) + \langle X, v\rangle - \rho(X)\}. \end{aligned}$$

Hence, $\rho^*(v) = +\infty$, unless $v(\Omega) = \langle \mathbf{1}, v\rangle = -1$.

Conversely, if $v(\Omega) = -1$, then $\langle X + m\mathbf{1}, v\rangle = \langle X, v\rangle + v(\Omega)m = \langle X, v\rangle - m$. (where the first equality follows from linearity of the pairing). Hence, translation invariance follows from $\rho(X) = \sup_{v \in \mathrm{dom}(\rho^*)}\{\langle X, v\rangle - \rho^*(v)\}$.

□

Föllmer and Schied [7] proved a version of Theorem 2.2.8 for $\mathbb{X} = L^\infty(\Omega, \mathcal{F}, P)$, $V = L^1(\Omega, \mathcal{F}, P)$. In this case, it is sufficient to assume that the acceptance set $\mathcal{A}_\rho$ of ρ is weak*-closed (i.e., closed with respect to the coarsest topology that makes all the linear functionals originating from the inner product, $\langle \cdot, v\rangle$ continuous) in order to derive a representation of ρ as above.

2.3 Two commonly used measures of financial risk

In this section, we present two measures of monetary risk which are frequently used in practice. One of these measures, called value at risk, is not coherent, or even convex in general. One may (rightfully so!) wonder why risk professionals use measures not satisfying the economically reasonable conditions of Definition 2.1.1 and 2.1.2. There are several reasons for this:

- *Old habits die hard:* These measures were used before the concepts coherent and convex risk measures were introduced. Hence, people are so used to the old measures that they are hesitant to implement others.
- *Simplicity:* As we will see in Section 2.3.2, value at risk is a very intuitive concept.
- *Good enough:* In many practical situations, the results attained are sufficient, though the measures in question are economically unreasonable in theory.

In the next subsection, we define value at risk and look at its limitations as a measure of financial risk.

2.3.1 Value at risk

Value at risk, VaR, is the most commonly used risk measure in practice. For a given portfolio, time horizon and probability λ, VaR is the maximum potential loss over the time period after excluding the λ worst cases. Like in Section 2.1, let X be a random variable representing a financial position. Note that X may represent one stock, a portfolio of stocks or the financial holdings of an entire firm. As before, negative values of $X(\omega)$ correspond to losses, and positive values to profit. Mathematically, VaR is defined as follows: Fix some level $\lambda \in (0,1)$ (typically close to 0), and define $Y := -X$. Note that for the random variable Y, losses are positive numbers. Then, VaR_λ is defined as the $(1-\lambda)$-quantile of Y

$$\mathrm{VaR}_\lambda(X) := F_Y^{-1}(1-\lambda) \tag{2.6}$$

where F_Y is the cumulative distribution function of Y. One can show (this is left as an exercise) that

$$\mathrm{VaR}_\lambda(X) = -\inf\{x \in \mathbb{R} : F_X(x) > \lambda\}. \tag{2.7}$$

That is, VaR is the maximum potential loss over the time period after excluding the λ worst cases.

Lemma 2.3.1 *$VaR_\lambda(X)$ is decreasing in λ.*

Proof: Left as an exercise to the reader. □

$$\mathrm{VaR}_\lambda(X) := \inf\{m \in \mathbb{R} | P(X+m<0) \leq \lambda\}, \tag{2.8}$$

i.e., $\mathrm{VaR}_\lambda(X)$ is the smallest amount of money that needs to be added to X in order for the probability of a loss to be less than λ.

Financial firms and banks often use VaR to quantify the risk of their investments. This allows the firms to monitor their current risk at any time, and hence measure their potential losses. In practise, firms will specify their VaR depending on the confidence level λ, but also depending on some time horizon. In practice, the definition of VaR (2.6) is typically not used directly for computing the value at risk, since this formula requires that we know the exact distribution of X. Instead, most banks, insurance firms etc. use historical data as an approximation to the exact distribution, and compute the quantile in (2.6) based on this. As an alternative, some firms use Monte Carlo methods based on a stochastic model of the financial markets. Monte Carlo methods are based on random sampling from the stochastic model. The Monte Carlo approach is more time-consuming, and usually involves additional work by an analyst in order to fit the parameters of the model to the relevant problem based on historical data. However, a drawback with the historical data method is that this technique implicitly assumes that the future distribution will be the same as the past one, no further randomness or adaptation to the general economic situation is included.

Note that in our mathematical definition of VaR above, we didn't mention time at all. However, for a practical interpretation of VaR, one should think of our random variable X as the profit/loss random variable for the financial position. So, if a bank wants to compute their one-month VaR, X is the (uncertain) difference between the current value of the banks financial holdings and the value a month from now. So how is VaR used and interpreted in practice? Say a portfolio of stocks has a one-day 2% VaR of NOK 10 million. Then, there is a 0.02 probability that the value of the portfolio will decrease by more than NOK 10 million during this day, assuming no trading. On average, the bank will expect to lose more than this 1 out of 50 days. Note that it is very important for the VaR calculation that there is no trading happening in the portfolio. If there is trading, the distribution of the portfolio will change, and hence also the VaR.

Despite its frequent use in practice, value at risk has some major drawbacks as a risk measure:

- In general, VaR_λ is not convex, see Föllmer and Knispel [6]. This means that diversification may increase the risk w.r.t. VaR, which is economically unreasonable.

- In addition, we see from equation (2.6) that VaR_λ ignores extreme losses which occur with small probability. This tail insensitivity makes it an unsuitable measure of risk in situations where the consequences of large losses are very bad.

These drawbacks of VaR as a risk measure is what lead to the development of the theory of convex and coherent risk measures. Nevertheless, value at risk is still widely used in practice, despite its deficiencies.

2.3.2 Average value at risk

Average value at risk (AVaR), also called *expected shortfall* (ES) or *conditional value at risk* (CVaR), was introduced to mend the deficiencies of value at risk. For $\lambda \in (0, 1]$, the average value at risk is defined as

$$\mathrm{AVaR}_\lambda(X) := \frac{1}{\lambda}\int_0^\lambda \mathrm{VaR}_\alpha(X)d\alpha. \tag{2.9}$$

Hence, average value at risk can be interpreted as the expected loss in a presupposed percentage of worst cases.
Note that

$$\mathrm{AVaR}_\lambda(X) \geq \mathrm{VaR}_\lambda(X),$$

(the proof is left as an exercise). So, when considering the same level λ, the average value at risk is always greater than or equal the value at risk.

Föllmer and Schied [9] prove that AVaR_λ is a coherent risk measure, with a dual representation

$$\mathrm{AVaR}_\lambda(X) = \max_{Q \in \mathcal{Q}_\lambda} E_Q[-X]$$

where $\mathcal{Q}_\lambda := \{Q << P | \frac{dQ}{dP} \leq \lambda\}$. That is, $\mathcal{Q}_\lambda$, is the set of all measures Q that are absolutely continuous w.r.t. the measure P given that the Radon-Nikodym derivative of Q w.r.t. P is less than or equal λ (see Shilling [23] for more on these measure theoretical concepts). Note also that for $\lambda = 1$, average value at risk reduces to $E_P[-X]$, i.e., the expected loss.

Other examples of convex risk measures are *shortfall risk* and *divergence risk measures*, but are beyond the scope of this chapter. We refer the interested reader to Föllmer and Knispel [6].

2.4 Exercises to Chapter 2

Exercise 2.1: Prove item 2. of Theorem 2.1.3.

Exercise 2.2: Prove items (*iii*) and (*iv*) of Theorem 2.1.7.

Exercise 2.3: Show the alternative representation of VaR in equation (2.7).

Make sure you understand the interpretation of this expression. Hint: Make a figure of the distribution.

Exercise 2.4: Prove Lemma 2.3.1.

Exercise 2.5: Prove that

$$\mathrm{AVaR}_\lambda(X) \geq \mathrm{VaR}_\lambda(X).$$

Chapter 3

Solutions to selected exercises

3.1 Solutions to exercises from Chapter 1

Solution to exercise 1.1:

Proof:
Follows from the definitions of convex set, conv($\cdot$), intersection, Cartesian product, interior, relative interior and closure. Statement (i) also uses the fact that any convex set must contain all convex combinations of its elements. This can be proved by induction, using that C is convex and that a convex combination of convex combinations is also a convex combination. $\square$

Solution to exercise 1.2:

3. Use Definition 1.1.17 and the mean value inequality, see for example *Kalkulus* by Lindstrøm [12], or any other basic calculus book.

4. Follows from Definition 1.1.17 by induction, and the fact that a convex combination of convex combinations is a convex combination.

5. Use Definition 1.1.17 and induction.

Solution to exercise 1.6:

Proof: One can rewrite the saddle point condition (1.21) as

$$f(x') = K(x', y') = g(y').$$

The theorem then follows from equation (1.20). $\square$

3.2 Solutions to exercises from Chapter 2

Solution to exercise 2.1:

Proof:

2. Again, we check Definition 2.1.1 for any $X, Y \in \mathbb{X}$:

(i) : If $0 \leq \lambda \leq 1$,

$$\begin{aligned}\rho(\lambda X + (1-\lambda)Y) &= \max\{\rho_1(\lambda X + (1-\lambda)Y), \\ &\quad \dots, \rho_n(\lambda X + (1-\lambda)Y)\} \\ &\leq \max\{\lambda\rho_1(X) + (1-\lambda)\rho_1(Y), \\ &\quad \dots, \lambda\rho_n(X) + (1-\lambda)\rho_n(Y)\} \\ &\leq \lambda \max\{\rho_1(X), \dots, \rho_n(X)\} \\ &\quad +(1-\lambda)\max\{\rho_1(Y), \dots, \rho_n(Y)\} \\ &= \lambda\rho(X) + (1-\lambda)\rho(Y).\end{aligned}$$

(ii) : If $X \leq Y$,

$$\begin{aligned}\rho(X) &= \max\{\rho_1(X), \dots, \rho_n(X)\} \\ &\geq \max\{\rho_1(Y), \dots, \rho_n(Y)\} \\ &= \rho(Y).\end{aligned}$$

(iii) : For $m \in \mathbb{R}$,

$$\begin{aligned}\rho(X+m) &= \max\{\rho_1(X+m), \dots, \rho_n(X+m)\} \\ &= \max\{\rho_1(X) - m, \dots, \rho_n(X) - m\} \\ &= \max\{\rho_1(X), \dots, \rho_n(X)\} - m \\ &= \rho(X) - m.\end{aligned}$$

□

Solution to exercise 2.2:

Proof:

(iii) Since $X \in \mathcal{A}_\rho$, $\rho(X) \leq 0$, but because $Y \in \mathbb{X}$ is such that $X \leq Y$, $\rho(Y) \leq \rho(X)$ (from the definition of a convex risk measure). Hence

$$\rho(Y) \leq \rho(X) \leq 0.$$

So $Y \in \mathcal{A}_\rho$ (from the definition of an acceptance set).

(iv) Let ρ be a coherent risk measure, and let $X, Y \in \mathcal{A}_\rho$ and $\alpha, \beta \geq 0$. Then, from the positive homogeneity and subadditivity of coherent risk measures (see Definition 2.1.2), in addition to the definition of $\mathcal{A}_\rho$

$$\rho(\alpha X + \beta Y) \leq \alpha\rho(X) + \beta\rho(Y) \leq \alpha \cdot 0 + \beta \cdot 0 = 0.$$

Hence $\alpha X + \beta Y \in \mathcal{A}_\rho$, so $\mathcal{A}_\rho$ is a convex cone (from the definition of a convex cone).

□

Solution to exercise 2.3:

Note that

$$\mathrm{AVaR}_\lambda(X) \geq \frac{1}{\lambda}\int_0^\lambda \mathrm{VaR}_\lambda(X)d\alpha = \mathrm{VaR}_\lambda(X),$$

where the first equality follows because VaR_λ is decreasing in λ.

Bibliography

[1] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[2] Nedic A. Bertsekas, D. P and A. E. Ozdaglar. *Convex analysis and optimization.* Athena Scientific, Belmont, 2003.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, Cambridge, 2009.

[4] G. Dahl. An introduction to convexity, 2000.

[5] K. R. Dahl, Convex duality and mathematical finance, master thesis, University of Oslo, 2012, https://www.duo.uio.no/bitstream/handle/10852/10746/KristinaRognlienDahlThesis.pdf?sequence=3

[6] Hans Föllmer and Thomas Knispel. Convex risk measures: Basic facts, law-invariance and beyond, asymptotics for large portfolios. In *Handbook of the Fundamentals of Financial Decision Making: Part II*, pages 507–554. World Scientific, 2013.

[7] Hans Föllmer and Alexander Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6, 2002.

[8] Hans Föllmer and Alexander Schied. Robust preferences and convex measures of risk. *Advances in Finance and Stochastics*, 2002.

[9] Hans Föllmer and Alexander Schied. *Stochastic Finance. An Introduction in Discrete Time.* De Gruyter, Berlin, 2008.

[10] Marco Frittelli and Emanuela Rosazza Gianin. Putting order in risk measures. *Journal of Banking and Finance*, 26, 2002.

[11] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I.* Springer, Berlin Heidelberg, 1993.

[12] Tom Lindstrøm. *Kalkulus.* Universitetsforlaget, Oslo, 2006.

[13] D. G. Luenberger. Convex programming and duality in normed space. *IEEE Transactions and Systems Science and Cybernetics*, 2, 1968.

[14] Hans-Jakob Luthi and Jörg Doege. Convex risk measures for portfolio optimization and concepts of flexibility. *Mathematical Programming*, 104, 2005.

[15] James Munkres. *Topology.* Prentice Hall, Upper Saddle River, 2000.

[16] Gert Pedersen. *Analysis Now.* Springer, New York, 2002.

[17] Ralph Tyrell Rockafellar. *Convex analysis.* Princeton University Press, Princeton, 1972.

[18] Ralph Tyrell Rockafellar. *Conjugate Duality and Optimization.* SIAM, Philadelphia, 1974.

[19] Ralph Tyrell Rockafellar. *Coherent approaches to risk in optimization under uncertainty.* INFORMS, 2007.

[20] Ralph Tyrell Rockafellar and Roger Jean-Baptiste Robert Wets. *Variational analysis.* Springer, Berlin, 2004.

[21] Andrzej Ruszczynski and Alexander Shapiro. Optimization of convex risk functions. *Mathematical of Operations Research*, 31, 2006.

[22] Bryan Rynne and Martin Youngston. *Linear functional analysis.* Springer, London, 2008.

[23] Réne Schilling. *Measures, integrals and martingales.* Cambridge University Press, Cambridge, 2005.

[24] R. J. Vanderbei. *Linear Programming: Foundations and Extensions.* Springer, Berlin Heidelberg, 2008.

[25] Gunter Matthias Ziegler. *Lectures on polytopes.* Springer, Heidelberg, 1995.