Universidade de São Paulo
Instituto de Matemática e Estatística
Bachalerado em Matemática Aplicada

Ariel Serranoni Soares da Silva

Relative Entropy Optimization and Applications in Statistical Learning

São Paulo
Dezembro de 2018

Relative Entropy Optimization and Applications in Statistical Learning

Monografia final da disciplina
MAP2080 – Trabalho de Formatura.

Supervisor: Prof. Dr. Marcel K. de Carli Silva

São Paulo
Dezembro de 2018

Agradecimentos

Minha eterna gratidão ao Marcel, pela sua paciência infinita e por testar todos os meus limites. Físico, emocional, e principalmente intelectual. À minha família, pela certeza de nunca estar sozinho. Aos meus amigos, pela alegria que trazem até aos dias mais cinzentos, e ao meu bebê Leo por todos os anos de amor incondicional.

Resumo

Otimização cônica é o principal objeto de estudo deste trabalho, que é composto de cinco capítulos. O Capítulo 1 é dedicado a familiarizar o leitor com alguns conceitos básicos que permeiam essa área. O Capítulo 2 apresenta otimização cônica de maneira geral e inclui a prova do resultado mais importante sobre o tema, o Teorema de Dualidade Forte. No Terceiro capítulo, apresentamos os cones de entropia relativa e os programas cônicos definidos sobre eles (REP), além de uma aplicação simples em estimação de distribuições discretas de probabilidade. No Capítulo 4, estudamos brevemente programas sobre o famoso cone de segunda ordem (SOCP), mostramos como esses programas podem ser formulados como REP e então introduzimos a regressão ridge como uma aplicação de SOCP em aprendizado estatístico. No Capítulo 5, definimos programas geométricos (GP), mostramos que esses problemas de otimização também podem ser formulados como REP e então apresentamos regressão logística como nossa aplicação de GP.

Palavras-chave: Otimização, Convexidade, Probabilidade.

Abstract

Conic optimization is the main matter of this text, which consists of five chapters. Chapter 1 is intended to familiarize the reader with the the basic concepts that pervade this area. Chapter 2 presents a general panorama of the the theory of conic programming and show the main result about this topic, the Strong Duality Theorem. In Chapter 3, we present the relative entropy cone, the conic programs defined over them (REP), and a simple application on discrete density estimation. In Chapter 4, we briefly introduce optimization problems over the famous second-order cone (SOCP), show how to cast a SOCP as a REP, and present ridge regression as an application of SOCP to statistical learning. In Chapter 5, we define geometric programs, show that these may also be formulated as REPs, and finally, display logistic regression as our application of GP.

Keywords: Optimization, Convexity, Probability.

Contents

Motivation

Optimization is the area of mathematics which studies the task of choosing a *best* solution over a set of *possible* solutions. The generality of this problem is demonstrated by the range of applications, that go from choosing the best route on your daily comute passing through physics, economics, and evolutionary biology topics. In the latter, the theory of the selfish gene [18] suggests optimization as the purpose of genes when trying to perpetuate themselves. This presents optimization trespassing the border of a human activity and putting it in the context of nature as a whole.

In [17], the study of mathematical programming is divided in three periods. In the first of them, one was only able to find extremum points and values of some sorts of functions, such as polynomials of second degree. The second period was initiated in 1646 by Fermat's work regarding extremum points of differentiable functions. Still in this period, the theory of Lagrange Multipliers first introduced constraints in optimization. However, optimization did not receive much atention until 1947, when Dantzig came up with the simplex method to solve linear programs. Thereafter, mathematical programming became increasingly popular. In the fifties, the work of Khun and Tucker concerning optimality conditions is considered the birth of nonlinear programming. In the same decade the survey of Ford and Fulkerson gave rise to combinatorial optimization and Gomory published his cutting plane method, which is considered to be the genesis of integer programming. Published in 1970, Rockafellar's book [30] is a cornerstone in convex analysis and optimization, which has drawn attention from the mathematical community due to three factors: diversity of its applications, interesting duality theory, and algorithmic efficiency. In this context, conic optimization became well-known and is described in [19] as an elegant framework for convex optimization. Nowadays, this area is heavily ramificated. Thus, it became convenient to investigate if it is possible to apply results and algorithms to multiple branches of optimization.

In parallel, the human race is attached to randomness since the dawn of its existence, and this narrative can also be partitioned in three stages. At first this concept was thought from a qualitative point of view. In the chinese empire, outcomes from games of chance were superstitiously interpretated and even related to destiny. Later in Greece, Aristotle, Epicurus, and Democritus approached the subject on their respective surveys. In the Medieval period, catholic philosophers questioned randomness in opposition to free-will and God's omniscience. The start of the second period was simultaneous with the birth of calculus and came through the work of, e.g, Galileo, Pascal, and Leibnitz. This era reached its apex with the first book on quantitative probability, The Doctrine of Chances, by De Moivre. However, the advances of this area were inhibited by the belief in a deterministic universe, which was dominant at that time. The advances on statistical mechanics, formalized by Gibbs in 1902 officially introduced probability in the field of physics. Also, the axiomatization proposed by Kolmogorov decisively setteled probability theory as a branch of mathematics. Thenceforth, research on this topic and its relatives heavily intensified and its applications pervaded almost all of science.

More recently, in virtue of the constant advances in technology, computers became capable of processing data in large scale. As a consequece, various research areas concernig the intersection of statistics and computer science - which encompass beautiful mathematical results, interesting applications, and philosophical depth - have been emerging. These mentioned factors make the study of a fraction of this intersection highly attractive. This text is an attempt to bring to the reader some of my enthusiasm in respect to this subject.

Chapter 1

Preliminaries

1.1 Basic Notation, Definitions, and Results

Before we start the mathematical discussion prepared for this text, it is essential to establish a common vocabulary of words and symbols in order to avoid confusion. We use mostly the standard mathematical notation and definitions. Thus, the experienced reader may feel encouraged skip this section and refer back to it when needed.

As usual, we denote the set of all natural numbers, which includes 0, by $\mathbb{N}$, the set of all integer numbers by $\mathbb{Z}$, the set of all rational numbers by $\mathbb{Q}$, and the set of all real numbers by $\mathbb{R}$. We also call the elements of the set $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ the *extended real numbers.*

If $k \in \mathbb{N}$, define $[k] := \{1, 2, \dots, k\}$. If A and B are sets, the set of all the functions from A to B is $B^A := \{f \mid f \colon A \to B\}$. Note that sets such as $\mathbb{R}^n := \mathbb{R}^{[n]}$ violate this definition, but we still adopt this usual notation for convenience. Moreover, the set provided by the case where $B = \mathbb{R}$, $A = [n] \times [n]$, and $f(i, j) = f(j, i)$ for each $i, j \in [n]$ will be denoted by $\mathbb{S}^n$. If $M \in \mathbb{S}^n$, we say that M is *positive-semidefinite* if $x^\top M x \geq 0$ for each $x \in \mathbb{R}^n$. In this case, we denote $M \in \mathbb{S}^n_+$.

Let $n, k \in \mathbb{N}$ and consider $S \subseteq [n]$ and $T \subseteq [k]$. If $A \in \mathbb{R}^{[n] \times [k]}$, then $A[S, T]$ denotes the restriction of A to $S \times T$. Similarly, if $A \in \mathbb{R}^{[n] \times [n]}$ then $A[S]$ is the restriction of A to $S \times S$.

For each $k \in \mathbb{N}$, if $S \in \{\mathbb{N}^k, \mathbb{Z}^k, \mathbb{Q}^k, \mathbb{R}^k\}$ and $x \in S$, we refer to the i-th coordinate of x by $x_i := x(i)$ for every $i \in [k]$. Similarly, if $C \subseteq [k]$, we denote $x_C := \{x_i : i \in C\}$. If $x_i = 1$ for each $i \in [k]$, then $x =: \mathbb{1}$. We define $S_+ := \{s \in S : s \geq 0\}$ and $S_{++} := \{s \in S : s > 0\}$, where we consider $x \geq 0$ if $x_i \geq 0$ for each $i \in [k]$, and analogously for $x > 0$. Also, if S, W are sets, $B \subseteq S$, and $f \in W^S$, then $W \supseteq f(B) := \{f(s) : s \in B\}$. Furthermore, the *pre-image* of $C \subseteq W$ is $f^{-1}(C) := \{s \in S : f(s) \in C\}$.

Let S be a set. A *partition* of S is a collection $\mathcal{C}$ of nonempty pairwise disjoint subsets of S whose union is S. That is, if $A, B \in \mathcal{C}$ are distinct, then $A \cap B = \varnothing$ and $\bigcup \mathcal{C} = S$. In particular, if $A \subseteq S$ then $\{A, S \setminus A\}$ is a partition of S. In this case, the set $A \setminus S$ is called the *complement* of A relative to S and we may denote $S \setminus A =: A^c$. The *cardinality* of S is denoted by $|S|$. If there is $n \in \mathbb{N}$ such that $|S| = n$, then S is *finite*. Otherwise, S is *infinite.*

A *partial order* on S is a binary relation $\leq$ such that, for each $a, b, c \in S$:

(i) $a \leq a$;

(ii) if $a \leq b$ and $b \leq a$, then $a = b$;

(iii) if $a \leq b$ and $b \leq c$, then $a \leq c$.

The properties stated above are called *reflexivity*, *antisymmetry*, and *transitivity*, respectively. If the relation $\leq$ is antisymmetric, transitive and reflexive, and also satisfies

$$a \leq b \text{ or } b \leq a, \text{ for each } a, b \in S;$$

we say that $\leq$ is a *total order* on S. This last property is called *totality.*

An *equivalence relation* on S is a binary relation $\sim$ such that, for each $a, b, c \in S$:

(i) $a \sim a$;

(ii) if $a \sim b$, then $b \sim a$;

(iii) if $a \sim b$ and $b \sim c$, then $a \sim c$.

The second of these properties is called *symmetry*, the others were already presented in the previous definition.

Basic concepts of mathematical analysis such as sequences of various types, limits, continuity, first- and second-order derivatives, and integrals are assumed to be familiar to the reader. In case of necessity, we suggest [29; 32] as useful sources.

Let S be a set and let $f\colon S \to \overline{\mathbb{R}}$ be a function. We define the *epigraph* of f as

$$\operatorname{epi}(f) \coloneqq \{s \oplus \alpha \in S \oplus \mathbb{R} : f(s) \leq \alpha\}.$$

A function $f\colon S \to \overline{\mathbb{R}}$ is *proper* if $f(s) < +\infty$ for some $s \in S$ and $f(x) > -\infty$ for each $x \in S$. This condition is meant to guarantee that $\operatorname{epi}(f) \neq \varnothing$ and that $\operatorname{epi}(f)$ does not contain a line parallel to the $f(x)$-axis. The *effective domain* of f is the set $\{s \in S : f(s) \text{ is finite}\}$ and is denoted by $\operatorname{dom}(f)$. Whenever possible, we omit such terminology.

We say that f is *homogeneous of degree n* if

$$f(\alpha x) = \alpha^n f(s), \text{ for each } s \in S \text{ and } \alpha \in \mathbb{R}.$$

If the latter holds only for $\alpha \in \mathbb{R}_{++}$, then f is *positively homogeneous of degree n*. We will refer to a homogeneous function of degree 1 as homogeneous and analogously for positively homogeneous. We note that every homogeneous function is continuous.

We also assume the reader to have some familiarity with finite-dimensional vector spaces over $\mathbb{R}$. If not, we recommend the reader to refer to the first chapter of [22].

Let V and W be vector spaces over $\mathbb{R}$. A *linear transformation* $T\colon V \to W$ is a function such that, for each $x, y \in V$ and $\alpha, \beta \in \mathbb{R}$,

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y).$$

Moreover, we denote

$$\operatorname{Im}(T) \coloneqq T(V) \text{ and } \operatorname{Null}(T) \coloneqq \{x \in V : T(x) = 0\}.$$

Note that $\operatorname{Im}(T)$ and $\operatorname{Null}(T)$ are linear subspaces of W and V, respectively. The *direct sum* of V and W is the vector space

$$V \oplus W \coloneqq \{(v, w) : v \in V \text{ and } w \in W\}$$

and we will abbreviate $(v, w) \coloneqq v \oplus w$. For each $v_1 \oplus w_1, v_2 \oplus w_2 \in V \oplus W$ and $\alpha \in \mathbb{R}$, we consider

$$v_1 \oplus w_1 + v_2 \oplus w_2 = (v_1 +_V v_2) \oplus (w_1 +_W w_2)$$

and

$$\alpha(v_1 \oplus w_2) = \alpha v_1 \oplus \alpha w_1.$$

Let V be a vector space. An *inner product* on V is a function $\langle \cdot, \cdot \rangle\colon V \times V \to \mathbb{R}$ such that, for each $v, x, z \in V$ and $\alpha, \beta \in \mathbb{R}$:

(i) $\langle v, v\rangle \geq 0$, where equality holds if and only if $v = 0$;

(ii) $\langle v, x\rangle = \langle x, v\rangle$;

(iii) $\langle \alpha v + \beta x, z\rangle = \alpha\langle v, z\rangle + \beta\langle x, z\rangle$.

An *Euclidean space* $\mathbb{E}$ is a finite-dimensional vector space over $\mathbb{R}$ equipped with an inner product.

If $\mathbb{E}$ and $\mathbb{Y}$ are Euclidean spaces, we define, for each $x_1 \oplus y_1, x_2 \oplus y_2 \in \mathbb{E} \oplus \mathbb{Y}$:

$$\langle x_1 \oplus y_1, x_2 \oplus y_2\rangle_{\mathbb{E}\oplus\mathbb{Y}} = \langle x_1, x_2\rangle_{\mathbb{E}} + \langle y_1, y_2\rangle_{\mathbb{Y}}.$$

In order to simplify notation, the domains of inner products and sums are going to be omitted from now on when dealing with direct sums.

If $S \subseteq \mathbb{E}$, we define the *orthogonal complement* of S as the subspace

$$S^{\perp} := \{x \in \mathbb{E} : \langle x, s\rangle = 0 \text{ for each } s \in S\}.$$

If $a \in \mathbb{E} \setminus \{0\}$ and $\beta \in \mathbb{R}$ the set $\{x \in \mathbb{E} : \langle a, x\rangle = \beta\}$ is a *hyperplane*. Similarly, the set $\{x \in \mathbb{E} : \langle a, x\rangle \leq \beta\}$ is a *half space*. If I is a finite index set, $a \in (\mathbb{E} \setminus \{0\})^I$, and $\beta \in \mathbb{R}^I$, then $\bigcap_{i\in I}\{x \in \mathbb{E} : \langle a_i, x\rangle \leq \beta_i\}$ is a *polyhedron*. A function $f\colon \mathbb{E} \to \mathbb{R}$ is *polyhedral* if $\operatorname{epi}(f)$ is a polyhedron.

Let $C \subseteq \mathbb{E}$ be a finite set and consider a function $\lambda\colon C \to \mathbb{R}$. The point $y = \sum_{x\in C} \lambda_x x$ is a *linear combination* of the elements of C and the set of all the linear combinations of elements of C is $\operatorname{span}(C)$. In the cases where $\sum_{x\in C} \lambda_x = 1$ or $\lambda\colon C \to \mathbb{R}_+$, y is an *affine* or *conic* combination of the elements of C, respectively. When both of these conditions hold simultaneously, y is a *convex combination* of the elements of C. If $S \subseteq \mathbb{E}$ is any set, the *affine hull* is the set of all the finite affine combinations of elements of S, and likewise for *conic hull* and *convex hull*. These sets will be denoted by $\operatorname{aff}(S)$, $\operatorname{cone}(S)$, and $\operatorname{conv}(S)$, respectively. The following result establishes that convex combinations are taken using at most $\dim(\mathbb{E})+1$ elements at time. This result can be generalized in several ways, we recommend the interested reader to consult [30] for further information on this topic.

Proposition 1 (Carathéodory)**.** Let $\mathbb{E}$ be an n-dimensional Euclidean space and let $S \subseteq \mathbb{E}$. If $y \in \operatorname{conv}(S)$, then there exists $S' \subseteq S$ with $|S| \leq n+1$ such that y is a convex combination of the elements of S'.

Proof. We have by the definition of a convex hull that there exists a finite $S' \subseteq S$ and $\lambda\colon S' \to \mathbb{R}_+$ such that $\sum_{x\in S} \lambda_x = 1$ and $y = \sum_{x\in S'} \lambda_x x$. Let $S' \subseteq S$ satisfying this property such that $|S'|$ is as small as possible. Assume by contradiction that $|S'| > n + 1$.

Fix $x_0 \in S'$ and note that $0 \in S' - x_0$. Thus, $S' - x_0$ is linearly dependent. This fact implies that there exists $\alpha\colon S' \setminus \{x_0\} \to \mathbb{R}$ not identically zero such that $\sum_{x\in S'\setminus\{x_0\}} \alpha_x(x - x_0) = 0$. Define $\alpha_{x_0} := -\sum_{x\in S'\setminus\{x_0\}} \alpha_x$. Then, we obtain that $\sum_{x\in S'} \alpha_x = 0$ and

$$\sum_{x\in S'} \alpha_x x = \sum_{x\in S'\setminus\{x_0\}} \alpha_x(x - x_0) = 0.$$

It follows that, for each $\beta \in \mathbb{R}$:

$$\begin{aligned} y = y - \beta 0 &= \sum_{x\in S'} \lambda_x x - \beta \sum_{x\in S'} \alpha_x x \\ &= \sum_{x\in S'} (\lambda_x - \beta\alpha_x)x. \end{aligned}$$

Since $\sum_{x\in S'} \alpha_x = 0$ and α is not identically zero, there exists x_1 such that $\alpha_{x_1} > 0$. Thus, we can consider $\beta := \min_{x\in S'}\{\frac{\lambda_x}{\alpha_x} : \alpha_x > 0\}$. Note that $\beta = \frac{\lambda_{x'}}{\alpha_{x'}}$ for some $x' \in S'$. Hence, we have that $\lambda_x - \beta\alpha_x \in \mathbb{R}_+$ for each $x \in S'$, $\sum_{x\in S'} \lambda_x - \beta\alpha_x = 1$, and $\lambda_{x'} - \beta\alpha_{x'} = 0$. Therefore, we have written y as a convex combination of $|S'| - 1$ points, contradicting the minimality of S'. □

Let $S \subseteq \mathbb{E}$. We say that S is *finitely generated* if $S = \operatorname{span}(C)$, $S = \operatorname{conv}(C)$, $S = \operatorname{cone}(C)$, or $S = \operatorname{aff}(C)$ for some finite set $C \subset \mathbb{E}$. As shown in Theorem 19.1 from [30], the set S is finitely generated if, and only if S is a polyhedron. In addition, the Minkowski-Weyl characterization for polyhedra states that a set $P \subseteq \mathbb{E}$ is a polyhedron if and only if $P = \operatorname{conv}(C_1) + \operatorname{cone}(C_2)$ for some finite sets $C_1, C_2 \subseteq \mathbb{E}$. A detailed explanation and proof of this result can be found in [11]. This description of P allows one to express any point $y \in P$ as

$$y = \sum_{x\in C_1} \lambda_x x + \sum_{x\in C_2} \alpha_x x$$

for some $\lambda\colon C_1 \to \mathbb{R}_+$ with $\sum_{x\in C_1} \lambda_x = 1$ and $\alpha\colon C_2 \to \mathbb{R}_+$.

A *norm* on a vector space V is a function $\|\cdot\|\colon V \to \mathbb{R}$ such that for each $x, y \in V$ and $\alpha \in \mathbb{R}$:

(i) $\|x\| \geq 0$, where equality holds if and only if $x = 0$;

(ii) $\|\alpha x\| = |\alpha|\|x\|$;

(iii) $\|x + y\| \leq \|x\| + \|y\|$.

Property (iii) is often referred to as the triangle inequality. The reader should note that if $\mathbb{E}$ is an Euclidean space, then

$$\|x\| := \langle x, x\rangle^{\frac{1}{2}}$$

is a norm on $\mathbb{E}$ and will be adopted as the standard norm of Euclidean spaces. This fact implies that if $\mathbb{E}$ and $\mathbb{Y}$ are Euclidean spaces, then, for each $x \oplus y \in \mathbb{E} \oplus \mathbb{Y}$:

$$\|x \oplus y\| = (\langle x \oplus y, x \oplus y\rangle)^{\frac{1}{2}} = (\langle x, x\rangle + \langle y, y\rangle)^{\frac{1}{2}} = (\|x\|^2 + \|y\|^2)^{\frac{1}{2}}.$$

Proposition 2. Let $\mathbb{E}$ be an Euclidean space and let $x, y \in \mathbb{E}$. Then $\langle x, y\rangle^2 \leq \|x\|^2\|y\|^2$.

Proof. Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(\lambda) := \|x - \lambda y\|$ for each $\lambda \in \mathbb{R}$. First, assume that $y \neq 0$. Then, for each $\lambda \in \mathbb{R}$:

$$\begin{aligned} 0 &\leq \|x - \lambda y\| \\ &= \langle x - \lambda y, x - \lambda y\rangle \\ &= \|x\|^2 - 2\lambda\langle x, y\rangle + \lambda^2\|y\|^2. \end{aligned}$$

Set $\lambda := \frac{\langle x,y\rangle}{\|y\|^2}$. It follows:

$$\begin{aligned} &\|x\|^2 - 2\frac{\langle x, y\rangle^2}{\|y\|^2} + \frac{\langle x, y\rangle^2}{\|y\|^2} \geq 0 \\ \iff & -\frac{\langle x, y\rangle^2}{\|y\|^2} \geq -\|x\|^2 \\ \iff & \|x\|^2\|y\|^2 \geq \langle x, y\rangle^2. \end{aligned}$$

Finally, the reader can easily verify that the inequality also hold when $y = 0$ because $\langle x, 0\rangle^2 = \|x\|0 = 0$. ☐

Theorem 3. Let $\mathbb{E}$ be an Euclidean space. If $f\colon \mathbb{E} \to \mathbb{R}$ is a linear function, then there exists a unique $z \in \mathbb{E}$ such that $f(x) = \langle z, x\rangle$ for each $x \in \mathbb{E}$.

Proof. If $f(x) = 0$ for each $x \in \mathbb{E}$, then let $z = 0$ and we are done. Otherwise, consider $M := \mathrm{Ker}(f)$ and $N := M^\perp$. Since $M \neq \mathbb{E}$, it follows that there exists $0 \neq y \in N$. Then, the linearity of f implies that $y_0 := \frac{y}{\|y\|}$ belongs to N as well. Let $z := f(y_0)y_0$, then

$$f(x) = \langle x, z\rangle, \text{ if } x = y_0 \text{ or } x \in M.$$

For an arbitrary $x \in \mathbb{E}$, define $x_0 := x - \alpha y_0$, where $\alpha = \frac{f(x)}{f(y_0)}$. Thus, it follows that $f(x_0) = 0$ and thus $x_0 \in M$. Applying f to $x = x_0 + \alpha y_0$ yields

$$f(x) = f(x_0 + \alpha y_0) = f(x_0) + \alpha f(y_0) = \langle x_0, z\rangle + \alpha\langle y_0, z\rangle = \langle x_0 + \alpha y_0, z\rangle = \langle x, z\rangle.$$

To prove that z is unique, assume there exist z_1 and z_2 such that $f(x) = \langle x, z_1\rangle = \langle x, z_2\rangle$ for each $x \in \mathbb{E}$. Thus, it follows that $\langle x, z_1 - z_2\rangle = 0$ for each $x \in \mathbb{E}$. In particular, if $x = z_1 - z_2$ then $\langle z_1 - z_2, z_1 - z_2\rangle = \|z_1 - z_2\|^2 = 0$. This implies that $z_1 - z_2 = 0$ and therefore $z_1 = z_2$. ☐

Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces and let $A\colon \mathbb{E} \to \mathbb{Y}$ a linear function. A linear transformation $T\colon \mathbb{Y} \to \mathbb{E}$ is said to be *adjoint* to A if

$$\langle A(x), y\rangle = \langle x, T(y)\rangle \text{ for each } x \in \mathbb{E} \text{ and } y \in \mathbb{Y}.$$

Proposition 4. Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces. If $A\colon \mathbb{E} \to \mathbb{Y}$ is a linear function, then there exists a unique linear function $T\colon \mathbb{Y} \to \mathbb{E}$ adjoint to A.

Proof. The function $f_y\colon \mathbb{E} \to \mathbb{R}$ given by $\langle A(\cdot), y\rangle$ is trivially linear for each $y \in \mathbb{Y}$. Hence, Theorem 3 gives us the existence of a unique $z_y \in \mathbb{E}$ such that $\langle A(x), y\rangle = \langle x, z_y\rangle$ for each $x \in \mathbb{E}$. Consider the function $T\colon \mathbb{Y} \to \mathbb{E}$ given by $T(y) := z_y$ for each $y \in \mathbb{Y}$. From the definition of T,

$$\langle A(x), y\rangle = \langle x, T(y)\rangle, \text{ for each } x \in \mathbb{E} \text{ and } y \in \mathbb{Y}.$$

It remains to show that T is linear. Let $x \in \mathbb{E}$, let $y_1, y_2 \in \mathbb{Y}$, and let $\alpha_1, \alpha_2 \in \mathbb{R}$. Then:

$$\begin{aligned}\langle x, T(\alpha_1 y_1 + \alpha_2 y_2)\rangle &= \langle A(x), \alpha_1 y_1 + \alpha_2 y_2\rangle \\ &= \langle A(\alpha_1 x), y_1\rangle + \langle A(\alpha_2 x), y_2\rangle \\ &= \langle x, \alpha_1 T(y_1) + \alpha_2 T(y_2)\rangle.\end{aligned}$$

Since x is arbitrary, the result follows. ☐

Because we proved that, for every linear function A, the adjoint transformation of A always exists and is unique, we will simply denote the transformation adjoint to A by A^*.

Proposition 5. Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces. If $A\colon \mathbb{E} \to \mathbb{Y}$ is a linear function, then $\mathrm{Null}(A^*)^\perp = \mathrm{Im}(A)$.

Proof. We start showing that $\mathrm{Im}(A) \supseteq \mathrm{Null}(A^*)^\perp$. Note that $\mathrm{Im}(A) \supseteq \mathrm{Null}(A^*)^\perp$ if, and only if $\mathrm{Null}(A^*)^{\perp\perp} = \mathrm{Null}(A^*) \supseteq \mathrm{Im}(A)^\perp$. Thus, it suffices to show that $\mathrm{Null}(A^*) \supseteq \mathrm{Im}(A)^\perp$. Let $z \in \mathrm{Im}(A)^\perp$. For each $x \in \mathbb{E}$, we have that $A(x) \in \mathrm{Im}(A)$. Then,

$$0 = \langle z, A(x)\rangle = \langle A^*(z), x\rangle.$$

Since the equality above holds for each $x \in \mathbb{E}$, we conclude that $A^*(z) = 0$ and then $z \in \operatorname{Null}(A^*)$.

Conversely, let $y \in \operatorname{Im}(A)$. Then, there exists $x \in \mathbb{E}$ such that $y = A(x)$. Hence, for each $z \in \operatorname{Null}(A^*)$,

$$\langle z, y\rangle = \langle z, A(x)\rangle = \langle A^*(z), x\rangle = \langle 0, y\rangle = 0.$$

Therefore, $x \in \operatorname{Null}(A^*)^{\perp}$. □

One important fact that arises from the proposition above is that if A is invertible, then so is A^*. The converse of this corollary can be obtained from our next result.

Proposition 6. Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces. If $A\colon \mathbb{E} \to \mathbb{Y}$ is a linear function, then $(A^*)^* = A$.

Proof. Let $x \in \mathbb{E}$ and $y \in \mathbb{Y}$. By definition,

$$\langle A(x)^{**}, y\rangle = \langle x, A^*(y)\rangle = \langle A(x), y\rangle.$$

Since the latter holds for each $x \in \mathbb{E}$ and $y \in \mathbb{Y}$, the result follows. □

Let $\mathbb{E}$ be an Euclidean space. Define the *unit ball* on $\mathbb{E}$ as $\mathbb{B} := \{x \in \mathbb{E} : \|x\| \leq 1\}$. Also consider $\mathbb{B}_< := \{x \in \mathbb{E} : \|x\| < 1\}$ and $\mathbb{B}_= := \{x \in \mathbb{E} : \|x\| = 1\}$. Note that $\mathbb{B} = \mathbb{B}_< \cup \mathbb{B}_=$. Let $S \subseteq \mathbb{E}$. The set S is *bounded* if there exists $\alpha \in \mathbb{R}_{++}$ such that $S \subseteq \alpha\mathbb{B}$. A point $x \in \mathbb{E}$ is an *accumulation point* of S if for every $\varepsilon \in \mathbb{R}_{++}$ we have that $((x + \varepsilon\mathbb{B}) \setminus \{x\}) \cap S \neq \varnothing$. A point $x \in \mathbb{E}$ is *adherent* to S if for every $\varepsilon \in \mathbb{R}_{++}$ we have that $(x + \varepsilon\mathbb{B}) \cap S \neq \varnothing$. A point $x \in S$ is an *interior point* of S if there exists $\varepsilon \in \mathbb{R}_{++}$ such that $x + \varepsilon\mathbb{B} \subseteq S$. The interior of S is the set $\operatorname{int}(S)$ of the interior points of S. We say that S is *open* if $S = \operatorname{int}(S)$. The set S is *closed* if S^c is open. The *closure* of S is the set $\overline{S}$ of all the points adherent to S and S is closed if and only if it contains all of its accumulation points. The set $\overline{S}$ can be equivalently defined as $\bigcap_{\varepsilon>0}(S + \varepsilon\mathbb{B})$ and S is closed if and only if $S = \overline{S}$. Moreover, S is *compact* if it is closed and bounded.

Next, we present some of the basic topologic properties of sets and functions. We will approach these results within the setting of Euclidean spaces. However, they also hold in a more general context. For a complete panorama of topology, we suggest [27] as a helpful reference.

Proposition 7. Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces, let $f\colon \mathbb{E} \to \mathbb{Y}$ be a continuous function, and let $S \subseteq \mathbb{Y}$. Then:

(i) if S is open, then $f^{-1}(S)$ is open;

(ii) if S is closed, then $f^{-1}(S)$ is closed.

Proof.

(i) Let $x \in f^{-1}(S)$, then we have that $f(x) \in S$. Since S is open, we have that there exists $\varepsilon > 0$ such that $f(x) + \varepsilon\mathbb{B} \subseteq S$ and, since f is continuous, we have the existence of $\delta > 0$ such that $y \in x + \delta\mathbb{B}$ implies $f(y) \in f(x) + \varepsilon\mathbb{B}$. Because $f(x) + \varepsilon\mathbb{B} \subseteq S$, we have that $f(y) \in S$ whenever $y \in x + \delta\mathbb{B}$. Hence, $y \in x + \delta\mathbb{B}$ implies $y \in f^{-1}(S)$.

(ii) Let $S \subseteq \mathbb{Y}$ be a closed set. By definition, we have that S^c is open. From item (i), we know that $f^{-1}(S^c)$ is open. Therefore $f^{-1}(S^c)^c = f^{-1}(S)$ is closed. □

Proposition 8. Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces. If $T\colon \mathbb{E} \to \mathbb{Y}$ is a linear function. Then:

(i) T is continuous;

(ii) if T is invertible, then T^{-1} is linear.

Proof.

(i) Let $x \in \mathbb{E}$ and $\varepsilon > 0$. Set $\delta := \varepsilon(\sup_{x\in\mathbb{B}} \|T(x)\|)^{-1}$. We will show that $T(x+\delta\mathbb{B}) \subseteq T(x)+\varepsilon\mathbb{B}$. Let $y \in x+\delta\mathbb{B}$. Then, there exists $v \in \mathbb{B}$ satisfying $y = x+\delta v$. Since T is linear it follows that $T(y) = T(x)+\delta T(v)$. Hence,

$$\|T(y) - T(x)\| = \|T(x) + \delta T(v) - T(x)\| = \|\delta T(v)\| = \delta\|T(v)\| \leq \varepsilon.$$

Therefore, $T(y) \in T(x)+\varepsilon\mathbb{B}$.

(ii) Let $w_1, w_2 \in \mathbb{Y}$ and $\alpha_1, \alpha_2 \in \mathbb{R}$. It follows:

$$\begin{aligned} T^{-1}(\alpha_1 w_1) + T^{-1}(\alpha_2 w_2) &= T^{-1}(T(T^{-1}(\alpha_1 w_1) + T^{-1}(\alpha_2 w_2))) \\ &= T^{-1}(T(T^{-1}(\alpha_1 w_1)) + T(T^{-1}(\alpha_2 w_2))) \\ &= T^{-1}(\alpha_1 w_1 + \alpha_2 w_2). \end{aligned}$$ □

A function $f\colon \mathbb{E} \to \mathbb{R}$ is *closed* if $\operatorname{epi}(f)$ is closed. The *closure* of f is the function $\overline{f}$ such that $\operatorname{epi}(\overline{f}) = \overline{\operatorname{epi}(f)}$. We note that this definition implies that $\overline{f}(x) \leq f(x)$ for each $x \in \mathbb{E}$. This occurs because $\operatorname{epi}(f) \subseteq \operatorname{epi}(\overline{f})$. The following proposition presents continuity as a sufficient condition on f that ensures its epigraph to be closed. Nevertheless, this demand is not the weakest possible. The interested reader may look at the eighth chapter of [30] for more information about this topic.

Proposition 9. Let $\mathbb{E}$ be a Euclidean space and let $f\colon \mathbb{E} \to \mathbb{R}$ be a continuous function. Then:

(i) $\operatorname{epi}(f)$ is closed;

(ii) $\operatorname{int}(\operatorname{epi}(f)) \neq \varnothing$.

Proof.

(i) Consider the function $g\colon \mathbb{E}\oplus\mathbb{R} \to \mathbb{R}$ given by $g(x\oplus t) := f(x) - t$ for each $(x \oplus t) \in \mathbb{E}\oplus\mathbb{R}$. Note that g is continuous. Also, we have that $x \oplus t \in \operatorname{epi}(f)$ if and only if $g(x \oplus t) \leq 0$ and thus $\operatorname{epi}(f) = g^{-1}((-\infty, 0])$. Then, since $(-\infty, 0]$ is closed, the result follows from Proposition 7.

(ii) Let $x \in \mathbb{E}$. Then $(x \oplus f(x)) \in \operatorname{epi}(f)$. Since f is continuous, let $\varepsilon, \delta \in \mathbb{R}_{++}0$ such that

$$y \in x + \delta\mathbb{B} \text{ implies } |f(y) - f(x)| \leq \varepsilon.$$

Since $\varepsilon \in \mathbb{R}_{++}$, we have that $f(x) + 2\varepsilon > f(x)$ and thus $(x \oplus (f(x) + 2\varepsilon)) \in \operatorname{epi}(f)$. Define $\gamma := \min\{\varepsilon, \delta\}$, we shall prove that $(x \oplus (f(x) + 2\varepsilon)) + \gamma\mathbb{B} \subseteq \operatorname{epi}(f)$.

Let $z \oplus t \in (x \oplus (f(x)+2\varepsilon)) + \gamma\mathbb{B}$. Then, it follows that $\|(z\oplus t) - (x \oplus (f(x)+2\varepsilon))\| \leq \gamma$ and thus we have that $\|z - x\| \leq \delta$ and $|t - (f(x) + 2\varepsilon)| \leq \varepsilon$. Since $\|z - x\| \leq \delta$ we have that $z \in x + \delta\mathbb{B}$. Hence, $|f(z) - f(x)| \leq \varepsilon$. Summing these inequalities yields:

$$-2\varepsilon \leq f(z) - t + 2\varepsilon \leq 2\varepsilon \iff -4\varepsilon + t \leq f(z) \leq t.$$

Thus, we have that $(z \oplus t) \in \operatorname{epi}(f)$. Therefore, $(x \oplus (f(x) + 2\varepsilon)) + \gamma\mathbb{B} \subseteq \operatorname{epi}(f)$. □

Proposition 10. Let $\mathbb{E}$ be an Euclidean space and let $\{S_i\}_{i\in I} \subseteq \mathbb{E}$ for each $i \in I$. Then:

(i) if S_i is open for each $i \in I$, then $\bigcup_{i\in I} S_i$ is open;

(ii) if S_i is closed for each $i \in I$, then $\bigcap_{i\in I} S_i$ is closed;

(iii) if S_i is open for each $i \in I$ and I is finite, then $\bigcap_{i\in I} S_i$ is open;

(iv) if S_i is closed for each $i \in I$ and I is finite, then $\bigcup_{i\in I} S_i$ is closed.

Proof.

(i) Let $x \in \bigcup_{i\in I} S_i$. Then, $x \in S_i$ for some $i \in I$. Since S_i is open, there exists $\varepsilon > 0$ such that $x + \varepsilon\mathbb{B} \subseteq S_i$. Thus, $x + \varepsilon\mathbb{B} \subseteq \bigcup_{i\in I} S_i$. Therefore, $\bigcup_{i\in I} S_i$ is open.

(ii) Since S_i is closed for each $i \in I$, we have that S_i^c is open for each $i \in I$. From Item (i), we have that $\bigcup_{i\in I} S_i^c$ is open. Thus, $(\bigcup_{i\in I} S_i^c)^c = \bigcap_{i\in I} S_i$ is closed.

(iii) Let $x \in \bigcap_{i\in I} S_i$. Since S_i is open for each $i \in I$, there exists ε_i such that $x + \varepsilon_i\mathbb{B} \subseteq S_i$. Setting $0 < \bar{\varepsilon} := \min\{\varepsilon_i : i \in I\}$ we obtain that $x + \bar{\varepsilon}\mathbb{B} \subset S_i$ for each $i \in I$. Thus, $x + \bar{\varepsilon}\mathbb{B} \subset \bigcap_{i\in I} S_i$. That is, $\bigcap_{i\in I} S_i$ is open.

(iv) Since S_i is closed for each $i \in I$, we have that S_i^c is open for each $i \in I$. From Item (iii), we have that $\bigcap_{i\in I} S_i^c$ is open and thus $(\bigcap_{i\in I} S_i^c)^c = \bigcup_{i\in I} S_i$ is closed. □

Theorem 11. Let V be a vector space, let $C \subset V$ be a compact set, and let $f\colon C \to \mathbb{R}$ be a continuous function. Then $\inf_{x\in C} f(x) = \min_{x\in C} f(x)$ and $\sup_{x\in C} f(x) = \max_{x\in C} f(x)$.

Proof. This proof uses that $f(C)$ is compact. Showing this result would require much additional terminology and it can be found in, for example, [29; 32].

Because $f(C) \subset \mathbb{R}$ is bounded, there exist $\alpha, \beta \in \mathbb{R}$ such that $\alpha = \inf f(C)$ and $\beta = \sup f(C)$. Since $f(C)$ is closed, we know that $f(C) = \overline{f(C)}$. Thus, it suffices to show that α and β belong to $\overline{f(C)}$. First, let $\varepsilon \in \mathbb{R}_{++}$ and note that if $\alpha + \varepsilon\mathbb{B} \cap f(C) = \varnothing$ then $\inf f(C) \geq \alpha + \varepsilon$. This implies that $\inf f(C) > \alpha$, which is a contradiction and thus, $\alpha \in f(C)$. Similarly, let $\varepsilon' \in \mathbb{R}_{++}$ and note that if $\beta + \varepsilon'\mathbb{B} \cap f(C) = \varnothing$ then $\sup f(C) \leq \beta - \varepsilon'$. This implies that $\inf f(C) < \beta$, which is a contradiction. So, we conclude that, $\beta \in f(C)$. □

Theorem 12. Let $A \in \mathbb{S}^n$. Then $A \in \mathbb{S}^n_+$ if and only if $\det(A[\{1, \ldots, k\}]) \geq 0$ for each $k \in [n]$.

Proof. We show this result for the case that $n = 2$. Let $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ and consider $f(x_1, x_2) := x^t Ax = ax_1^2 + 2bx_1x_2 + cx_2^2$. We seek for conditions under a, b and c such that f is non-negative throughout $\mathbb{R}^2$.

First, note that $f(x_1, 0) = ax_1^2$ is non-negative if and only if $a \geq 0$. Similarly, we have that $f(0, x_2) = c^2 x_2$ if and only if $c \geq 0$. Finally, consider the case where $x_1 \neq 0$ and $x_2 \neq 0$. We shall prove that $f(x_1, x_2) \geq 0$ if and only if $p(t) := at^2 + bt + c \geq 0$ for each $t \in \mathbb{R}$, where $t := \frac{x_1}{x_2}$. It follows:

$$\begin{aligned} & f(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2 \geq 0 \text{ for each } x_1 \neq 0, x_2 \neq 0 \\ \iff & \frac{f(x_1, x_2)}{y^2} = a\frac{x_1^2}{x_2^2} + 2b\frac{x_1x_2}{x_2^2} + c\frac{x_2^2}{x_2^2} \geq 0 \text{ for each } x_1 \neq 0, x_2 \neq 0 \\ \iff & p(t) = at^2 + 2bt + c \geq 0 \text{ for each } t \in \mathbb{R}. \end{aligned}$$

Moreover, note that p does not change signs if and only if $4b^2 - 4ac \leq 0$. That is,

$$b^2 - ac = -\det(A) \leq 0. \quad \square$$

1.2 Elements From Convex Analysis

Let $\mathbb{E}$ be an Euclidean space and let $x, y \in \mathbb{E}$. We define

$$[x, y] \coloneqq \{\lambda x + (1 - \lambda)y : 0 \leq \lambda \leq 1\}$$

to be the *line segment* between x and y. A subset C of $\mathbb{E}$ is said to be *convex* if $[x, y] \subseteq C$ for each $x, y \in C$. For any set S, we have that $\operatorname{conv}(S)$ is convex. If S is already convex, then $\operatorname{conv}(S) = S$. Also, $\operatorname{conv}(S) = \bigcap\{C \supseteq S : C \text{ is convex}\}$. In this sense, $\operatorname{conv}(S)$ can be regarded as the smallest convex set containing S. The unitary ball, polyhedra, hyperplanes, half-spaces, and the empty set are all examples of convex sets. We will now show some operations that preserve convexity.

Proposition 13. Let $\mathbb{E}$ be an Euclidean space and let $\{C_i\}_{i \in I} \subseteq \mathbb{E}$ be a family of convex sets. Then:

(i) $\bigcap_{i \in I} C_i$ is convex;

(ii) $\sum_{i \in I} C_i$ is convex if I is finite;

(iii) $\bigoplus_{i \in I} C_i$ is convex if I is finite.

Proof.

(i) Let $x, y \in \bigcap_{i \in I} C_i$ so that $x, y \in C_i$ for each $i \in I$. Since C_i is convex for each $i \in I$, we have that $[x, y] \subseteq C_i$ for each $i \in I$. Therefore, $[x, y] \subseteq \bigcap_{i \in I} C_i$.

(ii) Let $x, y \in \sum_{i \in I} C_i$ and let $\lambda \in [0, 1]$. Then:

$$\lambda x + (1 - \lambda)y = \lambda \sum_{i \in I} x_i + (1 - \lambda) \sum_{i \in I} y_i = \sum_{i \in I}(\lambda x_i + (1 - \lambda)y_i).$$

Since $\lambda x_i + (1 - \lambda)y_i \in C_i$ for each $i \in I$, the result follows.

(iii) Let $x, y \in \bigoplus_{i \in I} C_i$ and let $\lambda \in [0, 1]$. Then:

$$\lambda x + (1 - \lambda)y = \lambda \bigoplus_{i \in I} x_i + (1 - \lambda) \bigoplus_{i \in I} y_i = \bigoplus_{i \in I}(\lambda x_i + (1 - \lambda)y_i).$$

Since $\lambda x_i + (1 - \lambda)y_i \in C_i$ for each $i \in I$, the result follows. □

Proposition 14. Let $\mathbb{E}$, $\mathbb{Y}$ be Euclidean spaces, let $C \subseteq \mathbb{E}$ and $S \subseteq \mathbb{Y}$ both be convex sets, and let $A \colon \mathbb{E} \to \mathbb{Y}$ be a linear transformation. Then:

(i) $A(C)$ is convex;

(ii) $A^{-1}(S)$ is convex.

Proof.

(i) Let $x, y \in A(C)$ and $\lambda \in [0, 1]$. By the definition of $A(C)$, there exist $v, w \in C$ such that $A(v) = x$ and $A(w) = y$. Since C is convex we have that $\lambda v + (1 - \lambda)w \in C$. Thus:

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

(ii) Let $x, y \in A^{-1}(S)$ and $\lambda \in [0,1]$. By the definition of $A^{-1}(S)$, there exist v and $w \in S$ such that $A(x) = v$ and $A(y) = w$. Since S is convex:

$$\lambda v + (1-\lambda)w = \lambda A(x) + (1-\lambda)A(y) = A(\lambda x + (1-\lambda)y) \in S.$$

Therefore, $\lambda x + (1-\lambda)y \in A^{-1}(S)$. □

Some Topological Properties of Convex Sets

The next propositions concern about interiors and closures of convex sets, which present stronger properties when comparing to ordinary sets. However, it happens that many convex sets of interest may have empty interior. To avoid this inconvenient and obtain results that are valid for any convex set, one considers the *relative interior* of a convex set C as

$$\mathrm{ri}(C) := \{x \in C : \text{there exists } \varepsilon \in \mathbb{R}_{++} \text{ such that } (x + \varepsilon\mathbb{B}) \cap \mathrm{aff}(C) \subseteq C\}.$$

We note that

$$\mathrm{ri}(C) \subseteq C \subseteq \overline{C} \subseteq \mathrm{aff}(C).$$

This implies that $[x, y] \subseteq \mathrm{aff}(C)$ for any $x, y \in C$. We also observe that if C has nonempty interior, then the affine hull of C is the whole ambient space. Therefore, under this extra assumption, it is possible to achieve the same results replacing relative interiors with the usual concept of interior.

Proposition 15. Let $\mathbb{E}$ be an Euclidean space and let $C \subseteq \mathbb{E}$ be a convex set. If $x \in \mathrm{ri}(C)$ and $y \in \overline{C}$, then $(1-\lambda)x + \lambda y \in \mathrm{ri}(C)$ for each $\lambda \in [0,1)$.

Proof. Let $\lambda \in [0,1)$. We shall prove that there exists $\varepsilon \in \mathbb{R}_{++}$ such that

$$((1-\lambda)x + \lambda y) + \varepsilon\mathbb{B} \cap \mathrm{aff}(C) \subseteq C.$$

Since $y \in \overline{C}$, we have that $y + \varepsilon\mathbb{B} \neq \varnothing$ for each $\varepsilon \in \mathbb{R}_{++}$. Then:

$$\begin{aligned}
((1-\lambda)x + \lambda y) + \varepsilon\mathbb{B} \cap \mathrm{aff}(C) &\subseteq ((1-\lambda)x + \lambda(C + \varepsilon\mathbb{B})) + \varepsilon\mathbb{B} \cap \mathrm{aff}(C) \\
&= (1-\lambda)x + \lambda C + (1+\lambda)\varepsilon\mathbb{B} \cap \mathrm{aff}(C) \\
&= (1-\lambda)\left(x + \frac{(1+\lambda)\varepsilon}{1-\lambda}\mathbb{B}\right) + \lambda C \cap \mathrm{aff}(C).
\end{aligned}$$

Since $x \in \mathrm{ri}(C)$, we have that there exists $\varepsilon' \in \mathbb{R}_{++}$ such that $x + \varepsilon'\mathbb{B} \cap \mathrm{aff}(C) \subseteq C$. Thus, if $\varepsilon = \frac{1-\lambda}{1+\lambda}\varepsilon'$, then $(x + \frac{(1+\lambda)\varepsilon}{1-\lambda}\mathbb{B}) \cap \mathrm{aff}(C) \subseteq C$. Whence:

$$(1-\lambda)\left(x + \frac{(1+\lambda)\varepsilon}{1-\lambda}\mathbb{B}\right) + \lambda C \subseteq (1-\lambda)C + \lambda C = C.$$

So, we conclude that $((1-\lambda)x + \lambda y) + \bar{\varepsilon}\mathbb{B} \cap \mathrm{aff}(C) \subseteq C$, which is the desired result. □

The reader may also note that Proposition 15 implies that the interior of a convex set C is convex and is always nonempty. Moreover, we note that the closure of C is convex since it is the intersection of the convex sets $C + \varepsilon\mathbb{B}$, where $\varepsilon \in \mathbb{R}_{++}$. Polyhedral sets are important examples of convex objects that are always closed. Euclidean spaces, linear supspaces, affine sets, and the empty set are sets which are equal to their respective closures and relative interiors.

Proposition 16. Let $\mathbb{E}$ be an Euclidean space and let $C \subseteq \mathbb{E}$ be a convex set. Then $x \in \text{ri}(C)$ if, and only if there exists $\lambda > 1$ such that $(1-\lambda)y + \lambda x \in C$ for each $y \in C$.

Proof. First, assume that $x \in \text{ri}(C)$ and fix $y \in C$. Since $x \in \text{ri}(C)$, we know that $x + \varepsilon\mathbb{B} \cap \text{aff}(C) \subseteq C$ for some $\varepsilon \in \mathbb{R}_{++}$. For such ε, we have $x + \frac{\varepsilon}{\|x-y\|}(x-y) \in C$. Writing $x = y + x - y$ yields:

$$\begin{aligned} y + x - y + \frac{\varepsilon}{\|x-y\|}x - \frac{\varepsilon}{\|x-y\|}y &= y + \left(1 + \frac{\varepsilon}{\|x-y\|}\right)(x-y) \\ &= \left(\left(1 + \frac{\varepsilon}{\|x-y\|}\right)x - \frac{y\varepsilon}{\|x-y\|}\right) \in C. \end{aligned}$$

Conversely, assume that x is such that for each $y \in C$ there exists $\lambda > 1$ such that $(1-\lambda)y + \lambda x \in C$. Let $y \in \text{ri}(C)$. If $x = y$ we are done. Otherwise, let $\lambda > 1$ such that $(1-\lambda)y + \lambda x \in C$ and set $z := (1-\lambda)x + \lambda y$. Since $z \in C$, Proposition 15 gives us that $(1-\alpha)y + \alpha z \in \text{ri}(C)$ for each $\alpha \in [0,1)$. In particular, if $\alpha = \frac{1}{\lambda}$:

$$(1-\alpha)y + \alpha z = x.$$

Therefore, $x \in \text{ri}(C)$. □

This last proof can be replicated considering directions y belonging aff(C). In this context, it is possible to consider more vectors such that there exists $\lambda > 1$ satisfying $\lambda x + (1-\lambda)y \in C$ for each $x \in \text{ri}(C)$. On the other hand, it would be necessary to find a 'larger' set of directions satisfying this property in order to confirm that $x \in \text{ri}(C)$. Obviously, since both results are true one can always use the strongest implication of each.

Proposition 17. Let $\mathbb{E}$ be an Euclidean space and let $C \subseteq \mathbb{E}$ be a convex set. Then:

(i) $\overline{\text{ri}(C)} = \overline{C}$;

(ii) $\text{ri}(C) = \text{ri}(\overline{C})$.

Proof.

(i) First, we note that $\overline{\text{ri}(C)} \subseteq \overline{C}$ because $\text{ri}(C) \subseteq C$. Conversely, let $x \in \text{ri}(C)$ and $y \in \overline{C}$. By Proposition 15, it follows that $(1-\lambda)x + \lambda y \in \text{ri}(C)$ for each $\lambda \in [0,1)$. Hence, $y + \varepsilon\mathbb{B} \cap \text{ri}(C) \neq \varnothing$ for each $\varepsilon \in \mathbb{R}_{++}$. Therefore, $y \in \overline{\text{ri}(C)}$.

(ii) Similarly to the former item, we start noting that $\text{ri}(C) \subseteq \text{ri}(\overline{C})$ since $C \subseteq \overline{C}$. Conversely, let $x \in \text{ri}(\overline{C})$ and let $y \in \text{ri}(C)$. If $x = y$ we are done. Otherwise, Proposition 16 gives us that there exists $\alpha > 1$ such that $(1-\alpha)y + \alpha x \in \text{ri}(\overline{C}) \subseteq \overline{C}$. Define $z = (1-\alpha)y + \alpha x$. By Proposition 15, $(1-\lambda)y + \lambda z \in \text{ri}(C)$ for each $\lambda \in [0,1)$. In particular, for $\lambda = \frac{1}{\alpha}$:
$$(1-\lambda)y + \lambda z = x.$$
Therefore, $x \in \text{ri}(C)$. □

Proposition 18. Let $\mathbb{E}$ be an Euclidean space and let $\{C_i\}_{i\in I} \subseteq \mathbb{E}$ be a family of convex sets such that $\bigcap_{i\in I} \text{ri}(C_i) \neq \varnothing$. Then:

(i) $\overline{\bigcap_{i\in I} C_i} = \bigcap_{i\in I} \overline{C_i}$;

(ii) $\text{ri}(\bigcap_{i\in I} C_i) = \bigcap_{i\in I} \text{ri}(C_i)$, if I is finite.

Proof.

(i) First, let $x \in \bigcap_{i\in I} \overline{C_i}$. Then, $x \in \overline{C_i}$ for each $i \in I$. By hypothesis, there exists $y \in \bigcap_{i\in I} \operatorname{ri}(C_i)$. That is, $y \in \operatorname{ri}(C_i)$ for each $i \in I$. Thus, by Proposition 15, for each $\lambda \in [0,1)$ and $i \in I$ we have $(1-\lambda)y + \lambda x \in \operatorname{ri}(C_i)$. Since $\bigcap_{i\in I} \operatorname{ri}(C_i) \subseteq \bigcap_{i\in I} C_i$, it follows that $x + \varepsilon\mathbb{B} \cap \bigcap_{i\in I} C_i \neq \varnothing$ for each $\varepsilon \in \mathbb{R}_{++}$. Thus, $x \in \overline{\bigcap_{i\in I} C_i}$. Conversely, let $x \in \overline{\bigcap_{i\in I} C_i}$. By definition, we have that $x + \varepsilon\mathbb{B} \cap (\bigcap_{i\in I} C_i)$ for each $\varepsilon \in \mathbb{R}_{++}$. Hence, for each $i \in I$ and $\varepsilon \in \mathbb{R}_{++}$:
$$x + \varepsilon\mathbb{B} \cap C_i \neq \varnothing.$$
Thus, by definition $x \in \overline{C}_i$ for each $i \in I$. Therefore, $x \in \bigcap_{i\in I} \overline{C_i}$.

(ii) The inclusion $\operatorname{ri}(\bigcap_{i\in I} C_i) \subseteq \bigcap_{i\in I} \operatorname{ri}(C_i)$ is easy to prove and does not depend on the finiteness of I. Let $x \in \operatorname{ri}(\bigcap_{i\in I} C_i)$, then there exists $\varepsilon \in \mathbb{R}_{++}$ such that $x + \varepsilon\mathbb{B} \subseteq \bigcap_{i\in I} C_i$. In particular, $x + \varepsilon\mathbb{B} \subseteq C_i$ for each $i \in I$. Thus, $x \in \operatorname{ri}(C_i)$ for each $i \in I$. Therefore $x \in \bigcap_{i\in I} \operatorname{ri}(C_i)$. Conversely, let $x \in \bigcap_{i\in I} \operatorname{ri}(C_i)$. Then, there exists $\varepsilon\colon I \to \mathbb{R}_{++}$ such that $x + \varepsilon_i\mathbb{B} \subseteq C_i$ for each $i \in I$. Assuming I is finite we can define $\varepsilon' := \min\{\varepsilon_i : i \in I\}$ and note that $x + \varepsilon'\mathbb{B} \subseteq C_i$ for each $i \in I$. That is, $x + \varepsilon'\mathbb{B} \subseteq \bigcap_{i\in I} C_i$. Hence, $x \in \operatorname{ri}(\bigcap_{i\in I} C_i)$. □

Considering $C = \mathbb{Q}$ produces an example illustrating why Proposition 17 does not hold for non-convex sets. Similarly, setting $C_1 = \mathbb{Q}$ and $C_2 = \mathbb{R} \setminus \mathbb{Q}$ shows that Proposition 18 is not true in general. If some members of the family $\{C_i\}_{i\in I}$ are closed, the first item of this last result can be refined as follows.

Proposition 19. Let $\mathbb{E}$ be an Euclidean space and let $\{C_i\}_{i\in I} \subseteq \mathbb{E}$ be a family of convex sets. Let $I_0 \subseteq I$ such that C_i is closed for each $i \in I_0$. Assume that
$$S := (\bigcap_{i\in I\setminus I_0} \operatorname{ri}(C_i)) \cap (\bigcap_{i\in I_0} C_i) \neq \varnothing.$$
Then $\bigcap_{i\in I} \overline{C_i} = \overline{\bigcap_{i\in I} C_i}$.

Proof. Let $x \in \bigcap_{i\in I} \overline{C_i}$. By hypothesis, there exists $y \in S$. In particular, $y \in \operatorname{ri}(C_i)$ for each $i \in I \setminus I_0$. From Proposition 15, we obtain that $(1-\lambda)y + \lambda x \in \operatorname{ri}(C_i)$ for each $\lambda \in [0,1)$ and $i \in I \setminus I_0$. Thus, $x + \varepsilon\mathbb{B} \cap \bigcap_{i\in I\setminus I_0} C_i \neq \varnothing$ for each $\varepsilon \in \mathbb{R}_{++}$. Because $C_i = \overline{C_i}$ for $i \in I_0$, it follows that $x + \varepsilon\mathbb{B} \cap \bigcap_{i\in I} C_i \neq \varnothing$. Therefore, $x \in \overline{\bigcap_{i\in I} C_i}$.

Conversely, let $x \in \overline{\bigcap_{i\in I} C_i}$. By definition, we have that $x + \varepsilon\mathbb{B} \cap (\bigcap_{i\in I} C_i) \neq \varnothing$ for each $\varepsilon \in \mathbb{R}_{++}$. Hence, for each $i \in I$ and $\varepsilon \in \mathbb{R}_{++}$:
$$x + \varepsilon\mathbb{B} \cap C_i \neq \varnothing.$$
Thus, by definition $x \in \overline{C}_i$ for each $i \in I$. In other words, $x \in \bigcap_{i\in I} \overline{C_i}$. □

Proposition 20. Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces, let $C \subseteq \mathbb{E}$ be a convex set, and let $A\colon \mathbb{E} \to \mathbb{Y}$ be a linear function. Then:

(i) $A(\overline{C}) \subseteq \overline{A(C)}$;

(ii) $\operatorname{ri}(A(C)) = A(\operatorname{ri}(C))$.

Proof.

(i) Let $x \in A(\overline{C})$. By definition, there exists $y \in \overline{C}$ such that $A(y) = x$. Let $\varepsilon \in \mathbb{R}_{++}$. By Proposition 8, we have $\delta \in \mathbb{R}_{++}$ such that $z \in y + \delta\mathbb{B}$ implies that $f(z) \in x + \varepsilon\mathbb{B}$. Since $y \in \overline{C}$ we can assume that $z \in C$, obtaining that $f(z) \in A(C)$. Therefore, $x + \varepsilon\mathbb{B} \cap A(C) \neq \varnothing$ for each $\varepsilon \in \mathbb{R}_{++}$. That is, $x \in \overline{A(C)}$.

(ii) Let $x \in \operatorname{ri}(A(C)) \subseteq A(C)$ and assume that $x \notin A(\operatorname{ri}(C))$. Thus, for each $y \in C$ such that $A(y) = x$ we have that $i \in C \setminus \operatorname{ri}(C)$. Then, for each $\varepsilon \in \mathbb{R}_{++}$, there exists $z \in y + \varepsilon\mathbb{B}$ such that $z \in \mathbb{E} \setminus C$. Applying Proposition 8 we conclude that for each $\gamma \in \mathbb{R}_{++}$ there exists $w \in x + \gamma\mathbb{B}$ such that $w \in \mathbb{Y} \setminus A(C)$. Therefore, $x \notin \operatorname{ri}(A(C))$.

Conversely, let $x_1 \in A(\operatorname{ri}(C))$ and let $y_1 \in A(C)$. Consider $x_2 \in \operatorname{ri}(C)$ such that $A(x_2) = x_1$ and $y_2 \in C$ such that $A(y_2) = y_1$. By Proposition 16, we have that there exists $\lambda > 1$ such that $(1-\lambda)y_2 + \lambda x_2 \in C$. Thus,

$$(A((1-\lambda)y_2 + \lambda x_2) = (1-\lambda)A(y_2) + \lambda A(x_2) = (1-\lambda)y_1 + \lambda x_1 \in C.$$

Therefore, $x_1 \in \operatorname{ri}(A(C))$ by Proposition 16. □

Corollary 21. Let $\mathbb{E}$ be an Euclidean space, and let $\varnothing \neq C_1, C_2 \subseteq \mathbb{E}$ be convex sets. Then

(i) $\overline{C_1} + \overline{C_2} \subseteq \overline{C_1 + C_2}$;

(ii) $\operatorname{ri}(C_1) + \operatorname{ri}(C_2) = \operatorname{ri}(C_1 + C_2)$.

Proof. Consider the linear function $A \colon \mathbb{E} \oplus \mathbb{E} \to \mathbb{E}$ where $A(x_1 \oplus x_2) \coloneqq x_1 + x_2$ for each $x_1 \oplus x_2 \in \mathbb{E} \oplus \mathbb{E}$ and apply Proposition 20 to A and $C_1 \oplus C_2$, which is trivially nonempty and convex by Proposition 13. □

Proposition 22. Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces, let $A \colon \mathbb{E} \to \mathbb{Y}$ be a linear function, and let $C \subseteq \mathbb{E}$ and $S \subseteq \mathbb{Y}$ both be polyhedral sets. Then:

(i) $A(C)$ is polyhedral;

(ii) $A^{-1}(S)$ is polyhedral.

Proof.

(i) Writing $C = \{\sum_{i\in[m]} \lambda_i x_i : \lambda \colon I \to \mathbb{R}_+, \sum_{i\in[k]} \lambda_i = 1\}$, it it obvious that

$$A(C) = \{\sum_{i\in[m]} \lambda_i A(x)_i : \lambda \colon I \to \mathbb{R}_+, \sum_{i\in[k]} \lambda_i = 1\}.$$

Thus, $A(C)$ is polyhedral.

(ii) Since S is polyhedral, we have by definition $S = \bigcap_{i\in I}\{y \in \mathbb{Y} : \langle y, b_i\rangle \leq \beta_i\}$ Thus, $A^{-1}(S) = \bigcap_{i\in I}\{x \in \mathbb{E} : \langle A(x), b_i\rangle \leq \beta_i\}$. □

Corollary 23. Let $\mathbb{E}$ be an Euclidean space and let $C_1, C_2 \subseteq \mathbb{E}$ be polyhedral sets. Then $C_1 + C_2$ is polyhedral.

Proof. Immediate. □

The previous corollary is an obvious consequece of Proposition 22. Together with the first item of Corollary 21 (which we will improve in Chapter 2) and the tools from hyperplane separation, this result will be essential to derive the conditions that ensure conic duality to hold.

Hyperplane Separation

The notion of separation is fundamental to the development of the theory of convex analysis. If C_1 and C_2 are convex subsets of an Euclidean space $\mathbb{E}$, we say that a hyperplane $H := \{x \in \mathbb{E} : \langle a, x\rangle \leq \beta\}$ separates C_1 and C_2 if $\langle a, x_1\rangle \leq \beta$ for each $x_1 \in C_1$ and $\langle a, x_2\rangle \geq \beta$ for each $x_2 \in C_2$. If in addition we have $C_1 \cup C_2 \nsubseteq H$, then H separates C_1 and C_2 *properly.* Moreover, if $\langle a, x_2\rangle > \beta$ for each $x_2 \in C_2$, then H separates C_1 and C_2 *strongly.*

The following results will handle strong and proper separation. First, we show conditions that allow us to separate points from sets both properly and strongly. Then, we extend these results in order to separate a pair of nonempty convex sets. Later, we modify this condition considering the case where one of these sets is a polyhedron.

Hopefully, this will be sufficient to bring some insight about the meanings and consequences of separation. A detailed description of this theory including proofs may be found in Chapter 11 of [30].

Proposition 24. Let $\mathbb{E}$ be an Euclidean space, let $\{x \in \mathbb{E} : \langle a, x\rangle = \beta\} =: H \subset \mathbb{E}$ be a hyperplane, and let $\varnothing \neq C \subset \mathbb{E}$ be a convex set such that $C \subseteq \{x \in \mathbb{E} : \langle a, x\rangle \leq \beta\}$. Then $\operatorname{ri}(C) \cap H \neq \varnothing$ if, and only if $C \subseteq H$.

Proof. First, note that if $C \subseteq H$ then $\operatorname{ri}(C) \subseteq H$. Because $C \neq \varnothing$ we have that $\operatorname{ri}(C) \neq \varnothing$. Thus, $\operatorname{ri}(C) \cap H \neq \varnothing$.

On the other hand, assume that $\operatorname{ri}(C) \cap H \neq \varnothing$ and that there exists $x \in C \setminus H$. In this case, consider the line segment $[x, y]$ where $y \in \operatorname{ri}(C) \cap H$. By Proposition 16, there exists $\lambda > 1$ such that $(1 - \lambda)x + \lambda y =: z \in C$. For such z we have that $\langle a, z\rangle > \beta$, which is a contradiction. □

Theorem 25. Let $\mathbb{E}$ be an Euclidean space and let $C \subseteq \mathbb{E}$ be a closed convex set. If $y \in \mathbb{E} \setminus C$, then there exists $a \in \mathbb{E} \setminus \{0\}$ and $\beta \in \mathbb{R}$ such that $C \subseteq \{v \in \mathbb{E} : \langle a, v\rangle \leq \beta\}$ and $\langle a, y\rangle > \beta$.

Proof. Let $\bar{x} \in C$ and set $\delta := \|\bar{x} - y\|$. Then, $C' := C \cap (\bar{x} + \delta\mathbb{B}) \neq \varnothing$ is a intersection of a closed and a compact set and thus, C' is compact. By Theorem 11, there exists $z \in C'$ that minimizes the continuous function $f : C' \to \mathbb{R}_+$ given by $f(\cdot) := \| \cdot - y\|$. Set $a := y - z$ and $\beta := \langle a, z\rangle$. Note that if $w \in C \setminus C'$ then $\|w - y\| \geq \|v - y\|$ for any $v \in C'$. Thus, z also minimizes the extension of f to C. Let $z \neq x \in C$. Since C is convex we have $[x, z] \subseteq C$. So, for each $\lambda \in [0, 1]$:

$$\|\lambda x+(1-\lambda)z-y\|^2 = \|(z-y)+\lambda(x-z)\|^2 = \|z-y\|^2+2\lambda\langle x-z, z-y\rangle+\lambda^2\|x-z\| \geq \|z-y\|^2.$$

Hence, $\lambda\|x - z\|^2 \geq 2\langle x - z, a\rangle$. By sending λ to 0, we obtain that $\langle x - z, a\rangle \leq 0$. That is, $\langle a, x\rangle \leq \langle a, z\rangle = \beta$. On the other hand, $\langle y - z, y - z\rangle > 0$ is the same as $\langle a, y - z\rangle > 0$. Therefore, $\langle a, y\rangle > \beta$. □

Proposition 26. Let $\mathbb{E}$ be an Euclidean space, let $\varnothing \neq C$ be a convex set and let $x \in \mathbb{E}$. Then there exists a hyperplane separating C and x properly if, and only if $x \notin \operatorname{ri}(C)$.

Proof. First assume that $x \notin \operatorname{ri}(C)$. Note that if $x \notin \overline{C}$, the result is given by Theorem 25. Otherwise, let $y \in \mathbb{E} \setminus \overline{C}$ and $z \in \operatorname{ri}(C)$ such that $x \in [y, z]$. In this case, x minimizes the function specified in the construction of Theorem 25. Adopting the same $a \in \mathbb{E} \setminus \{0\}$ we obtain that $\langle a, x\rangle \geq \langle a, w\rangle$ for each $w \in \overline{C}$. Finally, we obtain from Proposition 24 that $C \nsubseteq H$ because $x \notin \operatorname{ri}(C)$.

Conversely, assume by contradiction that H is a hyperplane properly separating x and C and $x \in \operatorname{ri}(C)$. In this case, we would have that $C \subseteq H$ and, since $x \in C$, $C \cup \{x\} = C \subseteq H$. This contradicts the definition of proper separation. □

Proposition 27. Let $\mathbb{E}$ be an Euclidean space and let $\varnothing \neq C_1, C_2 \subseteq \mathbb{E}$ be convex sets. Then, there exists a hyperplane separating C_1 and C_2 properly if and only if there exists $a \in \mathbb{E}$ such that:

(i) $\inf\{\langle x, a\rangle : x \in C_1\} \geq \sup\{\langle x, a\rangle : x \in C_2\}$;

(ii) $\sup\{\langle x, a\rangle : x \in C_1\} > \inf\{\langle x, a\rangle : x \in C_2\}$.

Proof. First, assume that there exists $a \in \mathbb{E}$ satisfying (i) and (ii). Observe that $a \neq 0$ because otherwise we would have $\sup\{\langle x, a\rangle : x \in C_1\} = \inf\{\langle x, a\rangle : x \in C_2\} = 0$, which violates (ii). Set

$$\beta := \inf\{\langle x, a\rangle : x \in C_1\} + \frac{\sup\{\langle x, a\rangle : x \in C_2\} - \inf\{\langle x, a\rangle : x \in C_1\}}{2}$$

and consider $H = \{x \in \mathbb{E} : \langle x, a\rangle = \beta\}$. Then $C_2 \subseteq \{x \in \mathbb{E} : \langle x, a\rangle \leq \beta\}$ and $C_1 \subseteq \{x \in \mathbb{E} : \langle x, a\rangle \geq \beta\}$. Moreover, (ii) implies that $C_2 \not\subseteq H$. Thus, H separates C_1 and C_2 properly.

Conversely, Let H be a hyperplane separating C_1 and C_2 properly. By definition, we can assume that there exists $a \in \mathbb{E} \setminus \{0\}$ and $\beta \in \mathbb{R}$ such that $C_1 \subseteq \{x \in \mathbb{E} : \langle x, a\rangle \geq \beta\}$, $C_2 \subseteq \{x \in \mathbb{E} : \langle x, a\rangle \leq \beta\}$, and $C_1 \cup C_2 \not\subseteq H$. This implies that $\langle x, a\rangle \geq \beta$ for each $x \in C_1$. Hence, $\inf\{\langle x, a\rangle : x \in C_1\} \geq \beta$. On the other hand, because $C_2 \subseteq \{x \in \mathbb{E} : \langle x, a\rangle \leq \beta\}$, we have that $\sup\{\langle x, a\rangle : x \in C_2\} \leq \beta$. Thus, we conclude that

$$\sup\{\langle x, a\rangle : x \in C_1\} \geq \inf\{\langle x, a\rangle : x \in C_1\} \geq \sup\{\langle x, a\rangle : x \in C_2\} \geq \inf\{\langle x, a\rangle : x \in C_2\}.$$

The second inequality above corresponds to (i). Finally, since $C_1 \cup C_2 \not\subseteq H$ there exist $x_1 \in C_1$ and $x_2 \in C_2$ such that $\langle x_1, a\rangle > \langle x_2, a\rangle$. Therefore,

$$\sup\{\langle x, a\rangle : x \in C_1\} > \inf\{\langle x, a\rangle : x \in C_2\}.$$

The latter corresponds to (ii). □

Proposition 28. Let $\mathbb{E}$ be an Euclidean space, and let $\varnothing \neq C_1, C_2 \subseteq \mathbb{E}$ be convex sets. Then, there exists a hyperplane properly separating C_1 and C_2 if, and only if $\operatorname{ri}(C_1) \cap \operatorname{ri}(C_2) = \varnothing$.

Proof. Consider the convex set $C := C_1 - C_2$. By Corollary 21, we have that $\operatorname{ri}(C) = \operatorname{ri}(C_1) - \operatorname{ri}(C_2)$. Note that $0 \notin \operatorname{ri}(C)$ because $\operatorname{ri}(C_1) \cap \operatorname{ri}(C_2) = \varnothing$. By Proposition 26, there exists a hyperplane separating $\{0\}$ and C properly. Thus, by Proposition 27, we have that

(i) $0 \geq \sup\{\langle x, a\rangle : x \in C\}$;

(ii) $0 > \inf\{\langle x, a\rangle : x \in C\}$.

Note that item (i) is equivalent to $\inf\{\langle x, a\rangle : x \in C_1\} \geq \sup\{\langle x, a\rangle : x \in C_2\}$ while item (ii) is equivalent to $\sup\{\langle x, a\rangle : x \in C_1\} > \inf\{\langle x, a\rangle : x \in C_2\}$. Then, the result follows from Proposition 27. □

Proposition 29. Let $\mathbb{E}$ be an Euclidean space, let $C_1, C_2 \subseteq \mathbb{E}$ be polyhedral sets. If $C_1 \cap C_2 = \varnothing$ then there exists a hyperplane separating C_1 and C_2 strongly.

Proof. Consider $C := C_1 - C_2$ and note that C is polyhedral by Corollary 23. Note that C is closed and that $0 \notin C$. Thus, applying Theorem 25 to C and 0 yields the desired result. □

Proposition 30. Let $\mathbb{E}$ be an Euclidean space and let $C, P \subseteq \mathbb{E}$ be convex sets such that P is polyhedral. Then there exists a hyperplane separating C and P properly and not containing C if, and only if $\operatorname{ri}(C) \cap P = \varnothing$.

Proof. First, assume that $H \coloneqq \{x \in \mathbb{E} : \langle a, x\rangle \leq \beta\}$ is a hyperplane separating C and P properly and not containing C. From our definition of separation, we obtain that $\langle a, c\rangle \leq \beta$ for each $c \in C$ and $\langle a, p\rangle \geq \beta$ for each $p \in P$. Because C is not contained in H, we have that $\mathrm{ri}(C) \cap H = \varnothing$. Thus, $\langle a, c'\rangle < \beta$ for each $c' \in \mathrm{ri}(C)$. Since $\langle a, p\rangle \geq \beta$ for every $p \in P$, it follows that $\mathrm{ri}(C) \cap P = \varnothing$.

Conversely, consider $D \coloneqq P \cap \mathrm{aff}(C)$. If D is empty, then we can separate P and $\mathrm{aff}(C)$ strongly by Proposition 29. Because $C \subseteq \mathrm{aff}(C)$, the same hyperplane separates C and P as required. If $D \neq \varnothing$, we note that $\mathrm{ri}(D)\cap\mathrm{ri}(C) = \mathrm{ri}(P)\cap\mathrm{ri}(C) = \varnothing$. Thus, Proposition 28 gives us a hyperplane H separating D and C properly. Also, H cannot contain C because otherwise it follows $C \cup D \subseteq \mathrm{aff}(C) \subseteq H$, which contradicts the definition of proper separation.

To conclude the proof, we shall use H to construct a hyperplane satisfying all our requirements. Let H_C be the closed half-space delimited by H that contains C and consider $W \coloneqq \mathrm{aff}(C) \cap H_C$. By construction we have that $C \subseteq W$ and, consequently, $\mathrm{ri}(C) \subseteq \mathrm{W}$. Moreover, W is polyhedral and $P \cap \mathrm{ri}(W) = \varnothing$. If actually $P \cap W = \varnothing$, the result follows from Proposition 29. Otherwise, set $M \coloneqq W \setminus \mathrm{ri}(W) = H \cap \mathrm{aff}(C)$. Translating all sets if necessary, we assume that $0 \in P \cap W$. Now consider $K \coloneqq \mathrm{cone}(P) + M$ and note that K is polyhedral by Corollary 23 and also $K \cap \mathrm{ri}(C) = \varnothing$. Express K as the intersection of a finite collection $\{H_i\}_{i\in I} \subseteq \mathbb{E}$ of closed half-spaces. Observe that $P \subseteq H_i$ for each $i \in I$. Since $W \nsubseteq K$, we obtain from Proposition 24 there exists some $j \in I$ such that $H_j \cap \mathrm{ri}(W) = \varnothing$ and hence $H_j \cap \mathrm{ri}(C) = \varnothing$. Therefore, H_j separates C and P properly and does not contain C. □

Corollary 31. Let $\mathbb{E}$ be an Euclidean space and let $C \subseteq \mathbb{E}$ be a closed convex set. Then C is the intersection of the half spaces that contain it. That is,

$$C = \bigcap_{(a,\beta)\in X} H(a, \beta)$$

Where $H(a, \beta) \coloneqq \{x \in \mathbb{E} : \langle x, a\rangle \leq \beta\}$ for each $a \in \mathbb{E} \setminus \{0\}$ and $\beta \in \mathbb{R}$ and

$$X \coloneqq \{(a, \beta) \in (\mathbb{E} \setminus \{0\}) \times \mathbb{R} : C \subseteq H(a, \beta)\}.$$

Proof. The right-hand side trivially contains C. Conversely, let x belonging to the right-hand side and assume that $x \notin C$. By Theorem 25, there exists $c \in \mathbb{E} \setminus \{0\}$ and $\eta \in \mathbb{R}$ such that $\langle x, c\rangle > \eta$ and $\langle y, c\rangle \leq \eta$ for each $y \in C$. This implies that x does not belong to the intersection above because we found $(c, \eta) \in X$ such that and $x \notin H(c, \eta)$. □

Convex Functions

Let $\mathbb{E}$ be an Euclidean space. We say that a function $f : V \to \mathbb{R}$ is *convex* if $\mathrm{epi}(f)$ is convex. If $C \subseteq V$, we define the *indicator function* of C by:

$$\delta(x \mid C) \coloneqq \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{otherwise} \end{cases}$$

Note that $\mathrm{epi}(f) = C \oplus \mathbb{R}_+$ and therefore $\delta(\cdot \mid C)$ is convex if and only if C is convex. Moreover, $\delta(\cdot \mid C)$ is closed if C is closed and is polyhedral if C is a polyhedron. More generally, an arbitrary polyhedral convex function $p : \mathbb{E} \to \mathbb{R}$ is be expressed by the formula:

$$p(x) = g(x) + \delta(x \mid P) = \begin{cases} g(x), & \text{if } x \in P; \\ +\infty, & \text{otherwise.} \end{cases}$$

Where $g(x) = \max_{i\in I}\{\langle x, a_i\rangle - \beta_i\}$ for some finite collection $\{a_i\}_{i\in I} \subseteq \mathbb{E} \setminus \{0\}$ and P is a polyhedron. The reader may note that the epigraph of g is the intersection of finitely many half-spaces in $\mathbb{E} \oplus \mathbb{R}$. We next present two propositions that give us different criteria to decide if a arbitrary function f is convex or not. Also, we provide a formula to calculate values of the closure of a convex function.

Proposition 32. Let V be a vector space, let $C \subseteq V$ be a convex set, and let $f : C \to \mathbb{R}$. Then, f is convex if and only if:

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

Proof. Let $(x \oplus f(x)), (y \oplus f(y)) \in \operatorname{epi}(f)$. It follows:

$$\begin{aligned} & \lambda(x \oplus f(x)) + (1-\lambda)(y \oplus f(y)) \in \operatorname{epi}(f), \text{ for each } \lambda \in [0,1] \\ \iff & (\lambda x + (1-\lambda)y) \oplus (\lambda f(x) + (1-\lambda)f(y)) \in \operatorname{epi}(f), \text{ for each } \lambda \in [0,1] \\ \iff & f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \text{ for each } \lambda \in [0,1]. \end{aligned}$$ □

Proposition 33. Let $C \subseteq \mathbb{R}$ be a convex set and let $f : C \to \mathbb{R}$ be a twice continuously differentiable function. Then, f is convex if and only if $f''(x) \geq 0$ for each $x \in C$.

Proof. First, we prove the backwards implication by contrapositive. Assume that there exists $v \in C$ such that $f''(v) < 0$. Then, since f'' is continuous there exists a segment (α, β) such that $f''(v) < 0$ for each $v \in (\alpha, \beta)$. Thus, f' is strictly non-increasing in (α, β). Let $x, y \in (\alpha, \beta)$ with $x < y$, let $\lambda \in (0, 1)$ and define $z := (1-\lambda)x + \lambda y$. We have:

$$f(z) - f(x) = \int_x^z f'(t)dt > f'(z)(z-x) \tag{1.1}$$

and

$$f(y) - f(z) = \int_z^y f'(t)dt < f'(z)(y-z). \tag{1.2}$$

Moreover, note that

$$z - x = \lambda y + (1-\lambda)x - x = \lambda(y-x) \text{ and } (y-z) = (1-\lambda)y - (1-\lambda)x = (1-\lambda)(y-x). \tag{1.3}$$

Substituting in equations (2.1) and (2.2), it follows:

$$f(z) > f(x) + \lambda f'(z)(y-x) \tag{1.4}$$

and

$$f(z) > f(y) - (1-\lambda)f'(z)(y-x). \tag{1.5}$$

Now, we multiply (1.4) and (1.5) by $(1-\lambda)$ and λ, respectively and add them to obtain:

$$f(z) > (1-\lambda)f(x) + \lambda f(y).$$

Since $z = (1-\lambda)x + \lambda y$ it follows that f is not convex from Proposition 32.

Conversely, assume that $f''(v) \geq 0$ for each $v \in C$. Then, we know that f' is non-decreasing. Let $x, y \in C$ with $x < y$, let $\lambda \in (0,1)$, and consider $z := (1-\lambda)x + \lambda y$. Then:

$$f(z) - f(x) = \int_x^z f'(t)dt \leq f'(z)(z-x)$$

and

$$f(y) - f(z) = \int_z^y f'(t)dt \leq f'(z)(y-z).$$

Substituting the results from (1.3) in the expressions above yields:

$$f(z) \leq f(x) + \lambda f'(z)(y-z) \tag{1.6}$$

and

$$f(z) \leq f(y) - (1-\lambda)f'(z)(y-x). \tag{1.7}$$

Finally, we multiply (1.6) and (1.7) by $1-\lambda$ and λ, respectively and sum them to obtain:

$$f(z) \leq (1-\lambda)f(x) + \lambda f(y).$$

Since $z = (1-\lambda)x + \lambda y$ the result follows from Proposition 32. □

Proposition 34. Let $\mathbb{E}$ be an Euclidean space, let $f\colon \mathbb{E} \to \mathbb{R}$ be a convex function. Then

$$\overline{f}(y) = \lim_{\lambda \uparrow 1} f((1-\lambda)x + \lambda y)$$

for each $x \in \text{ri}(\text{dom}(f))$ and $y \in \overline{\text{dom}(f)}$.

Proof. Let $x \in \text{ri}(\text{dom}(f))$ and $y \in \overline{\text{dom}(f)}$. We want to prove that

$$\overline{f}(y) \leq \liminf_{\lambda \uparrow 1} f((1-\lambda)x + \lambda y) \leq \limsup_{\lambda \uparrow 1} f((1-\lambda)x + \lambda y) \leq \overline{f}(y).$$

Since $f(y) \geq \overline{f}(y)$ by definition, we automatically conclude that

$$\liminf_{\lambda \uparrow 1} \overline{f}((1-\lambda)x + \lambda y) \leq \liminf_{\lambda \uparrow 1} f((1-\lambda)x + \lambda y).$$

Because $\overline{f}(y)$ lower semi-continuous:

$$\overline{f}(y) \leq \liminf_{\lambda \uparrow 1} \overline{f}((1-\lambda)x + \lambda y) \leq \liminf_{\lambda \uparrow 1} f((1-\lambda)x + \lambda y).$$

It remains to show the right-hand side of our inequality. Let $\beta \in \mathbb{R}$ with $\beta \geq \overline{f}(y)$ so that $(y, \beta) \in \text{epi}(\overline{f}) = \overline{\text{epi}(f)}$. Also let $\alpha > f(x)$ and note that $(x, \alpha) \in \text{ri}(\text{epi}(f))$ by Proposition 16. By Proposition 15, we know that $(1-\lambda)(x,\alpha) + \lambda(y,\beta) \in \text{ri}(\text{epi}(f))$ for all $\lambda \in [0,1)$. Thus, since f is convex:

$$f((1-\lambda)x + \lambda y) < (1-\lambda)\alpha + \lambda\beta \text{ for each } \lambda \in [0,1).$$

Hence,

$$\limsup_{\lambda \uparrow 1} f((1-\lambda)x + \lambda y) \leq \limsup_{\lambda \uparrow 1} (1-\lambda)\alpha + \lambda\beta = \beta.$$

In particular, setting $\beta = \overline{f}(y)$ yields the desired result. □

Proposition 35. Let $\{f_i\}_{i\in I} \subseteq \mathbb{E}^{\mathbb{R}}$ be a finite family of convex functions. If f_i is closed for each $i \in I$ and $\sum_{i\in I} f_i$ is proper, then $\sum_{i \in I} f_i$ is a closed convex function. Otherwise, if

$$\bigcap_{i\in I} \text{ri}(\text{dom}(f_i)) \neq \varnothing,$$

then $\overline{\sum_{i\in I} f_i} = \sum_{i\in I} \overline{f_i}$.

Proof. Consider $f := \sum_{i\in I} f_i$ and let $x \in \operatorname{ri}(\operatorname{dom}(f))$ which is equal to $\bigcap_{i\in I} \operatorname{ri}(\operatorname{dom}(f_i))$ by Proposition 18. This set is nonempty by hypothesis. By proposition 34,

$$\overline{f}(y) = \lim_{\lambda\uparrow 1} f(\lambda y+(1-\lambda)x) = \lim_{\lambda\uparrow 1} \sum_{i\in I} f_i(\lambda y+(1-\lambda x)) = \sum_{i\in I} \lim_{\lambda\uparrow 1} f_i(\lambda y+(1-\lambda)x) = \sum_{i\in I} \overline{f_i}(y).$$

If each f_i is closed, then $\overline{f_i} = f_i$ for each $i \in I$ and we have the result. □

Also, we remark that the effective domain of the sum of a family $\{f_i\}_{i\in I} \subseteq \mathbb{R}^{\mathbb{E}}$ of functions is the intersection $\bigcap_{i\in I} \operatorname{dom}(f_i)$.

Corollary 31 can be refined to cover open sets as well. In this case, one obtains that the closure of a convex set C is the intersection of the half-spaces that contain it. This result also enlightens the idea to describe a convex set C as the intersection of all the half-spaces that contain it. To translate this property to the language of convex functions, we define the *convex conjugate*, also known as *Fenchel conjugate*, of a convex function $f\colon \mathbb{E} \to \mathbb{R}$ as the function

$$f^*(y) = \sup_{x\in\mathbb{E}}\{\langle x, y\rangle - f(x)\} = -\inf_{x\in\mathbb{E}}\{f(x) - \langle x, y\rangle\}.$$

The epigraph of f^* can be thought as the intersection of the half-spaces in $\mathbb{E}\oplus\mathbb{R}$ that contain $\operatorname{epi}(f)$. The next proposition explores some of the basic properties of this operator.

Proposition 36. Let $\mathbb{E}$ be an Euclidean space and let $f, g\colon \mathbb{E} \to \mathbb{R}$ be closed convex functions. Then:

(i) f^* is convex;

(ii) f^* is closed;

(iii) $\langle x, y\rangle \le f^*(y) + f(x)$;

(iv) if $g(x) \le f(x)$ for each x, then $f^* \ge g^*$;

(v) $f^{**} = \overline{f}$.

Proof.

(i) We will show that if $h\colon \mathbb{E}\times\mathbb{E} \to \mathbb{R}$ is a convex function, then the function

$$\phi(x) := \sup_{y\in\mathbb{E}}\{h(x, y)\}, \text{ for each } x \in \mathbb{E}$$

is also convex.

Let $x, z \in \mathbb{E}$ and $\lambda \in [0, 1]$. It follows:

$$\begin{aligned}\lambda\phi(x) + (1-\lambda)\phi(z) &= \lambda \sup_{y\in\mathbb{E}}\{h(x, y)\} + (1-\lambda)\sup_{y\in\mathbb{E}}\{h(z, y)\}\\ &= \sup_{y\in\mathbb{E}}\{\lambda h(x, y)\} + \sup_{y\in\mathbb{E}}\{(1-\lambda)h(z, y)\}\\ &\ge \sup_{y\in\mathbb{E}}\{\lambda h(x, y) + (1-\lambda)h(z, y)\}\\ &\ge \sup_{y\in\mathbb{E}}\{h(\lambda x + (1-\lambda)z, y)\}\\ &= \phi(\lambda x + (1-\lambda)z).\end{aligned}$$

Considering $h(x, y) = \langle x, y\rangle - f(y)$, Proposition 32 yields the desired result.

(ii) Consider the family of functions indexed by $\mathbb{E}$:

$$\phi_x(y) \coloneqq \langle x, y\rangle - f(x).$$

Let $y \oplus t \in \operatorname{epi}(f^*)$. By definition,

$$t \geq f^*(y) = \sup_{x\in\mathbb{E}}\{\phi_x(y)\}.$$

Thus, we have that $f^*(y) \geq \phi_x(y)$ for each $x \in \mathbb{E}$. That is, $y \oplus t \in \operatorname{epi}(\phi_x)$ for each $x \in \mathbb{E}$. Hence, $y \oplus t \in \bigcap_{x\in\mathbb{E}} \operatorname{epi}(\phi_x)$. Since each of the functions ϕ_x is affine, $\operatorname{epi}(f^*)$ is closed by Proposition 10.

(iii) Let $y \in \mathbb{E}$. By definition,

$$\begin{aligned} f^*(y) &= \sup_{x\in\mathbb{E}}\{\langle x, y\rangle - f(x)\} \\ &\geq \langle x, y\rangle - f(x), \text{ for each } x \in \mathbb{E}. \end{aligned}$$

Therefore, $f^*(y) + f(x) \geq \langle x, y\rangle$ for each $x, y \in \mathbb{E}$.

(iv) Let $y \in \mathbb{E}$. By definition,

$$\begin{aligned} f^*(y) &= \sup_{x\in\mathbb{E}}\{\langle y, x\rangle - f(x)\} \\ &\geq \langle x, y\rangle - f(x), \text{ for each } x \in \mathbb{E} \\ &\geq \langle x, y\rangle - g(x), \text{ for each } x \in \mathbb{E}. \end{aligned}$$

The latter implies that $f^*(y) \geq \sup_{x\in\mathbb{E}}\{\langle x, y\rangle - g(x)\} = g^*(y)$.

(v) Let $x, y \in \mathbb{E}$. By Item (iii), $f(x) \geq \langle x, y\rangle - f^*(y)$. This implies that $f(x) \geq \sup_{y\in\mathbb{E}}\{\langle x, y\rangle - f^*(y)\} =$ for each $x \in \mathbb{E}$. Thus, $\operatorname{epi}(f) \subseteq \operatorname{epi}(f^{**})$. Conversely, let $(x \oplus f^{**}(x)) \in \operatorname{epi}(f^{**})$ and assume that $(x \oplus f^{**}(x)) \notin \operatorname{epi}(f)$.

In this case, by Theorem 25 there exist $a = a_1 \oplus a_2 \in (\mathbb{E} \oplus \mathbb{R}) \setminus \{0\}$ and $\beta \in \mathbb{R}$ such that $\langle x \oplus f^{**}(x), a\rangle > \beta$ and $\langle y \oplus t, a\rangle \leq \beta$ for each $x \oplus t \in \operatorname{epi}(f)$. Thus,

$$\langle x - y, a_1\rangle + (f^{**}(x) - t)a_2 \geq 0 \text{ for each } x \oplus t \in \operatorname{epi}(f).$$

If $a_2 > 0$, this is a contradiction since t is unbounded from above and y is arbitrary. If $a_2 = 0$, setting $y = x - \lambda a_1$ for sufficiently big λ yields the desired contradiction. □

We remark that the same method applied to prove (ii) could be reproduced in (i) and that item (v) becomes $f^{**} = \overline{f}$ if f is not closed. Also, it is important to highlight the example where we want to compute the conjugate of indicator function of a convex set C. In this case, the conjugate of $\delta(\cdot \,|\, C)$ is the *support* function of C and its formula is given by

$$\delta^*(y \,|\, C) = \sup_{x\in\mathbb{E}}\{\langle x, y\rangle - \delta(x \,|\, C)\} = \sup_{x\in C}\{\langle x, y\rangle\}.$$

Let $f\colon \mathbb{E} \to \mathbb{R}$ be a polyhedral function and consider the polyhedron $\operatorname{epi}(f) =: P \subseteq \mathbb{E}\oplus\mathbb{R}$. Let $C_1 = \{(x_i \oplus \mu_i)\}_{i\in[k]}$ and $C_2 = \{(x_i \oplus \mu_i)\}_{i\in[m]\setminus[k]}$ be finite subsets of $\mathbb{E} \oplus \mathbb{R}$ such that

$P = \operatorname{conv}(C_1) + \operatorname{cone}(C_2)$. Then, for each $y \in P$ we have:

$$y = \sum_{i\in[m]} \lambda_i(x_i \oplus \mu_i) = \sum_{i\in[m]} (\lambda x_i \oplus \lambda_i\mu_i).$$

For some $\lambda\colon [m] \to \mathbb{R}_+$ with $\sum_{i\in[k]} \lambda_i = 1$. Thus we can write:

$$f(y) = \inf\left\{ \sum_{i\in[m]} \lambda_i\mu_i : \sum_{i\in[m]} \lambda_i x_i = y \right\}.$$

This formula can be used to study the conjugate of a polyhedral function.

Proposition 37. Let $\mathbb{E}$ be an Euclidean space and let $f\colon \mathbb{E} \to \mathbb{R}$ be a polyhedral function. Then f^* is polyhedral.

Proof. By definition,

$$\begin{aligned}
f^*(y) &= \sup_{x\in\mathbb{E}}\{\langle x, y\rangle - f(x)\} \\
&= \sup_{x\in\mathbb{E}}\left\{ \langle x, y\rangle - \inf\left\{ \sum_{i\in I} \lambda_i\mu_i : \sum_{i\in[m]} \lambda_i x_i = x \right\}\right\} \\
&= \sup_{\lambda}\left\{ \langle \sum_{i\in[m]} \lambda_i x_i, y\rangle - \sum_{i\in[m]} \lambda_i\mu_i \right\} \\
&= \sup\left\{ \sum_{i\in[m]} \lambda_i(\langle x_i, y\rangle - \mu_i) \right\} \\
&= \sup\left\{ \sum_{i\in[k]} \lambda_i(\langle x_i, y_i\rangle - \mu_i) + \sum_{i\in[m]\setminus[k]} \lambda_i(\langle x_i, y_i\rangle - \mu_i) \right\}.
\end{aligned}$$

If $\langle x_i, y\rangle - \mu_i \geq 0$ for some $i \geq k+1$, the supremum above will be infinite because λ_i is unbounded above. Otherwise, the supremum will be attained when $\lambda_i = 1$ for i such that $(\langle x_i, y_i\rangle - \mu_i) = \max_{j\in[k]}\{(\langle x_j, y_j\rangle - \mu_j)\}$. Defining the polyhedron

$$P' \coloneqq \{x \in \mathbb{E} : \langle x_i, y_i\rangle \leq \mu_i \ \text{ for each } i \in [m] \setminus [k]\}$$

we can write:

$$f^*(y) = \begin{cases} \max_{i\in[k]}\{\langle x_i, y_i\rangle - \mu_i\}, & \text{if } y \in P' \\ +\infty, & \text{otherwise.} \end{cases} \qquad \square$$

1.3 Optimization Problems

Definition. An *optimization problem* is an ordered pair $P = (X, f)$, where X is a set and $f\colon X \to \overline{\mathbb{R}}$ is an extended-real-valued function. The problem P is more commonly denoted as

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in X. \end{aligned}$$

The set X is called *feasible region* and the function f is called *objective function*. The elements of X are called *feasible points* or *feasible solutions*; everything else is *infeasible*. The optimization problem P is *feasible* if $X \neq \varnothing$. Otherwise it is *infeasible*. The *objective value* of $x \in X$ is $f(x)$. The *optimal value* of P is $\inf_{x \in X} f(x) \in \overline{\mathbb{R}}$ if P is feasible. Otherwise it is $+\infty$. A feasible solution $\bar{x}$ is *optimal* if $f(\bar{x})$ is the optimal value of the problem, i.e, if $f(\bar{x}) \leq f(x)$, for every $x \in X$. If the optimal value of P is $-\infty$, the problem is *unbounded*.

When we write

$$\begin{aligned} \text{maximize} \quad & f(x) \\ \text{subject to} \quad & x \in X \end{aligned}$$

we are referring to the optimization problem $(X, -f)$ and we use the same terminology as above.

Definition. Let $P = (X, f)$ and $Q = (Y, g)$ be optimization problems. A *homomorphism* from P to Q is a function $\phi\colon X \to Y$ such that $g(\phi(x)) \leq f(x)$ for each $x \in X$. The problems P and Q are said to be *equivalent* if there exists a homomorphism from P to Q and vice-versa.

Proposition 38. Let $P = (X, f)$, $Q = (Y, g)$, and $S = (Z, h)$ be optimization problems. If $\phi\colon X \to Y$ be a homomorphism from P to Q and $\psi\colon Y \to Z$ be a homomorphism from Q to S, then $\psi \circ \phi\colon X \to Z$ is a homomorphism from P to S.

Proof. Since ψ is a homomorphism from Q to S, we have that $h(\psi(y)) \leq g(y)$ for each $y \in Y$. Consider $Y \supseteq Y' := \phi(X)$. Then, $h(\psi(y')) \leq g(y')$ for each $y' \in Y'$. The latter implies that $h(\psi(\phi(x))) \leq g(\phi(x))$ for every $x \in X$. Since ϕ is a homomorphism from P to Q, we have that $g(\phi(x)) \leq f(x)$ for each $x \in X$. Hence, $h(\psi(\phi(x))) \leq f(x)$ for each $x \in X$. Therefore, $\psi \circ \phi$ is a homomorphism from P to S. □

Proposition 39. Let $P = (X, f)$ and $Q = (Y, g)$ be optimization problems. If $\phi\colon X \to Y$ is a bijective function such that $g(\phi(x)) = f(x)$ for each $x \in X$, then $\phi^{-1}\colon Y \to X$ is a homomorphism from Q to P.

Proof. Let $y \in Y$. Since the function ϕ is bijective, there exists a unique $x \in X$ such that $x = \phi^{-1}(y)$. Then:

$$g(y) = f(x) = f(\phi^{-1}(y)) \geq f(\phi^{-1}(y)).$$

Thus, ϕ^{-1} a homomorphism from Q to P. □

Corollary 40. Let $P = (X, f)$ and $Q = (Y, g)$ be optimization problems. If there exists a bijective function $\phi\colon X \to Y$ such that $g(\phi(x)) = f(x)$ for each $x \in X$ then P and Q are equivalent.

Proof. Immediate from Proposition 39. □

Based on these results, we now show that our concept of equivalence between optimization problems is indeed an equivalence relation in the formal sense.

Proposition 41. Consider, for every optimization problems A and B,

$$A \sim B \text{ if, and only if there exists a homomorphism from A to B and vice-versa.}$$

Then $\sim$ is an equivalence relation.

Proof. Let $P = (X, f)$, $Q = (Y, g)$, and $S = (Z, h)$ be optimization problems.

(i) For reflexivity, consider the function $\phi\colon X \to X$ given by $f(x) = x$ for each $x \in X$. Then, we have that $f(\phi(x)) = f(x)$ for each $x \in X$. Thus, $P \sim P$ by Corollary 40.

(ii) Symmetry follows from definition.

(iii) For transitivity, assume that $P \sim Q$ and $Q \sim S$. Then, let ϕ_1, ϕ_2 be homomorphisms from P to Q and from Q to P, respectively. Similarly, let ψ_1, ψ_2 be a homomorphism from Q to S and from S to Q, respectively. By Proposition 38 it follows that $\phi_1 \circ \psi_1$ is a homomorphism from P to S and $\psi_2 \circ \phi_2$ is a homomorphism from S to P. Therefore, $P \sim S$. □

Finally, our next propositions present the advantageous features that our construction of equivalence between optimization problems produces.

Proposition 42. Let $P = (X, f)$ and $Q = (Y, g)$ be equivalent optimization problems. If either P or Q has finite optimal value $\alpha \in \mathbb{R}$, then α is the optimal value of both problems.

Proof. Consider the homomorphisms $\phi\colon X \to Y$ and $\psi\colon Y \to X$, which exist by hypothesis. With no loss of generality, assume that P has optimal value α. Assume that the optimal value β of Q is different from α. If $\beta > \alpha$, for each $\bar{x} \in X$ such that $f(\bar{x}) < \alpha + \frac{\beta-\alpha}{2}$ we have that $g(\phi(\bar{x})) \leq \alpha + \frac{\beta-\alpha}{2}$, and then β is greater then the optimal value of Q. If $\beta < \alpha$, for each $\bar{y} \in Y$ such that $g(\bar{y}) \leq \beta + \frac{\alpha-\beta}{2}$ we have that $f(\psi(\bar{y})) \leq \beta + \frac{\alpha-\beta}{2}$ and thus α is not the optimal value of P. Therefore, $\alpha = \beta$. □

Proposition 43. Let $P = (X, f)$ and $Q = (Y, g)$ be equivalent optimization problems. Then P has an optimal solution if, and only if Q has an optimal solution.

Proof. Consider the homomorphisms $\phi\colon X \to Y$ and $\psi\colon Y \to X$, which exist by hypothesis.

Let $\alpha \in \mathbb{R}$ be the optimal value of P and assume that there exists $x^* \in X$ such that $f(x^*) = \alpha$. By definition, $g(\phi(x^*)) \leq \alpha$. By Proposition 42, we know that $g(y) \geq \alpha$ for each $y \in Y$. Thus, $g(\phi(x^*)) = \alpha$. That is, $\phi(x^*)$ is an optimal solution in Q.

Similarly, if α is the optimal value of Q and there exists $y^* \in Y$ such that $g(y^*) = \alpha$, then $f(\psi(y^*)) \leq \alpha$. By Proposition 42, we have that $f(x) \geq \alpha$ for each $x \in X$. Thus, we conclude that $f(\psi(y^*)) = \alpha$. Thus, $\psi(y^*)$ ia an optimal solution for P. □

Proposition 44. Let $P = (X, f)$ and $Q = (Y, g)$ be equivalent optimization problems. Then P and Q have the same outcome. That is:

(i) P is infeasible if and only if Q is infeasible.

(ii) P is unbounded if and only if Q is unbounded.

(iii) P has finite optimal value and does not have optimal solution if and only if Q has finite optimal value and does not have optimal solution.

(iv) P has finite optimal value and optimal solution(s) if and only if Q has finite optimal value and optimal solution(s).

Proof. Let $\phi\colon X \to Y$ be a homomorphism from P to Q and let $\psi\colon Y \to X$ be a homomorphism from Q to P. We will show each of the items in our statement.

For the first item, assume that P is infeasible and Q is feasible. Thus, there exists $y \in Y$ and we have by definition that $\psi(y) \in X$. This contradicts the hypothesis that $X = \varnothing$. Hence, Q is infeasible. Clearly, the converse is proven using the exact same reasoning.

For the second item, assume that P is unbounded. Then, the set $L_n := \{x \in X : f(x) \leq n\}$ is nonempty for each $n \in \mathbb{N}$. Consider sequence $\{v_n\}_{n\in\mathbb{N}}$ such that $v_n \in L_n$ for each $n \in \mathbb{N}$ and note that $\lim_{n\to\infty} f(v_n) = -\infty$. Set $w_n := \phi(v_n)$ for each $n \in \mathbb{N}$ so that w_n is always feasible in Q. Observe that $\lim_{n\to\infty} g(w_n) = -\infty$ and thus Q is unbounded. Again, the converse is shown by the exact same argument.

The two remaining items follow immediately from Propositions 42 and 43. □

1.4 Probability Spaces

The following sections are aimed at briefly introducing some crucial concepts in probability theory. These concepts are essential for a full understanding of the mathematical basis underlying the applications that will be shown at the end of the last three chapters of this text. Our treatment is rather superficial; for a deeper look into to the role that measure theory plays in probability, we refer the reader to [10; 31].

Definition. Let Ω be a set. A *σ-field* in Ω is a collection $\mathcal{F}$ of subsets of Ω satisfying:

(i) $\Omega \in \mathcal{F}$;

(ii) $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$;

(iii) $\bigcup_{i\in I} A_i \in \mathcal{F}$ for each countable family $\{A_i\}_{i\in I} \subseteq \mathcal{F}$.

In measure theory, an ordered pair $(\Omega, \mathcal{F})$ where Ω is a set and $\mathcal{F}$ is a σ-field in Ω is called a *measurable space*. It is also important to remark that for every set Ω, the collection $\mathcal{P}(\Omega) := \{X : X \subseteq \Omega\}$ of all of its subsets is a σ-field in Ω.

Definition. Let Ω be a set and let $\mathcal{C}$ be a nonempty collection of subsets of Ω. The *σ-field generated* by $\mathcal{C}$ is defined as

$$\sigma(\mathcal{C}) := \bigcap\{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-field}, \mathcal{F} \supseteq \mathcal{C}\}.$$

Proposition 45. Let Ω be a set and let $\mathcal{C}$ be a nonempty collection of subsets of Ω. Then $\sigma(\mathcal{C})$ is a σ-field in Ω.

Proof. First, note that $\Omega \in \sigma(\mathcal{C})$ since by definition Ω belongs to all σ-fields in Ω. Then, let $A \in \sigma(\mathcal{C})$. By definition, $A \in \bigcap\{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-field}, \mathcal{F} \supseteq \mathcal{C}\}$, which means that A belongs to all the σ-fields $\mathcal{F}$ that contain $\mathcal{C}$. Then, it follows that A^c also belongs to all the mentioned σ-fields and thus $A^c \in \sigma(\mathcal{C})$. Finally, Let $\{A_i\}_{i\in I} \subseteq \sigma(\mathcal{C})$. Let $\mathcal{F}$ be a σ-field such that $\mathcal{F} \supseteq \mathcal{C}$. Then, $\{A_i\}_{i\in I} \subseteq \mathcal{F}$ and $\bigcup_{i\in I} A_i \in \mathcal{F}$. Since $\mathcal{F}$ was arbitrary, this proves that $\bigcup_{i\in I} A_i \in \sigma(\mathcal{C})$. □

Now that it has been proved that $\sigma(\mathcal{C})$ is a σ-field, we are able to say that $\sigma(\mathcal{C})$ is the smallest σ-field that contains $\mathcal{C}$. Moreover, the collection $\mathcal{B}(\mathbb{R}^k) := \sigma(\{\prod_{i\in[k]}(-\infty, y_i] : y \in \mathbb{R}^k\})$ is the *Borel σ-field* in $\mathbb{R}^k$.

Definition. Let Ω be a set and let $\mathcal{F}$ be a σ-field of subsets of Ω. A *probability measure* is a function $\mathbb{P}\colon \mathcal{F} \to \mathbb{R}$ such that:

(i) $\mathbb{P}(\Omega) = 1$;

(ii) $\mathbb{P}(A) \geq 0$ for each $A \in \mathcal{F}$;

(iii) $\mathbb{P}(\bigcup_{i\in I} A_i) = \sum_{i\in I} \mathbb{P}(A_i)$ for each countable pairwise disjoint family $\{A_i\}_{i\in I} \subseteq \mathcal{F}$.

Definition. Let Ω be a set, let $\mathcal{F}$ be a σ-field in Ω, and let $\mathbb{P}\colon \mathcal{F} \to \mathbb{R}$ be a probability measure. The ordered triple $(\Omega, \mathcal{F}, \mathbb{P})$ is a *probability space*. The set Ω is its *sample space* and its elements are called *sample points*. Moreover, the elements of $\mathcal{F}$ are called *events*.

Definition. A *sequence of events* is a family $\{A_i\}_{i\in I}$ of sets where $I = \mathbb{N}$. The sequence is *increasing* if $A_{n+1} \supseteq A_n$ for each $n \in \mathbb{N}$ and it is *decreasing* if $A_{n+1} \subseteq A_n$ for each $n \in \mathbb{N}$. In either case, the sequence is called *monotone*. Define $\liminf A_i := \bigcup_{i\in\mathbb{N}} \bigcap_{m\leq i} A_m$ and $\limsup A_i := \bigcap_{i\in\mathbb{N}} \bigcup_{m\leq i} A_m$. If $\liminf A_i = \limsup A_n =: A$ the sequence is *convergent* and A is called the *limit* of $\{A_i\}_{i\in I}$.

Proposition 46. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If $\{A_n\}_{n\in\mathbb{N}}$ is a monotone convergent sequence in $\mathcal{F}$. Then

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\lim_{n\to\infty} A_n).$$

Proof. With no loss of generality, we consider a monotonically increasing sequence $\{A_n\}_{n\in\mathbb{N}}$, i.e, $A_n \subseteq A_{n+1}$, for each $n \in \mathbb{N}$. Define $B_n := A_{n+1} \setminus A_n$, for every $n \in \mathbb{N}$ and see that the sequence $\{B_n\}_{n\in\mathbb{N}}$ is pairwise disjoint and, moreover, $\bigcup_{n\leq M} B_n = A_M$, for each $M \in \mathbb{N}$, which implies that

$$\lim_{M\to\infty} \bigcup_{n\leq M} B_n = \lim_{M\to\infty} A_M$$

and then $\bigcup_{n\in\mathbb{N}} B_n = A := \lim_{n\to\infty} A_n$. It follows:

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(\lim_{M\to\infty} A_M)\\
&= \mathbb{P}\Big(\lim_{M\to\infty} \bigcup_{n\leq M} B_n\Big)\\
&= \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} B_n\Big)\\
&= \sum_{n\in\mathbb{N}} \mathbb{P}(B_n)\\
&= \lim_{M\to\infty} \sum_{n\leq M} \mathbb{P}(B_n)\\
&= \lim_{M\to\infty} \mathbb{P}(A_M).
\end{aligned}$$

□

Proposition 47. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $A \in \mathcal{F}$ such that $\mathbb{P}(A) > 0$. Define

$$\mathbb{P}_A(B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}, \text{ for each } B \in \mathcal{F}.$$

Then $(A, \mathcal{F} \cap A, \mathbb{P}_A)$ is a probability space.

Proof. First, we prove that $\mathcal{F} \cap A$ is a σ-field in A. Note that $A \in \mathcal{F} \cap A$ since $A \in \mathcal{F}$ and $A \cap A = A$. Moreover, if $B \in \mathcal{F} \cap A$ there exists $C \in \mathcal{F}$ such that $B = C \cap A$. Since $\Omega \setminus C \in \mathcal{F}$ it follows that $A \setminus B = (\Omega \setminus C) \cap A \in \mathcal{F} \cap A$. Consider a countable family $\{B_i\}_{i\in I} \subseteq \mathcal{F} \cap A$. Then, there exists a family $\{C_i\}_{i\in I} \subseteq \mathcal{F}$ such that $B_i = C_i \cap A$ for each $i \in I$. Since $\bigcup_{i\in I} C_i \in \mathcal{F}$, it follows that $(\bigcup_{i\in I} C_i) \cap A = \bigcup_{i\in I}(C_i \cap A) \in \mathcal{F} \cap A$.

It remains to show that $\mathbb{P}_A\colon \mathcal{F} \cap A \to \mathbb{R}$ is a probability measure. The statement that $\mathbb{P}_A$ is non negative is trivial. Moreover,

$$\mathbb{P}_A(A) = \frac{\mathbb{P}(A \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A)}{\mathbb{P}(A)} = 1.$$

Finally, let $\{B_i\}_{i\in I} \subseteq \mathcal{F} \cap A$ be a countable pairwise disjoint countable family. Then, there exists a pairwise disjoint countable family $\{C_i\}_{i\in I} \subseteq \mathcal{F}$ such that $B_i = C_i \cap A$ for each $i \in I$. It follows:

$$\mathbb{P}_A\Big(\bigcup_{i\in I} B_i\Big) = \frac{\mathbb{P}((\bigcup_{i\in I} C_i \cap A))}{\mathbb{P}(A)} = \frac{\sum_{i\in I} \mathbb{P}(C_i \cap A)}{\mathbb{P}(A)} = \sum_{i\in I} \frac{\mathbb{P}(C_i \cap A)}{\mathbb{P}(A)} = \sum_{i\in I} \mathbb{P}_A(B_i). \quad \square$$

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $A, B \in \mathcal{F}$ be events where $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$. The events A and B are said to be *independent* if

$$\mathbb{P}(B \cap A) = \mathbb{P}(B)\mathbb{P}(A).$$

The latter holds if and only if $\mathbb{P}_B(A) = \mathbb{P}(A)$ and $\mathbb{P}_A(B) = \mathbb{P}(B)$. Furthermore, $\mathbb{P}_A(B)$ is called the *conditional probability* of B *given* A and we will adopt the more common notation $\mathbb{P}(B \,|\, A)$.

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(\Lambda, \mathcal{A})$ be a measurable space. A *random variable* in $(\Omega, \mathcal{F})$ is a function $X \colon \Omega \to \Lambda$ such that

$$A \in \mathcal{A} \text{ implies } X^{-1}(A) \in \mathcal{F}.$$

In the measure-theoretic context, functions that satisfy the definition above are called *measurable functions*.

Proposition 48. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $(\Lambda, \mathcal{A})$ be a measurable space, and let $X \colon \Omega \to \Lambda$ be a random variable in $(\Omega, \mathcal{F}, \mathbb{P})$. Define

$$\mathbb{P}_X(A) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbb{P}(X^{-1}(A)), \text{ for each } A \in \mathcal{A}.$$

Then $\mathbb{P}_X$ is a probability measure in $(\Lambda, \mathcal{A})$. That is, $(\Lambda, \mathcal{A}, \mathbb{P}_X)$ is a probability space.

Proof. We shall prove that the function $\mathbb{P}_X \colon \mathcal{A} \to \mathbb{R}$ is a probability measure.

Let $A \in \mathcal{A}$. If $A = \Lambda$ then $\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(\Lambda)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in \Lambda\}) = 1$. The fact that $P_X(A) \geq 0$ for each $A \in \mathcal{A}$ is trivially implied by its definition. To complete the proof, consider a countable pairwise disjoint family $\{A_i\}_{i\in I} \subseteq \mathcal{A}$. Then so is $\{X^{-1}(A_i)\}_{i\in I} \subseteq \mathcal{F}$. Whence,

$$\mathbb{P}_X\Big(\bigcup_{i\in I} A_i\Big) = \mathbb{P}\Big(X^{-1}\Big(\bigcup_{i\in I} A_i\Big)\Big) = \mathbb{P}\Big(\bigcup_{i\in I} X^{-1}\Big(A_i\Big)\Big) = \sum_{i\in I} \mathbb{P}(X^{-1}(A_i)) = \sum_{i\in I} \mathbb{P}_X(A_i). \quad \square$$

Proposition 49. Let $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ be a probability space for each $i \in [3]$. Let $X \colon \Omega_1 \to \Omega_2$ be a random variable in $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$. If $Y \colon \Omega_2 \to \Omega_3$ be a random variable in $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$. Then $Y \circ X \colon \Omega_1 \to \Omega_3$ is a random variable in $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$.

Proof. Let $A \in \mathcal{F}_3$. Since Y is a random variable, it follows that $Y^{-1}(A) \in \mathcal{F}_2$. Again, since X is a random variable it follows that $X^{-1} \circ Y^{-1}(A) \in \mathcal{F}_1$. Therefore, $Y \circ X$ is a random variable in $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$. □

Definition. Let $X \colon \Omega \to \Lambda$ be a random variable in some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A set $S \subseteq \Lambda$ is called the *support* of X if for every set $W \subseteq \Lambda$ such that $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in W\}) = 1$ we have $W \supseteq S$. Random variables with countable support are often called *discrete*.

Proposition 50. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space , let $(\Lambda_1, \mathcal{A}_1)$ and $(\Lambda_2, \mathcal{A}_2)$ be measurable spaces, and let $X \colon \Omega \to \Lambda_1$ and $Y \colon \Omega \to \Lambda_2$ be random variables. Then:

(i) The function $\mathbb{P}_{X,Y}\colon \sigma(\mathcal{A}_1 \times \mathcal{A}_2) \to \mathbb{R}$ given by

$$\mathbb{P}_{X,Y}(A_1, A_2) := \mathbb{P}(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in (A_1, A_2)\})$$

is a probability measure.

(ii) The function $\mathbb{P}^*_X\colon \Lambda_1 \to \mathbb{R}$ given by

$$\mathbb{P}^*_X(A_1) := \mathbb{P}_{X,Y}(A_1, \Lambda_2)$$

$\mathbb{P}_{X,Y}$ and $\mathbb{P}^*_X$ is a probability measure.

Proof.

(i) First we note that $\mathbb{P}_{X,Y}(\Lambda_1, \Lambda_2) = \mathbb{P}(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in (\Lambda_1, \Lambda_2)\}) = \mathbb{P}(\Omega) = 1$ and that the non-negativity $P_{X,Y}$ is implied by the non-negativity of $\mathbb{P}$. Finally, let $\{A_i\}_{i\in I} \subseteq \sigma(\mathcal{A}_1 \times \mathcal{A}_2)$ be a countable pairwise disjoint family. It follows:

$$\begin{aligned}\mathbb{P}_{X,Y}\Big(\bigcup_{i\in I} A_i\Big) &= \mathbb{P}\Big(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in \bigcup_{i\in I} A_i\}\Big) \\ &= \sum_{i\in I} \mathbb{P}(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A_i\}) \\ &= \sum_{i\in I} \mathbb{P}_{X,Y}(A_i).\end{aligned}$$

(ii) By item (i), it follows that $\mathbb{P}^*_X(\Lambda_1) = \mathbb{P}_{X,Y}(\Lambda_1, \Lambda_2) = 1$ and that $\mathbb{P}^*_X$ is non-negative. Let $\{A_i\}_{i\in I} \subseteq \mathcal{A}_1$. Then:

$$\mathbb{P}^*_X\Big(\bigcup_{i\in I} A_i\Big) = \mathbb{P}_{X,Y}\Big(\bigcup_{i\in I} A_i, \Lambda_2\Big) = \sum_{i\in I} \mathbb{P}_{X,Y}(A_i, \Lambda_2) = \sum_{i\in I} \mathbb{P}^*_X(A_i).$$

To show that $\mathbb{P}_X(A) = \mathbb{P}^*_X(A)$ for each $A \in \mathcal{A}_1$ it suffices to note that

$$\{\omega \in \Omega : X(\omega) \in A\} = \{\omega \in \Omega : (X(\omega), Y(\omega)) \in (A, \Lambda_2)\}. \quad \square$$

Now that it has been proved that $\mathbb{P}_X$ and $\mathbb{P}^*_X$ are identical, we adopt $\mathbb{P}_X$ as our standard notation. Moreover, note that we can define $\mathbb{P}^*_Y$ in the exact same way. The probability measures $\mathbb{P}_{X,Y}$ and $\mathbb{P}_X$ are the *joint probability measure* of X and Y and the *marginal probability measure* of X. Next, we introduce the *conditional probability measure* of X given Y.

Proposition 51. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space , let $(\Lambda_1, \mathcal{A}_1)$ and $(\Lambda_2, \mathcal{A}_2)$ be measurable spaces, and let $X\colon \Omega \to \Lambda_1$ and $Y\colon \Omega \to \Lambda_2$ be random variables. Let $A_2 \in \mathcal{A}_2$ such that $\mathbb{P}_Y(A_2) > 0$ and define $\mathbb{P}_X(\cdot \,|\, Y \in A_2)\colon \mathcal{A}_1 \to \mathbb{R}$ such that

$$\mathbb{P}_X(A_1 \,|\, Y \in A_2) := \frac{\mathbb{P}_{X,Y}(A_1, A_2)}{\mathbb{P}_Y(A_2)} \text{ for each } A_1 \in \mathcal{A}_1.$$

Then $\mathbb{P}_X(\cdot \,|\, Y \in A_2)$ is a probability measure.

Proof. We first note that $\mathbb{P}_X(\cdot \,|\, Y \in A_2)$ is non-negative by definition. Moreover,

$$\mathbb{P}_X(\Lambda_1 \,|\, Y \in A_2) = \frac{\mathbb{P}_{X,Y}(\Lambda_1, A_2)}{\mathbb{P}_Y(A_2)} = \frac{\mathbb{P}_Y(A_2)}{\mathbb{P}_Y(A_2)} = 1.$$

Finally, consider a countable pairwise disjoint family $\{A_i\}_{i\in I}$. It follows:

$$\mathbb{P}_X\Big(\bigcup_{i\in I} A_i \,|\, Y \in A_2\Big) = \frac{\mathbb{P}_{X,Y}((\bigcup_{i\in I} A_i), A_2)}{\mathbb{P}_Y(A_2)} = \frac{\sum_{i\in I}\mathbb{P}_{X,Y}(A_i, A_2)}{\mathbb{P}_Y(A_2)} = \sum_{i\in I}\mathbb{P}_X(A_i|Y\in A_2). \qquad \square$$

Obviously, $\mathbb{P}_Y(\cdot\,|\,X \in A_1)$ is defined in the exact same manner. Also, the results in Propositions 50 and 51 may easily be extended for a finite collection of random variables.

Until the end of this section we present some of the theory that involves the case where X outputs real vectors, or, in other words, X is *real-valued.*

Definition. Let $X\colon \Omega \to \mathbb{R}^k$ be a random variable. The function $F_X\colon \mathbb{R}^k \to [0,1]$ defined as

$$F_X(a) := \mathbb{P}(\{\omega\in\Omega : X(\omega)\leq a\}) = \mathbb{P}\Big(X^{-1}\Big(\prod_{i\in[k]}(-\infty, a_i]\Big)\Big), \quad \text{for each } a\in\mathbb{R}^k$$

is the *cumulative distribution function* of the random variable X.

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P}_1)$ and $(\Omega, \mathcal{F}, \mathbb{P}_2)$ be probability spaces. Let $X\colon \Omega \to \mathbb{R}^k$ and $Y\colon \Omega \to \mathbb{R}^k$ be random variables. We say that the random variables X and Y are *identically distributed* if $F_X(a) = F_Y(a)$ for each $a\in\mathbb{R}^k$.

Proposition 52. Let $X\colon\Omega\to\mathbb{R}^k$ be a random variable in some probability space $(\Omega,\mathcal{F},\mathbb{P})$. Then F_X satisfies:

(i) $\lim_{a\to\infty} F_X(a) = 1$;

(ii) $\lim_{a\to-\infty} F_X(a) = 0$;

(iii) F_X is monotonically increasing;

(iv) F_X is right continuous, i.e, for each $x\in\mathbb{R}^k$, $\lim_{a\downarrow x} F_X(a) = F_X(x)$.

Proof. For the first two items, it suffices to notice that

$$\lim_{a\downarrow-\infty} X^{-1}((-\infty,a]) = \varnothing \text{ and } \lim_{a\uparrow\infty} X^{-1}((-\infty,a]) = \mathbb{R}^k.$$

The third item follows from the fact that $a_1 \leq a_2$ implies $(-\infty, a_1] \subseteq (-\infty, a_2]$. The final item is shown using Propositions 46 and 48. Let $\{x_n\}_{n\in\mathbb{N}}$ be a decreasing sequence in $\mathbb{R}^k$ such that $x_n \downarrow x$. Now, consider the sets $\{A_n\}_{n\in\mathbb{N}} := \{a\in\mathbb{R}^k : a\leq x_n\}$. Since x_n is a decreasing sequence it is immediate that A_n is decreasing as well. Furthermore,

$$A_n = \bigcap_{i=1}^{n} A_i \text{ for each } n\in\mathbb{N}.$$

Therefore,

$$\{a\in\mathbb{R}^k : a\leq x\} = \bigcap_{n\in\mathbb{N}} A_n.$$

Thus,

$$\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty}\mathbb{P}_X((A_n)) = \mathbb{P}_X(\lim_{n\to\infty} A_n) = \mathbb{P}_X(\{a\in\mathbb{R}^k : a\leq x\}) = F_X(x). \qquad \square$$

Definition. Let $X\colon \Omega \to \mathbb{R}^k$ be a random variable in some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $S \subset [k]$. The *marginal cumulative distribution* of X_S is the function

$$F_{X_S}(a) := \lim_{a_{S^c} \to \infty} F_X(a), \text{ for each } a \in \mathbb{R}^k.$$

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X\colon \Omega \to \mathbb{R}^k$ be a random variable, let $S \subset [k]$, and let $B \in \mathcal{B}(\mathbb{R}^S)$ such that $\mathbb{P}_{X_S}(B) > 0$. The *conditional cumulative distribution* of X given $X_S \in B$ is the function

$$F_X(a \,|\, X_S \in B) := \frac{\mathbb{P}_X(X \leq a \cap X_S \in B)}{\mathbb{P}_X(X_S \in B)}, \text{ for each } a \in \mathbb{R}^k.$$

Two distinct coordinates i and j of X are *independent* if

$$F_{X_{\{i,j\}}}(a) = F_{X_i}(a)F_{X_j}(a), \text{ for each } a \in \mathbb{R}^k.$$

If the latter property hold for each distinct $i, j \in [k]$, then the coordinates of X are said to be *pairwise independent.*

Moreover, if

$$\prod_{i \in [k]} F_{X_i}(a_i) = F_X(a_i), \text{ for each } a \in \mathbb{R}^k$$

The coordinates of X are said to be *jointly independent.*

Similarly, if $S \subseteq [k]$ is such that

$$\{i, j\} \cap S = \varnothing \text{ and } F_{X_{\{i,j\}}}(a \,|\, X_S \in B) = F_{X_i}(a \,|\, X_S \in B)F_{X_j}(a \,|\, X_S \in B) \text{ for each } a \in \mathbb{R}^k$$

the coordinates of X are said to be *conditionally independent* given X_S.

Definition. Let $X\colon \Omega \to \mathbb{R}^k$ be a random variable in some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The *expected value* of X is defined as:

$$E(X) := \int_\Omega X d\mathbb{P} = \int_{\mathbb{R}^k} X d\mathbb{P}_X$$

and it is said to exist if and only if the integral exists.

A straightforward fact that arises from the latter definition is that if X and Y are identically distributed random variables, then they have the same expected value. This happens because $F_X(a) = F_Y(a)$ for each $a \in \mathbb{R}^k$ implies that $\mathbb{P}_X(B) = \mathbb{P}_Y(B)$, for every $B \in \mathcal{B}(\mathbb{R}^k)$.

In a measure-theoretic context, $E(X)$ is the *Lebesgue integral* of X. Next, we show that the expectation functional is linear. Then, the result known as the law of the unconscious statistician is going to be proved.

Proposition 53. Let $X\colon \Omega \to \mathbb{R}^k$ and $Y\colon \Omega \to \mathbb{R}^k$ be random variables in some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let α, $\beta \in \mathbb{R}$. Then,

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y).$$

Proof.

$$E(\alpha X + \beta Y) = \int_\Omega (\alpha X + \beta Y) d\mathbb{P} = \alpha \int_\Omega X d\mathbb{P} + \beta \int_\Omega Y d\mathbb{P} = \alpha E(X) + \beta E(Y)$$

□

Proposition 54. Let $X\colon \Omega \to \mathbb{R}^k$ be a random variable in some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $n \in \mathbb{N}$, and $g\colon \mathbb{R}^k \to \mathbb{R}^n$ be measurable function. Define $Y \coloneqq g(X)$. Then,

$$E(Y) = \int_\Omega g(X) d\mathbb{P}.$$

Proof. By definition we have that:

$$E(Y) = \int_{\mathbb{R}^k} g(X) d\mathbb{P}_X = \int_\Omega g(X(\omega)) d\mathbb{P}$$

□

The special case $g(x) = x$ gives us $E(X) = \int_{\mathbb{R}^k} x d\mathbb{P}_X$

Definition. Let X be a random variable. The *variance* of X is defined as:

$$\operatorname{Var}(X) \coloneqq E((X - E(X))^2)$$

Similarly to the expected value, the variance of X is said to exist if and only if the integral exists.

There is an important thing to note about the relation between the existence of the expected value and variance of a random variable. If the variance of a real-valued random variable is finite, then the expected value of X is finite. However, the converse may not be true. For instance, if X is a random variable having the Pareto distribution [1] with parameters α and x_m and $\alpha \in (1, 2]$, then, the expected value of X is finite but its variance does not exist.

Definition. Let $X\colon \Omega \to \mathbb{R}^k$ be a random variable in some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $B \in \mathcal{B}(\mathbb{R}^k)$ such that $\mathbb{P}_X(\cdot \mid X_S \in B) > 0$ and let $S \subset [k]$. The *conditional expectation* of $X|X_S$ is defined as:

$$E(X \mid X_S \in B) \coloneqq \int_\Omega (X \mid X_S \in B) d\mathbb{P}$$

Similarly, define the *conditional variance* of X given X_S as:

$$E((X - E(X \mid X_S \in B))^2 \mid X_S \in B)$$

As in the previous cases, these values are said to exist if and only if the appropriated integrals exist.

1.5 Learning from Data

As V.Vapnik defines in [35], statistical learning is "a theory that explores ways to estimate functional dependence from a given collection of data". As Vapnik himself concedes shortly after, this is a very general definition which covers important topics in statistical theory such as the pattern recognition, regression, and density estimation problems. Consequently, it is hard to apply an unified framework for all problems in learning theory. In the aforementioned book, the general model of a learning problem is described as an interplay between three abstract objects.

The first of them is named *generator* and is usually denoted as G. It considers a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X\colon \Omega \to \mathbb{R}^k$. Then, it independently generates a finite family $\{\omega_i\}_{i\in I} \subseteq \Omega$ according to the probability measure $\mathbb{P}$. Finally, the generator returns the collection $\{X(\omega_i)\}_{i\in I}$.

The second element of the learning problem is called *target operator* or *supervisor* and is denoted as T^*. It is a function that transforms each input $X(\omega_i)$ provided by G to some output y_i belonging to some space S that we will leave unspecified for now. It is presumed that such operator exists and does not change but is unknown. As in [35], it is considered that the outputs y_i are conditionally independent given the inputs and are distributed according to some probability measure.

The final ingredient of a learning problem is denominated *learning machine*. It observes a finite family of ordered pairs $\{(X(\omega_i), y_i)\}_{i\in I}$ of inputs generated by G and outputs returned by T^*. The trials $(X(\omega_i), y_i)$ are assumed to be pairwise independent and identically distributed according to a function $\mathbb{P}_{X,Y}$. The learning machine then constructs an approximation of T^* pursuing one of these two goals:

(i) Imitate the target operator T^*, i.e, given the generator G, predict the outputs returned by the supervisor in the best possible way with respect to a fixed criterion.

(ii) Identify the target operator, i.e, construct an operator which is close to the target with respect to a fixed criterion.

Despite the obvious similarity between the goals described above, they have some slight structural differences. For example, if the learning machine is eventually able to identify the target operator then it should return precise estimations for the outputs of the supervisor. However, the converse may not be true. For this reason, the identification problem is seen as more difficult by those who study learning theory.

From this point of view, the learning problem can be summarized as "a problem of choosing an appropriate function from a given set". We'll next introduce decision problems and their application for these tasks should arise naturally.

1.6 Decision-Theoretic Framework

In this section we present some small amount of theory that will be the key to connect statistical learning problems with the optimization theory developed so far. Our goal is to construct decision problems from the definition of an experiment and establish a parallel between them and optimization problems in the sense they were defined. We now define an experiment:

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The choice of a finite family $\{\omega_i\}_{i\in I} \subseteq \Omega$ is called an *experiment*. For each $i \in I$, the choice of ω_i is called the i-th *trial* of the experiment.

Consider a research group that must make some decision about some experiment. In the definition of a decision problem presented in [15], De Groot considers the case where the outcomes of the experiment are unknown. Now, we present this definition and also consider the case where the research group has given data to work on. However, we regard that in both cases it is assumed that there exists a well defined set Ω of all possible outcomes and that there is a probability measure $\mathbb{P}$ well defined in a convenient σ-field $\mathcal{F}$ in Ω. The reader should observe that at this point nothing new was defined since the objects just mentioned follow the standard definitions already given.

To model the possible decisions, assume that there exists a set $\mathcal{D}$ of all possible decisions d that can be made by the researchers, alike the set Ω just presented, it is supposed that $\mathcal{D}$ is defined along with a appropriate σ-field $\mathcal{A}$ of its subsets. It is also assumed that there exists some $\sigma(\mathcal{F} \times \mathcal{A})$-measurable function $L\colon \Omega \times \mathcal{D} \to \mathbb{R}_+$ that represents the reward obtained by the scientists for each pair (ω, d) consisting of a outcome ω of the experiment and a decision d made by the group. The function L is not unique and may be chosen accordingly to the preferences of the one(s) involved in the problem instance. Indeed, we refer to the seventh chapter of [15] for a detailed discussion about this topic.

Obviously, it is assumed that no matter who the scientists are, they want to get the best out of the experiment. However, the presence of randomness in the experiment is an inconvenience that must be handled concerning that the objective is to simply choose $d \in \mathcal{D}$ that minimizes the loss function. Towards this objective, the uncertain parameter ω must somehow be managed.

The most popular way to get around this uncertainty problem is to define, for each decision $d \in \mathcal{D}$, the associated *expected risk* $\rho(d)$ as the conditional expectation in Ω of the outcome given the decision d. Specifically, due to lemma 54, the formula to the risk $\rho(d)$ is given by

$$\rho(d) := \int_\Omega L(\omega, d) d\mathbb{P}.$$

To extend this definition to the case where the result of the experiment is given, we define the *empirical risk* as

$$\rho(d) := \sum_{i \in I} L(\omega_i, d).$$

Where $\{\omega_i\}_{i \in I}$ is the outcome of the experiment.

Based on the previous discussion, we now formally define a decision problem.

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $(\mathcal{D}, \mathcal{A})$ be a measurable space, and let $L \colon \Omega \times \mathcal{D} \to \mathbb{R}_+$ be a $\sigma(\mathcal{F} \times \mathcal{A})$-measurable function. A *decision problem* is the optimization problem $(\mathcal{D}, \rho)$, where ρ is a risk function obtained from L. The elements of $\mathcal{D}$ are called *decisions*. The decision problem is *feasible* if $\mathcal{D} \neq \varnothing$. Otherwise it is *infeasible*. The *risk of* d is $\rho(d)$. The *Bayes risk* of the optimization problem is $\inf_{d \in \mathcal{D}} \rho(d) \in \bar{\mathbb{R}}_+$. A decision d^* is a *Bayes decision* if $\rho(d^*)$ is the Bayes risk, i.e, if $\rho(d^*) \leq \rho(d)$ for each $d \in \mathcal{D}$. The function L is called *loss function* and the set $\mathcal{D}$ is named *decision space*.

This correspondence between the definitions of optimization and decision problems leads to some kind of symmetry between their terminology. For example, pairs like Bayes risk and optimal value and Bayes decision and optimal solution share the same meaning. In some sense, a decision problem is an optimization problem with a random component. In the case where the result of the experiment is unknown, the randomness in the problem is handled by applying the expectation functional to the random component of the loss function. When the result of the experiment is known, we use it in order to calculate the risk function ρ.

While applying the described framework to learning problems, the previously mentioned research group plays the role of the learning machine in one of the following manners:

To relate prediction and decision problems, consider the experiment of observing finitely many pairs $\{(X(\omega_i), y_i)\}_{i \in I}$ taking values in some set $\mathbb{R}^k \times S$ and the decision of choosing the best function, with respect to a fixed loss $L \colon (\mathbb{R}^k \times S) \times \mathcal{D} \to \mathbb{R}_+$ to predict a new value of y given a new value of X. Without any further assumption, the decision space is considered to be the set of all functions d from $\mathbb{R}^k$ to S. In this context, the optimal decision d^* is chosen based on the minimization the empirical risk function obtained from L. We note that this setting is fine with the definition of a decision problem because of Proposition 48.

On the other hand, we will relate imitation and decision problems by considering a finite collection $\{X(\omega_i)\}_{i \in I}$ taking values in some set S. Then, we aim to attribute probabilities to each element s os S in a 'realistic' way. The disparity between the 'real' probability of each element s and our choice is assumed to be represented by some loss function $L \colon S \times \mathcal{D} \to \mathbb{R}_+$, that we want to minimize. Identically to the former paragraph, Proposition 48 allows one to consider S as the random component of the domain of L.

Chapter 2

Conic Optimization and Duality

In Chapter 1, we defined an optimization problem as an ordered pair (S, f) where S is a set and f is an extended-real valued function. This is a truly general definition and it gives no clue for us to actually solve the problem. Depending on the peculiarities of S and f, this task can be extremely difficult or even impossible. So mathematicians focus on special classes of problems that have some interesting properties such as duality theories, algorithms to find exact or approximate solutions, useful applications, and so on. Attention will now be turned to conic programming, which has all these features.

In this context, we are going to deal with the optimization of a linear functional over the intersection of a cone and an affine subspace of the ambient space. This chapter starts with some basic definitions and results that show how to construct convex cones as epigraphs of real-valued functions. Then, we use the recession cone of a convex set C to investigate conditions that imply that the image of C under some linear transformation is closed. Next, we explore some properties of the dual cone of a convex set. Later, we will use proper cones to construct a partial order on its ambient space. Further, we will define conic optimization problems and their duals, and, finally, show some capital results to understand the relation between these pairs of problems.

2.1 Conic Programming Scenario

Definition. Let $\mathbb{E}$ be an Euclidean space. A set $K \subseteq \mathbb{E}$ is a *cone* if for each $x \in K$ and $\alpha \in \mathbb{R}_{++}$ we have $\alpha x \in K$. That is, K is a cone if $\alpha K \subset K$ for each $\alpha \in \mathbb{R}_{++}$. A cone is *polyhedral* if it is a polyhedron.

Proposition 55. Let $\mathbb{E}$ be an Euclidean space, let $K \subseteq \mathbb{E}$ be a cone and let $f \colon K \to \mathbb{R}$ be a function. Then $\operatorname{epi}(f)$ is a cone if and only if f is positively homogeneous.

Proof. First assume that f is positively homogeneous. Let $x \oplus t \in K \oplus \mathbb{R}$ and let $\alpha \in \mathbb{R}_{++}$. It follows:

$$\begin{aligned} x \oplus t \in \operatorname{epi}(f) &\iff f(x) \leq t \\ &\iff \alpha f(x) = f(\alpha x) \leq \alpha t \\ &\iff \alpha x \oplus \alpha t \in \operatorname{epi}(f). \end{aligned}$$

Conversely, if $\operatorname{epi}(f)$ is a cone then $x \oplus t \in \operatorname{epi}(f)$ implies that $\alpha x \oplus \alpha t \in \operatorname{epi}(f)$ for each $\alpha \in \mathbb{R}_{++}$. For each $x \in K$, it is clear that $x \oplus f(x) \in \operatorname{epi}(f)$. Thus, $\alpha x \oplus \alpha f(x) \in \operatorname{epi}(f)$ for every $\alpha \in \mathbb{R}_{++}$. So it follows that $f(\alpha x) \leq \alpha f(x)$ for each $x \in K$ and $\alpha \in \mathbb{R}_{++}$. On the other hand, we also have that $\alpha x \oplus f(\alpha x) \in \operatorname{epi}(f)$. Then, $x \oplus \frac{f(\alpha x)}{\alpha} \in \operatorname{epi}(f)$. That is, $\alpha f(x) \leq f(\alpha x)$. Combining these two inequalities yields $\alpha f(x) \leq f(\alpha x) \leq \alpha f(x)$. Therefore, f is positively homogeneous. $\square$

Proposition 56. Let $\mathbb{E}$ be an Euclidean space and let $\varnothing \neq K \subseteq \mathbb{E}$ be a closed cone. Then $0 \in K$.

Proof. Let $x \in K$. By definition it is true that $\alpha x \in K$ for each $\alpha \in \mathbb{R}_{++}$. In this case, for each $\alpha \in \mathbb{R}_{++}$ we have that $\alpha\mathbb{B} \cap K \neq \varnothing$. Thus, 0 is a accumulation point of K. Since K is closed, we have that $0 \in K$. □

Proposition 57. Let $\mathbb{E}$ be an Euclidean space and $K \subseteq \mathbb{E}$ a cone. Then, K is convex if and only if it is closed under addition.

Proof. First suppose K is closed under addition. Let $x, y \in K$ then choose $\alpha \in [0,1]$. Since K is a cone we have $\alpha x \in K$ and $(1-\alpha)y \in K$. Then, $\alpha x + (1-\alpha)y \in K$.

Conversely, assume K is convex and let $x, y \in K$. We have $\frac{1}{2}(x+y) \in K$ and again, the definition of a cone gives us that $x + y \in K$. □

Corollary 58. Let $\mathbb{E}$ be an Euclidean space and let $K \subseteq \mathbb{E}$ be a cone. If $f\colon K \to \mathbb{R}$ is a positively homogeneous function, then f is convex if and only if:

$$f(x+y) \leq f(x) + f(y), \text{ for each } x, y \in K.$$

Proof. Immediate from Propositions 55 and 57. □

Recession Cones and Closedness of Convex Linear Images

Definition. Let $\mathbb{E}$ be an Euclidean space and let $\varnothing \neq C \subseteq \mathbb{E}$ be a convex set. The *recession cone* of C is the set:

$$0^+C \coloneqq \{y \in \mathbb{E} : x + \alpha y \in C \text{ for each } x \in C \text{ and } \alpha \in \mathbb{R}_+\}.$$

Moreover, the subspace $\operatorname{lin}(C) \coloneqq 0^+C \cap -(0^+C)$ is the *lineality space* of C and C is *pointed* if $\operatorname{lin}(C) = \{0\}$.

Proposition 59. Let $\mathbb{E}$ be an Euclidean space and let $f\colon \mathbb{E} \to \mathbb{R}$ be positively homogeneous a function. Then, $\operatorname{epi}(f)$ is pointed if, and only if

$$f(x) \leq t \text{ implies } f(-x) > t, \text{ for each } 0 \neq (x \oplus t) \in \mathbb{E} \oplus \mathbb{R}.$$

Proof. To prove this result, it suffices to note that:

$$\begin{aligned}
\operatorname{epi}(f) \text{ is pointed} &\iff \operatorname{epi}(f) \cap (-\operatorname{epi}(f)) = \{0\} \\
&\iff (x \oplus t) \in \operatorname{epi}(f) \text{ and } (x \oplus t) \neq 0 \text{ implies } (-x \oplus -t) \notin \operatorname{epi}(f) \\
&\iff f(x) \leq t \text{ and } x \neq 0 \text{ implies } f(-x) > t.
\end{aligned}$$

□

Proposition 60. Let $\mathbb{E}$ be an Euclidean space and let $\varnothing \neq C \subseteq \mathbb{E}$ be a closed convex set. Then:

(i) 0^+C is a closed convex cone;

(ii) 0^+C=$\{0\}$ if and only if C is bounded;

(iii) For each $y \in \mathbb{E}$, if there exists $x \in C$ such that $x + \lambda y \in C$ whenever $\lambda \in \mathbb{R}_+$, then $y \in 0^+C$;

(iv) $0^+C = 0^+\operatorname{ri}(C)$ and $\operatorname{lin}(C) = \operatorname{lin}(\operatorname{ri}(C))$.

Proof.

(i) By definition, $x \in 0^+C$ implies that $y + \gamma x \in C$ for fixed $y \in C$ and $\gamma \in \mathbb{R}_+$. Hence, if $\alpha \in \mathbb{R}_{++}$, then $y + \frac{\gamma}{\alpha}(\alpha x) \in C$. This implies that $\alpha x \in 0^+C$ and then 0^+C is a cone. To prove that 0^+C is convex, let $x_1, x_2 \in 0^+C$ and fix $\lambda \in [0,1]$. Then, for any $\alpha \in \mathbb{R}_{++}$ and $y \in C$, we have that

$$\alpha((1-\lambda)x_1 + \lambda x_2)) + y = (1-\lambda)(y + \alpha x_1) + \lambda(y + \alpha x_2)$$

belongs to C by convexity. Finally, let $x \in \overline{0^+C}$. By definition, $(x + \varepsilon\mathbb{B}) \cap 0^+C \neq \varnothing$ for each $\varepsilon \in \mathbb{R}_{++}$. Define $\varepsilon_k := \frac{1}{k}$ for every $k \in \mathbb{Z}_{++}$ and let $x_k \in (x + \varepsilon_k\mathbb{B}) \cap 0^+C$ so that x is the limit of the sequence $\{x_k\}_{k \in \mathbb{Z}_{++}}$. Because x_k is always in 0^+C, it follows that

$$y + \gamma x_k \in C, \text{ for each } y \in C \text{ and } \gamma \in \mathbb{R}_+.$$

Since C is closed and the limit of $\{x_k\}$ is x, we obtain that $y + \lambda x \in C$ whenever $\lambda \in \mathbb{R}_+$. This implies by definition that $x \in 0^+C$. Since the converse is trivial, we conclude that $\overline{0^+C} = 0^+C$. That is, 0^+C is closed.

(ii) If C is bounded, then there exists $M \in \mathbb{R}_{++}$ such that $C \subseteq M\mathbb{B}$. Assume that there exists $0 \neq y \in 0^+C$. Then, we have that $x + \lambda y \in C$ for every $\lambda \in \mathbb{R}_+$. On the other hand:

$$\|x + \lambda y\|^2 = \|x\|^2 + 2\lambda\langle x, y\rangle + \|y\|^2.$$

Thus, for $\lambda \geq \frac{M^2}{2\langle x,y\rangle}$ we have that $\|x + \lambda y\| > M$, which contradicts our assumption that $y \in 0^+C$ since $C \subseteq M\mathbb{B}$.

Conversely, Assume that $0^+C \neq \{0\}$ and we will show that C is not bounded. Let $0 \neq y \in 0^+C$. Then, for each $x \in C$ and $\lambda \in \mathbb{R}_+$ we have that $x + \lambda y \in C$. Let $M \in \mathbb{R}_{++}$. We already know that if $\lambda \geq \frac{M^2}{2\langle x,y\rangle}$ then $\|x + \lambda y\| > M$. That is, there is no $M \in \mathbb{R}_{++}$ such that $C \subseteq M\mathbb{B}$. Therefore, C is not bounded.

(iii) Let $0 \neq y \in \mathbb{E}$ such that there exists $x \in C$ such that $x + \alpha y \in C$ for each $\alpha \in \mathbb{R}_+$ and assume that there is some $x' \neq x$ such that there exists $\alpha' \in \mathbb{R}_+$ such that $x' + \alpha y \notin C$ for $\alpha \geq \alpha'$. Fix $\lambda > \alpha'$ so that $z := x' + \lambda y \notin C$. Then, by Theorem 25 there exists $a \in \mathbb{E} \setminus \{0\}$ and $\beta \in \mathbb{R}$ such that $\langle a, z\rangle > \beta$ and $\langle a, v\rangle \leq \beta$ for each $v \in C$.

Since $x + \alpha y \in C$ whenever $\alpha \in \mathbb{R}_+$, we have that

$$\langle x + \alpha y, a\rangle = \langle x, a\rangle + \alpha\langle y, a\rangle \leq \beta \text{ for each } \alpha \in \mathbb{R}_+.$$

Thus, $\langle y, a\rangle \leq 0$. On the other hand, we have that $\langle x', a\rangle \leq \beta$ because $x' \in C$. Then:

$$\langle x' + \lambda y, a\rangle = \langle x', a\rangle + \lambda\langle a, y\rangle \leq \beta.$$

Which is a contradiction.

(iv) Let $y \in 0^+\mathrm{ri}(C)$. By definition we have that $x + \lambda y \in C$ whenever $\lambda \in \mathbb{R}_+$ and $x \in \mathrm{ri}(C)$. Since $\mathrm{ri}(C) \subseteq C$, the result follows. Conversely, let $y \in 0^+C$. In particular, we have that $x + \lambda y \in C$ for each $x \in \mathrm{ri}(C)$ and $\lambda \in \mathbb{R}_+$. Since $x \in \mathrm{ri}(C)$, it follows from Proposition 15 that $x + \lambda y$ actually lies in $\mathrm{ri}(C)$. Thus, $y \in 0^+\mathrm{ri}(C)$. Finally, we note that

$$\mathrm{lin}(\mathrm{ri}(C)) = 0^+\mathrm{ri}(C) \cap -(0^+\mathrm{ri}(C)) = 0^+C \cap -(0^+C) = \mathrm{lin}(C). \quad \square$$

There are several possible corollaries from Proposition 60. In particular, if K is a closed convex cone, then $0^+K = K$. Thus, K is pointed if and only if $K \cap -K = \{0\}$. In some sense, the recession cone of a convex set C can be understood as the largest closed convex cone contained in C.

Proposition 61. Let $\mathbb{E}$ be an Euclidean space and let $\{C_i\}_{i\in I}$ be a family of closed convex sets such that $\bigcap_{i\in I} C_i \neq \varnothing$. Then

$$0^+\bigcap_{i\in I} C_i = \bigcap_{i\in I} 0^+C_i, \text{ and } \operatorname{lin}(\bigcap_{i\in I} C_i) = \bigcap_{i\in I} \operatorname{lin}(C_i).$$

Proof. Let x belong to the left-hand side of the first equation. By definition, for each $y \in \bigcap_{i\in I} C_i$ we have that $x + \lambda y \in \bigcap_{i\in I} C_i$ for every $\lambda \in \mathbb{R}_{++}$. From Proposition 60, we conclude that $x \in 0^+C_i$ for each $i \in I$. That is, $x \in \bigcap_{i\in I} 0^+C_i$. Conversely, let x belong to the right-hand side. This means that $x \in 0^+C_i$ for each $i \in I$. Let $y \in \bigcap_{i\in I} C_i$, which is nonempty by hypothesis. Then, $y + \lambda x \in C_i$ for each $\lambda \in \mathbb{R}_+$ and $i \in I$. This means that $y + \lambda x \in \bigcap_{i\in I} C_i$ for each $\lambda \in \mathbb{R}_+$, or equivalently $x \in 0^+\bigcap_{i\in I} C_i$. Finally, we note that the result for $\operatorname{lin}(C)$ follows immediately by associativity. □

Proposition 62. Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces, let $A\colon \mathbb{E} \to \mathbb{Y}$ be a linear function, let $C \subseteq \mathbb{E}$ and $V \subseteq \mathbb{Y}$ both be closed convex sets such that V is compact. If $Z \coloneqq C \cap A^{-1}(V)$ is nonempty, then Z is closed, convex, and we have $0^+Z = 0^+C \cap \operatorname{Null}(A)$ and $\operatorname{lin}(Z) = \operatorname{lin}(C) \cap \operatorname{Null}(A)$. Moreover, Z is compact if and only if $0^+C \cap \operatorname{Null}(A) = \{0\}$.

Proof. First, note that Z is closed and convex by Propositions 13 and 10, respectively. Now, we prove that $\operatorname{Null}(A) = 0^+A^{-1}(V)$. The inclusion $\operatorname{Null}(A) \subseteq 0^+A^{-1}(V)$ is obvious. Conversely, if there is $y \in 0^+A^{-1}(V)$ such that $A(y) \neq 0$. Then, $A(x + \lambda y) = A(x) + \lambda A(y) \in V$ for each $\lambda \in \mathbb{R}_+$ and $x \in A^{-1}(V)$, which implies that V is not bounded. Whence, $\operatorname{Null}(A) = 0^+A^{-1}(C)$. Then, by Proposition 61 it follows that $0^+Z = 0^+C \cap \operatorname{Null}(A)$ and $\operatorname{lin}(Z) = \operatorname{lin}(C) \cap \operatorname{Null}(A)$. Moreover, by Proposition 60, Z is compact if and only if $0^+C \cap \operatorname{Null}(A) = \{0\}$. □

Proposition 63. Let $\mathbb{E}$ be an Euclidean space, let $\varnothing \neq C \subseteq \mathbb{E}$ be a convex set, and let $S \subseteq \operatorname{lin}(C)$ be a subspace of $\mathbb{E}$. Then

$$C = S + (C \cap S^\perp).$$

Proof. To show that $S + (C \cap S^\perp) \subseteq C$, let $x \coloneqq y + z$ for some $y \in S$ and $z \in C \cap S^\perp$. Since $S \subseteq \operatorname{lin}(C)$ we have that $y \in \operatorname{lin}(C)$. Therefore, $y + z = x \in C$ because $z \in C$. Conversely, since $\mathbb{E} = S + S^\perp$ and $C \subseteq \mathbb{E}$, we can write any $x \in C$ as $y + z$, where $y \in S$ and $z \in S^\perp$. We know that $y \in \operatorname{lin}(C)$. Using that $x \in C$, we obtain $x - y = z \in C$. Hence, $z \in C \cap S^\perp$, which implies that $C \subseteq S + (C \cap S^\perp)$. □

Proposition 64. Let $\mathbb{E}$ be an Euclidean space and let $\{C_i\}_{i\in\mathbb{N}} \subseteq \mathbb{E}$ be a monotonically decreasing sequence of nonempty, closed and convex sets. Consider:

$$R \coloneqq \bigcap_{i\in\mathbb{N}} 0^+C_i, \text{ and } L \coloneqq \bigcap_{i\in\mathbb{N}} \operatorname{lin}(C_i).$$

If $R = L$, then $\bigcap_{i\in\mathbb{N}} C_i$ is nonempty.

Proof. First, note that the sequence $\{\operatorname{lin}(C_i)\}_{i\in\mathbb{N}}$ is monotonically decreasing and that $\operatorname{lin}(C_i)$ is a subspace of $\mathbb{E}$ for each $i \in \mathbb{N}$. Thus, there exists $i_0 \in \mathbb{N}$ such that $\operatorname{lin}(C_i) = L$ whenever $i \geq i_0$.

Now, we shall prove that there exists i_1 such that we have $0^+C_i \cap L^\perp = \{0\}$ for $i \geq i_1$. Assume that $0^+C_i \cap L^\perp \neq \{0\}$. Then, there exists $0 \neq x \in 0^+C_i \cap L^\perp$. Since 0^+C_i is a cone and $L^\perp$ is a subspace, we may assume that $\|x\| = 1$. So we conclude that $0^+C_i \cap L^\perp \cap \mathbb{B}_= \neq \varnothing$ for each $i \in \mathbb{N}$. Since this set is compact for each $i \in \mathbb{N}$, we know that

$$\bigcap_{i \in \mathbb{N}} 0^+C_i \cap L^\perp \cap \mathbb{B}_= = R \cap L^\perp \cap \mathbb{B}_= \neq \varnothing.$$

Because $L = R$, we obtain that $L \cap L^\perp \cap \mathbb{B}_= \neq \varnothing$. This is a contradiction since $L \cap L^\perp = \{0\}$.

Fix $i \geq \max\{i_0, i_1\}$, we have just shown that:

$$0^+C_i \cap L^\perp = \{0\}.$$

From Proposition 61,

$$0^+(0^+C_i \cap L^\perp) = 0^+C_i \cap L^\perp = \{0\}.$$

Thus, Proposition 60 implies that $D_i := C_i \cap L^\perp$ is compact. Then $\cap_{j \geq i} D_j$ is nonempty, which gives us that $\bigcap_{i \in \mathbb{N}} C_i \neq \varnothing$ because $D_i \subseteq C_i$ for each $i \geq \max\{i_0, i_1\}$. □

Proposition 65. Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces, let $A \colon \mathbb{E} \to \mathbb{Y}$ be a linear function, and let $\varnothing \neq C \subseteq \mathbb{E}$ be a closed convex set. If $0^+C \cap \operatorname{Null}(A) \subseteq \operatorname{lin}(C)$, then $A(C)$ is closed.

Proof. Let $x \in \overline{A(C)}$. By definition, for each $\varepsilon \in \mathbb{R}_{++}$ we have $x + \varepsilon\mathbb{B} \cap A(C) \neq \varnothing$. For each $k \in \mathbb{Z}_{++}$, consider $\varepsilon_k := \frac{1}{k}$ and $x_k \in x + \varepsilon_k\mathbb{B}$. Observe that the limit of the sequence $\{x_k\}_{k \in \mathbb{Z}_{++}}$ is x and that $x_k \in x + \bigcap_{i \leq k} \varepsilon_i \mathbb{B}$. Consider, for each $k \in \mathbb{Z}_{++}$:

$$C_k := \{y \in C : A(y) \in x + \varepsilon_k\mathbb{B}\} = C \cap A^{-1}(x + \varepsilon_k\mathbb{B}).$$

Note that $x_k \in C_k$ for each $k \in \mathbb{Z}_{++}$ and thus C_k is always nonempty. Moreover,

$$\bigcap_{k \in \mathbb{Z}_{++}} C_k = \{y \in C : A(y) \in x + \varepsilon_k\mathbb{B},\ \text{for each } k \in \mathbb{Z}_{++}\} = \{y \in C : A(y) = x\}.$$

Thus, it suffices to show that $\bigcap_{k \in \mathbb{Z}_{++}} C_k$ is nonempty. By Proposition 62 we have that, $0^+C_k = 0^+C \cap \operatorname{Null}(A)$, that $\operatorname{lin}(C_k) = \operatorname{lin}(C) \cap \operatorname{Null}(A)$, and that C_k is closed and convex for each $k \in \mathbb{Z}_{++}$. Also, by hypothesis we know that $0^+C \cap \operatorname{Null}(A) \subseteq \operatorname{lin}(C)$, which implies that

$$0^+C \cap \operatorname{Null}(A) \subseteq \operatorname{lin}(C) \cap \operatorname{Null}(A).$$

Since the converse is always true, we conclude that these sets are actually equal. Thus, applying Proposition 64 we obtain that $\bigcap_{k \in \mathbb{Z}_{++}} C_k \neq \varnothing$ and thus $x \in A(C)$. Therefore $\overline{A(C)} \subseteq A(C)$. That is, $A(C)$ is closed. □

Corollary 66. Let $\mathbb{E}$ be an Euclidean space, let $\{C_i\}_{i \in I} \subseteq \mathbb{E}$ be a finite family of nonempty convex sets. If $x_i \in C_i$ for each $i \in I$ and $\sum_{i \in I} x_i = 0$ implies that $x_i \in \operatorname{lin}(C_i)$ for each $i \in I$, then $\sum_{i \in I} \overline{C_i} = \overline{\sum_{i \in I} C_i}$.

Proof. Consider the Euclidean space $\mathbb{E}^I$ and the linear transformation $A \colon \mathbb{E}^I \to \mathbb{E}$ where $A(x) = \sum_{i \in I} x_i$. Then $\operatorname{Null}(A) = \{x \in \mathbb{E}^I : \sum_{i \in I} x_i = 0\}\}$. Moreover, if $C := \bigoplus_{i \in I} C_i$, then trivially $0^+C = \bigoplus_{i \in I} 0^+C_i$ and, needless to say, $\operatorname{lin}(C) = \bigoplus_{i \in I} \operatorname{lin} C_i$. Thus, our hypothesis implies that $0^+C \cap \operatorname{Null}(A) \subseteq \operatorname{lin}(C)$. Therefore, Proposition 65 yields the desired result. □

Definition. Let $\mathbb{E}$ be an Euclidean space and $K \subseteq \mathbb{E}$ be a cone. If K is closed, convex, pointed, and has nonempty interior then K is a *proper* cone.

Dual Cones and Their Properties

Definition. Let $\mathbb{E}$ be an Euclidean space and let $\varnothing \neq K \subseteq \mathbb{E}$. The *dual cone* of K is the set $K^* := \{\, y \in \mathbb{E} : \langle x, y\rangle \geq 0 \text{ for every } x \in K \,\}$.

Writing K^* as

$$\bigcap_{x\in K} \{\, y \in \mathbb{E} : \langle x, y\rangle \geq 0\}$$

makes it clear that K^* is an intersection of half-spaces. Thus, K^* is always closed and convex. Moreover, the homogeneity of the inner-product gives us that K^* is a cone even when K is not. Also, if K is a convex cone, then the support function of K is given by:

$$\delta^*(x \,|\, K) = \begin{cases} 0, \text{ if } \langle x, y\rangle \leq 0 \text{ for each } y \in K; \\ +\infty \text{ otherwise.} \end{cases}$$

Which is the same as $\delta(\cdot \,|\, -K^*)$. Together with Proposition 37, this implies that the dual cone of a polyhedral cone is polyhedral. The set $-K^*$ is commonly called the *polar cone* of K. We now present basic propositions on the basic properties of dual cones. Then, we will show two operations on proper cones and figure how these relate with duality. Finally, we prove some theorems of alternative relating convex cones with their respective duals. When combined with Corollary 66, these results will culminate in Proposition 79, which is essential to our presentation of conic optimization duality.

Proposition 67. Let $\mathbb{E}$ be an Euclidean space and $C, S \subseteq \mathbb{E}$, . If $C \subseteq S$, then $S^* \subseteq C^*$.

Proof. Let $x \in S^*$. Then, by definition $\langle s, x\rangle \geq 0, \forall s \in S$. Since $C \subseteq S$ we know that $\langle c, x\rangle \geq 0$ for each $c \in C$. Hence, $x \in C^*$. □

Theorem 68. Let $\mathbb{E}$ be an Euclidean space and $K \subseteq \mathbb{E}$ a closed convex cone. Then $K^{**} = K$.

Proof. We first prove that $K \subseteq K^{**}$ using the definition of a dual cone. For every $x \in K$ it holds that $\langle x, y\rangle \geq 0$ for each $y \in K^*$. Thus $K \subseteq K^{**}$. The converse will be shown by contradiction. Let $w \in K^{**}$ and assume that $w \notin K$. By Theorem 25, there exists $c \in \mathbb{E}\setminus\{0\}$ such that

$$\langle w, c\rangle < 0 \text{ and } \langle x, c\rangle \geq 0 \text{ for each } x \in K.$$

This fact implies that $c \in K^*$. Since $\langle w, c\rangle \leq 0$ it follows that $w \notin K^{**}$. □

Analogously to Corollary 31 and Proposition 36, if the convex cone K from the latter is not assumed to be closed, one obtains that $K^{**} = \overline{K}$.

Proposition 69. Let $\mathbb{E}$ be an Euclidean space and let $K \subseteq \mathbb{E}$ be a cone with nonempty interior, then K^* is a pointed cone. Moreover, if K is pointed then K^* has nonempty interior.

Proof. Since K has nonempty interior, choose $x \in K$ and $\delta > 0$ such that $x + \delta\mathbb{B} \subseteq K$. Simultaneously, suppose that there exists $y \in \mathbb{E}$ such that $\mathbb{R}y \subseteq K^*$. Then,

$$\langle \alpha y, x + \delta v\rangle = \alpha(\langle y, x\rangle + \delta\langle y, v\rangle) \text{ for each unitary } v \text{ and for every } \alpha \in \mathbb{R}.$$

This is impossible to happen as there must exist $x_0 \in x + \delta\mathbb{B}$ such that $\alpha\langle y, x_0\rangle < 0$ for sufficiently large α. Thus, we have the result. The final part of the proof follows by symmetry from Theorem 68. □

Corollary 70. If $\mathbb{E}$ is an Euclidean space and $K \subseteq \mathbb{E}$ is a proper cone, then K^* is also a proper cone.

Proof. Immediate from Theorem 68 and Proposition 69. □

Proposition 71. Let $\mathbb{E}$ be an Euclidean space and let $K \subseteq \mathbb{E}$ be a proper cone. Then

(i) $K^* = \operatorname{int}(K)^*$;

(ii) $\operatorname{int}(K^*) = \{x \in \mathbb{E} : \langle x, y\rangle > 0 \text{ for each } y \in K \setminus \{0\}\}$.

Proof.

(i) The inclusion $K^* \subseteq \operatorname{int}(K)^*$ follows from Proposition 67 because $\operatorname{int}(K) \subseteq K$. Conversely, let $x \in \operatorname{int}(K)^*$ and assume that $x \notin K^*$. Thus, there exists $y \in K$ such that $\langle x, y\rangle < 0$. Since the function $\langle x, \cdot\rangle$ is continuous, there exists $\varepsilon \in \mathbb{R}_{++}$ such that $\langle x, z\rangle < 0$ for each $z \in y + \varepsilon\mathbb{B}$. Moreover, for each $\varepsilon \in \mathbb{R}_{++}$ we have that $y + \varepsilon\mathbb{B} \cap \operatorname{int}(K) \neq \varnothing$. So we conclude that there exists $\bar{z} \in \operatorname{int}(K)$ such that $\langle x, \bar{z}\rangle < 0$, which contradicts our assumption that $x \in \operatorname{int}(K)^*$.

(ii) First we note that K^* has nonempty interior by Proposition 69. Let $x \in \operatorname{int}(K^*)$ and assume by contradiction that there exists $y \in K\setminus\{0\}$ such that $\langle x, y\rangle = 0$. Let $\varepsilon \in \mathbb{R}_{++}$, we shall show that there exists $z \in x + \varepsilon\mathbb{B}$ such that $\langle z, y\rangle < 0$. Set $z := x - \frac{\varepsilon}{\|y\|}y$. Clearly, $z \in x + \varepsilon\mathbb{B}$ and also

$$\langle z, y\rangle = \langle x - \frac{\varepsilon}{\|y\|}y, y\rangle = -\frac{\varepsilon}{\|y\|}\langle y, y\rangle = -\varepsilon\|y\| < 0.$$

This contradicts our assumption that $x \in \operatorname{int}(K^*)$. Conversely, let x such that $\langle x, y\rangle > 0$ for each $y \in K \setminus \{0\}$. Note that this property holds if, and only if $\langle x, y\rangle > 0$ for each $y \in K \cap \mathbb{B}_=$. Considering the continuous function $\langle x, \cdot\rangle$, we obtain from Theorem 11 is minimized at some $\bar{y} \in K \cap \mathbb{B}_=$ assuming value $\varepsilon \in \mathbb{R}_{++}$. Let $\delta \in \mathbb{R}_{++}$ such that $\delta \leq \varepsilon$ and $z \in x + \delta\mathbb{B}$ implies that $\langle z, \bar{y}\rangle \in [0, 2\varepsilon]$. We need to show that $x + \delta\mathbb{B} \subseteq K^*$. So, let $z \in x + \delta\mathbb{B}$ and note that

$$\langle z, y\rangle \geq \langle x, y\rangle + \delta \inf_{w\in\mathbb{B}_=}\{\langle w, y\rangle\} \geq \varepsilon - \delta \geq 0 \text{ for each } y \in \mathbb{B}_= \cap K.$$

Thus, we conclude that $z \in K^*$. □

Proposition 72. Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces, let $K \subseteq \mathbb{E}$ be a proper cone and let $A\colon \mathbb{E} \to \mathbb{Y}$ be an invertible linear function. Consider $\mathbb{Y} \supseteq L := A(K)$. Then L is a proper cone.

Proof. We will show that L has each of the properties of a proper cone:

(i) Fix $y \in L$ and $\alpha \in \mathbb{R}_{++}$. By hypothesis, we have that there exists $x \in K$ such that $A(x) = y$. To show that $\alpha y \in L$, we note that $\alpha x \in K$. Thus, $A(\alpha x) = \alpha A(x) = \alpha y \in L$. Therefore L is a cone.

(ii) Let $y_1, y_2 \in L$. By hypothesis, there exist $x_1, x_2 \in K$ such that $A(x_1) = y_1$ and $A(x_2) = y_2$. Since K is convex, it follows from Proposition 57 that $x_1 + x_2 \in K$. Thus, $A(x_1 + x_2) = A(x_1) + A(x_2) = y_1 + y_2 \in L$. Therefore L is convex.

(iii) Since A is linear, we know that A^{-1} is linear as well by Proposition 8. In particular, we have that A^{-1} is continuous. Then, note that $L = A^{-1}(K)$. Hence, L is closed by Proposition 7.

(iv) Since K has nonempty interior, let $x \in \text{int}(K)$ and $\varepsilon > 0$ such that $x + \varepsilon\mathbb{B} \subseteq K$. Since A^{-1} is continuous, we have that there exists $\delta > 0$ such that $A^{-1}(A(x) + \delta\mathbb{B}) \subseteq x + \varepsilon\mathbb{B} \subseteq K$. Because $A^{-1}(L) = K$, it follows that $A(x)+\delta\mathbb{B} \subseteq L$. Thus $A(x) \in \text{int}(L)$, which implies that $\text{int}(L) \neq \varnothing$.

(v) First, observe that $0 \in L$ since $0 \in K$ and A is linear. Then, because K is pointed we have that $x \in K \setminus \{0\}$ implies $-x \notin K$. Hence, since A is invertible and $L = f(K)$ it follows that $y \in L \setminus \{0\}$ implies $-y \notin L$. Thus, $-L = \{0\}$ and then $L \cap -L = \{0\}$. Therefore L is pointed. □

Proposition 73. Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces. Let $K \subseteq \mathbb{E}$ and $L \subseteq \mathbb{Y}$ be proper cones. Then $K \oplus L \subseteq \mathbb{E} \oplus \mathbb{Y}$ is a proper cone.

Proof. We will show that $K \oplus L$ satisfies each of the properties of a proper cone:

(i) Let $(z \oplus t) \in (K \oplus L)$ and fix $\alpha \in \mathbb{R}_{++}$. Then $\alpha(z \oplus t) = (\alpha z \oplus \alpha t) \in (K \oplus L)$ since $\alpha z \in K$ and $\alpha t \in L$. Thus, $K \oplus L$ is a cone.

(ii) The convexity of $K \oplus L$ follows from proposition 13.

(iii) Since K and L are closed, the sets $\mathbb{E}\setminus K$ and $\mathbb{Y}\setminus L$ are open. Thus, the set $(\mathbb{E}\setminus K)\oplus(\mathbb{Y}\setminus L)$ is also open. Then, we note that:

$$K \oplus L = (\mathbb{E} \setminus (\mathbb{E} \setminus K)) \oplus (\mathbb{Y} \setminus (\mathbb{Y} \setminus L)) = (\mathbb{E} \times \mathbb{Y}) \setminus (((\mathbb{E} \setminus K) \oplus \mathbb{Y}) \cap (\mathbb{E} \oplus (\mathbb{Y} \setminus L))).$$

Since $(\mathbb{E} \oplus (\mathbb{Y} \setminus L))$ and $((\mathbb{E} \setminus K) \oplus \mathbb{Y})$ are open, their intersection is also open by Proposition 10 and thus $K \oplus L$ is closed.

(iv) To prove that $K \oplus L$ is pointed, note that $K \cap -K = \{0\}$ and $L \cap -L = \{0\}$ imply that $(K \oplus L) \cap -(K \oplus L) = \{0 \oplus 0\}$.

(v) The facts that $\text{int}(K) \neq \varnothing$ and $\text{int}(L) \neq \varnothing$ give the existence of $(x_1 \oplus x_2) \in (K \oplus L)$ such that $x_1 \in \text{int}(K)$ and $x_2 \in \text{int}(L)$. Then, there exist ε_1 and ε_2 such that $x_1 + \varepsilon_1\mathbb{B} \subseteq K$ and $x_2 + \varepsilon_2\mathbb{B} \subseteq L$. Let $\bar{\varepsilon} := \min\{\varepsilon_1, \varepsilon_2\}$. We will show that $(x_1 \oplus x_2) + \bar{\varepsilon}\mathbb{B} \subseteq (K \oplus L)$. Let $(z \oplus t) \in (x_1 \oplus x_2) + \bar{\varepsilon}\mathbb{B}$. Then:

$$\|(z \oplus t) - (x_1 \oplus x_2)\|^2 = \|z - x_1\|^2 + \|t - x_2\| \leq \bar{\varepsilon}^2.$$

Thus, $\|z-x_1\| \leq \varepsilon_1$ and $\|t-x_2\| \leq \varepsilon_2$. Hence, $z \in (x_1+\varepsilon_1\mathbb{B}) \subseteq K$ and $t \in (x_2 + \varepsilon_2\mathbb{B}) \subseteq L$. Therefore $z \oplus t \in K \oplus L$ and $\text{int}(K \oplus L) \neq \varnothing$. □

The next results study how this operations that preserve the property of being a proper cone relate with duality.

Proposition 74. Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces. Let $K \subseteq \mathbb{E}$ and $L \subseteq \mathbb{Y}$ both be convex sets. Then $(K \oplus L)^* = K^* \oplus L^*$.

Proof. The inclusion $(K \oplus L)^* \supseteq K^* \oplus L^*$ is trivial. The converse will be shown by contradiction. Let $x \oplus y \in (K \oplus L)^*$ and assume that $x \oplus y \notin K^* \oplus L^*$. Then, $x \notin K^*$ or $y \notin L^*$. With no loss of generality, assume that $x \notin K^*$ and $y = 0$, belonging to L^*. Since $x \notin K^*$ there is $k \in K$ such that $\langle x, k\rangle < 0$. Then, clearly $\langle x \oplus y, k \oplus l\rangle < 0$ for any $l \in L$. Therefore, $x \oplus y \notin (K \oplus L)^*$. □

Proposition 75. Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces, let $K \subseteq \mathbb{E}$ and $L \subseteq \mathbb{Y}$ both be convex sets, and let $A \colon \mathbb{E} \to \mathbb{Y}$ be an invertible linear transformation. Then:

(i) $(A^*(L))^* = A^{-1}(L^*)$;

(ii) $A(K)^* = (A^*)^{-1}(K^*)$.

Proof.

(i) By definition:

$$\begin{aligned} x \in (A^*(L))^* &\iff \langle A^*(x), y\rangle \geq 0 \text{ for each } y \in L \\ &\iff \langle x, A(y)\rangle \geq 0 \text{ for each } y \in L \\ &\iff A(x) \in L^* \\ &\iff x \in A^{-1}(L^*). \end{aligned}$$

(ii) By definition:

$$\begin{aligned} x \in A(K)^* &\iff \langle x, A(y)\rangle \geq 0 \text{ for each } y \in K \\ &\iff \langle A^*(x), y\rangle \geq 0 \text{ for each } y \in K \\ &\iff A(x) \in K^* \\ &\iff x \in (A^*)^{-1}(K^*). \end{aligned}$$

□

Proposition 76. Let $\mathbb{E}$ be an Euclidean space, let $S \subseteq \mathbb{E}$ be a subspace of $\mathbb{E}$, and let $K \subseteq \mathbb{E}$ be a convex cone. Then exactly one of the following two statements is true:

(i) There is no hyperplane separating S and K properly;

(ii) There exists x such that $x \in S^{\perp}$, $x \in -K^*$, and $x \notin K^*$.

Proof. By Proposition 27, there exists a hyperplane separating S and K properly if, and only if there exists $x \in \mathbb{E} \setminus \{0\}$ such that

$$\inf_{s \in S}\{\langle x, s\rangle\} \geq \sup_{k \in K}\{\langle x, k\rangle\}$$

and

$$\sup_{s \in S}\{\langle x, s\rangle\} > \inf_{k \in K}\{\langle x, k\rangle\}.$$

This inequalities are equivalent to

$$-\,\delta(-x \,|\, S^{\perp}) \geq \delta(x \,|\, -K^*) \tag{2.1}$$

and

$$\delta(x \,|\, S^{\perp}) > -\delta(-x \,|\, -K^*) \tag{2.2}$$

Analyzing each possible case, we easily conclude that (2.1) and (2.2) hold if, and only if $x \in S^{\perp}$, $x \in -K^*$, and $x \notin K^*$. □

Proposition 76 has many generalizations and corollaries. We now present few of them:

Corollary 77. Let $\mathbb{E}$ be an Euclidean space and let $\{K_i\}_{i \in I} \subseteq \mathbb{E}$ be a finite family of convex cones. Then exactly one of the following two statements is true:

(i) There exists $y \in \bigcap_{i\in I} \operatorname{ri}(K_i)$;

(ii) There exists a family $\{x_i\}_{i\in I}$ such that $\sum_{i\in I} x_i = 0$, $x_i \in -K^*$ for each $i \in I$, and $x_i \notin K_i^*$ for some $i \in I$.

Proof. Consider the Euclidean space $\mathbb{E}^I$ and the cone $\bigoplus_{i\in I} K_i \subseteq \mathbb{E}^I$. By Corollary 20, we have that $\operatorname{ri}(K) = \bigoplus_{i\in I} \operatorname{ri}(K_i)$. Set $S := \{x \in \mathbb{E}^I : x_i = x_j \text{ for each } i,j \in I\}$. Then , for each $y \in \mathbb{E}^I$:

$$\langle x, y\rangle = \sum_{i\in I} \langle x_i, y_i\rangle = |I|\langle x_i, \sum_{i\in I} y_i\rangle.$$

The latter expression is zero for each $x \in S$ if and only if $\sum_{i\in I} y_i = 0$. Thus, $S^\perp = \{y \in \mathbb{E}^I : \sum_{i\in I} y_i = 0\}$ Applying Proposition 76 for L and K, we conclude that exactly one of the following statements is true

(i) There exists $y \in S \cap \operatorname{ri}(K)$;

(ii) There exists x such that $x \in S^\perp$, $x \in -K^*$, and $x \notin K^*$.

Note that $K^* = \bigoplus_{i\in I} K_i^*$ by Proposition 74. Also observe that $S \cap \operatorname{ri}(K) \neq \varnothing$ if, and only if $\bigcap_{i\in I} \operatorname{ri}(K_i) \neq \varnothing$. Thus, the former alternatives are equivalent to

(i) There exists $y \in \bigcap_{i\in I} \operatorname{ri}(K_i)$;

(ii) There exists a family $\{x_i\}_{i\in I}$ such that $\sum_{i\in I} x_i = 0$, $x_i \in -K^*$ for each $i \in I$, and $x_i \notin K_i^*$ for some $i \in I$. □

Corollary 78. Let $\mathbb{E}$ be an Euclidean space and let $\{K_i\}_{i\in I} \subseteq \mathbb{E}$ be a finite family of convex cones. If $\bigcap_{i\in I} \operatorname{ri}(K_i) \neq \varnothing$ then $\sum_{i\in I} K_i^*$ is closed.

Proof. Since $\bigcap_{i\in I} \operatorname{ri}(K_i) \neq \varnothing$, we know that item (ii) of Corollary 77 is false. Thus, we can apply Corollary 66 to the family $\{K_i^*\}_{i\in I}$, obtaining the desired result. □

Corollaries 77 and 78 can obviously be proved considering a finite family of polyhedral cones. The following proposition is a refinement of these conditions and will be essential to our proof of strong duality in the end of this chapter.

Proposition 79. Let $\mathbb{E}$ be an Euclidean space, Let $\{K_i\}_{i\in I} \subseteq \mathbb{E}$ be a finite family of convex cones. Assume that there exists $I_0 \subseteq I$ such that K_i is polyhedral for each $i \in I_0$. If $\bigcap_{i\in I_0} K_i \cap \bigcap_{i\in I\setminus I_0} \operatorname{ri}(K_i) \neq \varnothing$, then $\sum_{i\in I} K_i^*$ is closed.

Proof. We already know that the result is valid when $I_0 = I$ and $I_0 = \varnothing$. Then, we conclude that the result is true for the families $\{K_i\}_{i\in I_0}$ and $\{K_i\}_{i\in I\setminus I_0}$. Hence, it suffices to show the result for cones $K, K_p \subseteq \mathbb{E}$, where K_p is polyhedral and $\operatorname{ri}(K) \cap K_p \neq \varnothing$.

In this context, let $S = \{x \in E^2 : x_1 = x_2\}$. We know that there exists a hyperplane properly separating S and $K \oplus K_p$ if, and only if there exists a hyperplane properly separating K and K_p. By Proposition 30, this happens if, and only if $\operatorname{ri}(K) \cap K_p = \varnothing$. Since $\operatorname{ri}(K) \cap K_p \neq \varnothing$ by hypothesis, we conclude that there is no hyperplane separating S and $K \oplus K_p$ properly. Applying Proposition 76 for these sets, we conclude that item (ii) is false.

Just as in Corollaries 77 and 78, we obtain that the statement "There exists $x \in K$ and $p \in K_p$ such that $x + p = 0$, $x \in -K^*, p \in -K_p^*$, and $x \notin K_i^*$ or $p \notin K_p^*$" is false. Then, Applying Corollary 66 to K and K_p yields the desired result. □

2.2 Cone Partial Order

We will next present the last concept needed to introduce conic programming. Defining a partial order on a proper cone brings the solid advantage that its properties allow one to generalize the concept of inequalities to Euclidean spaces. This will be our main tool to describe the feasible region of conic programs.

Definition. Let $\mathbb{E}$ be an Euclidean space, $K \subseteq \mathbb{E}$ be a closed, convex, and pointed cone and $x, y \in \mathbb{E}$. Then, the cone K induces an order in $\mathbb{E}$ as follows:

$$x \succeq_K y \iff x - y \in K.$$

Moreover,

$$x \succ_K y \iff x - y \in \operatorname{int}(K).$$

The expression $x \succeq_K y$ may be read as x is greater or equal to y in K.

The reader should notice that $x \in K$ if, and only if $x \succeq_K 0$. Thus, $K = \{x \in \mathbb{E} \colon x \succeq_K 0\}$. Also note that the order does not depend on $\langle \cdot, \cdot \rangle$. Therefore, it could be said that the cone K induces an order in the vector space V. Here, this will not be done for simplicity. Moreover, the order considered at the very beginning of this text for $\mathbb{R}^n$ can be seen as the special case of this definition where $\mathbb{E} = \mathbb{R}^n$ equipped with any inner product and $K = \mathbb{R}^n_+$.

Proposition 80. Let $\mathbb{E}$ be an euclidean space and let $K \subseteq \mathbb{E}$ be a proper cone. Then, $\succeq_K$ is a partial order on $\mathbb{E}$.

Proof. For reflexivity, let $x \in \mathbb{E}$. Since K is closed, we have by Proposition 56 that $x - x = 0 \in K$. Thus, $x \succeq_K x$. Antisymmetry follows from the fact that, whenever $x, y \in \mathbb{E}$, $x \succeq_K y$ and $y \succeq_K x$, we have $x - y \in K$, and $y - x = -(x - y) \in K$. Since K is pointed, this implies that $x - y = 0$. For transitivity, let $x, y, z \in K$ such that $x \succeq_K y$ and $y \succeq_K z$. We have :

$$x - y \in K \text{ and } y - z \in K$$

Since K is convex $(x - y) + (y - z) = x - z \in K$. Hence, $x \succeq_K z$. □

Hopefully, Proposition 80 makes it clear why a proper cone K being required to be convex, pointed and closed contributes for the definition of this partial order. Requiring that $\operatorname{int}(K) \neq \varnothing$ allows us to consider strict inequalities. Furthermore, we can similarly consider $x \preceq_K y$ if, and only if $-x \succeq_K -y$. Also, we remark that Corollary 70 allows us to define a partial order in K^* as well. Next, we present an example illustrating why '$\succeq_K$' is not a total order on $\mathbb{E}$:

Example. Let $\mathbb{E}$ be $\mathbb{R}^m$ and K be $\mathbb{R}^m_+$. Then, take $x = e_1$ and $y = e_2$ and see that $x - y \notin K$ and $y - x \notin K$ as both have a negative coordinate.

2.3 Conic Optimization Problems and Their Duals

In the present section, the mathematical object named conic optimization problem is firstly introduced. Then, we define the so called dual problem in a purely syntactic way.

There are more intuitive approaches for the construction of this theory. Refer to [16] for an algebraic approach, to [33] for the Lagrangian approach, and [30] for a more general one. Also, [7] provides several interpretations for duality theory in the convex optimization context. Anyway, we still adopt this construction once it makes it clearer why many symmetric properties of the primal-dual pair of conic problems are true.

Definition. Let $\mathbb{E}_1, \mathbb{E}_2, \mathbb{E}_3$ and $\mathbb{Y}_1, \mathbb{Y}_2, \mathbb{Y}_3$ be Euclidean spaces. Let $K \subseteq \mathbb{E}_1$ and $L \subseteq \mathbb{Y}_1$ both be proper cones. Let $K_p \subseteq \mathbb{E}_2$ and $L_p \subseteq \mathbb{Y}_2$ both be polyhedral cones. Consider $c_1 \oplus c_2 \oplus c_3 \in \mathbb{E}_1 \oplus \mathbb{E}_2 \oplus \mathbb{E}_3$ and $b_1 \oplus b_2 \oplus b_3 \in \mathbb{Y}_1 \oplus \mathbb{Y}_2 \oplus \mathbb{Y}_3$. Let $A\colon \mathbb{E}_1 \oplus \mathbb{E}_2 \oplus \mathbb{E}_3 \to \mathbb{Y}_1 \oplus \mathbb{Y}_2 \oplus \mathbb{Y}_3$ be a linear function.

A *conic optimization problem* is an optimization problem of the form:

$$\begin{aligned} \text{minimize} \quad & \langle x_1 \oplus x_2 \oplus x_3, c_1 \oplus c_2 \oplus c_3 \rangle \\ \text{subject to} \quad & A(x_1 \oplus x_2 \oplus x_3) \succeq_{L^* \oplus L_p^* \oplus \{0\}} b_1 \oplus b_2 \oplus b_3, \\ & x_1 \oplus x_2 \oplus x_3 \in K \oplus K_p \oplus \mathbb{E}_3. \end{aligned} \tag{2.3}$$

The set $G \coloneqq \{x_1 \oplus x_2 \oplus x_3 \in K \oplus K_p \oplus \mathbb{E}_3 \colon A(x_1 \oplus x_2 \oplus x_3) \succeq_{L^* \oplus L_p^* \oplus \{0\}} b_1 \oplus b_2 \oplus b_3\}$ is the feasible set of (2.3). According to the notation presented in the Preliminaries, a conic optimization problem can be represented simply as $(G, \langle \cdot, c_1 \oplus c_2 \oplus c_3 \rangle)$.

We now define the dual problem of (2.3).

Definition. Consider the conic optimization problem (2.3). The *dual problem* of (2.3) is the conic optimization problem

$$\begin{aligned} \text{maximize} \quad & \langle b_1 \oplus b_2 \oplus b_3, y_1 \oplus y_2 \oplus y_3 \rangle \\ \text{subject to} \quad & A^*(y_1 \oplus y_2 \oplus y_3) \preceq_{K^* \oplus K_p^* \oplus \{0\}} c_1 \oplus c_2 \oplus c_3, \\ & y_1 \oplus y_2 \oplus y_3 \in L \oplus L_p \oplus \mathbb{Y}_3. \end{aligned} \tag{2.4}$$

When these problems are presented in pairs, (2.3) can be referred as the *primal* problem. To avoid excessive repetition, we also refer to a conic optimization problem as conic program or simply conic problem.

Here, the advantages from our definitions arise as Theorems 68 and 6 clearly give us that the dual of (2.4) is (2.3) and then these problems' labels can be exchanged with no loss of generality. We now focus on studying some relations between the so-called *primal-dual* pair of conic problems. To simplify the notation on the following propositions, we will denote $x \coloneqq x_1 \oplus x_2 \oplus x_3$, $y \coloneqq y_1 \oplus y_2 \oplus y_3$, $c \coloneqq c_1 \oplus c_2 \oplus c_3$, and $b \coloneqq b_1 \oplus b_2 \oplus b_3$ on (2.3) and (2.4) until the end of this section.

Theorem 81 (Weak duality)**.** Let α be the optimal value of (2.3) and let β be the optimal value of (2.4). If x is feasible in (2.3) and y is feasible in (2.4), then $\langle b, y \rangle \leq \langle c, x \rangle$. In particular, $\alpha \geq \beta$. Moreover, if $\langle x, c \rangle = \langle b, y \rangle$ then x and y are optimal solutions for their respective problems and $\alpha = \beta$.

Proof. Since y is a feasible point in (2.4), we have that $A^*(y) \succeq_{K^* \oplus K_p^* \oplus \{0\}} c$. By definition, this is equivalent to $c - A^*(y) \in K^* \oplus K_p^* \oplus \{0\}$. That is,

$$\langle c - A^*(y), x \rangle = \langle c, x \rangle - \langle A^*(y), x \rangle \geq 0.$$

Similarly, since x is feasible in (2.3), we have that $A(x) - b \in L^* \oplus L_p^* \oplus \{0\}$, which is equivalent to

$$\langle A(x) - b, y \rangle = \langle A(x), y \rangle - \langle b, y \rangle \geq 0.$$

Thus, we conclude that $\langle b, y \rangle \leq \langle c, x \rangle$ by the definition of an adjoint operator. Obviously, this implies that $\alpha \geq \beta$. Finally, assume that $\langle x, c \rangle = \langle y, b \rangle$ and note that our latter calculations imply that $\langle x, c \rangle \leq \langle \bar{x}, c \rangle$ for each $\bar{x}$ feasible in (2.3). That is, x is optimal. Symmetrically, our calculations imply that $\langle y, b\rangle \geq \langle \bar{y}, b\rangle$ for each $\bar{y}$ feasible in (2.4). Hence, y is optimal as well. Clearly, this implies that $\alpha = \beta$. □

Corollary 82. Let α be the optimal value of (2.3) and let β be the optimal value of (2.4). Let x be a feasible solution in (2.3) and y be a feasible solution in (2.4). Then x and y are optimal in their respective problems and $\alpha = \beta$ if, and only if,

$$\langle x, c - A^*(y)\rangle = \langle A(x) - b, y\rangle = 0.$$

Proof. From Theorem 81, we obtain that, whenever x is feasible in (2.3) and y is feasible in (2.4):

$$\langle x, c\rangle \geq \langle A^*(y), x\rangle = \langle y, A(x)\rangle \geq \langle b, y\rangle.$$

In particular, if x and y are optimal and $\alpha = \beta$, then

$$\langle c, x\rangle = \langle b, y\rangle.$$

This forces $\langle x, c\rangle = \langle x, A^*(y)\rangle$ and thus $\langle x, c - A^*(y)\rangle = 0$.

Symmetrically, $\langle A(x), y\rangle = \langle b, y\rangle$ implies that $\langle A(x) - b, y\rangle = 0$. Conversely, assume that

$$\langle x, c - A^*(y)\rangle = 0 \text{ and } \langle A(x) - b, y\rangle = 0.$$

Then, $\langle x, c\rangle = \langle x, A^*(y)\rangle$ and $\langle A(x), y\rangle = \langle b, y\rangle$. Applying the definition of an adjoint operator and Theorem 81 produces the desired result. □

Corollary 83. If the optimal value of the problem (2.3) is $-\infty$, then (2.4) is infeasible. Similarly, if the optimal value of (2.4) is $+\infty$, then (2.3) is infeasible.

Proof. Trivial from Theorem 81. □

2.4 Conditions for Strong Duality

Strong duality is the most important result in conic optimization and it states that the optimal values of the primal and dual problems coincide given that the primal problem is feasible and has finite optimal value. Unfortunately, this result, which is also know as zero duality-gap, does not hold in general. However, there are several conditions that, being satisfied by the primal problem, guarantee strong duality to hold. These conditions are called *constraint qualifications* and the most common of them is *Slater's condition.*

As noted in [14], when K is polyhedral, strong duality still hold regardless of the slater condition being satisfied. This motivated us to show this result in a scenario which includes both polyhedral and non-polyhedral constraints. Then, we will obtain that strong duality holds with an intermediate assumption.

We say that (2.3) satisfies the *restricted*, or *weak* Slater's condition if there exists

$$x_1 \oplus x_2 \oplus x_3 \in \operatorname{int}(K) \oplus K_p \oplus \mathbb{E}_3 \text{ such that } A(x_1 \oplus x_2 \oplus x_3) - b_1 \oplus b_2 \oplus b_3 \in \operatorname{int}(L) \oplus L_p \oplus \{0\}.$$

In this case, x is a *restricted Slater point.*

The following Proposition is the key to simplify our proof of strong duality.

Proposition 84. Let $\mathbb{E}$ be an Euclidean space, let $K \subseteq \mathbb{E}$ be a proper cone, let $K_p \subseteq \mathbb{E}$ be a polyhedral cone and $S \subseteq \mathbb{E}$ a linear subspace. Assume that $\operatorname{int}(K) \cap K_p \cap S \neq \varnothing$. Then $(K \cap K_p \cap S)^* = (K^* + K_p^* + S^\perp)$.

Proof. The inclusion $(K \cap K_p \cap S)^* \supseteq (K^* + K_p^* + S^\perp)$ is easy to prove. Let $a + b + c \in (K^* + K_p^* + S^\perp)$ and fix $x \in (K \cap K_p \cap S)$. By definition:

$$\langle a + b + c, x \rangle = \langle a, x \rangle + \langle b, x \rangle + \langle c, x \rangle \geq 0.$$

Since $a \in K^*, b \in K_p^*$ and $c \in S^\perp$ the result is shown.

Conversely, we show that $(K \cap K_p \cap S) \supseteq (K^* + K_p^* + S^\perp)^*$ then, the desired result will be given by Propositions 67 and 79 along with Theorem 68. Let $x \in (K^* + K_p^* + S^\perp)^*$. By definition,

$$\langle a + b + c, x \rangle = \langle a, x \rangle + \langle b, x \rangle + \langle c, x \rangle \geq 0 \text{ for each } a + b + c \in (K^* + K_p^* + S^\perp).$$

We want to conclude $x \in (K \cap K_p \cap S)$. Note that if $b = c = 0$ we have $\langle a, x \rangle \geq 0$. Similarly, $a = b = 0$ and $a = c = 0$ imply, respectively, that $\langle c, x \rangle \geq 0$ and $\langle b, x \rangle \geq 0$. Furthermore, since $S^\perp$ is a linear subspace, it is true that $-c \in S^\perp$. Hence, $\langle x, c \rangle = 0$ for each $c \in S^\perp$ if and only if $x \in (S^\perp)^\perp = S$. Therefore, $x \in (K \cap K_p \cap S)$. ☐

Theorem 85 (Strong duality)**.** Consider the optimization problem (2.3). If (2.3) is bounded below and has a restricted Slater point, then the optimal values of (2.3) and its dual (2.4) are equal and (2.4) has an optimal solution.

Proof. Let $\alpha \in \mathbb{R}$ be the optimal value of (2.3) and consider the following objects:

(i) $\mathbb{K} := K \oplus \mathbb{E}_2 \oplus \mathbb{E}_3 \oplus \mathbb{R}_+$;

(ii) $\mathbb{K}_p := \mathbb{E}_1 \oplus K_p \oplus \mathbb{E}_3 \oplus \mathbb{R}_+$;

(iii) $S := \{x_1 \oplus x_2 \oplus x_3 \oplus t \in \mathbb{E}_1 \oplus \mathbb{E}_2 \oplus \mathbb{E}_3 \oplus \mathbb{R} : A(x_1 \oplus x_2 \oplus x_3) \succeq_{L^* \oplus L_p^* \oplus \{0\}} t(b_1 \oplus b_2 \oplus b_3)\}$;

(iv) $c := c_1 \oplus c_2 \oplus c_3 \oplus -\alpha$.

Note that $(\mathbb{K} \cap \mathbb{K}_p \cap S)$ corresponds to the feasible region of (2.3) and that $c \in (\mathbb{K} \cap \mathbb{K}_p \cap S)^*$ because α is the optimal value of this problem. Moreover, $\mathbb{K}^* = K^* \oplus \{0\} \oplus \{0\} \oplus \mathbb{R}_+$ and $\mathbb{K}_p^* = \{0\} \oplus K_p^* \oplus \{0\} \oplus \mathbb{R}_+$ by Proposition 74 and

$$S^\perp = \{A^*(y_1 \oplus y_2 \oplus y_3) \oplus (-\langle b_1 \oplus b_2 \oplus b_3, y_1 \oplus y_2 \oplus y_3 \rangle) : y_1 \oplus y_2 \oplus y_3 \in L \oplus L_p \oplus \mathbb{Y}_3\}.$$

Since there is a restricted Slater point by hypothesis, we can apply Proposition 84 to conclude that there exists $z \in \mathbb{K}^*$, $w \in \mathbb{K}_p^*$ and $v \in S^\perp$ such that $c = z + v + w$. The last coordinate of this equation gives us:

$$-\alpha = (\beta + \gamma) - \langle b_1 \oplus b_2 \oplus b_3, y_1 \oplus y_2 \oplus y_3 \rangle \text{ for some } \beta, \gamma \in \mathbb{R}_+ \text{ and } y_1 \oplus y_2 \oplus y_3 \in L \oplus L_p \oplus \mathbb{Y}_3$$

This equality implies that $\alpha \leq \langle b_1 \oplus b_2 \oplus b_3, y_1 \oplus y_2 \oplus y_3 \rangle$. Since $\alpha \geq \langle b_1 \oplus b_2 \oplus b_3, y_1 \oplus y_2 \oplus y_3 \rangle$ by Theorem 81, the result follows. Of course, $y_1 \oplus y_2 \oplus y_3$ in an optimal solution for (2.4). ☐

We conclude this Chapter with a brief review of our path to Theorem 85. In Section 1.2, Proposition 20 has shown that closures and linear images of convex sets do not always commute. Without this inconvenient, our use of Proposition 79 should be a triviality, allowing us to remove remove Slater's condition requirement on Proposition 84 and preserving the exact same result. In this imaginary circumstance, strong duality would hold regardless of any constraint qualification being satisfied.

In the real world, we were forced to handle this obstacle. Using recession cones as our main tool, we determined conditions that fill the gap left by Proposition 20. Applying these to the summation linear function produced Corollary 66. Later, Proposition 79 allowed us to connect our previous achievements with duality theory. In fact, this result completely reflects the importance of Slater's condition in our endeavor. The proof of Proposition 84 relies on this construction when we use Theorem 68. With all this machinary set, we were able to turn the Strong Duality Theorem into an impressively simple and elegant result.

Chapter 3

Relative Entropy Optimization

In the first chapters of this text we introduced the basics of conic programming. From now on, our work will be focused on some specific cones. The first section of this chapter covers the definition and basic properties of the well known, e.g. [21], exponential cone. Then, we define the one-dimensional relative entropy cone, prove its basic properties and expose a linear bijection that relates these two mentioned cones and show that the same bijection actually also connects their dual cones. Later, we derive the properties of the n-dimensional relative entropy cone from the results established in Chapter 2 and present the so-called relative entropy programs (REP). Finally, we display a simple application of REP to statistical learning.

3.1 The Exponential Cone

Definition. Let $n \in \mathbb{N}$. The *exponential cone* is defined as:

$$\mathbb{G}_n := \left\{ (x \oplus \theta \oplus \beta) \in \mathbb{R}^n \oplus \mathbb{R}_+ \oplus \mathbb{R}_+ : \theta \sum_{i \in [n]} \exp\left(\frac{-x_i}{\theta}\right) \leq \beta \right\}.$$

Where we consider $0 \exp(\frac{\alpha}{0}) = 0$ for each $\alpha \in \mathbb{R}_+$.

The convention mentioned in the previous definition is designed to preserve the continuity of the function $\theta \exp(\frac{-x}{\theta})$, which is obviously not defined for $\theta = 0$.

Proposition 86. The function $\overline{f} \colon \mathbb{R}^n \oplus \mathbb{R}_+ \to \mathbb{R}_+$ given by:

$$\overline{f}(x \oplus \theta) := \theta \sum_{i \in [n]} \exp\left(\frac{-x_i}{\theta}\right) \text{ for each } (x \oplus \theta) \in \mathbb{R}^n \oplus \mathbb{R}_+$$

where we consider $0 \exp(\frac{-\alpha}{0}) = 0$ for each $\alpha \in \mathbb{R}_+$ is convex.

Proof. Consider the function $g \colon \mathbb{R} \to \mathbb{R}_+$ given by $g(\alpha) := \exp(-\alpha)$ for each $\alpha \in \mathbb{R}$. Calculating the second-order derivative of g we obtain that $g'' = g$ and thus g'' is strictly positive in $\mathbb{R}$. Then, since g is convex by Proposition 33, by Proposition 32:

$$g((1-\lambda)\alpha_1 + \lambda\alpha_2) \leq (1-\lambda)g(\alpha_1) + \lambda g(\alpha_2), \text{ for each } \alpha_1, \alpha_2 \in \mathbb{R} \text{ and } \lambda \in [0,1].$$

Set $\alpha_1 := \frac{x}{\theta}$, $\alpha_2 := \frac{z}{\mu}$ and $\lambda := \frac{\mu}{\theta+\mu}$ for some $x, z \in \mathbb{R}$ and $\theta, \mu \in \mathbb{R}_{++}$. Substituting these values in the previous expression yields:

$$\exp\left(-\left(\frac{\theta}{\theta+\mu}\frac{x}{\theta} + \frac{\mu}{\theta+\mu}\frac{z}{\mu}\right)\right) \leq \frac{\theta}{\theta+\mu}\exp\left(\frac{-x}{\theta}\right) + \frac{\mu}{\theta+\mu}\exp\left(\frac{-z}{\mu}\right).$$

Multiplying both sides by $\theta + \mu$ gives us that:

$$(\theta + \mu)\exp\left(\frac{-z - x}{\theta + \mu}\right) \leq \theta \exp\left(\frac{-x}{\theta}\right) + \mu \exp\left(\frac{-z}{\mu}\right).$$

Thus, the function $h\colon \mathbb{R}\oplus\mathbb{R}_{++} \to \mathbb{R}_+$ given by $h(x\oplus\theta) := \theta\exp(\frac{-x}{\theta})$ is convex by Corollary 58. Then, by Proposition 35 the function $f\colon \mathbb{R}^n\oplus\mathbb{R}_{++} \to \mathbb{R}_+$ defined as $f(y \oplus \theta) := \sum_{i\in[n]} h(y_i \oplus \theta)$ is convex. Finally, applying Proposition 34 we obtain that $\overline{f}$ is exactly the closure of f. This implies that $\overline{f}$ is convex because $\operatorname{epi}(\overline{f}) = \overline{\operatorname{epi}(f)}$ by definition and $\operatorname{epi}(f)$ is convex. □

Proposition 87. The set $\mathbb{G}_n$ is a proper cone.

Proof. First, note that $\mathbb{G}_n = \operatorname{epi}(\overline{f})$ where $\overline{f}$ is the function considered in Proposition 86, which gives us that $\mathbb{G}_n$ is convex. This Proposition also tells us that $\mathbb{G}_n$ is closed because $\operatorname{epi}(\overline{f})$ is the closure of a set. Moreover, noting that $\overline{f}$ is positively homogeneous we obtain that $\mathbb{G}_n$ is a cone by Proposition 55. We also have that $\mathbb{G}_n$ has nonempty interior by Proposition 9 because $\overline{f}$ is continuous. Finally, we note that $-\mathbb{G}_n = \{0\}$ since $\mathbb{G}_n \subseteq \mathbb{R}^n_+ \oplus \mathbb{R}_+$ and $0 \in \mathbb{G}_n$. This implies that $\mathbb{G}_n \cap -\mathbb{G}_n = \{0\}$. That is, $\mathbb{G}_n$ is pointed. □

We note that we can also obtain that $\mathbb{G}_n$ is closed from Proposition 9. This cone is useful, for example, to study geometric programs [8], which will be explored in more detail in Chapter 5. Next, we calculate the dual cone of $\mathbb{G}_n$.

Proposition 88. The dual cone of $\mathbb{G}_n$ is the set

$$\mathbb{G}_n^* = \left\{(y \oplus \lambda \oplus \alpha) \in \mathbb{R}^n \oplus \mathbb{R} \oplus \mathbb{R} : -\sum_{i\in[n]} y_i \log\left(\left|\frac{y_i}{\alpha}\right|\right) + |y_i| \leq \lambda\right\}.$$

Proof. Let $y \oplus \lambda \oplus \alpha \in \mathbb{R}^n \oplus \mathbb{R} \oplus \mathbb{R}$. By definition, $(y \oplus \lambda \oplus \alpha) \in \mathbb{G}_n^*$ if and only if

$$\langle x \oplus \theta \oplus \beta, y \oplus \lambda \oplus \alpha\rangle = \langle x, y\rangle + \theta\lambda + \beta\alpha \geq 0 \text{ for each } (x \oplus \theta \oplus \beta) \in \mathbb{G}_n.$$

Assume that $\theta \in \mathbb{R}_{++}$ and note that it does not change our result by Proposition 71. Let $t \in \mathbb{R}^n$ such that $t_i := \frac{x_i}{\theta}$ for each $i \in [n]$. Then, for each $(x \oplus \theta \oplus \beta) \in \mathbb{G}_n$:

$$\begin{aligned}
\langle x, y\rangle + \theta\lambda + \beta\alpha \geq 0 &\iff \langle x, y\rangle + \theta\lambda + \alpha\left(\theta \sum_{i\in[n]} \exp\left(\frac{-x_i}{\theta}\right)\right) \geq 0 \\
&\iff \lambda \geq \frac{-\langle x, y\rangle}{\theta} - \alpha \sum_{i\in[n]} \exp\left(\frac{-x_i}{\theta}\right) \\
&\iff \lambda \geq -\langle t, y\rangle - \alpha \sum_{i\in[n]} \exp(-t_i).
\end{aligned}$$

Analyzing each term of the sum above as a function of t_i, we obtain that $t_i y_i + \alpha\exp(-t_i)$ is minimized when $t_i = -\log(|\frac{y_i}{\alpha}|)$. Thus:

$$\min_{t_i\in\mathbb{R}}\{t_i y_i + \alpha\exp(-t_i)\} = y_i \log\left(\left|\frac{y_i}{\alpha}\right|\right) + |y_i|.$$

Summing for each $i \in [n]$ yields the desired result. □

The conventions we adopted on the undefined cases will be exposed below.

3.2 The Relative Entropy Cone

The relative entropy cone is defined as the epigraph of the relative entropy function, which is explored in detail in [13]. As in [9], this cone will be defined as a subset of $\mathbb{R}_+ \oplus \mathbb{R}_+ \oplus \mathbb{R}$ and then this definition is extended to the n-dimensional case by taking direct sums of its first version.

Before we start, it is important to explicitly declare that the usual, e.g [13], conventions that $0\log(0) = 0$, $0\log(\frac{0}{0}) = 0$, and $\alpha\log(\frac{\alpha}{0}) = \infty$ for each $\alpha \in \mathbb{R}_{++}$ will be adopted through the remainder of this text. These conventions are meant to preserve the continuity of the function $x\log(\frac{x}{y})$, originally well-defined in $\mathbb{R}_{++}\oplus\mathbb{R}_{++}$. The next proposition aims at showing that the convexity of $x\log(\frac{x}{y})$ is also preserved by our conventions.

Definition. The 1-dimensional *relative entropy cone* is defined as:

$$\mathbb{H}_1 := \left\{ (x \oplus y \oplus \delta) \in \mathbb{R}_+ \oplus \mathbb{R}_+ \oplus \mathbb{R} : x\log\left(\frac{x}{y}\right) \leq \delta \right\}.$$

Where we consider $0\log(0) = 0$, $0\log\left(\frac{0}{0}\right) = 0$, and $\alpha\log\left(\frac{\alpha}{0}\right) = \infty$ for each $\alpha \in \mathbb{R}_{++}$.

Proposition 89. The function $\overline{f}\colon \mathbb{R}_+ \oplus \mathbb{R}_+ \to \mathbb{R}$ given by:

$$\overline{f}(x \oplus y) := x\log\left(\frac{x}{y}\right)$$

where we consider $0\log(0) = 0$, $0\log\left(\frac{0}{0}\right) = 0$, and $\alpha\log\left(\frac{\alpha}{0}\right) = \infty$ for each $\alpha \in \mathbb{R}_{++}$ is convex.

Proof. Consider the function $g\colon \mathbb{R}_{++} \to \mathbb{R}$ given by $g(\alpha) := \alpha\log(\alpha)$ for each $\alpha \in \mathbb{R}_{++}$. Calculating the second-order derivative of g we obtain that $g''(\alpha) = \frac{1}{\alpha}$ for each $\alpha \in \mathbb{R}_{++}$. Then, it follows from Proposition 33 that g is convex. Thus, by Proposition 32:

$$g((1-\lambda)\alpha_1 + \lambda\alpha_2) \leq (1-\lambda)g(\alpha_1) + \lambda g(\alpha_2), \text{ for each } \alpha_1, \alpha_2 \in \mathbb{R}_{++} \text{ and } \lambda \in [0,1].$$

Set $\alpha_1 := \frac{x_1}{y_1}$, $\alpha_2 := \frac{x_2}{y_2}$, and $\lambda := \frac{y_2}{y_1+y_2}$ for some $x_1, x_2, y_1, y_2 \in \mathbb{R}_{++}$. Substituting these values in the previous equations yields:

$$\left(\frac{x_1}{y_1+y_2} + \frac{x_2}{y_1+y_2}\right)\log\left(\frac{x_1}{y_1+y_2} + \frac{x_2}{y_1+y_2}\right) \leq \frac{x_1}{y_1+y_2}\log\left(\frac{x_1}{y_1}\right) + \frac{x_2}{y_1+y_2}\log\left(\frac{x_2}{y_2}\right).$$

Multiplying both sides by $y_1 + y_2$ we get:

$$(x_1+x_2)\log\left(\frac{x_1+x_2}{y_1+y_2}\right) \leq x_1\log\left(\frac{x_1}{y_1}\right) + x_2\log\left(\frac{x_2}{y_2}\right).$$

Thus, by Proposition 58 we have that the function $f\colon \mathbb{R}_{++} \oplus \mathbb{R}_{++} \to \mathbb{R}$ given by $f(x \oplus y) = x\log(\frac{x}{y})$ is convex. Applying Proposition 34 we obtain that $\overline{f}$ is the closure of f, which yields the desired result. □

Proposition 90. The set $\mathbb{H}_1$ is a proper cone.

Proof. First, we note that $\mathbb{H}_1 = \operatorname{epi}(\overline{f})$ where $\overline{f}$ is the function from Proposition 89, which gives us that $\mathbb{H}_1$ is convex. Moreover, this result also gives us that $\mathbb{H}_1$ is closed because it is the closure of a set. Since $\overline{f}$ is positively homogeneous and continuous, we have that $\mathbb{H}_1$ is a cone and has nonempty interior by Propositions 55 and 9, respectively. It remains to show that $\mathbb{H}_1$ is pointed. Note that $0 \in \mathbb{H}_1$ and also that the only element $(x \oplus y \oplus \delta)$ of $\mathbb{H}_1$ with non-positive x and y is 0. This implies that $-\mathbb{H}_1 = \{0\}$. Thus $\mathbb{H}_1 \cap -\mathbb{H}_1 = \{0\}$. That is, $\mathbb{H}_1$ is pointed. □

Alike in Proposition 87, we can obtain that $\mathbb{H}_1$ is closed from Proposition 9. Next, we show a linear correspondence between $\mathbb{G}_1$ and $\mathbb{H}_1$.

Proposition 91. Let $(x \oplus y \oplus \delta) \in \mathbb{R} \oplus \mathbb{R}_+ \oplus \mathbb{R}_+$. Then $(x \oplus y \oplus \delta) \in \mathbb{G}_1$ if and only if $(y \oplus \delta \oplus x) \in \mathbb{H}_1$.

Proof. By definition:

$$\begin{aligned}(x \oplus y \oplus \delta) \in \mathbb{G}_1 &\iff y \exp\left(\frac{-x}{y}\right) \leq \delta \\ &\iff -x \leq y \log\left(\frac{\delta}{y}\right) \\ &\iff y \log\left(\frac{y}{\delta}\right) \leq x \\ &\iff (y \oplus \delta \oplus x) \in \mathbb{H}_1. \end{aligned}$$ □

Corollary 92. Let $(x \oplus y \oplus \delta) \in \mathbb{R}_+ \oplus \mathbb{R}_+ \oplus \mathbb{R}_+$. Then $(x \oplus y \oplus \delta) \in \mathbb{G}_1^*$ if and only if $(y \oplus \delta \oplus x) \in \mathbb{H}_1^*$.

Proof. Let T be the permutation derived on Proposition 91. Noting that $(T^*)^{-1} = T$ and applying Proposition 75 yield the desired result. □

Definition. We define the n-dimensional relative entropy cone as:

$$\mathbb{H}_n := \left\{ (x \oplus y \oplus \delta) \in \mathbb{R}^n_+ \oplus \mathbb{R}^n_+ \oplus \mathbb{R}^n : x_i \log\left(\frac{x_i}{y_i}\right) \leq \delta_i, \text{ for each } i \in [n] \right\}.$$

Corollary 93. The set $\mathbb{H}_n$ is a proper cone.

Proof. Let $H := \bigoplus_{i\in[n]} \mathbb{H}_1^i \subseteq (\mathbb{R}_+ \oplus \mathbb{R}_+ \oplus \mathbb{R})^n$ and note that H is a proper cone by Propositions 73 and 90. Consider the function $A\colon (\mathbb{R}_+ \oplus \mathbb{R}_+ \oplus \mathbb{R})^n \to \mathbb{R}^n_+ \oplus \mathbb{R}^n_+ \oplus \mathbb{R}^n$ such that $A((x \oplus y \oplus \delta)_i) = x_i \oplus y_i \oplus \delta_i$ for each $(x \oplus y \oplus \delta) \in (\mathbb{R}_+ \oplus \mathbb{R}_+ \oplus \mathbb{R})^n$. Note that A is a linear bijection and, as such, is invertible. Moreover $A(H) = \mathbb{H}_n$. Thus, we obtain that $\mathbb{H}_n$ is a proper cone from Proposition 72. □

3.3 Relative Entropy Programming

Now, we use the fact that $\mathbb{H}_n$ is a proper cone in order to define relative entropy programs. The previous Corollary shows that for any $n \in \mathbb{N}$, the n-dimensional relative entropy cone is produced using copies of $\mathbb{H}_1$ as primitive building blocks. This implies that direct sums of multi-dimensional relative entropy cones equivalent to $\mathbb{H}_m$ for some appropriate $m \in \mathbb{N}$. In this scenario, it turns out that it is irrelevant to consider direct sums of relative entropy cones to define (REP).

Definition. Let $A\colon \mathbb{R}^n \oplus \mathbb{R}^n \oplus \mathbb{R}^n \to \mathbb{R}^m \oplus \mathbb{R}^m \oplus \mathbb{R}^m$ be a linear function, let $c \in \mathbb{R}^n \oplus \mathbb{R}^n \oplus \mathbb{R}^n$, and let $b \in \mathbb{R}^m \oplus \mathbb{R}^m \oplus \mathbb{R}^m$. A *relative entropy program* (REP) is defined as:

$$\begin{aligned} \text{minimize} \quad & \langle x, c \rangle \\ \text{subject to} \quad & A(x) \succeq_{\mathbb{H}_m^*} b, \\ & x \in \mathbb{H}_n. \end{aligned}$$

In this case, the feasible set is given by $G := \{ x \in \mathbb{R}^n \oplus \mathbb{R}^n \oplus \mathbb{R}^n : A(x) \preceq_{\mathbb{H}_m^*} b, x \in \mathbb{H}_n \}$. According to the notation introduced in the Preliminaries, a relative entropy problem is simply represented by $(G, \langle c, \cdot \rangle)$.

3.4 Application: The Density Estimation Problem

Until the end of this chapter, we are working on a specific instance of the identification problem presented in the Preliminaries. Let $k \in \mathbb{N}$ and consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to [k]$. As shown by Proposition 48, the random variable X induces a probability measure on $[k]$. This measure is assumed to be unknown. Obviously, this means that $\mathbb{P}$ is unknown as well. Our goal is: given a set $\mathcal{D}$ of probability measures and a finite collection $\{\omega_i\}_{i \in I} \subseteq \Omega$, estimate the probability measure $\mathbb{P}_X$.

First, we will represent each probability measure on $[k]$ as an element β of $[0,1]^k$ such that $\sum_{i \in [k]} \beta_i = 1$. That is, we will consider the decision space $\mathcal{D}_\beta := \{\beta \in [0,1]^k : \langle \mathbb{1}, \beta \rangle = 1\}$. Then, we choose the loss-function $L : \Omega \times \mathcal{D}_\beta \to \mathbb{R}_+$ given by:

$$L(\omega, \beta) := -\log(\beta_{X(\omega)}), \text{ for each } \omega \in \Omega \text{ and } \beta \in \mathcal{D}_\beta.$$

Set $x_i := X(\omega_i)$ for each $i \in I$, $n_j := |\{i \in I : X(\omega_i) = j\}|$ for each $j \in [k]$, and $n \in \mathbb{R}^k$ in the obvious way. Calculating the empirical risk associated with L and the collection $\{x_i\}_{i \in I}$ yields:

$$\rho(\beta) = -\sum_{i \in I} \log(\beta_{x_i}) = -\sum_{j \in [k]} n_j \log(\beta_j).$$

Finally, our optimal solution is obtained by minimizing ρ. Equivalently, one can solve the following REP:

$$\begin{aligned} \text{minimize} \quad & \langle -n, \delta \rangle \\ \text{subject to} \quad & \langle \mathbb{1}, \beta \rangle = 1, \\ & (\mathbb{1} \oplus \frac{1}{\beta} \oplus \delta) \in \mathbb{H}_k. \end{aligned}$$

Chapter 4

Second-Order Cone Programming

We start this chapter by presenting a family of cones, which are constructed as the epigraph of some of the well-known p-norms in $\mathbb{R}^n$. Then, we will show their basic properties, just as done for $\mathbb{G}_n$ and $\mathbb{H}_n$. These features guarantee that the conic programming theory developed in Chapter 2 also hold for this family of cones. Next, we formally define an optimization problem over the second-order cone. Later, it is shown that this class of problems can be represented as a special case of REP. Finally, we introduce ridge regression as an application of SOCP to the statistical learning theory shortly described in the Preliminaries of this text.

4.1 The Second-Order Cone

Definition. Let $p \geq 1$. The p-norm of $x \in \mathbb{R}^n$ is defined as:

$$\|x\|_p := \left(\sum_{i \in [n]} |x_i|^p \right)^{\frac{1}{p}}.$$

Definition. The n-dimensional *p-th order cone* $\mathbb{L}_n^p \subset \mathbb{R}^n \oplus \mathbb{R}_+$ is defined as:

$$\mathbb{L}_n^p := \{x \oplus \lambda \in \mathbb{R}^n \oplus \mathbb{R}_+ : \|x\|_p \leq \lambda\}.$$

Proposition 94. Each of the sets $\mathbb{L}_n^p$ is proper a cone.

Proof. Fix $n \in \mathbb{N} \setminus \{0\}$ and $p \geq 1$. Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ given by:

$$f(x) := \|x\|_p, \text{ for each } x \in \mathbb{R}^n.$$

Note that $\mathbb{L}_n^p = \operatorname{epi}(f)$. Thus, it follows from Proposition 55 that $\mathbb{L}_n^p$ is a cone since f is positively homogeneous. Moreover, from the triangle inequality:

$$f(x+y) = \|x+y\|_p \leq \|x\|_p + \|y\|_p = f(x) + f(y), \text{for each } x, y \in \mathbb{R}^n.$$

Hence, $\mathbb{L}_n^p$ is convex by Corollary 58. Also, since f is a composition of continuous functions we have that f is continuous. Then, by Proposition 9 we know that $\mathbb{L}_n^p$ is closed and has non-empty interior. It remains to show that $\mathbb{L}_n^p$ is pointed. Let $x \oplus \lambda \in \mathbb{L}_n^p$ and note that $-x \oplus -\lambda \notin \mathbb{L}_n^p$ unless $x = 0$ and $\lambda = 0$. Hence, we conclude that $-\mathbb{L}_n^p = \{0\}$. Finally, noting that $0 \in \mathbb{L}_n^p$ as well gives us that $\mathbb{L}_n^p$ is pointed. □

In the proof above, the reader may have noted that the only point where we used the definition of the p-norm in $\mathbb{R}^n$ to obtain that f was continuous, and we did not use at all any specific property of $\mathbb{R}^n$. Also, it is widely known that every norm on any vector space V is continuous. Therefore, the epigraph of every norm in any vector space V is a proper cone.

The norm induced by the standard inner-product in $\mathbb{R}^n$ corresponds to the p-norm where $p = 2$ and, from now on, we will consider the cone produced by this case. For simplicity, we denote $\|\cdot\|_2 =: \|\cdot\|$ and $\mathbb{L}_n := \mathbb{L}_n^2$ for the rest of the chapter. This proper cone is the *second-order* cone and the next proposition makes use of this fact and is the final piece for us to define conic problems over $\mathbb{L}_n$.

Proposition 95. The dual cone of $\mathbb{L}_n$ is $\mathbb{L}_n$.

Proof. Let $y \oplus \alpha \in \mathbb{L}_n^*$. Then $\langle y, x\rangle + \alpha\lambda \geq 0$ for each $x \oplus \lambda \in \mathbb{L}_n$. In particular, if $x = -y$ and $\lambda = \|y\|$ we have that:

$$-\|y\|^2 + \alpha\|y\| \geq 0.$$

Therefore, $\alpha \geq \|y\|$. That is, $y \oplus \alpha \in \mathbb{L}_n$. Conversely, let $y \oplus \alpha \in \mathbb{L}_n$. Then, by Proposition 2:

$$\langle x, y\rangle + \alpha\lambda \geq -\|x\|\|y\| + \alpha\lambda \text{ for each } x \oplus \lambda \in \mathbb{L}_n.$$

From this last inequality, we conclude that $y \oplus \alpha \in \mathbb{L}_n^*$. □

4.2 Second-Order Cone Programming

Now that is known that $\mathbb{L}_n$ is a proper cone, we are finally able to present the so called second order conic programs.

Definition. Let K and L be direct sums of second order cones contained in suitable Euclidean spaces $\mathbb{E}$ and $\mathbb{Y}$, respectively. Also let a linear function $A\colon \mathbb{E} \to \mathbb{Y}$.

A *second order conic problem* (SOCP) is defined as:

$$\begin{aligned} \text{minimize} \quad & \langle x, c\rangle \\ \text{subject to} \quad & A(x) \succeq_L b, \\ & x \in K. \end{aligned}$$

Where $c \in \mathbb{E}$ and $b \in \mathbb{Y}$. Moreover, in this case the feasible set is given by the set

$$\{x \in \mathbb{R}^n \oplus \mathbb{R} : A(x) \succeq_L b, x \in K\}.$$

4.3 SOCP as a Special Case of REP

At this point, we are ready to formulate a SOCP as a REP. The main idea behind the following derivation is to write the second order cone $\mathbb{L}_n$ as a convenient direct sum of copies of $\mathbb{L}_2$ and then transform each term in this sum into a relative entropy cone $\mathbb{H}_1$. After that, direct sums of $\mathbb{H}_1$ are taken in order to "reconstruct" $\mathbb{L}_n$.

First, we show how to write the cone $\mathbb{L}_n$ as a convenient direct sum of copies of $\mathbb{L}_2$. In [3], the following construction was named tower of variables. Before we start, it is important to say that we can suppose without loss of generality that exists $j \in \mathbb{N}$ such that $n = 2^j$. The reader should note that this can only be done because if $x \oplus \lambda \in \mathbb{L}_n$, then $x \oplus 0 \oplus \lambda \in \mathbb{L}_{n+1}$.

By definition, we have that $x \oplus \lambda \in \mathbb{L}_n$ if and only if $\|x\| \leq \lambda$, i.e :

$$\sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} \leq \lambda.$$

The assumption that there exists $j \in \mathbb{N}$ such that $n = 2^j$ allows us to form exactly $\frac{j}{2}$ pairs of coordinates of x. By repeating the same process, at the end of the j-th step there is only one variable left. This idea can be stated as the following result:

Theorem 96. Let $j \in \mathbb{N}$ and let $x \oplus \lambda \in \mathbb{R}^{2^j} \oplus \mathbb{R}_+$. Then $x \oplus \lambda \in \mathbb{L}_{2^j}$ if and only if there exists $y \in \mathbb{R}_+^{2^j-2}$ such that

$$(w_{2k}, w_{2k+1}) \oplus w_k \in \mathbb{L}_2, \text{ for each } k \in [2^j - 1], \tag{4.1}$$

where $w := \lambda \oplus y \oplus x$ and we are identifying $\mathbb{R} \oplus \mathbb{R}^{2^j-2} \oplus \mathbb{R}^{2^j}$ with $\mathbb{R}^{2^{j+1}-1}$ in the obvious way.

Proof. We shall prove this result by induction in j. First, assume that $j = 2$, fix $x \oplus \lambda \in \mathbb{L}_4$ and define $y \in \mathbb{R}_+^2$ such that $y_1^2 = x_1^2 + x_2^2$ and $y_2^2 = x_3^2 + x_4^2$. It is clear that $(x_1, x_2) \oplus y_1 \in \mathbb{L}_2$ and $(x_3, x_4) \oplus y_2 \in \mathbb{L}_2$. Then, it is left to prove that $(y_1, y_2) \oplus \lambda \in \mathbb{L}_2$. By definition:

$$\sqrt{y_1^2 + y_2^2} = \sqrt{x_1^2 + x_2^2 + x_3^2 + x_4^2} \leq \lambda.$$

Thus, $(y_1, y_2) \oplus \lambda \in \mathbb{L}_2$. Conversely, let $x \oplus \lambda \in \mathbb{R}^4 \oplus \mathbb{R}_+$ and assume that there exists $y \in \mathbb{R}_+^2$ satisfying (4.1). That is, $(y_1, y_2) \oplus \lambda$, $(x_1, x_2) \oplus y_1$, and $(x_3, x_4) \oplus y_2$ all belong to $\mathbb{L}_2$. Thus,

$$\sqrt{x_1^2 + x_2^2 + x_3^2 + x_4^2} \leq \sqrt{y_1^2 + y_2^2} \leq \lambda.$$

Therefore, $x \oplus \lambda \in \mathbb{L}_4$.

Finally, assuming that the result is true for $j = n$, we will show that it also holds for $j = n + 1$. Let $x \oplus \lambda \in \mathbb{L}_{2^{n+1}}$ and define $s_1 := (x_1, \ldots, x_{2^n})$ and $s_2 := (x_{2^n+1}, \ldots, x_{2^{n+1}})$. By the induction hypothesis, the fact that $s_i \oplus \|s_i\| \in \mathbb{L}_{2^n}$ implies the existence of y_i such that (4.1) holds for $i \in [2]$. Now set $y := \bigoplus_{l=1}^{k-1} (y_1[2^{l-1}, 2^l - 1], y_2[2^{l-1}, 2^l - 1])$. It is left to show that $(\|s_1\|, \|s_2\|) \oplus \lambda \in \mathbb{L}_2$. It follows:

$$\sqrt{\|s_1\|^2 + \|s_2\|^2} = \sqrt{x_1^2 + \ldots + x_{2^{n+1}}^2} = \|x\| \leq \lambda.$$

And then $(\|s_1\|, \|s_2\|) \oplus \lambda \in \mathbb{L}_2$ and (4.1) holds for y.

Conversely, let $x \oplus \lambda \in \mathbb{R}^{2^{n+1}} \oplus \mathbb{R}_+$ and fix y such that (4.1) holds. In particular, we have that $(w_2, w_3) \oplus w_1 \in \mathbb{L}_2$. Note that $w_1 = \lambda$. Applying the induction hypothesis, it follows

$$\|x\| \leq \sqrt{w_2^2 + w_3^2} \leq \lambda.$$

Therefore, $x \oplus \lambda \in \mathbb{L}_{2^{n+1}}$. □

The idea of the previous prove may be replicated for any of the p-norms in $\mathbb{R}^n$. However, the following results, which show how to transform each copy of $\mathbb{L}_2$ in a copy of $\mathbb{H}_1$, are particular of the case we are focusing on.

Theorem 97. Let $x \oplus \lambda \in \mathbb{R}^2 \oplus \mathbb{R}_+$. Then:

$$x \oplus \lambda \in \mathbb{L}_2 \text{ if and only if } \begin{bmatrix} \lambda - x_1 & x_2 \\ x_2 & \lambda + x_1 \end{bmatrix} \in \mathbb{S}_+^2.$$

That is, $\mathbb{L}_2 = \left\{ x \oplus \lambda \in \mathbb{R}^2 \oplus \mathbb{R} : \begin{bmatrix} \lambda - x_1 & x_2 \\ x_2 & \lambda + x_1 \end{bmatrix} \in \mathbb{S}_+^2 \right\} =: A.$

Proof. Let $x \oplus \lambda \in \mathbb{L}_2$. By definition,

$$\|x\| = \sqrt{x_1^2 + x_2^2} \leq \lambda \text{ implies } (\lambda + x_1)(\lambda - x_1) - x_2^2 = \det \begin{bmatrix} \lambda - x_1 & x_2 \\ x_2 & \lambda + x_1 \end{bmatrix} \geq 0.$$

Since $x \oplus \lambda \in \mathbb{L}_2$, we also have that $\lambda - x_1 \geq 0$ and $\lambda + x_1 \geq 0$. Thus, by Theorem 12 it follows that $\mathbb{L}_2$ is contained in the right-hand side.

Conversely, let $x \oplus \lambda \in A$. By Theorem 12:

$$\begin{bmatrix} \lambda - x_1 & x_2 \\ x_2 & \lambda + x_1 \end{bmatrix} \in \mathbb{S}^2_+ \text{ implies } \det \begin{bmatrix} \lambda - x_1 & x_2 \\ x_2 & \lambda + x_1 \end{bmatrix} = (\lambda - x_1)(\lambda + x_1) - x_2^2 = \lambda^2 - x_1^2 - x_2^2 \geq 0.$$

Therefore, $\|x\| \leq \lambda$ and then $x \oplus \lambda \in \mathbb{L}_2$. □

Theorem 98. Let $\begin{bmatrix} a & c \\ c & b \end{bmatrix} \in \mathbb{S}^2$. Then $\begin{bmatrix} a & c \\ c & b \end{bmatrix} \in \mathbb{S}^2_+$ if and only if there exists $\eta \in \mathbb{R}_+$ such that

$$\begin{cases} \eta \log\left(\frac{\eta}{a}\right) + \eta \log\left(\frac{\eta}{b}\right) - 2\eta \leq 2c \\ \\ \eta \log\left(\frac{\eta}{a}\right) + \eta \log\left(\frac{\eta}{b}\right) - 2\eta \leq -2c \end{cases}.$$

Proof. By definition:

$$\begin{bmatrix} a & c \\ c & b \end{bmatrix} \in \mathbb{S}^2_+ \text{ if and only if } z^\top \begin{bmatrix} a & c \\ c & b \end{bmatrix} z = az_1^2 + bz_2^2 + 2cz_1z_2 \geq 0, \text{ for each } z \in \mathbb{R}^2.$$

Equivalently,

$$az_1^2 + bz_2^2 + 2cz_1z_2 \geq 0 \text{ and } az_1^2 + bz_2^2 - 2cz_1z_2 \geq 0, \text{ for each } z \in \mathbb{R}^2_+.$$

Now, consider the variables $w_i = \log(z_i); i = 1, 2$. Substituting into the preceding equation, it follows:

$$a \exp(2w_1) + b \exp(2w_2) + 2c \exp(w_1 + w_2) \geq 0 \text{ and } a \exp(2w_1) + b \exp(2w_2) - 2c \exp(w_1 + w_2) \geq 0.$$

Dividing both sides by $\exp(w_1 + w_2)$ it is easy to see that

$$a \exp(w_1 - w_2) + b \exp(w_2 - w_1) \geq -2c \text{ and } a \exp(w_1 - w_2) + b \exp(w_2 - w_1) \geq 2c, \forall w \in \mathbb{R}^2.$$

Thus,

$$\begin{bmatrix} a & c \\ c & b \end{bmatrix} \in \mathbb{S}^2_+ \iff \inf_{w \in \mathbb{R}^2} \{a \exp(w_1 - w_2) + b \exp(w_2 - w_1)\} \geq \max\{2c, -2c\}.$$

Using the result noted in the appendix of [9], we have that:

$$\inf_{w \in \mathbb{R}^2} \{a \exp(w_1 - w_2) + b \exp(w_2 - w_1)\} = \sup_{\eta \in \mathbb{R}_+} \left\{\eta \log\left(\frac{\eta}{a}\right) + \eta \log\left(\frac{\eta}{b}\right) + 2\eta\right\}.$$

Which yields the desired result. □

Corollary 99. Let $x \oplus \lambda \in \mathbb{R}^2 \oplus \mathbb{R}_+$. Then, $x \oplus \lambda \in \mathbb{L}_2$ if and only if there exists $\eta \in \mathbb{R}_+$

such that:

$$\begin{cases} \delta_a + \delta_b - 2\eta \leq 2x_2; \\ \delta_a + \delta_b - 2\eta \leq -2x_2; \\ (\eta \oplus x_1 - \lambda \oplus \delta_a) \in \mathbb{H}_1; \\ (\eta \oplus x_1 + \lambda \oplus \delta_b) \in \mathbb{H}_1. \end{cases}$$

Proof. From Theorem 97, we have that:

$$x \oplus \lambda \in \mathbb{L}_2 \text{ if and only if } \begin{bmatrix} \lambda - x_1 & x_2 \\ x_2 & \lambda + x_1 \end{bmatrix} \in \mathbb{S}^2_+.$$

Applying Theorem 98 gives us that the former matrix is positive semidefinite if, and only if there exists $\eta \in \mathbb{R}_+$ such that:

$$\begin{cases} \eta \log\left(\frac{\eta}{x_1 - \lambda}\right) + \eta \log\left(\frac{\eta}{x_1 + \lambda}\right) - 2\eta \leq 2x_2; \\ \eta \log\left(\frac{\eta}{x_1 - \lambda}\right) + \eta \log\left(\frac{\eta}{x_1 + \lambda}\right) - 2\eta \leq -2x_2. \end{cases}$$

Or, equivalently:

$$\begin{cases} \delta_a + \delta_b - 2\eta \leq 2x_2; \\ \delta_a + \delta_b - 2\eta \leq -2x_2; \\ (\eta \oplus x_1 - t \oplus \delta_a) \in \mathbb{H}_1; \\ (\eta \oplus x_1 + t \oplus \delta_b) \in \mathbb{H}_1. \end{cases}$$

□

4.4 Application: Ridge Regression

4.4.1 The Regression Problem

Until the end of this chapter, we are working on a particular instance of the general learning problem introduced in the Preliminaries. Specifically, we will handle the case of the prediction problem where the output provided by the supervisor is quantitative, which is commonly named *regression*.

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable $X\colon \Omega \to \mathbb{R}^k$ and a finite collection $\{\omega_i\}_{i\in I} \subseteq \Omega$. Let $x_i := X(\omega_i)$ for each $i \in I$. Assume that we observed a finite family of pairs $\{(x_i, y_i)\}_{i\in I}$, where x_i is an input provided by the generator and y_i is the corresponding output returned from the target operator $T^*\colon \mathbb{R}^k \to \mathbb{R}$. As mentioned in the previous paragraph, our goal is to choose a function $d\colon \mathbb{R}^k \to \mathbb{R}$ which makes good approximations for future outputs of T^*. To restrict our search a bit, we consider that d is affine, which obliges it to be of the form:

$$d(z') = \beta'_0 + \sum_{j\in[k]} \beta'_j z'_j = \langle \beta'_0 \oplus \beta', 1 \oplus z' \rangle, \text{ for each } z' \in \mathbb{R}^k.$$

By considering $z := 1 \oplus z'$ and $\beta := \beta'_0 \oplus \beta'$ we obtain the following simplification of the formula for d.

$$d(z') = \langle \beta, z \rangle, \text{ for each } z' \in \mathbb{R}^k.$$

At this point, it is important to remark that this last formula for d allows us to acknowledge the decision space as $\mathcal{D}_\beta := \mathbb{R}^{k+1}$, since Theorem 3 implies that each $\beta \in \mathbb{R}^{k+1}$ uniquely represents an affine function $d\colon \mathbb{R}^k \to \mathbb{R}$, where we are identifying $\mathbb{R}^{k+1}$ with $\mathbb{R} \oplus \mathbb{R}^k$ in the obvious way.

Under these hypotheses, the problem of anticipating future outcomes from a generator function is labeled *linear regression* and there is a vast theory around it, specially when the quadratic loss function, given by:

$$L((x,y),d) \coloneqq (y - d(x))^2 = (y - \langle \beta, 1 \oplus x \rangle)^2$$

is adopted.

Applying the expectation operator to the function L, we obtain the result that is known in the literature as the bias-variance decomposition of the squared deviation loss.

Proposition 100. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X\colon \Omega \to \mathbb{R}^k$ be a random variable. Consider the loss function $L\colon (\mathbb{R}^k \times \mathbb{R}) \times \mathcal{D}_\beta \to \mathbb{R}$ given by:

$$L((x,y),d) \coloneqq (y - d(x))^2 = (y - \langle \beta, 1 \oplus x \rangle)^2, \text{ where } x = X(\omega) \text{ for some } \omega \in \Omega.$$

Define $E(d(x)) \coloneqq \mu$. Then, $E(L((x,y),d) = E(\mu - y)^2 + \operatorname{Var}(d(x))$.

Proof.

$$\begin{aligned}
E(L((x,y),d)) &= E((d(x) - y)^2) \\
&= E((d(x) - \mu + \mu - y)^2) \\
&= E(((d(x) - \mu) + (\mu - y))^2) \\
&= E((d(x) - \mu)^2 + 2(d(x) - \mu)(\mu - y) + (\mu - y)^2) \\
&= E((d(x) - \mu)^2) + 2(\mu - y)(E(d(x)) - \mu) + E(\mu - y)^2 \\
&= E((d(x) - \mu)^2) + E(\mu - y)^2 \\
&= \operatorname{Var}(d(x)) + E(\mu - y)^2.
\end{aligned}$$

□

Proposition 100 motivated statisticians to restrict the search to unbiased affine estimators since, in this case, the optimal solution for the regression problem can be obtained analytically. The following result is a derivation of this solution and a very similar proof to the one we present can be found in [24].

Theorem 101. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X\colon \Omega \to \mathbb{R}^k$ be a random variable, and consider a finite collection $\{\omega_i\}_{i \in I} \subseteq \Omega$. Let $x_i \coloneqq X(\omega_i)$ for each $i \in I$. Consider a family $\{(x_i, y_i)\}_{i \in I}$, where $y_i \in \mathbb{R}$ for each $i \in I$. Define the empirical risk function $\rho\colon \mathcal{D}_\beta \to \mathbb{R}$ given by:

$$\rho(\beta) \coloneqq \sum_{i \in I} (y_i - \langle \beta, 1 \oplus x_i \rangle)^2, \text{ for each } \beta \in \mathcal{D}_\beta.$$

Define $\mathcal{U} \coloneqq \{\beta \in \mathcal{D}_\beta : E(\langle \beta, x \rangle - y) = 0\}$. Define $\mathbf{X} \in \mathbb{R}^{I \times k}$ such that $\mathbf{X}(i,j) = X_{j-1}(\omega_i)$ for each $j \in [k]$ with $j \geq 2$ and $\mathbf{X}(1,i) = 1$ for each $i \in I$. Consider $Y \in \mathbb{R}^I$ so that $Y_i = y_i$ for each $i \in I$. Then, the optimal solution for the optimization problem $(\rho, \mathcal{U})$ is:

$$\beta^* = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top Y.$$

Proof. Rewriting the function ρ using matrix notation yields:

$$\rho(\beta) = (Y - \mathbf{X}\beta)^\top (Y - \mathbf{X}\beta).$$

Calculating the first and second-order derivatives of ρ, we obtain the following results:

$$\frac{\partial \rho}{\partial \beta} = 2\mathbf{X}^\top (y - \mathbf{X}\beta)$$

$$\frac{\partial^2 \rho}{\partial \beta \partial \beta^t} = 2\mathbf{X}^\top \mathbf{X}.$$

As the data matrix $\mathbf{X}$ has full rank and it is positive definite [25], we set the first derivative to zero

$$\mathbf{X}^\top (y - \mathbf{X}\beta) = 0$$

and then we have the unique solution

$$\beta^* = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top y. \quad \square$$

4.4.2 The Curse of Dimensionality

At first sight, the result given by the preceding theorem seems to be the endpoint for linear regression theory as it gives an analytical expression for its solution. However, when the dimension of the data set is high, the classical method can lead to estimators with extremely high variance, which is not good to make predictions. One way to get around this problem, known as curse of dimensionality, was proposed by Kennard and Hoerl in [23].

The main idea behind ridge regression is to add penalties to the quadratic risk function proportional to the length of the coefficient vector beta that represents the estimator d. That is, adopt the following risk function:

$$\bar{\rho}(\beta) := \sum_{i \in I} (y_i - \langle \beta, 1 \oplus x_i \rangle)^2 + \lambda \|\beta\|^2, \ \lambda \in \mathbb{R}_+.$$

The risk function $\bar{\rho}$, obviously, encourages the vector β to be 'smaller' with respect to the to the 2-norm in the proportion that λ is 'bigger'. Fortunately, the ridge regression problem

$$\begin{aligned} \text{minimize} \quad & \bar{\rho}(\beta) \\ \text{subject to} \quad & \beta \in \mathbb{R}^{k+1} \end{aligned}$$

has also a analytic solution, given by:

$$\beta^* = (\mathbf{X}^\top \mathbf{X} + \lambda I)^{-1} \mathbf{X}^\top Y.$$

4.4.3 A Conic Formulation For Ridge Regression

As briefly argued at the end of the last subsection, the new risk function $\bar{\rho}$ is minimized by some coefficient vector β^* with smaller length than the ordinary least squares estimator. Instinctively, one may think that in this case, we could restrict our search to vectors that have their length bounded by some constant and still have the same optimal solution as the new problem. This hypothetic intuitive view of the ridge regression problem is confirmed by the following proposition:

Proposition 102. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X \colon \Omega \to \mathbb{R}^k$ be a random variable, and consider a finite collection $\{\omega_i\}_{i \in I} \subseteq \Omega$. Let $x_i := X(\omega_i)$ for each $i \in I$. Consider the family $\{(x_i, y_i)\}_{i \in I}$, where $y_i \in \mathbb{R}$ for each $i \in I$ and the ridge empirical risk function $\bar{\rho} \colon \mathcal{D}_\beta \to \mathbb{R}$ given by:

$$\bar{\rho}(\beta) := \sum_{i \in I} (y_i - \langle \beta, 1 \oplus x_i \rangle)^2 + \lambda \|\beta\|, \ \text{for each } \beta \in \mathcal{D}_\beta.$$

Then, for each $\lambda \geq 0$ there exists $\alpha \geq 0$ such that the optimization problems $(\bar{\rho}, \mathbb{R}^{k+1})$

and $(\rho, \alpha\mathbb{B})$ have the same optimal solution.

Proof. Define $L_\lambda := \inf_{\beta \in \mathbb{R}^{k+1}} \sum_{i \in I} (y_i - \langle \beta, 1 \oplus x_i \rangle)^2 + \lambda \|\beta\|$ for each $\lambda \in \mathbb{R}_+$. We first note that L_λ is the optimal value of the optimization problem $(\mathbb{R}^{k+1}, \bar{\rho})$. Thus, the existence of an analytical solution for the ridge regression problem guarantees that there exists $\beta^* \in \mathbb{R}^{k+1}$ such that $\sum_{i \in I}(y_i - \langle \beta^*, 1 \oplus x_i \rangle)^2 + \lambda \|\beta^*\| = L_\lambda$. Then:

$$L_\lambda = \sum_{i \in I} (y_i - \langle \beta^*, 1 \oplus x_i \rangle)^2 + \lambda \|\beta^*\| \implies L_\lambda \geq \lambda \|\beta^*\| \implies \|\beta^*\| \leq \frac{L_\lambda}{\lambda}.$$

Therefore, for $\alpha \geq \frac{L_\lambda}{\lambda}$ we have that $\beta^* \in \alpha\mathbb{B}$ or, equivalently, $(\beta^*, \alpha) \in \mathbb{L}^2_{k+1}$. □

Based on the previous result, we are now able write the ridge regression problem as

$$\begin{aligned} \text{minimize} \quad & \rho(\beta) \\ \text{subject to} \quad & (\beta, \alpha) \in \mathbb{L}^2_{k+1} \end{aligned}$$

for some suitable $\alpha \in \mathbb{R}_+$ given by Proposition 102. Thus, adopting the definitions of $\mathbf{X}$ and Y from Theorem 101, we may rewrite the function ρ as $\rho(\beta) = \|Y - \beta\mathbf{X}\|_2$. Therefore, the ridge regression problem is equivalent to the following SOCP:

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & (Y - \beta\mathbf{X} \oplus t) \in \mathbb{L}^2_{|I|}, \\ & (\beta \oplus \alpha) \in \mathbb{L}^2_{k+1}. \end{aligned}$$

Chapter 5

Geometric Programming

Unlike the families of optimization problems presented in Chapters 3 and 4, geometric programs (GP) are not conic problems by definition. Thus, we start this chapter presenting monomials and posynomials, which are the building blocks of a GP. Then, we present geometric programs and show that they can also be expressed as REP. Finally, we introduce logistic regression as an application of GP to statistical learning.

5.1 Monomials and Posynomials

The introduction of monomials and posynomials usually employs excessive indexing. As an attempt of clarifying the notation as much as possible, we define the function $\log\colon \mathbb{R}^n_{++} \to \mathbb{R}^n$ such that $(\log(x))_i = \log(x_i)$ for each $x \in \mathbb{R}^n_{++}$.

Similarly, consider the function $\exp\colon \mathbb{R}^n \to \mathbb{R}^n_{++}$ so that $(\exp(x))_i = \exp(x_i)$ for each $x \in \mathbb{R}^n$. Moreover, if $x, a \in \mathbb{R}^n$, we denote $x^a \coloneqq \prod_{i\in[n]} x_i^{a_i}$.

Definition. A *monomial* is a function $f\colon \mathbb{R}^n_{++} \to \mathbb{R}$ of the form

$$f(x) = \gamma x^a \text{ for each } x \in \mathbb{R}^n_{++}.$$

The number $\gamma \in \mathbb{R}_+$ is the *coefficient* of the monomial and $a \in \mathbb{R}^n$ is the vector of *exponents* of f.

Anyone that has already taken an algebra course might have feel strange about the latter definition. By taking a deeper look into it, it should become clear that all we are doing is generalizing the more commonly used definition of a monomial function by allowing real exponents instead of requiring them to be positive integers. In the same sense, the following definition is meant to generalize the well-known definition of a polynomial.

Definition. A *posynomial* is a function $f\colon \mathbb{R}^n_{++} \to \mathbb{R}$ of the form:

$$f(x) = \sum_{a\in F} \gamma(a)x^a \text{ for each } x \in \mathbb{R}^n_{++}$$

for some finite $F \subset \mathbb{R}^n$ and $\gamma\colon F \to \mathbb{R}_+$. That is, a posynomial is the sum of finitely many monomials.

To avoid indexing, it is important to understand the fact that each posynomial is uniquely represented by some finite set F and a function $\gamma\colon F \to \mathbb{R}_+$. Moreover, we remark that when dealing with a collection of posynomials, it can always be assumed without any loss of generality that all of them are composed by the same number of monomials, since one can always add monomials with coefficient 0.

The term posynomial suggests a combination of the words polynomial and positive. We now present some properties of this special type of function that follow directly from its definition. In the next section, these results will enable the standard definition of a geometric optimization problem to be generalized.

Proposition 103. Let $f, g\colon \mathbb{R}^n_{++} \to \mathbb{R}$ be posynomials. Then:

(i) $(f+g)(x)$ is a posynomial;

(ii) $(fg)(x)$ is a posynomial;

(iii) $(\frac{f}{g})(x)$ is a posynomial if g is a monomial.

Proof. (i) By definition,

$$\begin{aligned}(f+g)(x) &= \sum_{a\in F}\gamma(a)x^a + \sum_{a\in G}\delta(a)x^a \\ &= \sum_{a\in F\cup G}(\gamma(a)+\delta(a))x^a.\end{aligned}$$

Since F and G are finite, we have that $F\cup G$ is finite. Thus, $(f+g)(x)$ is a posynomial.

(ii) Similarly to the latter, we write:

$$(fg)(x)\sum_{a\in F}\gamma(a)x^a\sum_{b\in G}\delta(b)x^b = \sum_{a\in F}\sum_{b\in G}\gamma(a)\delta(b)x^{a+b}$$

Note that the function fg is the sum of $|F||G|$ monomials. Hence, fg is a posynomial.

(iii) The function

$$\left(\frac{f}{g}\right)(x) = \frac{\sum_{a\in F}\gamma(a)x^a}{\lambda x^v} = \sum_{a\in F}\frac{\gamma(a)}{\lambda}x^{a-v}$$

is trivially a posynomial. □

It is also trivial to see that cf is a posynomial whenever c is a nonnegative constant. Also, the second property immediately implies that f^k is a posynomial for each nonnegative integer k.

5.2 Geometric Optimization Problems

Definition. Let $f_0\colon \mathbb{R}^n_{++} \to \mathbb{R}$ be a posynomial, let $\mathcal{F}_{\leq}$ be a finite family of posynomials such that $f\colon \mathbb{R}^n_{++} \to \mathbb{R}$ for each $f \in \mathcal{F}_{\leq}$, and let $\mathcal{F}_{=}$ be a finite family monomials such that $g\colon \mathbb{R}^n_{++} \to \mathbb{R}$ for every $g \in \mathcal{F}_{=}$.

A *geometric program* (GP) is an optimization problem of the form

$$\begin{aligned}&\text{minimize} && f_0(x) \\ &\text{subject to} && f(x)\leq 1 \quad \forall f\in\mathcal{F}_{\leq}, \\ &&& g(x) = 1 \quad \forall g\in\mathcal{F}_{=}, \\ &&& x\in\mathbb{R}^n_{++}.\end{aligned} \tag{5.1}$$

Problem (5.1) is also known as a standard form GP and it can be generalized in several different ways. For example, if f is a posynomial and λ is a positive constant, the constraint $f(x) \leq \lambda$ can be expressed as $h(x) \leq 1$ where $h(x) = \frac{f(x)}{\lambda}$. More generally, if we consider a monomial g, the inequality $f(x) \leq g(x)$ can be rewritten as $h(x) \leq 1$ where $h(x) = \frac{f(x)}{g(x)}$

since $h(x)$ is also a posynomial by Proposition 103. Obviously, a similar approach can be easily applied to the equality constraints. Also, a minimization program can be transformed into a maximization one by taking the multiplicative inverse of the objective function, which is also a posynomial.

5.3 GP as a Special Case of REP

The main goal of this chapter is to show that any GP may be cast as a REP. This process will be done in two steps. The first of them is to make a change of variable in the standard form GP, obtaining a convex optimization problem. Then, this new version of (5.1) will be shown to be equivalent to a REP.

Following the steps shortly described in the past paragraph, we start rewriting the definition of a GP so that our change of variable will become evident. The only difference between this formulation and (5.1) is the fact that the definitions of posynomials and monomials is used in order to write the standard form GP in terms of them. It is also important to say that we will represent each monomial $g \in \mathcal{F}_=$ as a exponent $b \in \mathbb{R}^n$ and its image under an appropriate function $\delta\colon \mathbb{R}^n \to \mathbb{R}_+$, which stands for the coefficient. Similarly, each function $f \in F_0 \cup \mathcal{F}_\leq$ will be represented by the finite set F containing each of its exponents and a function $\gamma_F\colon F \to \mathbb{R}_+$, just as described previously. Thus, we begin with:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{a\in F_0} \gamma_{F_0}(a) x^a \\
\text{subject to} \quad & \sum_{a\in F} \gamma_F(a) x^a \leq 1 \qquad \forall F \in \mathcal{F}_\leq, \\
& \delta(b) x^b = 1 \qquad \forall b \in \mathcal{F}_=, \\
& x \in \mathbb{R}^n_{++}.
\end{aligned}
\tag{5.2}
$$

Consider the variable $y \coloneqq \log(x)$. Substituting y into (5.2), we obtain the following equivalent problem:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{a\in F_0} \gamma_{F_0}(a) \exp(\langle y, a\rangle) \\
\text{subject to} \quad & \sum_{a\in F} \gamma_F(a) \exp(\langle y, a\rangle) \leq 1 \quad \forall F \in \mathcal{F}_\leq, \\
& \delta(b) \exp(\langle y, b\rangle) = 1 \quad \forall b \in \mathcal{F}_=, \\
& y \in \mathbb{R}^n.
\end{aligned}
\tag{5.3}
$$

Now, define $w_a \coloneqq \langle y, a\rangle$, for each $a \in F_0 \cup \mathcal{F}_\leq$ and $v_b = \langle y, b\rangle$ for each $b \in \mathcal{F}_=$. This last step gives us the problem:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{a\in F_0} \gamma_{F_0}(a) \exp(w_a) \\
\text{subject to} \quad & \sum_{a\in F} \gamma_F(a) \exp(w_a) \leq 1 \qquad \forall F \in \mathcal{F}_\leq, \\
& \delta(b) \exp(v_b) = 1 \qquad \forall b \in F_=, \\
& w_a = \langle y, a\rangle \ \forall a \in F,\ \forall F \in F_0 \cup \mathcal{F}_\leq, \\
& v_b = \langle y, b\rangle \ \forall b \in F_=, \\
& y \in \mathbb{R}^n.
\end{aligned}
\tag{5.4}
$$

Finally, we note that the exponential function is monotone in each coordinate and also that each of the coefficients γ and δ are nonnegative. Hence, we can introduce variables z_a as an upper-bound to $\exp(w_a)$ for each $a \in F$, for each $F \in \{F_0, \mathcal{F}_{\leq}\}$. Similarly, define u_b bounding $\exp(v_b)$ for each $b \in \mathcal{F}_=$ in order to obtain the equivalent optimization problem:

$$
\begin{aligned}
\text{minimize} \quad & \langle \gamma_0, z_0 \rangle \\
\text{subject to} \quad & \langle \gamma_F, z_F \rangle \leq 1 && \forall F \in \mathcal{F}_{\leq}, \\
& \langle \delta(b), u_b \rangle = 1 && \forall b \in \mathcal{F}_=, \\
& w_a = \langle y, a \rangle && \forall a \in F,\ \forall F \in F_0 \cup \mathcal{F}_{\leq}, \\
& v_b = \langle y, b \rangle && \forall b \in \mathcal{F}_=, \\
& w_a \leq \log(z_a) && \forall a \in F,\ \forall F \in F_0 \cup \mathcal{F}_{\leq}, \\
& v_b \leq \log(u_b) && \forall b \in \mathcal{F}_=, \\
& y \in \mathbb{R}^n.
\end{aligned}
\tag{5.5}
$$

Where

$$\gamma_0 \coloneqq \bigoplus_{a \in F_0} \gamma(a), z_0 \coloneqq \bigoplus_{a \in F_0} z_a$$

and

$$\gamma_F \coloneqq \bigoplus_{a \in F} \gamma(a) z_F \coloneqq \bigoplus_{a \in F} z_a \text{ for each } F \in \mathcal{F}_{\leq}.$$

Therefore, the problem 5.1 can be expressed as:

$$
\begin{aligned}
\text{minimize} \quad & \langle \gamma_0, z_0 \rangle \\
\text{subject to} \quad & \langle \gamma_F, z_F \rangle \leq 1 && \forall F \in \mathcal{F}_{\leq}, \\
& \langle \delta(b), u_b \rangle = 1 && \forall b \in \mathcal{F}_=, \\
& w_a = \langle y, a \rangle && \forall a \in F,\ \forall F \in F_0 \cup \mathcal{F}_{\leq}, \\
& v_b = \langle y, b \rangle && \forall b \in \mathcal{F}_=, \\
& (1, z_a, -w_a) \in \mathbb{H}_1 && \forall a \in F,\ \forall F \in F_0 \cup \mathcal{F}_{\leq}, \\
& (1, u_b, -v_b) \in \mathbb{H}_1 && \forall b \in \mathcal{F}_=, \\
& y \in \mathbb{R}^n.
\end{aligned}
\tag{5.6}
$$

This last problem is an REP. Because we just applyed invertible functions which preserve objective values, all the optimization problems displayed above are equivalent by Corollary 40.

In [21], the author adopts a similar approach to write a GP as a conic program in the exponential cone. Together with Proposition 91, this suggests reinforces the relation between $\mathbb{H}_n$ and $\mathbb{G}_n$.

5.4 Application: Logistic Regression

5.4.1 The Binary Classification Problem

This section is devoted to the case of the prediction problem where the outcome provided by the target operator is qualitative and takes values in the set $\{0, 1\}$. Namely, we will be working on a *binary classification problem*.

Similarly to the approach proposed for the Ridge Regression problem, we start by considering a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable $X\colon \Omega \to \mathbb{R}^k$, and finite family $\{\omega_i\}_{i \in I} \subseteq \Omega$. Assume that we observed a finite family of pairs $\{(x_i, y_i)\}_{i \in I}$, where each of these pairs consists of a input $x_i \coloneqq X(\omega_i)$ provided by the generator and an output y_i, returned from the target operator $T^*\colon \mathbb{R}^k \to \{0,1\}$. Our goal is to construct a function $d \in \mathcal{D} := \{0,1\}^{\mathbb{R}^k}$ which makes good approximations for future outputs of T^*.

Since in this problem the range of T^* is discrete, it would be unhandy to evaluate the error on an eventual prediction by squared deviation. An alternative that is more suitable for this context is to consider the loss function $L\colon (\mathbb{R}^k \times \{0,1\}) \times \mathcal{D} \to \mathbb{R}_+$ given by:

$$L((x,y),d) := \begin{cases} 1, & \text{if } d(x) \neq y; \\ 0, & \text{otherwise.} \end{cases}$$

To put it in words, the loss function we are adopting attributes 1 in case of a wrong estimative and 0 in case of a right one. In this scenario, the following proposition, which is a simple calculation of the expected risk $\rho(d)$ induced by the loss function L, gives some direction to solve the problem:

Proposition 104. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X\colon \Omega \to \mathbb{R}^k$ be a random variable, and consider the loss function $L\colon (\mathbb{R}^k \times \{0,1\}) \times \mathcal{D} \to \mathbb{R}_+$ given by:

$$L((x,y),d) := \begin{cases} 1, & \text{if } d(x) \neq y; \\ 0, & \text{otherwise.} \end{cases}$$

Then, the expected risk $\rho(d)$ equals $\mathbb{P}(d(x) \neq y)$.

Proof. We compute the expected value of L:

$$\begin{aligned} E(L((x,y),d)) &= 0 \cdot \mathbb{P}(d(x) = y) + 1 \cdot \mathbb{P}(d(x) \neq y) \\ &= \mathbb{P}(d(x) \neq y). \end{aligned}$$

□

Note that $\mathbb{P}(d(x) \neq y) = \mathbb{P}(d(x) = 1 \cap y = 0) + \mathbb{P}(d(x) = 0 \cap y = 1)$. Using this fact to rewrite the latter result, we obtain:

$$\mathbb{P}(d(x) \neq y) = \begin{cases} \mathbb{P}(y = 0 \,|\, X = x), & \text{if } d(x) = 1; \\ \mathbb{P}(y = 1 \,|\, X = x), & \text{if } d(x) = 0. \end{cases}$$

5.4.2 Logistic Regression as a Plug-in Classifier

Proposition 104 guides us in the direction of a simple method to estimate the outcome of a qualitative variable y. Nonetheless, applying this method requires an estimative for $\mathbb{P}(Y = 1 \,|\, X = x) =: p(x)$ and then we decide based on whether $p(x) \geq \frac{1}{2}$.

An advantageous feature that ridge regression presents and would be convenient to preserve is the fact that the estimator is chosen to be affine. As mentioned in the occasion, this assumption enables us to conveniently apply Theorem 3 to consider the decision space to be $\mathbb{R}^{k+1}$, where we are identifying $\mathbb{R}^{k+1}$ with $\mathbb{R} \oplus \mathbb{R}^k$ in the obvious way.

The simplest procedure to replicate this property would be to simply assume that $p(x)$ is an affine function of x. In this case, we would be able to assume that $p(x) = \langle \beta, 1 \oplus x \rangle$ for some $\beta := \beta_0 \oplus \bar{\beta} \in \mathbb{R} \oplus \mathbb{R}^k$. However, this would imply $p(x) \in (-\infty, \infty)$, which is undesirable since $p(x)$ is defined as a probability. To get around this problem, [38] suggests the following parametrization:

$$p(x) = \frac{\exp(\langle \beta, x' \rangle)}{1 + \exp(\langle \beta, x' \rangle)} = \frac{1}{1 + \exp(-\langle \beta, x' \rangle)}.$$

Where $x' := 1 \oplus x$.

Note that, with this parametrization, we guarantee that $p(x) \in (0,1)$ for each $x \in \mathbb{R}^k$ and then our requirement is satisfied. Moreover, the decision space that we wanted is retained, allowing us to look after an optimal vector $\beta^* \in \mathbb{R}^{k+1}$. It is also important to remark that the function $F\colon \mathbb{R} \to \mathbb{R}$ given by:

$$F(\theta) \coloneqq \frac{\exp(\theta)}{1+\exp(\theta)} = \frac{1}{1+\exp(\theta)}$$

is the *logistic function* and also has applications in, for example, differential equations [37]. This fact motivates the problem that we are working on to be called *logistic regression*.

Even though the choice of the logistic function for the parametrization was influenced by a medical application due to the symmetric sigmoid shape of its graphic, when we look at its inverse applied to our problem, we obtain:

$$\log\left(\frac{p(x)}{1-p(x)}\right) = \langle \beta, x' \rangle = \log\left(\frac{\mathbb{P}(Y=1 \mid X=x)}{\mathbb{P}(Y=0 \mid X=x)}\right).$$

This provides the statistical interpretation for the parametrization as the latter is the logarithm of the odd ratio of the event $\{Y=1|X=x\}$.

We are now left with the problem of estimating the β coefficients of the function $p(x)$. The standard way to solve this final problem is the maximum likelihood method, which simply consists in finding the coefficient vector β that maximizes the likelihood function, which in this case is given by:

$$\begin{aligned} S(\{(x_i,y_i)\}_{i\in I},\beta) &= \prod_{i\in I} p(x_i)^{y_i}(1-p(x_i))^{1-y_i} \\ &= \prod_{i\in I}\left(\frac{\exp(\langle\beta,x'\rangle)}{1+\exp(\langle\beta,x_i'\rangle)}\right)^{y_i}\left(\frac{1}{1+\exp(\langle\beta,x_i'\rangle)}\right)^{1-y_i}. \end{aligned}$$

Intuitively, to maximize the likelihood function is to find the coefficient vector β that makes the sample $\{(x_i,y_i)\}_{i\in I}$ as plausible as possible.

5.4.3 A GP Formulation for Logistic Regression

At this point, all that is left to do is to maximize the function $S(\{x_i,y_i\}_{i\in I},\beta)$. First, define the sets $I_0 \coloneqq \{i \in I : y_i = 0\}$ and $I_1 \coloneqq I \setminus I_0$ and note that $I_0 \cap I_1 = \varnothing$ and $I_0 \cup I_1 = I$. Also consider $x_i' \coloneqq 1 \oplus x_i$ for each $i \in I$. Then:

$$S(\{(x_i,y_i)\}_{i\in I},\beta) = \prod_{i\in I_1} p(x_i) \prod_{i\in I_0}(1-p(x_i)).$$

Next, we apply the natural logarithm to both sides of the previous equation, obtaining:

$$
\begin{aligned}
\log(S(\{(x_i, y_i)\}_{i\in I}, \beta)) &= \log\left(\prod_{i\in I_1} p(x_i) \prod_{i\in I_0}(1 - p(x_i))\right) \\
&= \log\left(\prod_{i\in I_1}\left(\frac{\exp(\langle\beta, x_i'\rangle)}{1+\exp(\langle\beta, x_i'\rangle)}\right)\prod_{i\in I_0}\left(\frac{1}{1+\exp(\langle\beta, x_i'\rangle)}\right)\right) \\
&= \sum_{i\in I_1}\log\left(\frac{\exp(\langle\beta, x_i'\rangle)}{1+\exp(\langle\beta, x_i'\rangle)}\right) + \sum_{i\in I_0}\log\left(\frac{1}{1+\exp(\langle\beta, x_i'\rangle)}\right) \\
&= \sum_{i\in I_1}\log(\exp(\langle\beta, x_i'\rangle)) - \sum_{i\in I_1}\log(1+\exp(\langle\beta, x_i'\rangle)) - \sum_{i\in I_0}\log(1+\exp(\langle\beta, x_i'\rangle)) \\
&= \sum_{i\in I_1}\log(\exp(\langle\beta, x_i'\rangle)) - \sum_{i\in I}\log(1+\exp(\langle\beta, x_i'\rangle)) \\
&= \sum_{i\in I_1}\langle\beta, x_i'\rangle - \sum_{i\in I}\log(1+\exp(\langle\beta, x_i'\rangle)).
\end{aligned}
$$

And then we derived the following optimization problem:

$$
\begin{array}{ll}
\text{maximize} & \sum_{i\in I_1}\log(\exp(\langle\beta, x_i'\rangle)) - \sum_{i\in I}\log(1+\exp(\langle\beta, x_i'\rangle)) \\
\text{subject to} & \beta \in \mathbb{R}^{k+1}.
\end{array}
$$

Finally, we consider the homomorphism $b \coloneqq \exp(\beta)$. Substituting in the objective function of the preceding problem yields:

$$
\begin{aligned}
&\sum_{i\in I_1}\langle\beta, x_i'\rangle - \sum_{i\in I}\log(1+\exp(\langle\beta, x_i'\rangle)) \\
&= \sum_{i\in I_1}\langle\log(b), x_i'\rangle - \sum_{i\in I}\log(1+\exp(\langle\log(b), x_i'\rangle)) \\
&= \sum_{i\in I_1}\log(b^{x_i'}) - \sum_{i\in I}\log(1+\exp(\log(b^{x_i'}))) \\
&= \log(\prod_{i\in I_1} b^{x_i'}) - \log(\prod_{i\in I}(1+b^{x_i'})) \\
&= \log\left(\frac{\prod_{i\in I_1} b^{x_i'}}{\prod_{i\in I}(1+b^{x_i'})}\right).
\end{aligned}
$$

Since the natural logarithm is a strictly increasing function, it suffices to analyze its argument, which is clearly nonnegative. Set $f\colon \mathbb{R}^{k+1}_{++} \to \mathbb{R}$ given by:

$$
f(b) \coloneqq \frac{\prod_{i\in I_1} b^{x_i'}}{\prod_{i\in I}(1+b^{x_i'})} \text{ for each } b \in \mathbb{R}^{k+1}_{++}.
$$

Note that the multiplicative inverse of f is a posynomial. Therefore, the logistic regression problem is equivalent to the following GP:

$$\begin{array}{ll} \text{minimize} & \dfrac{1}{f(b)} \\ \text{subject to} & b \in \mathbb{R}^{k+1}_{++}. \end{array}$$

Concluding Remarks

Convex analysis is the most important tool to understand conic optimization from our theoretical point of view. In this text we touched only on the crucial aspects to our purposes, ommiting other interesting topics. Most of the theory of convex functions and Fenchel conjugates, and deeper consequences of Carathéodory's Theorem are among the points that remained undiscussed. All these can be found in [5; 7; 30]. Also, there are several alternative approaches for convex analysis, we recommend [2] for a introduction of the subject in potentially infinite-dimensional vector spaces, and [28] for the study of convexity in metric spaces.

Many of the results we presented also deserve further attention. For example, the theorems of alternative at the end of Section 2.1 may be refined concerning any nonempty convex set C. In this case, one should derive conditions depending on the *barrier cone* of C and its dual, that happens to be 0^+C. In special, Proposition 76 seems to be an endless source of potentially interesting corollaries and refining it could extend the scope of its applications. We also presented several results concerning closed convex sets, e.g Theorems 68, 25, and Proposition 65. These proofs can easily be replicated regarding (relatively) open sets as well. Considering this modification, one obtains the same results in a slightly more general setting. Finally, most of the theory we covered is formulated in [30] in terms of convex functions.

The cones we presented in Chapter 3 can be in a more general context. Actually, both $\mathbb{G}_n$ and $\mathbb{H}_n$ are conceived as the epigraph of the *perspective* of a convex function and, in this sense, can be regarded as members of a more comprehensive family of cones. The book [7] defines and presents basic results on perspectives of convex functions. Additionally, in [9] the authors doubt if semidefinite programs can also be formulated as a special case of REP. In Chapter 4, we used the semidefinite cone to write $\mathbb{L}_2$ as $\mathbb{H}_2$. Those results show that we can write $\mathbb{S}^2_+$ as $\mathbb{H}_2$. If we were able to write $\mathbb{S}^n_+$ as copies of $\mathbb{S}^2_+$, this would be sufficient to formulate semidefinite programs as REP.

As mentioned above, we approached conic programming from a theoretical perspective. For this reason, algorithmic aspects were outside of the scope of this text. The main algorithms for solving conic problems are the so-called interior-point methods. These are studied in [20] while [26] contains tons of information on how to solve conic problems in practice.

Other variants of optimization problems are abundant in literature. Linear programs (LP) are by far the most popular 'family' of optimization problems. They are indeed a particular case of conic problems defined over the nonnegative orthant. However, familirity with this theory is absolutely helpful to improve on intuition and we recommend [34] for the interested reader. Integer programming requires linear programming as a prerequisite and we suggest [11] as a reference. Combinatorial optimization also demands some knowledge of graph theory added to LP. For these we indicate [6] and [12], respectively. Non-linear programming (NLP) is based on multidimensional real-analysis even though there are structural similarities beetween LP and NLP, specially in duality theory. The book [4] provides more information about this topic. More generally, our definition of optimization is quite unpopular. For this reason, the properties embedded in the definition can be further explored. Also, other formulation 'tricks' apart from invertible functions preserving objective values may be formalized using our conception of equivalence.

Despite of the simplicity of the applications we presented, it is pleasant to see that they coincide with "the three main learning problems" pointed out in [36] and they could be studied in more general ways.

For instance, similar types of linear regression that aim to shrink the coefficient vector have already been extensively explored. The addition of an 'extra' term to the loss function is often called a *penalization*. See, e.g [24; 35] for lasso regression. To unify the approach to this family of regression problems, it would be convenient that each of the penalization terms could be represented in terms of $\mathbb{H}_n$. Because it is common to penalize the loss function proportionally to some of the p-norms of the coefficient vector β, expressing each of the cones $\mathbb{L}_n^p$ in terms of $\mathbb{H}_n$ would help handling this task.

Our approach to logistic regression could also have been more general in at least two ways. The first one would be to admit more outcomes, in this case we would solve a *multinomial classification* problem. The second possibility is to modify our loss function so that our decision function changes its bounds or we could even claim that there is not enough evidence to decide the value of a future outcome. In this scenario, we could still use the same method to estimate $p(x)$.

Anyway, it seems to be possible to combine this problems to produce more sofisticated models. Also, in virtue of the scarce literature on REP, there is much work to be done in the applications of this problems, including other areas of science.

Bibliography

[1] Barry C Arnold. *Pareto distribution.* Wiley Online Library, 2015. 32

[2] Viorel Barbu and Teodor Precupanu. *Convexity and optimization in Banach spaces.* Springer Science & Business Media, 2012. 75

[3] Aharon Ben-Tal and Arkadi Nemirovski. On polyhedral approximations of the second-order cone. *Mathematics of Operations Research*, 26(2):193–205, 2001. 60

[4] Dimitri P Bertsekas. *Nonlinear programming.* Athena scientific Belmont, 1999. 75

[5] Dimitri P Bertsekas, Angelia Nedi, Asuman E Ozdaglar, et al. Convex analysis and optimization. 2003. 75

[6] John Adrian Bondy and Uppaluri Siva Ramachandra Murty. Graph theory, volume 244 of graduate texts in mathematics, 2008. 75

[7] Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge university press, 2004. 47, 75

[8] Stephen Boyd, Seung-Jean Kim, Lieven Vandenberghe, and Arash Hassibi. A tutorial on geometric programming. *Optimization and engineering*, 8(1):67, 2007. 54

[9] Venkat Chandrasekaran and Parikshit Shah. Relative entropy optimization and its applications. *Mathematical Programming*, 161(1-2):1–32, 2017. 55, 62, 75

[10] Kai Lai Chung. *A course in probability theory.* 2001. 26

[11] Michele Conforti, Gerard Cornuejols, and Giacomo Zambelli. *Integer Programming.* Springer Publishing Company, Incorporated, 2014. ISBN 3319110071, 9783319110073. 6, 75

[12] William J Cook, WH Cunningham, WR Pulleyblank, and A Schrijver. Combinatorial optimization. *Oberwolfach Reports*, 5(4):2875–2942, 2009. 75

[13] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing).* Wiley-Interscience, 2006. ISBN 0471241954. 55

[14] Marcel Kenji de Carli Silva. Geometric ramifications of the lovász theta function and their interplay with duality. 2013. 49

[15] Morris H. DeGroot. *Optimal statistical decisions.* Wiley classics library. Wiley-Interscience, wiley classics library ed edition, 2004. ISBN 9780471680291,0-471-68029-X. 33, 34

[16] Hamza Fawzi. Topics of convex optimization - lecture notes. (6), 2017. URL http://www.damtp.cam.ac.uk/user/hf323/lent2017_topics_convex_optimisation/lecture6.pdf. 47

[17] Christodoulos A Floudas and Panos M Pardalos. *Encyclopedia of optimization*, volume 1. Springer Science & Business Media, 2001. 1

[18] Andy Gardner and John J Welch. A formal theory of the selfish gene. *Journal of evolutionary biology*, 24(8):1801–1813, 2011. 1

[19] François Glineur. Conic optimization: an elegant framework for convex optimization. *Belgian Journal of Operations Research, Statistics and Computer Science*, 41(1-2):5–28, 2001. 1

[20] François Glineur. *Topics in convex optimization: interior-point methods, conic duality and approximations.* PhD thesis, Polytechnic College of Mons, 2001. 75

[21] François Glineur. An extended conic formulation for geometric optimization, 2000. 53, 70

[22] Paul Richard Halmos. *Finite-dimensional vector spaces.* Springer Science & Business Media, 2012. 4

[23] A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67, 1970. 65

[24] Robert Tibshirani Jerome Friedman, Trevor Hastie. *The Elements of Statistical Learning.* 2001. 64, 76

[25] Richard A Johnson, Richard A Johnson, and Dean W Wichern. *Applied multivariate stastistical analysis.* Prentice-Hall of India Private Limited, 2003. 65

[26] Hans D Mittelmann. The state-of-the-art in conic optimization software. In *Handbook on Semidefinite, Conic and Polynomial Optimization*, pages 671–686. Springer, 2012. 75

[27] J Munkres. Topology. featured titles for topology series, 2000. 8

[28] Athanase Papadopoulos. Metric spaces, convexity and nonpositive curvature (irma lectures in mathematics and theoretical physics, vol. 6). 75

[29] Murray H Protter. *Basic elements of real analysis.* Springer Science & Business Media, 2006. 4, 10

[30] Ralph Tyrell Rockafellar. *Convex analysis.* Princeton university press, 2015. 1, 5, 6, 9, 16, 47, 75

[31] George G. Roussas. *An Introduction to Measure-Theoretic Probability.* Academic Press, 2 edition, 2014. ISBN 978-0-12-800042-7. 26

[32] Walter Rudin et al. *Principles of mathematical analysis*, volume 3. McGraw-hill New York, 1964. 4, 10

[33] Alexander Shapiro. On duality theory of conic linear problems. In *Semi-infinite programming*, pages 135–165. Springer, 2001. 47

[34] Bertsimas Tsitskilis. *Introduction To Linear Optimization.* 75

[35] Vladimir N. Vapnik. *Statistical Learning Theory.* Adaptive and learning systems for signal processing, communications, and control. Wiley, 1998. 32, 33, 76

[36] Vladimir Naumovich Vapnik. An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5):988–999, 1999. 76

[37] PF Verhulst. Recherches mathématiques sur la loi d'accroissement de la population, nouveaux mémoires de l'académie royale des sciences et belles-lettres de bruxelles, 18, art. 1, 1-45. 1845. 72

[38] Strother H. Walker and David B. Duncan. Estimation of the probability of an event as a function of several independent variables. *Biometrika*, 54(1/2):167–179, 1967. ISSN 00063444. URL http://www.jstor.org/stable/2333860. 71