MINISTRY OF EDUCATION
AND TRAINING

VIETNAM ACADEMY
OF SCIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY OF SCIENCE AND TECHNOLOGY

Nguyen Viet Anh

SEPARATION THEOREMS AND RELATED PROBLEMS

MASTER THESIS IN MATHEMATICS

Hanoi, 2022

MINISTRY OF EDUCATION AND TRAINING

VIETNAM ACADEMY OF SCIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY OF SCIENCE AND TECHNOLOGY

Nguyen Viet Anh

SEPARATION THEOREMS AND RELATED PROBLEMS

Major: Applied Mathematics
Code: 8 46 01 12

MASTER THESIS IN MATHEMATICS

ADVISOR:
Dr. Le Xuan Thanh

Hanoi, 2022

Commitment

This thesis is done by my own study under the supervision of Dr. Le Xuan Thanh. It has not been defensed in any council and has not been published on any media. The results as well as the ideas of other authors are all specifically cited. I take full responsibility for my commitment.

Hanoi, October 2022

Nguyen Viet Anh

Acknowledgements

Firstly, I am extremely grateful for my advisor - Dr. Le Xuan Thanh - who devotedly guided me to learn some interesting fields in Optimization and taught me to enjoy the topic of my master thesis. He shared his research experience and career opportunities to me, and help me to find a way for my research plan.

In the time I study here, I sincerely thank all of my lecturers for teaching and helping me, and to the Institute of Mathematics, Hanoi for offering me facilitation in a professional working environment.

I would like to say thanks for the help of Graduate University of Science and Technology, Vietnam Academy of Science and Technology in the time of my master program.

Especially, I really appreciate my family and my friends for their supporting in my whole life.

Hanoi, October 2022

Nguyen Viet Anh

Contents

Introduction

An important topic in the field of optimization theory is separation involving convex sets. A number of separation theorems concerning different types of separation between convex sets have been conducted in literature. Also a number of important results in convex analysis, optimization theory, and functional analysis base on these separation theorems. Namely, the homogeneous Farkas lemma, which gives a condition that is necessary and sufficient for the feasibility of a particular case of homogeneous linear systems, can be obtained from a separation theorem. The theory of duality in convex programming and the construction of convex barrier functions can also be obtained from the separation theorems. Additionally, a cornerstone in functional analysis - the Hahn-Banach theorem - can be derived from a separation theorem.

With the aim of understanding the importance of the separation theorems, we use Chapter 6 in [1] as the main reference, and study some types of separation between two convex sets, together with their applications in the related problems mentioned above. In Chapter 1 we recall some preliminaries for the contents in the sequel chapters. In Chapter 2 we recall some popular separation concepts including general separation, strict separation, strong separation, and proper separation. These concepts are considered in both settings of finite dimensional Euclidean vector spaces and general vector spaces without any equipped topology. It is worth noting that, in this thesis, we only consider vector spaces over the field of real numbers. In Chapter 3 we present detail arguments to derive the homogeneous Farkas lemma, the theorem on dual cone, the construction of a barrier convex function for convex optimization problem, and the Hahn-Banach theorem from the separation theorems.

Chapter 1

Preliminaries

In this chapter, we recall some preliminaries in convex analysis, that will be used in the sequel chapters. Throughout this chapter (except for the last section), E is a vector space equipped with a norm $\|\cdot\|$ induced by an inner product $\langle \cdot, \cdot \rangle$. In the last section of this chapter, we will consider E as a general vector space without any equipped topology.

1.1 Affine sets

Definition 1.1. (Affine set, see e.g. [2]). *A subset $A \subset E$ is called an affine set if for every $\mathbf{a}, \mathbf{b} \in A$ and $\lambda \in \mathbb{R}$ we have $\lambda \mathbf{a} + (1-\lambda)\mathbf{b} \in A$.*

Given two distinct points $\mathbf{a}, \mathbf{b} \in E$, we define the line through these points as the set of form $\{\mathbf{x} \in E \mid \mathbf{x} = \lambda \mathbf{a} + (1-\lambda)\mathbf{b} \text{ for some } \lambda \in \mathbb{R}\}$. It is not hard to see that such a line is an affine set, and a subset $A \subset E$ is affine if and only if the line through any pair of distinct points in A is also contained in A.

Definition 1.2. (Hyperplane, see e.g. [1]). *A hyperplane in E is a set of form*

$$H(\mathbf{a}, \alpha) = \{\mathbf{x} \in E \mid \langle \mathbf{a}, \mathbf{x} \rangle = \alpha\}$$

for some $\mathbf{a} \in E \backslash \{\mathbf{0}\}$ and $\alpha \in \mathbb{R}$.

It is also not hard to see that a hyperplane is an affine set.

Definition 1.3. (Affine hull, see e.g. [2]). *Given a subset $A \subset E$. The affine hull of A, denoted $\operatorname{aff}(A)$, is the smallest affine set in E containing A (in sense of set inclusion).*

The following proposition is a well-known result about the structure of the affine hull.

Proposition 1.4. (See e.g. [2]) *For a given subset $A \subset E$, its affine hull* $\text{aff}(A)$ *coincides the set of all affine combinations of its points, i.e.,*

$$\text{aff}(A) = \{\theta_1 \mathbf{x}^1 + \ldots + \theta_k \mathbf{x}^k \mid \mathbf{x}^1, \ldots, \mathbf{x}^k \in A, \theta_1 + \ldots + \theta_k = 1\}.$$

Definition 1.5. (Relative interior, see e.g. [3]). *Given a subset $A \subset E$. The relative interior of A, denoted* $\text{relint}(A)$*, is the set*

$$\{\mathbf{x} \in A \mid \exists \epsilon > 0 : B(\mathbf{x}, \epsilon) \cap \text{aff}(A) \subset A\},$$

in which $B(\mathbf{x}, \epsilon) = \{\mathbf{y} \in E \mid \|\mathbf{y} - \mathbf{x}\| < \epsilon\}$.

Roughly speaking, the relative interior of a subset of $\mathbb{R}^n$ is the interior of that set relative to its affine hull.

1.2 Convex sets

Definition 1.6. (Convex set, see e.g. [3]). *A subset $C \subset E$ is called a convex set if for every $\mathbf{a}, \mathbf{b} \in C$ and $\lambda \in [0, 1]$ we have $\lambda\mathbf{a} + (1 - \lambda)\mathbf{b} \in C$.*

Given two distinct points $\mathbf{a}, \mathbf{b} \in E$, we define the line segment $[\mathbf{a}, \mathbf{b}]$ between these points as the set $\{\mathbf{x} \in E \mid \mathbf{x} = \lambda\mathbf{a} + (1 - \lambda)\mathbf{b} \text{ for some } \lambda \in [0, 1]\}$. It is not hard to see that such a line segment is a convex set, and a subset $C \subset E$ is convex if and only if the line segment between any pair of distinct points in C is also contained in C. It is also not hard to see that a hyperplane in E is a convex set.

Similar to the affine hull, we have the following concept.

Definition 1.7. (Convex hull, see e.g. [2]). *Given a subset $C \subset E$. The convex hull of C, denoted* $\text{conv}(C)$*, is the smallest convex set in E containing C (in sense of set inclusion).*

The following proposition is a well-known result about structure of the convex hull.

Proposition 1.8. (See e.g. [2]) *For a given subset $C \subset E$, its convex hull* $\text{conv}(C)$ *coincides the set of all convex combinations of its points, i.e.,*

$$\text{conv}(C) = \{\theta_1 \mathbf{x}^1 + \ldots + \theta_k \mathbf{x}^k \mid \mathbf{x}^1, \ldots, \mathbf{x}^k \in A, \theta_1, \ldots, \theta_k \geq 0, \theta_1 + \ldots + \theta_k = 1\}.$$

The following proposition provides some useful properties of convex sets.

Proposition 1.9. *(i) The closure $\overline{C}$ of any convex set $C \subset E$ is also convex.*

(ii) Let C_1 and C_2 be convex sets in E. Then $C_1 \cap C_2$, $C_1 + C_2$, $C_1 - C_2$ are also convex.

Proof. (i) Let $\lambda \in [0,1]$ and $\mathbf{x}, \mathbf{y} \in \overline{C}$. There exist sequences $\{\mathbf{x}^n\}, \{\mathbf{y}^n\}$ in C such that $\mathbf{x}^n \to \mathbf{x}$ and $\mathbf{y}^n \to \mathbf{y}$ as $n \to \infty$. Since C is convex, we have $\lambda \mathbf{x}^n + (1-\lambda)\mathbf{y}^n \in C$ for all $n \in \mathbb{N}$. Taking $n \to \infty$ we have $\lambda \mathbf{x} + (1-\lambda)\mathbf{y} \in \overline{C}$, which shows that $\overline{C}$ is convex.

(ii) Let $\mathbf{x}^1, \mathbf{x}^2 \in C_1 \cap C_2$, and $\theta \in [0,1]$. Since $\mathbf{x}^1, \mathbf{x}^2 \in C_1$, by convexity of C_1 we have $\theta \mathbf{x}^1 + (1-\theta)\mathbf{x}^2 \in C_1$. Similarly, since $\mathbf{x}^1, \mathbf{x}^2 \in C_2$, by convexity of C_2 we have $\theta \mathbf{x}^1 + (1-\theta)\mathbf{x}^2 \in C_2$. Thus, $\theta \mathbf{x}^1 + (1-\theta)\mathbf{x}^2 \in C_1 \cap C_2$, which proves the convexity of $C_1 \cap C_2$.

Let $\lambda \in [0,1]$ and $\mathbf{u}, \mathbf{v} \in C_1 + C_2$. Since $\mathbf{u}, \mathbf{v} \in C_1 + C_2$, there exist $\mathbf{u}^1, \mathbf{v}^1 \in C_1$ and $\mathbf{u}^2, \mathbf{v}^2 \in C_2$ such that $\mathbf{u} = \mathbf{u}^1 + \mathbf{u}^2, \mathbf{v} = \mathbf{v}^1 + \mathbf{v}^2$. Since $\mathbf{u}^1, \mathbf{v}^1 \in C_1$, by convexity of C_1 we have $\lambda \mathbf{u}^1 + (1-\lambda)\mathbf{v}^1 \in C_1$. Similarly, since $\mathbf{u}^2, \mathbf{v}^2 \in C_2$, by convexity of C_2 we have $\lambda \mathbf{u}^2 + (1-\lambda)\mathbf{v}^2 \in C_2$. Therefore we have

$$\begin{aligned}\lambda \mathbf{u} + (1-\lambda)\mathbf{v} &= \lambda(\mathbf{u}^1 + \mathbf{u}^2) + (1-\lambda)(\mathbf{v}^1 + \mathbf{v}^2)\\ &= (\lambda \mathbf{u}^1 + (1-\lambda)\mathbf{v}^1) + (\lambda \mathbf{u}^2 + (1-\lambda)\mathbf{v}^2) \in C_1 + C_2.\end{aligned}$$

Thus $C_1 + C_2$ is convex. By similar arguments we obtain convexity of the set $C_1 - C_2$. □

Additionally, the following proposition gives some non-trivial properties of convex sets in finite dimensional spaces.

Proposition 1.10. *(i) Any nonempty convex set in $\mathbb{R}^n$ has nonempty relative interior.*

(ii) Let $C_1, C_2 \subset \mathbb{R}^n$ be nonempty convex sets. Then we have

$$\operatorname{relint}(C_1 - C_2) = \operatorname{relint}(C_1) - \operatorname{relint}(C_2).$$

For the proof of Proposition 1.10(i), we refer to Proposition 1.9 in [2]. For the proof of Proposition 1.10(ii), we refer to Corollary 2.87 in [4].

The following proposition gives an additional property of points in relative interior of a convex set.

Proposition 1.11. *Let C be a nonempty convex set in E, $\mathbf{x} \in \operatorname{relint}(C)$, and $\mathbf{y} \in C$. Then there exists $t > 0$ for which $\mathbf{x} + t(\mathbf{x} - \mathbf{y}) \in C$.*

Proof. For any $t \in \mathbb{R}$, we have $\mathbf{x} + t(\mathbf{x} - \mathbf{y}) = (1+t)\mathbf{x} - t\mathbf{y}$ is an affine combination of $\mathbf{x}$ and $\mathbf{y}$ (since the sum of coefficients in this combination is $1 + t - t = 1$). Furthermore, since $\mathbf{x} \in \operatorname{relint}(C) \subset C$ and $\mathbf{y} \in C$, this affine combination is in affine hull of C, that is

$$\mathbf{x} + t(\mathbf{x} - \mathbf{y}) \in \operatorname{aff}(C). \tag{1.1}$$

Since $\mathbf{x} \in \operatorname{relint}(C)$, there exists $r > 0$ such that $B(\mathbf{x}, r) \cap \operatorname{aff}(C) \subset C$. By choosing t such that $0 < t < \frac{r}{\|\mathbf{x}-\mathbf{y}\|}$ we have

$$\mathbf{x} + t(\mathbf{x} - \mathbf{y}) \in B(\mathbf{x}, r). \tag{1.2}$$

For such choice of t we have both (1.1) and (1.2), and consequently

$$\mathbf{x} + t(\mathbf{x} - \mathbf{y}) \in B(\mathbf{x}, r) \cap \operatorname{aff}(C) \subset C.$$

□

We will need the following result in the sequel.

Lemma 1.12. *Let C be a nonempty convex set in E and $\bar{\mathbf{x}} \in \overline{C} \backslash \operatorname{relint}(C)$. Then there exists a sequence $\{\mathbf{x}^k \mid k \in \mathbb{N}\} \subset \operatorname{aff}(C)$ with $\mathbf{x}^k \notin \overline{C}$ and $\mathbf{x}^k \to \bar{\mathbf{x}}$ as $k \to \infty$.*

Proof. Note that $\operatorname{relint}(C)$ is non-empty by Proposition 1.10(i), therefore we can take $\mathbf{x}^0$ as a point in $\operatorname{relint}(C)$. We shall begin with showing that $(1+t)\bar{\mathbf{x}} - t\mathbf{x}^0 \notin \overline{C}$ for all $t > 0$. Indeed, assume the contrary that $(1+t)\bar{\mathbf{x}} - t\mathbf{x}^0 \in \overline{C}$ for some $t > 0$. This, together with the fact that $\mathbf{x}^0 \in \operatorname{relint}(C)$, ensures that the following affine combination

$$\bar{\mathbf{x}} = \frac{t}{t+1}\mathbf{x}^0 + \frac{1}{t+1}\left((t+1)\bar{\mathbf{x}} - t\mathbf{x}^0\right)$$

is in $\operatorname{relint}(C)$. However, this contradicts the assumption $\bar{\mathbf{x}} \notin \operatorname{relint}(C)$.

Now, by choosing $t = \frac{1}{k}$ for $k = 1, 2, \ldots$, we obtain $\mathbf{x}^k := \left(1 + \frac{1}{k}\right)\bar{\mathbf{x}} - \frac{1}{k}\mathbf{x}^0 \notin C$. Each $\mathbf{x}^k$ is an affine combination of $\bar{\mathbf{x}} \in \overline{C} \backslash \operatorname{relint}(C)$ and $\mathbf{x}^0 \in \operatorname{relint}(C)$, hence it is in $\operatorname{aff}(C)$. By letting $k \to \infty$, we have

$$\mathbf{x}^k := \left(1 + \frac{1}{k}\right)\bar{\mathbf{x}} - \frac{1}{k}\mathbf{x}^0 \to \bar{\mathbf{x}}.$$

□

1.3 Conic sets

Definition 1.13. (See e.g. [3]). *(i) A subset $K \subset E$ is called a cone if for every $\mathbf{a} \in K$ and $\lambda \geq 0$ we have $\lambda\mathbf{a} \in K$.*

(ii) A conic combination of points $\mathbf{x}^1, \ldots, \mathbf{x}^k \in E$ is a point of form

$$\lambda_1\mathbf{x}^1 + \ldots + \lambda_k\mathbf{x}^k$$

with $\lambda_1, \ldots, \lambda_k \geq 0$.

(iii) The conic hull of a given subset $C \subset E$, denoted $\operatorname{cone}(C)$ is the set of all conic combinations of points in C.

Similar to the case of affine hulls, we have the following well-known result about conic hulls.

Proposition 1.14. (See e.g. [3]). *The conic hull* $\text{cone}(C)$ *of a subset* $C \subset E$ *is the smallest convex cone containing* C *(in sense of set inclusion).*

1.4 Projection on convex sets

Proposition 1.15. (See e.g. [1]). *Let* $C \subset \mathbb{R}^n$ *be a nonempty closed convex set. Let* $\mathbf{x} \in \mathbb{R}^n$*. Then there exists uniquely a vector* $\mathbf{x}^* \in C$ *such that*

$$\|\mathbf{x} - \mathbf{x}^*\| = \min_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|.$$

Proof. *Existence.* Firstly, we observe that the function $f(\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|$ is continuous on $\mathbb{R}^n$. Indeed, let $\mathbf{y}^0$ be an arbitrary vector in $\mathbb{R}^n$ and $\{\mathbf{y}^n \mid n \in \mathbb{N}\}$ a sequence in $\mathbb{R}^n$ converging to $\mathbf{y}^0$, *i.e.,* $\|\mathbf{y}^n - \mathbf{y}^0\| \to 0$ as $n \to \infty$. For any $n \in \mathbb{N}$ we have

$$\|\mathbf{y}^n - \mathbf{y}^0\| = \|(\mathbf{x}-\mathbf{y}^n) - (\mathbf{x}-\mathbf{y}^0)\| \geq \Big| \|\mathbf{x} - \mathbf{y}^n\| - \|\mathbf{x} - \mathbf{y}^0\| \Big| = \Big| f(\mathbf{y}^n) - f(\mathbf{y}^0) \Big| \geq 0.$$

It follows that $f(\mathbf{y}^n) \to f(\mathbf{y}^0)$ as $n \to \infty$, *i.e.,* $f(\mathbf{y})$ is continuous at $\mathbf{y}^0$. Since $\mathbf{y}^0$ is chosen arbitrarily in $\mathbb{R}^n$, we obtain the continuity of f on $\mathbb{R}^n$.

Now, fix $\mathbf{y}^* \in C$ and define

$$C_{y*} = \{\mathbf{y} \in C \mid \|\mathbf{x} - \mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}^*\|\}.$$

Since C is closed, so is C_{y*}. Clearly, C_{y*} is bounded, so it is compact. Since f is continuous, by Bolzano-Weierstrass theorem, f achieves its minimum on the compact set C_{y*} at some $\mathbf{x}^* \in C_{y*} \subset C$, *i.e.,*

$$\|\mathbf{x} - \mathbf{x}^*\| = \min_{\mathbf{y} \in C_{y*}} \|\mathbf{x} - \mathbf{y}\|.$$

Furthermore, for any $\mathbf{y} \notin C_{y*}$, by definition of C_{y*} we have $\|\mathbf{x} - \mathbf{y}\| > \|\mathbf{x} - \mathbf{y}^*\|$. It means that

$$\min_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\| = \min_{\mathbf{y} \in C_{y*}} \|\mathbf{x} - \mathbf{y}\| = \|\mathbf{x} - \mathbf{x}^*\|.$$

Uniqueness. Assume that $\mathbf{x}^1$ and $\mathbf{x}^2$ are minimizers of f over C. That means $\mathbf{x}^1, \mathbf{x}^2 \in C$ and

$$\|\mathbf{x} - \mathbf{x}^1\| = \|\mathbf{x} - \mathbf{x}^2\| = \min_{\mathbf{y} \in C} f(\mathbf{y}) = \min_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\| := m.$$

Let $\bar{\mathbf{x}} = \frac{1}{2}(\mathbf{x}^1 + \mathbf{x}^2)$. Since $\mathbf{x}^1, \mathbf{x}^2 \in C$ and C is convex, we have $\bar{\mathbf{x}} \in C$, thus $\|\mathbf{x} - \bar{\mathbf{x}}\| \geq m$. Note that

$$\|\mathbf{x}^1 - \mathbf{x}^2\|^2 = \|(\mathbf{x}^1 - \mathbf{x}) - (\mathbf{x}^2 - \mathbf{x})\|^2$$

$$= \|\mathbf{x}^1 - \mathbf{x}\|^2 + \|\mathbf{x}^2 - \mathbf{x}\|^2 - 2\langle \mathbf{x}^1 - \mathbf{x}, \mathbf{x}^2 - \mathbf{x}\rangle$$

and

$$\begin{aligned}\|\mathbf{x}^1 - \mathbf{x}\|^2 + \|\mathbf{x}^2 - \mathbf{x}\|^2 + 2\langle \mathbf{x}^1 - \mathbf{x}, \mathbf{x}^2 - \mathbf{x}\rangle &= \|(\mathbf{x}^1 - \mathbf{x}) + (\mathbf{x}^2 - \mathbf{x})\|^2 \\ &= 4\left\|\frac{1}{2}(\mathbf{x}^1 + \mathbf{x}^2) - \mathbf{x}\right\|^2 \\ &= 4\|\bar{\mathbf{x}} - \mathbf{x}\|^2,\end{aligned}$$

it follows that

$$\|\mathbf{x}^1 - \mathbf{x}^2\|^2 + 4\|\bar{\mathbf{x}} - \mathbf{x}\|^2 = 2\|\mathbf{x}^1 - \mathbf{x}\|^2 + 2\|\mathbf{x}^2 - \mathbf{x}\|^2,$$

and therefore we have

$$\begin{aligned}0 \le \|\mathbf{x}^1 - \mathbf{x}^2\|^2 &= 2\|\mathbf{x}^1 - \mathbf{x}\|^2 + 2\|\mathbf{x}^2 - \mathbf{x}\|^2 - 4\|\bar{\mathbf{x}} - \mathbf{x}\|^2 \\ &\le 2m^2 + 2m^2 - 4m^2 = 0.\end{aligned}$$

So we must have $\|\mathbf{x}^1 - \mathbf{x}^2\| = 0$, and consequently, $\mathbf{x}^1 = \mathbf{x}^2$. □

Thanks to Proposition 1.15, we can define the projection of a vector $\mathbf{x} \in \mathbb{R}^n$ onto a nonempty closed convex set $C \subset \mathbb{R}^n$ to be $\operatorname{argmin}_{\mathbf{y}\in C}\|\mathbf{x} - \mathbf{y}\|$, denoted by $\operatorname{proj}_C(\mathbf{x})$. The next proposition is a characterization of the projection onto closed convex sets.

Proposition 1.16. (See e.g. [1]) *Given a nonempty closed convex set $C \subset \mathbb{R}^n$ and let $\mathbf{x} \in \mathbb{R}^n$. A vector $\mathbf{z} \in C$ is the projection $\operatorname{proj}_C(\mathbf{x})$ if and only if*

$$\langle \mathbf{x} - \mathbf{z}, \mathbf{y} - \mathbf{z}\rangle \le 0 \quad \forall \mathbf{y} \in C. \tag{1.3}$$

Proof. Sufficiency. Assume that $\mathbf{z}$ is the projection of $\mathbf{x}$ onto C. Since (1.3) holds with $\mathbf{y} = \mathbf{z}$, we consider an arbitrary $\mathbf{y} \in C\backslash\{\mathbf{z}\}$. Since $\mathbf{x}, \mathbf{z} \in C$ and C is convex, for any $\alpha \in (0, 1)$ we have

$$\mathbf{z} + \alpha(\mathbf{y} - \mathbf{z}) = \alpha\mathbf{y} + (1 - \alpha)\mathbf{z} \in C.$$

Recall $\mathbf{z} = \operatorname{proj}_C(\mathbf{x}) = \operatorname{argmin}_{\mathbf{y}\in C}\|\mathbf{x} - \mathbf{y}\|$, we have

$$\begin{aligned}\|\mathbf{x} - \mathbf{z}\|^2 &\le \|\mathbf{x} - (\mathbf{z} + \alpha(\mathbf{y} - \mathbf{z}))\|^2 \\ &= \|\mathbf{x} - \mathbf{z}\|^2 + \alpha^2\|\mathbf{y} - \mathbf{z}\|^2 - 2\alpha\langle \mathbf{x} - \mathbf{z}, \mathbf{y} - \mathbf{z}\rangle,\end{aligned}$$

which implies

$$\langle \mathbf{x} - \mathbf{z}, \mathbf{y} - \mathbf{z}\rangle \le \frac{\alpha}{2}\|\mathbf{x} - \mathbf{z}\|^2.$$

This inequality holds for arbitrary $\alpha \in (0,1)$, therefore by letting $\alpha \to 0^+$ we obtain (1.3).

Necessity. Let $\mathbf{z} \in C$ satisfying (1.3). For any $\mathbf{y} \in C$ such that $\mathbf{y} \neq \mathbf{z}$, we have $\|\mathbf{z}-\mathbf{y}\| > 0$ and $\langle \mathbf{x}-\mathbf{z}, \mathbf{y}-\mathbf{z}\rangle \leq 0$, hence

$$\begin{aligned}\|\mathbf{x}-\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{z}\|^2 &= \|(\mathbf{x}-\mathbf{z}) + (\mathbf{z}-\mathbf{y})\|^2 - \|\mathbf{x}-\mathbf{z}\|^2\\ &= \|\mathbf{z}-\mathbf{y}\|^2 + 2\langle \mathbf{x}-\mathbf{z}, \mathbf{z}-\mathbf{y}\rangle\\ &= \|\mathbf{z}-\mathbf{y}\|^2 - 2\langle \mathbf{x}-\mathbf{z}, \mathbf{y}-\mathbf{z}\rangle\\ &> 0.\end{aligned}$$

From $\|\mathbf{x}-\mathbf{y}\|^2 > \|\mathbf{x}-\mathbf{z}\|^2$ for any $\mathbf{y} \in C\backslash\{\mathbf{z}\}$, we derive

$$\mathbf{z} = \text{argmin}_{\mathbf{y}\in C}\|\mathbf{x}-\mathbf{y}\| = \text{proj}_C(\mathbf{x}).$$

□

Another important property of projection mapping onto closed convex sets is given in the following proposition.

Proposition 1.17. (See e.g. [1]). *Let C be a closed convex set in $\mathbb{R}^n$. Then* proj_C *is nonexpansive in the following sense*

$$\|\text{proj}_C(\mathbf{x}^1) - \text{proj}_C(\mathbf{x}^2)\| \leq \|\mathbf{x}^1 - \mathbf{x}^2\| \quad \forall \mathbf{x}^1, \mathbf{x}^2 \in \mathbb{R}^n. \tag{1.4}$$

Consequently, proj_C *is a continuous mapping.*

Proof. Let $\mathbf{x}^1$, x^2 be arbitrary points in E. We first observe that the inequality (1.4) holds when $\text{proj}_C(\mathbf{x}^1) = \text{proj}_C(\mathbf{x}^2)$. Therefore, we consider the case in which the projections of $\mathbf{x}^1$ and $\mathbf{x}^2$ are distinct.

In view of the inequality (1.3) with $\mathbf{x} = \mathbf{x}^1, \mathbf{y} = \text{proj}_C(\mathbf{x}^2)$, we obtain

$$\langle \mathbf{x}^1 - \text{proj}_C(\mathbf{x}^1), \text{proj}_C(\mathbf{x}^2) - \text{proj}_C(\mathbf{x}^1)\rangle \leq 0. \tag{1.5}$$

We now apply the inequality (1.3) again with $\mathbf{x} = \mathbf{x}^2, \mathbf{y} = \text{proj}_C(\mathbf{x}^1)$, we obtain

$$\langle \mathbf{x}^2 - \text{proj}_C(\mathbf{x}^2), \text{proj}_C(\mathbf{x}^1) - \text{proj}_C(\mathbf{x}^2)\rangle \leq 0. \tag{1.6}$$

Adding (1.5) and (1.6) gives

$$\langle \mathbf{x}^1 - \mathbf{x}^2 + \text{proj}_C(\mathbf{x}^2) - \text{proj}_C(\mathbf{x}^1), \text{proj}_C(\mathbf{x}^2) - \text{proj}_C(\mathbf{x}^1)\rangle \leq 0.$$

Then we have

$$\|\text{proj}_C(\mathbf{x}^2) - \text{proj}_C(\mathbf{x}^1)\|^2 \leq \langle \mathbf{x}^2 - \mathbf{x}^1, \text{proj}_C(\mathbf{x}^2) - \text{proj}_C(\mathbf{x}^1)\rangle$$

$$\leq \|\mathbf{x}^2 - \mathbf{x}^1\| \|\text{proj}_C(\mathbf{x}^2) - \text{proj}_C(\mathbf{x}^1)\|.$$

Note that $\text{proj}_C(\mathbf{x}^1) \neq \text{proj}_C(\mathbf{x}^2)$, then by dividing both sides of above inequality by $\|\text{proj}_C(\mathbf{x}^2) - \text{proj}_C(\mathbf{x}^1)\|$, we obtain the inequality (1.4). It means that the projection mapping proj_C is nonexpansive. The continuity of proj_C follows as a consequence of its nonexpansiveness. □

We close this section with a computational result on the distance from a point to a hyperplane in a finite dimensional space $\mathbb{R}^n$.

Lemma 1.18. *Let $H := H(\mathbf{a}, \alpha) = \{\mathbf{u} \in \mathbb{R}^n \mid \langle \mathbf{a}, \mathbf{u} \rangle = \alpha\}$ be a hyperplane in $\mathbb{R}^n$. Then for any $\mathbf{x} \in \mathbb{R}^n$ we have*

$$\min\{\|\mathbf{x} - \mathbf{y}\| \mid \mathbf{y} \in H\} = \frac{|\langle \mathbf{a}, \mathbf{x} \rangle - \alpha|}{\|\mathbf{a}\|}.$$

Proof. We first note that, as a hyperplane, H is a closed convex set in $\mathbb{R}^n$. Hence, by Proposition 1.15, for fixed $\mathbf{x} \in \mathbb{R}^n$, $\min\{\|\mathbf{x} - \mathbf{y}\| \mid \mathbf{y} \in H\}$ is achieved. Furthermore, as $H = H(\mathbf{a}, \alpha)$ is a hyperplane, we must have $\mathbf{a} \neq \mathbf{0}$. For any $\mathbf{y} \in H$, by Cauchy-Schwartz inequality, we obtain

$$\|\mathbf{a}\|.\|\mathbf{x} - \mathbf{y}\| \geq |\langle \mathbf{a}, \mathbf{x} - \mathbf{y} \rangle| = |\langle \mathbf{a}, \mathbf{x} \rangle - \alpha|,$$

or equivalently

$$\|\mathbf{x} - \mathbf{y}\| \geq \frac{|\langle \mathbf{a}, \mathbf{x} \rangle - \alpha|}{\|\mathbf{a}\|}.$$

We observe that

$$\mathbf{y}^* := \mathbf{x} - \frac{|\langle \mathbf{a}, \mathbf{x} \rangle - \alpha|}{\|\mathbf{a}\|^2}\mathbf{a}$$

satisfying $\langle \mathbf{a}, \mathbf{y}^* \rangle = \alpha$ and $\|\mathbf{x} - \mathbf{y}^*\| = \frac{|\langle \mathbf{a}, \mathbf{x} \rangle - \alpha|}{\|\mathbf{a}\|}$. It readily follows that

$$\min\{\|\mathbf{x} - \mathbf{y}\| \mid \mathbf{y} \in H\} = \frac{|\langle \mathbf{a}, \mathbf{x} \rangle - \alpha|}{\|\mathbf{a}\|}.$$

□

1.5 Convex and concave functions

Definition 1.19. (See e.g. [3]). *A function $f : E \to \mathbb{R} \cup \{+\infty\}$ is said to be convex on a convex set $C \subset E$ if*

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y} \in C, \lambda \in [0, 1].$$

A function $g : E \to \mathbb{R} \cup \{-\infty\}$ is said to be concave on a convex set $C \subset E$ if $-g$ is convex on C.

It is well-known that the pointwise infimum of a set of linear functions is concave. This result is stated more precisely in the following proposition.

Proposition 1.20. (See e.g. [3]). *Let $C \subset \mathbb{R}^n$ be a convex set. For each α in an index set $I \subset \mathbb{R}$, let $f_\alpha : C \to \mathbb{R}$ be a linear function. Then*

$$\begin{aligned} f : C &\to \mathbb{R} \\ \mathbf{x} &\mapsto \inf_{\alpha \in I} f_\alpha(\mathbf{x}) \end{aligned}$$

is a concave function on C.

Proof. For any $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in [0,1]$, we have

$$\begin{aligned} f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= \inf_{\alpha \in I} f_\alpha(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \\ &= \inf_{\alpha \in I} \left(\lambda f_\alpha(\mathbf{x}) + (1-\lambda) f_\alpha(\mathbf{y})\right) \qquad \text{(since each } f_\alpha \text{ is linear)} \\ &\geq \inf_{\alpha \in I} \left(\lambda f_\alpha(\mathbf{x})\right) + \inf_{\alpha} \left((1-\lambda) f_\alpha(\mathbf{y})\right) \\ &= \lambda \inf_{\alpha \in I} f_\alpha(\mathbf{x}) + (1-\lambda) \inf_{\alpha \in I} f_\alpha(\mathbf{y}) \\ &= \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}). \end{aligned}$$

This proves the concavity of f. □

We will also need the following result.

Proposition 1.21. *Let $g : E \to \mathbb{R}$ be a concave function on a convex set $C \subset E$. Let $f : \mathbb{R} \to \mathbb{R}$ be a concave non-decreasing function on $\mathbb{R}$. Then the composition function $h(\mathbf{x}) := f(g(\mathbf{x}))$ is also a concave function on C.*

Proof. Let $\mathbf{x}, \mathbf{y}$ be arbitrary point in C and $\lambda \in [0,1]$. Since C is convex, we have $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ is also in C. By concavity of g on C, we have

$$g(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \geq \lambda g(\mathbf{x}) + (1-\lambda) g(\mathbf{y}).$$

Since f is non-decreasing, it follows that

$$h(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) = f(g(\lambda\mathbf{x} + (1-\lambda)\mathbf{y})) \geq f(\lambda g(\mathbf{x}) + (1-\lambda) g(\mathbf{y})). \tag{1.7}$$

By concavity of f we have

$$f(\lambda g(\mathbf{x}) + (1-\lambda) g(\mathbf{y})) \geq \lambda f(g(\mathbf{x})) + (1-\lambda) f(g(\mathbf{y})) = \lambda h(\mathbf{x}) + (1-\lambda) h(\mathbf{y}). \tag{1.8}$$

From (1.7) and (1.8) we obtain

$$h(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \geq \lambda h(\mathbf{x}) + (1-\lambda) h(\mathbf{y}).$$

This proves the concavity of h on C. □

The following proposition gives us an important and non-trivial property of convex and concave functions on finite dimensional spaces.

Proposition 1.22. *Let C be a nonempty open convex set. If $f : C \subset \mathbb{R}^n \to \mathbb{R}$ is a convex (or concave) function, then it is continuous on C.*

The proof of Proposition 1.22 can be found in e.g. [2], Proposition 2.3.

1.6 Algebraic interior and algebraic closure

In this section, E is a general vector space without any equipped topology. We first note that, as defined in the previous sections, the following concepts do not depend on any topology equipped on the underlying vector space:

- affine sets and affine hull (Definition 1.1 and Definition 1.3),
- convex sets and convex hull (Definition 1.6 and Definition 1.7),
- cones and conic hull (Definition 1.13 and Definition 1.14),
- convex and concave functions (Definition 1.19).

These concepts are also valid in infinite dimensional vector spaces. However, the concept of relative interior (as defined in Definition 1.5), as well as the concept of projection onto convex sets (as defined in Section 1.4), depends on the equipped norm of the underlying vector space.

Let us focus on the concept of relative interior. Figure 1.1 illustrates this concept on an example in $\mathbb{R}^2$ with the usual Euclidean norm. Let $\mathbf{x}^1, \mathbf{x}^2$ be distinct points in $\mathbb{R}^2$, and A the line segment between these points:

$$A = \{\mathbf{x} \mid \mathbf{x} = \theta\mathbf{x}^1 + (1-\theta)\mathbf{x}^2, \theta \in [0,1]\}.$$

Then aff(A) is the line passing through $\mathbf{x}^1$ and $\mathbf{x}^2$. If we take $\mathbf{x}$ as a point inside the line segment, then one can choose $r > 0$ small enough so that $B(\mathbf{x}, r) \cap \text{aff}(A) \subset A$ (as illustrated in Figure 1.1), and therefore such $\mathbf{x}$ is a relative interior of A. However, if we take $\mathbf{x}$ to be either $\mathbf{x}^1$ or $\mathbf{x}^2$, then such r does not exists, so both $\mathbf{x}^1$ and $\mathbf{x}^2$ are not relative interior points of A.

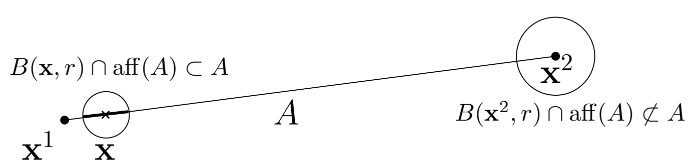

Figure 1.1: Relative interior of a line segment.

With a closer look into this example, we see that $B(\mathbf{x}, r) \cap \text{aff}(A)$ is an open line segment containing $\mathbf{x}$ in the middle. Therefore, intrinsically, in order to have $\mathbf{x}$ as a relative interior of A, the condition $B(\mathbf{x}, r) \cap \text{aff}(A) \subset A$ can be replaced by requiring the following: "*Every line $\ell \in \text{aff}(A)$ through $\mathbf{x}$ contains an open line segment in A such that $\mathbf{x}$ is in the interior of the line segment*". The main advantage of this new condition is that: it depends only on the algebraic structure of the underlying vector space E, and is independent of any norm as well as any topology. In this direction, we can generalize the concept of relative interior to the case in general vector spaces. That gives rise to the following concepts.

Definition 1.23. (Relative algebraic interior and relative algebraic closure, see e.g. [1]). *Let A be a subset in a general vector space E.*

(i) The relative algebraic interior of A, denoted $\text{rai}(A)$, *is defined by*

$$\{\mathbf{x} \in A \mid \forall \mathbf{y} \in \text{aff}(A)\ \exists r > 0 \ s.t.\ [\mathbf{x} - r(\mathbf{y} - \mathbf{x}), \mathbf{x} + r(\mathbf{y} - \mathbf{x})] \subseteq A\}. \tag{1.9}$$

In case $\text{aff}(A) = E$, *we call the above set the algebraic interior of A, and denote* $\text{ai}(A)$ *instead of* $\text{rai}(A)$.

(ii) The relative algebraic closure of A, denoted $\text{rac}(A)$, *is defined by*

$$\{\mathbf{y} \in \text{aff}(A) \mid \exists \mathbf{x} \in A \ s.t.\ [\mathbf{x}, \mathbf{y}) \subset A\}.$$

In case $\text{aff}(A) = E$, *we call the above set the algebraic closure of A, and denote* $\text{ac}(A)$ *instead of* $\text{rac}(A)$.

Concerning the notations in Definition 1.23, for $\mathbf{u}, \mathbf{v} \in E$ we define

$$\begin{aligned}
[\mathbf{u}, \mathbf{v}] = [\mathbf{v}, \mathbf{u}] &= \{\mathbf{w} \in E \mid \exists \lambda \in [0,1] \text{ s.t. } \mathbf{w} = \lambda \mathbf{u} + (1-\lambda)\mathbf{v}\}, \\
(\mathbf{u}, \mathbf{v}) = (\mathbf{v}, \mathbf{u}) &= \{\mathbf{w} \in E \mid \exists \lambda \in (0,1) \text{ s.t. } \mathbf{w} = \lambda \mathbf{u} + (1-\lambda)\mathbf{v}\}, \\
[\mathbf{u}, \mathbf{v}) = (\mathbf{v}, \mathbf{u}] &= \{\mathbf{w} \in E \mid \exists \lambda \in (0,1] \text{ s.t. } \mathbf{w} = \lambda \mathbf{u} + (1-\lambda)\mathbf{v}\}.
\end{aligned}$$

It is worth noting that the condition (1.9) can be equivalently replaced by

$$\{\mathbf{x} \in A \mid \forall \mathbf{y} \in \text{aff}(A)\ \exists r > 0 \text{ s.t. } [\mathbf{x}, \mathbf{x} + r(\mathbf{y} - \mathbf{x})) \subseteq A\},$$

or

$$\{\mathbf{x} \in \text{aff}(A) \mid \forall \mathbf{y} \in A\ \exists \mathbf{z} \in A \text{ s.t. } \mathbf{x} \in (\mathbf{y}, \mathbf{z})\}.$$

The following proposition can be seen as a generalization of Proposition 1.9 (i).

Proposition 1.24. *For any convex set $C \subset E$ we have* $\text{ai}(C)$ *and* $\text{ac}(C)$ *are also convex.*

Proof. See Lemma 5.2 in [1]. □

We will use the following useful result in the sequel chapters.

Proposition 1.25. *Let C be a convex set in E and $\mathbf{x} \in \operatorname{ai}(C)$, $\mathbf{y} \in \operatorname{ac}(C)$. Then $[\mathbf{x}, \mathbf{y}) \subset \operatorname{ai}(C)$.*

Proof. See Lemma 5.5 in [1]. □

The following proposition can be seen as a generalization of Proposition 1.10 (ii).

Proposition 1.26. *For any convex sets $C, D \subset E$ with nonempty relative algebraic interiors we have $\operatorname{rai}(C + D) = \operatorname{rai}(C) + \operatorname{rai}(D)$.*

Proof. See Lemma 5.11 in [1]. □

Chapter 2

Separation between two convex sets

This chapter is devoted to presenting some separation theorems related to two convex sets. We will recall in Section 2.1 some separation concepts, then present the theorems as well as their corollaries in Section 2.2. We first consider the separation concepts and separation theorems in the setting finite dimensional Euclidean vector spaces, and then in the setting of general vector spaces.

2.1 Separation concepts

In this section, we recall some concepts involving separation between two convex sets. The finite dimensional versions of these concepts are presented in Subsection 2.1.1, while their generalizations in the setting of infinite dimensional spaces are given in Subsection 2.1.2.

2.1.1 In $\mathbb{R}^n$

For the sake of simplicity, we will consider them in the setting of $\mathbb{R}^n$ with the usual inner product and its induced norm. However, it is worth noting that the concepts in this subsection are valid for finite dimensional Euclidean vector spaces.

Definition 2.1. (Half-space in $\mathbb{R}^n$, see e.g. [1]). *Let $H := H(\mathbf{a}, \xi)$ be a hyperplane in $\mathbb{R}^n$. The two following closed sets*

$$\bar{H}^+(\mathbf{a}, \xi) = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x} \rangle \geq \xi\}, \quad \bar{H}^-(\mathbf{a}, \xi) = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x} \rangle \leq \xi\}$$

are called the closed half-spaces associated with H, while the two following open sets

$$H^+(\mathbf{a}, \xi) = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x} \rangle > \xi\}, \quad H^-(\mathbf{a}, \xi) = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x} \rangle < \xi\}.$$

are called the open half-spaces associated with H.

Definition 2.2. (Separation concepts in finite dimensional spaces, see e.g. [1]). *Given nonempty convex sets $C, D \subset E$, and let $H = H(\mathbf{a}, \xi)$ be a hyperplane in $\mathbb{R}^n$.*

(i) The sets C and D are said to be separated by the hyperplane H if $C \subseteq \bar{H}^+(\mathbf{a}, \xi)$ and $D \subseteq \bar{H}^-(\mathbf{a}, \xi)$, i.e.,

$$\langle \mathbf{a}, \mathbf{x} \rangle \geq \xi \geq \langle \mathbf{a}, \mathbf{y} \rangle \quad \forall \mathbf{x} \in C, \mathbf{y} \in D.$$

In this case we say that H is a separating hyperplane for C and D.

(ii) The sets C and D are said to be strictly separated by the hyperplane H if $C \subseteq H^+(\mathbf{a}, \xi)$ and $D \subseteq H^-(\mathbf{a}, \xi)$, i.e.,

$$\langle \mathbf{a}, \mathbf{x} \rangle > \xi > \langle \mathbf{a}, \mathbf{y} \rangle \quad \forall \mathbf{x} \in C, \mathbf{y} \in D.$$

In this case we say that H is a strictly separating hyperplane for C and D.

(iii) The sets C and D are said to be strongly separated by the hyperplane H if there exist $\beta > \xi > \gamma$ such that $C \subseteq \bar{H}^+(\mathbf{a}, \beta)$, $D \subseteq \bar{H}^-(\mathbf{a}, \gamma)$, i.e.,

$$\langle \mathbf{a}, \mathbf{x} \rangle \geq \gamma > \xi > \beta \geq \langle \mathbf{a}, \mathbf{y} \rangle \quad \forall \mathbf{x} \in C, \mathbf{y} \in D.$$

In this case we say that H is a strongly separating hyperplane for C and D.

(iv) The sets C and D are said to be properly separated by the hyperplane H if the two following conditions hold:

- *H separates C and D.*
- *C and D are not both included in H.*

In this case we say that H is a proper separating hyperplane for C and D.

In Figure 2.1, the set C is a closed circle (including its boundary) and the set D is a closed square (including its boundary) in $\mathbb{R}^2$. An edge of the square D is included in the hyperplane $H(\mathbf{a}, \xi)$ and it is tangent to the circle C. In this case, C and D are separated by the hyperplane $H(\mathbf{a}, \xi)$. We observe furthermore that in this case C and D cannot be either strictly separated or strongly separated.

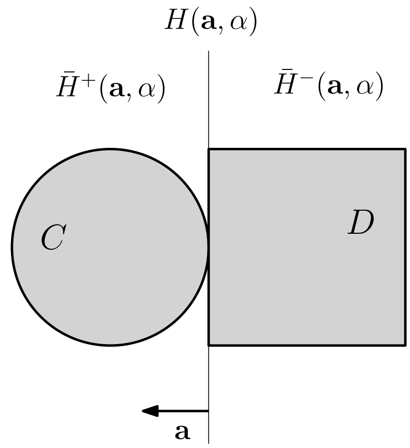

Figure 2.1: Separation of two sets by a hyperplane.

In Figure 2.2, the set C is an open circle (excluding its boundary) and the set D is an open square (excluding its boundary) in $\mathbb{R}^2$. An edge of the boundary of the square D is included in the hyperplane $H(\mathbf{a}, \xi)$ and it is tangent to the boundary of the circle C. In this case, C and D are strictly separated by the hyperplane $H(\mathbf{a}, \xi)$. We observe furthermore that in this case C and D are separated, but cannot be strongly separated.

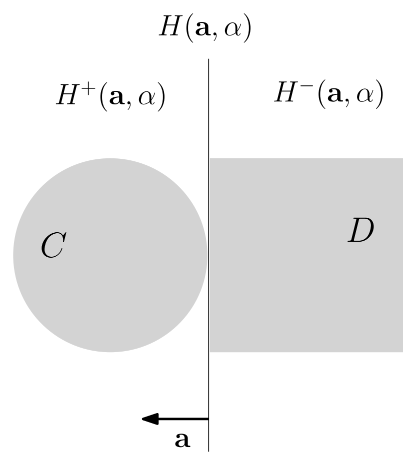

Figure 2.2: Strict separation of two sets by a hyperplane.

In Figure 2.3, the set C is a closed circle (including its boundary), the set D is a closed square (including its boundary) in $\mathbb{R}^2$, and they are disjoint. In this case, $H(\mathbf{a}, \xi)$ strongly separates C and D, and these sets are both separated and strictly separated.

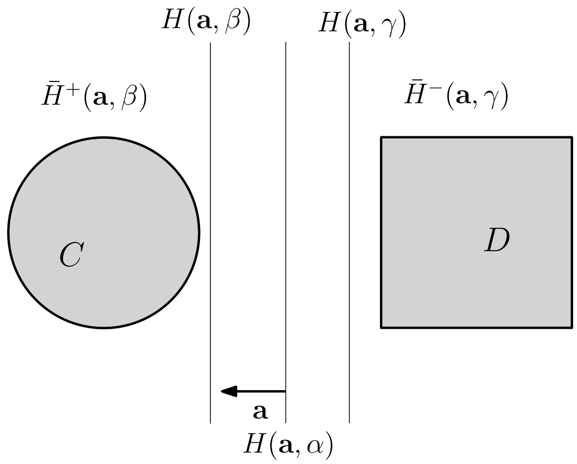

Figure 2.3: Strong separation of two sets by a hyperplane.

In Figure 2.4(i), the sets C and D are two line segments lying on the same hyperplane $H(\mathbf{a}, \xi)$. Since this hyperplane is included in both half-spaces $\bar{H}^{+}(\mathbf{a}, \xi)$ and $\bar{H}^{-}(\mathbf{a}, \xi)$, it follows that $C \subset \bar{H}^{+}(\mathbf{a}, \xi)$ and $D \subset \bar{H}^{-}(\mathbf{a}, \xi)$. It means that $H(\mathbf{a}, \xi)$ separates C and D. However, in this case C and D do not both lie in the hyperplane $H(\mathbf{a}, \xi)$, so they are not properly separated by this hyperplane. Intuitively, we can see that C and D can be still separated by a hyperplane which is orthogonal to $H(\mathbf{a}, \xi)$. Therefore, when saying that two convex sets are properly separated, we must emphasize the separating hyperplane in the proper separation.

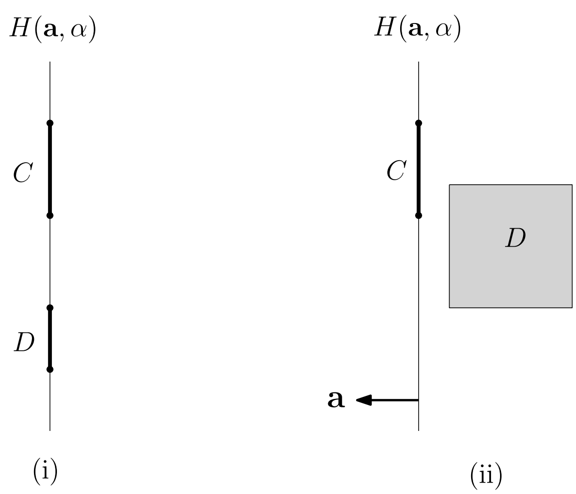

Figure 2.4: (i) Not proper separation. (ii) Proper separation.

In Figure 2.4(ii), the set C is a line segment lying on the hyperplane $H(\mathbf{a}, \xi)$, while the set D is a closed square lying entirely in half-space $\bar{H}^{-}(\mathbf{a}, \xi)$. Since the hyperplane $H(\mathbf{a},\xi)$ is included in the half-space $\bar{H}^+(\mathbf{a},\xi)$, it follows that C is contained in $\bar{H}^+(\mathbf{a},\xi)$. Thus, $H(\mathbf{a},\xi)$ is a separating hyperplane for C and D. It is obvious that D is not contained in that hyperplane. Therefore, in this case, C and D are properly separated by the hyperplane $H(\mathbf{a},\xi)$.

2.1.2 In general vector spaces

Let E be a general vector space.

Definition 2.3. (See e.g. [1]). *A subset $H \subset E$ is called a hyperplane if it is of the form*

$$H = \{\mathbf{x} \in E \mid h(\mathbf{x}) = \xi\}$$

for some $\xi \in \mathbb{R}$ and some nontrivial linear functional $h : E \to \mathbb{R}$.

Roughly speaking, a hyperplane in E is the level set of a nontrivial linear functional. We denote $H := H(h,\xi)$ to indicate the linear functional h and the level ξ defining the hyperplane.

Definition 2.4. (See e.g. [1]). *Given a hyperplane $H := H(h,\xi)$ in E. The two following sets*

$$\bar{H}^+(h,\xi) = \{\mathbf{x} \in E : h(\mathbf{x}) \geq \xi\}, \quad \bar{H}^-(h,\xi) = \{\mathbf{x} \in E : h(\mathbf{x}) \leq \xi\}$$

are called the (algebraically) closed half-spaces associated with H, while the two following sets

$$H^+(h,\xi) = \{\mathbf{x} \in E : h(\mathbf{x}) > \xi\}, \quad H^-(h,\xi) = \{\mathbf{x} \in E : h(\mathbf{x}) < \xi\}$$

are called the (algebraically) open half-spaces associated with H.

It is worth mentioning that the terms 'closed' and 'open' in the above definition do not rely on any topology of the underlying space E. These terms are there to emphasize the similarity of the concepts to the ones in the setting of finite dimensional spaces.

Definition 2.5. (See e.g. [1]). *Given a hyperplane $H = H(h,\xi)$ in E and two nonempty convex sets $C, D \subset E$.*

(i) The sets C and D are said to be separated by the hyperplane H if $C \subseteq \bar{H}^+(h,\xi)$ and $D \subseteq \bar{H}^-(h,\xi)$, i.e.,

$$h(\mathbf{x}) \geq \xi \geq h(\mathbf{y}) \quad \forall \mathbf{x} \in C, \mathbf{y} \in D.$$

In this case we say that H separates C and D.

(ii) The sets C and D are said to be strictly separated by the hyperplane H if $C \subseteq H^+(h, \xi)$ and $D \subseteq H^-(h, \xi)$, i.e.,

$$h(\mathbf{x}) > \xi > h(\mathbf{y}) \quad \forall \mathbf{x} \in C, \mathbf{y} \in D.$$

In this case we say that H strictly separates C and D.

(iii) The sets C and D are said to be strongly separated by the hyperplane H if there exist $\beta > \xi > \gamma$ such that $C \subseteq \bar{H}^+(h, \beta)$, $D \subseteq \bar{H}^-(h, \gamma)$, i.e.,

$$h(\mathbf{x}) \geq \gamma > \xi > \beta \geq h(\mathbf{y}) \quad \forall \mathbf{x} \in C, \mathbf{y} \in D.$$

In this case we say that H strongly separates C and D.

(iv) The sets C and D are said to be properly separated by the hyperplane H if the two following conditions hold:

- *H separates C and D.*
- *C and D are not both contained in H.*

In this case we say that H properly separates C and D.

The following proposition gives us an important property of hyperplanes in general vector spaces.

Proposition 2.6. *Any hyperplane $H \subset E$ is a proper maximal affine subset of E.*

Proof. See Lemma 6.27 in [1]. □

2.2 Separation theorems

This section presents some results concerning the separation, strong separation, proper separation between two convex sets. The results in the setting of finite dimensional spaces are discussed in Subsection 2.2.1, while the ones in the setting of infinite dimensional spaces are given in Subsection 2.2.2.

2.2.1 In $\mathbb{R}^n$

Although the results in this subsection hold for finite dimensional Euclidean vector spaces, for the sake of simplicity we will consider them in the setting of $\mathbb{R}^n$ with the usual inner product and its induced norm.

It is worth noting the equivalence of the two following facts:

- Two given convex sets $C, D \in \mathbb{R}^n$ are separable from each other.

- The point $\mathbf{0}$ can be separated from the convex set $C - D$.

Therefore, in the following, we first discuss about separation of a single point from a closed convex set, and then draw the results concerning separation between convex sets.

Theorem 2.7. (See e.g. [1]). *If $C \subset \mathbb{R}^n$ is a nonempty convex set and $\bar{\mathbf{x}} \notin \text{relint}(C)$, then $\bar{\mathbf{x}}$ can be separated from $\overline{C}$. That is, there exists a hyperplane $H(\mathbf{a}, \xi)$ containing $\bar{\mathbf{x}}$ such that $\overline{C} \subset \bar{H}^+(\mathbf{a}, \xi)$, or equivalently, $\langle \mathbf{a}, \mathbf{x} \rangle \geq \langle \mathbf{a}, \bar{\mathbf{x}} \rangle = \xi$ for all $\mathbf{x} \in \overline{C}$.*

Proof. Since C is convex, by Proposition 1.9(i), its closure $\overline{C}$ is also convex. Since $\bar{\mathbf{x}} \notin \text{relint}(C)$, either $\bar{\mathbf{x}} \notin \overline{C}$ or $\bar{\mathbf{x}} \in \overline{C} \backslash \text{relint}(C)$.

We first consider the former case in which $\bar{\mathbf{x}} \notin \overline{C}$. Proposition 1.16 gives us the inequality

$$\langle \bar{\mathbf{x}} - \text{proj}_{\overline{C}}(\bar{\mathbf{x}}), \mathbf{x} - \text{proj}_{\overline{C}}(\bar{\mathbf{x}}) \rangle \leq 0 \quad \forall \mathbf{x} \in \overline{C}. \tag{2.1}$$

Let $\mathbf{a} = \text{proj}_{\overline{C}}(\bar{\mathbf{x}}) - \bar{\mathbf{x}}$. Clearly, $\mathbf{a} \neq \mathbf{0}$ since $\bar{\mathbf{x}} \notin \overline{C}$ by our assumption. By rewriting

$$\mathbf{x} - \text{proj}_{\overline{C}}(\bar{\mathbf{x}}) = \mathbf{x} - \bar{\mathbf{x}} - (\text{proj}_{\overline{C}}(\bar{\mathbf{x}}) - \bar{\mathbf{x}}) = \mathbf{x} - \bar{\mathbf{x}} - \mathbf{a},$$

from inequality (2.1) we derive $\langle \mathbf{a}, \mathbf{x} - \bar{\mathbf{x}} - \mathbf{a} \rangle \geq 0$. Then $\langle \mathbf{a}, \mathbf{x} - \bar{\mathbf{x}} \rangle \geq \langle \mathbf{a}, \mathbf{a} \rangle = \|\mathbf{a}\|^2 > 0$, and we have

$$\langle \mathbf{a}, \mathbf{x} \rangle > \langle \mathbf{a}, \bar{\mathbf{x}} \rangle \quad \forall \mathbf{x} \in \overline{C}.$$

Let $\xi = \langle \mathbf{a}, \bar{\mathbf{x}} \rangle$, then the theorem is proved in this case.

In the latter case $\bar{\mathbf{x}} \in \overline{C} \backslash \text{relint}(C)$, by Lemma 1.12 there exists a sequence $\{\mathbf{x}^k \mid k \in \mathbb{N}\}$ of points not in $\overline{C}$ such that $\mathbf{x}^k \to \bar{\mathbf{x}}$. By (2.1), we obtain

$$\langle \mathbf{x}^k - \text{proj}_{\overline{C}}(\mathbf{x}^k), \mathbf{x} - \text{proj}_{\overline{C}}(\mathbf{x}^k) \rangle \leq 0 \quad \forall \mathbf{x} \in \overline{C}.$$

Note that $\mathbf{x}^k \notin \overline{C}, \text{proj}_{\overline{C}}(x^k) \in \overline{C}$, so $\mathbf{x}^k \neq \text{proj}_{\overline{C}}(\mathbf{x}^k)$. Hence we can define

$$\mathbf{a}^k := \frac{1}{\|\text{proj}_{\overline{C}}(\mathbf{x}^k) - \mathbf{x}^k\|} \left(\text{proj}_{\overline{C}}(\mathbf{x}^k) - \mathbf{x}^k\right) \neq \mathbf{0},$$

and get

$$\langle \mathbf{a}^k, \mathbf{x} - \text{proj}_{\overline{C}}(\mathbf{x}^k) \rangle \geq 0 \quad \forall \mathbf{x} \in \overline{C}. \tag{2.2}$$

Since $\|\mathbf{a}^k\| = 1$, the sequence $\{\mathbf{a}^k \mid k \in \mathbb{N}\}$ is bounded in $\mathbb{R}^n$, so it has a convergent subsequence $\{\mathbf{a}^{k_i}\}$, *i.e.*, $\mathbf{a}^{k_i} \to \mathbf{a}$. Then $\|\mathbf{a}\| = 1$, hence $\mathbf{a} \neq \mathbf{0}$. From the continuity of $\text{proj}_{\overline{C}}$ and the fact that $\mathbf{x}^k \to \bar{\mathbf{x}}$, we derive $\text{proj}_{\overline{C}}(\mathbf{x}^k) \to \text{proj}_{\overline{C}}(\bar{\mathbf{x}}) = \bar{\mathbf{x}}$. Observe (2.6) and let $k_i \to \infty$, we obtain

$$\langle \mathbf{a}, \mathbf{x} \rangle \geq \langle \mathbf{a}, \bar{\mathbf{x}} \rangle \quad \forall \mathbf{x} \in C.$$

By letting $\xi = \langle \mathbf{a}, \bar{\mathbf{x}} \rangle$, the theorem is proved. $\square$

The previous theorem implies the existence of a so-called support hyperplane to a convex set, which is defined as follows.

Definition 2.8. (Support hyperplane, see e.g. [1]). *A hyperplane $H(\mathbf{a}, \xi)$ is called a support hyperplane of a convex set $C \subset \mathbb{R}^n$ at a point $\mathbf{x} \in \overline{C}$ if $\mathbf{x} \in H(\mathbf{a}, \xi)$ and $C \subset \bar{H}^+(\mathbf{a}, \xi)$, i.e., $\langle \mathbf{a}, \mathbf{y} \rangle \geq \langle \mathbf{a}, \mathbf{x} \rangle = \xi$ for all $\mathbf{y} \in \overline{C}$.*

Since $\bar{H}^+(\mathbf{a}, \xi)$ is closed, the condition $C \subset \bar{H}^+(\mathbf{a}, \xi)$ in the above definition is equivalent to $\overline{C} \subset \bar{H}^+(\mathbf{a}, \xi)$. With this definition, the above theorem can be restated as follows.

Theorem 2.9. (Support hyperplane theorem, see e.g. [1]). *For any point $\mathbf{x} \in \overline{C} \backslash \mathrm{relint}(C)$ in which $C \subseteq \mathbb{R}^n$ is a nonempty convex set, there exists a support hyperplane to C at $\mathbf{x}$.*

Now we discuss the results concerning separation between convex sets.

Theorem 2.10. (First separation theorem, see e.g. [1]). *Any nonempty disjoint convex sets $C, D \subset \mathbb{R}^n$ can be separated by a hyperplane $H(\mathbf{a}, \xi)$ in the sense that*

$$\langle \mathbf{a}, \mathbf{x} \rangle \geq \xi \geq \langle \mathbf{a}, \mathbf{y} \rangle \quad \forall \mathbf{x} \in C, \mathbf{y} \in D.$$

Proof. Let $A := C - D = \{\mathbf{x} - \mathbf{y} \mid \mathbf{x} \in C, \mathbf{y} \in D\}$. By Proposition 1.9(ii), A is convex. Since C and D are disjoint, we have $\mathbf{0} \notin A$, and hence $\mathbf{0} \notin \mathrm{relint}(A)$. By Theorem 2.7, there exists a hyperplane $H(\mathbf{a}, 0)$ containing $\mathbf{0}$ such that $\langle \mathbf{a}, \mathbf{s} \rangle \geq \langle \mathbf{a}, \mathbf{0} \rangle = 0$ for all $\mathbf{s} \in \overline{A}$. In particular, $\langle \mathbf{a}, \mathbf{s} \rangle \geq 0$ for all $\mathbf{s} \in A \subset \overline{A}$. Since $A = C - D$, it follows that $\langle \mathbf{a}, \mathbf{x} - \mathbf{y} \rangle \geq 0$, or equivalently, $\langle \mathbf{a}, \mathbf{x} \rangle \geq \langle \mathbf{a}, \mathbf{y} \rangle$ for all $\mathbf{x} \in C, \mathbf{y} \in D$. By choosing ξ such that

$$\inf_{\mathbf{x} \in C} \langle \mathbf{a}, \mathbf{x} \rangle \geq \xi \geq \sup_{\mathbf{y} \in D} \langle \mathbf{a}, \mathbf{y} \rangle,$$

the hyperplane $H(\mathbf{a}, \xi)$ separates C and D. $\square$

Theorem 2.11. (Strong separation theorem, see e.g. [1]). *Any nonempty disjoint closed convex sets $C, D \subset \mathbb{R}^n$ can be strongly separated if one of the sets is compact.*

Proof. Without loss of generality, let us assume that D is compact. By the definition of strong separation between two convex sets, we observe that the theorem is equivalent to the existence of a hyperplane $H(\mathbf{a}, \xi)$ satisfying the condition

$$\inf_{\mathbf{x} \in C} \langle \mathbf{a}, \mathbf{x} \rangle > \xi > \max_{\mathbf{y} \in D} \langle \mathbf{a}, \mathbf{y} \rangle. \tag{2.3}$$

Let $A := C - D$, then A is closed. Indeed, let $\{\mathbf{u}^k \mid k \in \mathbb{N}\}$ be a sequence of points in A that converges to some $\mathbf{u} \in \mathbb{R}^n$. For every $k \in \mathbb{N}$, we represent $\mathbf{u}^k = \mathbf{x}^k - \mathbf{y}^k$ where $\mathbf{x}^k \in C, \mathbf{y}^k \in D$. Since D is compact, we can extract a convergent subsequence $\{y^{k_i}\}$ of $\{\mathbf{y}^k\}$. Assume that $y^{k_i} \to \mathbf{y} \in D$. Since $\mathbf{x}^{k_i} - \mathbf{y}^{k_i} \to \mathbf{u}$ and $\mathbf{y}^{k_i} \to \mathbf{y}$, it follows that $\mathbf{x}^{k_i} \to \mathbf{x} := \mathbf{u} + \mathbf{y}$. Note that C is closed, hence $\mathbf{x} \in C$. Therefore $\mathbf{u} = \mathbf{x} - \mathbf{y} \in C - D = A$. This implies that A is a closed set.

Since C and D are disjoint, we have $\mathbf{0} \notin A = \overline{A}$. By Proposition 1.9(ii), A is convex. By similar arguments as in the first part of the proof of Theorem 2.7, we must have a nonzero vector $\mathbf{a}$ satisfying

$$\langle \mathbf{a}, \mathbf{u} \rangle \geq \langle \mathbf{a}, \mathbf{a} \rangle = \|\mathbf{a}\|^2 > 0 \quad \forall \mathbf{u} \in A = \overline{A},$$

or equivalently,

$$\langle \mathbf{a}, \mathbf{x} - \mathbf{y} \rangle \geq \|\mathbf{a}\|^2 > 0 \quad \forall \mathbf{x} \in C, \mathbf{y} \in D.$$

Hence we have

$$\langle \mathbf{a}, \mathbf{x} \rangle > \langle \mathbf{a}, \mathbf{x} \rangle - \frac{1}{2}\|\mathbf{a}\|^2 \geq \langle \mathbf{a}, \mathbf{y} \rangle + \frac{1}{2}\|\mathbf{a}\|^2 > \langle \mathbf{a}, \mathbf{y} \rangle$$

for all $\mathbf{x} \in C$ and $\mathbf{y} \in D$. Then, by choosing $\xi = \max_{\mathbf{y} \in D} \langle \mathbf{a}, \mathbf{y} \rangle + \frac{1}{2}\|\mathbf{a}\|^2$, we obtain (2.3) and the theorem is proved. □

As a remark, the compactness condition in Theorem 2.11 cannot be omitted. For example, in $\mathbb{R}^2$ let us consider the two convex sets

$$C = \{(x, y) \mid y \leq 0\}, \quad D = \left\{(x, y) \mid x > 0, y \geq \frac{1}{x}\right\}.$$

Figure 2.5 illustrates the two sets in this counter-example. Neither C or D is compact. We can observe that the only hyperplane separating C and D is the x-axis. Since this axis coincides the boundary of C, these sets cannot be strongly separated.

An important corollary of Theorem 2.11 concerns representation of convex sets as follows.

Corollary 2.12. (See e.g. [1]). *Any nonempty closed convex set in* $\mathbb{R}^n$ *coincides with the intersection of all closed half-spaces containing it.*

Proof. Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set and define D as

$$D = \bigcap_{(\mathbf{a}, \xi)} \{\bar{H}^+(\mathbf{a}, \xi) \mid C \subseteq \bar{H}^+(\mathbf{a}, \xi)\}.$$

We need to show that $C = D$. Indeed, since C is contained in each half-space forming D, it is also contained in the intersection of the half-spaces. Therefore $C \subseteq D$. It remains to show that $D \subseteq C$.

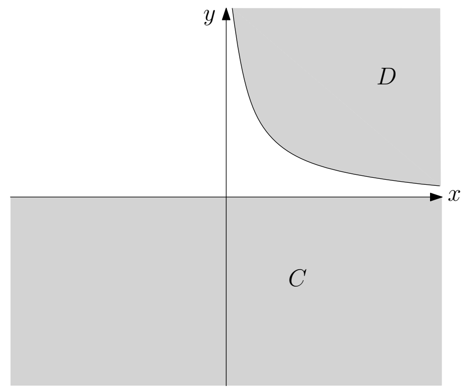

Figure 2.5: An example of two convex sets that cannot be strongly separated.

Since D is the intersection of closed sets, it is also closed. Moreover, it follows from Proposition 1.9(ii) that D is convex, since it is the intersection of half-spaces that are also convex. Assume the contrary that D is not a subset of C. Then there exists a point $\mathbf{x}^0 \in D \backslash C$. Applying Theorem 2.11 to the compact convex set $\{\mathbf{x}^0\}$ and the closed convex set C, there exists a hyperplane $H := H(\mathbf{a}, \xi)$ such that $\mathbf{x}^0 \in H^-(\mathbf{a}, \xi)$ and $C \subseteq H^+(\mathbf{a}, \xi) \subset \bar{H}^+(\mathbf{a}, \xi)$. By construction of D, the half-space $\bar{H}^+(\mathbf{a}, \xi)$ is one of the closed half-spaces intersected to obtain D, so $D \subseteq \bar{H}^+(\mathbf{a}, \xi)$. Since $\mathbf{x}^0 \in D$, we have $\mathbf{x}^0 \in \bar{H}^+(\mathbf{a}, \xi)$, but this contradicts $\mathbf{x}^0 \in H^-(\mathbf{a}, \xi)$. This contradiction proves the corollary. □

We now come to the results concerning proper separation between convex sets. The following lemma is useful in the proof of the results.

Lemma 2.13. (See e.g. [1]). *Two nonempty convex sets $C, D \in \mathbb{R}^n$ can be properly separated if and only if $\mathbf{0}$ is properly separated from $K := C - D$.*

Proof. Sufficiency. Recall that the set $K = C - D$ is convex thanks to Proposition 1.9(ii). Let $H(\mathbf{a}, \xi)$ be a hyperplane properly separating C and D such that $C \subseteq \bar{H}^+(\mathbf{a}, \xi), D \subseteq \bar{H}^-(\mathbf{a}, \xi)$. Without loss of generality, assume that C does not lie on $H(\mathbf{a}, \xi)$. Then we have

$$\langle \mathbf{a}, \mathbf{x} \rangle \geq \xi \geq \langle \mathbf{a}, \mathbf{y} \rangle \quad \forall \mathbf{x} \in C, \mathbf{y} \in D,$$

and $\langle \mathbf{a}, \mathbf{x}^0 \rangle > \xi$ for some $\mathbf{x}^0 \in C$. This means that

$$\langle \mathbf{a}, \mathbf{x} - \mathbf{y} \rangle \geq 0 \quad \forall \mathbf{x} \in C, \mathbf{y} \in D,$$

or equivalently

$$\langle \mathbf{a}, \mathbf{z} \rangle \geq 0 \quad \forall \mathbf{z} \in K = C - D,$$

and furthermore we have $\langle \mathbf{a}, \mathbf{z}^0 \rangle > 0$ for $\mathbf{z}^0 = \mathbf{x}^0 - \mathbf{y}^0$ for some $\mathbf{y}^0 \in D$. This implies that the hyperplane $H(\mathbf{a}, 0)$ properly separates the origin $\mathbf{0}$ and the convex set K.

Necessity. Suppose that there exists a hyperplane $H(\mathbf{a}, \xi)$ properly separating the origin $\mathbf{0}$ with the convex set $K = C - D$, and that $K \subseteq \bar{H}^+(\mathbf{a}, \xi)$. Then $\langle \mathbf{a}, \mathbf{x} - \mathbf{y} \rangle \geq \xi \geq 0$ for all $\mathbf{x} \in C, \mathbf{y} \in D$. The proper separation means that

- either K is included in the hyperplane $H(\mathbf{a}, \xi)$ while the origin $\mathbf{0}$ is not,
- or K is not included in the hyperplane $H(\mathbf{a}, \xi)$ (but it is still contained in the half-space $\bar{H}^+(\mathbf{a}, \xi)$).

In the former case, since $\mathbf{0} \notin H(\mathbf{a}, \xi)$ we have $\xi > 0$, and since $K \subset H(\mathbf{a}, \xi)$ we have $\langle \mathbf{a}, \mathbf{x} - \mathbf{y} \rangle = \xi$ for all $\mathbf{x} \in C, \mathbf{y} \in D$. In this case we obtain for any $\mathbf{x} \in C$ and any $\mathbf{y} \in D$ that

$$\langle \mathbf{a}, \mathbf{x} \rangle = \xi + \langle \mathbf{a}, \mathbf{y} \rangle > \frac{\xi}{2} + \langle \mathbf{a}, \mathbf{y} \rangle > \langle \mathbf{a}, \mathbf{y} \rangle,$$

so the hyperplane $H(\mathbf{a}, \beta)$ with $\beta = \frac{\xi}{2} + \langle \mathbf{a}, \mathbf{y} \rangle$ properly separates C and D.

In the latter case, since K is not included in the hyperplane $H(\mathbf{a}, \xi)$, there exists $\mathbf{z}^0 \in K$ such that $\langle \mathbf{a}, \mathbf{z}^0 \rangle > \xi$. Since $\mathbf{z}^0 \in K = C - D$, there exist $\mathbf{x}^0 \in C$ and $\mathbf{y}^0 \in D$ such that $\mathbf{z}^0 = \mathbf{x}^0 - \mathbf{y}^0$. So we have

$$\langle \mathbf{a}, \mathbf{z}^0 \rangle = \langle \mathbf{a}, \mathbf{x}^0 - \mathbf{y}^0 \rangle > \xi,$$

or equivalently,

$$\langle \mathbf{a}, \mathbf{x}^0 \rangle > \xi + \langle \mathbf{a}, \mathbf{y}^0 \rangle. \tag{2.4}$$

From the fact that $\langle \mathbf{a}, \mathbf{x} - \mathbf{y} \rangle \geq \xi$ for all $\mathbf{x} \in C, \mathbf{y} \in D$, we have

$$\langle \mathbf{a}, \mathbf{x} \rangle \geq \xi + \langle \mathbf{a}, \mathbf{y} \rangle \quad \forall \mathbf{x} \in C, \mathbf{y} \in D.$$

So we obtain

$$\inf_{\mathbf{x} \in C} \langle \mathbf{a}, \mathbf{x} \rangle \geq \xi + \sup_{\mathbf{y} \in D} \langle \mathbf{a}, \mathbf{y} \rangle.$$

This, together with (2.4), means that any hyperplane $H(\mathbf{a}, \gamma)$ with

$$\inf_{\mathbf{x} \in C} \langle \mathbf{a}, \mathbf{x} \rangle \geq \gamma \geq \xi + \sup_{\mathbf{y} \in D} \langle \mathbf{a}, \mathbf{y} \rangle$$

properly separates C and D. □

In relation with Lemma 2.13 we have the following result.

Lemma 2.14. (See e.g. [1]). *Let C be a nonempty convex set in $\mathbb{R}^n$. Then the origin $\mathbf{0}$ and the set C can be properly separated if and only if $\mathbf{0} \notin \operatorname{relint}(C)$.*

Proof. *Necessity.* Since $\mathbf{0} \notin \operatorname{relint}(C)$, either $\mathbf{0} \notin \overline{C}$ or $\mathbf{0} \in \overline{C} \setminus \operatorname{relint}(C)$. We first consider the former case in which $\mathbf{0} \notin \overline{C}$. By Proposition 1.9(i), since C is convex, so is its clossure $\overline{C}$. Proposition 1.16 gives us the inequality

$$\langle \mathbf{0} - \operatorname{proj}_{\overline{C}}(\mathbf{0}), \mathbf{x} - \operatorname{proj}_{\overline{C}}(\mathbf{0}) \rangle \leq 0 \quad \forall \mathbf{x} \in \overline{C},$$

or equivalently

$$\langle \mathbf{a}, \mathbf{x} - \mathbf{a} \rangle \geq 0 \quad \forall \mathbf{x} \in \overline{C}, \tag{2.5}$$

in which $\mathbf{a} = \operatorname{proj}_{\overline{C}}(\mathbf{0})$. Clearly, $\mathbf{a} \neq \mathbf{0}$ since $\mathbf{0} \notin \overline{C}$ by our assumption. From the inequality (2.5), we derive $\langle \mathbf{a}, \mathbf{x} \rangle \geq \langle \mathbf{a}, \mathbf{a} \rangle = \|\mathbf{a}\|^2 > 0$ for all $\mathbf{x} \in \overline{C}$. This implies that the set C and $\{\mathbf{0}\}$ are properly separated.

We now consider the latter case in which $\mathbf{0} \in \overline{C} \setminus \operatorname{relint}(C)$. By Lemma 1.12, there exists a sequence $\{\mathbf{x}^k \mid k \in \mathbb{N}\} \subset \operatorname{aff}(C)$ with $\mathbf{x}^k \notin \overline{C}$ and $\mathbf{x}^k \to \mathbf{0}$ as $k \to \infty$. By Proposition 1.16, we obtain

$$\langle \mathbf{x}^k - \operatorname{proj}_{\overline{C}}(\mathbf{x}^k), \mathbf{x} - \operatorname{proj}_{\overline{C}}(\mathbf{x}^k) \rangle \leq 0 \quad \forall \mathbf{x} \in \overline{C}.$$

Note that $\mathbf{x}^k \notin \overline{C}, \operatorname{proj}_{\overline{C}}(\mathbf{x}^k) \in \overline{C}$, so $\mathbf{x}^k \neq \operatorname{proj}_{\overline{C}}(\mathbf{x}^k)$ and hence we have

$$\mathbf{a}^k := \frac{1}{\|\operatorname{proj}_{\overline{C}}(\mathbf{x}^k) - \mathbf{x}^k\|} \left(\operatorname{proj}_{\overline{C}}(\mathbf{x}^k) - \mathbf{x}^k \right) \neq \mathbf{0},$$

and furthermore we obtain

$$\langle \mathbf{a}^k, \mathbf{x} - \operatorname{proj}_{\overline{C}}(\mathbf{x}^k) \rangle \geq 0 \quad \forall \mathbf{x} \in \overline{C}. \tag{2.6}$$

We observe that $\mathbf{0} \in \overline{C} \subset \operatorname{aff}(C)$, so $\operatorname{aff}(C)$ is a linear subspace of $\mathbb{R}^n$. Moreover, both $\mathbf{x}^k$ and $\operatorname{proj}_{\overline{C}}(\mathbf{x}^k)$ are in $\operatorname{aff}(C)$, thus $\mathbf{a}^k \in \operatorname{aff}(C)$ for all $k \in \mathbb{N}$. Since $\|\mathbf{a}^k\| = 1$, the sequence $\{\mathbf{a}^k \mid k \in \mathbb{N}\}$ is bounded in $\operatorname{aff}(C) \subset \mathbb{R}^n$, so it has a convergent subsequence $\{\mathbf{a}^{k_i}\} \subset \operatorname{aff}(C)$, *i.e.* $\mathbf{a}^{k_i}$ converges to some $\mathbf{a} \in \mathbb{R}^n$. Since $\operatorname{aff}(C)$ is a closed set, $\mathbf{a} \in \operatorname{aff}(C)$. Moreover, note that $\|\mathbf{a}\| = 1$, hence $\mathbf{a} \neq \mathbf{0}$. From the continuity of $\operatorname{proj}_{\overline{C}}$ (see Proposition 1.17) and the fact that $\mathbf{x}^k \to \mathbf{0}$, we derive $\operatorname{proj}_{\overline{C}}(\mathbf{x}^k) \to \operatorname{proj}_{\overline{C}}(\mathbf{0}) = \mathbf{0}$. Keeping (2.6) in mind and let $i \to \infty$ we obtain

$$\langle \mathbf{a}, \mathbf{x} \rangle \geq \langle \mathbf{a}, \mathbf{0} \rangle = 0 \quad \forall \mathbf{x} \in \overline{C}.$$

Assume that $\langle \mathbf{a}, \mathbf{x} \rangle = 0$ for all $\mathbf{x} \in C$. Since $\mathbf{a} \in \operatorname{aff}(C)$, by Proposition 1.4, $\mathbf{a}$ can be represented as an affine combination of some vectors $\mathbf{v}^1, \ldots, \mathbf{v}^m \in C$, that is

$$\mathbf{a} = \sum_{i=1}^{m} \lambda_i \mathbf{v}^i$$

in which $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ and $\lambda_1 + \ldots + \lambda_m = 1$. By our assumption that $\langle \mathbf{a}, \mathbf{x} \rangle = 0$ for all $\mathbf{x} \in C$, taking $\mathbf{x}$ as $\mathbf{v}^1, \ldots, \mathbf{v}^m$ we have $\langle \mathbf{a}, \mathbf{v}^i \rangle = 0$ for all $i = 1, \ldots, m$. Hence we obtain

$$\|\mathbf{a}\|^2 = \langle \mathbf{a}, \mathbf{a} \rangle = \sum_{i=1}^{m} \lambda_i \langle \mathbf{a}, \mathbf{v}^i \rangle = 0,$$

which contradicts the fact that $\|\mathbf{a}\| = 1$. Since the assumption is false, there exists $\mathbf{x}^0 \in C$ such that $\langle \mathbf{a}, \mathbf{x}^0 \rangle > 0$. This shows the proper separation of the sets $\{\mathbf{0}\}$ and C.

Sufficiency. Assume that $\{\mathbf{0}\}$ and C are properly separated. Then there exists $\mathbf{a} \in \mathbb{R}^n$ such that $\langle \mathbf{a}, \mathbf{x} \rangle \geq 0$ for all $\mathbf{x} \in C$ and $\langle \mathbf{a}, \mathbf{x}^0 \rangle > 0$ for some $\mathbf{x}^0 \in C$. If on the contrary $\mathbf{0} \in \mathrm{relint}(C)$, by Proposition 1.11, there exists $t > 0$ such that

$$\mathbf{0} + t(\mathbf{0} - \mathbf{x}^0) = -t\mathbf{x}^0 \in C.$$

Then $\langle \mathbf{a}, -t\mathbf{x}^0 \rangle \geq 0$, or equivalently $\langle \mathbf{a}, \mathbf{x}^0 \rangle \leq 0$, which is a contradiction. Hence $\mathbf{0} \notin \mathrm{relint}(C)$. □

We come up with the following theorem on proper separation between convex sets.

Theorem 2.15. (Proper separation theorem, see e.g. [1]). *We can properly separate two nonempty convex sets $C, D \subset \mathbb{R}^n$ if and only if their relative interiors are disjoint.*

Proof. Since $\mathrm{relint}(C)$ and $\mathrm{relint}(D)$ are disjoint, we have $\mathbf{0} \notin \mathrm{relint}(C) - \mathrm{relint}(D)$. By Proposition 1.10, we have $\mathrm{relint}(C) - \mathrm{relint}(D) = \mathrm{relint}(C - D)$. Thus $\mathbf{0} \notin \mathrm{relint}(C - D)$. Note that C and D are convex, so is $C - D$ (cf. Proposition 1.9(ii)). Hence, by Lemma 2.14, the origin $\mathbf{0}$ and the convex set $C - D$ can be properly separated. It then follows from Lemma 2.13 that C and D can be properly separated. □

2.2.2 In general vector spaces

Throughout this subsection, E is a general vector space without any equipped topology. We start with the following concept.

Definition 2.16. (See e.g. [1]). *Two nonempty convex sets $C, D \subset E$ are called complementary convex sets if they are disjoint and $C \cup D = E$.*

We say that complementary convex sets $C, D \subset E$ separate two given nonempty convex sets $A, B \subset E$ if A is contained in one of the complementary convex sets while B is included in the other, i.e., either $A \subset C, B \subset D$ or $A \subset D, B \subset C$. In this case we also say that A and B are complementarily convex separated (by C and D).

Since complementary convex sets are disjoint, if they separate two given nonempty convex sets $A, B \subset E$, then A and B are also disjoint. The following lemma states that the reverse direction also holds. This is a nontrivial result in order to come up with the sequel theorems in this subsection.

Lemma 2.17. (See e.g. [1]). *If two nonempty convex sets $A, B \subset E$ are disjoint, then they are complementarily convex separated.*

Proof. Let $\mathcal{G}$ be the set of disjoint convex subsets $(C, D) \subset E \times E$ such that $A \subset C$ and $B \subset D$. We introduce a relation $\preceq$ on $\mathcal{G}$ by defining $(C, D) \preceq (C', D')$ if $C \subset C'$ and $D \subset D'$. Since the set inclusion $\subset$ is a partial relation on E, so is $\preceq$ on $\mathcal{G}$. Furthermore, if $\mathcal{F}$ is a totally ordered subset of $\mathcal{G}$, then by taking the union of all sets in $\mathcal{F}$ we obtain an upper bound for elements in $\mathcal{F}$. This property follows from the similar one of the set inclusion relation. It is worth noting that, due to the nested structure of elements in $\mathcal{F}$, the upper bound is a pair of disjoint convex sets in E. By the well-known Zorn's lemma, we obtain a maximal element $(C^*, D^*) \in \mathcal{G}$. It means that

- C^* and D^* are convex and disjoint,
- $C^* \supset$ A, $D^* \supset B$,
- if C and D are convex sets satisfying $C \supset C^*$ and $D \supset D^*$, then we have $C = C^*$ and $D = D^*$.

It is left to prove that $C^* \cup D^* = E$. Indeed, assume the contrary that there exists $\mathbf{x} \in E \backslash (C^* \cup D^*)$. By the maximality of (C^*, D^*), we have

$$\operatorname{conv}(C^* \cup \{\mathbf{x}\}) \cap D^* \neq \varnothing \quad \text{and} \quad \operatorname{conv}(D^* \cup \{\mathbf{x}\}) \cap C^* \neq \varnothing.$$

Therefore we can pick

$$\mathbf{y}^1 \in \operatorname{conv}(C^* \cup \{\mathbf{x}\}) \cap D^* \quad \text{and} \quad \mathbf{y}^2 \in \operatorname{conv}(D^* \cup \{\mathbf{x}\}) \cap C^*.$$

By that choice of $\mathbf{y}^1$, there exists $\mathbf{x}^1 \in C^*$ such that $\mathbf{y}^1 \in (\mathbf{x}, \mathbf{x}^1)$. Similarly, by the choice of $\mathbf{y}^2$, there exists $\mathbf{x}^2 \in D^*$ such that $\mathbf{y}^2 \in (\mathbf{x}, \mathbf{x}^2)$. Let $\mathbf{z}$ be the intersection of the line segments $[\mathbf{x}^1, \mathbf{y}^2]$ and $[\mathbf{x}^2, \mathbf{y}^1]$ (as illustrated in Figure 2.6). Note that $\mathbf{x}^1 \in C^*$ and $\mathbf{y}^2 \in C^*$, by convexity of C^* we have $\mathbf{z} \in C^*$. Similarly, since $\mathbf{z}^2 \in D^*$ and $\mathbf{y}^1 \in D^*$, by convexity of D^* we have $\mathbf{z} \in D^*$. Therefore, $\mathbf{z} \in C^* \cap D^*$, so C^* and D^* are not disjoint. This contradicts the construction of these sets. This contradiction means that $C^* \cup D^* = E$ as desired. $\square$

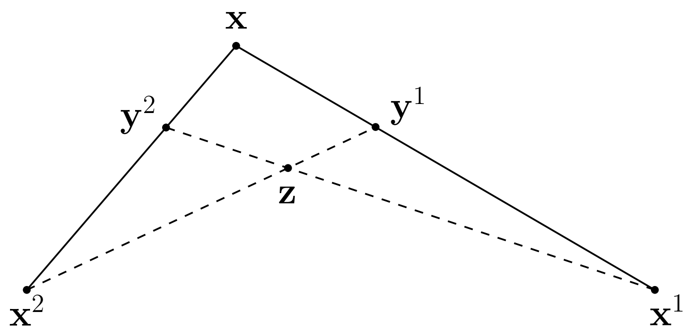

Figure 2.6: Illustration for the proof of Lemma 2.17.

The following lemma gives us a closer look at structure of complementary convex sets. Note that it also holds in the setting of finite dimensional spaces, which has obvious geometric intuition.

Lemma 2.18. (See e.g. [1]). *Let C and D be complementary convex sets in E. Let $L := \mathrm{ac}(C) \cap \mathrm{ac}(D)$. Then either $L = E$ or L is a hyperplane in E. The former case holds if and only if the algebraic interiors of C and D are both empty, or equivalently,* $\mathrm{ac}(C) = \mathrm{ac}(D) = E$. *If the latter case holds, then the following also holds:*

(i) the algebraic interiors of C and D are both nonempty,
(ii) $\mathrm{ai}(C)$, $\mathrm{ai}(D)$ *are the algebraically open half-spaces associated with L,*
(iii) $\mathrm{ac}(C)$, $\mathrm{ac}(D)$ *are the algebraically closed half-spaces associated with L.*

Proof. By Proposition 1.24, since C and D are convex, so are $\mathrm{ac}(C)$ and $\mathrm{ac}(D)$. Thus, as intersection of two convex sets, L is convex. Furthermore, L is nonempty. Indeed, since both C and D are nonempty, we can choose $\mathbf{x} \in C$ and $\mathbf{y} \in D$. Since C and D are disjoint, there exists $\mathbf{z} \in (\mathbf{x}, \mathbf{y})$ such that $[\mathbf{x}, \mathbf{z}) \subset C$ and $(\mathbf{z}, \mathbf{y}] \subset D$. By definition of algebraic closure, we have $\mathbf{z} \in \mathrm{ac}(C)$ and $\mathbf{z} \in \mathrm{ac}(D)$. Hence $\mathbf{z} \in L$, which implies $L \neq \varnothing$.

We now show that

$$\mathrm{ac}(C) = E \backslash \mathrm{ai}(D). \tag{2.7}$$

Indeed, pick any $\mathbf{x} \in E \backslash \mathrm{ai}(D)$. Following the definition of algebraic interior, there exists $\mathbf{u} \in E$ such that for all $r > 0$ we have $[\mathbf{x}, \mathbf{x} + r(\mathbf{u} - \mathbf{x})) \not\subset D$. By letting $\mathbf{v} = \mathbf{x} + r(\mathbf{u} - \mathbf{x})$, this is equivalent to say that for all $\mathbf{v} \in E$ with $\mathbf{x} \in (\mathbf{u}, \mathbf{v})$ we have $[\mathbf{x}, \mathbf{v}) \subset E \backslash D = C$. Thus, $\mathbf{x} \in \mathrm{ac}(C)$. Since $\mathbf{x}$ is chosen arbitrarily in $E \backslash \mathrm{ai}(D)$, we obtain $E \backslash \mathrm{ai}(D) \subset \mathrm{ac}(C)$. Conversely, pick any $\mathbf{y} \in \mathrm{ac}(C)$. Then, following the definition of algebraic closure, there exists $\mathbf{z} \in C$ such that $[\mathbf{z}, \mathbf{y}) \subset C$. Hence $\mathbf{y} \notin \mathrm{ai}(D)$, since otherwise we would have $[\mathbf{y}, \mathbf{z}) \subset D$, which would lead to

$(\mathbf{y}, \mathbf{z}) \subset C \cap D$, contradicting the fact that C and D are disjoint. So we obtain the reverse inclusion $\mathrm{ac}(C) \subset E \backslash \mathrm{ai}(D)$, and therefore (2.7) holds.

Since the sets C and D have equal roles, by similar arguments we obtain

$$\mathrm{ac}(D) = E \backslash \mathrm{ai}(C). \tag{2.8}$$

It follows immediately from (2.7) and (2.8) that $L = E$ if and only if both $\mathrm{ai}(C)$ and $\mathrm{ai}(D)$ are empty, or equivalently, $\mathrm{ac}(C) = \mathrm{ac}(D) = E$.

Now we consider the case that $L \subsetneq E$. In this case we need to show that L is a hyperplane.

Firstly, we observe that L is an affine set. Indeed, let $\mathbf{x}, \mathbf{y}$ are arbitrary points in L, and $\mathbf{z} \in E$ such that $\mathbf{y} \in (\mathbf{x}, \mathbf{z})$. Assume the contrary that $\mathbf{z} \notin L = \mathrm{ac}(C) \cap \mathrm{ac}(D)$. If $\mathbf{z} \notin \mathrm{ac}(C)$, then by (2.7) we have $\mathbf{z} \in \mathrm{ai}(D)$. However, in this case, since $\mathbf{x} \in \mathrm{ac}(D)$, it follows from Proposition 1.25 that $\mathbf{y} \in \mathrm{ai}(D)$. In turn, by (2.7) this means that $\mathbf{y} \notin \mathrm{ac}(C)$. This contradicts our setting that $\mathbf{y} \in L = \mathrm{ac}(C) \cap \mathrm{ac}(D) \subset \mathrm{ac}(C)$. This contradiction proves that $\mathbf{z} \in L$, which implies that L is affine.

Since we are considering the case that $L \subsetneq E$, we can pick some $\mathbf{p} \notin L$. Since a hyperplane in E is a maximal affine set in E (cf. Proposition 2.6), to show that L is a hyperplane it suffices to prove $E = \mathrm{aff}(L \cup \{\mathbf{p}\})$. Indeed, since $\mathbf{p} \notin L = \mathrm{ac}(C) \cap \mathrm{ac}(D)$, we may assume without loss of generality that $\mathbf{p} \notin \mathrm{ac}(D)$. Hence, by (2.8) we have $\mathbf{p} \in \mathrm{ai}(C)$. Now, let us take $\mathbf{r} \in L$ and consider $\mathbf{q} = 2\mathbf{r} - \mathbf{p}$. By this choice, $\mathbf{r} \in (\mathbf{p}, \mathbf{q})$. Observe that if $\mathbf{q} \in \mathrm{ac}(C)$, then again by Proposition 1.25 we have $\mathbf{r} \in \mathrm{ai}(C) = E \backslash \mathrm{ac}(D)$, contradicting $r \in L \subset \mathrm{ac}(D)$. Hence, we must have $\mathbf{q} \in E \backslash \mathrm{ac}(C) = \mathrm{ai}(D)$. Therefore, if we take an arbitrary point $\mathbf{x} \in C \backslash L$, then the line segment $[\mathbf{x}, \mathbf{q}]$ must intersect L, so $\mathbf{x} \in \mathrm{aff}(L \cup \{\mathbf{p}\})$. With the similar argument, if we pick an arbitrary point $\mathbf{y} \in D \backslash L$, then $\mathbf{y} \in \mathrm{aff}(L \cup \{p\})$. Altogether, we have $E = \mathrm{aff}(L \cup \{p\}$ as desired.

It follows from (2.7) and (2.8) that $\mathrm{ai}(C)$, $\mathrm{ai}(D)$, L are pairwise disjoint, and their union is E. The arguments (i), (ii), (iii) follows immediately. $\square$

Now we come to the first separation theorem in the setting of general vector spaces.

Theorem 2.19. (See e.g. [1]). *Let $C, D \subset E$ be nonempty convex sets such that $\mathrm{ai}(C) \neq \varnothing$. Then C and D can be separated by a hyperplane H in E if and only if $\mathrm{ai}(C) \cap D = \varnothing$. In this case, $\mathrm{ai}(C)$ is contained in one of the algebraically open half-spaces associated with H.*

Proof. Necessity. Let C and D be separated by a hyperplane H in such a way that $C \subseteq \bar{H}^+$ and $D \subseteq \bar{H}^-$. Since $\mathrm{ai}(C) \neq \varnothing$, we have $\mathrm{aff}(C) = E$. Since a hyperplane in E is also an affine set, it follows that C must not be contained in H. So we can pick a point $\mathbf{y} \in C \cap H^+$. Hence, if there were $\mathbf{x} \in \mathrm{ai}(C) \cap H$, then by definition of algebraic interior we would have a point $\mathbf{z} \in C$ such that $\mathbf{x} \in (\mathbf{y}, \mathbf{z})$. Keeping in mind that $\mathbf{y} \in H^+$ and $\mathbf{x} \in H$, this would imply furthermore that $z \in C \cap H^-$. However, this contradicts the fact that $C \cap H^- = \varnothing$ (since we assume $C \subseteq \bar{H}^+$, it follows that C and H^- are disjoint). This contradiction ensures that $\mathrm{ai}(C) \cap H = \varnothing$. This, together with the fact that $C \subset \bar{H}^+$, implies $\mathrm{ai}(C) \subseteq H^+$, *i.e.*, $\mathrm{ai}(C)$ is contained in the open half-space H^+ associated with H. Since $H^+ \cap \bar{H}^- = \varnothing$ and $D \subset \bar{H}^-$, it follows that $\mathrm{ai}(C) \cap D = \varnothing$.

Sufficiency. Assume that $\mathrm{ai}(C) \cap D = \varnothing$. Since C is convex, by Proposition 1.24 we have $\mathrm{ai}(C)$ is convex. Applying Lemma 2.17 for disjoint convex sets $\mathrm{ai}(C)$ and D, there exists complementary convex sets C' and D' such that $\mathrm{ai}(C) \subseteq C'$ and $D \subseteq D'$. Let $\mathbf{x} \in \mathrm{ai}(C)$. Then by definition of algebraic interior, for any $\mathbf{y} \in E$ there is $\mathbf{u} \in C$ such that $\mathbf{x} \in (\mathbf{u}, \mathbf{y})$. By Proposition 2.6, $[\mathbf{x}, \mathbf{u}) \subseteq \mathrm{ai}(C)$, so we can assume that $\mathbf{u} \in \mathrm{ai}(C)$. Since $\mathrm{ai}(C) \subset C'$, we have $\mathbf{x}, \mathbf{u} \in C'$. Since C' is convex, it follows that $[\mathbf{x}, \mathbf{u}) \in C'$. Hence we obtain $\mathbf{x} \in \mathrm{ai}(C')$. Since $\mathbf{x}$ is chosen arbitrarily in $\mathrm{ai}(C)$, we come up with $\mathrm{ai}(C) \subseteq \mathrm{ai}(C')$.

Since C' and D' are complementary convex sets, by Lemma 2.18 the set $H := \mathrm{ac}(C') \cap \mathrm{ac}(D')$ is a hyperplane separating C' and D'. Since $\mathrm{ai}(C) \subseteq C'$ and $D \subseteq D'$, the hyperplane H also separates $\mathrm{ai}(C)$ and D. Without loss of generality, we assume that $\mathrm{ai}(C) \subseteq \bar{H}^+$ and $D \subseteq \bar{H}^-$.

We now show that $C \subseteq \bar{H}^+$. Indeed, since $H^- \cap \bar{H}^+ = \varnothing$ and $H^- \cup \bar{H}^+ = E$, if we assume the contrary, then $C \cap H^- \neq \varnothing$ and therefore we can pick some $\mathbf{x} \in C \cap H^-$. Pick $\mathbf{y} \in \mathrm{ac}(C) \subseteq \bar{H}^+$. By Proposition 1.25, $(\mathbf{y}, \mathbf{x})$ contains a point $\mathbf{z} \in \mathrm{ai}(C) \cap H^-$. However, since $\mathrm{ai}(C) \subseteq \bar{H}^+$, we have $\mathrm{ai}(C) \cap H^- = \varnothing$, which contradicts the existence of $\mathbf{z}$.

We have shown that $C \subset \bar{H}^+$ and $D \subset \bar{H}^-$. This means that C ans D are separated by H. $\square$

For the second separation theorem in the setting of general vector spaces, we need the result stated in the following lemma. It is worth noting that this lemma generalizes Lemma 2.13.

Lemma 2.20. (See e.g. [1]). *Two nonempty convex sets C and D in E can be properly separated if and only if the set $\{\mathbf{0}\}$ and the convex set $K := C - D$ can be properly separated.*

Proof. Necessity. Let $H := H(h, \xi)$ be a hyperplane properly separating C and D such that $C \subseteq \bar{H}^+, D \subseteq \bar{H}^-$. Without loss of generality, assume that C does not lie on H. Then we have

$$h(\mathbf{x}) \geq \xi \geq h(\mathbf{y}) \quad \forall \mathbf{x} \in C, \mathbf{y} \in D,$$

and $h(\mathbf{x}^0) > \xi$ for some $\mathbf{x}^0 \in C$. This implies that the hyperplane $H(h, \mathbf{0})$ properly separates the sets $\{\mathbf{0}\}$ and K.

Sufficiency. Suppose that there exists a hyperplane $H(h, \xi)$ properly separating the sets $\{\mathbf{0}\}$ and K such that $K \subseteq \bar{H}^+(h, \xi)$. Then $h(\mathbf{x} - \mathbf{y}) \geq \xi \geq 0$ for all $\mathbf{x} \in C, \mathbf{y} \in D$. By the proper separation, either $\xi > 0$ or $h(\mathbf{x}^0 - \mathbf{y}^0) > \xi$ for some $\mathbf{x}^0 \in C, \mathbf{y}^0 \in D$.

In the former case ($\xi > 0$), we have

$$h(\mathbf{x}) \geq \xi + h(\mathbf{y}) > \frac{\xi}{2} + h(\mathbf{y}) > h(\mathbf{y}) \quad \forall \mathbf{x} \in C, \mathbf{y} \in D.$$

Observe furthermore that

$$h(\mathbf{x}) \geq \frac{\xi}{2} + \sup_{\mathbf{y} \in D} h(\mathbf{y}) > h(\mathbf{y}) \forall \mathbf{x} \in C, \mathbf{y} \in D,$$

which implies that the hyperplane $H(h, \beta)$ with $\beta = \frac{\xi}{2} + \sup_{\mathbf{y} \in D} h(\mathbf{y})$ properly separates the sets C and D.

In the latter case, from the inequality

$$h(\mathbf{x}) \geq \xi + h(\mathbf{y}) \geq h(\mathbf{y}) \quad \forall \mathbf{x} \in C, \mathbf{y} \in D$$

and the fact that $h(\mathbf{x}^0) > \xi + h(\mathbf{y}^0) \geq h(y^0)$ for some $\mathbf{x}^0 \in C, y^0 \in D$, we derive that any hyperplane $H(h, \beta)$ with $\beta \in \mathbb{R}$ satisfying

$$\inf_{\mathbf{x} \in C} h(\mathbf{x}) \geq \beta \geq \sup_{\mathbf{y} \in D} h(\mathbf{y})$$

properly separates C and D. □

We come up with the following separation theorem which can be seen as a generalization of Theorem 2.15 (proper separation theorem in the setting of finite dimensional spaces) to the setting of general vector spaces.

Theorem 2.21. (Proper separation theorem in general vector spaces, see e.g. [1]). *Let C and D be nonempty convex sets in E such that both* $\operatorname{rai}(C)$ *and* $\operatorname{rai}(D)$ *are nonempty. Then C and D can be properly separated if and only if* $\operatorname{rai}(C) \cap \operatorname{rai}(D) = \varnothing$.

Proof. Let $K := C - D$. Since both C and D are convex, it follows from Proposition 1.9(ii) that K is also convex. By Proposition 1.26, we have $\operatorname{rai}(K) = \operatorname{rai}(C - D) =$

$\text{rai}(C) - \text{rai}(D)$. Then it is readily to see that $\text{rai}(C) \cap \text{rai}(D) = \varnothing$ if and only if $\mathbf{0} \notin \text{rai}(K)$. By using Lemma 2.20, it is left to prove that the sets $\{\mathbf{0}\}$ and K are properly separated if and only if $\mathbf{0} \notin \text{rai}(K)$.

Necessity. Let H be a hyperplane properly separating $\{\mathbf{0}\}$ from K in such a way that $\mathbf{0} \in \bar{H}^-$ and $K \subseteq \bar{H}^+$. Since $\mathbf{0} \in \bar{H}^-$, there are two following cases.

- If $\mathbf{0} \notin H$, then $\mathbf{0}$ must be in H^-. Since $\text{rai}(K) \subseteq K \subseteq \bar{H}^+$ and note that H^- is disjoint with $\bar{H}^+$, it follows that $\mathbf{0}$ does not belong to $\text{rai}(K)$.

- If $\mathbf{0} \in H$, then there exists $\mathbf{x} \in K \backslash H$ due to the proper separation between $\{\mathbf{0}\}$ and K. If $\mathbf{0} \in \text{rai}(K)$, then there would be some $\mathbf{y}$ in K such that $\mathbf{0} \in (\mathbf{x}, \mathbf{y})$. Then $\mathbf{y}$ must be in H^-. However, this contradicts with $\mathbf{y} \in K$, since $K \subset \bar{H}^+$ and $\bar{H}^+ \cap H^- = \varnothing$. Therefore, in this case $\mathbf{0}$ must not be in $\text{rai}(K)$.

In both cases above, we have $\mathbf{0} \notin \text{rai}(K)$ as desired.

Sufficiency. We are given that $\mathbf{0} \notin \text{rai}(K)$. Let L be the affine hull of K. There are two following cases.

- $\mathbf{0}$ is not in L. In this case, let

  $$\mathcal{G} = \{F \subset E \mid F \text{ is affine, } F \supset L, \mathbf{0} \notin F\}.$$

  Clearly, $\mathcal{G}$ is partially ordered by set inclusion. By the well-known Zorn's lemma, $\mathcal{G}$ contains a maximal element H. As a maximal element in $\mathcal{G}$, we have $H \supset L$, H does not contain $\mathbf{0}$, and H is affine. If H is not a hyperplane, then it is not a maximal affine set in E. Since H does not contain $\mathbf{0}$, it follows that $H' := \text{aff}(\{\mathbf{0}\} \cup H) \neq E$, and hence there exists $\mathbf{x} \notin H'$. Take $\tilde{H} := \text{aff}(\{\mathbf{x}\} \cup H)$. Then $\tilde{H}$ is an affine set, $\tilde{H} \supset H$, and $\tilde{H}$ does not contain $\mathbf{0}$. The existence of $\tilde{H}$ contradicts the maximality of H in $\mathcal{G}$. This contradiction means that H must be a hyperplane. Since H is a hyperplane containing L but not $\mathbf{0}$, we have proper separation between $\mathbf{0}$ and L. Since $K \subseteq L$, this implies the proper separation between $\mathbf{0}$ and K.

- $\mathbf{0}$ is in L. In this case, we have L as a vector subspace of E. Note that $\mathbf{0} \notin \text{rai}(K)$. By applying Theorem 2.19 to the sets $\{\mathbf{0}\}$ and K relative to L, we obtain a hyperplane P in L separating $\mathbf{0}$ and K such that $\text{rai}(K) \subseteq P^+$. Since the translation of P to the one containing $\mathbf{0}$ also satisfies the same separation properties, we can assume that $\mathbf{0} \in P$. By the well-known Zorn's lemma, there exists H as a maximal linear subspace of E such that $H \supset P$ and $P = H \cap L$. If H is not a hyperplane in E, then one can pick some $\mathbf{x} \notin H$ and obtain $H' := \text{span}(\{\mathbf{x}\} \cup H) \supsetneq H$. We observe that $H' \cap L = P$. Indeed, for any

$\mathbf{y} = \xi\mathbf{x} + \mathbf{h} \in H' \cap L$ with $\xi \in \mathbb{R}, \mathbf{h} \in H$, we have $\mathbf{y} \in H$, then $\xi\mathbf{x} \in H$. Since $\mathbf{x} \in E \backslash H$, we must have $\xi = 0$, then $\mathbf{y} \in H \cap L = P$. Hence $H' \cap L \subseteq P$. Obviously $P \subseteq H' \cap L$, so we obtain $H \cap L = P$, which contradicts to the maximality of H. Therefore, H is a hyperplane, and as the previous case, it is easy to see that H separates $\mathbf{0}$ and K.

In both cases above, we have the proper separation between $\mathbf{0}$ and K as desired. □

We close this subsection with a result on proper separation between a convex set and an affine set.

Theorem 2.22. (See e.g. [1]). *Let $C \subset E$ be a nonempty convex set and $M \subset E$ an affine set satisfying* $\text{rai}(C) \cap M = \varnothing$. *Then there exists a hyperplane $H \supseteq M$ such that* $\text{rai}(C) \cap H = \varnothing$.

Proof. Since M is an affine set, clearly $\text{rai}(M) = M$. It readily follows that $\text{rai}(M)$ and $\text{rai}(C)$ are disjoint. By Theorem 2.21, there exists a hyperplane $H(h, \xi)$ properly separating C and M. Then we have

$$h(\mathbf{x}) \geq \xi \geq h(\mathbf{y}) \quad \forall \mathbf{x} \in M, \mathbf{y} \in C.$$

We claim that $h(\mathbf{x})$ is constant on the affine set M. Indeed, assume the contrary that there exists $\mathbf{x}^*, \mathbf{y}^* \in M$ such that $h(\mathbf{x}^*) \neq h(\mathbf{y}^*)$. Since M is affine, for any $t \in \mathbb{R}$ we have $t\mathbf{x}^* + (1-t)\mathbf{y}^* = \mathbf{y}^* + t(\mathbf{x}^* - \mathbf{y}^*) \in M$. Since h is linear, we have $h(\mathbf{y}^* + t(\mathbf{x}^* - \mathbf{y}^*)) = h(\mathbf{y}^*) + t(h(\mathbf{x}^*) - h(\mathbf{y}^*)) \geq \xi$ for all $t \in \mathbb{R}$. By letting $t \to -\infty$, we obtain a contradiction.

Let $h(\mathbf{x}) = \beta$ for some $\beta \in R$ and for all $\mathbf{x} \in M$. If $\beta = \xi$, which implies $M \subseteq H$, we are done. Otherwise, from the fact that $\beta > \xi \geq h(\mathbf{y})$ for all $\mathbf{y} \in C$, we derive the hyperplane $H(h, \beta)$ containing M and does not intersect the set $\text{rai}(C)$. □

Chapter 3

Some related problems

In this chapter, we present some results related to the separation theorems mentioned in the previous chapter. Namely, in Section 3.1 we will show that the well-known homogeneous Farkas lemma can be viewed as a consequence of the strong separation theorem. In Section 3.2 we will present a particular case in duality theory that bases also on the strong separation theorem. Section 3.3 presents the use of the first separation theorem in constructing a barrier convex function for the feasible set of a convex optimization problem. The connection between the well-known Hahn-Banach theorem with proper separation of convex sets in general vector spaces is presented in Section 3.4.

3.1 Homogeneous Farkas lemma

Homogeneous Farkas lemma is a result on the solvability of a finite system of homogeneous linear inequalities. It is named after the Hungarian mathematician Gyula Farkas who gave the first proof for the result. In the setting of $\mathbb{R}^n$ with the usual inner product $\langle \cdot, \cdot \rangle$, the lemma is stated as follows.

Lemma 3.1 (Homogeneous Farkas lemma). (See e.g. [5]). *Let* $\mathbf{a}, \mathbf{a}^1, \ldots, \mathbf{a}^m$ *be vectors in* $\mathbb{R}^n \setminus \{\mathbf{0}\}$. *Then the following system of homogeneous linear inequalities in* $\mathbf{x} \in \mathbb{R}^n$

$$(F) \qquad \begin{cases} \langle \mathbf{a}, \mathbf{x} \rangle & < 0 \\ \langle \mathbf{a}^i, \mathbf{x} \rangle & \geq 0 \quad (i = 1, \ldots, m) \end{cases}$$

is infeasible if and only if there exist non-negative numbers $\lambda_1, \ldots, \lambda_m$ *such that*

$$\mathbf{a} = \sum_{i=1}^{m} \lambda_i \mathbf{a}^i. \tag{3.1}$$

Roughly speaking, the representation (3.1) means that $\mathbf{a}$ belongs to the conic hull of vectors $\mathbf{a}^1, \ldots, \mathbf{a}^m$. With that point of view, the homogeneous Farkas lemma has an obvious geometric illustration as follows. In Figure 3.1, we are given three vectors $\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3$ in $\mathbb{R}^2$, as well as a vector $\mathbf{x} \in \mathbb{R}^2$ such that $\langle \mathbf{a}^1, \mathbf{x}\rangle \geq 0$, $\langle \mathbf{a}^2, \mathbf{x}\rangle \geq 0$, $\langle \mathbf{a}^3, \mathbf{x}\rangle \geq 0$. On the left, we have a vector $\mathbf{a}$ in the conic hull of vectors $\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3$. In that case, we can easily see that $\langle \mathbf{a}, \mathbf{x}\rangle \geq 0$, and therefore the system (F) in this context is infeasible (since its first inequality is violated). On the right, we have a vector $\mathbf{a}$ satisfying that $\langle \mathbf{a}, \mathbf{x}\rangle < 0$. In that case, the system (F) is feasible, and we can easily see that $\mathbf{a}$ is not in the convex cone generated by vectors $\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3$.

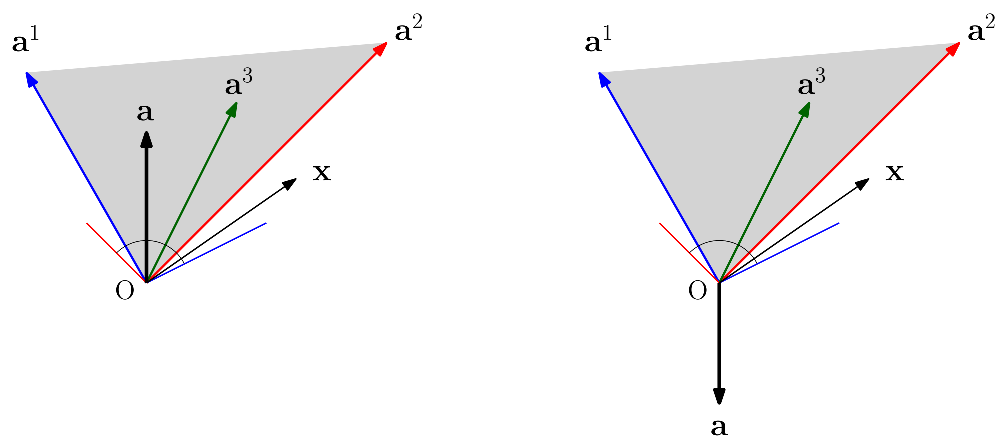

Figure 3.1: Illustration of homogeneous Farkas lemma.

Unlike the obvious illustration above, it is not trivial to prove the homogeneous Farkas lemma. In this section we present a proof of the lemma using the theorem on strong separation of convex sets. For the proof we need the following results.

Lemma 3.2. *The conic hull of any set of linearly independent vectors in $\mathbb{R}^n$ is closed.*

Proof. Let $V = \text{cone}(\mathbf{v}^1, \ldots, \mathbf{v}^\ell)$ in which $\mathbf{v}^1, \ldots, \mathbf{v}^\ell$ are linearly independent vectors in $\mathbb{R}^n$. We need to prove that V is closed. Indeed, let $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ be a sequence of vectors in V converging to some vector $\mathbf{x}$. What we need to show now is $\mathbf{x} \in V$.

For each $k \in \mathbb{N}$, since $\mathbf{x}^k \in V$, we can represent

$$\mathbf{x}^k = \xi_1^k \mathbf{v}^1 + \ldots + \xi_\ell^k \mathbf{v}^\ell$$

in which $\xi_1^k, \dots, \xi_\ell^k \geq 0$. Hence, $\mathbf{x}^k$ lies in the subspace $W = \text{span}(\mathbf{v}^1, \dots, \mathbf{v}^\ell)$ spanned by vectors $\mathbf{v}^1, \dots, \mathbf{v}^\ell$. Since finite dimensional subspaces of $\mathbb{R}^n$ are closed, $\mathbf{x}$ must also lie in W. Since $\mathbf{v}^1, \dots, \mathbf{v}^\ell$ are linearly independent, there exists unique $\xi_1, \dots, \xi_\ell$ such that

$$\mathbf{x} = \xi_1 \mathbf{v}^1 + \dots + \xi_\ell \mathbf{v}^\ell.$$

Now we prove that $\xi_i^k \to \xi_i$ for each $i = 1, \dots, \ell$. In the following we will show the proof in case $i = 1$, the other cases of i can be shown similarly. Let $F = \text{span}(\mathbf{v}^2, \dots, \mathbf{v}^\ell)$ be the subspace spanned by vectors $\mathbf{v}^2, \dots, \mathbf{v}^\ell$. Then F is a finite dimensional subspace of $\mathbb{R}^n$, hence it is closed. Since $\mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^\ell$ are linearly independent, we have $\mathbf{v}^1 \notin F$. Let $\mathbf{u} = \mathbf{v}^1 - \text{proj}_F(\mathbf{v}^1)$. Clearly, $\mathbf{u} \neq \mathbf{0}$ since $\mathbf{v}^1 \notin F$ while $\text{proj}_F(\mathbf{v}^1) \in F$. It follows that

$$\|\mathbf{u}\| > 0. \tag{3.2}$$

Since $\text{proj}_F(\mathbf{v}^1) \in F$ and F is a subspace of $\mathbb{R}^n$, for any $\mathbf{z} \in F$ and $\lambda \in \mathbb{R}$ we have $\text{proj}_F(\mathbf{v}^1) + \lambda \mathbf{z} \in F$. Then, by definition of $\text{proj}_F(\mathbf{v}^1)$ we obtain

$$\begin{aligned}
\|\mathbf{u}\|^2 &= \|\mathbf{v}^1 - \text{proj}_F(\mathbf{v}^1)\|^2 \\
&\leq \|\mathbf{v}^1 - (\text{proj}_F(\mathbf{v}^1) + \lambda \mathbf{z})\|^2 \\
&= \|(\mathbf{v}^1 - \text{proj}_F(\mathbf{v}^1)) - \lambda \mathbf{z}\|^2 \\
&= \|\mathbf{u} - \lambda \mathbf{z}\|^2 \\
&= \langle \mathbf{u} - \lambda \mathbf{z}, \mathbf{u} - \lambda \mathbf{z} \rangle \\
&= \|\mathbf{u}\|^2 - 2\lambda \langle \mathbf{u}, \mathbf{z} \rangle + \lambda^2 \|\mathbf{z}\|^2,
\end{aligned}$$

or equivalently

$$2\lambda \langle \mathbf{u}, \mathbf{z} \rangle \leq \lambda^2 \|\mathbf{z}\|^2.$$

By letting

$$\lambda = \frac{1}{\|\mathbf{z}\|^2 + 1} \langle \mathbf{u}, \mathbf{z} \rangle,$$

we obtain

$$2|\langle \mathbf{u}, \mathbf{z} \rangle|^2 \leq \frac{\|\mathbf{z}\|^2}{\|\mathbf{z}\|^2 + 1} |\langle \mathbf{u}, \mathbf{z} \rangle|^2$$

or equivalently

$$\frac{\|\mathbf{z}\|^2 + 2}{\|\mathbf{z}\|^2 + 1} |\langle \mathbf{u}, \mathbf{z} \rangle|^2 \leq 0,$$

which implies that

$$\langle \mathbf{u}, \mathbf{z} \rangle = 0 \quad \forall \mathbf{z} \in F. \tag{3.3}$$

As a consequence, we have

$$\begin{aligned}\langle \mathbf{v}^1, \mathbf{u}\rangle &= \langle (\mathbf{v}^1 - \mathrm{proj}_F(\mathbf{v}^1)) + \mathrm{proj}_F(\mathbf{v}^1), \mathbf{u}\rangle \\ &= \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathrm{proj}_F(\mathbf{v}^1), \mathbf{u}\rangle \\ &= \|\mathbf{u}\|^2. \end{aligned} \tag{3.4}$$

The last equality follows from (3.3) and the fact that $\mathrm{proj}_F(\mathbf{v}^1) \in F$. By Cauchy-Schwartz inequality, we see furthermore that

$$\begin{aligned}\|\mathbf{x}^k - \mathbf{x}\|\|\mathbf{u}\| &\geq |\langle \mathbf{x}^k - \mathbf{x}, \mathbf{u}\rangle| \\ &= |\langle (\xi_1^k - \xi_1)\mathbf{v}^1 + (\xi_2^k - \xi_2)\mathbf{v}^2 + \ldots + (\xi_\ell^k - \xi_\ell)\mathbf{v}^\ell, \mathbf{u}\rangle| \\ &= |(\xi_1^k - \xi_1)\langle \mathbf{v}^1, \mathbf{u}\rangle + (\xi_2^k - \xi_2)\langle \mathbf{v}^2, \mathbf{u}\rangle + \ldots + (\xi_\ell^k - \xi_\ell)\langle \mathbf{v}^\ell, \mathbf{u}\rangle| \\ &= |\xi_1^k - \xi_1||\langle \mathbf{v}^1, \mathbf{u}\rangle|. \end{aligned} \tag{3.5}$$

The last equality is because of (3.3) and the fact that $\mathbf{a}^2, \ldots, \mathbf{a}^\ell \in F$. Combining (3.5) with (3.4) we obtain

$$\|\mathbf{x}^k - \mathbf{x}\|\|\mathbf{u}\| \geq |\xi_1^k - \xi_1|\|\mathbf{u}\|^2.$$

Keeping (3.2) in mind, it follows that

$$\|\mathbf{x}^k - \mathbf{x}\| \geq |\xi_1^k - \xi_1|\|\mathbf{u}\|.$$

As $\mathbf{x}^k \to \mathbf{x}$ by our assumption, letting $k \to \infty$ we have $\|\mathbf{x}^k - \mathbf{x}\| \to 0$. Together with (3.2), it follows from the above inequality that $|\xi_1^k - \xi_1| \to 0$ as $k \to \infty$, or equivalently, $\xi_1^k \to \xi_1$.

Now we have $\xi_i^k \to \xi_i$ for $i = 1, \ldots, \ell$. Since $\xi_1^k, \ldots, \xi_\ell^k \geq 0$ for all $k \in \mathbb{N}$, we have $\xi_i \geq 0$. Thus $\mathbf{x} = \xi_1\mathbf{v}^1 + \ldots + \xi_\ell\mathbf{v}^\ell$ is a conic combination of $\mathbf{v}^1, \ldots, \mathbf{v}^\ell$, *i.e.*, $\mathbf{x} \in V$. This proves the closedness of V. □

Proposition 3.3. *Let* $K := \mathrm{cone}(\mathbf{a}^1, \ldots, \mathbf{a}^m)$. *Then* K *is a closed convex cone.*

Proof. *Conic property of* K. Let $\mathbf{x} \in K$ and $\theta \geq 0$. Since $\mathbf{x} \in K$, it admits the following representation

$$\mathbf{x} = \lambda_1\mathbf{a}^1 + \ldots + \lambda_m\mathbf{a}^m$$

for some $\lambda_1, \ldots, \lambda_m \geq 0$. Then we have

$$\theta\mathbf{x} = \theta\lambda_1\mathbf{a}^1 + \ldots + \theta\lambda_m\mathbf{a}^m.$$

Since θ is also non-negative, the coefficients $\theta\lambda_i$ $(i = 1, \ldots, m)$ in the above representation are non-negative. Therefore $\theta\mathbf{x} \in K$ by definition of K.

Convexity of K. Let $\mathbf{x}, \mathbf{y} \in K$ and $\theta \in [0, 1]$. Since $\mathbf{x}, \mathbf{y} \in K$, they admits the following representations

$$\mathbf{x} = \lambda_1 \mathbf{a}^1 + \ldots + \lambda_m \mathbf{a}^m, \quad \mathbf{y} = \mu_1 \mathbf{a}^1 + \ldots + \mu_m \mathbf{a}^m$$

for some $\lambda_1, \ldots, \lambda_m \geq 0$ and $\mu_1, \ldots, \mu_m \geq 0$. Then we have

$$\begin{aligned} \mathbf{z} &= \theta \mathbf{x} + (1-\theta)\mathbf{y} \\ &= \theta(\lambda_1 \mathbf{a}^1 + \ldots + \lambda_m \mathbf{a}^m) + (1-\theta)(\mu_1 \mathbf{a}^1 + \ldots + \mu_m \mathbf{a}^m) \\ &= (\theta\lambda_1 + (1-\theta)\mu_1)\mathbf{a}^1 + \ldots + (\theta\lambda_m + (1-\theta)\mu_m)\mathbf{a}^m \end{aligned}$$

Since $\theta \in [0, 1]$ and $\lambda_i \geq 0, \mu_i \geq 0$ $(i = 1, \ldots, m)$, we have $\theta\lambda_i + (1-\theta)\mu_i \geq 0$ for all $i = 1, \ldots, m$. Therefore $z \in K$ by definition of K, which confirms convexity of K.

Closedness of K. Let

$$\mathcal{I} = \{J \subset \{1, \ldots, m\} \mid \mathbf{a}^j \ (j \in J) \text{ are linearly independent}\},$$

and

$$C = \bigcup_{J \in \mathcal{I}} \operatorname{cone}\left(\{\mathbf{a}^j \mid j \in J\}\right).$$

Roughly speaking, C is the union of conic hulls of linearly independent subsets of $\{\mathbf{a}^1, \ldots, \mathbf{a}^m\}$. This is a finite union (*i.e.* $|\mathcal{I}|$ is finite) since the index set $\{1, \ldots, m\}$ is finite. For each $J \in \mathcal{I}$ we have $\{\mathbf{a}^j \mid j \in J\} \subset \{\mathbf{a}^1, \ldots, \mathbf{a}^m\}$, hence $\operatorname{cone}\left(\{\mathbf{a}^j \mid j \in J\}\right) \subset \operatorname{cone}\left(\{\mathbf{a}^1, \ldots, \mathbf{a}^m\}\right) = K$. Therefore $C \subseteq K$.

We now show that $K \subseteq C$. Indeed, let $\mathbf{x}$ be an arbitrary nonzero vector in K. Then it can be represented as

$$\mathbf{x} = \xi_1 \mathbf{a}^1 + \ldots + \xi_m \mathbf{a}^m \tag{3.6}$$

where $\xi_i \geq 0$ for $i = 1, \ldots, m$. Since $\mathbf{x} \neq \mathbf{0}$, we have $(\xi_1, \ldots, \xi_m) \neq (0, \ldots, 0)$. The terms with zero coefficients can be removed from the sum on the right hand side of (3.6). By renumbering the indices, without loss of generality we can assume that $\mathbf{x}$ admits a shorten representation

$$\mathbf{x} = \xi_1 \mathbf{a}^1 + \ldots + \xi_k \mathbf{a}^k \tag{3.7}$$

with $k \leq m$ and $\xi_i > 0$ for $i = 1, \ldots, k$. If $\mathbf{a}^1, \ldots, \mathbf{a}^k$ are linearly independent, then $\mathbf{x} \in C$ by definition of C. Otherwise, there exists $(\beta_1, \ldots, \beta_k) \neq (0, \ldots, 0)$ such that

$$\mathbf{0} = \beta_1 \mathbf{a}^1 + \ldots + \beta_k \mathbf{a}^k. \tag{3.8}$$

By multiplying both sides of (3.8) with -1 if needed, we can assume furthermore that there exists at least one positive coefficient in $\beta_1, \dots, \beta_k$. For any $s \in \mathbb{R}$, from (3.7) and (3.8) we have

$$\begin{aligned}\mathbf{x} = \mathbf{x} - s \cdot \mathbf{0} &= (\xi_1 \mathbf{a}^1 + \dots + \xi_k \mathbf{a}^k) - s(\beta_1 \mathbf{a}^1 + \dots + \beta_k \mathbf{a}^k) \\ &= (\xi_1 - s\beta_1)\mathbf{a}^1 + \dots + (\xi_k - s\beta_k)\mathbf{a}^k. \qquad (3.9)\end{aligned}$$

Let us take

$$s := s^* = \min \left\{ \frac{\xi_i}{\beta_i} \mid i \in \{1, \dots, k\} \text{ with } \beta_i > 0 \right\}$$

and let I^* be the set of indices where the above minimum is attained. Since all coefficients ξ_i $(i = 1, \dots, k)$ are positive, it follows from the choice of s^* that $s^* > 0$. Then the following holds.

- For any $i \in \{1, \dots, k\}$ with $\beta_i < 0$, since $\xi_i > 0$ and $s^* > 0$, we have $\xi_i - s^*\beta_i > 0$.
- For any $i \in \{1, \dots, k\}$ with $\beta_i = 0$, since $\xi_i > 0$, we have $\xi_i - s^*\beta_i = \xi_i > 0$.
- For $i \in \{1, \dots, k\}$ with $\beta_i > 0$: if $i \in I^*$, then $\xi_i - s^*\beta_i = 0$, otherwise $\xi_i - s^*\beta_i > 0$ (by definition of s^* and I^*).

Therefore, by substituting $s = s^*$ in (3.9) and then removing the terms having zero coefficients, we obtain a representation of $\mathbf{x}$ as a conic combination of a proper subset of $\{\mathbf{a}^1, \dots, \mathbf{a}^k\}$ with positive coefficients. Removing the vectors that are not in the proper subset, and as long as the remaining vectors are still linearly dependent, we can repeat the above procedure. This process stops when the remaining vectors are linearly independent, and we obtain a representation of $\mathbf{x}$ as a conic combination of some linearly independent vectors in $\{\mathbf{a}^1, \dots, \mathbf{a}^m\}$. This means $\mathbf{x} \in C$. Since $\mathbf{x}$ is chosen arbitrarily in K, we come up with $K \subseteq C$.

We have proved that $C \subseteq K$ and $K \subseteq C$, so $K = C$. Recall that, by construction, C is the union of a finite number of sets, each of such sets is the conic hull of some linearly independent vectors in $\{\mathbf{a}^1, \dots, \mathbf{a}^m\}$. By Lemma 3.2, such conic hulls are closed. Since the union of a finite number of closed sets is also closed, we obtain the closedness of C. Since $K = C$, we also have the closedness of K. □

We are now ready for the proof of the homogeneous Farkas lemma.

Proof of Lemma 3.1.

'If' part. Assume that there exist $\lambda_i \geq 0$ $(i = 1, \dots, m$ such that $\mathbf{a} = \sum_{i=1}^m \lambda_i \mathbf{a}^i$.

If the system of inequalities (F) is feasible, then

$$0 > \langle \mathbf{a}, \mathbf{x} \rangle = \sum_{i=1}^{m} \lambda_i \langle \mathbf{a}^i, \mathbf{x} \rangle \geq 0,$$

which is a contradiction. Therefore the system (F) must be infeasible.

'Only if' part. By Proposition 3.3 the set

$$K = \text{cone}(\mathbf{a}^1, \ldots, \mathbf{a}^m) = \left\{ \sum_{i=1}^{m} \lambda_i \mathbf{a}^i \mid \lambda_1, \ldots, \lambda_m \geq 0 \right\}$$

is a closed convex cone. What we need to show is that there exists non-negative numbers $\lambda_1, \ldots, \lambda_m$ such that $\mathbf{a} = \lambda_1 \mathbf{a}^1 + \ldots + \lambda_m \mathbf{a}^m$, i.e., we need to show that $\mathbf{a} \in K$. Assume the contrary that $\mathbf{a} \notin K$. Since $\{\mathbf{a}\}$ is compact, by Theorem 2.11 (strong separation theorem), there exists a vector $\mathbf{e} \in \mathbb{R}^n$ such that $\langle \mathbf{e}, \mathbf{a} \rangle > 0$ and that $\langle \mathbf{e}, \mathbf{u} \rangle \leq 0$ for all $\mathbf{u} \in K$. Let $\mathbf{x}^* = -\mathbf{e}$, we obtain

$$\begin{aligned} &\langle \mathbf{a}, \mathbf{x}^* \rangle < 0, \\ &\langle \mathbf{u}, \mathbf{x}^* \rangle \geq 0 \quad \forall \mathbf{u} \in K. \end{aligned}$$

Note that $\mathbf{a}^1, \ldots, \mathbf{a}^m \in K$, so respectively replacing $\mathbf{u}$ by these vectors we get

$$\begin{aligned} &\langle \mathbf{a}, \mathbf{x}^* \rangle < 0, \\ &\langle \mathbf{a}^i, \mathbf{x}^* \rangle \geq 0 \quad (i = 1, \ldots, m). \end{aligned}$$

This means that $\mathbf{x}^*$ is a solution of (F), which contradicts the infeasibility of this system. The contradiction means that $\mathbf{a}$ must be in K. □

3.2 Dual cone

In this section, we present a particular case in duality theory. For that we recall the following concept.

Definition 3.4. (Dual cone, see e.g. [1]). *Given a nonempty set $K \subseteq \mathbb{R}^n$. The set*

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

is called the dual cone of K.

The following proposition gives an important property of the concept of dual cone.

Proposition 3.5. *If $K \subseteq \mathbb{R}^n$ is a nonempty set, then its dual cone K^* is a closed convex set.*

Proof. *Conic property of K^*.* Let $\mathbf{y} \in K^*$ and $\theta \geq 0$. Since $\mathbf{y} \in K^*$, we have

$$\langle \mathbf{x}, \mathbf{y} \rangle \geq 0 \quad \forall \mathbf{x} \in K.$$

Since $\theta \geq 0$, it follows that

$$\langle \mathbf{x}, \theta\mathbf{y} \rangle = \theta\langle \mathbf{x}, \mathbf{y} \rangle \geq 0 \quad \forall \mathbf{x} \in K.$$

This means $\theta\mathbf{y} \in K^*$, hence K^* is conic.

Convexity of K^*. Let $\mathbf{y}^1, \mathbf{y}^2 \in K^*$ and $\theta \in [0,1]$. Since $\mathbf{y}^1, \mathbf{y}^2 \in K^*$, for all $\mathbf{x} \in K$ we have

$$\langle \mathbf{x}, \mathbf{y}^1 \rangle \geq 0 \quad \text{and} \quad \langle \mathbf{x}, \mathbf{y}^2 \rangle \geq 0.$$

Since $\theta \in [0,1]$, we have $\theta \geq 0$ and $1 - \theta \geq 0$. It follows that for all $\mathbf{x} \in K$ we have

$$\theta\langle \mathbf{x}, \mathbf{y}^1 \rangle \geq 0 \quad \text{and} \quad (1-\theta)\langle \mathbf{x}, \mathbf{y}^2 \rangle \geq 0.$$

Therefore we have

$$\langle \mathbf{x}, \theta\mathbf{y}^1 + (1-\theta)\mathbf{y}^2 \rangle = \theta\langle \mathbf{x}, \mathbf{y}^1 \rangle + (1-\theta)\langle \mathbf{x}, \mathbf{y}^2 \rangle \geq 0 \quad \forall \mathbf{x} \in K.$$

It means $\theta\mathbf{y}^1 + (1-\theta)\mathbf{y}^2 \in K^*$, hence K^* is convex.

Closedness of K^*. By the continuity of the function $f(\mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle$, we observe that for any sequence $\{\mathbf{y}^k \mid k \in \mathbb{N}\} \subset K^*$ such that $\mathbf{y}^k \to \bar{\mathbf{y}}$ we have $\langle \mathbf{x}, \mathbf{y}^k \rangle \to \langle \mathbf{x}, \bar{\mathbf{y}} \rangle$ as $k \to \infty$. Since $\langle \mathbf{x}, \mathbf{y} \rangle \geq 0$ for all $\mathbf{x} \in K$, we have $\langle \mathbf{x}, \bar{\mathbf{y}} \rangle \geq 0$ for all $\mathbf{x} \in K$. Hence $\bar{\mathbf{y}} \in K^*$, which implies K^* is a closed set. □

The main result in this section is as follows.

Theorem 3.6. (See e.g. [1]). *If $K \subseteq \mathbb{R}^n$ is a closed convex cone, then $K = K^{**}$.*

Proof. Firstly, we observe from definition of K^* that if $\mathbf{x} \in K$, then for any $\mathbf{y} \in K^*$ we have $\langle \mathbf{x}, \mathbf{y} \rangle \geq 0$, hence $\mathbf{x} \in K^{**}$. It means $K \subseteq K^{**}$. Now we prove the reserve inclusion $K^{**} \subseteq K$.

Suppose that the reserve inclusion does not hold. Then there exists $\bar{\mathbf{x}} \in K^{**} \backslash K$. Since $\bar{\mathbf{x}} \notin K$, by applying Theorem 2.11 (strong separation theorem) on the closed convex set K and the compact set $\{\bar{\mathbf{x}}\}$, there exists a vector $\mathbf{a} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ satisfying

$$\langle \mathbf{a}, \bar{\mathbf{x}} \rangle < \langle \mathbf{a}, \mathbf{z} \rangle \quad \forall \mathbf{z} \in K. \tag{3.10}$$

Since K is a cone, $\mathbf{0} \in K$. Taking $\mathbf{z} = \mathbf{0}$ in (3.10) gives $\langle \mathbf{a}, \bar{\mathbf{x}} \rangle < 0$. On the other hand, for every $\mathbf{z} \in K$, by conic property of K we have $t\mathbf{z} \in K$ for all $t > 0$. From (3.10), for all $t > 0$ we have

$$\langle \mathbf{a}, \bar{\mathbf{x}} \rangle < \langle \mathbf{a}, t\mathbf{z} \rangle = t\langle \mathbf{a}, \mathbf{z} \rangle.$$

Dividing both sides of above inequality by t, then letting $t \to +\infty$, we obtain

$$\langle \mathbf{a}, \mathbf{z} \rangle \geq 0 \quad \forall \mathbf{z} \in K,$$

which implies $\mathbf{a} \in K^*$. Note that $\bar{\mathbf{x}} \in K^{**}$, then $\langle \mathbf{a}, \bar{\mathbf{x}} \rangle \geq 0$, which contradicts the fact that $\langle \mathbf{a}, \bar{\mathbf{x}} \rangle < 0$ proved above. Therefore, $K^{**} \subseteq K$, hence $K^{**} = K$. $\square$

3.3 Convex barrier function

For a constrained optimization problem, a barrier function is a continuous one whose value on a point increases to $+\infty$ as the point approaches the boundary of the feasible region of the problem. The main motivation of the use of such function is to replace constraints by a penalizing term in the objective function, hence transform the constrained optimization problem under consideration to an unconstrained one. More precisely, let us consider the following constrained optimization problem for instance:

$$\min f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in C \subset \mathbb{R}^n \tag{3.11}$$

in which f is a continuous function on $\mathbb{R}^n$. With penalty method, this problem is equivalently reformulated as the following unconstrained one:

$$\min f(\mathbf{x}) + p(\mathbf{x})$$

in which

$$p(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in C \\ +\infty & \text{if } \mathbf{x} \notin C \end{cases}$$

is the penalizing term that replaces the constraint $\mathbf{x} \in C$. In this form, we get rid of the constraint but face up with discontinuity of the penalty function $p(\mathbf{x})$. For dealing with this issue, one can use a barrier function $b(\mathbf{x})$ as a continuous approximation of the penalty function and formulate a new optimization problem

$$\min f(\mathbf{x}) + \mu b(\mathbf{x}) \tag{3.12}$$

in which $\mu > 0$ is a free parameter. This problem is just an approximation, not equivalent to the original one. However, as $\mu \to 0$, (3.12) becomes an even-better approximation to (3.11).

This section is devoted to the construction of such a barrier function which is not only continuous but also convex. The correction of that construction bases mainly on strong separation of convex sets.

From now, for convenience we denote $\partial C = \overline{C} \backslash \text{relint}(C)$. The following proposition gives us an important example for a concave function.

Proposition 3.7. *Let C be a nonempty open convex set in $\mathbb{R}^n$. Then the function*

$$d_C(\mathbf{x}) := \min\{\|\mathbf{x} - \mathbf{y}\| \mid \mathbf{y} \in \partial C\}$$

is a concave function on C that vanishes on ∂C.

Proof. The vanishment of $d_C(\mathbf{x})$ on ∂C is obvious from definition of $d_C(\mathbf{x})$, so it is left to show that $d_C(\mathbf{x})$ is a concave function on C. Let $\mathbf{x}$ be an arbitrary point in C. Let H^+ be an open half-space containing C. Then H^+ admits the following representation

$$H^+ = \{\mathbf{u} \in \mathbb{R}^n \mid \langle \mathbf{a}, \mathbf{u} \rangle > \xi\}$$

for some $\mathbf{a} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ and $\xi \in \mathbb{R}$. It follows that

$$\bar{H}^- := \mathbb{R}^n \backslash H^+ = \{\mathbf{u} \in \mathbb{R}^n \mid \langle \mathbf{a}, \mathbf{u} \rangle \leq \xi\}$$

is a closed half-space that does not contain any point of C. Furthermore, H^+ and $\bar{H}^-$ have common boundary

$$\partial H^+ = \partial \bar{H}^- = \{\mathbf{u} \in \mathbb{R}^n \mid \langle \mathbf{a}, \mathbf{u} \rangle = \xi\} =: H.$$

Let $\mathbf{z}$ be an arbitrary point in H. Since $H \subset \bar{H}^-$, we have $\mathbf{z} \in \bar{H}^-$. Since $\bar{H}^-$ does not contain any point of C, we have $\mathbf{z} \notin C$. This, together with the fact that $\mathbf{x} \in C$ by our choice, implies that the line segment

$$[\mathbf{x}, \mathbf{z}] := \{\lambda \mathbf{x} + (1 - \lambda)\mathbf{z} \mid \lambda \in [0, 1]\}$$

intersects the boundary ∂C of C at some point $\mathbf{y_z}$. Since $\mathbf{y_z} \in [\mathbf{x}, \mathbf{z}]$, we have

$$\|\mathbf{x} - \mathbf{y_z}\| \leq \|\mathbf{x} - \mathbf{z}\|. \tag{3.13}$$

This holds for arbitrary choice of $\mathbf{z}$ in H, hence it also holds for

$$\mathbf{z} = \mathbf{z}^* := \operatorname{argmin}_{\mathbf{v} \in H} \|\mathbf{x} - \mathbf{v}\|.$$

We come up with the following inequalities

$$\begin{aligned}
d_C(\mathbf{x}) &= \min\{\|\mathbf{x} - \mathbf{y}\| \mid \mathbf{y} \in \partial C\} \\
&\leq \|\mathbf{x} - \mathbf{y}_{\mathbf{z}^*}\| && (\text{since } \mathbf{y}_{\mathbf{z}^*} \in \partial C) \\
&\leq \|\mathbf{x} - \mathbf{z}^*\| && (\text{by } (3.13)) \\
&= d_H(\mathbf{x}) && (\text{by definition of } \mathbf{z}^*) \\
&= d_{H^+}(\mathbf{x}) && (\text{since } H = \partial H^+).
\end{aligned}$$

Since the inequality $d_C(\mathbf{x}) \le d_{H^+}(\mathbf{x})$ holds for arbitrary open half-space H^+ containing C, we obtain

$$d_C(\mathbf{x}) \le \inf_{H^+ \in \mathcal{H}} d_{H^+}(\mathbf{x}), \tag{3.14}$$

in which $\mathcal{H}$ is the set of all open half-space containing C.

Now, let $\mathbf{y_x}$ be a point in ∂C that is closest to $\mathbf{x}$, then $d_C(\mathbf{x}) = \|\mathbf{x} - \mathbf{y_x}\|$. Since $\mathbf{y_x} \in \partial C$, thanks to Theorem 2.9 there exists a support hyperplane to C at $\mathbf{y_x}$. Let us denote that support hyperplane by $H_\mathbf{x}$. As a hyperplane, $H_\mathbf{x}$ admits the following representation

$$H_\mathbf{x} = \{\mathbf{u} \in \mathbb{R}^n \mid \langle \mathbf{b}, \mathbf{u} \rangle = \beta\}$$

for some $\mathbf{b} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ and $\beta \in \mathbb{R}$. Since $H_\mathbf{x}$ is the support hyperplane to the open convex set C at $\mathbf{y_x}$, the open half-space $H_\mathbf{x}^+ = \{\mathbf{u} \in \mathbb{R}^n \mid \langle \mathbf{b}, \mathbf{u} \rangle > \beta\}$ contains C (so $H_\mathbf{x}^+ \in \mathcal{H}$). Note that $\mathbf{y_x}$ belongs to not only ∂C but also $H_\mathbf{x} = \partial H_\mathbf{x}^+$, and that $d_C(\mathbf{x}) = \|\mathbf{x} - \mathbf{y_x}\|$, we have

$$d_C(\mathbf{x}) = d_{H_\mathbf{x}^+}(\mathbf{x}). \tag{3.15}$$

Since $H_\mathbf{x}^+ \in \mathcal{H}$, it follows from (3.14) and (3.15) that

$$d_C(\mathbf{x}) = \inf_{H^+ \in \mathcal{H}} d_{H^+}(\mathbf{x}). \tag{3.16}$$

Note that this equality holds for arbitrarily chosen point $\mathbf{x} \in C$. Let us again consider an arbitrary open half-space $H^+ = \{\mathbf{u} \in \mathbb{R}^n \mid \langle \mathbf{a}, \mathbf{u} \rangle > \xi\} \in \mathcal{H}$, which contains C by definition of $\mathcal{H}$. On one hand, we have $\mathbf{x} \in H^+$ (since $\mathbf{x} \in C$ and $C \subset H^+$), and hence $\langle \mathbf{a}, \mathbf{x} \rangle > \xi$. On the other hand, we have

$$\begin{aligned} d_{H^+}(\mathbf{x}) &= \frac{|\langle \mathbf{a}, \mathbf{x} \rangle - \xi|}{\|\mathbf{a}\|} \qquad \text{(by Lemma 1.18)} \\ &= \frac{\langle \mathbf{a}, \mathbf{x} \rangle - \xi}{\|\mathbf{a}\|} \qquad \text{(since } \langle \mathbf{a}, \mathbf{u} \rangle > \xi\text{)}. \end{aligned}$$

So the right hand side of (3.16) is the infimum of linear functions. Therefore, by Proposition 1.20, $d_C(\mathbf{x})$ is a concave function on C. □

We are now ready for the main result of this section.

Theorem 3.8. *Let C be a nonempty open convex set in $\mathbb{R}^n$. Then the function $b(\mathbf{x}) := -\ln d_C(\mathbf{x})$ is a convex barrier function on C.*

Proof. We need to show the followings.

(i) $b(\mathbf{x})$ is continuous on C.

(ii) $b(\mathbf{x})$ is convex on C.

(iii) $b(\mathbf{x}) \to +\infty$ as $\mathbf{x} \to \partial C$.

For the proof of (i), we first note from Proposition 3.7 that $d_C(\mathbf{x})$ is a concave function on C. Hence, thanks to Proposition 1.22, $d_C(\mathbf{x})$ is continuous on C. Since $b(\mathbf{x})$ is the composition of the continuous functions $-\ln(\cdot)$ on $\mathbb{R}_+$ and $d_C(\mathbf{x})$ on C, it is continuous.

For the proof of (ii), we note that $\ln(\cdot)$ is a concave non-decreasing function on $\mathbb{R}_+$. By Proposition 1.21, the composition $\ln(d_C(\mathbf{x}))$ of $\ln(\cdot)$ with the concave function $d_C(\mathbf{x})$ on C is also concave on C. As a consequence, $b(\mathbf{x}) = -\ln(d_C(\mathbf{x}))$ is convex on C.

By definition of $d_C(\mathbf{x})$, we have $d_C(\mathbf{x}) = 0$ for $\mathbf{x} \in \partial C$. Thanks to the continuity of $d_C(\mathbf{x})$ on C, $\mathbf{x} \to \partial C$ we have $d_C(\mathbf{x}) \to 0$ and hence $\ln(d_C(\mathbf{x})) \to -\infty$. Therefore $b(\mathbf{x}) = -\ln(d_C(\mathbf{x})) \to +\infty$ as $\mathbf{x} \to \partial C$, *i.e.*, $b(\mathbf{x})$ is a barrier function on C. This completes the proof of (iii). □

3.4 Hahn-Banach theorem

In this section, we present a proof of Hahn-Banach theorem - a cornerstone result in functional analysis - by using theorem on proper separation between a convex set and an affine set in general vector spaces.

Firstly, let us recall the statement of Hahn-Banach theorem. For that, let E be a general vector space (without any equipped topology). By *functional* on E we mean a scalar-valued function defined on E, *i.e.*, a function $p : E \to \mathbb{R}$. Such functional p is said to be *linear* if

(i) $p(\mathbf{u} + \mathbf{v}) = p(\mathbf{u}) + p(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in E$, and

(ii) $p(\lambda \mathbf{v}) = \lambda p(\mathbf{v})$ for all $\lambda \in \mathbb{R}, \mathbf{v} \in E$.

Such functional p is said to be *sub-additive* if $p(\mathbf{u}+\mathbf{v}) \leq p(\mathbf{u})+p(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in E$. It is said to be *positive homogeneous* if $p(\lambda \mathbf{v}) = \lambda p(\mathbf{v})$ for all $\mathbf{v} \in E$ and for all $\lambda \geq 0$. We say that the functional p is *sublinear* if it is sub-additive and positive homogeneous. It is not hard to see that a sublinear functional is convex.

Theorem 3.9. (Hahn-Banach theorem). *Let E be a vector space, F a linear subspace of E, and p a sublinear functional on E. If $f : F \to \mathbb{R}$ is a linear functional satisfying $f(\mathbf{v}) \leq p(\mathbf{v})$ for all $\mathbf{v} \in F$, then there exists a linear functional g on E such that $g(\mathbf{v}) = f(\mathbf{v})$ for all $\mathbf{v} \in F$ and $g(\mathbf{v}) \leq p(\mathbf{v})$ for all $\mathbf{v} \in E$.*

Roughly speaking, the theorem states that a linear functional on a linear subspace and majorized by the sublinear functional p can be extended to a linear functional on the whole space which is still majorized by p.

Proof. Let

$$A := \{(\mathbf{x}, \xi) \in E \times \mathbb{R} \mid p(\mathbf{x}) < \xi\},$$
$$B := \{(\mathbf{x}, \xi) \in F \times \mathbb{R} \mid f(\mathbf{x}) = \xi\}.$$

Since p is a sublinear functional on E, it is convex. Hence A is also convex. Since f is a linear functional, B is a linear subspace of $E \times R$. Hence B is an affine set. Furthermore, we have

$$\text{rai}(A) = \{(\mathbf{x}, \xi) \in E \times \mathbb{R} \mid p(\mathbf{x}) < \xi\}, \tag{3.17}$$

which is also a convex set. We claim that $\text{rai}(A) \cap B = \varnothing$. Indeed, assume the contrary that there exists $(\mathbf{y}, \xi) \in \text{rai}(A) \cap B \subseteq F \times \mathbb{R}$. It follows from (3.17) that $p(\mathbf{y}) < \xi$. However, since $(\mathbf{y}, \xi) \in B$, we have $f(\mathbf{y}) = \xi$, contradicting the fact that $p(\mathbf{y}) < \xi$ we have shown. This contradiction ensures our claim.

By applying Theorem 2.22 to the convex set $\text{rai}(A)$ and the affine set B in the vector space $E \times \mathbb{R}$, we have a hyperplane H containing B and disjoint from $\text{rai}(A)$. Such hyperplane H admits the following form

$$H = \{(\mathbf{x}, \xi) \in E \times \mathbb{R} \mid h(\mathbf{x}) + m\xi = 0\}$$

which corresponds to a nonzero linear functional (h, m) on $E \times \mathbb{R}$ defined by

$$(h, m)(\mathbf{x}, \xi) := h(\mathbf{x}) + m\xi.$$

Suppose without loss of generality that $A \subseteq \bar{H}^-$. Then it holds that

$$h(\mathbf{x}) + mf(\mathbf{x}) = 0 \quad \forall \mathbf{x} \in F \tag{3.18}$$

and

$$h(\mathbf{x}) + m\xi < 0 \quad \forall (\mathbf{x}, \xi) \in E \times \mathbb{R} : p(\mathbf{x}) < \xi. \tag{3.19}$$

If $m > 0$, we can fix $\mathbf{x} \in E$ and let $\xi \to +\infty$ to obtain a contradiction with (3.19). If $m = 0$, then it follows from (3.19) that $h(\mathbf{x}) < 0$ for all $\mathbf{x} \in E$, while (3.18) gives us $h(\mathbf{x}) = 0$ for $\mathbf{x} \in F \subset E$. This contradiction means that m must be nonzero, and hence $m < 0$.

We may assume that $m = -1$ due to linearity of (h, m). Then we obtain from (3.18) that $h(\mathbf{x}) = f(\mathbf{x})$ for all $\mathbf{x} \in F$, while from (3.19) we have

$$h(\mathbf{x}) < \xi \quad \forall (\mathbf{x}, \xi) \in E \times \mathbb{R} : p(\mathbf{x}) < \xi. \tag{3.20}$$

Let $g : E \to \mathbb{R}$ be define by $g(\mathbf{x}) := h(\mathbf{x})$. Obviously, $g(\mathbf{x})$ is a linear functional and $g(\mathbf{x}) = f(\mathbf{x})$ for all $\mathbf{x} \in F$. To complete the proof, it is left to show that $g(\mathbf{x}) \le p(\mathbf{x})$ for all $\mathbf{x} \in E$. Indeed, assume the contrary that we can pick $\mathbf{x}^0 \in E$ such that $g(\mathbf{x}^0) > p(\mathbf{x}^0)$, or equivalently $h(\mathbf{x}^0) > p(\mathbf{x}^0)$. By letting $\mathbf{x} = \mathbf{x}^0, \xi = h(\mathbf{x}^0)$, we derive $h(\mathbf{x}^0) < h(\mathbf{x}^0)$, which is a contradiction. $\square$

Conclusions

In this thesis, we studied some types of separation between two convex sets in real vector spaces, and presented their applications in some related problems in convex analysis as well as functional analysis. Namely, in Chapter 1 we recalled some related preliminaries concerning affine sets, convex sets, conic sets, projection on convex sets, convex functions, and algebraic interior as well as algebraic closure of convex sets. In Chapter 2, we first recalled the definition of general separation, strict separation, strong separation, and proper separation between two convex sets. These separation types are considered in the setting of finite dimensional Euclidean vector spaces as well as in the setting of general vector spaces without any equipped topology. In the former setting, we presented some theorems about conditions for the general separation, strong separation, and proper separation involving two convex sets. In the latter setting, we presented some theorems about conditions for general separation and proper separation between two convex sets, and a particular case of proper separation between a convex set and an affine set.

In Chapter 3 we have shown that a number of results in convex analysis and functional analysis can be obtained from the separation theorems. Namely, we gave detail arguments to derive the homogeneous Farkas lemma from the strong separation theorem involving two convex sets in finite dimensional Euclidean vector spaces. It can be also derived from the theorem that a closed convex cone in finite dimensional Euclidean vector spaces is dual to its dual cone. We also presented the use of the general separation theorem in constructing a barrier convex function for the feasible set of a convex optimization problem. Also in this chapter we have presented how to derive a well-known result in functional analysis - the Hahn-Banach theorem - from the theorem on the proper separation of convex sets in general vector spaces.

Bibliography

[1] O. Güler. *Foundation of Optimization*. Springer Science+Business Media, 2010.

[2] H. Tuy. *Convex Analysis and Global Optimization (second edition).* Springer International Publishing, 2016.

[3] S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[4] B. S. Mordukhovich and N. M. Nam. *Convex Analysis and Beyond. Volume I: Basic Theory*. Springer Natural Switzerland, 2022.

[5] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications.* Society for Industrial and Applied Mathematics, 2001.