M. Walk

Theory of Duality in Mathematical Programming

Mathematische Lehrbücher und Monographien
Herausgegeben von der Akademie der Wissenschaften der DDR
Karl-Weierstraß-Institut für Mathematik

II. Abteilung
Mathematische Monographien
Band 72

Theory of Duality in Mathematical Programming

von M. Walk

Theory of Duality in Mathematical Programming

by Manfred Walk

with 9 figures

Akademie-Verlag Berlin
1989

Verfasser:

Prof. Dr. Manfred Walk

Friedrich-Schiller-Universität Jena
Sektion Mathematik

ISBN 3-05-500404-3 Akademie-Verlag Berlin
ISBN 3-211-82057-4 Springer-Verlag Wien—New York
ISBN 0-387-82057-4 Springer-Verlag New York—Wien

ISSN 0076-5430

Erschienen im Akademie-Verlag Berlin, Leipziger Str. 3—4, Berlin, DDR-1086

Lizenznummer: 202 · 100/403/88
Printed in the German Democratic Republic
Gesamtherstellung: VEB Druckhaus „Maxim Gorki", Altenburg, DDR-7400
Lektor: Dr. Reinhard Höppner
LSV 1085
Bestellnummer: 763 741 2 (9064)
03800

Preface

The theory of mathematical programming has developed considerably during the last years due to the stimulus given by an increasing demand from management and technology and it is therefore understandable that courses on mathematical programming now belong to standard teaching programs of Universities and Institutes.

Our aim in writing this book is to introduce students of mathematics, science, economics and management to the qualitative theory of mathematical programming in vector spaces. Prerequisite for the study of this book is a basic knowledge of analysis and linear algebra. We also apply some elementary ideas of functional analysis for a more rigorous construction of proofs and for some generalizations of the finite dimensional theory enclosing Banach-spaces. The problems are presented in such a way that the theory can be generalized to apply to topological vector spaces.

Many different concrete programming problems possess one and the same theoretical basis, which we can formulate as a principle of duality. Many authors have generalized the well known relations of duality of linear programming for problems of non-linear programming. In this connection, the Fenchel theory of conjugate functions, and the relations of duality following on from this theory, plays an important role.

The main theme of this book is to represent such relations of duality and to consider a general theoretical basis for different special programming problems. In so doing, we do not claim to cover all programming problems but rather discuss some of the main ones by concentrating our enquiries on specially chosen Lagrangian forms. For linear programming we extend our investigations by constructing the theory in Banach spaces. In infinite dimensional Banach spaces we give examples to demonstrate the difference between finite dimensional and infinite dimensional theory and give a generalization of the transportation- and potential problem in finite directed graphs. We are not concerned here with constructions of algorithms for calculation of solutions.

In the first chapter the reader will find a collection of relations between convex sets, hyperplanes and extremal points which are required in the following chapters. Furthermore, we construct relations between convex functions and convex sets and describe the Fenchel theory of conjugate functions. In the second chapter, we formulate relations of duality and prove equivalence theorems with respect to saddle-point, duality and minimax theorems. Special structures of equivalence statements are very useful for well known linear programming problems to be embedded in. The third chapter deals with the embedding of different programming problems in connection with specially chosen Lagrangian forms and in addition to equivalence theorems, the chapter also contains existence theorems.

The basic conception of this book is an extended version of a course which the author gave at the Rangoon Arts and Science University during 1974/1975. The main sources were the University text the author prepared based on this course with the assistance of Professor Chit Swe.

M. Walk

Contents

1. Convex sets and convex functions

1.1. Convex sets

In the theory of optimization, the idea of convexity plays a central role. We shall begin our study of this concept with convex sets and their properties. In this monograph, we would restrict our discussions to real finite dimensional vector spaces, even though it is possible to extend most of the ideas and theorems to an abstract topological vector space. Inquiries of the following kind, one may find in all standard books on mathematical programming, especially works by VOGEL [32], KARLIN [20], BERGE and GHOUILA-HOURI [4], and KREKO [23].

Let $\mathbb{R}^n$ be a vector space of dimension n.

Definition 1.1.1. A set $C \subseteqq \mathbb{R}^n$ is *convex*, if for any two elements $u, v \in C$ and any real number λ, $0 \leqq \lambda \leqq 1$, we have

$$x = \lambda u + (1 - \lambda)\, v \in C. \tag{1.1.1}$$

An element $x = \sum_{i=1}^{r} \lambda_i x^i$ is a *convex combination* of vectors $x^1, x^2, \ldots, x^r \in \mathbb{R}^n$ if the real numbers $\lambda_1, \lambda_2, \ldots, \lambda_r$ are nonnegative and $\sum_{i=1}^{r} \lambda_i = 1$.

Theorem 1.1.1. *Let $C \subseteqq \mathbb{R}^n$ be a convex set. Every convex combination of vectors of C is an element of C.*

Proof: For $r = 1$, the assertion of the theorem is trivial.
Assume that the theorem is true for the value $r - 1$ with $r > 1$. Now consider vectors $x^1, x^2, \ldots, x^r \in C$ and nonnegative numbers $\lambda_1, \lambda_2, \ldots, \lambda_r$,where $\sum_{i=1}^{r} \lambda_i = 1$. If $\sum_{i=1}^{r-1} \lambda_i = 0$, then

$$x = \sum_{i=1}^{r} \lambda_i x^i = x^r$$

which is an element of C. If $\lambda = \sum_{i=1}^{r-1} \lambda_i \neq 0$ then we have

$$x = \sum_{i=1}^{r} \lambda_i x^i = \lambda \sum_{i=1}^{r-1} \tilde{\lambda}_i x^i + \lambda_r x^r$$

with $\tilde{\lambda}_i = \dfrac{\lambda_i}{\lambda}$, $i = 1, 2, \ldots, r - 1$. Obviously $\tilde{\lambda}_i \geqq 0$ and $\sum_{i=1}^{r-1} \tilde{\lambda}_i = 1$. By induction hypothesis $\tilde{x} = \sum_{i=1}^{r-1} \tilde{\lambda}_i x^i \in C$. Moreover, $\sum_{i=1}^{r} \lambda_i = 1$ implies $0 < \lambda \leqq 1$ and $\lambda_r = 1 - \lambda$. Thus, by Definition 1.1.1 $x = \lambda \tilde{x} + \lambda_r x^r \in C$. □

Theorem 1.1.2. *Let C_1 and C_2 be convex sets. If σ_1 and σ_2 are any two real numbers, then the set $C = \sigma_1 C_1 + \sigma_2 C_2 = \{x \in \mathbb{R}^n : x = \sigma_1 x^2 + \sigma_2 x^2,\ x^i \in C_i\}$ is also convex.*

Proof: Consider any two elements $u, v \in C$. Then there exist elements $u^1, v^1 \in C_1$ and corresponding elements $u^2, v^2 \in C_2$ such that

$$u = \sigma_1 u^1 + \sigma_2 u^2, \quad v = \sigma_1 v^1 + \sigma_2 v^2.$$

Now, for any real number λ, $0 \leqq \lambda \leqq 1$,

$$\begin{aligned} x &= \lambda u + (1 - \lambda)\, v = \lambda(\sigma_1 u^1 + \sigma_2 u^2) + (1 - \lambda)\,(\sigma_1 v^1 + \sigma_2 v^2) \\ &= \sigma_1\big(\lambda u^1 + (1 - \lambda)\, v_1\big) + \sigma_2\big(\lambda u^2 + (1 - \lambda)\, v^2\big) = \sigma_1 x^1 + \sigma_2 x^2 \end{aligned}$$

where x^1, x^2 are elements of convex sets C_1 and C_2. Consequently, $x \in C$. $\square$

Regarding Theorem 1.1.2 we notice the following special cases:

(i) *If C_1 is a convex set and x^0 any element of $\mathbb{R}^n$, then the set*

$$C = C_1 + \{x^0\} = \{x \in \mathbb{R}^n : x = x^1 + x^0,\ x^1 \in C_1\}$$

is also convex.

(ii) *If C_1 is a convex set and σ any real number, then the set*

$$C = \sigma C_1 = \{x \in \mathbb{R}^n : x = \sigma x^1,\ x^1 \in C_1\}$$

is also convex.

(iii) *If C_1 is a convex set, then the set*

$$C = -C_1 = \{x \in \mathbb{R}^n : x = -x^1,\ x^1 \in C_1\}$$

is also convex.

Definition 1.1.2. Given $A \subseteqq \mathbb{R}^n$,

$$\operatorname{Conv}(A) = \left\{x \in \mathbb{R}^n : x = \sum_{i=1}^{r} \lambda_i x^i,\ x^i \in A,\ \lambda_i \geqq 0,\ \sum_{i=1}^{r} \lambda_i = 1,\ r = 1, 2, \ldots\right\}; \tag{1.1.2}$$

is called the *convex hull* of A.

Obviously Conv (A) is a closed set if the set A is finite. In case the set A is infinite, the convex hull Conv (A) is generally not closed. We may consider, for example,

$$A = \left\{\frac{1}{n} : n = 1, 2, \ldots\right\} \subseteqq \mathbb{R}.$$

One can easily find

$$\operatorname{Conv}(A) = \{x \in \mathbb{R} : 0 < x \leqq 1\}$$

which is not closed.

Theorem 1.1.3. *The convex hull* Conv (A) *of a set A is the smallest convex set, containing the set A.*

Proof: We shall first show that Conv (A) is convex. Consider $u, v \in$ Conv (A). Then u and v may be represented in the following forms:

$$u = \sum_{i=1}^{r} \varrho_i u^i, \quad u^i \in A, \quad \varrho_i \geqq 0, \quad \sum_{i=1}^{r} \varrho_i = 1,$$

$$v = \sum_{i=1}^{s} \sigma_i v^i, \quad v^i \in A, \sigma_i \geqq 0, \quad \sum_{i=1}^{s} \sigma_i = 1.$$

Now, for any real number λ, $0 \leqq \lambda \leqq 1$,

$$\begin{aligned} x &= \lambda u + (1 - \lambda) v \\ &= \sum_{i=1}^{r} \lambda \varrho_i u^i + \sum_{i=1}^{s} (1 - \lambda) \sigma_i v^i = \sum_{k=1}^{r+s} \tau_k z^k \end{aligned}$$

where

$$\tau_k = \begin{cases} \lambda \varrho_k, & k = 1, \ldots, r, \\ (1 - \lambda) \sigma_{k-r}, & k = r + 1, r + 2, \ldots, r + s, \end{cases}$$

$$z^k = \begin{cases} u^k \quad, & k = 1, \ldots, r, \\ v^{k-r}, & k = r + 1, \ldots, r + s. \end{cases}$$

We can observe that $z^k \in A$, $\tau_k \geqq 0$, $k = 1, \ldots, r + s$ and

$$\sum_{k=1}^{r+s} \tau_k = \lambda \sum_{k=1}^{r} \varrho_k + (1 - \lambda) \sum_{k=r+1}^{r+s} \sigma_{k-r} = \lambda + (1 - \lambda) = 1.$$

Thus $x \in$ Conv (A) implying that Conv (A) is convex. To complete the proof, we still have to show that for every convex set K containing A, Conv $(A) \subseteqq K$. This is quite clear because K is convex and from $A \subseteqq K$, it follows by Theorem 1.1.1 that every convex combination of elements of A belongs to K. □

The converse of Theorem 1.1.3 now follows:

Theorem 1.1.4. *If L is the smallest convex set containing A, then*

$$L = \text{Conv}\,(A).$$

Proof: L has the following properties:

(i) $A \subseteqq L$.

(ii) $L \subseteqq K$ for every convex set K containing A.

Since Conv (A) is a convex set containing A, we have by (ii),

$$L \subseteqq \text{Conv}\,(A).$$

On the other hand, Theorem 1.1.3 together with (i) implies that

$$\text{Conv}\,(A) \subseteqq L.$$

Hence the proof of the theorem. □

Theorem 1.1.5. *The convex hull of a set A is the intersection of all convex sets which contain A:*

$$\operatorname{Conv}(A) = \bigcap_{A \subset K} K.$$

K being convex and containing A.

Proof. It is easily seen that the intersection of any collection of convex sets is convex. Thus, the set $\bigcap_{A \subset K} K$ is convex and it contains A. Furthermore, for any convex set K containing A, we have $\bigcap_{A \subset K} K$, i.e. $\bigcap_{A \subset K} K$ is the smallest convex set which contains A.

Now, by Theorem 1.1.4,

$$\bigcap_{A \subset K} K = \operatorname{Conv}(A). \ \square$$

The Theorems 1.1.1 to 1.1.5 are given only on the basis of the Definition 1.1.1 of convex sets. However, the proofs do not depend on the fact that we are restricting our considerations to finite dimensional Euclidean spaces. We should notice that if the discussions were in the more general framework of any linear space, the Theorems 1.1.1 to 1.1.5 are still true.

Besides convexity the following theorems are concerned with boundedness in finite dimensional spaces.

We will first deal with a theorem which says that in the case of an n-dimensional Euclidean space, the convex hull Conv (A) can be generated by all convex linear combinations of $n + 1$ or fewer elements of A.

Theorem 1.1.6. *If A is a subset of $\mathbb{R}^n$, then $x \in \operatorname{Conv}(A)$ can be represented by*

$$x = \sum_{i=1}^{r} \lambda_i x^i, \quad x^i \in A, \quad \lambda_i > 0, \quad \sum_{i=1}^{r} \lambda_i = 1, \quad r \leqq n + 1.$$

Proof: For any $x \in \operatorname{Conv}(A)$,

$$x = \sum_{i=1}^{r} \lambda_i x^i, \quad x^i \in A, \quad \lambda_i \geqq 0, \quad \sum_{i=1}^{r} \lambda_i = 1.$$

We can assume $\lambda_i > 0$. If there is one $\lambda_i = 0$, we will drop this term off. If $r \leqq n + 1$, then there is nothing to prove. Now, consider the case $r > n + 1$. In this case, we can construct a new representation for x in which at most $r - 1$ elements of A appear. Such a construction is possible, because in an n-dimensional vector space, $n + 1$ vectors are always linearly dependent.

Let us consider vectors $z^i = x^i - x^r$, $i = 1, 2, \ldots, r - 1$. Under the assumption $r > n + 1$, the number of vectors z^i is more than n and in consequence the vectors z^i are linearly dependent. Thus there exist real numbers σ_i, $i = 1, \ldots, r - 1$, which are not all zero such that

$$\sum_{i=1}^{r-1} \sigma_i z^i = 0.$$

Putting

$$\sigma_r = -\sum_{i=1}^{r-1} \sigma_i$$

we get

$$0 = \sum_{i=1}^{r-1} \sigma_i z^i = \sum_{i=1}^{r-1} \sigma_i x^i - x^r \sum_{i=1}^{r-1} \sigma_i = \sum_{i=1}^{r} \sigma_i x^i$$

and

$$\sum_{i=1}^{r} \sigma_i = 0 .$$

Now, for any $\varepsilon > 0$

$$x = \sum_{i=1}^{r} \lambda_i x^i = \sum_{i=1}^{r} (\lambda_i - \varepsilon\sigma_i)\, x^i . \tag{1.1.3}$$

Since at least one σ_i is positive, it is possible to calculate

$$\min \left\{ \frac{\lambda_i}{\sigma_i} \colon \sigma_i > 0 \right\} = \frac{\lambda_{i_0}}{\sigma_{i_0}} .$$

Here the index i_0 may not be uniquely defined. Now, if we put $\varepsilon = \dfrac{\lambda_{i_0}}{\sigma_{i_0}}$, then

$$\lambda_{i_0} - \varepsilon\sigma_{i_0} = 0 ,$$

$$\lambda_i - \varepsilon\sigma_i \neq 0 , \quad \text{for} \quad i \neq i_0$$

and

$$\sum_{r=1}^{r} (\lambda_\iota - \varepsilon\sigma_i) = 1 .$$

Now, from (1.1.3), for $\varepsilon = \dfrac{\lambda_{i_0}}{\sigma_{i_0}}$, we get the new representation

$$x = \sum_{\substack{i=1 \\ i \neq i_0}}^{r} (\lambda_i - \varepsilon\sigma_i)\, x^i .$$

In this form the element x will be represented by a convex linear combination of at most $r - 1$ elements of A. This construction, we can continue till we get the representation of x as a convex linear combination of at most $n + 1$ elements of A. □

Theorem 1.1.7. *Every set $\{x^k : k = 1, \ldots, r\}$ of r elements from $\mathbb{R}^n$ where $r \geqq n + 2$ can be divided into two parts such that the convex hulls of these parts have a nonempty intersection.*

Proof: The proof is based on the fact that there always exist a nontrivial solution for a homogeneous system of $n + 1$ linear algebraic equations in r unknowns, if $r \geqq n + 2$. Consider the set $\{x^k : k = 1, \ldots, r\}$, $r \geqq n + 2$. Then, there exists a nontrivial solution $\lambda_1, \lambda_2, \ldots, \lambda_r$ of the system.

$$\sum_{k=1}^{r} \lambda_k x^k = 0 , \quad \sum_{k=1}^{r} \lambda_k = 0 . \tag{1.1.4}$$

Define the two sets of integers

$$I = \{k \colon \lambda_k > 0\} , \quad J = \{1, 2, \ldots, r\} \setminus I .$$

One can easily see that I and J are nonempty and because of (1.1.4), we find

$$\sum_{k \in I} \lambda_k = -\sum_{k \in J} \lambda_k > 0 .$$

Furthermore, we get from (1.1.4) an element z in the form

$$z = \frac{\sum\limits_{k\in I} \lambda_k x^k}{\sum\limits_{k\in I} \lambda_k} = \frac{\sum\limits_{k\in J} -\lambda_k x^k}{\sum\limits_{k\in J} -\lambda_k}. \tag{1.1.5}$$

Clearly,

$$\sigma_k = \frac{\lambda_k}{\sum\limits_{k\in I} \lambda_k} > 0 \quad \text{if} \quad k \in I; \quad \sum_{k\in I} \sigma_k = 1,$$

$$\sigma_k = \frac{-\lambda_k}{\sum\limits_{k\in J} -\lambda_k} \geqq 0 \quad \text{if} \quad k \in J; \quad \sum_{k\in J} \sigma_k = 1.$$

The equation (1.1.5) means that the element z belongs to the convex hull of $\{x^k : k \in I\}$ as well as the convex hull of $\{x^k : k \in J\}$.

Theorem 1.1.8. (Theorem of Helly). *Let C_i, $i = 1, 2, \ldots, r$, $r \geqq n + 1$ be a finite class of convex sets in $\mathbb{R}^n$. If the intersection of any $n + 1$ of the sets $C_1, C_2, \ldots, C_r$ is nonempty, then the intersection of all r sets is nonempty.*

Proof: If $r = n + 1$, the assertion of the theorem is trivial.

Now assume that the theorem is true for $r - 1 \geqq n + 1$. This means if any $n + 1$ of the sets $C_1, \ldots, C_r$ have a nonempty intersection, the intersection of any $r - 1$ of them always has a nonempty intersection.

Now construct the set

$$\{x^k : k = 1, 2, \ldots, r\} \tag{1.1.6}$$

with elements $x^k \in \bigcap\limits_{\substack{i=1\\ i\neq k}}^{r} C_i$.

Let $\{x^k : k \in I\}$ and $\{x^j : j \in J\}$ be a partition of the set (1.1.6) as given in Theorem 1.1.7. Every element x^k with $k \in I$ belongs to all sets C_j, $j \in J$, because for any $k \in I$ and $j \in J$, we have $k \neq j$. Consequently, for the convex hull of the set $\{x^k : k \in I\}$ we have

$$\operatorname{Conv}(\{x^k : k \in I\}) \subseteqq \bigcap_{j\in J} C_j. \tag{1.1.7}$$

Similarly we find that every element x^j with $j \in J$ belongs to all sets C_k, $k \in I$, and we have

$$\operatorname{Conv}(\{x^j : j \in J\}) \subseteqq \bigcap_{k\in I} C_k. \tag{1.1.8}$$

Then it follows from Theorem 1.1.7 that

$$\operatorname{Conv}(\{x^k : k \in I\}) \cap \operatorname{Conv}(\{x^j : j \in J\}) \neq \emptyset.$$

Let

$$z \in \operatorname{Conv}(\{x^k : k \in I\}) \cap \operatorname{Conv}(\{x^j : j \in J\}).$$

Because of (1.1.7) and (1.1.8)

$$z \in \left(\bigcap_{k\in I} C_k\right) \cap \left(\bigcap_{j\in J} C_j\right) = \bigcap_{i=1}^{r} C_i$$

which proves the theorem. $\square$

At last we will show that the convex hull of a compact set is compact. For this reason we need the following lemma:

Lemma 1.1.1. *If the sets $A \subseteqq \mathbb{R}^n$ and $B \subseteqq \mathbb{R}^m$ are compact, then the product*

$$A \times B = \{(x, y)\} : x \in A, y \in B\}$$

is compact.

Proof: In a finite dimensional Euclidean space, a set is compact if and only if it is closed and bounded. For further discussions in the monograph we use the Euclidean norm. The following discussion shows that $A \times B$ is closed and bounded.

A and B being compact there exist numbers α and β such that

$$\sum_{i=1}^{n} \xi_i^2 = \|x\|^2 \leqq \alpha \quad \text{for any} \quad x \in A,$$

$$\sum_{i=1}^{m} \eta_i^2 = \|y\|^2 \leqq \beta \quad \text{for any} \quad y \in B.$$

If we recognize that a norm for elements $(x, y) \in A \times B$ is

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

we get for any element $(x, y) \in A \times B$

$$\|(x, y)\|^2 = \|x\|^2 + \|y\|^2 \leqq \alpha + \beta.$$

Thus $A \times B$ is bounded.

The product $A \times B$ is also closed. Suppose $(x^k, y^k) \in A \times B$ is a converging sequence with a limit (x, y) i.e.

$$\lim_{k\to\infty} \|(x^k, y^k) - (x, y)\| = \lim_{k\to\infty} (\|x^k - x\|^2 + \|y^k - y\|^2)^{\frac{1}{2}} = 0.$$

This means the sequences $\{x^k\}$ and $\{y^k\}$ converge respectively to the limits x and y. Since A and B are closed $(x, y) \in A \times B$. □

The Lemma 1.1.1 and the fact that a continuous image of a compact set is compact lead us to the following theorem:

Theorem 1.1.9. *Let $A \subseteqq \mathbb{R}^n$ be compact. Then the convex hull* Conv (A) *is compact.*

Proof: We can easily observe that

$$S = \left\{x \in \mathbb{R}^{n+1} : \sum_{i=1}^{n+1} \xi_i = 1, \xi_i \geqq 0\right\}$$

is compact. By Lemma 1.1.1, the set

$$Y = S \times \underbrace{A \times \cdots \times A}_{n+1}$$

is compact.

We assign to each point $y \in Y$, by a continuous mapping $\psi \colon Y \to \mathbb{R}^n$, the value

$$\psi(y) = \psi(x, a^1, \ldots, a^{n+1}) = \sum_{i=1}^{n+1} \xi_i a^i \in \mathbb{R}^n.$$

Then the image of the continuous mapping,

$$\{\psi(Y)\} = \left\{z \in \mathbb{R}^n : z = \sum_{i=1}^{n+1} \xi_i a^i,\, a^i \in A,\, \xi_i \geqq 0,\, \sum_{i=1}^{n+1} \xi_i = 1\right\}$$

is compact. By Theorem 1.1.6, we have $\{\psi(Y)\} = \operatorname{Conv}(A)$. Consequently $\operatorname{Conv}(A)$ is compact. □

1.2. Hyperplanes and separation theorems

We will now present some principal relations between hyperplanes and convex sets. In the theory of duality of convex programming, such relations in the form of separation theorems play a central role especially in the proof of existence of optimal solutions. Of course, we apply the separation theorems also to an immediate characterization of convex sets. It is in this context, we will find them in the following discussions.

Definition 1.2.1. Let γ be a real number and $a \in \mathbb{R}^n$ any non zero fixed vector. The set $H \subseteqq \mathbb{R}^n$ given by

$$H = \{x \in \mathbb{R}^n : \langle a, x\rangle - \gamma = 0\}$$

is a *hyperplane* in $\mathbb{R}^n$.

Here $\langle a, x\rangle = \sum_{i=1}^{n} \alpha_i \xi_i$ denotes the scalar product of vectors $a = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ and $x = (\xi_1, \xi_2, \ldots, \xi_n)$. Further we notice that the Euclidean norm of vector x may be expressed in the form:

$$\|x\| = \left(\sum_{i=1}^{n} \xi_i^2\right)^{\frac{1}{2}} = (\langle x, x\rangle)^{\frac{1}{2}}.$$

For $\gamma = 0$, the set H will form a $(n-1)$-dimensional linear subspace of $\mathbb{R}^n$; otherwise, we get a linear manifold which arises by parallel translation of the subspace.

Geometrically in $\mathbb{R}^3$ a linear subspace $\langle a, x\rangle = 0$ represents a plane through the origin whereas the linear manifold $\langle a, x\rangle - \gamma = 0$ still represents a plane obtained by parallel displacement.

To prove the separation theorems, we need the following lemma:

Lemma 1.2.1. *If a closed convex set C is not empty and zero is not in C, then there exist exactly one point $x^0 \in C$ such that*

$$\|x^0\| = \min \{\|x\| : x \in C\}.$$

Proof: The proof of this lemma is based on the following well known facts:

(i) A continuous function on a compact set always attains its minimum.
(ii) Compactness in $\mathbb{R}^n$ means being closed and bounded.

Under the assumption that C is convex, the minimal element x^0 is uniquely defined.

Let $S = \{x \in \mathbb{R}^n : \|x\| \leqq \varrho\}$ be a closed sphere with centre at the origin, radius $\varrho > 0$ being chosen sufficiently great so that the set $A = C \cap S$ is not empty. Since A is bounded and closed, A is compact. Consequently, for the continuous function $\varphi(x) = \|x\|$, a vector $x^0 \in A$ exists such that

$$\varphi(x^0) = \|x^0\| = \min \{\varphi(x) : x \in A\}.$$

Because zero does not belong to C, it cannot be an element of A and thus $\|x^0\| > 0$. Furthermore, $\|x^0\| < \varrho$ and $\|x\| > \varrho$ for all $x \in C$, which are not in S.

Hence a fortiori

$$\|x^0\| \leqq \|x\|$$

for any $x \in C$. Consequently,

$$\varphi(x^0) = \|x^0\| = \min \{\varphi(x) : x \in C\}. \tag{1.2.1}$$

To complete the proof, we have to show that x^0 is uniquely defined. Let us assume that this is not so, and $x^1 \in C$ is another element such that

$$\|x^0\| = \|x^1\| = \min \{\varphi(x) : x \in C\}$$

and $x^0 \neq x^1$. Convexity of C implies that $x^2 = \frac{1}{2}(x^0 + x^1) \in C$. Finally we get

$$\|x^2\|^2 = \frac{1}{4}\|x^0 + x^1\|^2 < \frac{1}{4}(\|x^0 + x^1\|^2 + \|x^0 - x^1\|^2)$$

$$= \frac{1}{2}(\|x^0\|^2 + \|x^1\|^2) = \|x^0\|^2.$$

This means $\|x^2\| < \|x^0\|$ in contradiction to (1.2.1). □

From Lemma 1.2.1, we get a first separation theorem which says that a closed convex set not containing the origin can be strictly separated from the origin by a hyperplane.

Theorem 1.2.1. *If the convex set C is not empty and if the closure $\bar{C}$ of C does not contain the origin, then there exists a vector $a \neq 0$, such that*

$$\langle a, x\rangle \geqq \|a\|^2 > 0 \tag{1.2.2}$$

for any $x \in \bar{C}$.

Proof: We notice that the vector a which satisfies the assertion of the theorem can be chosen from $\bar{C}$. By Lemma 1.2.1, there exists a uniquely defined element $x^0 \in \bar{C}$ such that

$$\|x^0\| = \min \{\|x\| : x \in \bar{C}\}.$$

Choose $a = x^0$ and consider any $x \in \bar{C}$ and $0 < \lambda \leqq 1$. Because $\bar{C}$ is convex,

$$\lambda x + (1 - \lambda)\, x^0 = x^0 + \lambda(x - x^0) \in \bar{C}.$$

x^0 being a minimal element,

$$\|x^0\|^2 \leqq \|x^0 + \lambda(x - x_0)\|^2 = \|x^0\|^2 + 2\lambda\langle x^0, x - x^0\rangle + \lambda^2 \|x - x^0\|^2.$$

Dividing by 2λ,

$$\langle x^0, x\rangle = \langle a, x\rangle \geqq \|a\|^2 - \frac{\lambda}{2}\|x - a\|^2.$$

Now, by letting λ converge to zero, we get the relation $\langle a, x\rangle \geqq \|a\|^2 > 0$ for any $x \in \bar{C}$. □

Theorem 1.2.1 can be generalized in a certain sense. Instead of a strict separation of a convex set from the origin, we consider two nonempty convex sets which are separated from each other by a positive distance and show that these two convex sets are separated strictly by a suitable hyperplane.

Definition 1.2.2. Let C_1 and C_2 be two nonempty convex sets. Then

$$\delta(C_1, C_2) = \inf \{\|x - y\|; x \in C_1, y \in C_2\}$$

denotes the *distance* between C_1 and C_2.

Theorem 1.2.2. *If C_1 and C_2 are two nonempty convex sets and the distance $\delta(C_1, C_2)$ is positive, then there exists a hyperplane*

$$H = \{x \in \mathbb{R}^n : \langle a, x\rangle - \gamma = 0\}$$

such that

$$\begin{aligned} &\langle a, x\rangle - \gamma > 0 \quad \text{for any } x \in C_1, \\ &\langle a, x\rangle - \gamma < 0 \quad \text{for any } x \in C_2. \end{aligned} \tag{1.2.3}$$

Proof: $\delta(C_1, C_2) > 0$ implies that the set $C = C_1 - C_2$ has a positive distance from zero i.e. the closure $\overline{C}$ does not contain zero. Since $C = C_1 - C_2$ is also convex, by Theorem 1.2.1, it follows that there exists a vector $a \in \mathbb{R}^n$, $a \neq 0$, such that for any $z \in C$

$$\langle a, z\rangle \geqq \|a\|^2 > 0.$$

Here $z = x - y$, $x \in C_1$, $y \in C_2$ and we get

$$\langle a, x\rangle \geqq \langle a, y\rangle + \|a\|^2 \tag{1.2.4}$$

for any $x \in C_1$ and $y \in C_2$. Thus we have

$$\inf \{\langle a, x\rangle : x \in C_1\} \geqq \sup \{\langle a, y\rangle : y \in C_2\} + \|a\|^2$$

i.e.

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

Let γ be such that

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

Then, we can see that

$$H = \{x \in \mathbb{R}^n : \langle a, x\rangle - \gamma = 0\}$$

is the required hyperplane, satisfying the conditions

$$\begin{aligned} &\langle a, x\rangle - \gamma > 0 \quad \text{for any } x \in C_1, \\ &\langle a, y\rangle - \gamma < 0 \quad \text{for any } y \in C_2. \ \square \end{aligned}$$

In fact existence of separating hyperplanes is assured under more stringent conditions than those given in Theorems 1.2.1 and 1.2.2. A convex set which contains the origin on the boundary or two convex sets having a zero distance between them and having at the most boundary points in common can be separated by a suitable hyperplane. But in such cases, we cannot expect a strict separation but the separating hyperplane contains the origin in the former case and boundary points of the two convex sets in the latter.

Theorem 1.2.3. *Let C be a nonempty convex set with zero on the boundary of C. Then there exists a vector $a \in \mathbb{R}^n$, $a \neq 0$, such that for any $x \in \overline{C}$*

$$\langle a, x\rangle \geqq 0 \tag{1.2.5}$$

[Notice that zero may not be a point of C, if C is an open set.]