Mathematical Programming

(Linear Optimization and Extensions)

Dr. Tim Hoheisel

McGill University
Department of Mathematics and Statistics
Burnside Hall, Room 1114
805 Sherbrooke Street West
Montréal Quebec H3A0B9

e-mail: tim.hoheisel@mcgill.ca

Lecture Notes, Fall 2017

Last updated: September 11, 2017

"Since the building of the universe is perfect
and is created by the wisdom creator, nothing arises in the universe
in which one cannot see the sense of some maximum or minimum. "

(L. Euler)

Contents

Introduction

1 Preliminaries

1.1 Review of linear algebra

1.1.1 The vector space $\mathbb{R}^n$

Recall that $\mathbb{R}^n$, the set of all n-tuples $(x_i) := (x_i)_{i=1,\dots,n}$ with real entries $x_i \in \mathbb{R}$, is a real vector space with the *linear operations*

$$x + y := (x_i) + (y_i) = (x_i + y_i) \quad \text{and} \quad \lambda x := \lambda(x_i) = (\lambda x_i) \quad (x, y \in \mathbb{R}^n,\ \lambda \in \mathbb{R}).$$

We think of the elements of $\mathbb{R}^n$ as column vectors. A subset $U \subset \mathbb{R}^n$ is a *subspace* if and only if

$$0 \in U \quad \text{and} \quad \lambda x + \mu y \in U \ (x, y \in U,\ \lambda, \mu \in \mathbb{R}).$$

A mapping $F : \mathbb{R}^n \to \mathbb{R}^m$ is said to be *linear* if it is interchangeable with the linear operations, i.e.

$$F(\lambda x + \mu y) = \lambda F(x) + \mu F(y) \quad (\lambda, \mu \in \mathbb{R},\ x, y \in \mathbb{R}^n).$$

We call the vectors $x_1, \dots, x_p \in \mathbb{R}^n$ *linearly independent* if the (linear) equation

$$0 \overset{!}{=} \sum_{i=1}^{p} \lambda_i x_i \ (= \lambda_1 x_i + \cdots + \lambda_p x_p)$$

only admits the trivial solution $\lambda_1 = \cdots = \lambda_p = 0$.

The *span* or *linear hull of* of $X \subset \mathbb{R}^n$ is the set

$$\operatorname{span} X := \left\{ \sum_{i=1}^{p} \lambda_i x_i \mid p \in \mathbb{N},\ \lambda_i \in \mathbb{R}\ x_i \in X\ (i = 1, \dots, p) \right\}$$

of all *linear combinations* of of points in X. It is easily seen to be a subspace of $\mathbb{R}^n$, in fact, the smallest subspace containing X.

For a subspace $U \subset \mathbb{R}^n$ we say that $x_1, \dots, x_p \in U$ form a *basis* of U if the following hold:

i) $x_1, \dots, x_p$ are linearly indpendent;

ii) $\operatorname{span}\{x_1, \dots, x_p\} = U$.

It is known that all bases of a subspace $U \subset \mathbb{R}^n$ have the same length (number of vectors) which is called the *dimension* of the subspace and denoted by $\dim U$. In particular, $\dim \mathbb{R}^n = n$.

We recall some important properties of bases of subspaces in $\mathbb{R}^n$ (and hence every finite dimensional real vector space) below:

Theorem 1.1.1 (Bases in finite dimensional spaces) *Let $U \subset \mathbb{R}^n$ be a p-dimensional subspace.*

a) Let $a_1, \dots, a_p \in U$. Then the following are equivalent:

i) $\operatorname{span}\{a_1, \dots, a_p\} = U$*;*

ii) $a_1, \dots, a_p$ *are linearly independent;*

iii) $\{a_1, \dots, a_p\}$ *is a basis of* U*.*

b) (Basis Completion Theorem) For $r < p$ let $a_1, \dots, a_r \in U$ be linearly independent. Then there exist vectors $b_{r+1}, \dots, b_p \in U$ such that $\{a_1, \dots, a_r, b_{r+1}, \dots, b_p\}$ is a basis of U.

The canonical *scalar product* on $\mathbb{R}^n$ is given by

$$\langle \cdot,\, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}, \quad \langle x,\, y \rangle = x^T y \left(= \sum_{i=1}^{n} x_i y_i \right).$$

Note that the mapping

$$\| \cdot \|_2 : \mathbb{R}^n \to \mathbb{R}, \quad \|x\|_2 := \sqrt{x^T x}$$

is a *norm* on $\mathbb{R}^n$, called the *Euclidean norm*, i.e.

N1: $\|x\|_2 \geq 0$ and $\|x\|_2 = 0 \quad \Longleftrightarrow \quad x = 0 \quad (x \in \mathbb{R}^n)$ (definiteness);

N2: $\|\alpha x\|_2 = |\alpha| \cdot \|x\|_2 \quad (x \in \mathbb{R}^n, \alpha \in \mathbb{R})$ (absolute homogeneity);

N3: $\|x + y\|_2 \leq \|x\|_2 + \|y\|_2 \quad (x, y \in \mathbb{R}^n)$ (triangle inequality).

If no ambiguity arises we will drop the subscript and simply write $\| \cdot \|$ for the Euclidean norm.

Note that the inner product $\langle \cdot,\, \cdot \rangle$ and the Euclidean norm satisfy

$$\|x \pm y\|^2 = \|x\|^2 \pm 2 \langle x,\, y \rangle + \|y\|^2 \quad (x, y \in \mathbb{R}^n),$$

and obey the *Cauchy-Schwarz inequality*

$$| \langle x,\, y \rangle | \leq \|x\| \cdot \|y\| \quad (x, y \in \mathbb{R}^n)$$

where equality holds if and only if x and y are linearly dependent.

For a subspace $U \subset \mathbb{R}^n$ its *orthogonal complement* is defined by

$$U^\perp := \left\{ v \in \mathbb{R}^n \;\middle|\; v^T u = 0 \ (u \in U) \right\}$$

as the set of all vectors in $\mathbb{R}^n$ that are orthogonal to all points in U. It is easliy seen to be a subspace of $\mathbb{R}^n$.

1.1.2 Matrices

We denote the set of real $m \times n$-matrices by $\mathbb{R}^{m\times n}$. Note that every matrix $A \in \mathbb{R}^{m\times n}$ induces a linear mapping $x \in \mathbb{R}^n \mapsto Ax \in \mathbb{R}^m$.

For $A \in \mathbb{R}^{m\times n}$ its *transpose* is denoted by $A^T \in \mathbb{R}^{n\times m}$. Recall that for A, B such that AB exists, we have

$$(AB)^T = B^T A^T.$$

The *range* (or *image*) and the *kernel* (or *null space*) of A are given by

$$\operatorname{rge} A := \{Ax \mid x \in \mathbb{R}^n\} \subset \mathbb{R}^m \quad \text{and} \quad \ker A = \{x \in \mathbb{R}^n \mid Ax = 0\} \subset \mathbb{R}^n.$$

Note that $\operatorname{rge} A$ is a subspace of the image set $\mathbb{R}^m$ of the linear mapping $x \mapsto Ax$, and $\ker A$ is a subspace of its preimage space $\mathbb{R}^n$ linked through the *Range-Nullity Theorem*

$$n = \dim(\operatorname{rank} A) + \dim(\ker A). \tag{1.1}$$

The dimension of $\operatorname{rge} A$ is called the *rank* of A and is denoted by $\operatorname{rank} A$, is the i.e. $\operatorname{rank} A := \dim(\operatorname{rge} A)$. On the other hand it is also the maximum number of linearly independent row or column vectors of A. For $A \in \mathbb{R}^{m\times n}$ we thus have $\operatorname{rank} A \le \min\{m, n\}$.

The dimension of $\ker A$ is called the *defect* of A and is denoted by def A. With these abreviations (1.1) reads

$$n = \operatorname{rank} A + \operatorname{def} A \tag{1.2}$$

which we will refer to as the *rank formula.*

We recall the relation between the four fundamental subspaces associated with any matrix which are intimately linked through orthogonality.

Theorem 1.1.2 (Fundamental subspaces) *Let $A \in \mathbb{R}^{m\times n}$. Then the following hold:*

a) $(\operatorname{rge} A)^\perp = \ker A^T$ *and* $(\ker A^T)^\perp = \operatorname{rge} A$.

b) $(\ker A)^\perp = \operatorname{rge} A^T$ *and* $(\operatorname{rge} A^T)^\perp = \ker A$.

A square matrix $A \in \mathbb{R}^{n\times n}$ is called *invertible* or *nonsingular* if there exists $B \in \mathbb{R}^{n\times n}$ such that $AB = I = BA$. In this case B is called the *inverse (matrix)* of A and is denoted by A^{-1}. Note that if A is invertible, we have $(A^{-1})^T = (A^T)^{-1}$.

Proposition 1.1.3 (Invertibility characterizations) *Let $A \in \mathbb{R}^{n\times n}$. Then the following are equivalent:*

i) *A is invertible.*

ii) $\operatorname{rank} A = n$.

iii) $\ker A = \{0\}$.

iv) *The column vectors of A are linearly independent.*

v) *The column vectors of A span $\mathbb{R}^n$.*

vi) *The row vectors of A are linearly independent.*

vii) *The row vectors of A span $\mathbb{R}^n$.*

viii) *For every $b \in \mathbb{R}^n$ there exists a (unique) solution of '$Ax = b$'.*

It is often very useful to either think of a matrix $A \in \mathbb{R}^{m\times n}$ in terms of its columns $a_1, \ldots, a_n \in \mathbb{R}^m$, i.e. $A = [a_1, \ldots, a_m]$ or its row vectors $\hat{a}_1, \ldots, \hat{a}_m \in \mathbb{R}^n$, i.e. $A = \begin{pmatrix} \hat{a}_1^T \\ \vdots \\ a_m^T \end{pmatrix}$. A matrix-vector multiplication Ax for $x \in \mathbb{R}^n$ then can be interpreted as

$$\sum_{j=1}^{n} x_j a_i = Ax = \begin{pmatrix} \hat{a}_1^T x \\ \vdots \\ a_m^T x \end{pmatrix} \in \mathbb{R}^m.$$

Eigenvectors, eigenvalues and spectral decomposition

Let $A \in \mathbb{R}^{n\times n}$. Then $\lambda \in \mathbb{R}$ is called an *eigenvalue* of A if there exists $x \in \mathbb{R}^n \setminus \{0\}$ such that $Ax = \lambda x$. In this case x is called *eigenvector* of A to the eigenvalue λ.

The following result is a strong statement about the eigenvalues and eigenvectors of symmetric matrices.

Theorem 1.1.4 (Spectral theorem) *Let $A \in \mathbb{R}^{n\times m}$ be* symmetric*, i.e. $A^T = A$. Then there exists an orthogonal matrix $Q = [q_1, \ldots, q_n] \in \mathbb{R}^{n\times n}$ (i.e. $Q^T Q = I$) such that*

$$A = Q \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} Q^T = \sum_{i=1}^{n} \lambda_i q_i q_i^T.$$

The spectral theorem is equivalent to saying that every symmetric matrix has an orthonormal basis of eigenvectors and only real eigenvalues. It has many interesting consequences, for example it helps to characterize definiteness properties of symmetric matrices in terms of their eigenvalues, see Exercise 2. Recall that we call a symmetric matrix $A \in \mathbb{R}^{n\times n}$ *positive semidefinite* if

$$x^T Ax \geq 0 \quad (x \in \mathbb{R}^n),$$

and *positive definite* if

$$x^T Ax > 0 \quad (x \in \mathbb{R}^n \setminus \{0\}).$$

1.2 Optimization terminology

1.2.1 Suprema and infima

For a nonempty subset $S \subset \mathbb{R}$ its supremum, denoted by $\sup S$, is the smallest value $\tau \in (-\infty, +\infty]$ such that $\tau \geq s$ for all $s \in S$. In other words, $\sup S$ is largest (possibly improper) cluster point of sequences $\{x_k \in S\}$. In particular, $\sup S = +\infty$ if and only if S is unbounded from above. Moreover, we set $\sup \emptyset := -\infty$.

Analogously, the infimum of $\emptyset \neq S \subset \mathbb{R}$, denoted by $\inf S$, is the largest value $\sigma \in [-\infty, +\infty)$ such that $\sigma \leq s$ for all $s \in S$, and we put $\inf \emptyset := +\infty$.

Equipped with these definitions, for $f : \mathbb{R}^n \to \mathbb{R}$ and $X \subset \mathbb{R}^n$ we define

$$\sup_X f := \sup_{x\in X} f(x) := \sup \{ f(x) \mid x \in X \}$$

the supremum of f over X and, analogously,

$$\inf_X f := \inf_{x\in X} f(x) := \inf \{ f(x) \mid x \in X \},$$

the infimum of f over X.

1.2.2 Optimization problems

An *optimization problem* is described by an *objective function* $f : V \to \mathbb{R}$ on some real normed vector space V (or something even more general settings) and a *constraint set* (or *feasible set*) $X \subset V$. It consists in minimizing or maximizing f over X. In our study we will only be concerned with finite dimensional problems, i.e. where V is a finite dimensional real vector space. In this case V is isomorphic to $\mathbb{R}^n$ so there is no loss in generality to set $V = \mathbb{R}^n$ for the remainder.

The minimization problem described by $f : \mathbb{R}^n \to \mathbb{R}$ and $X \subset \mathbb{R}^n$ reads

$$\text{minimize } f(x) \quad \text{subject to} \quad x \in X. \tag{1.3}$$

A point $x \in X$ is called *feasible point* of (1.3). If $X = \emptyset$ we call (1.3) *infeasible.*

The maximization problem associated with f and X is given analogously by

$$\text{maximize } f(x) \quad \text{subject to} \quad x \in X. \tag{1.4}$$

On the one hand, (1.3) describes the task of computing $\inf_X f$, the infimum of f over X. On the other hand, one is usually more interested in finding points where this infimum is acutally attained (if they exist). These points are called *minimizers (minima) of f over X* or *solutions* of (1.3) and the set of all minimizers of f over X is given by

$$\operatorname{argmin}_X f := \operatorname{argmin}_{x \in X} f(x) := \left\{ x \in X \;\middle|\; f(x) = \inf_X f \right\}.$$

Analogously, (1.4) aims at computing

$$\operatorname{argmax}_X f := \operatorname{argmax}_{x \in X} f(x) := \left\{ x \in X \;\middle|\; f(x) = \sup_X f \right\}$$

the set of all *maximizers* of f over X, or at least one of them.

In what follows, we will abbreviate the words 'maximize' and 'minimize' by 'max' and 'min', respectively, and we will write 's.t.' instead of 'subject to'. Hence, e.g., (1.3) will read

$$\min f(x) \quad \text{s.t.} \quad x \in X.$$

Recall from calculus the important sufficient condition for (1.3) and (1.4) to have solutions which we formulate in our new terminology.

At this, recall that a subset $K \subset \mathbb{R}^n$ is said to be *compact* if it is *bounded* and *closed.*

Theorem 1.2.1 (Existence of Extrema) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous and $X \subset \mathbb{R}^n$ nonempty and compact. Then*

$$\operatorname{argmin}_X f \neq \emptyset \quad \textit{and} \quad \operatorname{argmax}_X f \neq \emptyset,$$

i.e. f takes its minimum and maximum over X.

Exercise 6 gives another sufficient condition which guarantees that a continuous function f takes its minimum over $\mathbb{R}^n$.

1.3 Convex sets

We start with the central definition of this section.

Definition 1.3.1 (Convex sets) *A set $C \subset \mathbb{R}^n$ is called* convex *if*

$$\lambda x + (1 - \lambda) y \in C \quad (x, y \in C,\ \lambda \in (0, 1)). \tag{1.5}$$

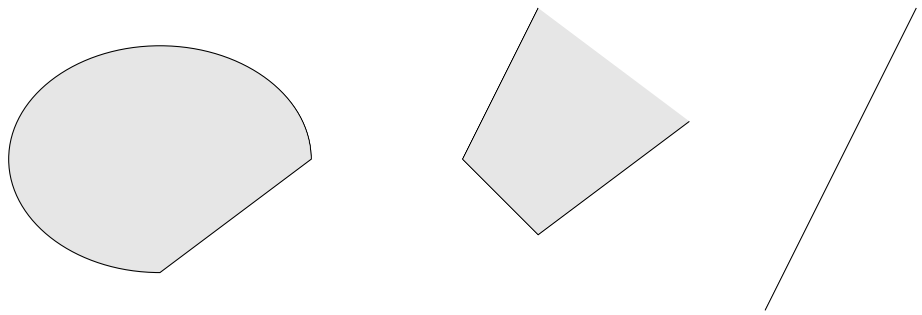

Figure 1.1: Convex sets in $\mathbb{R}^2$

In other words, a convex set is simply a set which contains all connecting lines of points from the set, see Figure (1.1) for examples.
A vector of the form

$$\sum_{i=1}^{r} \lambda_i x_i \quad \sum_{i=1}^{r} \lambda_i = 1,\ \lambda_i \geq 0\ (i = 1, \ldots, r)$$

is called a *convex combination* of the points $x_1, \ldots, x_r \in \mathbb{R}^n$. It is easily seen that a set $C \subset \mathbb{R}^n$ is convex if and only if it contains all convex combinations of its elements.

Below is a list of important classes of convex sets as well as operations that preserve convexity.

Example 1.3.2 (Convex sets)

a) (Subspaces) Every subspace of $\mathbb{R}^n$ (in particular $\mathbb{R}^n$ itself) is convex since convex combinations are special cases of linear combinations.

b) (Minkowski sum) The Minkowski sum

$$A + B := \{a + b \mid a \in A, b \in B\}$$

of two convex sets $A, B \subset \mathbb{R}^n$ is convex: For $x, y \in A + B$ there exist $a, a' \in A$ and $b, b' \in B$ such that $x = a + b$ and $y = a' + b'$. Then for $\lambda \in [0, 1]$ we have

$$\begin{aligned} \lambda x + (1 - \lambda) y &= \lambda(a + b) + (1 - \lambda)(a' + b') \\ &= \lambda a + (1 - \lambda) a' + \lambda b + (1 - \lambda) b' \end{aligned}$$

By convexity of A and B, respectively, we see that $\lambda a + (1 - \lambda) a' \in A$ and $\lambda b + (1 - \lambda) b' \in B$, hence $\lambda x + (1 - \lambda) y \in A + B$.
We point out that the Minkowski sum of two subspaces is known to be a subspace and can also be written as span $(A \cup B)$.

c) *(Intersection of convex sets) Arbitrary intersections of convex sets (in the same space) are convex, see Exercise 5a).*

d) *(Hyperplane) For $s \in \mathbb{R}^n$ and $\gamma \in \mathbb{R}$ the set*

$$\left\{x \in \mathbb{R}^n \mid s^T x = \gamma\right\}$$

is called a hyperplane. *It is a convex set, which is easily verified elementary or as a special case of Exercise 5b) with $A = s^T$ and $D = \{\gamma\}$.*

e) *(Half-spaces) Sets of the form*

$$\left\{x \in \mathbb{R}^n \mid s^T x \geq \gamma\right\}, \quad \left\{x \in \mathbb{R}^n \mid s^T x > \gamma\right\}$$

with $s \in \mathbb{R}^n, \gamma \in \mathbb{R}$, called closed *and* open half-spaces*, respectively, are convex. This can be verified easily by elementary calculations or, again, as a special case of Exercise 5b) with $D = [\gamma, +\infty)$ and $D = (\gamma, +\infty)$, respectively.*

f) *(Intervalls) The intervalls (closed, open, half-open) are exactly the convex sets in $\mathbb{R}$.*

◇

We continue with an important concept for convex sets which will play a central role in our theoretical analysis of the feasible set of linear programs.

Definition 1.3.3 *Let $C \subset \mathbb{E}$ be convex. A point $x \in C$ is said to be an* extreme point *if the following implication holds true for all $x_1, x_2 \in C$:*

$$\lambda x_1 + (1-\lambda)x_2 = x \quad, \ \lambda \in (0,1) \Rightarrow \quad x_1 = x_2.$$

The set of all extreme points of S is denoted by $\operatorname{ext} S$.

In other words: an extreme point of a convex set is a point in the set which cannot be represented as a convex combination of two points in the set that differ from the point in question. Figure 1.2 shows a convex set and its extreme points.
There are several ways of characterizing the fact that x is an extreme point of a convex set $C \subset \mathbb{E}$, see Exercise 7. In particular, it is shown there that we can restrict ourselves to the case $\lambda = \frac{1}{2}$ in the defining implication, see Definition 1.3.3.

Example 1.3.4 (Extreme points)

a) *(Subspaces) Nontrivial subspaces (or half-spaces) have no extreme points.*

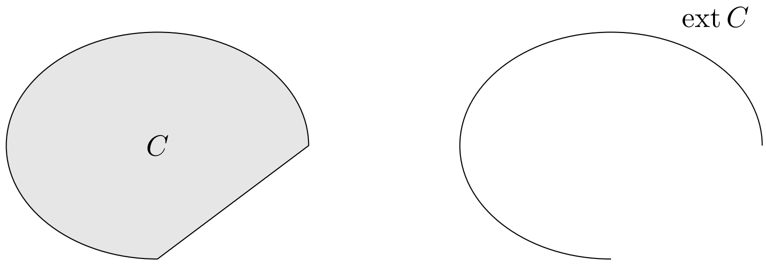

Figure 1.2: A convex set and its extreme points

b) (Cones) Let $K \subset \mathbb{R}^n$ be a cone, i.e.

$$\lambda x \in K \quad (x \in K, \lambda \geq 0).$$

Then $\operatorname{ext} K = \{0\}$ *or* $\operatorname{ext} K = \emptyset$ *(e.g. if K is a subspace).*

c) (Unit ball) Let $\mathbb{B} := \{x \in \mathbb{R}^n \mid \|x\|_2 \leq 1\}$ *be the closed unit ball. Using the identity*

$$\frac{1}{2}\|x+y\|^2 = \|x\|^2 + \|y\|^2 - \frac{1}{2}\|x-y\|^2, \quad (x, y \in \mathbb{E}) \tag{1.6}$$

(and Exercise 7) we realize that $\operatorname{ext} \mathbb{B} = \{x \mid \|x\|_2 = 1\}\,(= \operatorname{bd} \mathbb{B})$.

A given convex set does not necessarily have extreme points as Example 1.3.4 a) shows. A sufficient condition for a (nonempty) convex set to have extreme points is *compactness*, see Exercise 8.

1.3.1 Projection on convex sets

For a set $S \subset \mathbb{R}^n$ and a given point $x \in \mathbb{R}^n$ we want to assign to x the subset of points in S which have the shortest distance to it. We formalize this in the following definition.

Definition 1.3.5 (Projection on a set) *Let $S \subset \mathbb{R}^n$ be nonempty and $x \in \mathbb{R}^n$. Then we define the* projection *of x on S by*

$$P_S(x) := \operatorname{argmin}_{y \in C} \|x - y\|.$$

Observe that no changes occur if we subsitute $y \mapsto \|y - x\|$ for $y \mapsto \frac{1}{2}\|y - x\|^2$ in the above definition.

In general, the projection $P_S(x)$ of x on S is a subset of $\mathbb{R}^n$. We will now give sufficient conditions for this subset to be nonempty and also show when it contains at most one point.

For the proof we use Theorem 1.2.1.

Lemma 1.3.6 *Let $x \in \mathbb{E}$ and $S \subset \mathbb{E}$. Then the following hold:*

a) If S is closed then $P_S(x)$ is nonempty.

b) If S is convex then $P_S(x)$ has at most one element.

Proof:

a) Let $w \in S$. Defining the set

$$D := \{y \mid \|y - x\| \leq \|w - x\|\} \cap S,$$

we have

$$\operatorname{argmin}_{y \in S} \|y - x\| = \operatorname{argmin}_{y \in D} \|y - x\|.$$

As an intersection of a closed and a compact set D is compact. Hence, by Theorem 1.2.1, the continuous function $y \mapsto \|y - x\|$ takes its minimum on D. Therefore, it also takes its minimum on C, which proves the assertion.

b) Define $f : \mathbb{R}^n \to \mathbb{R},\ f(v) = \frac{1}{2}\|v - x\|^2$ and assume that $v_1, v_2 \in \operatorname{argmin}_S \frac{1}{2}\|(\cdot) - x\|^2$, i.e. $f^* := f(v_1) = f(v_2) = \inf_S f$. Then we have $\bar{v} := \frac{v_1+v_2}{2} \in S$, by convexity of C. Now compute that

$$\begin{aligned}
f(\bar{v}) &= \frac{1}{2}\left\|x - \frac{v_1 + v_2}{2}\right\|^2 \\
&= \frac{1}{8}\|(x - v_1) + (x - v_2)\|^2 \\
&\overset{(1.6)}{=} \frac{1}{4}\left(\|x - v_2\|^2 + \|x - v_2\|^2 - \frac{1}{2}\|(x - v_1) - (x - v_2)\|^2\right) \\
&= \frac{1}{2}\left[f(v_1) + f(v_2)\right] - \frac{1}{8}\|v_1 - v_2\|^2 \\
&= f^* - \frac{1}{8}\|v_1 - v_2\|^2.
\end{aligned}$$

Hence, necessarily, $v_1 = v_2$. Thus, $P_S(x) = \operatorname{argmin}_S \frac{1}{2}\|(\cdot) - x\|^2$ has at most one element.

□

An immediate consequence is the fact that the projection on a closed convex set is single-valued. Figure 1.3 illustrates this fact.

Corollary 1.3.7 (Projection on closed convex sets) *Let $C \subset \mathbb{R}^n$ be nonempty, closed and convex. Then P_C is a mapping $\mathbb{R}^n \to C$ with $x = P_C(x)$ if and only if $x \in C$.*

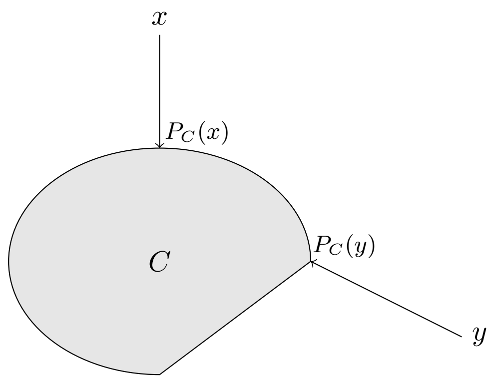

Figure 1.3: Projection on a closed convex set

The following theorem gives an important characterization of the projection on a closed convex set in terms of a variational inequality.

Theorem 1.3.8 (Projection Theorem) *Let $C \subset \mathbb{R}^n$ be nonempty, closed and convex and let $x \in \mathbb{R}^n$. Then $\bar{v} = P_C(x)$ if and only if*

$$\bar{v} \in C \quad \text{and} \quad \langle \bar{v} - x,\, v - \bar{v}\rangle \geq 0 \quad (v \in C). \tag{1.7}$$

Proof: First, assume that $\bar{v} = P_C(x) \in C$ and define $f : \mathbb{R}^n \to \mathbb{R},\ f(v) = \frac{1}{2}\|v - x\|^2$. By convexity of C, we have $\bar{v} + \lambda(v - \bar{v}) \in C$ for all $v \in C$ and $\lambda \in (0, 1)$. This implies

$$\frac{1}{2}\|\bar{v} - x\|^2 = f(\bar{v}) \leq f(\bar{v} + \lambda(v - \bar{v})) = \frac{1}{2}\|(\bar{v} - x) + \lambda(v - \bar{v})\|^2 \quad (v \in C,\ \lambda \in (0, 1)),$$

which, in turn, gives

$$\begin{aligned} 0 &\leq \frac{1}{2}\|(\bar{v} - x) + \lambda(v - \bar{v})\|^2 - \frac{1}{2}\|\bar{v} - x\|^2 \\ &= \lambda \langle \bar{v} - x,\, v - \bar{v}\rangle + \frac{\lambda^2}{2}\|v - \bar{v}\|^2 \end{aligned}$$

for all $v \in C$ and $\lambda \in (0, 1)$. Dividing by λ yields

$$0 \leq \langle \bar{v} - x,\, v - \bar{v}\rangle + \frac{\lambda}{2}\|y - \bar{v}\|^2.$$

Letting $\lambda \downarrow 0$ gives the desired inequality in (1.7).

In order to see the converse implication, let $\bar{v} \in \mathbb{E}$ such that (1.7) holds. For $v \in C$ we hence obtain

$$0 \geq \langle x - \bar{v},\, v - \bar{v}\rangle$$

$$\begin{aligned} &= \langle x - \bar{v},\, v - x + x - \bar{v} \rangle \\ &= \|x - \bar{v}\|^2 + \langle x - \bar{v},\, v - x \rangle \\ &\geq \|x - \bar{x}\|^2 - \|x - \bar{v}\| \cdot \|v - x\|, \end{aligned}$$

where the last inequality is due to the Cauchy-Schwarz inequality. As $v \in C$ was chosen arbitrarily, this yields

$$\|x - \bar{v}\| \leq \|x - v\| \quad (v \in C)$$

i.e. $\bar{v} = P_C(x)$. □

A geometrical interpretation of the projection theorem is as follows: The angle between $P_C(x) - x$ and $v - P_C(x)$ cannot exceed 90° for all $v \in C$.

1.3.2 A basic separation theorem

We commence with a key observation which is a simple consequence of the projection theorem.

Theorem 1.3.9 (Separation theorem) *Let $C \subset \mathbb{R}^n$ be nonempty, closed and convex, and let $x \notin C$. Then there exists $s \in \mathbb{R}^n$ with*

$$\langle s,\, x \rangle > \sup_{v \in C} \langle s,\, v \rangle \, .$$

Proof: Put $s := x - P_C(x) \neq 0$. Then the projection theorem yields

$$0 \geq \langle x - P_C(x),\, v - P_C(x) \rangle = \langle s,\, v - x + s \rangle = \langle s,\, v \rangle - \langle s,\, x \rangle + \|s\|^2 \quad (v \in C).$$

Thus,

$$\langle s,\, x \rangle - \|s\|^2 \geq \langle s,\, v \rangle \quad (v \in C),$$

hence, s fulfills the requirements of the theorem. □

We would like to note some technicalities about the former theorem.

Remark 1.3.10

a) The vector s can always be substituted for $-s$ and thus, under the same assumptions, there exists $s \in \mathbb{R}^n$ such that $\langle s,\, x \rangle < \inf_{v \in C} \langle s,\, v \rangle$.

b) By positive homogeneity, we can assume w.l.o.g. that $\|s\| = 1$.

It is not quite clear yet why the above theorem was labeled *separation* theorem. In the situation of the theorem, define $\gamma := \frac{1}{2}(\langle s,\, x\rangle + \sup_{y\in C} \langle s,\, x\rangle)$. Then

$$x \in \left\{z \;\middle|\; s^T z > \gamma\right\} \quad \text{and} \quad C \subset \left\{z \;\middle|\; s^T z < \gamma\right\},$$

i.e. $\{x\}$ and C lie in two distinct open half-spaces induced by the hyperplane $H = \left\{z \;\middle|\; s^T z = \gamma\right\}$. We say that H separates the set C from the point $x \notin C$. This situation is illustrated in Figure 1.4.

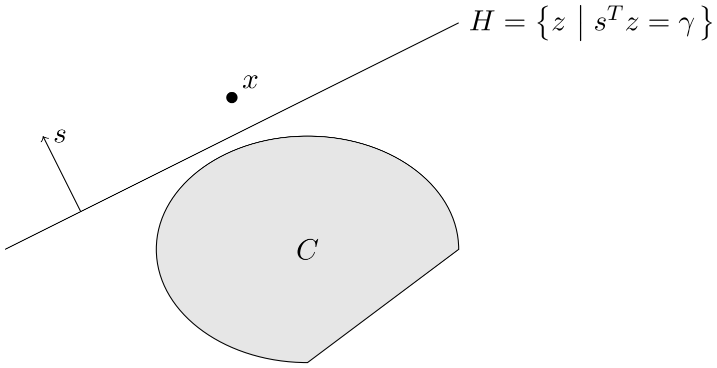

Figure 1.4: Separation of a point from a closed convex set

1.3.3 Polyhedra

The next example of a convex set merits its own definition since we are going to need it frequently throughout. From now on we employ the following notation for vectors $x = (x_i), y = (y_i) \in \mathbb{R}^n$:

$$x \geq y \quad :\Leftrightarrow \quad x_i \geq y_i \quad (i = 1, \ldots, n).$$

and

$$x > y \quad :\Leftrightarrow \quad x_i \geq y_i \; (i = 1, \ldots, n) \quad \text{and} \quad \exists\, j :\; x_j > y_j.$$

Definition 1.3.11 (Polyhedra) *Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Then the set*

$$P := \{x \in \mathbb{R}^n \mid Ax \geq b\}$$

is called a polyhedron. *A bounded polyhedron is called* polytope.

Note that we can always incorporate inequalities of the type $Bx \le c$ by multiplying by -1:

$$Bx \le c \quad \Longleftrightarrow \quad (-B)x \ge -c.$$

Likewise equalities of the form $Bx = c$ are covered by using two inequalities:

$$Bx = b \quad \Longleftrightarrow \quad Bx \le c \quad \text{and} \quad Bx \ge c \quad \Longleftrightarrow \quad \begin{pmatrix} B \\ -B \end{pmatrix} x \ge \begin{pmatrix} c \\ c \end{pmatrix}.$$

Writing the matrix $A \in \mathbb{R}^{m\times n}$ using its rows $a_i^T \in \mathbb{R}^n$ $(i = 1, \ldots, m)$, i.e.

$$A = \begin{pmatrix} \vdots \\ a_i^T \ (i = 1, \ldots, m) \\ \vdots \end{pmatrix},$$

we see that the polyhedron $P = \{x \in \mathbb{R}^n \mid Ax \ge b\}$ is simply the intersection of (finitely many) closed half-spaces $\left\{x \in \mathbb{R}^n \;\middle|\; a_i^T x \ge b\right\}$, i.e.,

$$P = \bigcap_{i=1}^{m} \left\{x \in \mathbb{R}^n \;\middle|\; a_i^T x \ge b\right\}.$$

Clearly, this reasoning also works in the converse direction, i.e. every finite intersection of closed half-spaces is a polyhedron. We have hence proven the following result, where the second statement uses that finite intersections of closed sets are closed, and the same holds for convexity, see to Example 1.3.2.

Proposition 1.3.12 (Convexity of polyhedra) *A set $P \in \mathbb{R}^n$ is the intersection of finitely many closed half-spaces if and only it is a polyhedron. In particular, every polyhedron is closed and convex, hence compact if bounded.*

We illustrate our recent remarks by a concrete example of a polyhedron in $\mathbb{R}^2$.

Example 1.3.13 *We consider the polyhedron $P = \left\{x \in \mathbb{R}^2 \mid Ax \ge b\right\}$ defined by*

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ -1 & -1 \\ -1 & 3 \end{pmatrix} \quad \textit{and} \quad b = \begin{pmatrix} 0 \\ 0 \\ -4 \\ -8 \end{pmatrix}.$$

With the rows

$$a_1^T = (1 \quad -1), \quad a_2^T = (1 \quad 1), \quad a_3^T = (-1 \quad -1), \quad a_4^T = (1 \quad 3)$$

of A we thus have

$$P = \bigcap_{i=1}^{4} \left\{x \in \mathbb{R}^2 \;\middle|\; a_i^T x \ge b_i\right\}.$$

See Figure 1.5 for an illustration.

◇

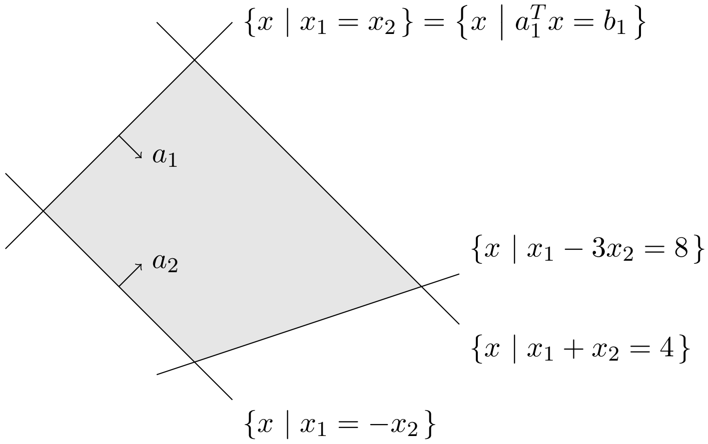

Figure 1.5: Illustration of Example 1.3.13

Exercises to Chapter 1

1. **(Linear independence, bases etc.)**
   a) Let $z_1, \dots, z_r \in \mathbb{R}^n \setminus \{0\}$ such that $z_i^T z_j = 0$ for all i, j with $i \neq j$. Show that $z_1, \dots, z_r$ are linearly independent.
   b) For $b_1 = (1, 0, 1)^T$ and $b_2 = (0, 1, 0)^T$ find $b_3 \in \mathbb{R}^3$ such that b_1, b_3, b_3 form a basis of $\mathbb{R}^3$. Is b_3 uniquely determined?
   c) For $z \in \mathbb{R}^n$ determine $\operatorname{rank}(zz^T)$.

2. **(Positive (semi)definiteness)** Let $A \in \mathbb{R}^{n \times n}$ be symmetric with the eigenvalues $\lambda_1, \dots, \lambda_n$ (see Theorem 1.1.4). Show the following:
   a) $x^T A x \geq 0 \ (x \in \mathbb{R}^n) \iff \lambda_i \geq 0 \ (i = 1, \dots, n)$;
   b) $x^T A x \, 0 \ (x \in \mathbb{R}^n \setminus \{0\}) \iff \lambda_i > 0 \ (i = 1, \dots, n)$;

3. **(Symmetric matrices, quadratic functions and infima)** Let $A \in \mathbb{R}^{n \times n}$ be symmetric and $b \in \mathbb{R}^n$.
   a) Show that $\ker A \cap \operatorname{rge} A = \{0\}$.
   b) Prove that $\ker A + \operatorname{rge} A = \mathbb{R}^n$.
   c) For $q : \mathbb{R}^n \to \mathbb{R}, \ q(x) = \frac{1}{2} x^T A x + b^T x$ show that the following are equivalent:
      i) $\inf_{\mathbb{R}^n} q > -\infty$;
      ii) A is positive semidefinite (i.e. $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$) and $b \in \operatorname{rge} A$;

iii) $\operatorname{argmin}_{\mathbb{R}^n} q \neq \emptyset$.

Hint: You may use (if needed) without proof that a positive semidefinite matrix has only nonnegative eigenvalues.

4. **(Minimizing a linear function over the unit ball)** Let $g \in \mathbb{R}^n \setminus \{0\}$. Compute the solution of the optimization problem

$$\min\ \langle g,\, d\rangle \quad \text{s.t.} \quad \|d\| \leq 1.$$

5. **(Convexity preserving operations)**
   a) (Intersection) Let I be an arbitrary index set (possibly uncountable) and let $C_i \subset \mathbb{R}^n$ $(i \in I)$ be a family of convex sets. Show that $\bigcap_{i\in I} C_i$ is convex.
   b) (Linear images and preimages) Let $A \in \mathbb{R}^{m\times n}$ and let $C \subset \mathbb{R}^n$, $D \subset \mathbb{R}^m$ be convex. Show that

   $$A(C) := \{Ax \mid x \in C\} \quad \text{and} \quad A^{-1}(D) = \{x \mid Ax \in D\}$$

   are convex.

6. **(Level-boundedness and existence of minimizers)** Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous such that for all $\alpha \in \mathbb{R}$ the *sublevel set*

$$\operatorname{lev}_\alpha f := \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$$

is bounded (possibly empty). Show that f takes a minimum on $\mathbb{R}^n$.

7. **(Characterization of extreme points)** Let $C \subset \mathbb{E}$ be convex. Show that for $x \in C$ the following are equivalent:
   i) For all $x_1, x_2 \in C$ we have that $\frac{1}{2}x_1 + \frac{1}{2}x_2 = x$ imnplies $x_1 = x_2$.
   ii) x is an extreme point of C.
   iii) $x = \sum_{i=1}^r \lambda_i x_i$ for some $r \in \mathbb{N}$, $x_i \in C$ $(i = 1,\dots,r)$, $\lambda_i \geq 0$ $(i = 1,\dots,r)$ and $\sum_{i=1}^r \lambda_i = 1$ implies $x_i = x$ for all $i = 1,\dots,r$.
   iv) $C \setminus \{x\}$ is convex.

8. **(Existence of extreme points)** Let $S \subset \mathbb{E}$ be nonempty and compact. Then $\operatorname{ext} S \neq \emptyset$.

9. **(Projection on subspaces)** Let $U \subset \mathbb{R}^n$ be a subspace. Then it is known that for every $x \in \mathbb{R}^n$ there exist unique vectors $u \in U$ and $u' \in U^\perp$ such that $x = u + u'$. Show the following:
   a) $P_U(x) = u$.

b) $P_U : \mathbb{R}^n \to U$ is linear.

10. **(Separation of convex sets)** Let $C \subset \mathbb{R}^n$ be convex and closed and $D \subset \mathbb{R}^n$ convex and compact such that $C \cap D = \emptyset$. Show that there exists $s \in \mathbb{R}^n \setminus \{0\}$ such that
$$\inf_{v \in C} \langle s, v \rangle > \sup_{w \in D} \langle s, w \rangle .$$

11. **(Convex Optimization)** For a convex function $f : \mathbb{R}^n \to \mathbb{R}$ and a convex set $X \subset \mathbb{R}^n$ consider the optimization problem
$$\min f(x) \quad \text{s.t.} \quad x \in X. \tag{1.8}$$
Prove the following:
   a) $\operatorname{argmin}_X f \subset X$ is convex.
   b) If $\bar{x}$ is a local minimizer[1] of (1.8) if and only if $\bar{x} \in \operatorname{argmin}_X f$.

12. **(Nonexpansiveness of projection)** Let $S \subset \mathbb{R}^n$ be nonempty, closed and convex. Show that
$$\|P_S(x) - P_S(y)\| \leq \|x - y\| \quad (x, y \in S).$$

[1]There exists $\varepsilon > 0$ such that $f(\bar{x}) \leq f(x)$ for all $x \in B_\varepsilon(\bar{x}) \cap X$.