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 2016

Last updated: December 9, 2016

"*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 $U \in \mathbb{R}^n$ 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

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

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 $a_1, \ldots, a_p \in U$. Then the following are equivalent:

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

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

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

b) (Steinitz's Exchange Theorem) Let $\{a_1, \ldots, a_p\}$ be a basis of U and $b_1, \ldots, b_l \in U$ linearly independent. Then there exist indices $i_{l+1}, \ldots, i_p$ such that $\{b_1, \ldots, b_l, a_{i_{l+1}}, \ldots, a_{i_p}\}$ is a basis of U.

c) (Basis Completion Theorem) For $r < p$ let $a_1, \ldots, a_r \in U$ be linearly independent. Then there exist vectors $b_{r+1}, \ldots, b_p \in U$ such that $\{a_1, \ldots, a_r, b_{r+1}, \ldots, b_p\}$ 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 $\mathbb{R}^n$, call the *Euclidean norm*, i.e.

N1: $\|x\|_2 \geq 0$ and $\|x\|_2 = 0 \iff x = 0 \quad (x \in \mathbb{R}^n)$ (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 obey the *Cauchy-Schwarz inequality*

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

and 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

$$U^\perp := \{ v \in \mathbb{R}^n \mid v^T u = 0 \ (u \in U) \}$$

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 $A^T \in \mathbb{R}^{m\times n}$. Recall that for A, B such that AB exists, we have

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

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

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

Note that $\operatorname{im} A$ is a subspace of the image set $\mathbb{R}^m$ of the linear mapping $x \in\mapsto Ax$, 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 $\operatorname{im} A$ is called the *rank* of A and is denoted by $\operatorname{rank} A$, is the i.e. $\operatorname{rank} A := \dim(\operatorname{im} A)$. 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 $\operatorname{rank} A \leq \min\{m, n\}$.

The dimension of $\ker A$ is called the *defect* of A and is denoted by $\operatorname{def} A$. 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 $A \in \mathbb{R}^n$. Then the following hold:*

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

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

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 $a_1, \dots, a_n \in \mathbb{R}^m$, i.e. $A = [a_1, \dots, a_m]$ or its row vectors $\hat{a}_1, \dots, \hat{a}_m \in \mathbb{R}^n$, 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

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

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 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

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

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

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

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

$$\underset{X}{\textit{argmin}}\, f := \underset{x \in X}{\textit{argmin}}\, f(x) := \left\{ x \in X \;\middle|\; f(x) = \inf_X f \right\}.$$

Analogously, (1.4) aims at computing

$$\underset{X}{\textit{argmax}}\, f := \underset{x \in X}{\textit{argmax}}\, f(x) := \left\{ x \in X \;\middle|\; f(x) = \sup_X f \right\}$$

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 5 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*

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

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.

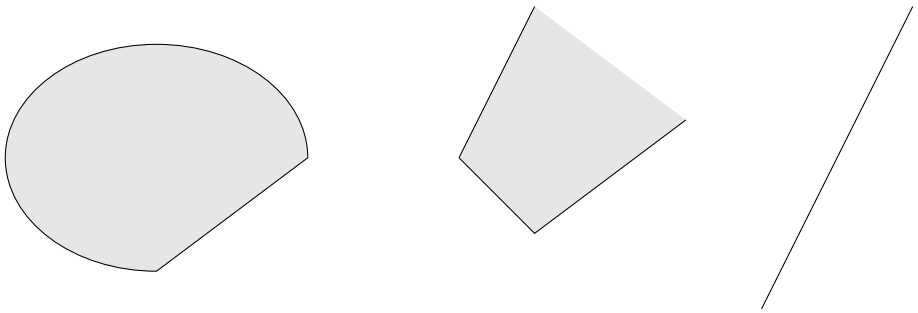

Figure 1.1: Convex sets in $\mathbb{R}^2$

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 asconvex 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 $A, B \in \mathbb{R}^n$ 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 $\operatorname{span}(A \cup B)$.

c) *(Intersection of convex sets) Arbitrary intersections of convex sets are convex, see Exercise 4a).*

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

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

is called a hyperplane. *It is a convex set, which is easily verified elementary or as a special case of Exercise 4b) with $A = s^T$ and $D = \{\gamma\}$.*

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

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

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 4b) with $D = [\gamma, +\infty)$ and $D = (\gamma, +\infty)$, respectively.*

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

$\diamond$

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 $x_1, x_2 \in S$:*

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

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

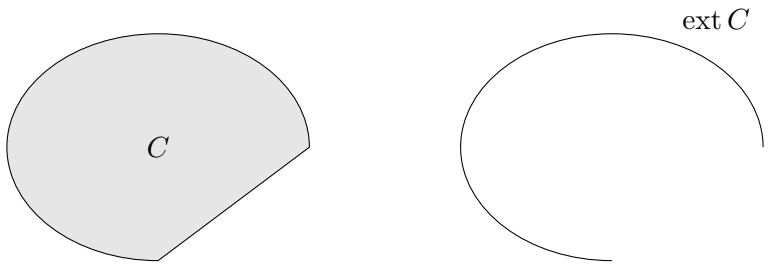

Figure 1.2: A convex set and its extreme points

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 6. 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.

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 6) 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 7.

1.3.1 Projection on convex sets

For a set $S \subset \mathbb{R}^n$ and a given point in $x \in \mathbb{R}^n$ 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 $S \subset \mathbb{E}$. The 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 $v_1, v_2 \in \operatorname{argmin}_S \frac{1}{2}\|(\cdot) - x\|^2$. Then we have $\bar{v} := \frac{v_1 + v_2}{2} \in S$, by convexity of C. A short computation shows (cf. (1.6)) that

$$f(\bar{v}) = \frac{1}{2}\left[f(v_1) + f(v_2)\right] - \frac{1}{8}\|v_1 - v_2\|^2 = \min_{x \in S} f(x) - \frac{1}{8}\|v_1 - v_2\|^2,$$

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.

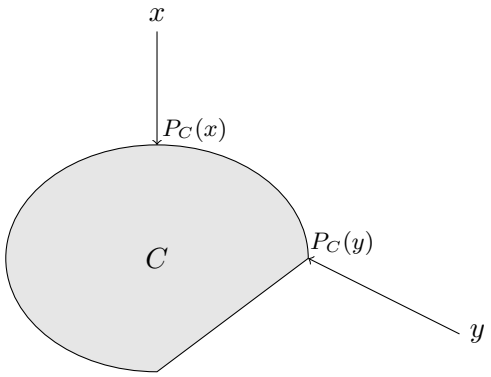

Figure 1.3: Projection on a closed convex set

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 $x \in C$.*

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

For its proof, observe that by the definition of the Euclidean norm $\|\cdot\| := \|\cdot\|_2$ and the canonical scalar product $\langle \cdot, \cdot \rangle$ we have

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

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*

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

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

$$\begin{aligned} 0 &\geq \langle x - \bar{v},\, v - \bar{v} \rangle \\ &= \langle x - \bar{v},\, v - x + x - \bar{v} \rangle \\ &= \|x - \bar{v}\|^2 + \langle x - \bar{v},\, v - x \rangle \\ &\geq \|x - \bar{x}\|^2 - \|x - \bar{v}\| \cdot \|v - x\|, \end{aligned}$$

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.

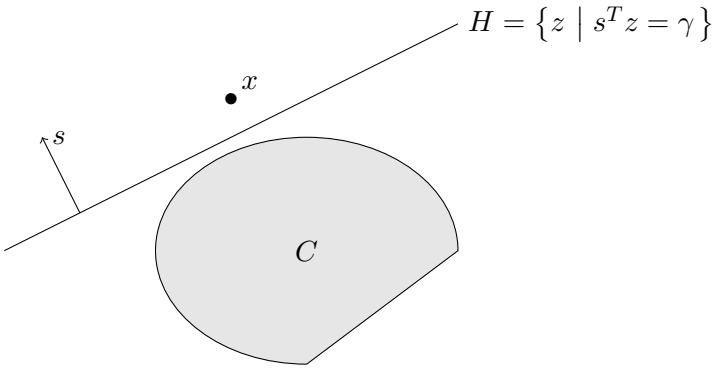

Figure 1.4: Separation of a point from a closed convex set

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 $Bx \leq c$ by multiplying by -1:

$$Bx \leq c \quad \Longleftrightarrow \quad (-B)x \geq -c.$$

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

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

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

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

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 $P = \left\{x \in \mathbb{R}^2 \mid Ax \geq b\right\}$ 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

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

See Figure 1.5 for an illustration.

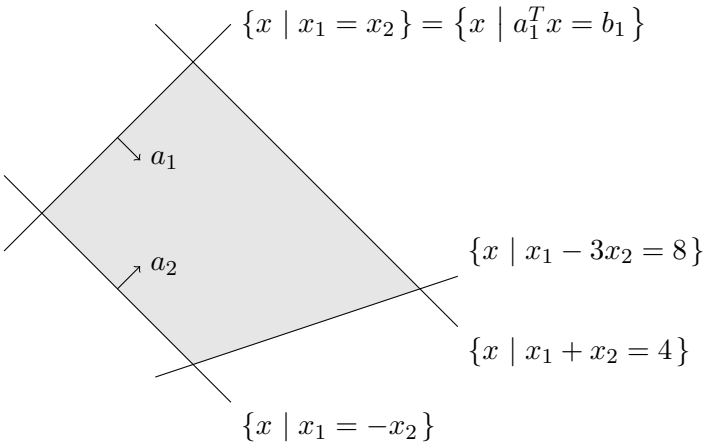

Figure 1.5: Illustration of Example 1.3.13

◇

We now want to study the extreme points, see Definition 1.3.3, of a polyhedron. To this end, we first establish the notion of a vertex and a basic (feasible) point of a polyhedron.

We start with the notion of a *vertex* of a polyhedron which does not play a major role, but which we rather provide for completeness.

Definition 1.3.14 (Vertex of polyhedron) *Let $P := \{x \in \mathbb{R}^n \mid Ax \geq b\}$ be a polyhedron. Then $x \in P$ is called a* vertex *of P if there exists $c \in \mathbb{R}^n$ such that*

$$c^T x < c^T y \quad (y \in P \setminus \{x\}).$$

In other words, x is a vertex of P if and only if there exists a hyperplane (namely $\left\{y \;\middle|\; c^T y = c^T x\right\}$) which meets P only at x and such that P lies entirely in one of the corresponding half-spaces (namely $\left\{y \;\middle|\; c^T y > c^T x\right\}$).

We continue with the definition of a basic (feasible) point of a polyhedron. At this, with a slight abuse of language, we call a set of (affine) linear constraints

$$s_i^T x = b_i \ (i \in I)$$

linearly independent if the corresponding vectors s_i $(i \in I)$ are linearly independent.

Definition 1.3.15 (Basic points of polyhedra) *For some finite index sets I_1, I_2 let*

$$P = \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{rcll} a_i^T x & = & b_i & (i \in I_1) \\ a_i^T x & \geq & b_i & (i \in I_2) \end{array} \right\}$$

be a polyhedron and for $x \in \mathbb{R}^n$ define

$$I(x) := \left\{ i \in I_1 \cup I_2 \;\middle|\; a_i^T x = b_i \right\},$$

the set of indices of active constraints at x. A point $\bar{x} \in \mathbb{R}^n$ is called a basic point *of P if the following holds:*

i) $a_i^T \bar{x} = b_i$ $(i \in I_1)$ (i.e. x satisfies the equality constraints);

ii) There are n linearly independent active constraints, i.e. there exists and index set $K \subset I(\bar{x})$ such that

$$|K| = n \quad \text{and} \quad a_i \ (i \in K) \text{ are linearly independent.}$$

In addition, we call $\bar{x}$ a basic feasible point *of P if it is a basic point with $\bar{x} \in P$.*

Clearly, if the number of constraints that defines a polyhedron in $\mathbb{R}^n$ is less than n, it cannot have a basic (feasible) point.

We point out that the question whether a point x is a basic point of a polyhedron may depend on its representation, see Exercise 10. This is not the case for basic *feasible* points as will come out of the following main result of this section which says that there is no difference between the extreme points, the vertices and the basic feasible points of a polyhedron.

As a preparation we provide an auxiliaty result which will turn out to be very useful in other situations of our study.

Lemma 1.3.16 (Active constraint lemma) *For a finite index M let $s_i \in \mathbb{R}^n$ $(i \in M)$, $b \in \mathbb{R}^{|M|}$ and $\bar{x} \in \mathbb{R}^n$. Then for the set*

$$\bar{I} := \left\{ i \in M \;\middle|\; s_i^T \bar{x} = b_i \right\}$$

the following are equivalent:

i) There exist n vectors $\{s_i \mid i \in \bar{I}\}$ *that are linearly independent.*

ii) span $\{s_i \mid i \in \bar{I}\} = \mathbb{R}^n$.

iii) The linear system of equations

$$s_i^T x = b_i \quad (i \in I)$$

has a unique solution (namely $\bar{x}$*).*

Proof: The equivalence of i) and ii) follows immediately from Theorem 1.1.1 a).

To prove the equivalence of i) (and ii)) to iii) consider the matrix

$$S := \begin{pmatrix} \vdots \\ -s_i^T \ (i \in \bar{I}) - \\ \vdots \end{pmatrix} \in \mathbb{R}^{|I| \times n}.$$

By the rank formula we have $\ker S = \{0\}$ if and only if $\operatorname{rank} S = n$ which is equivalent to i). On the other hand, the system '$Sx \stackrel{!}{=} b$' has exactly one solution (namely $\bar{x}$) if $\ker S = \{0\}$.

This concludes the proof.

□

We now give the main result of this section, which states that there is no difference between the extreme points, vertices and basic feasible points of a poyhedron.

Theorem 1.3.17 *Let* $P \subset \mathbb{R}^n$ *be a polyhedron and let* $\bar{x} \in P$*. Then the following are equivalent:*

i) $\bar{x}$ *is a vertex of* P*.*

ii) $\bar{x}$ *is extreme point of* P*.*

iii) $\bar{x}$ *is a basic feasible point of* P *(in any representation in the sense of Definition 1.3.15.*

Proof: i)⇒ii)': If $\bar{x}$ is a vertex of P there exists $c \in \mathbb{R}^n$ such that $c^T\bar{x} < c^T y$ for all $y \in P \setminus \{\bar{x}\}$. Now assume that $y, z \in P \setminus \{\bar{x}\}$ and $\lambda \in (0,1)$ with $\bar{x} = \lambda y + (1-\lambda)z$. Then

$$c^T\bar{x} = \lambda c^T y + (1-\lambda)c^T z > \lambda c^T\bar{x} + (1-\lambda)c^T\bar{x} = c^T\bar{x},$$

which is a contradiction; hence $\bar{x}$ is an extreme point of P.

ii)⇒iii)': (Contraposition) Assume that $\bar{x} \in P$ is not a basic feasible solution of P for P defined by a (finite) family of equalities and inequalities of the form

$$a_i^T x \geq b_i \quad \text{and} \quad a_i^T x_i = b_i. \tag{1.8}$$

Then for $\bar{I} := \left\{ i \mid a_i^T \bar{x} = b_i \right\}$ (the index set of active constraints) we cannot find n indices in $\bar{I}$ such that the corresponding vectors a_i are linearly independent. Therefore, for

$$\bar{A} := \begin{pmatrix} \vdots \\ a_i^T \ (i \in \bar{I}) \\ \vdots \end{pmatrix} \in \mathbb{R}^{|\bar{I}| \times n} \tag{1.9}$$

we have $\operatorname{rank} \bar{A} < n$, thus by the rank formula (1.2), we have $\ker \bar{A} \neq \{0\}$. Hence, there exists $d \in \mathbb{R}^n \setminus \{0\}$ such that $a_i^T d = 0$ $(i \in \bar{I})$. Now choose $\varepsilon > 0$ sufficiently small such that

$$\varepsilon |a_i^T d| < a_i^T \bar{x} - b_i \quad (i \notin \bar{I})$$

and put

$$y := \bar{x} + \varepsilon d \quad \text{and} \quad z := \bar{x} - \varepsilon d.$$

Then

$$a_i^T y = a_i^T \bar{x} + \varepsilon a_i^T d = b_i \quad (i \in \bar{I})$$

and, by the choice of $\varepsilon > 0$, we have

$$a_i^T y = a_i^T \bar{x} + \varepsilon a_i^T d > a_i^T \bar{x} - \varepsilon |a_i^T d| > b_i.$$

Therefore, $y \in P$ and analogously we can show that $z \in P$. On the other hand, we have $\frac{1}{2}(y + z) = \bar{x}$, and $y \neq z$, hence $\bar{x}$ is not an extreme point of P which yields the desired implication.

iii)⇒i)': Assume that $\bar{x}$ is a basic feasible solution of P for P defined through (1.8). Putting $c := \sum_{i \in \bar{I}} a_i$, we have

$$c^T \bar{x} = \sum_{i \in \bar{I}} a_i^T \bar{x} = \sum_{i \in I} b_i.$$

On the other, hand we have

$$c^T x = \sum_{i \in \bar{I}} a_i^T x \geq \sum_{i \in \bar{I}} b_i = c^T \bar{x} \quad (x \in P). \tag{1.10}$$

And equality in (1.10) for some $x \in P$ holds if and only if $a_i^T x = b_i$ for all $i \in \bar{I}$. Suppose there is $x' \in P$ that satisfies this. Defining $\bar{A}$ as in (1.9), we thus have $\bar{A}(x' - \bar{x}) = 0$,

hence $\operatorname{rank} A = n - \operatorname{def} A < n$, which contradicts the fact that $\bar{x}$ is a basic feasible point (and hence $\operatorname{rank} \bar{A} = n$). Therefore such $x' \in P$ cannot exist, and thus $c^T x < c^T \bar{x}$ for all $x \in P$, thus $\bar{x}$ is a vertex of P. □

Exercises to Chapter 1

1. **(Linear independence, bases etc.)**
   a) Let $z_1, \ldots, z_r \in \mathbb{R}^n \setminus \{0\}$ such that $z_i^T z_j = 0$ for all i, j with $i \neq j$. Show that $z_1, \ldots, z_r$ 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. **(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 $\ker A \cap \operatorname{im} A = \{0\}$.
   b) Prove that $\ker A + \operatorname{im} A = \mathbb{R}^n$.
   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 $b \in \operatorname{im} A$;
      iii) $\mathit{argmin}_{\mathbb{R}^n} q \neq \emptyset$.

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

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

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

4. **(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.

5. **(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 $\mathbb{R}^n$.

6. **(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 $r \in \mathbb{N}$, $x_i \in C$ $(i = 1, \ldots, r)$, $\lambda_i \geq 0$ $(i = 1, \ldots, r)$ and $\sum_{i=1}^r \lambda_i = 1$ implies $x_i = x$ for all $i = 1, \ldots, r$.
   iv) $C \setminus \{x\}$ is convex.

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

8. **(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.

9. (**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

$$\inf_{v \in C} \langle s, v \rangle > \sup_{w \in D} \langle s, w \rangle .$$

10. **(Basic points of polyhedra)** Consider the polyhedron

$$P = \left\{ x \in \mathbb{R}^2 \mid x_1 + x_2 = 0,\ x_2 - x_1 \geq 1 \right\}.$$

    a) Draw a sketch of P and mark all basic feasible points.
    b) Is $\bar{x} = (0, 0)$ a basic point of P (in the given representation)?
    c) Find a representation of P such that $\bar{x}$ *is* a basic point of P.

2 Linear Programming Theory

2.1 LP terminology and examples

In this section we want to lay out the notational ground for linear programming and give motivational examples.
A *linear program (LP)* is an optimization problem where the objective function is linear with the feasible set being a polyhedron, i.e. every linear program is of the the form

$$\min (or\ \max)\ c^T x \quad \text{s.}t. \quad x \in P$$

for $c \in \mathbb{R}^n$ and $P \subset \mathbb{R}^n$ a polyhedron (i.e. defined by fintely many linear equalities and inequalities).

We now give some instances of this problem class. The first one is a standard example which, in its two-dimensional form, was taken from [3].

Example 2.1.1 (The diet problem) *A farmer needs to feed his cows and has two feed resources:* concentrate feed *and* fresh clover*.*

The nutritional value of these food components and their cost per unit U is given in Table 2.1.

	carbohydrates	protein	vitamins	cost
1 U concentrate	20 U	15 U	5 U	10 $
1 U clover	20 U	3 U	10 U	7 $
Consumption/Day	60 U	15 U	20 U	

Table 2.1: Diet problem data

The farmer clearly wants to minimize his expenses while making sure that the cows get all the necessary nutrients. Setting x_1 as the amount of concentrate (units) and x_2 as the amount of clover (units) this results in the following optimization problem:

$$\begin{aligned} \min\ 10x_1 + 7x_2 \quad s.t. \quad & 20x_1 + 20x_2 \geq 60, \\ & 15x_1 + 3x_2 \geq 15, \\ & 5x_1 + 10x_2 \geq 20, \\ & x_1, x_2 \geq 0. \end{aligned} \tag{2.1}$$

Since this a problem in only two variables, we are actually able to solve it graphically, see 2.1.

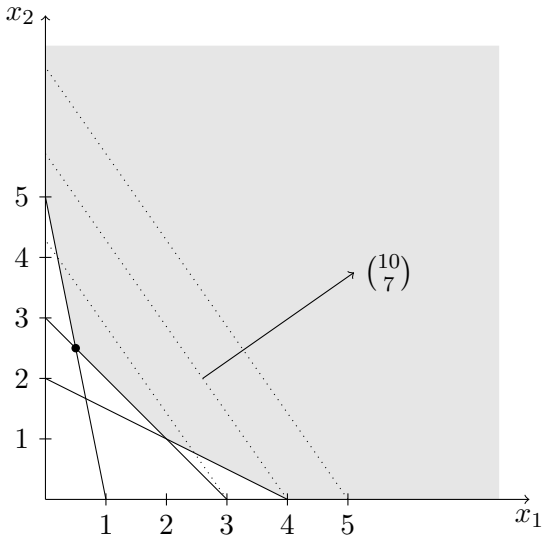

Figure 2.1: Graphical solution of the diet problem (2.1)

Every constraint defines a (closed) half-space. The intersection of these half-spaces is the feasible set of (2.1)*, which is hence a polyhedron.*

Since the objective function $x \mapsto f(x) = 10x_1 + 7x_2$ *is linear, its level sets are lines with the slope* $-\frac{10}{7}$ *to which the gradient* $\nabla f(x) = \binom{10}{7}$ *is perpendicular for all* $x \in \mathbb{R}^2$*.*

The solution of of (2.1) *lies at the point where the level set to the smallest level (here, the one closest to the origin) meets the feasible set. This yields the point* $\bar{x} = (0.5, 2.5)$*.* ◇

It is not by coincidence that the solution of the above problem lies in one of the corners (extreme points) of the feasible set. We will learn later on from a fundamental theorem that, if a linear program has a solution, then the solution set contains an extreme point.

We now present a second example which goes a simple example of a classification problem which occurs in [6].

Example 2.1.2 (Separation of points - rabbit or weasel?) *A computer-controlled rabbit trap is to be programmed such that it catches rabbits, but releases weasels. The trap can weigh the animal and determine the area of its shadow. These two parameters were collected for a number of specimen of rabbits and weasels . This is depicted, using '○' for rabbits and '⋆' for weasels, in Figure 2.2.*

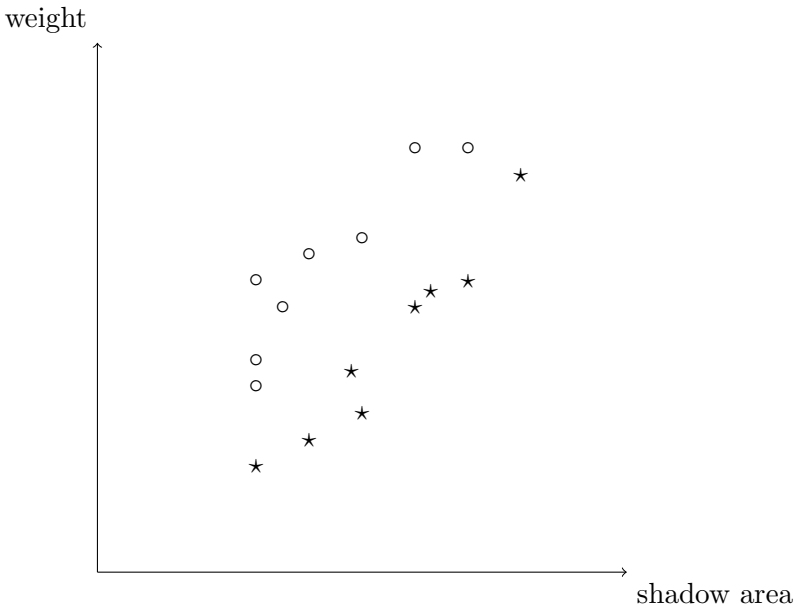

Figure 2.2: Rabbit or weasel?

Obviously, neither of the characteristics (weight or shadow area) alone can be used to tell apart rabbits from weasels. Our hope is that we can find a line which seperates the circles from the stars in the sense that all stars lie on one side of the line, and all circles lie on the other.

Mathematically speaking, we are given points

$$p_i = (x_i, y_i) \in \mathbb{R}^2 \ (i = 1, \ldots, m) \quad \textit{and} \quad q_j = (x_j, y_j) \in \mathbb{R}^2 \ (j = 1, \ldots, n)$$

and we try to find a linear classfier*, i.e. an affine linear mapping $x \in \mathbb{R} \mapsto ax + b$ with $a, b \in \mathbb{R}$ such that*

$$\begin{aligned} y_i &> ax_i + b \quad (i = 1, \ldots, m), \\ y_j &< ax_j + b \quad (j = 1, \ldots, n) \end{aligned}$$

We now introduce a new variable $\delta \in \mathbb{R}$, which measures the gap *between the left and right hand side of each inequality above.*

This gap we try to make as large as possible ending up with the linear program

$$\begin{aligned} \max_{(\delta,a,b)} \quad \delta \quad \text{s.t.} \qquad y_i &\geq ax_i + b + \delta \quad (i = 1, \ldots, m), \\ y_j &\leq ax_j + b + \delta \quad (j = 1, \ldots, n). \end{aligned} \tag{2.2}$$

$\diamond$

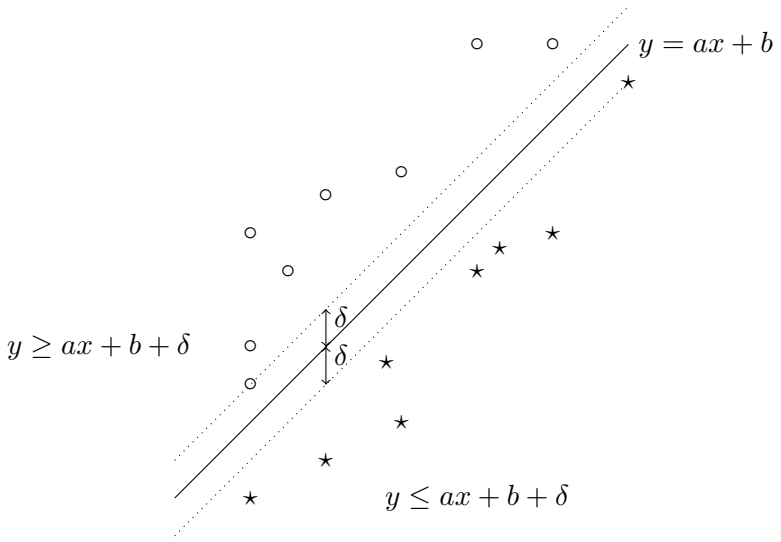

Figure 2.3: Linear classifier

We point out that if we are provided with $d > 2$ characteristics (in addition to weight and area of shadow) our apporach still works: Lines simply translate to hyperplanes in $\mathbb{R}^d$.

We now return to our theroretical study of linear programs: For notational convenience and conceptual uniformity, we prefer a particular form of a linear program which merits its own definition.

Definition 2.1.3 (LP standard form) *A linear program*

$$\min c^T x \quad s.t. \quad Ax = b,\ x \geq 0, \tag{2.3}$$

with $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ *and* $c \in \mathbb{R}^n$ *is said to have* standard form.

We will now show that an arbitrary LP can be transformed to standard form:

Consider a single constraint

$$m_j^T x \leq b_j$$

for some vector $m_j \in \mathbb{R}^n$ and a scalar $b_j \in \mathbb{R}$. Introducing a *slack variable* $s_j \geq 0$ this can be written as

$$m_j^T x + s_j = b_j,\ \ s_j \geq 0.$$

Inequalities of the form

$$m_j^T x \geq b_j,$$

are dealt with analogously, since it can be written as

$$-m_j^T x \leq -b_j.$$

In case a variable x_i has no sign restriction (we call this a *free variable*) we split x_i in its *positive part* x_i^+ and its *negative part* x_i^- such that

$$x_i = x_i^+ - x_i^-, \;\; x_i^+ \geq 0, \;\; x_i^- \geq 0.$$

Finally, every maximization problem

$$\max \;\; c^T x \quad \text{u.d.N.} \quad Ax = b, \;\; x \geq 0$$

is obviously equivalent to the minimization problem

$$\min \;\; -c^T x \quad \text{u.d.N.} \quad Ax = b, \;\; x \geq 0,$$

in the sense that $\bar{x}$ solves one of the above problems if and only if it solves the other.

We illustrate the above reformulation techniques in the following example.

Example 2.1.4 *Consider the LP*

$$\begin{array}{lll} \max \; x_1 - x_2 - x_3 & s.t. & x_1 + 2x_2 \leq 0, \\ & & x_1 + x_2 + x_3 \geq 0, \\ & & 2x_1 + x_3 = 1, \\ & & x_1 \geq 0, \; x_2 \geq 0. \end{array} \tag{2.4}$$

Introducing slack variables for the first two inequalities and changing the siqn in the objective function yields the optimization problem

$$\begin{array}{lll} \min \; -x_1 + x_2 + x_3 & s.t. & x_1 + 2x_2 + s_1 = 0, \\ & & x_1 + x_2 + x_3 - s_2 = 0, \\ & & 2x_1 + x_3 = 1, \\ & & x_1 \geq 0, \; x_2 \geq 0, \; s_1 \geq 0, \; s_2 \geq 0. \end{array}$$

equivalent to (2.4). *Splitting the free variables x_3 in its positive and negative part yields the equivalent program*

$$\begin{array}{lll} \min \; -x_1 + x_2 + x_3^+ - x_3^- & s.t. & x_1 + 2x_2 + s_1 = 0, \\ & & x_1 + x_2 + x_3^+ - x_3^- - s_2 = 0, \\ & & 2x_1 + x_3^+ - x_3^- = 1, \\ & & x_1 \geq 0, \; x_2 \geq 0, \; x_3^+ \geq 0, \; x_3^- \geq 0, \; s_1 \geq 0, \; s_2 \geq 0, \end{array}$$

which has standard form. ◇

The above remarks show that, without loss of generality, we can always assume that a linear program has standard form.

2.2 Polyhedra in standard form

In this section we take a closer look at the geometric and algebraic properties of the feasible set of an LP in standard form.

Definition 2.2.1 (Polyhedra in standard form) *For $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$ we call the set*

$$P := P(A,b) := \{x \in \mathbb{R}^n \mid Ax = b,\ x \geq 0\}$$

a polyhedron in standard form.

With the above terminology, an LP in standard form is the minimization of a linear functional subject to a polyhedron in standard form as its feasible set.

Basic points of poyhedra in standard form

The next result gives a characterization of basic feasible points of a polyhedron in standard form under the assumption that the defining matrix has full row rank.

Note that, by Definition 1.3.15, $\bar{x}$ is a basic point of $P = P(A,b)$ if $A\bar{x} = b$ and there exist n linearly indpendent constraints among '$Ax \overset{!}{=} x, x \overset{!}{\geq} 0$' active at $\bar{x}$.

Theorem 2.2.2 (Basic points of polyhedra in standard form) *Let $P = P(A,b)$ be a polyhedron in standard form defined by $A = [a_1, \ldots, a_n] \in \mathbb{R}^{m\times n}$ such that* $\operatorname{rank} A = m$ *(in particular $n \geq m$). Then for $\bar{x} \in \mathbb{R}^n$ the following are equivalent:*

i) $\bar{x}$ is a basic point of P.

ii) We have $A\bar{x} = b$ and there exists $J \subset \{1, \ldots, n\}$ with $|J| = m$ such that

I) the vectors a_j $(j \in J)$ are linearly independent;

II) $\bar{x}_j = 0$ if $j \notin J$.

Proof: 'ii)⇒i)': By assumption we have

$$b = A\bar{x} = \sum_{j\in J} \bar{x}_j a_j.$$

Since the vectors a_j $(j \in J)$ are linearly independent, the linear system

$$b = Ax, \quad e_j^T x = 0\ (j : \bar{x} = 0)$$

has a unique solution. By the active constraint lemma, there hence exist n linearly independent constraints active at $\bar{x}$, i.e. $\bar{x}$ is a basic point of P.

'i)⇒ii)': Assume that $\bar{x}$ is a basic solution of P, thus there exist m linearly independent constraints among

$$Ax = b, \quad x_j \geq 0 \ (j = 1, \ldots, n)$$

active at $\bar{x}$. By the active constraint lemma, the linear system (of the constraints active at $\bar{x}$)

$$Ax = b, x_j = 0 \ (j : \bar{x}_j = 0)$$

has hence a unique solution. Equivalently, the system

$$\sum_{j:\bar{x}_j \neq 0} x_j a_j = b$$

has a unique solution. Therefore, the vectors a_j $(j : \bar{x} \neq 0)$ are linearly independent. Now, set $\hat{J} := \{j \mid \bar{x}_j \neq 0\}$. If $|\hat{J}| = m$, the proof is finished by setting $J := \hat{J}$. Otherwise, since A has m linearly independent rows, hence also m linearly independent columns (row rank = column rank), we simply complete $\left\{a_i \;\middle|\; i \in \hat{J}\right\}$ to a basis of span $\{a_1, \ldots, a_n\}$ and the resulting index set fulfills all criteria.

□

The full rank assumption on A

The crucial assumption for the characterization of basic points of polyedra in standard form is that the defining matrix has full (row) rank. This assumption will also be critical later on, but fortunately the next result shows that we can assume without loss of generality that it is fulfilled.

Proposition 2.2.3 (Full rank assumption) *Let $P = P(A, b)$ be a nonempty polyhedron in standard form with*

$$A = \begin{pmatrix} - & a_1^T & - \\ & \vdots & \\ - & a_m^T & - \end{pmatrix} \in \mathbb{R}^{m \times n}$$

such that $\operatorname{rank} A = k < m$ *and $a_{i_1}, \ldots, a_{i_k}$ are linearly independent. Then for (the polyhedron in standard form)*

$$Q := \left\{x \in \mathbb{R}^n \;\middle|\; a_{i_j}^T x = b_{i_j} \ (j = 1, \ldots, k), \quad x \geq 0\right\}$$

we have $Q = P$.

Proof: W.l.o.g. the first k row vectors $a_1, \ldots, a_k$ of A are linearly independent (otherwise rearrange the rows): The inclusion $P \subset Q$ is obvious since P has all the constraints of Q.

For the converse inclusion it suffices to show that $x \in Q$ satsifies

$$a_j^T x = b_j \quad (j = k+1, \ldots, m). \tag{2.5}$$

Since $\operatorname{rank} A = k$ and $a_1, \ldots, a_k$ are linearly independent, the latter vectors are a basis for $\operatorname{span}\{a_1, \ldots, a_m\}$, in particular

$$a_j \in \operatorname{span}\{a_1, \ldots, a_k\} \quad (j = k+1, \ldots, m).$$

Hence, for all $j = k+1, \ldots, m$ there exist scalars $\lambda_1^j, \ldots, \lambda_k^j$ such that

$$a_j = \sum_{i=1}^{k} \lambda_i^j a_i.$$

With $\hat{x} \in P$m which exists as P is nonempty, it follows that

$$b_j = a_j^T \hat{x} = \sum_{i=1}^{k} \lambda_i^j a_i^T \hat{x} = \sum_{i=1}^{k} \lambda_i^j a_i^T x = a_j^T x \quad (j = k+1, \ldots, m),$$

which proves (2.5), and hence completes the proof. □

2.3 The fundamental theorem of linear programming

Recall from Example 2.1.1 that the solution of the diet problem lies in an extreme point of the feasible set. This is part of the *fundamental theorem of linear programming* that we prove now.

Theorem 2.3.1 (Fundamental Theorem of Linear Programming) *Let $P = P(A, b)$ be a polyhedron in standard form with $A \in \mathbb{R}^{m \times n}$. Then the following hold:*

a) If P is nonempty it has a basic feasible point.

b) P has at most finitely many basic feasible points.

c) If the linear program

$$\min c^T x \quad \text{u.d.N.} \quad x \in P \tag{2.6}$$

has a solution then also a basic feasible point of P is a solution of (2.6).

Proof:

a) If $0 \in P$ then obviously $\bar{x} = 0$ is a basic feasible point of P. Otherwise choose $\bar{x} \in P$ such that the number of positive entries is minimal. Then the index set $\hat{J} := \{j \,|\, \bar{x}_j > 0\}$ is nonempty. We claim that the column vectors $a_j \, (j \in \hat{J})$ are linearly independent: If this were not the case there exist scalars $\gamma_j \, (j \in \hat{J})$ such that

$$\sum_{j \in \hat{J}} \gamma_j a_j = 0 \tag{2.7}$$

and $\gamma_j \neq 0$ for at least one $j \in \hat{J}$. W.l.o.g. we can assume that $\gamma_j < 0$ for at least one index $j \in \hat{J}$ (otherwise we simply multiply equation (2.7) by -1). As $\bar{x}_j > 0$ for all $j \in \hat{J}$ we have

$$x_j(\delta) := \bar{x}_j + \delta\gamma_j \geq 0$$

for all $j \in \hat{J}$ and all $\delta > 0$ sufficiently small. Since $\gamma_j < 0$ for at least one index $j \in \hat{J}$ there exists a minimal $\bar{\delta} > 0$ such that $x_j(\bar{\delta}) \geq 0$ for all $j \in \hat{J}$ and $x_j(\bar{\delta}) = 0$ for at least one index $j \in \hat{J}$. The vector $\bar{x}$ defined componentwise by

$$x_j^* := \begin{cases} x_j(\bar{\delta}), & \text{if } j \in \hat{J}, \\ 0, & \text{else} \end{cases}$$

lies in P as $\bar{x} \geq 0$ and

$$A\bar{x} = \sum_{j=1}^{n} x_j^* a_i = \sum_{j \in \hat{J}} \bar{x}_j a_j + \bar{\delta} \sum_{j \in \hat{J}} \gamma_j a_j = A\bar{x} = b.$$

On the other hand, $\bar{x}$ has, by construction, one zero component more than $\bar{x}$, which contradicts the choice of $\bar{x}$. Hence, nonvanishing scalars in the sense of (2.7) cannot exist, therefore the vectors $a_j \, (j \in \hat{J})$ are linearly independent.

If $|\hat{J}| = m$ Theorem 2.2.2 yields that $\bar{x}$ is a basic feasible point. Otherwise, we simply complete $a_j \, (j \in \hat{J})$ by vectors from $a_j \, (j \notin \hat{J})$ and the resulting index set fulfills the properties of Theorem 2.2.2 ii).

b) See Exercise 4.

c) By assumption, the optimal value

$$f^* := \inf\{c^T x \,|\, x \in P\}$$

of (2.6) is finite. Consider the modified LP

$$\min c^T x \quad \text{s.t.} \quad x \in \bar{P} \tag{2.8}$$

with

$$\bar{P} := \{x \in \mathbb{R}^n \mid Ax = b, c^T x = f^*, x \geq 0\}.$$

Then $\bar{P}$ is nonempty, since we assume that (2.6) has a solution, which lies in $\bar{P}$. If c is linearly dependent of the row vectors of A we have $P = \bar{P}$ (cf. the proof of Proposition 2.2.3). Hence, every feasible point, in particular every basic feasible point of (2.6) is a solution of it.

If, in turn, c is linearly indpendent of the rows of A, we can apply part a) to $\bar{P}$ (as $\bar{A} := \begin{pmatrix} A \\ c^T \end{pmatrix} \in \mathbb{R}^{m+1 \times n}$ has rank $m+1$) and infer that $\bar{P}$ has a basic feasible point, say $\bar{x}$. Due to Theorem 1.3.17, $\bar{x}$ is an extreme point of $\bar{P}$. We claim that $\bar{x}$ is already an extreme point of P: To this end, assume this were not the case. Then there exist $x_1, x_2 \in P$ with $x_1 \neq x_2$ and $\lambda \in (0,1)$ such that

$$\bar{x} = \lambda x_1 + (1-\lambda)x_2.$$

Since $x_1, x_2 \in P$ we have

$$f^* \leq c^T x_1 \quad \text{and} \quad f^* \leq c^T x_2.$$

On the other hand $\bar{x} \in \bar{P}$, hence $f^* = c^T \bar{x}$. This implies

$$f^* = c^T x_1 = c^T x_2,$$

thus, $x_1, x_2 \in \bar{P}$. As $\bar{x}$ is an extreme point of $\bar{P}$ this yields $x_1 = x_2$ which contradicts the choice of x_1 and x_2. Hence, $\bar{x}$ is an extreme, hence a basic feasible point of P. Since $\bar{x} \in \bar{P}$ we have

$$c^T \bar{x} = f^* \leq c^T x \quad (x \in P),$$

i.e. $\bar{x}$ solves (2.6).

□

We would like to point out that a polyhedron $P(A,b)$ in standard form can be seen to have a basic feasible point in the sense of Definition 1.3.15 even without the full rank assumption on A, cf. [1, Corollary 2.2], but in order to get the full characterization in terms of Theorem 2.2.2, we need the full rank assumption.

2.4 Duality theory

In this section we establish a *duality theory* for linear programming which in combination with the fundamental theorem (Theorem 2.3.1) forms the conceptual backbone of linear optimization and will be exploited throughout for theoretical and computational purposes.

2.4.1 Motivation

Consider the equality constrained optimization problem

$$\min f(x) \quad \text{s.t.} \quad g_i(x) = 0 \ (i = 1, \dots, m), \tag{2.9}$$

where $f, g_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \dots, m)$ are continuously differentiable. In multivariate calculus the method of Lagrange (local) multipliers is used to determine (local) minimizers of (2.9).

Theorem 2.4.1 (Method of Lagrange multipliers) *Let $\bar{x}$ be (local) minimizer of (2.9) such that the vectors $\nabla g_1(\bar{x}), \dots, \nabla g_m(\bar{x})$ are linearly independent. Then there exists $\bar{\lambda} \in \mathbb{R}^m$ such that*

$$0 = \nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_i \nabla g_i(\bar{x}).$$

The underlying idea of the method of Lagrange multipliera is as follows: Instead of dealing with the constrained problem, i.e. enforcing the equality constraints to hold, we allow for a violation of the constraints but associate each constraint with a Lagrange multiplier λ_i. This leads to the *Lagrangian*

$$L : (x, \lambda) \mapsto f(x) + \lambda^T g(x).$$

of (2.9). Minimizing L with respect to x (while fixing λ) yields the optimality condition

$$0 = \nabla_x L(x, \lambda) \quad \Longleftrightarrow \quad \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0.$$

We then look for a multiplier λ such that the solution of the above system is also feasible for the constrained problem.

We now apply this idea of 'incorporating the hard equality constraints into the objective function by introducing a multiplier for each constraint' to a linear program in standard form

$$\min \ c^T x \quad \text{s.t.} \quad Ax = b, \ x \geq 0, \tag{2.10}$$

which we call the *primal problem*. Assume that $\bar{x}$ solves (2.23), i.e.

$$c^T \bar{x} = \inf \left\{ c^T x \mid Ax = b, \ x \geq 0 \right\} =: f^*.$$

We now introduce the *relaxed problem*

$$\min \ c^T x + \lambda^T (b - Ax) \quad \text{s.t.} \quad x \geq 0. \tag{2.11}$$

We define

$$g : \lambda \mapsto \inf \left\{ c^T x + \lambda^T (b - Ax) \mid x \geq 0 \right\} \in [-\infty, \infty)$$

as the *optimal value function* of (2.11). Clearly, since, $\bar{x}$ is feasible for both (2.23) and (2.11), we have

$$g(\lambda) \leq c^T\bar{x} + \lambda^T(b - A\bar{x}) = c^T\bar{x} = f^* \quad (\lambda \in \mathbb{R}^m).$$

In other words, the function g is everywhere a lower bound for the optimal value f^* of the primal problem (2.23). Hence, the problem

$$\max_{\lambda\in\mathbb{R}^m} \; g(\lambda) \tag{2.12}$$

seeks for the tightest (i.e. largest) lower bound in this sense. By definition of g, we have

$$g(\lambda) = \inf\left\{c^Tx + \lambda^T(b - Ax) \mid x \geq 0\right\} = \lambda^Tb + \inf\left\{(c - A^T\lambda)^Tx \mid x \geq 0\right\}.$$

Now, observe that

$$\inf\left\{(c - A^T\lambda)^Tx \mid x \geq 0\right\} = \begin{cases} 0 & \text{if} \quad c - A^T\lambda \geq 0, \\ -\infty & \text{else}, \end{cases}$$

cf. Exercise 5. Hence, (2.12) can be written as

$$\max_{\lambda\in\mathbb{R}^m} \; b^T\lambda \quad \text{s.t.} \quad A^T\lambda \leq c. \tag{2.13}$$

This problem is called the *dual problem* to (2.23).

2.4.2 Duality and optimality in linear programming

Motivated by the foregoing section consider the primal linear program

$$\min \; c^Tx \quad \text{s.t.} \quad Ax = b, \; x \geq 0. \tag{P}$$

and its dual program

$$\max \; b^Ty \quad \text{s.t.} \quad A^Ty \leq c. \tag{D}$$

uniquely defined by the data $(A, b, c) \in \mathbb{R}^{m\times n} \times \mathbb{R}^m \times \mathbb{R}^n$. We put

$$\inf(P) := \inf\left\{c^Tx \mid Ax = b, \; x \geq 0\right\} \in [-\infty, +\infty]$$

and

$$\sup(D) := \sup\left\{b^Ty \;\middle|\; A^Ty \leq c\right\} \in [-\infty, +\infty].$$

Note that by our convention for infima and suprema from Section 1.2.1, we have that $\inf(D) = \infty$ if (P) has no feasible points (i.e. is infeasible), and analogously we have $\sup(D) = -\infty$ if (D) is infeasible.

From the way we derived the dual program the next result follows immediately. However, since there is a very elementary independent proof, we provide it for completeness.

Theorem 2.4.2 (Weak duality theorem) *Let $x \in \mathbb{R}^n$ be feasible for (P), and let $y \in \mathbb{R}^m$ be feasible for (D). Then*

$$b^T y \leq c^T x,$$

In particular, we have

$$\sup(D) \leq \inf(P).$$

Proof: By feasibilty of x and (y, s) for (P) and (D), respectively, we infer that

$$b^T y = (Ax)^T y = x^T (A^T y) \leq c^T x,$$

since $x \geq 0$ and $A^T y \leq c$. □

We call (P) *unbounded* if $\inf(P) = -\infty$ and, analogously, we call (D) unbounded if $\sup(D) = +\infty$. From the weak duality the next result about unbounded LPs follows immediately, cf. Exercise 8

Corollary 2.4.3 (Unbounded LPs) *Let (P) and (D) be defined as above. Then the following hold:*

a) If (P) is unbounded then (D) is infeasible.

b) If (D) is unbounded then (P) is infeasible.

In particular, (P) and (D) cannot both be unbounded.

Given $(A, b, c) \in \mathbb{R}^{m \times n} \times \mathbb{R}^m \times \mathbb{R}^n$ we define

$$\Delta(A, b, c) := \inf(P) - \sup(D)$$

and call it the *duality gap* between the optimal primal and dual value. In view of the weak duality theorem, and using the convention that $\infty + \infty = \infty$, we immediately see that

$$\Delta(A, b, c) \geq 0 \quad \left((A, b, c) \in \mathbb{R}^{m \times n} \times \mathbb{R}^m \times \mathbb{R}^n\right).$$

From the weak duality theorem we immediately obtain a sufficient optimality condition, namely a zero duality gap, for both the primal and dual program.

Corollary 2.4.4 (Sufficient optimality condition for LP) *Let $\bar{x} \in \mathbb{R}^n$ be feasible for (P) and $\bar{y} \in \mathbb{R}^m$ be feasible for (D) such that*

$$c^T \bar{x} = b^T \bar{y}.$$

Then $\bar{x}$ solves (P) and $\bar{y}$ solves (D).

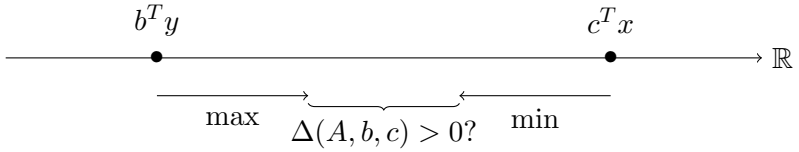

Figure 2.4: Duality gap in linear programming

Proof: Let $x \in \mathbb{R}^n$ be feasible for (P). By weak duality and the assumptions on $\bar{x}$ and $\bar{y}$, we have

$$c^T \bar{x} = b^T \bar{y} \leq c^T x.$$

Since x was an arbitrary feasible point of (P), we find that $\bar{x}$ solves (P).

The statement for the dual program can be proven analogously.

□

We now want to show that also the converse implication in Corollary 2.4.4 holds true and that the primal program is solvable if and only if the dual is. The key result at this is the *Farkas Lemma*. Its proof relies on separation of closed convex sets from points in its complement, see Theorem 1.3.9. In fact, the convex set that we are dealing with here is also even convex *cone*. Recall from Example 1.3.4 c) that a set $K \in \mathbb{R}^n$ is said to be a cone if

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

Proposition 2.4.5 (Farkas Lemma) *Let $B \in \mathbb{R}^{m \times n}$ and $h \in \mathbb{R}^n$. Then the following statements are equivalent:*

i) The polyhedron $P = \left\{ x \in \mathbb{R}^m \mid B^T x = h,\ x \geq 0 \right\}$ is nonempty.

ii) We have $h^T d \geq 0$ for all $d \in \mathbb{R}^n$ with $Bd \geq 0$.

Proof: 'i)⇒ii)': Take $x \in P$ and d such that $Bd \geq 0$. Then

$$h^T d = (B^T x)^T d = x^T (Bd) \geq 0,$$

which proves the first implication.
'ii)⇒i)': (Contraposition) Suppose i) does not hold. Then $h \in \mathbb{R}^n$ is not an element of the closed, convex cone $K = \left\{ B^T x \mid x \geq 0 \right\} \subset \mathbb{R}^n$. Hence, by basic separation, we find $s \in \mathbb{R}^n$ such that

$$s^T h < s^T y \quad (y \in K).$$

From this we infer that

$$s^T h < 0 \leq s^T y \quad (y \in K). \tag{2.14}$$

Here, the first inequality is due to the fact that $0 \in K$, and the second one must hold since K is a cone, i.e. $\lambda y \in K$ for any $y \in K$ and $\lambda \geq 0$. Choosing $y = b_i = B^T e_i \in K$ $(i = 1, \ldots, m)$ (the columns of B^T) in (2.14), yields

$$s^T h < 0 \leq s^T b_i \quad (i = 1, \ldots, m),$$

i.e. $Bs \geq 0$ but $s^T h < 0$, i.e. ii) does not hold, which concludes the proof. □

We are now in a position to show the main result of this section, which was already foreshadowed above.

Theorem 2.4.6 (Strong duality theorem) *The primal program (P) has a solution $\bar{x}$ if and only if the dual program has a solution $\bar{y}$. In this case we have $c^T \bar{x} = b^T \bar{y}$, i.e. there is no duality gap.*

Proof: As for the first statement we only prove here that solvability of the dual program implies solvabilty of the primal program. The converse implication follows from this by observing that the dual of the dual program is the primal program, cf. Exercise 6.

To this end, let $\bar{y}$ solve the dual program

$$\max\ b^T y \quad \text{s.t.} \quad A^T y \leq c. \tag{D}$$

Let $a_1, \ldots, a_n \in \mathbb{R}^m$ be the column vectors of A. Then (D) is equivalent to

$$\max\ b^T y \quad \text{s.t.} \quad a_i^T y \leq c_i \ \ (i = 1, \ldots, n). \tag{2.15}$$

We define the index set $\bar{I} := \left\{ i \mid a_i^T \bar{y} = c_i \right\}$ of active constraints at $\bar{y}$, put $l := |\bar{I}|$ and consider two cases:

$l = 0$: In this case, we must have $b = 0$, since otherwise we could perturb $\bar{y}$ slightly to get a better objective value and still stay feasible. Hence, in this case, $\bar{x} := 0$ is feasible for the primal program (P) and we have $c^T \bar{x} = 0 = b^T \bar{y}$. Due to Corollary 2.4.4, $\bar{x}$ solves (P). Moreover, both programs have the same optimal value.

$l > 0$: We define a matrix $B \in \mathbb{R}^{l \times m}$ by

$$B := \begin{pmatrix} & \vdots & \\ — & a_i^T \ (i \in \bar{I}) & — \\ & \vdots & \end{pmatrix}.$$

Now, let d be any vector with $Bd \leq 0$. Then we have $a_i^T(\bar{y} + \varepsilon d) \leq c_i$ for all $i \in \bar{I}$ and $\varepsilon > 0$. Moreover, since $a_i^T \bar{y} < c_i$ for $i \notin \bar{I}$, we have $a_i^T(\bar{y} + \varepsilon d) \leq c_i$ for all $i \notin \bar{I}$ and $\varepsilon > 0$ sucfficiently small. Hence, $\bar{y} + \varepsilon d$ is feasible for (2.15). Since (2.15) is a maximization problem which is solved by $\bar{y}$ we must have $b^T(\bar{y} + \varepsilon d) \leq b^T y$. Putting $h = -b$, this implies $h^T d \geq 0$ for any d with $Bd \leq 0$. The Farkas Lemma (with $-B$ instead of B) hence yields that there exists a vector $\hat{x} \in \mathbb{R}^l$ with $\hat{x} \geq 0$ and $-B^T \hat{x} = h$, where the latter reads

$$\sum_{i \in \bar{I}} \hat{x}_i a_i = b. \tag{2.16}$$

Now define the vector $\bar{x} \in \mathbb{R}^n$ componentwise by

$$\bar{x}_i := \begin{cases} \hat{x}_i & \text{if} \quad i \in \bar{I}, \\ 0 & \text{if} \quad i \notin \bar{I}. \end{cases}$$

Then, as $\hat{x} \geq 0$, we have $\bar{x} \geq 0$. Moreover, due to (2.16) and the definition of $\bar{x}$, we have

$$A\bar{x} = \sum_{i \in \bar{I}} \hat{x}_i a_i = b,$$

hence, all in all, $\bar{x}$ is feasible for the primal program (P). Using the definitions of $\bar{x}$ and $\bar{I}$, respectively, as well as (2.16), we compute that

$$c^T \bar{x} = \sum_{i \in \bar{I}} c_i \hat{x}_i = \sum_{i \in \bar{I}} (a_i^T y) \hat{x}_i = \left(\sum_{i \in \bar{I}} \hat{x}_i a_i \right)^T \bar{y} = b^T \bar{y}.$$

From Corollary 2.4.4 we thus infer infer that $\bar{x}$ solves (P) and the objective values of (P) and (D) are equal.

This completes the proof.

□

We derived the strong duality theorem using the Farkas Lemma. In fact, both statements are equivalent, see Exercise 7.

The next result, which is a consequence of strong duality, gives a characterization for feasible points of the primal and dual program, respectively, to be solutions of the respective program in terms of a *complementarity* condition.

Corollary 2.4.7 (Complementarity of primal and dual solutions) *Let $\bar{x}$ be feasible for the primal program (P) and $\bar{y}$ for the dual program (D). Then the following are equivalent:*

i) $\bar{x}$ solves (P) and $\bar{y}$ solves (D).

ii) We have

$$[c - A^T\bar{y}]_i \cdot \bar{x}_i = 0 \quad (i = 1, \ldots, n). \tag{2.17}$$

Proof: 'i)⇒ ii)': By the feasibilty of $\bar{x}$ and $\bar{y}$ for (P) and (D), respectively, we have $\bar{x}_i \geq 0$ and $[c - A^T\bar{y}]_i \geq 0$ for all $i = 1, \ldots, n$. This implies

$$[c - A^T\bar{y}]_i \cdot \bar{x}_i \geq 0 \quad (i = 1, \ldots, n). \tag{2.18}$$

Since $\bar{x}$ and $\bar{y}$ solve the respective programs, from the strong duality theorem it follows that there is no duality gap, hence

$$\begin{aligned} 0 &= c^T\bar{x} - b^T\bar{y} \\ &= c^T\bar{x} - (A\bar{x})^T\bar{y} \\ &= (c - A^T\bar{y})^T\bar{x} \\ &= \sum_{i=1}^{n} [c - A^T\bar{y}]_i \cdot \bar{x}_i. \end{aligned}$$

Since every summand in the above sum is nonnegative, see (2.18), and its sum is zero, each single summand must be zero, which proves (2.17) and thus the desired implication.

'ii)⇒i):' From (2.17) we infer

$$\begin{aligned} c^T\bar{x} &= \sum_{i=1}^{n} c_i\bar{x}_i \\ &= \sum_{i=1}^{n} \left[A^T\bar{y}\right]_i \bar{x}_i \\ &= \bar{x}^T A^T\bar{y} \\ &= b^T\bar{y}. \end{aligned}$$

By Corollary 2.4.4 (and feasibility) this implies that $\bar{x}$ solves (P) and $\bar{y}$ solves (D). This concludes the proof.

□

As of now we know how to express unboundedness and feasibility of the the primal linear program (P) and its dual (D) in terms of the extended-real valued numbers $\inf(D) \geq \sup(D)$. However, the question remains if it is also possible to express solvability of either program in terms of these values. The answer to this is given in the next theorem, which also can be viewed as an existence result for solutions of linear programs, which we have not provided yet.

Theorem 2.4.8 (Existence of LP solutions) *Let (P) and (D) be given as above. Then the following statements hold:*

a) $\inf(P) \in \mathbb{R}$ *if and only if (P) has a solution.*

b) $\sup(D) \in \mathbb{R}$ *if and only if (D) has a solution.*

Proof: We only proof part a). Part b) is left to the reader as an exercise: Clearly, if (P) has a solution $\bar{x}$ then $\inf(P) = c^T\bar{x} \in \mathbb{R}$.

To show the converse implication, let $f^* := \inf(P) \in \mathbb{R}$ and assume that (P) has no solution, i.e.

$$c^T x > f^* \quad (x : Ax = b,\ x \geq 0). \tag{2.19}$$

Thus, the system

$$\begin{pmatrix} c^T \\ -A \end{pmatrix} x = \begin{pmatrix} f^* \\ -b \end{pmatrix}$$

has no solution $x \geq 0$. Applying the Farkas Lemma to

$$B := \begin{pmatrix} c^T \\ -A \end{pmatrix} \quad \text{and} \quad h := \begin{pmatrix} f^* \\ -b \end{pmatrix}$$

yields $d = \binom{y}{\alpha} \in \mathbb{R}^{m+1}$ such that

$$0 > h^T d = \alpha f^* - b^T y \tag{2.20}$$

and

$$0 \leq Bd = \alpha c - A^T y. \tag{2.21}$$

Since $\inf(P) \in \mathbb{R}$ there exists a point $x \in \mathbb{R}^n$ feasible for (P), i.e. $Ax = b$ and $x \geq 0$. Multiplying (2.21) by x^T hence yields

$$\alpha c^T x - b^T y \geq 0.$$

By (2.20) we therefore have

$$\alpha c^T x > \alpha f^*,$$

which, in view of (2.19), implies $\alpha > 0$. Dividing (2.20) and (2.21) by α yields

$$f^* < b^T \bar{y} \quad \text{and} \quad A^T \bar{y} \leq c.$$

for $\bar{y} := \frac{y}{\alpha}$. Hence, $\bar{y}$ is dually feasible with a dual objective value strictly greater than the optimal primal value f^*. This contradicts the weak dualtiy theorem and hence the proof is complete. □

Exercises to Chapter 2

1. **(LP standard form)** Bring the linear program

$$\begin{aligned} \max\ x_1 + 2x_2 + 3x_3 \quad \text{s.}t. \quad & x_1 + x_3 \geq 0, \\ & x_1 - x_3 \leq 0, \\ & x_1 + x_2 = 1, \\ & x_1, x_2 \geq 0. \end{aligned}$$

   to standard form

$$\min c^T x \quad \text{u.d.N.} \quad Ax = b,\ x \geq 0,$$

   with $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$ and write down A, b, c and $n, m \in \mathbb{N}$ explicitly.

2. **(LP modelling)** A hauling company has trucks (of the same size) a the locations A and B, namely 18 at location A and 12 at location B. These are to be sent to the terminals R,S and T where there are 11, 10 and 9 trucks needed, respectively. The distances (in km) from the locations to the terminals are given in the following table:

	Terminal R	Terminal S	Terminal T
Location A	5	4	9
Location B	7	8	10

   The trucks are to be routed such that the kilometers driven are minimal and such that the needs at every terminal are met.

   a) Formulate this problem as a linear program.

   b) Use linear algebra to reduce it to a two-dimensional problem in order to solve it graphically.

3. **(Discrete Chebyshev approximation)**

   For $A \in \mathbb{R}^{m\times n}$ $(m \geq n)$ and $b \in \mathbb{R}^m$ consider the optimization problems

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|_\infty. \tag{2.22}$$

   where $\|z\|_\infty := \max_{i=1,\ldots,m} |z_i|$ $(z \in \mathbb{R}^m)$ and

$$\min_{(x,\delta)\in\mathbb{R}^n\times\mathbb{R}} \delta \quad \text{u.}d.N. \quad -\delta e \leq Ax - b \leq \delta e \tag{2.23}$$

   where $e := (1, \ldots, 1)^T \in \mathbb{R}^m$.

   a) Show that (2.22) and (2.23) are equivalent in the sense that $\bar{x}$ solves (2.22) if and only if $(\bar{x}, \bar{\delta})$ with $\bar{\delta} := \|A\bar{x} - b\|_\infty$ solves (2.23).

b) Write (2.23) as an LP in standard form.

*c) In the *discrete Chebyshev approximation*, given a function $f : [a, b] \to \mathbb{R}$, we try to approximate f by a polynomial of maximum degree $r \in \mathbb{N}$ by solving the problem

$$\min_{p \in \Pi_r} \max_{i=0,1,\ldots,s} |f(t_i) - p(t_i)| \tag{2.24}$$

where

$$\Pi_r := \left\{ p : \mathbb{R} \to \mathbb{R} \;\middle|\; \exists\, a_0, a_1, \ldots, a_r :\; p(t) = \sum_{i=0}^{r} a_i t^i \;(t \in \mathbb{R}) \right\}.$$

Write (2.24) in the form of (2.22).

4. **(Completing the proof of the fundamental theorem)** Let $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$ with $\operatorname{rank} A = m$ and let $P = P(A, b)$ be the polyhedron in standard form defined by (A, b). Show that P has at most finitely many basic points.

5. **(Derivation of the dual problem)** For $v \in \mathbb{R}^n$ compute $\inf \left\{ v^T x \mid x \geq 0 \right\}$.

6. **(Dual of the dual is the primal program)** For $(A, b, c) \in \mathbb{R}^{m \times n} \times \mathbb{R}^m \times \mathbb{R}^n$ consider the primal linear program

$$\min\; c^T x \quad \text{s.t.} \quad Ax = b,\; x \geq 0. \tag{P}$$

and its dual

$$\max\; b^T y \quad \text{s.t.} \quad A^T y \leq c. \tag{D}$$

Show that the dual of the dual program is (equivalent to) the primal,

7. **(Equivalence of Farkas Lemma and strong duality)** Show that the Farkas Lemma (Proposition 2.4.5) follows from the strong duality theorem (Theorem 2.4.6).

8. **(Duality overview)** Consider the primal program linear program (P) and its dual (D) and show:

   a) If (P) is unbounded then (D) is infeasible.

   b) If (D) is unbounded then (P) is infeasible.

   Afterwards fill in the table below.

	$\inf(P) = -\infty$	$\inf(P) \in \mathbb{R}$	$\inf(P) = +\infty$
$\sup(D) = -\infty$			
$\sup(D) \in \mathbb{R}$			
$\sup(D) = +\infty$			

Use the symbol '✓' if the case for the tuple $(\inf(P), \sup(D))$ can occur, and '×' if this case cannot occur. Verify the positive case by an example, and cite a result from the lecture that applies in the negative case.

9. **(Generalized duality)** For $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$ consider the optimization problem

$$\min\ f(x) \quad \text{s.t.} \quad g(x) \leq 0,\ h(x) = 0. \tag{2.25}$$

Then define the *Lagrangian function*

$$L : (x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \mapsto f(x) + \lambda^T g(x) + \mu^T h(x)$$

and the *dual function*

$$q : (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p \mapsto \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu).$$

We call the optimziation problem

$$\max_{\lambda, \mu}\ q(\lambda, \mu) \quad \text{s.t.} \quad \lambda \geq 0 \tag{2.26}$$

the *dual problem* of (2.25).

a) *(Weak duality)* Let x be feasible for (2.25) and let (λ, μ) be feasible for (2.26). Show that

$$q(\lambda, \mu) \leq f(x).$$

b) *(Failure of strong duality)* Find functions f, g, h such that both (2.25) and (2.26) are feasible but

$$\sup\{q(\lambda, \mu) \mid \lambda \geq 0\} < \inf\{f(x) \mid g(x) \leq 0,\ h(x) = 0\}.$$

3 The Simplex Algorithm

3.1 The simplex iteration

We assume that the given linear program has standard form

$$\min c^T x \quad \text{s.t.} \quad Ax = b,\, x \geq 0 \tag{3.1}$$

with $A \in \mathbb{R}^{m\times n}, c \in \mathbb{R}^n$ und $b \in \mathbb{R}^m$ and we assume w.l.o.g. (cf. Proposition 2.2.3) that $\operatorname{rank} A = m$. We recall that the dual program to (3.1) is given by

$$\max b^T y \quad \text{s.t.} \quad A^T y \leq c, \tag{3.2}$$

see the extensive discussion in Section 2.4. Furthermore, we put

$$P := P(A, b) = \{x \in \mathbb{R}^n \mid Ax = b,\ x \geq 0\}.$$

When searching for solutions of (3.1), the fundamental Theorem 2.3.1, tells us that we can restrict ourselves to basic feasible points of P. The general strategy of the simplex iteration is as follows: *Given a basic feasible point of P, construct another basic feasible point whose objective value is smaller than for the one before.*
Hence, assume that x is a basic feasible point of P and let $J \subset \{1, \ldots, n\}$ be the index set with $|J| = m$ and such that

$$a_j\ (j \in J) \text{ are linearly independent}, \quad x_j = 0\ (i \notin J),$$

which exists by Theorem 2.2.2. We denote the complement of J by K, i.e. $K = \{1, \ldots, n\} \setminus J$ and define the matrices

$$B := B_J := [a_j\ (j \in J)] \in \mathbb{R}^{m\times m} \quad \text{(basis matrix)}$$

and

$$N := N_J := [a_k\ (k \in K)] \in \mathbb{R}^{m\times n-m} \quad \text{(non-basis matrix)}.$$

We think of the index sets J and K as ordered, so that the order of the rows of B and N are fixed. Note that the basis matrix is invertible, since it is square and its columns are linearly independent by construction.

We partition vectors $z \in \mathbb{R}^n$ accordingly by

$$z_J := (z_j)_{j\in J} \in R^m \quad z_K := (z_k)_{k\in K} \in \mathbb{R}^{n-m}.$$

Using these conventions, we have

$$Az = \sum_{j \in J} z_j a_j + \sum_{k \in K} z_k a_k = Bz_J + Nz_K.$$

In particular, for the basic feasible point $x \in P$, we have

$$Bx_J = b \quad \text{and} \quad x_K = 0. \tag{3.3}$$

On order to get used to this new terminology, we consider an example to which we will come back later on.

Example 3.1.1 *Let*

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 4 \\ 6 \\ 2 \\ 3 \end{pmatrix}, \quad c = (-2, -3, -4, 0, 0, 0, 0)^T.$$

It is easily checked that the vector $x = (2, 0, 0, 2, 6, 0, 3)^T$ is a basic feasible basis point of $P = P(A, b)$ with the index set $J = \{1, 4, 5, 7\}$ (thus $K = \{2, 3, 6\}$). Hence, we have

$$B = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad N = \begin{pmatrix} 1 & 1 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$x_J = (2, 2, 6, 3)^T$ und $x_K = (0, 0, 0)^T$. For c we have $c_J = (-2, 0, 0, 0)^T$ and $c_K = (-3, -4, 0)^T$. ◇

We continue with our development of the simplex iteration. In addition to our basic feasible point x of P, we now consider an arbtirary feasible point $z \in P$ in order to compare the objective values of z and x. We choose a representation which gives us some insight as to how to choose z in order to decrease the objective value:

Since $z \in P$ we have

$$Bz_J + Nz_K = Az = b. \tag{3.4}$$

Recalling that B is invertible, see the discussion above, we thus have

$$z_J = B^{-1}b - B^{-1}Nz_K.$$

Inserting this term in the objective function of (3.1) while using (3.3) (and $(A^{-1})^T = (A^T)^{-1}$) yields

$$\begin{aligned} c^T z &= c_J^T z_J + c_K^T z_K \\ &= c_J^T \left(B^{-1}b - B^{-1}Nz_K\right) + c_K^T z_K \\ &= c_J^T x_J + (c_K^T - c_J^T B^{-1} N) z_K \\ &= c^T x + (c_K - N^T (B^T)^{-1} c_J)^T z_K. \end{aligned} \tag{3.5}$$

Now, put

$$y := (B^T)^{-1} c_J \in \mathbb{R}^m,$$

i.e. y is the unique solution of the linear equation

$$B^T y = c_J. \tag{3.6}$$

Herewith, define

$$u_k := c_k - a_k^T y \quad (k \in K). \tag{3.7}$$

With these abbreviations and (3.5) we have

$$c^T z = c^T x + \sum_{k \in K} u_k z_k. \tag{3.8}$$

From (3.8) we know extract a sufficient optimality condition for the basic feasible point x, which will later on be used as a stopping criterion.

Lemma 3.1.2 (Sufficient optimality condition) *Let u_k $(k \in K)$ be defined by* (3.7) *(and* (3.6)*). If*

$$u_k \geq 0 \quad (k \in K) \tag{3.9}$$

then x is a solution of the primal program (3.1)*. Moreover, y solves the dual program* (3.2)*.*

Proof: Since (3.8) holds for any vector z feasible for (3.1), we infer from (3.8), since $z_k \geq 0$ for all $k \in K$, that $c^T z \geq c^T x$ for all $z \in P$, hence x solves (3.1).

From $0 \leq u_k = c_k - a_k^T y$ $(k \in K)$ we infer that $a_k^T y \leq c_k$ for all $k \in K$. Moreover, the definition of y (see (3.6)) yields $a_j^T y = c_j$ for all $j \in J$, hence, all in all $a_j^T y \leq c_j$ for all $j = 1, \ldots, n$, i.e. $A^T y \leq c$. Thus, y is feasible for the dual program (3.2). Due to (3.3) and (3.6) we have

$$c^T x = c_J^T x_J = y^T B x_J = y^T b,$$

hence Corollary 2.4.4 implies that y solves (3.2). □

If (3.9) is satisfied, we can stop our search for a solution, since we know from Lemma 3.1.2 that the (basic feasible) point x solves (3.1). Hence, now assume that this were not the case, assume that

$$u_k < 0 \quad \text{for at least one} \quad k \in K. \tag{3.10}$$

Let $r \in K$ be such an index, i.e. $u_r < 0$. Then from (3.8) we can extract a strategy for how to reduce the the objective value: *Try to find a vector z feasible for* (3.1) *for which $z_k = 0$ for all $k \in K \setminus \{r\}$; if then $z_r > 0$, we have $c^T z < c^T x$.* For this vector $z = z(t)$ $(t \geq 0)$ we hence make the ansatz

$$z_r(t) := t, \quad z_k(t) := 0 \quad (k \in K \setminus \{r\}). \tag{3.11}$$

If $z(t)$ is feasible then by (3.8) we have

$$c^T z(t) = c^T x + t u_r. \tag{3.12}$$

For the ansatz (3.11) the feasibility condition $Az(t) \stackrel{!}{=} b$ yields

$$B z_J(t) + t a_r \stackrel{!}{=} b,$$

cf. (3.4). Thus, we can compute the components for the indices from J by

$$z_J(t) = B^{-1}(b - t a_r) = x_J - t B^{-1} a_r,$$

where we also use (3.3). Defining the vector $d = (d_j)_{j \in J} \in \mathbb{R}^m$ as the unique solution of

$$Bd = a_r, \tag{3.13}$$

this reduces to

$$z_J(t) = x_J - td. \tag{3.14}$$

Then the vector $z \in \mathbb{R}^n$ defined through (3.11) and (3.14) fulfills the condition $Az(t) = b$. We still need to satsify the nonnegativity condition $z(t) \stackrel{!}{\geq} 0$. In a special case, we obtain the unboundedness of (3.1) (i.e. $\inf(P) = -\infty$), which will lead to another stopping criterion:

Lemma 3.1.3 (Unboundedness of (P)) *Let $d \in \mathbb{R}^m$ be defined by* (3.13)*. If*

$$d_j \leq 0 \quad (j \in J) \tag{3.15}$$

then the primal program (3.1) *is unbounded. In particular,* (3.1) *has no solution and* (3.2) *is infeasible.*

Proof: Exercise 1. □

Should we arrive at a (basic feasible) point x such that $d \in \mathbb{R}^m$ (constructed through (3.13) as described above) fulfills $d \leq 0$, Lemma 3.1.3 tells us that we do not need to keep on searching for a solution of the LP (3.1) as it does not exist. Thus, constructing our simplex iteration we can therefore assume that, in addition to (3.10), we have

$$d_j > 0 \quad \text{for at least one} \quad j \in J. \tag{3.16}$$

By the construction of $z(t)$ from above (see (3.11), (3.14)) the missing feasibility condition $z(t) \geq 0$ is going to be satisfied if and only if

$$t \geq 0 \quad \text{and} \quad x_j - td_j \geq 0 \quad (j \in J : d_j > 0),$$

which, in turn, is equivalent to

$$0 \leq t \leq \frac{x_j}{d_j} \quad (j \in J : d_j > 0). \tag{3.17}$$

All in all, $z(t)$ $(t \geq 0)$ as constructed above, which by construction satisfies $c^T z(t) \leq c^T x$, is feasible for (3.1) if and only if (3.17) holds. Recall, however, that we would like to construct a new basic feasible point that reduces the objective value. As $x_k = 0$ $(k \in K)$ and in view of (3.11), we we observe that $z(t)$ potentially has a non-zero component (namlely $z_r(t)$) more than x. Hence, we have to enforce a new zero component. For theses purposes, we define

$$s := \mathit{argmin} \left\{ \frac{x_j}{d_j} \mid j \in J : \ d_j > 0 \right\} \quad \text{and} \quad \hat{t} := \frac{x_s}{d_s}. \tag{3.18}$$

Then, in view of (3.14), we have $z_s(\hat{t}) = 0$. We then define an updated vector $\hat{x} := z(\hat{t})$. The following result shows that this vector has all the desired properties.

Theorem 3.1.4 (Simplex iteration) *Let x be a basic feasible point of $P = P(A, b)$ with corresponding index set $J \subset \{1, \ldots, n\}$ (cf. Theorem 2.2.2), its complement $K := \{1, \ldots, n\} \setminus J$ and the corresponding basis matrix $B = [a_j \ (j \in J)] \in \mathbb{R}^{m \times m}$. Assume that for the numbers u_k $(k \in K)$ defined by (3.7) (and (3.6)) we have*

$$u_k < 0 \quad \textit{for at least one} \quad k \in K.$$

In addition, for $r \in K$ with $u_r < 0$ assume that the vector $d \in \mathbb{R}^m$ defined by (3.13) we have

$$d_j > 0 \quad \textit{for at least one} \quad j \in J.$$

Finally, for $s \in J$ and $\hat{t}$ defined by (3.18) we define $\hat{x} \in \mathbb{R}^n$ by

$$\hat{x}_j := \begin{cases} x_j - \hat{t} d_j & \textit{if} \quad j \in J, \ j \neq s, \\ \hat{t} & \textit{if} \quad j = r, \\ 0 & \textit{else}. \end{cases}$$

Then the following hold:

a) $\hat{x}$ is a basic feasible point of P with the corresponding index set

$$\hat{J} := (J \cup \{r\}) \setminus \{s\}.$$

b) We have $c^T\hat{x} \leq c^T x$.

Proof: Since $\hat{x} = z(\hat{t})$ and $\hat{t} \geq 0$, we already know from the proof of Lemma 3.1.3 that $\hat{x}$ is feasible for (3.1). Moreover, statement b) is also valid, see (3.12) in combination with (3.10). Hence, it remains to be proven that $\hat{x}$ is a basic point of P: By definition we have

$$\hat{x}_j = 0 \quad (j \notin (J \cup \{r\}) \setminus \{s\} = \hat{J})$$

and, clearly $|\hat{J}| = m$. It remains to be shown that the vectors a_j $(j \in \hat{J})$ are linearly independent: For these purposes, let $\gamma_j \in \mathbb{R}$ $(j \in \hat{J})$ such that

$$0 = \sum_{j \in \hat{J}} \gamma_j a_j.$$

By the definition of $d \in \mathbb{R}^m$ in (3.13), we infer that

$$\begin{aligned} 0 &= \sum_{j \in J \setminus \{s\}} \gamma_j a_j + \gamma_r a_r \\ &= \sum_{j \in J \setminus \{s\}} \gamma_j a_j + \gamma_r B d \\ &= \sum_{j \in J \setminus \{s\}} \gamma_j a_j + \gamma_r \left(\sum_{j \in J} d_j a_j \right) \\ &= \sum_{j \in J \setminus \{s\}} (\gamma_j + \gamma_r d_j) a_j + \gamma_r d_s a_s. \end{aligned}$$

Since, by assumption, x is a (feasible) basic point of with the associated index set J (and $s \in J$) all vectors occurring in the last sum are linearly independent. Thus, we infer

$$\gamma_j + \gamma_r d_j = 0 \ (j \in J \setminus \{s\}) \quad \text{and} \quad \gamma_r d_s = 0.$$

Since by the definition of s in (3.18), we have $d_s > 0$, it follows that $\gamma_r = 0$ and consequently $\gamma_j = 0$ for all $j \in J \setminus \{s\}$. Therefore, $\gamma_j = 0$ for all $j \in \hat{J}$, which proves the linear independence of a_j $(j \in \hat{J})$ and hence completes the proof.

□

Given a basic feasible point x of the feasible set P, which is not yet a solution of the of the underlying primal program, Theorem 3.1.4 tells us, in case that the program is

bounded, how to construct a new basic feasible point $\hat{x}$ whose objective value is at most as big as for its predecessor, i.e. $c^T\hat{x} \leq c^T x$. Unfortunately, thus far, we cannot exclude the case $c^T\hat{x} = c^T x$. This will occur (cf. (3.12)) if and only if $\hat{t} = 0$. In fact, this case can occur in (3.18), but only if the original basic feasible point x fulfills

$$x_j = 0 \quad \text{for at least one} \quad j \in J.$$

A basic point x with this property we call *degenerate*. Otherwise we call it *nondegenerate*.

Corollary 3.1.5 *Under the assumptions of Theorem 3.1.4, if x is nondegenerate, then $c^T\hat{x} < c^T x$.*

Note however, that even if x is degenerate, $\hat{t} > 0$ is still possible!

3.2 The simplex method

The simplex method, in its basic form, simply consists in a sequence of simplex iterations, see Algorithm 3.2.1.
The following theorem summarizes the results from the foregoing section with regard to Algorithm 3.2.1.

Theorem 3.2.1 (Simplex method)

a) Alle iterates x^ℓ ($\ell = 0, 1, 2 \dots$) produced by Algorithm 3.2.1 are basic feasible points of $P = P(A, b)$ and it holds that

$$c^T x^{l+1} \leq x^l \quad (\ell = 0, 1, 2 \dots).$$

*b) If Algorithm 3.2.1 stops at step **(S3)**, the vector x^ℓ solves the linear program* (3.1) *and y^ℓ solves its dual program* (3.2).

*c) If Algorithm 3.2.1 stops at **(S6)** then has no solution.*

d) If all iterates x^ℓ are nondegenerate basic (feasible) points, then Algorithm (3.2.1) *stops after finitely many iterations either at a solution of* (3.1) *or with the finding that* (3.1) *has no solution.*

Proof: Statements a)–c) follow immediately from Theorem 3.1.4, Lemma 3.1.2 and Lemma 3.1.3.

In order to prove statement d) observe that, by Corollary 3.1.5, we have

$$c^T x^{\ell+1} < c^T x^\ell \quad (\ell = 0, 1, 2, \dots)$$

Algorithm 3.2.1 The simplex method

(S0) Choose a basic feasible point x_0 of $P := P(A, b)$ with a corresponding index set J_0 with $|J_0| = m$, set $K_0 := \{1, \ldots, n\} \setminus J_0$, define

$$B_0 := [a_j \ (j \in J_0)] \in \mathbb{R}^{m \times m},$$

and set $\ell := 0$.

(S1) Compute $y_\ell \in \mathbb{R}^m$ as the solution of the linear system

$$B_\ell^T y = c_{J_\ell}.$$

(S2) Set

$$u_k^\ell := c_k - (a_k)^T y_\ell, \quad k \in K_\ell.$$

(S3) If $u_k^\ell \geq 0$ for all $k \in K_\ell$: STOP.

(S4) Choose $r_\ell \in K_\ell$ such that $u_{r_\ell}^\ell < 0$.

(S5) Compute $d^\ell = (d_j^\ell)_{j \in J_\ell} \in \mathbb{R}^m$ as the solution of the linear system

$$B_\ell d = a_{r_\ell}.$$

(S6) If $d_j^\ell \leq 0$ for all $j \in J_\ell$: STOP.

(S7) Determine $s_\ell \in J_\ell$ and t_ℓ by

$$s_\ell := \mathit{argmin} \left\{ \frac{x_j^\ell}{d_j^\ell} \;\middle|\; j \in J_\ell,\, d_j^\ell > 0 \right\} \quad \text{and} \quad t_\ell := \frac{x_{s_\ell}^\ell}{d_{s_\ell}^\ell},$$

and set

$$\begin{aligned} x_j^{\ell+1} &:= \begin{cases} x_j^\ell - t_\ell d_j^\ell, & j \in J_\ell,\, j \neq s_\ell, \\ t_\ell, & j = r_\ell, \\ 0, & \text{else}, \end{cases} \\ J_{\ell+1} &:= (I_\ell \cup \{r_\ell\}) \setminus \{s_\ell\}, \\ K_{\ell+1} &:= \{1, \ldots, n\} \setminus J_{\ell+1}, \\ B_{\ell+1} &:= [a_j \ (j \in J_{\ell+1})], \end{aligned}$$

set $\ell \leftarrow \ell + 1$ and go to **(S1)**.

Consequently, every basic feasible point x^ℓ can occur at most once. But since there are only finitely many basic (feasible) points, cf. Theorem 2.3.1 b), Algorithm 3.2.1 must stop after finitely many iterations. The rest of the statement is clear from b) and c), respectively. □

The first question that arises naturally about the simplex method from Algorithm 3.2.1 is how to obtain the initial basic feasible point x_0. We will devote the whole next section to this nontrivial issue of not having a basic feasible point to start the simplex algorithm.

Moreover, in **(S4)** and **(S7)** there is still a certain degree of freedom as for the choice of the indices r_ℓ and s_ℓ. With a clever choice of these indices one can even drop the undesirable assumption in part d) that all iterates x_ℓ be nondegenerate. One of these choices is to take r_ℓ as the smallest index $k \in K_\ell$ such that $u_k^\ell < 0$ and to choose s_ℓ as the smallest index $j \in J_l$ such that

$$j \in \underset{i \in J_l, d_i^l}{argmin} \frac{x_i^l}{d_i^l > 0}.$$

This method is known as *Bland's rule* .

3.3 Initializing the simplex method

The simplex method from Algorithm 3.2.1 can only be started if one has a basic feasible point of $P = P(A, b)$. In certain situations, one of which is illustrated by the following example, it is very easy to find such a point.

Example 3.3.1 *Consider the linear program*

$$\min\ c^T x \quad \text{s.t.} \quad Ax \leq b,\ x \geq 0 \tag{3.19}$$

for $b \geq 0$. After introducing a slack variable $z \geq 0$ we obtain an equivalent linear program in standard form

$$\min\ c^T x \quad \text{s.t.} \quad [A\ I_m]\begin{pmatrix} x \\ z \end{pmatrix} = b,\ \begin{pmatrix} x \\ z \end{pmatrix} \geq 0. \tag{3.20}$$

Since $b \geq 0$ the point $(x_0, z_0) := (0, b)$ is a basic feasible point with the index set $J = \{n+1, \ldots, n+m\}$ (cf. Theorem 2.2.2) since the associated columns of $[A\ I_m]$ are the unit vectors e_i $(i = 1, \ldots, m)$ in $\mathbb{R}^m$.

◇

In general, one will not be given a basic feasible point that easily. This section is hence devoted to presenting two different methods for dealing with this situation where a basic feasible point of a given linear program in standard form

$$\min\ c^T x \quad \text{s.t.} \quad Ax = b,\ x \geq 0 \tag{3.21}$$

is not known.

3.3.1 The two-phases method

The first method we are going to present is based on the following idea: We apply the simplex method to an auxiliary linear program (based on the given data A, b, c that define the underlying LP in standard form (3.21)) for which a basic a feasible point can easily be found. The auxiliary LP is to be designed such that its solution leads to a basic feasible point of the original LP. The following theorem describes how this is done.

Theorem 3.3.2 (Two-phases method) *In* (3.21) *let* $b \geq 0$ *and consider the auxiliary linear program*

$$\min_{x,z} e^T z \quad \text{s.t.} \quad Ax + z = b, \quad x, z \geq 0. \tag{3.22}$$

where $e := (1, 1, \ldots, 1)^T \in \mathbb{R}^m$. *Then the following hold:*

a) The vector $\binom{x}{z} := \binom{0}{b}$ *is a basic feasible point for* (3.22) *(with the associated index set* $J := \{n+1, \ldots, n+m\}$*).*

b) The LP (3.22) *has a solution.*

c) Let $\binom{\bar{x}}{\bar{z}}$ *be a basic feasible point of* (3.22) *which solves* (3.22). *If* $\bar{z} \neq 0$ *then* (3.21) *is infeasible. If, in turn,* $\bar{z} = 0$ *and* $\operatorname{rank} A = m$ *then* $\bar{x}$ *is a basic feasible solution for* (3.21).

Proof:

a) Use the same reasoning (which is unaffected by a different objective function) as in Example 3.3.1.

b) Since the objective function $(x, z) \mapsto e^T z$ is bounded by 0 from below on the feasible set of (3.22) and the problem is feasible (see a)) and in standard form, we get the desired statement from Theorem 2.4.8 a).

c) First, consider the case that $\bar{z} \neq 0$: Suppose $\bar{x} \in \mathbb{R}^n$ were feasible for (3.22). Then $\binom{\bar{x}}{0} \in \mathbb{R}^{n+m}$ is feasible for (3.21) with

$$e^T 0 = 0 < e^T \bar{z},$$

which contradicts the fact that $\binom{\bar{x}}{\bar{z}}$ solves (3.22).

Now, consider the case that $\bar{z} = 0$ and assume that $\operatorname{rank} A = m$: Since $\binom{\bar{x}}{0}$ is feasible for (3.22), clearly, $\bar{x}$ is feasible for (3.21). Moreover, since $\binom{\bar{x}}{0}$ is a basic feasible point for (3.22), by Theorem 2.2.2, the columns of A with indices i such that $\bar{x}_i > 0$ must be linearly indpendent. Since $\operatorname{rank} A = m$ we can complete them, if necessary, with other columns of A to m linearly independent vectors. This shows with Theorem 2.2.2 that $\bar{x}$ is a basic feasible point of (3.21).

□

The meaning of Theorem 3.3.2 for finding a basic feasible point of (3.21) is clear: We first apply the simplex method to the auxiliary program 3.22. Since this program has a solution by part b), we will end up with a basic feasible point $\binom{\bar{x}}{\bar{z}}$ that solves (3.22). By item c), if $\bar{z} \neq 0$, then (3.21) is infeasible. Otherwise, if rank $A = m$, $\bar{x}$ is a basic feasible point of (3.21).

3.3.2 The big-M method

Without further ado, we present the central result for the second approach for dealing with the problem of not knowing a basic feasible point of (3.21).

Theorem 3.3.3 (Big-M method) *In* (3.21) *in let* $b \geq 0$ *and consider the linear program*

$$\min\ c^T x + Me^T z \quad \text{s.t.} \quad Ax + z = b,\ x, z \geq 0 \qquad Lin(M)$$

parameterized by $M \geq 0$. *Then the following hold:*

a) The vector $\binom{0}{b}$ *is a basic feasible point of Lin(M).*

b) If $\binom{\bar{x}}{0}$ *solves Lin(M) then* $\bar{x}$ *solves* (3.21).

c) If the LP (3.21) *has a solution there exists a number* $\bar{M}$ *such that for all* $M > \bar{M}$ *the linear program Lin(M) has a solution and every such solution* $\binom{\bar{x}}{\bar{z}}$ *has* $\bar{z} = 0$.

Proof:

a) Clearly, $\binom{0}{b}$ is feasible for Lin(M). Now observe that $[A\ I_m] \in \mathbb{R}^{m \times (n+m)}$ has full rank m, hence basic (feasible) points are characterized through Theorem 2.2.2.

Putting $J := \{n+1, \ldots, n+m\}$ the corresponding columns of $[A\ I_m]$ are the standard unit vectors in $\mathbb{R}^m$ hence linearly independent. Moreover, the components of $\binom{0}{b}$ for the indices not in J are zero. All in all, $\binom{0}{b}$ is a basic feasible point of Lin(M).

b) By assumption $\bar{x}$ is feasible for (3.21). If $\bar{x}$ is an arbitrary feasible point of (3.21) then $\binom{\bar{x}}{0}$ is feasible for Lin(M). Since we assume that $\binom{\bar{x}}{0}$ solves Lin(M) we infer that

$$c^T \bar{x} = c^T \bar{x} + Me^T 0 \leq c^T x + Me^T 0 = c^T \bar{x}.$$

Therefore $\bar{x}$ solves (3.21).

c) With Exercise 3 the dual of Lin(M) can be written as

$$\max\ b^T y \quad \text{s.t.} \quad A^T y \le c, \quad y_i \le M\ (i = 1, \dots, m). \tag{3.23}$$

By assumption (3.21) has a solution and hence, by strong duality, so has his dual

$$\max\ b^T y \quad \text{s.t.} \quad A^T y \le c. \tag{3.24}$$

Let $\bar{y}$ such a solution and put

$$\bar{M} := \max_{i=1,\dots,m} \bar{y}_i.$$

Since the feasible set of (3.23) is contained in the feasible set of (3.24) and $\bar{y}$ is by construction feasible for (3.23) for all $M > \bar{M}$, $\bar{y}$ also solves (3.23). Thus, by strong duality, Lin(M) also has a solution.

To prove the remaining assertion, for $M > \bar{M}$, let $\binom{\bar{x}}{\bar{z}}$ solve Lin(M). By Corollary 2.4.7 the complementarity conditions

$$\left[\begin{pmatrix} c \\ Me \end{pmatrix} - \begin{pmatrix} A^T \\ I_m \end{pmatrix} \bar{y}\right]_i \cdot \begin{pmatrix} \bar{x} \\ \bar{z} \end{pmatrix}_i = 0 \quad (i = 1, \dots, n+m).$$

The lower block yields

$$(M - \bar{y}_i) \cdot \bar{z}_i = 0\ (i = 1, \dots, m).$$

Using the fact that $M > \bar{M}$, the definition of $\bar{M}$ and $\bar{z} \ge 0$ this yields that $\bar{z} = 0$.

□

The interpretation of Theorem 3.3.3 is the following: Instead of applying the simplex method to our original problem (3.21), we apply it to Lin(M) for which we trivially find a basic feasible point to initialize the simplex algorithm. We start with an arbitrary $M > 0$, say $M = 100$. If The Algorithm (3.2.1) stops in **(S3)** with $\bar{z} \neq 0$ or in **(S6)**, we multiply M with a factor (say 10) and start over.

Exercises to Chapter 3

1. Prove Lemma 3.1.3.
2. **(Example of a simplex iteration)** For the data (A, b, c) from Example 3.1.1 and the basic feasible x point given there with the index sets J, K and the basis and non-basis matrices B and N, respectively, compute the vector $\hat{x}$ as described by Theorem 3.1.4. Are x or $\hat{x}$ degenerate?
3. **(Dual of Big-M problem)** Show that the dual of the problem Lin(P) is equivalent to

$$\max\ b^T y \quad \text{s.t.} \quad A^T y \le c, \quad y_i \le M\ (i = 1, \dots, m).$$

4 Convex functions

4.1 Definition and examples

In optimization (and also other areas such as measure theory, calculus of variations etc.) it is often expedient to consider real-valued functions that also take the value $+\infty$ (sometimes even $-\infty$). These we call *extended real-valued* functions. We use the (very reasonable) conventions

$$\infty + \infty = \infty \quad \text{and} \quad \alpha \cdot \infty = \infty \quad (\alpha > 0).$$

For $f : \mathbb{R}^n \to \overline{\mathbb{R}} := \{\pm\infty\}$ we define its *domain*

$$\operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < \infty\}$$

and its *epigraph*

$$\operatorname{epi} f := \left\{(x, \alpha) \in \mathbb{R}^{n+1} \mid f(x) \leq \alpha\right\}.$$

A mosty simple, yet very useful and prominent extended real-valued function is the following: For a set $C \subset \mathbb{R}^n$ its *indicator function* is defined by

$$\delta_C : x \in \mathbb{R}^n \mapsto \begin{cases} 0 & \text{if} \quad x \in C, \\ +\infty, & \text{else} \end{cases}$$

For instance it can be used to reformulate a constrained minimization problem

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

into an unconstrained one

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

A particularly important class of extended real-valued functions is defined below.

Definition 4.1.1 (Convex functions) *Let $C \subset \mathbb{R}^n$ be convex and $f : C \subset \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Then we say that f is* convex on C *if*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \quad (x, y \in \mathbb{R}^n, \lambda \in (0,1)).$$

For $C = \mathbb{R}^n$ we simply call f convex. *We call f* concave *(on C) if $-f$ is convex (on C).*

Clearly, the above definition coincides with the well-known notion of convexity for ordinary real-valued functions. Note that in the definition above we excluded the case $\lambda = 0$, since we did not want to deal with the case $0 \cdot \infty$.
The epigraph and the domain allow for very handy characterizations of convexity of and extended-real valued functions.

Proposition 4.1.2 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Then the following are equivalent:*

i) f is convex.

ii) f is convex on $\operatorname{dom} f$.

iii) $\operatorname{epi} f$ *is a convex set.*

Proof: See Exercise 1. □

We proceed with a short list of convex functions.

Example 4.1.3 (Convex functions)

a) The functions $\exp : \mathbb{R} \to \mathbb{R}$, $-\log : (0, \infty) \to \mathbb{R}$ *are convex. (Look at the epigraphs!)*

b) If $f : C \to \mathbb{R}$ is convex with a convex set $C \subset \mathbb{R}^n$ then

$$x \in \mathbb{R}^n \mapsto \begin{cases} f(x) & \text{if} \quad x \in C, \\ +\infty, & \text{else} \end{cases}$$

is convex (on $\mathbb{R}^n$).

c) (Affine functions) A function $f : \mathbb{R}^n \to \mathbb{R}$ of the form

$$f(x) = Ax - b \quad (A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m)$$

is called affine (linear). *All affine functions, hence all linear functions ($b = 0$) $\mathbb{R}^n \to \mathbb{R}$ are convex.*

d) (Indicator of convex sets) For a set $C \subset \mathbb{R}^n$ its indicator function δ_C is convex if and only if C is convex.

e) (Norms) Any Norm $\|\cdot\|$ on $\mathbb{R}^n$ is convex.

f) (Quadratic functions) For $Q \in \mathbb{R}^{n \times n}$ symmetric, $c \in \mathbb{R}^n$ and $\gamma \in \mathbb{R}$ the function

$$f : x \in \mathbb{R}^n \mapsto \frac{1}{2} x^T Q x + c^T x + \gamma$$

is convex if (and only if) Q is positive semidefinite, cf. Exercise 4.

We continue with a short (and incomplete) list of functional operations that preserve convexity.

Proposition 4.1.4 (Convexity preserving operations)

a) (Pointwise supremum) For an index set I let $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex for all $i \in I$. Then

$$f : x \in \mathbb{R}^n \mapsto \sup_{i \in I} f_i(x)$$

is a convex function.

b) (Positive combinations) For $i = 1, \ldots, n$ let $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex and $\lambda_i \geq 0$. Then $\sum_{i=1}^n \lambda_i f_i$ is convex.

c) (Composition with affine mapping) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex and $G : \mathbb{R}^m \to \mathbb{R}^n$ affine. Then $f \circ G$ is convex.

Proof:

a) Observe that $\operatorname{epi} f = \bigcap_{i \in I} \operatorname{epi} f_i$. Since, by Proposition 4.1.2, $\operatorname{epi} f_i$ ($i \in I$) is convex and, by Example 1.3.2 d), the intersection of convex sets is convex, $\operatorname{epi} f$ is convex, hence again by Proposition 4.1.2, f is convex.

b) Straightforward.

c) Exercise 2.

□

The importance of convexity in optimization stems from the fact that *local minimizers* are already *global minimizers* which we define for extended real-valued functions below.

Definition 4.1.5 (Local/global minimizers) *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $X \subset \mathbb{R}^n$. Then $\bar{x} \in X$ is said to be a*

i) local minimizer *of f over X if there exists $\varepsilon > 0$ such that*

$$f(\bar{x}) \leq f(x) \quad (x \in B_\varepsilon \cap X).$$

ii) global minimizer *of f over X if*

$$f(\bar{x}) \leq f(x) \quad (x \in X).$$

Proposition 4.1.6 (Minimizers in convex optimization) *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex. Then every local minimizer of f (over $\mathbb{R}^n$) is a global minimizer of f (over $\mathbb{R}^n$).*

Proof: Let $\bar{x}$ be a local minimizer of f and assume there were $\hat{y} \in \mathbb{R}^n$ such that $f(\hat{x}) < f(\bar{x})$. Then by convexity we have

$$f(\lambda\bar{x} + (1-\lambda)\hat{x}) \leq \lambda f(\bar{x}) + (1-\lambda)f(\hat{x}) < f(\bar{x}) \quad (\lambda \in (0,1)).$$

Noticing that $\lambda\bar{x} + (1-\lambda)\hat{x} \to \bar{x}$ as $\lambda \uparrow 1$, this contradicts the fact that $\bar{x}$ is a local minimizer of f. Hence, $\hat{x}$ with the above properties does not exist, i.e. $\bar{x}$ is a global minimizer of f. □

4.2 Smooth and nonsmooth convex functions

In the smooth case, convexity of a function can be expressed by a variational inequality involving the gradient.

Proposition 4.2.1 *Let $C \subset \mathbb{R}^n$ be open and convex and let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on C. Then f is convex on C if and only if*

$$f(x) \geq f(\bar{x}) + \langle \nabla f(\bar{x}),\, x - \bar{x} \rangle \quad (x, \bar{x} \in C). \tag{4.1}$$

Proof: Let $x, \bar{x} \in C$ and $\lambda \in (0,1)$. By convexity of f (and C) and the mean-value theorem there exists τ_λ on the connecting line between $\bar{x} + \lambda(x - \bar{x})$ and $\bar{x}$ such that

$$\lambda \nabla f(\tau_\lambda)^T (x - \bar{x}) = f(\bar{x} + \lambda(x - \bar{x})) - f(\bar{x}) \leq \lambda(f(x) - f(\bar{x})).$$

This implies

$$f(x) - f(\bar{x}) \geq \nabla f(\tau_\lambda)^T (x - \bar{x}).$$

Letting $\lambda \downarrow 0$ we have $\tau_\lambda \to \bar{x}$, hence by continuous differentiability of f, we infer

$$f(x) - f(\bar{x}) \geq \nabla f(\bar{x})^T (x - \bar{x}),$$

which proves one implication.

Now let (4.1) hold. Then take $x, y \in C$, $\lambda \in (0,1)$ and put $z := \lambda x + (1-\lambda)y$. Due to (4.1) we have

$$f(x) - f(z) \geq \nabla f(z)^T + (x - z)$$

and

$$f(y) - f(z) \geq \nabla f(z)^T (y - z).$$

Multiplying these inequalities by λ and $1 - \lambda$, respectively, and adding them gives

$$\lambda f(x) + (1-\lambda) f(y) - f(\lambda x + (1 - \lambda y)) \geq 0,$$

i.e. f is convex. □

Even in the real-valued case, a convex function can be far from differentiable (e.g. every norm is nondifferentiable at 0). Nevertheless there is a generalized notion for the gradient of a convex function even at points of nondifferentiability which is inspired by the characterization (4.1) in the smooth case.

Definition 4.2.2 (The subdifferential of a convex function) *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $\bar{x} \in \operatorname{dom} f$. A vector $v \in \mathbb{R}^n$ is called a* subgradient of f at $\bar{x}$ *if*

$$f(x) \geq f(\bar{x}) + \langle v,\, x - \bar{x} \rangle \quad (x \in \operatorname{dom} f). \tag{4.2}$$

The set of all subgradients of f at $\bar{x}$ is denoted by $\partial f(\bar{x})$ and called the subdifferential *of f at $\bar{x}$.*

Equation (4.2) is called the *subgradient inequality.*

Example 4.2.3 (Subdifferential of the indicator function) *Let $C \in \mathbb{R}^n$ be convex. Then the indicator function δ_C is convex, see Example 4.1.3 and we have* $\operatorname{dom} \delta_C = C$. *For $\bar{x} \in \operatorname{dom} f$ we compute*

$$\begin{aligned} \partial \delta_C(\bar{x}) &= \{v \mid f(x) \geq f(\bar{x}) + \langle v,\, x - \bar{x} \rangle \ (x \in \operatorname{dom} f)\} \\ &= \{v \mid 0 \geq \langle v,\, x - \bar{x} \rangle \ (x \in C)\} \\ &=: N_C(\bar{x}). \end{aligned}$$

This object is the so called normal cone *to the set C at $\bar{x}$. It is a central object of study in variational geometry of convex sets as well as for optimality conditions in convex optimization.*

Proposition 4.2.4 (Nonsmooth Fermat's rule) *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex and $\bar{x} \in \operatorname{dom} f$. Then $\bar{x}$ is a minimizer of f if and only if $0 \in \partial f(\bar{x})$.*

Proof: Exercise 3. □

Exercises to Chapter 4

1. **(Characterization of convex functions)** Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Show that the following are equivalent:

   i) f is convex.

ii) f is convex on its domain $\operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < +\infty\}$.

iii) The epigraph $\operatorname{epi} f := \left\{(x, \alpha) \in \mathbb{R}^{n+1} \mid f(x) \leq \alpha\right\}$ of f is a convex set.

2. **(Composition of convex and affine mappings)** Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex and $G : \mathbb{R}^m \to \mathbb{R}^n$ affine. Show that $f \circ G$ is convex.

3. **(Nonsmooth Fermat's rule)** Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $\bar{x} \in \operatorname{dom} f$. Show that $\bar{x}$ is a minimizer of f if and only if $0 \in \partial f(\bar{x})$.

4. **(Convexity of quadratic functions)** Let $Q \in \mathbb{R}^{n \times n}$ be symmetric. Show that $f : x \in \mathbb{R}^n \mapsto \frac{1}{2} x^T Q x$ is convex if Q is positive semidefinite.

5. **(Subdifferential of absolute value function)** For $f := |\cdot| : \mathbb{R} \to \mathbb{R}$ compute $\partial f(0)$.

5 Sensitivity Analysis

Consider the linear program in standard form

$$\min c^T x \quad \text{s.t.} \quad Ax = b,\ x \geq 0 \tag{5.1}$$

and its dual

$$\max b^T y \quad \text{s.t.} \quad A^T y \leq c \tag{5.2}$$

for the given data $(A, b, c) \in \mathbb{R}^{m\times n} \times \mathbb{R}^m \times \mathbb{R}^n$. In many situations one has incomplete or approximate knowledge of that data, and it is important to know how the optimal value and the optimal solution(s) of the associated LP change as A, b or c change, e.g. to have control of worst-case scenarios.

For these purposes we first need to enlarge our mathematical toolbox and study *convex functions*.

5.1 Sensitivity of optimal values

5.1.1 Global dependence of the right-hand side

Let $P(b) := \{x \in \mathbb{R}^n \mid Ax = b,\ x \geq 0\}$ be the feasible set of the standard linear program (5.1) parameterized by the right-hand side $b \in \mathbb{R}^m$ and assume that $\operatorname{rank} A = m$. Moreover, let

$$S := \{b \mid P(b) \neq \emptyset\}$$

and observe that

$$S = \{Ax \mid x \geq 0\} = A(\mathbb{R}^n_+),$$

hence, in particular, S is convex (see Chapter 1, Exercise 4b)).

We now define the *optimal value function*

$$\varphi : \mathbb{R}^m \to \overline{\mathbb{R}}, \quad \varphi(b) = \inf_{x\in P(b)} c^T x, \tag{5.3}$$

which assigns to $b \in \mathbb{R}^m$ the optimal value of the optimization problem (5.1).

We want to study analytical properties of the function φ.

Theorem 5.1.1 (Optimal value function of the right-hand side) *Consider the function φ defined by* (5.3). *Assuming that $\left\{y \;\middle|\; A^T y \leq c\right\}$ is nonempty the following hold:*

a) φ *is a function* $\mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ *with* $\operatorname{dom} \varphi = S$.

b) For all $b \in \operatorname{dom} \varphi$ *there exists* $\bar{x} \in P(b)$ *such that* $\varphi(b) = c^T \bar{x}$.

c) φ *is convex.*

Proof:

a) By assumption the dual program (5.2) is feasible, hence (5.1) is bounded for any $b \in \mathbb{R}^m$ (weak duality!), hence $\varphi(b)$ is nowhere $-\infty$. Moreover, $\varphi(b) < \infty$ if and only if $P(b) \neq \emptyset$, hence $\operatorname{dom} \varphi = S$.

b) This follows immediately from a) and Theorem 2.4.8.

c) In view of Proposition 4.1.2 and a) we only need to show that φ is convex on S. Therefore, let $b, b' \in \operatorname{dom} \varphi$ and $\lambda \in (0,1)$. By part b) there exist $v \in P(b)$ and $w \in P(b')$ such that $\varphi(b) = c^T v$ and $\varphi(b') = c^T w$. Now put $y := \lambda v + (1-\lambda) w$. Then

$$Ay = \lambda Av + (1-\lambda) Aw = \lambda b + (1-\lambda) b',$$

i.e. $y \in P(\lambda b + (1-\lambda) b')$. Therefore,

$$\varphi(\lambda b + (1-\lambda) b') = \inf_{x \in P(\lambda b + (1-\lambda) b')} c^T x \leq c^T y = \lambda c^T v + (1-\lambda) c^T w = \lambda \varphi(b) + (1-\lambda) \varphi(b').$$

This proves the convexity of φ.

□

With a stronger version of the fundamental theorem (which we only prove for linear programs in standard form), one can show that φ is not only convex, but in fact a pointwise maximum of finitely many linear functions (which is convex by Proposition 4.1.4), cf. [1, Section 5.2].

In the next result we study the subdifferential of the optimal value function (with respect to the right-hand side).

Theorem 5.1.2 (Subdifferential of the optimal value function) *Let* φ *be defined by* (5.3) *such that* $\left\{ y \mid A^T y \leq c \right\}$ *is nonempty. The for all* $\bar{b} \in S$ *we have*

$$\partial \varphi(\bar{b}) = \operatorname{argmax} \left\{ \bar{b}^T y \mid A^T y \leq c \right\},$$

i.e. the subdifferential of the optimal value function equals the set of all dual solutions.

Proof: First, let $v \in \mathit{argmax}\left\{\bar{b}^T y \mid A^T y \leq c\right\}$. By strong duality we have $\bar{b}^T v = \varphi(\bar{b})$. Now take an arbitrary $b \in S$. Then $\varphi(b) \geq v^T b$ by the definition of φ and weak duality. All in all we obtain

$$\varphi(\bar{b}) + v^T(b - \bar{b}) \leq \varphi(b) \quad (b \in S),$$

i.e. $v \in \partial\varphi(\bar{b})$.

In turn, let $v \in \partial\varphi(\bar{b})$, i.e.

$$\varphi(\bar{b}) + v^T(b - \bar{b}) = v^T b \leq \varphi(b) \quad (b \in S). \tag{5.4}$$

Now take $x \geq 0$ and put $b := Ax$. Then we have $b \in S$. Hence, we obtain

$$(A^T v)^T x = b^T v \leq \varphi(b) - \varphi(\bar{b}) + \bar{b}^T v \leq c^T x - \varphi(\bar{b}) + \bar{b}^T v,$$

where the second inequality uses (5.4). Since $x \geq 0$ was chosen arbitrarily, we obtain

$$(A^T v)^T x \leq c^T x - \varphi(\bar{b}) + \bar{b}^T v \quad (x \geq 0),$$

which implies $A^T v \leq c$ (see Exercise 1), i.e. v is dually feasible. Since $0 \in S$ with $\varphi(0) \leq 0$, (5.4) yields in this case

$$0 \geq \varphi(0) \geq \varphi(\bar{b}) - v^T \bar{b},$$

i.e. $\bar{b}^T v \geq \varphi(\bar{b})$, which by Corollary 2.4.4 implies that v solves the dual problem. □

5.1.2 Global dependence on the cost vector

In this section we fix the right-hand side b and the coefficient matrix A of (5.1) and let the cost vector c vary. This means the primal feasible set $P = \{x \mid Ax = b,\ x \geq 0\}$ remains unchanged and we assume that it is nonempty throughout. We define the dual feasible set

$$Q(c) := \left\{y \mid A^T y \leq c\right\}.$$

and

$$T := \{c \in \mathbb{R}^n \mid Q(c) \neq \emptyset\}. \tag{5.5}$$

We define the optimal value function

$$\phi : \mathbb{R}^n \to \overline{\mathbb{R}}, \quad \phi(c) = \inf\left\{c^T x \mid Ax = b,\ x \geq 0\right\}. \tag{5.6}$$

Theorem 5.1.3 (Optimal value function of the cost vector) *For the linear program in standard form* (5.1) *be feasibkle with* $\operatorname{rank} A = m$ *let* T *and* ϕ *be defined* (5.5) *and* (5.6)*, respectively. Moreover, let* $x_1, \ldots, x_N$ *be the basic reasible points of* P*. Then the following hold:*

a) $T \subset \mathbb{R}^n$ *is convex.*

b) We have

$$\phi(c) = \begin{cases} \min\limits_{i=1,\dots,N} c^T x_i & \text{if} \quad c \in T, \\ -\infty & \text{else.} \end{cases}$$

In particular, ϕ *is concave.*

Proof:

a) Let $c, d \in T$ i.e. there exist $y, z \in \mathbb{R}^m$ such that $A^T y \leq y$ and $A^T z \leq d$, respectively. Therefore, for $\lambda \in (0,1)$, we have

$$A^T(\lambda y + (1-\lambda)) = \lambda A^T y + (1-\lambda) A^T z \leq \lambda c + (1-\lambda) d,$$

i.e. $\lambda y + (1-\lambda) z \in Q(\lambda c + (1-\lambda) d)$. Thus, $\lambda c + (1-\lambda) d \in T$ and therefore T is convex.

b) If $c \notin T$ then the dual problem (5.2) is infeasible. Since the primal problem is assumed feasible, we must have $\inf(P) = \phi(c) = -\infty$.

If, in turn, $c \in T$ and the dual problem is not unbounded (as the primal is feasible), the dual problem has a solution, hence by strong duality, also the primal problem. Thus, by the fundamental theorem (Th. 2.3.1) one of the basic feasible points must be a solution and this gives the expression of ϕ in this case.

The concavity of ϕ, i.e. the convexity of $-\phi$ follows from Proposition 4.1.4 a) and Example 4.1.3 b).

□

Exercises to Chapter 5

1. Let $a, c \in \mathbb{R}^n$ and $\gamma \in \mathbb{R}$ such that $a^T x \leq c^T x + \gamma$ for all $x \geq 0$. Show that $a \leq c$.

6 Newton's method

This chapter is devoted to one of the most powerful tools for numerically solving non-linear equations of the form

$$F(x) = 0$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is assumed to be continuously differentiable. In order to establish our method, we need some technical preparations.

6.1 Review of differentiation in several variables

For $D \subset \mathbb{R}^n$ open and $F : D \to \mathbb{R}^m$ recall that F is called *differentiable* at $x \in D$ if there exists $L(x) \in \mathbb{R}^{m \times n}$ such that

$$\lim_{h \to 0} \frac{F(x+h) - F(x) - L(x)h}{\|h\|} = 0.$$

We put $F'(x) := L(x)$ and call it the *derivative* or *Jacobian* of F at x. The j-th partial derivative of its i-th component function at x we denote by $D_j F_i(x)$. Thus,

$$F'(x) = \begin{pmatrix} D_1F_1(x) & \cdots & D_nF_1(x) \\ \vdots & \dots & \vdots \\ D_1F_m(x) & \cdots & D_nF_m(x) \end{pmatrix} \in \mathbb{R}^{m \times n}.$$

Hence, for $f : D \to \mathbb{R}$ differentiable at $x \in D$ its derivative $f'(x)$ is an n-dimensional row vector. Its gradient in turn, given by $\nabla f(x) = f'(x)^T$, is a column vector.

In particular,

$$F'(x) = \begin{pmatrix} \nabla F_1(x)^T \\ \vdots \\ \nabla F_m(x)^T \end{pmatrix} \in \mathbb{R}^{m \times n}.$$

If f is twice differentiable at $x \in D$ (i.e. if ∇f is differentiable at x), we denote its second derivative at x by $\nabla^2 f(x)$ and call it its *Hessian* at x. If f is even twice *continuously* differentiable at $x \in D$, the Hessian at x is symmetric[1], hence the notation is justified.

[1] See *Schwarz's Theorem*

6.2 Matrix norms

Norms on $\mathbb{R}^{m\times n}$ are called *matrix norms*. The next result shows how to construct a particular class of matrix norms from given vector norms; these we call *operator norms*.

Proposition 6.2.1 (Operator norms) *Let $\|\cdot\|_X$ be $\|\cdot\|_Y$ (vector) norms on $X := \mathbb{R}^n$ and $Y := \mathbb{R}^m$, respectively. Then*

$$A \in \mathbb{R}^{m\times n} \mapsto \|A\|_{X,Y} := \sup_{x\neq 0} \frac{\|Ax\|_Y}{\|x\|_X}$$

is a norm on $\mathbb{R}^{m\times n}$ with

$$\|A\|_{X,Y} = \sup_{\|x\|_X=1} \|Ax\|_Y = \sup_{\|x\|_X\leq 1} \|Ax\|_Y. \tag{6.1}$$

Proof: Exercise 2.

□

Note that an operator norm $\|\cdot\|_{X,Y}$ always has $\|I\|_{X,Y} = 1$.

We point out that all the suprema in Proposition 6.2.1 are in fact maxima, which can be seen by a compactness argument (which clearly relies on $\mathbb{R}^n$ being finite-dimensional).

If, in the setting of Proposition 6.2.1, X and Y are equipped with the same norm $\|\cdot\|_*$ (tailored to the respective dimension), we set $\|\cdot\|_* := \|\cdot\|_{X,Y}$, hoping that no ambiguity will arise.

This notation is already used in the below result which gives a list of the most prominent operator norms on $\mathbb{R}^{m\times n}$.

Proposition 6.2.2 *Let $A \in \mathbb{R}^{m\times n}$. Then we have*

$$\begin{aligned}
\|A\|_1 &= \max_{j=1,\dots,n} \sum_{i=1}^{m} |a_{ij}| \quad (\textit{maximum absolute column sum});\\
\|A\|_2 &= \sqrt{\lambda_{\max}(A^TA)} \quad (\textit{spectral norm});\\
\|A\|_\infty &= \max_{i=1,\dots,m} \sum_{j=1}^{n} |a_{ij}| \quad (\textit{maximum absolute row sum}).
\end{aligned}$$

Proof: We only prove the statement for the spectral norm. For the other assertions see Exercise 3.

As $A^TA \in \mathbb{R}^{n\times n}$ symmetric positive semidefinite there exists an orthonormal basis[2] (w.r.t. $\|\cdot\|_2$) $v_1, \dots, v_n \in \mathbb{R}^n$ comprised of eigenvectors to the (not necessarily distinct)

[2] $\|v_i\| = 1$ and $v_i^T v_j = 0$ $(i \neq j)$ for all $i, j = 1, \dots, n$.

eigenvalues $0 \leq \lambda_1 \leq \cdots \leq \lambda_n = \lambda_{\max}$. Hence, for $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$ there exist $\mu_1, \ldots, \mu_n \in \mathbb{R}$ with

$$x = \sum_{i=1}^{n} \mu_i v^i.$$

With the orthogonal [3] matrix $V = (v_1, \ldots, v_n)$ a $\mu = (\mu_1, \ldots, \mu_n)^T$ we hence have

$$1 = \|x\|_2 = \|V\mu\|_2 = \|\mu\|_2,$$

where the last equality uses that orthogonal matrices are (2-)norm preserving. Thus,

$$\|Ax\|_2^2 = x^T A^T A x = \left(\sum_{i=1}^{n} \mu_i v^i\right)^T A^T A \left(\sum_{j=1}^{n} \mu_j v^j\right) = \sum_{i=1}^{n} \mu_i^2 \lambda_i \leq \lambda_{\max} \|\mu\|_2^2 = \lambda_{\max}.$$

Taking the square root and then the supremum yields

$$\|A\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2 \leq \sqrt{\lambda_{\max}}. \tag{6.2}$$

On the other hand, with the normed eigenvector $v = v_n$ (to the eigenvalue $\lambda_n = \lambda_{\max}$) we obtain

$$\|Av\|_2^2 = v^T A^T A v = \lambda_{\max} \|v\|_2^2 = \lambda_{\max},$$

hence, the supremum in (6.2) is taken and we have

$$\|A\|_2 = \sqrt{\lambda_{\max}}.$$

□

The following result summarizes some important properties of operator norms.

Proposition 6.2.3 *Let $\|\cdot\|_X, \|\cdot\|_Y, \|\cdot\|_Z$ on $\mathbb{R}^n, \mathbb{R}^m$ and $\mathbb{R}^p$, respectively. Then for all $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$ the following hold:*

a) $\|Ax\|_Y \leq \|A\|_{X,Y} \|x\|_X \quad (x \in \mathbb{R}^n)$ *(compatibility).*

b) $\|AB\|_{Z,Y} \leq \|A\|_{X,Y} \|B\|_{Z,X}$ *(submulitiplicativity).*

Proof: Exercise 4. □

The central result for our purposes is the following, which has an analogon in infinite dimension.

[3] Recall that a matrix $Q \in \mathbb{R}^n$ is called *orthogonal* if $QQ^T = Q^T Q = I$.

Proposition 6.2.4 (Banach Lemma) *Let $C \in \mathbb{R}^{n\times n}$ with $\|C\| < 1$ where $\|\cdot\|$ is a submultiplicative matrix norm. Then $I + C$ is invertible and we have*

$$\|(I + C)^{-1}\| \leq \frac{1}{1 - \|C\|}.$$

Proof: By the triangle inequality, the fact that $\|C\| < 1$ and the geometric series, we have

$$\left\| \sum_{i=0}^{n} (-C)^i \right\| \leq \sum_{i=0}^{n} \|C\|^i \to_{n\to\infty} \frac{1}{1 - \|C\|}.$$

Hence, the sequence $\{A_n \in \mathbb{R}^{m\times n}\}$ with $A_n := \sum_{i=0}^{n}(-C)^i$ is bounded and w.l.o.g. has a limit A. Moreover, observe that (telescoping sum)

$$A_n(I + C) = I + (-1)^n C^{n+1} \quad (n \in \mathbb{N}).$$

Hence,

$$A(I + C) = \lim_{n\to\infty} A_n(I + C) = \lim_{n\to\infty} I + (-1)^n C^{n+1} = I,$$

where the last identity uses the fact that $\lim_{k\to\infty} C^k \to 0$, cf. Exercise 5. Therefore $A = (I + C)^{-1}$ and we have

$$\|(I + C)^{-1}\| \leq \sum_{i=0}^{\infty} \|C\|^i = \frac{1}{1 - \|C\|}.$$

□

We actually employ the Banach Lemma in the following form.

Corollary 6.2.5 *Let $A, B \in \mathbb{R}^{m\times n}$ with $\|U - BA\| < 1$ for some submulitplicative norm $\|\cdot\|$. Then A and B are invertible with*

$$\|B^{-1}\| \leq \frac{\|A\|}{1 - \|I - BA\|}.$$

Proof: Set $C := BA - I$, apply the Banach Lemma . □

6.3 Convergence rates

In order to measure and compare the speed of convergence of numerical algorithms we now introduce various convergence rates.

Definition 6.3.1 (Convergence rates) *Let $\{x^k \in \mathbb{R}^n\} \to \bar{x}$ and $\|\cdot\|$ an arbitrary norm on $\mathbb{R}^n$. Then $\{x^k\}$ converges (at least)*

i) linearly *to $\bar{x}$ if the exists $c \in (0,1)$ such that*

$$\|x^{k+1} - \bar{x}\| \leq c\|x^k - \bar{x}\| \quad (k \in \mathbb{N}).$$

ii) superlinearly *to $\bar{x}$ if*

$$\lim_{k\to\infty} \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} = 0.$$

iii) quadratically *to $\bar{x}$ if there exists $C > 0$ such that*

$$\|x^{k+1} - \bar{x}\| \leq C\|x^k - \bar{x}\|^2 \quad (k \in \mathbb{N}).$$

We point out that we explicitly demand the sequence $\{x^k\}$ in Definition 6.3.1 to be convergent, which clearly, is redundant in case of linear (and hence superlinear) convergence.

In case of quadratic convergence one always has to ensure that the sequence at hand is, in fact, convergent. For instance, the sequence $\{x_k \in \mathbb{R}\}$ defined by

$$x_k := \begin{cases} 1, & \text{if} \quad k \in 2\mathbb{N}, \\ 2, & \text{if} \quad k \in 2\mathbb{N}+1 \end{cases}$$

with $\bar{x} := 0$ and $C := 5$ fulfills $\|x_{k+1} - \bar{x}\| \leq C\|x_k - \bar{x}\|^2$ $(k \in \mathbb{N})$ but $\{x_k\}$ does *not* converge to $\bar{x}$.

A very useful tool for formulating convergence rates are the *Landau symbols*.

Definition 6.3.2 (Landau symbols) *Let $\{a_k > 0\}, \{b_k > 0\} \downarrow 0$. Then we define:*

a) $a_k = o(b^k) \quad :\Longleftrightarrow \quad \lim_{k\to\infty} \frac{a_k}{b_k} = 0$; *(little 'O')*

b) $a_k = O(b_k) \quad :\Longleftrightarrow \quad \exists\, C > 0 \;\forall\, k \in \mathbb{N} :\; a_k \leq C b_k.$ *(big 'O')*

Using Landau notation, a sequence $\{x^k\} \to \bar{x}$ converges superlinearly if (and only if)

$$\|x^{k+1} - \bar{x}\| = o(\|x^k - \bar{x}\|),$$

and it converges quadratically if (and only if)

$$\|x^{k+1} - \bar{x}\| = O(\|x^k - \bar{x}\|^2).$$

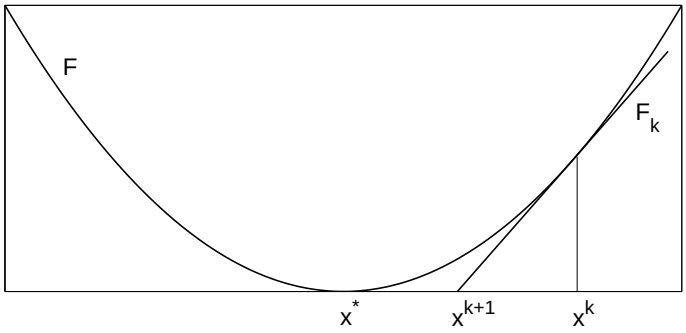

Figure 6.1: Newton's method

6.4 The Newton iteration

Recall that our goal is to establish a numerical method for solving

$$F(x) = 0, \tag{6.3}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is (at least) continuously differentiable. The basic idea is shockingly simple and relies on *linearization*, one of the most basic principles in mathematics:

Suppose $\bar{x}$ is a root of F and x^k is our current approximation for it. Then consider a local, linear approximation

$$x \mapsto F_k(x) := F(x^k) + F'(x^k)(x - x^k)$$

of F at x^k. Now, compute x^{k+1} as a root of F_k, see figure 6.1 for an illustration for the case $n = 1$.

If $F'(x^k)$ is invertible this leads to the iteration

$$x^{k+1} = x^k - F'(x^k)^{-1} F(x^k).$$

For numerical reasons[4] one does not explicitly invert a matrix, but one will rather compute the *Newton direction* d^k as a solution of the *Newton equation*

$$F'(x^k)d = -F(x^k) \tag{6.4}$$

and then update $x^{k+1} := x^k + d^k$. This yields Algorithm 6.4.1.

In order to prove well-definedness and local convergence of Algorithm 6.4.1 we need a series of auxiliary results. In what follows $\|\cdot\|$ describes both a vector norm on $\mathbb{R}^n$ and

[4] E.g. $F'(x^k)$ ill-conditioned

Algorithm 6.4.1 Local Newton method for equations

(S0) Choose $x^0 \in \mathbb{R}^n, \varepsilon > 0$ and set $k := 0$.
(S1) If $\|F(x^k)\| \leq \varepsilon$: STOP.
(S2) Compute d^k as a solution of

$$F'(x^k)d = -F(x^k).$$

(S3) Set $x^{k+1} := x^k + d^k, k \leftarrow k + 1$ and go to **(S1)**.

a submultiplicative matrix norm on $\mathbb{R}^{n\times n}$ that are compatible, which is the case, e.g., if we deal with an operator norm, see Proposition 6.2.3.

The first result states that, in particular, invertibility of the Jacobian of a continuously differentiable function at some point is a *local property*, i.e. is valid on a whole neighborhood of said point.

Lemma 6.4.1 *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable at $\bar{x} \in \mathbb{R}^n$ such that $F'(\bar{x})$ is invertible. Then there exist $\varepsilon > 0$ such that $F'(x)$ is invertible for all $x \in B_\varepsilon(\bar{x})$. Moreover, there exists $c > 0$ such that*

$$\|F'(x)^{-1}\| \leq c \quad (x \in B_\varepsilon(\bar{x})).$$

Proof: Since F' is continuous at $\bar{x}$ there exists $\varepsilon 0$ such that

$$\|F'(\bar{x}) - F'(x)\| \leq \frac{1}{2\|F'(\bar{x})^{-1}\|} \quad (x \in B_\varepsilon(\bar{x})).$$

Hence,

$$\|I - F'(\bar{x})^{-1}F'(x)\| \leq \|F'(\bar{x})^{-1}\| \cdot \|F'(\bar{x}) - F'(x)\| \leq \frac{1}{2} \quad (x \in B_\varepsilon(\bar{x})).$$

By Corollary 6.2.5 we see that $F'(x)$ is invertible for all $x \in B_\varepsilon(\bar{x})$ with

$$\|F'(x)^{-1}\| \leq \frac{\|F'(\bar{x})^{-1}\|}{\|I - F'(\bar{x})^{-1}F'(x)\|} \leq 2\|F'(\bar{x})^{-1}\|.$$

the assertion now follows with $c := 2\|F'(\bar{x})^{-1}\|$. □

Before we present the next result, we observe that using the Landau symbols from Section 6.3 that we can express the fact that F is differentiable at $\bar{x} \in \mathbb{R}^n$ if and only if

$$\|F(x_k) - F(\bar{x}) - F'(\bar{x})(x_k - \bar{x})\| = o(\|x_k - \bar{x}\|)$$

for all sequences $\{x_k\} \to \bar{x}$.

The following lemma shows how this can be sharpened if one assumes that the derivative F' is continuous or even locally Lipschitz continuous at $\bar{x}$

Recall that we say that $G : \mathbb{R}^n \to \mathbb{R}^m$ is *locally Lipschitz (continuous)* at $\bar{x} \in \mathbb{R}^n$ if there exists $L = L(\bar{x}) > 0$ and $\varepsilon > 0$ such that

$$\|G(x) - G(y)\| \leq L\|x - y\|$$

for all $x, y \in B_\varepsilon(\bar{x})$. The number $L > 0$ is called modulus of local Lipschitz continuity of G at $\bar{x}$. Note that every functions that is continuously differentiable function at some point is also locally Lipschitz at that point, see Exercise 7.

Its proof uses the well known fact that if $x : [a, b] \to \mathbb{R}^n$ is continuous and $\|\cdot\|$ is a norm on $\mathbb{R}^n$ then $\| \int_a^b x(t)\, dt\| \leq \int_a^b \|x(t)\|\, dt$.

Lemma 6.4.2 *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable and $\{x^k\}$ such that $x^k \to \bar{x}$. Then the following hold:*

a) We have

$$\|F(x^k) - F(\bar{x}) - F'(x^k)(x^k - \bar{x})\| = o(\|x^k - \bar{x}\|).$$

b) If F' is in addition locally Lipschitz continuous we have

$$\|F(x^k) - F(\bar{x}) - F'(x^k)(x^k - \bar{x})\| = O(\|x^k - \bar{x}\|^2)$$

Proof:

a) We have

$$\begin{aligned}
&\|F(x^k) - F(\bar{x}) - F'(x^k)(x^k - \bar{x})\| \\
&\quad\leq \|F(x^k) - F(\bar{x}) - F'(\bar{x})(x^k - \bar{x})\| + \|F'(x^k)(x^k - \bar{x}) - F'(\bar{x})(x^k - \bar{x})\| \\
&\quad\leq \|F(x^k) - F(\bar{x}) - F'(\bar{x})(x^k - \bar{x})\| + \|F'(x^k) - F'(\bar{x})\| \cdot \|x^k - \bar{x}\|
\end{aligned}$$

The first summand is $o(\|x^k - \bar{x}\|)$ as F is differentiable at $\bar{x}$. For the second one this follows from the continuity of F'.

b) Let $L > 0$ be the local Lipschitz constant of F' on a neighborhood of $\bar{x}$. Then the mean value theorem in integral form yields for all $k \in \mathbb{N}$ sufficiently large:

$$\begin{aligned}
&\|F(x^k) - F(\bar{x}) - F'(x^k)(x^k - \bar{x})\| \\
&\quad= \|\int_0^1 F'(\bar{x} + t(x^k - \bar{x}))\, dt(x^k - \bar{x}) - F'(x^k)(x^k - \bar{x})\| \\
&\quad\leq \int_0^1 \|F'(\bar{x} + t(x^k - \bar{x})) - F'(x^k)\|\, dt\|x^k - \bar{x}\|
\end{aligned}$$

$$
\begin{aligned}
&\leq L\|x^k - \bar{x}\| \int_0^1 \|(t-1)(x^k - \bar{x})\|\, dt \\
&= L\|x^k - \bar{x}\|^2 \int_0^1 |t-1|\, dt \\
&= \frac{L}{2}\|x^k - \bar{x}\|^2 \\
&= O(\|x^k - \bar{x}\|^2)
\end{aligned}
$$

□

We are now in a position to formulate the convergence theorem for Newton's method from Algorithm 6.4.1.

Theorem 6.4.3 (Local convergence of Newton's method) *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable and let $\bar{x}$ be a root of F such that $F'(\bar{x})$ is invertible. Then there exists $\varepsilon > 0$ such that for every $x^0 \in B_\varepsilon(\bar{x})$ the following hold:*

a) The local Newton method from Algorithm 6.4.1 is well-defined and generates a sequence $\{x^k\}$ convergent to $\bar{x}$.

b) The rate of convergence is at least superlinear.

c) If in addition F' is locally Lipschitz continuous the rate of convergence is quadratic.

Proof: By Lemma 6.4.1 there exist $\varepsilon_1 > 0$ and $c > 0$ such that $F'(x)$ is regular and

$$\|F'(x)^{-1}\| \leq c$$

for all $x \in B_{\varepsilon_1}(\bar{x})$. Moreover, by Lemma 6.4.2 a) there exists $\varepsilon_2 > 0$ such that

$$\|F(x) - F(\bar{x}) - F'(x)(x - \bar{x})\| \leq \frac{1}{2c}\|x - \bar{x}\|$$

for all $x \in B_{\varepsilon_2}(\bar{x})$. Set $\varepsilon := \min\{\varepsilon_1, \varepsilon_2\}$. Then for an arbitrary $x^0 \in B_\varepsilon(\bar{x})$ the next iterate x^1 is well-defined and

$$
\begin{aligned}
\|x^1 - \bar{x}\| &= \|x^0 - \bar{x} - F'(x^0)^{-1}F(x^0)\| \\
&\leq \|F(x^0)^{-1}\| \cdot \|F(x^0) - F(\bar{x}) - F'(x^0)(x^0 - \bar{x})\| \qquad (6.5) \\
&\leq \tfrac{1}{2}\|x^0 - \bar{x}\|.
\end{aligned}
$$

Hence $x^1 \in B_\varepsilon(\bar{x})$. Inductively we obtain

$$\|x^k - \bar{x}\| \leq \left(\frac{1}{2}\right)^k \|x^0 - \bar{x}\|$$

for all $k \in \mathbb{N}$. Thus the sequence $\{x^k\}$ is well-defined and converges to $\bar{x}$, which proves a).

In order to prove b) and c) we observe that analogous to (6.5) we can establish the inequalities

$$\begin{aligned} \|x^{k+1} - \bar{x}\| &= \|x^k - \bar{x} - F'(x^k)^{-1}F(x^k)\| \\ &\leq \|F'(x^k)^{-1}\| \cdot \|F(x^k) - F(\bar{x}) - F'(x^k)(x^k - \bar{x})\| \\ &\leq c\|F(x^k) - F(\bar{x}) - F'(x^k)(x^k - \bar{x})\|. \end{aligned}$$

By Lemma 6.4.2 we obtain superlinear and quadratic convergence of $\{x^k\}$, respectively. □

Exercises to Chapter 6

1. **(Frobenius norm)**
   a) Show that the mapping
   $$(A, B) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n} \mapsto \operatorname{tr}(A^T B)$$
   defines a scalar product on $\mathbb{R}^{m\times n}$.
   b) Give an explicit expression for the matrix norm $\|\cdot\|_F$ induced by the scalar product from a). Is this norm an operator norm in the sense of Proposition 6.2.1?

2. **(Operator norms)** Let $\|\cdot\|_X$ be $\|\cdot\|_Y$ (vector) norms on $X := \mathbb{R}^n$ and $Y := \mathbb{R}^m$, respectively. Show that
   $$A \in \mathbb{R}^{m\times n} \mapsto \|A\|_{X,Y} := \sup_{x\neq 0} \frac{\|Ax\|_Y}{\|x\|_X}$$
   is a norm on $\mathbb{R}^{m\times n}$ with
   $$\|A\|_{X,Y} = \sup_{\|x\|_X=1} \|Ax\|_Y = \sup_{\|x\|_X\leq 1} \|Ax\|_Y. \tag{6.6}$$

3. **(Examples of operator norms)** Let $A \in \mathbb{R}^{m\times n}$. Show the following:
   a) $\|A\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^m |a_{ij}|$;
   b) $\|A\|_\infty = \max_{i=1,\ldots,m} \sum_{j=1}^n |a_{ij}|$.

4. **(Properties of operator norms)**

   Let $\|\cdot\|_X, \|\cdot\|_Y, \|\cdot\|_Z$ on $\mathbb{R}^n, \mathbb{R}^m$ and $\mathbb{R}^p$, repsectively. Show that for all $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$ the following hold:

   a) $\|Ax\|_Y \leq \|A\|_{X\to Y}\|x\|_X \quad (x \in)R^n$;

   b) $\|AB\|_{Z\to Y} \leq \|A\|_{X\to Y}\|B\|_{Z\to X}$.

5. Let $A \in \mathbb{R}^{n\times n}$ with $\|A\|^k \to 0$ for a submultiplivative matrix norm on $\mathbb{R}^{m\times n}$. Prove that $A^k \to 0$.

6. **(Examples of locally Lipschitz functions)** Find a a function $f : \mathbb{R} \to \mathbb{R}$ which

   a) is continuous but not locally Lipschitz continuous;

   b) is locally Lipschitz continuous but not differentiable.

7. **(Local Lipschitz continuity of smooth functions)** Let $G : \mathbb{R}^n \to \mathbb{R}^m$ continuously differentiable. Show that G is locally Lipschitz continuous at every point $\bar{x} \in \mathbb{R}^n$.

8. **(Global vs. local Lipschitz continuity)** Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz at every point in the compact set $K \subset \mathbb{R}^n$. Show that there exists $L > 0$ such that

$$\|F(x) - F(y)\| \leq L\|x - y\| \quad (x, y \in K).$$

7 Interior-point methods for linear programs

The paper [4] by Karmakar was the first one to establish a solution method for linear programs that differs conceputally from the simplex method developed by Dantzig in 1947 and that we studied extensively in Chapter 3. Karmakar proposed an algorithm for the numerical solution of linear programs using *barrier method techniques* which had been neglected for quite some time and never used in the linear setting. These kinds of algorithms are nowadays called *interior-point methods*. We refer the interested reader to the survey article [9] for an account of the historical development of interior-point algorithms.

It cannot be stated in general which strategy, simplex or interior-point, is superior. That depends on the respective structure and size of the linear program at hand. What can be said with certainty is that certain interior-point methods have better *complexity*: As we will see in Section 7.4, it can be shown that the number of iterations that some interior-point method need to solve a linear program is bounded by a polynomial expression in the number of variables n of the linear program under reasonable assumption. This kind of complexity is called *polynomial*. An analogous result does not exist for simplex-type methods, in fact, as shown by infamous examples given by Klee and Minty in [5], where the simplex method may need a number of iterations which is *exponential* in the dimension n of the given linear program.

7.1 An auxiliary problem

As a preparation for the interior-point methods to be developed in this chapter we consider optimality conditions for problems of the form

$$\min\ f(x) \quad \text{s.t.} \quad Ax = b,\ x \in X \tag{7.1}$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is continuously differentiable on the open set $X \subset \operatorname{dom} f$, $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

Lemma 7.1.1 *Let $\bar{x}$ be a local minimizer of (7.1). Then there exists $\bar{\lambda} \in \mathbb{R}^m$ such that the tuple $(\bar{x}, \bar{\lambda})$ satisfies the optimality condition*

$$\nabla f(x) + A^T\lambda = 0, \quad Ax = b, \quad x \in X. \tag{7.2}$$

Proof: Note that $\bar{x}$ is feasible for (7.1), hence $\bar{x} \in X$ and $A\bar{x} = b$. Since X is open, $\bar{x}$ is even a local minimizer of the problem

$$\min\ f(x) \quad \text{s.t.} \quad Ax = b.$$

By a similar reasoning as in the proof of Proposition 2.2.3, we can assume that A has full (row) rank. Hence, we can apply the Lagrangian multiplier rule from Theorem 2.4.1 and infer that there exists $\bar{\lambda}$ with

$$\nabla f(\bar{x}) + A^T \bar{\lambda} = 0.$$

This proves the result. □

Under certain convexity assumption in (7.1) a much stronger result can be shown, where we rely on our findings for convex functions from Chapter 4.

Corollary 7.1.2 *Let f and X in (7.1) be convex. Then the following are equivalent:*

i) $\bar{x}$ is a local minimizer of (7.1);

ii) $\bar{x}$ is a global minimizer of (7.1);

iii) There exists $\bar{\lambda} \in \mathbb{R}^m$ such that $(\bar{x}, \bar{\lambda})$ satisfies (7.2).

Proof: Observe that (7.1) is equivalent to

$$\min \tilde{f} := f + \delta_{A^{-1}(\{b\}) \cap X}.$$

Since $\delta_{A^{-1}(\{b\}) \cap X}$ is convex (as the indicator of a convex set) and f is convex by assumption, so is $\tilde{f}$. The equivalence of i) and ii) therefore follows from Proposition 4.1.6. Since, by Lemma 7.1.1, i) always implies iii), it suffices to show that iii) implies ii): For these purposes let $(\bar{x}, \bar{\lambda})$ solve (7.2) and take $x \in X$ with $Ax = b$ (i.e. x feasible for (7.2)). Then we have

$$\begin{aligned} f(x) &\geq f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) \\ &= f(\bar{x}) - \bar{\lambda}^T A(x - \bar{x}) \\ &= f(\bar{x}), \end{aligned}$$

where the inequality is due to Propostion 4.2.1, the first equality uses the fact that $(\bar{x}, \bar{\lambda})$ satisfies (7.2) and the last one uses $Ax = b = A\bar{x}$.

□

7.2 The central path

We recall the primal linear program

$$\min c^T x \quad \text{s.t.} \quad Ax = b,\ x \geq 0, \tag{P}$$

determined by the data (A, b, c) with $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. Its dual can be written equivalently using a slack variable $s \geq 0$ in the following form

$$\max b^T y \quad \text{s.t.} \quad A^T y + s = c, s \geq 0, \tag{D'}$$

and we will also refer to it as the dual of (P).

They main motivation for our subsequent study stems from the following result.

Theorem 7.2.1 *The following are equivalent:*

a) The primal program (P) has a solution $\bar{x}$.

b) The solution (D') has a solution $(\bar{y}, \bar{s})$.

c) The optimality conditions

$$\begin{array}{rcll} A^T y + s & = & c, & \\ Ax & = & b, & \\ x_i s_i & = & 0 & (i = 1, \ldots, n), \\ x, s & \geq & 0 & \end{array} \tag{7.3}$$

have a solution $(\bar{x}, \bar{y}, \bar{s})$.

Proof: The equivalence of i) and ii) is immediately clear form strong duality (see Theorem 2.4.6). The equivalence of i)/ii) and iii) is a consequence of Corollary 2.4.7. □

The siginficance of Theorem 7.2.1 the following: If the primal (and hence the dual) linear program has a solution, then finding the primal and dual solution is equivalent to finding a solution of of the system (7.28), which is in fact equivalent to solving

$$0 \stackrel{!}{=} \begin{pmatrix} A^T y + s - c \\ Ax - b \\ \min\{x, s\} \end{pmatrix},$$

where the 'min' has to be interpreted component-wise. Unfortunately this is nonsmooth equation, and we are not in a position to deal with it. Hence, we consider the perturbed

system

$$\begin{array}{rcll} A^T y + s & = & c, & \\ Ax & = & b, & \\ x_i s_i & = & \tau & (i = 1, \ldots, n), \\ x, s & > & 0 & \end{array} \tag{7.4}$$

for some parameter $\tau > 0$. This system is usually called the *central path conditions*. As of yet it is not even clear how solvability of this system can be guaranteed. Considering for instance the linear program

$$\min\ x_1 + x_2 \quad \text{s.t.} \quad x_1 + x_2 = 0,\ x_1, x_2 \geq 0,$$

it can be seen immediately that the its associated central path conditions do not necessarily have a solution although the LP has one, cf. Exercise 2. In this section we will study in-depth the following mapping

$$0 < \tau \mapsto \{(x, y, s) \mid (x, y, s) \text{ solves } (7.4)\}.$$

Clearly, we hope to establish conditions under which this mapping is well-defined in the sense that for every $\tau > 0$ sufficiently small there exists a unique (x_τ, y_τ, s_τ) that solves (7.4). If this is the case, the mapping

$$\tau \mapsto (x_\tau, y_\tau, s_\tau)$$

is called the *central path* in the literature. In view of the above example, existence of the central path for a given linear program is not guaranteed in general. Hence, we now want to study this issue thoroughly. For these purposes, consider the function

$$\mathrm{lb} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}, \quad \mathrm{lb}(x) = \begin{cases} -\sum_{i=1}^n \log(x_i) & \text{if} \quad x > 0, \\ +\infty & \text{else.} \end{cases}$$

This function is called the *log-barrier function* and it is is strictly convex and differentiable on its open domain $\operatorname{dom} \mathrm{lb} = \mathbb{R}^n_+$ with $\nabla \mathrm{lb}(x) = (x_i^{-1})_{i=1}^n$ and has the property that it blows up as soon as x approaches the boundary of the positive orthant, cf. Exercise 4. For $\tau > 0$ we consider the so-called *log-barrier problem*

$$\min\ c^T x + \tau \mathrm{lb}(x) \quad \text{s.t.} \quad Ax = b,\ x > 0. \tag{7.5}$$

The connection between this problems and the central path conditions is clarified in the next result.

In its proof and also for the remainder we use the following standard notation:

$$\operatorname{diag}(x) := \begin{pmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_n \end{pmatrix} \in \mathbb{R}^{n \times n} \quad (x \in \mathbb{R}^n).$$

Theorem 7.2.2 *Let $\tau > 0$. Then the following are equivalent:*

i) The log-barrier problem (7.5) *has a solution x_τ;*

ii) The central path conditions (7.4) *have a solution (x_τ, y_τ, s_τ).*

Proof: First observe that the objective function $\phi : x \mapsto c^T x + \tau \mathrm{lb}(x)$ of the log-barrier problem (7.5) is obviously convex (as a sum of convex functions), and on the feasible set $\{x > 0 \mid Ax = b\}$ it is differentiable with $\nabla\phi(x) = c - \tau X^{-1}e$ where $X := \mathrm{diag}\,(x)$. Corollary 7.1.2 yields that x_τ solves (7.5) if and only if there exists $y_\tau \in \mathbb{R}^m$ such that (x_τ, y_τ) solves the optimality conditions

$$c - \tau X^{-1}e - A^T y = 0, \quad Ax = b, \quad x > 0.$$

Setting $s_\tau := \tau X^{-1}e > 0$ the triple (x_τ, y_τ, s_τ) solves the central path conditions (7.4). □

Theorem 7.2.2 tells us that we can trace back solvability of the central path conditions (7.4) to solvabilty of the log-barrier problem (7.5). Hence, we now concentrate for the moment on the latter. For these purposes, we define the *primal-dual strictly feasible set*

$$\mathcal{F} := \left\{(x, y, s) \;\middle|\; Ax = b,\ A^T y + s = c,\ x, s > 0\right\}.$$

In the result next we will give characterization for solvability of the log-barrier problem using the set $\mathcal{F}$.

Theorem 7.2.3 (Solvability of the log-barrier problem) *The following are equivalent:*

i) The primal-dual strictly feasible set $\mathcal{F}$ is nonempty.

ii) The log-barrier problem (7.5) *has a solution for all $\tau > 0$.*

Proof: Obviously, if (x_τ, y_τ, s_τ) solves the central path conditions (7.4) for some $\tau > 0$ then this triple lies in $\mathcal{F}$. Therefore, if $\mathcal{F}$ is empty then the central path conditions cannot have a solution for any $\tau > 0$. Hence, by Theorem 7.2.2, in this situation, also the log-barrier problem (7.5) has not solution. Hence, the condition $\mathcal{F} \neq \emptyset$ is necessary condition for (7.5) to have a solution for any $\tau > 0$.

We will now show that this is also sufficient: To this end, let $\tau > 0$ and $(\hat{x}, \hat{y}, \hat{s}) \in \mathcal{F}$, i.e.

$$\begin{aligned} A^T\hat{y} + \hat{s} &= s, \\ A\hat{x} &= b, \\ \hat{x}, \hat{s} &> 0. \end{aligned} \tag{7.6}$$

In what follows let B_τ denote the objective function of (7.5), i.e.

$$B_\tau : x \in \mathbb{R}^n \mapsto c^T x + \mathrm{lb}(x).$$

We will show that the (sublevel) set (cf. Exercise 5, Chapter 1)

$$\mathcal{L}_\tau = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0, B_\tau(x) \leq B_\tau(\hat{x})\} \, (= \mathrm{lev}_{B_\tau(\hat{x})} B_\tau)$$

is compact. It is clearly closed, since it is the preimage of the closed set $\{b\} \times \mathbb{R}^n_+ \times (-\infty, B_\tau(\hat{x}))$ under the continuous mapping $x > 0 \mapsto (Ax, x, B_\tau(x))$. Hence, we only need to show the boundedness of $\mathcal{L}_\tau$. For $x \in \mathcal{L}_\tau$ (7.6) implies

$$\begin{aligned} B_\tau(x) &= c^T x + \tau \mathrm{lb}(x) \\ &= c^T x - \hat{y}^T (\underbrace{Ax - b}_{=0}) + \tau \mathrm{lb}(x) \\ &= c^T x - x^T A^T \hat{y} + b^T \hat{y} + \tau \mathrm{lb}(x) \\ &= c^T x - x^T (c - \hat{s}) + b^T \hat{y} + \tau \mathrm{lb}(x) \\ &= x^T \hat{s} + b^T \hat{y} + \tau \mathrm{lb}(x). \end{aligned}$$

Hence, the condition

$$B_\tau(x) \leq B_\tau(\hat{x})$$

is equivalent to

$$\hat{s}^T x + \tau \mathrm{lb}(x) \leq B_\tau(\hat{x}) - b^T \hat{y}. \tag{7.7}$$

Now assume there were $\{x^k \in \mathcal{L}_\tau\}$ unbounded. Then, in particular, $x > 0$ and $x_i^k \to \infty$ for some $i \in \{1, \ldots, n\}$. Therefore $\hat{s}^T x^k + \tau \mathrm{lb}(x^k) \to \infty$, which contradicts (7.7). Therefore $\mathcal{L}_\tau$ is bounded, hence compact.

Since B_τ is continuous (and finite-valued) on $\mathcal{L}_\tau$ the optimization problem

$$\min \; B_\tau(x) \quad \text{s.t.} \quad x \in \mathcal{L}_\tau$$

has a solution x_τ. But by the definition of $\mathcal{L}_\tau$ the vector x_τ also solves the log-barrier problem (7.5).

□

We are no in a position to prove the the main result of this section.

Theorem 7.2.4 (Existence of the central path) *Let the primal-dual strictly feasible set $\mathcal{F}$ be nonempty. Then the central path conditions* (7.4) *have a solution (x_τ, y_τ, s_τ) for all $\tau > 0$ where the $x-$ and $s-$component are uniquely determined. If, in addition,* $\mathrm{rank}\, A = m$, *then also the y-component is unique.*

Proof: Since $\mathcal{F}$ is nonempty, by Theorem 7.2.3, the log-barrier problem (7.5) has a solution x_τ for all $\tau > 0$. Hence, by Theorem 7.2.2, the central path conditions (7.4) have a solution (x_τ, y_τ, s_τ) for all $\tau > 0$ where x_τ is the solution of the log-barrier problem (7.5). Since the objective function $x \mapsto c^T x + \mathrm{lb}(x)$ is strictly convex (cf. Exercise 4), x_τ is uniquely determined, see Exercise 3. By the condition $x_i s_i = \tau \; i(= 1, \ldots, n)$ we get the uniqueness of s_τ. If, in addition, A has full row rank then y_τ is uniquely determined by $A^T y_\tau + s_\tau = c$, namely $y_\tau = (AA^T)^{-1} A(c - s - \tau)$.

□

7.3 A general interior-point method for linear programming

The basic idea of the general inner-point method to be described here is to solve the central path conditions (7.4) numerically using Newton's method, see Algorithm 6.4.1 where we neglect the inequalities for the moment.

For theses purposes, for $\tau > 0$, we define the function $F_\tau : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ by

$$F_\tau(x, y, s) := \begin{pmatrix} A^T y + s - c \\ Ax - b \\ XSe - \tau e \end{pmatrix}, \tag{7.8}$$

where $X := \mathrm{diag}\,(x)$ and $S := \mathrm{diag}\,(s)$. Then obviously

$$F_\tau(x, y, s) = 0, \; x, s > 0 \quad \Longleftrightarrow \quad (x, y, s) \text{ solves } (7.4).$$

The critical assumption, other then continuous differentiability, for local convergence of Newton's method is that the Jacobian of the function in question be invertible at the root, cf. Theorem 6.4.3. Hence we now study invertibility of the Jacobian of F at points of interest.

Theorem 7.3.1 *Let $(x, y, s) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ with $x, s > 0$, let τ and define F_τ by (7.8). If $\mathrm{rank}\, A = m$ then $F'_\tau(x, y, s)$ is invertible.*

Proof: We have

$$F'_\tau(x, y, s) = \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \in \mathbb{R}^{(n+m+n)\times(n+m+n)}.$$

Now Let $q = (q_1, q_2, q_3) \in \mathbb{R}^{n+m+n}$ with $F'_\tau(x, y, s)q = 0$, i.e.

$$A^T q_2 + q_3 \;\; = \;\; 0,$$

$$\begin{aligned} Aq_1 &= 0, \\ Sq_1 + Xq_3 &= 0. \end{aligned}$$

Multiplying the first equation by q_1^T and using the second equation yields

$$0 = q_1^T A^T q + {}_2 + q_1^T q_3 = q_1^T q_3.$$

Hence, multiplying the third equation by $q_1^T X^{-1}$ gives

$$0 = q_1^T X^{-1} S q_1 + q_1^T X^{-1} X q_3 = q_1^T X^{-1} S q_1.$$

But as $X^{-1}S = \operatorname{diag}(\frac{s_1}{x_1}, \ldots, \frac{s_n}{x_n})$ is positive definite this implies $q_1 = 0$. The third equation thus yields $q_3 = 0$ as X is invertible. Since A^T has full (column) rank we infer from the first equation that $A^T q_2 = 0$, hence $q_2 = 0$. This proves that $\ker F'_\tau(x, y, s) = \{0\}$, i.e. $F'_\tau(x, y, s)$ is invertible (since square). □

The condition in Theorem 7.3.1 that A have full rank has been a standard assumption in our study in Chapters 2 and 3, and in view of Proposition 2.2.3 it is uncritical at least from a theoretical viewpoint. The fact that we need x and s to be strictly positive (which roughly speaking corresponds to $\tau > 0$) justifies retroactively once more that we focus on solving the central path conditions (7.4) rather then the actual optimality conditions (7.28).

All in all, Theorem 7.3.1 ensures that solving the nonlinear equation $F_\tau(x, y, s) = 0$ numerically using Newton's method is tractable under reasonable assumptions. Now assume that (x^k, y^k, s^k) is our current iterate. The Newton equation for computing the new Newton direction then reads

$$F'_\tau(x^k, y^k, s^k) \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} = -F_\tau(x^k, y^k, s^k),$$

which is equivalent to

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S_k & 0 & X_k \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} = \begin{pmatrix} c - A^T y^k - s^k \\ b - Ax^k \\ \tau e - X_k S_k e \end{pmatrix}. \tag{7.9}$$

If $(\Delta x^k, \Delta y^k, \Delta s^k)$ is a solution of (7.9) we put

$$(x^{k+1}, y^{k+1}, s^{k+1}) := (x^k, y^k, s^k) + t_k(\Delta x^k, \Delta y^k, \Delta s^k)$$

where $t_k > 0$ is a step-size (which in the local Newton method is chosen $t_k = 1$). The new iterate $(x^{k+1}, y^{k+1}, s^{k+1})$ inherits some very desirable properties provided the foregoing had them.

Lemma 7.3.2 *If $(\Delta x^k, \Delta y^k, \Delta s^k)$ solves (7.9) and (x^k, y^k, s^k) satisfies the linear equations $A^T y + s = c$ and $Ax = b$ then this also holds for the vector $(x^{k+1}, y^{k+1}, s^{k+1}) := (x^k, y^k, s^k) + t_k(\Delta x^k, \Delta y^k, \Delta s^k)$.*

Proof: We have

$$\begin{aligned} A^T x^{k+1} + s^{k+1} &= A^T(y^k + t_k \Delta y^k) + s^k + t_k \Delta s^k \\ &= A^T y^k + s^k + t_k \underbrace{(A^T \Delta y^k + \Delta s^k)}_{=0} \\ &= c. \end{aligned}$$

Moreover,

$$Ax^{k+1} = Ax^k + t_k \underbrace{A\Delta x^k}_{=0} = b.$$

□

Incdutively, by Lemma 7.3.2, provided that (x^0, y^0, s^0) satsifies the linear equations of the central path conditions (7.4), this holds for all iterates (x^k, y^k, s^k) generated by the procedure described above. In this case we can substitute the first two entries of the right-hand side of the Newton equation (7.9) by zeros. This explains the the linear system (7.10) in the general algorithm below, where also we come across the (primal-dual) strictly feasible set

$$\mathcal{F} := \{(x, y, s) \mid Ax = b,\ A^T y + s = c,\ x, s > 0\}$$

again, which we used in Section 7.2 to characterize solvability of the central path conditions.

Algorithm 7.3.1 general interior-point method

(S0) Choose $(x^0, y^0, s^0) \in \mathcal{F}$, $\varepsilon \in (0, 1)$ and put $k := 0$.

(S1) If $\mu_k := \frac{(x^k)^T s^k}{n} \leq \varepsilon$: STOP.

(S2) Choose $\sigma_k \in [0, 1]$ and determine $(\Delta x, \Delta y, \Delta s)$ as a solution of

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S_k & 0 & X_k \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \sigma_k \mu_k e - X_k S_k e \end{pmatrix}. \tag{7.10}$$

(S3) Put $(x^{k+1}, y^{k+1}, s^{k+1}) := (x^k, y^k, s^k) + t_k(\Delta x^k, \Delta y^k, \Delta s^k)$, $k \leftarrow k + 1$ and got to **(S1)**. At this, choose $t_k > 0$ such that $x^{k+1}, s^{k+1} >$ gilt.

Remark 7.3.3 (On Algorithm 7.3.1)

a) *As was argued above, by the choice of* $(x^0, s^0, y^0) \in \mathcal{F}$*, Lemma 7.3.2 implies that all iterates* (x^k, y^k, s^k) *satisfy the linear constraints* $A^T y = c$*,* $Ax = b$*. Moreover, by the choice of the setp-size* t_k *all iterates have* $x^k, s^k > 0$*. Therefore all iterates* $(x^k, y^k s^k)$ *lie in* $\mathcal{F}$ *for all* $k \in \mathbb{N}$*. In particular,* x^k *is feasible for the primal and* y^k *for the dual problem.*

b) *From a) we infer that*

$$\begin{aligned}(x^k)^T s^k &= (x^k)^T (c - A^T y^k) \\ &= c^T x^k - (Ax^k)^T y^k \\ &= c^T x^k - b^T y^k,\end{aligned}$$

i.e. $(x^k)^T s^k$ *is the duality gap the goes with* x^k, y^k *which are primally and dually feasible, respectively. Therefore the strong (or weak) duality theorem justify the termination criterion in* ***(S1)****. The term*

$$\mu_k := \frac{(x^k)^T s^k}{n} \quad (k \in \mathbb{N})$$

we call the weighted duality gap.

c) *Assume that* $\operatorname{rank} A = m$*. If* $(x^k, y^k, s^k) \in \mathcal{F}$ *then, by Theorem 7.3.1, the linear system* (7.10) *has a unique solution. Moreover, as* $x^k, s^k > 0$*, there exists* $t_k > 0$ *such that also* $x^{k+1}, s^{k+1} > 0$*. By Lemma 7.3.2,* $(x^{k+1}, y^{k+1}, s^{k+1}) \in \mathcal{F}$*. By the choice of* $(x^0, y^0, s^0) \in \mathcal{F}$ *Algorithm 7.3.1 is well-defined.*

d) *Algorithm 7.3.1 has basically two degrees of freedom, namely the step-size* t_k *an the* centering parameter σ_k*. Different choices yield different interior-point methods some of which we will study on more detail on what follows.*

e) *The term* $\sigma_k \mu_k$*, roughly speaking, plays the role of the parameter* $\tau > 0$ *in the central path conditions* (7.4)*: The choice* $\sigma_k = 0$ *corresponds to a Newton step for the equation* $F_0(x, y, s) = 0$ *(which is after all what we acutally want to solve). This however may lead to very small step-sizes since we want to ensure* $x^{k+1}, s^{k+1} > 0$ *as well.*

The other extreme $\sigma_k = 1$ *yields a Newton step for* $F_{\mu_k}(x, y, s) = 0$*. This choice does not lead us closer to the actual central path condtions, but allows for larger step-sizes.*

We will now show how the weighted dual gap μ_k can be reduced in every iteration. Here we make use of the following notation: For $t > 0$ we set

$$(x^k(t), y^k(t), s^k(t)) := (x^k + t\Delta x^k, y^k + t\Delta y^k, s^k + t\Delta s^k)$$

and

$$\mu_k(t) := \frac{x^k(t)^T s^k(t)}{n}$$

for an arbitrary step size $t > 0$.

Lemma 7.3.4 *Let $(\Delta x^k, \Delta y^k, \Delta s^k)$ be a solution of (7.10). Then the following hold:*

a) $(\Delta x^k)^T \Delta s^k = 0$.

b) $\mu_k(t) = (1 - t(1 - \sigma_k))\mu_k$.

Proof:

a) The proof of a) immediately from the first part of the proof of Theorem 7.3.1.

b) From (7.10) we infer, in particular, that

$$S_k \Delta x^k + X_k \Delta s^k = -X_k S_k e + \sigma_k \mu_k e.$$

Summation over the n rows of this equation yields

$$(s^k)^T \Delta x^k + (x^k)^T \Delta s^k = -(1 - \sigma_k)(x^k)^T s^k.$$

With a) this implies

$$\begin{aligned} x^k(t)^T s^k(t) &= (x^k + t\Delta x^k)^T (s^k + t\Delta s^k) \\ &= (x^k)^T s^k + t[(\Delta x^k)^T s^k + (x^k)^T \Delta s^k] + t^2 (\Delta x^k)^T \Delta s^k \\ &= [1 - t(1 - \sigma_k)](x^k)^T s^k, \end{aligned}$$

which proves the assertion.

□

We proceed with a convergence result for Algorithm 7.3.1 where we use the following extension of the Landau symbols from Definition 6.3.2: For nonnegative sequences $\{a_n \geq 0\}$ and $\{b_n \geq 0\}$ we define

$$a_n = O(b_n) \quad :\Longleftrightarrow \quad \exists_{C>0} \forall_{n \in \mathbb{N}}\ a_n \leq C b_n.$$

Theorem 7.3.5 *Let $\varepsilon \in (0, 1)$ and let $\{(x^k, y^k, s^k)\}$ and $\{\mu_k\}$ be generated by Algorithm 7.3.1. Moreover, assume that there exist $\delta > 0$ und $\omega > 0$ such that*

$$\mu_{k+1} \leq \Big(1 - \frac{\delta}{n^\omega}\Big)\mu_k, \quad k = 0, 1, 2, \ldots \tag{7.11}$$

and let $\kappa > 0$ *such that the starting point* (x^0, y^0, s^0) *satsifies*

$$\mu_0 \leq \frac{1}{\varepsilon^\kappa}. \tag{7.12}$$

Then there exists and index $K(= K(n)) \in \mathbb{N}$ *such that*

$$K = O(n^\omega |\log(\varepsilon)|)$$

and

$$\mu_k \leq \varepsilon \quad (k \geq K)$$

Proof: Since the logarithm is montonically increasing, from (7.11) we infer that

$$\log \mu_{k+1} \leq \log\Big(\Big(1 - \frac{\delta}{n^\omega}\Big)\mu_k\Big) = \log\Big(1 - \frac{\delta}{n^\omega}\Big) + \log \mu_k \quad (k \in \mathbb{N}).$$

Applying this formula multiple times and using (7.12) we obtain

$$\log u_k \leq k \log\Big(1 - \frac{\delta}{n^\omega}\Big) + \log \mu_0 \leq k \log\Big(1 - \frac{\delta}{n^\omega}\Big) + \kappa \log \frac{1}{\varepsilon} \quad (k \in \mathbb{N}).$$

Since

$$\log(1 + \beta) \leq \beta \quad (\beta > -1)^1$$

we thus obtain

$$\log \mu_k \leq k\Big(-\frac{\delta}{n^\omega}\Big) + \kappa \log \frac{1}{\varepsilon}.$$

Therefore we have $\mu_k \leq \varepsilon$ if

$$k\Big(-\frac{\delta}{n^\omega}\Big) + \kappa \log \frac{1}{\varepsilon} \leq \log \varepsilon.$$

A simple computation shows that the latter is satsified for all $k \in \mathbb{N}$ which have

$$k \geq (1 + \kappa)\frac{n^\omega}{\delta} \log \frac{1}{\varepsilon} = (1 + \kappa)\frac{n^\omega}{\delta} |\log \varepsilon|.$$

Putting

$$K := \min\left\{ k \in \mathbb{N} \;\middle|\; k \geq (1 + \kappa)\frac{n^\omega}{\delta} |\log \varepsilon| \right\}$$

gives the desired result.

□

The crucial assumption in Theorem 7.3.5 for an upper bound on the polynomial complexity of Algorithm 7.3.5 is (7.11). In the next section we hence study a realization of (7.3.1) in which this condition is naturally satsified.

[1] By Taylor's Theorem there exists θ between 1 and $\beta + 1$ such that $\log(\beta + 1) = \log 1 + \beta - \frac{1}{2\theta^2}\beta^2$.

7.4 Polynomial complexity of a path-following method

In this section we want to investigate an interior-point method which is a special instance of Algorithm (7.3.1) and which generates iterates in a neighborhood of the central path. For these purposes, recall again the (primal-dual) strictly feasible set Menge

$$\mathcal{F} := \{(x, y, s) \mid Ax = b,\ A^T y + s = c,\ x, s > 0\}.$$

In the literature there are several different vicinities of the central path used which yield different theoretical and numerical properties of the resulting interior-point method, see e.g. [8]. A general rule of thumb is that choosing a relatively tight neighborhood may lead to algorithms that have theoretically appealling properties but may convergence slowly in practice. In our study we focus on the set

$$\mathcal{N}(\gamma) := \{(x, y, s) \in \mathcal{F}^o \mid x_i s_i \geq \gamma\mu \quad \forall i = 1, \ldots, n\},$$

as a neighborhood of the central path where $\gamma \in (0, 1)$ and $\mu := \frac{x^T s}{n}$. Using our notational convention from the foregoing section that

$$(x^k(t), y^k(t), s^k(t)) := (x^k + t\Delta x^k, y^k + t\Delta y^k, s^k + t\Delta s^k) \quad (t > 0)$$

and

$$\mu_k(t) := \frac{x^k(t)^T s^k(t)}{n} \quad (t > 0)$$

we can now formulate an new algorithm which is a realization of Algorithm (7.3.1) with particular choices of the step-size and centering parameter. It is alled a *feasible path-following method* as its iterates try to stay in a neighborhood (namely $\mathcal{N}(\gamma)$) of the central path and that satisfy the linear equations throughout (cf. Lemma 7.3.2). See for Algorithm 7.4.1 for the explicit realization.

As was already mentioned above, the feasible path-following method from Algorithm (7.4.1) falls into the category of methods that Algorithm (7.3.1) constitutes. Thus all results proven in Section 7.3, in particular the convergence result Theorem 7.3.5, hold for Algorithm 7.4.1. However, as we will see, the critical assumption (7.11) will always be fulfilled off-hand. We also point out that the choice of the step-size $t_k > 0$ in **(S2)** could be substituted for other rules without affecting the convergence properties of the algorithm.

The main result of this section will show that Algorithm 7.4.1 has polynomial complexity, i.e. it terminates after finitely many iterations where the number of needed iterations is bounded by $p(n)$ where p is a polynomial and n the dimension of the problem variable. Before we can prove this result, however, we need a series of technical lemmas. Here, we always write $\|\cdot\| := \|\cdot\|_2$.

Algorithm 7.4.1 Feasible path-following method

(S0) Choose $\gamma \in (0,1), 0 < \sigma_{\min} < \sigma_{\max} < 1, \varepsilon \in (0,1), w^0 := (x^0, y^0, s^0) \in \mathcal{N}(\gamma)$ and put $k := 0$.

(S1) If $\mu_k := \frac{(x^k)^T s^k}{n} \leq \varepsilon$: STOP.

(S2) Choose $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$ and determine $\Delta w := (\Delta x, \Delta y, \Delta s)$ as a solution of

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S_k & 0 & X_k \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \sigma_k \mu_k e - X_k S_k e \end{pmatrix}. \tag{7.13}$$

Choose t_k as the largest $t \in [0,1]$ such that

$$(x^k(t), y^k(t), s^k(t)) \in \mathcal{N}(\gamma).$$

(S3) Put $w^{k+1} := w^k + t_k \Delta w^k$, $k \leftarrow k+1$and go to **(S1)**.

Lemma 7.4.1 *Let $u, v \in \mathbb{R}^n$ with $u^T v \geq 0$. Then*

$$\|UVe\| \leq 2^{-3/2} \|u+v\|^2,$$

where $U = \operatorname{diag}(u)$ and $V = \operatorname{diag}(v)$.

Proof: First, observe that for scalars $\alpha, \beta \in \mathbb{R}$ we have

$$\frac{1}{4}(\alpha+\beta)^2 = \frac{1}{4}(\alpha-\beta)^2 + \alpha\beta \geq \alpha\beta. \tag{7.14}$$

Moreover, it holds that

$$0 \leq u^T v = \sum_{i: u_i v_i \geq 0} u_i v_i + \sum_{i: u_i v_i < 0} u_i v_i = \sum_{i \in \mathcal{P}} |u_i v_i| - \sum_{i \in \mathcal{M}} |u_i v_i|, \tag{7.15}$$

where we partitioned $\{1, \ldots, n\}$ into the (disjoint) subsets

$$\mathcal{P} := \{i \mid u_i v_i \geq 0\} \quad \text{and} \quad \mathcal{M} := \{i \mid u_i v_i < 0\}.$$

Since for an arbitrary vector $x \in \mathbb{R}^d$ we always have

$$\|x\| \leq \|x\|_1,$$

equation (7.14) and (7.15) imply

$$\begin{aligned} \|UVe\| &= (\|[u_i v_i]_{i \in \mathcal{P}}\|^2 + \|[u_i v_i]_{i \in \mathcal{M}}\|^2)^{1/2} \\ &\leq (\|[u_i v_i]_{i \in \mathcal{P}}\|_1^2 + \|[u_i v_i]_{i \in \mathcal{M}}\|_1^2)^{1/2} \end{aligned}$$

$$\begin{aligned}
&\overset{(7.15)}{\leq} && (2\|[u_i v_i]_{i\in\mathcal{P}}\|_1^2)^{1/2} \\
&= && \sqrt{2}\|[u_i v_i]_{i\in\mathcal{P}}\|_1 \\
&\overset{(7.14)}{\leq} && \sqrt{2}\left\|\left[\frac{1}{4}(u_i+v_i)^2\right]_{i\in\mathcal{P}}\right\|_1 \\
&= && 2^{-3/2}\sum_{i\in\mathcal{P}}(u_i+v_i)^2 \\
&\leq && 2^{-3/2}\|u+v\|^2.
\end{aligned}$$

This concludes the proof.

□

In the next result we make use of the following notation: If $(\Delta x^k, \Delta y^k, \Delta s^k)$ is a solution of (7.13) we put

$$\Delta X_k := \operatorname{diag}(\Delta x^k) \quad \text{and} \quad \Delta S_k := \operatorname{diag}(\Delta s^k).$$

Lemma 7.4.2 *For $\gamma > 0$ let $(x^k, y^k, s^k) \in \mathcal{N}(\gamma)$. Then*

$$\|\Delta X_k \Delta S_k e\| \leq 2^{-3/2}(1+1/\gamma)n\mu_k.$$

Proof: From the third block-row of (7.13) we obtain

$$S_k\Delta x^k + X_k\Delta s^k = \sigma_k\mu_k e - X_k S_k e.$$

Multiplying this equation by $(X_kS_k)^{-1/2} = \operatorname{diag}\left(\frac{1}{\sqrt{x_1 s_1}}, \ldots, \frac{1}{\sqrt{x_n s_n}}\right)$ and using the abbreviations

$$D_k := X_k^{1/2} S_k^{-1/2} = \operatorname{diag}\left(\sqrt{\frac{x_1}{s_1}}, \ldots, \sqrt{\frac{x_n}{s_n}}\right)$$

yields

$$D_k^{-1}\Delta x^k + D_k\Delta s^k = (X_kS_k)^{-1/2}(\sigma_k\mu_k e - X_kS_k e). \tag{7.16}$$

Applying Lemma 7.4.1 to $u := D_k^{-1}\Delta x^k$ and $v := D_k\Delta s^k$ (note that we have $u^Tv = 0$ due to Lemma 7.3.4 a)) we infer

$$\begin{aligned}
\|\Delta X_k\Delta S_k e\| &= \|(D_k^{-1}\Delta X_k)(D_k\Delta S_k)e\| \\
&\overset{\text{Lemma 7.4.1}}{\leq} 2^{-3/2}\|D_k^{-1}\Delta x^k + D_k\Delta s^k\|^2 \\
&= 2^{-3/2}\|(X_kS_k)^{-1/2}(\sigma_k\mu_k e - X_kS_k e)\|^2.
\end{aligned}$$

From $(x^k)^Ts^k = n\mu_k, e^Te = n$ and $x_i^k s_i^k \geq \gamma\mu_k$ we thus obtain

$$\|\Delta X_k\Delta S_k e\| \leq 2^{-3/2}\|\sigma_k\mu_k(X_kS_k)^{-1/2}e - (X_kS_k)^{1/2}e\|^2$$

$$
\begin{aligned}
&= 2^{-3/2}\Big((x^k)^T s^k - 2\sigma_k\mu_k e^T e + \sigma_k^2\mu_k^2 \sum_{i=1}^{n} \frac{1}{x_i^k s_i^k}\Big) \\
&\leq 2^{-3/2}\left(n\mu_k - 2\sigma_k\mu_k n + \sigma_k^2\mu_k^2 \frac{n}{\gamma\mu_k}\right) \\
&= 2^{-3/2}\left(1 - 2\sigma_k + \frac{\sigma_k^2}{\gamma}\right) n\mu_k \\
&\leq 2^{-3/2}\,(1 + 1/\gamma)\, n\mu_k,
\end{aligned}
$$

where the last inequality uses $\sigma_k \in (0,1)$. This proves the result.

□

We next give a lower bound for the step-size t_k that is chosen in **(S2)** of Algorithm 7.4.1. This can be seen as the essential ingredient for the proof of polynomial complexity of Algorithm 7.4.1.

Lemma 7.4.3 *For $\gamma > 0$ let $(x^k, y^k, s^k) \in \mathcal{N}(\gamma)$. Then we have*

$$(x^k(t), y^k(t), s^k(t)) \in \mathcal{N}(\gamma) \quad (t \in [0, \bar{t}_k])$$

where

$$\bar{t}_k := 2^{3/2}\gamma\frac{\sigma_k}{n}\frac{1-\gamma}{1+\gamma}.$$

Proof: From the third block-row of (7.13) we infer that

$$s_i^k \Delta x_i^k + x_i^k \Delta s_i^k = \sigma_k\mu_k - x_i^k s_i^k \quad (i = 1, \ldots, n). \tag{7.17}$$

In addition, Lemma 7.4.2 implies

$$|\Delta x_i^k \Delta s_i^k| \leq \|\Delta X_k \Delta S_k e\| \leq 2^{-3/2}(1 + 1/\gamma)n\mu_k \quad (i = 1, \ldots, n). \tag{7.18}$$

Using the fact that $x_i^k s_i^k \geq \gamma\mu_k$ as well as equation (7.17) and (7.18) we infer that

$$
\begin{aligned}
x_i^k(t)s_i^k(t) &= (x_i^k + t\Delta x_i^k)(s_i^k + t\Delta s_i^k) \\
&= x_i^k s_i^k + t(x_i^k \Delta s_i^k + s_i^k \Delta x_i^k) + t^2 \Delta x_i^k \Delta s_i^k \\
&\geq x_i^k s_i^k (1-t) + t\sigma_k\mu_k - t^2|\Delta x_i^k \Delta s_i^k| \\
&\geq \gamma(1-t)\mu_k + t\sigma_k\mu_k - t^2 2^{-3/2}(1 + 1/\gamma)n\mu_k
\end{aligned}
$$

for all $i = 1, \ldots, n$ an all $t \in [0,1]$. Hence, the condition

$$x_i^k(t)s_i^k(t) \geq \gamma\mu_k(t) \overset{\text{Lem. } 7.3.4}{=} \gamma(1 - t(1-\sigma_k))\mu_k \quad (i = 1, \ldots, n) \tag{7.19}$$

is satisfied provided that

$$\gamma(1-t)\mu_k + t\sigma_k\mu_k - t^2 2^{-3/2}(1+1/\gamma)n\mu_k \geq \gamma(1-t+t\sigma_k)\mu_k.$$

Elementary transformations show that the latter is equivalent to

$$t\sigma_k\mu_k(1-\gamma) \geq t^2 2^{-3/2} n\mu_k(1+1/\gamma),$$

which in turn is equivalent to

$$t \leq 2^{3/2}\gamma\frac{\sigma_k}{n}\frac{1-\gamma}{1+\gamma}.$$

Hence, (7.19) is satsified for all $t \in [0, \bar{t}_k]$ and in order to prove the assertion

$$(x^k(t), y^k(t), s^k(t) \in \mathcal{N}(\gamma)(t \in [0, \bar{t}_k])$$

it remains to be seen that

$$(x^k(t), y^k(t), s^k(t)) \in \mathcal{F} \quad (t \in [0, \bar{t}_k]).$$

For these purposes, we first observe that the linear equations

$$Ax^k(t) = b \quad \text{and} \quad A^T y^k(t) + s^k(t) = c \quad (t \in [0, \bar{t}_k])$$

are satsified, cf. the discussion before Algorithm 7.3.1. Observing[2] that $\bar{t}_k < 1$ and[3] $\mu_k > 0$, from (7.19) we get for all $t \in [0, \bar{t}_k]$:

$$x_i^k(t)s_i^k(t) \geq \gamma(1-t(1-\sigma_k))\mu_k > 0. \tag{7.20}$$

Noticing that $x^k(0) = x^k > 0$ and $s^k(0) = s^k > 0$ this implies $x^k(t), s^k(t) > 0$ for all $t \in [0, \bar{t}_k]$.

□

The above result can be used to show that Algorithm 7.4.1 always satisfies condition (7.11) with $\omega = 1$.

Lemma 7.4.4 *Let $\{(x^k, y^k, s^k)\}$ be generated by Algorithmus 7.4.1. Then*

$$\mu_{k+1} \leq \Big(1 - \frac{\delta}{n}\Big)\mu_k \quad (k = 0, 1, 2, \dots)$$

for a constant δ independent of k.

[2] $\gamma \in (0,1)$ and $\gamma(1-\gamma) \leq \frac{1}{4}$.
[3] $\mu_k = \frac{(x^k)^T s_k}{n}$ and $x^k, s^k > 0$.

Proof: Due to Lemma 7.4.3 and the choice of t_k in **(S3)** of Algorithm 7.4.1, we have

$$t_k \geq \bar{t}_k = 2^{3/2}\gamma\frac{\sigma_k}{n}\frac{1-\gamma}{1+\gamma} \quad (k \in \mathbb{N}).$$

Therefore Lemma 7.3.4 b) implies

$$\begin{aligned} \mu_{k+1} &= \mu_k(t_k) \\ &= (1 - t_k(1-\sigma_k))\mu_k \\ &\leq (1 - \tfrac{2^{3/2}}{n}\gamma\tfrac{1-\gamma}{1+\gamma}\sigma_k(1-\sigma_k))\mu_k. \end{aligned} \tag{7.21}$$

The quadratic function $\sigma \mapsto \sigma(1-\sigma)$ takes its minimum on the intervall $[\sigma_{\min}, \sigma_{\max}]$ at one of the end points $\sigma_{\min}$ or $\sigma_{\max}$. Therefore

$$\sigma_k(1-\sigma_k) \geq \min\{\sigma_{\min}(1-\sigma_{\min}), \sigma_{\max}(1-\sigma_{\max})\} > 0 \quad (\sigma_k \in [\sigma_{\min}, \sigma_{\max}]).$$

Putting

$$\delta := 2^{3/2}\gamma\frac{1-\gamma}{1+\gamma}\min\{\sigma_{\min}(1-\sigma_{\min}), \sigma_{\max}(1-\sigma_{\max})\},$$

the assertion follows from (7.21).

□

The polynomial complexity of Algorithm (7.4.1) is now easily established.

Theorem 7.4.5 *Let $\{(x^k, y^k, s^k\}$ be generated by Algorithm 7.4.1, where (x^0, y^0, s^0) satisfies*

$$\mu_0 \leq \frac{1}{\varepsilon^\kappa}.$$

Then there exists $K = K(n) \in \mathbb{N}$ with $K = O(n|\log(\varepsilon)|)$ and

$$\mu_k \leq \varepsilon \quad (k \geq K).$$

Proof: Follows from Theorem 7.3.5 and Lemma 7.4.4. □

7.5 Mehrotra's predictor-corrector method

In the foregoing section we were able to prove polynomial convergence of a particular interior-method, and with comparably(!) small effort at that. In practice, however, interior-point methods are rarely implemented in the form of Algorithm 7.4.1, since, for example, already finding a starting point $(x^0, y^0, s^0) \in \mathcal{N}(\gamma)$ causes difficulties. In

Algorithm 7.5.1 Mehrotra's predictor-corrector method

(S0) Choose $(x^0, y^0, s^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ with $x^0 > 0, s^0 > 0$, choose $\eta \in (0,1)$ such that $\eta \approx 1$,and put $k := 0$.

(S1) If an appropriate stopping criterion is satisfied: STOP.

(S2) (Predictor step)
Compute $(\Delta x^{k,P}, \Delta y^{k,P}, \Delta s^{k,P})$ as a solution of the linear system

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S_k & 0 & X_k \end{pmatrix} \begin{pmatrix} \Delta x^P \\ \Delta y^P \\ \Delta s^P \end{pmatrix} = \begin{pmatrix} c - A^T y^k - s^k \\ b - Ax^k \\ -X_k S_k e \end{pmatrix}. \tag{7.22}$$

Put

$$\begin{aligned} t_{k,P}^{\text{prim}} &:= \min\Big\{1, \min_{i:\Delta x_i^{k,P}<0} \frac{-x_i^k}{\Delta x_i^{k,P}}\Big\}, \\ t_{k,P}^{\text{dual}} &:= \min\Big\{1, \min_{i:\Delta s_i^{k,P}<0} \frac{-s_i^k}{\Delta s_i^{k,P}}\Big\}, \\ \mu_k &:= \frac{(x^k)^T s^k}{n}, \\ \mu_{k,P} &:= \frac{\big(x^k + t_{k,P}^{\text{prim}} \Delta x^{k,P}\big)^T \big(s^k + t_{k,P}^{\text{dual}} \Delta s^{k,P}\big)}{n}, \\ \sigma_k &:= \big(\mu_{k,P}/\mu_k\big)^3. \end{aligned}$$

(S3) (Corrector step)
Compute $(\Delta x^k, \Delta y^k, \Delta s^k)$ as a solution of

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S_k & 0 & X_k \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} = \begin{pmatrix} c - A^T y^k - s^k \\ b - Ax^k \\ -X_k S_k e - \Delta X_{k,P} \Delta S_{k,P} e + \sigma_k \mu_k e \end{pmatrix}. \tag{7.23}$$

Put

$$\begin{aligned} t_{k,\max}^{\text{prim}} &:= \min\Big\{1, \min_{i:\Delta x_i^k<0} \frac{-x_i^k}{\Delta x_i^k}\Big\}, \\ t_{k,\max}^{\text{dual}} &:= \min\Big\{1, \min_{i:\Delta s_i^k<0} \frac{-s_i^k}{\Delta s_i^k}\Big\}, \end{aligned}$$

and set

$$\begin{aligned} t_k^{\text{prim}} &:= \min\big\{1, \eta t_{k,\max}^{\text{prim}}\big\}, \\ t_k^{\text{dual}} &:= \min\big\{1, \eta t_{k,\max}^{\text{dual}}\big\} \end{aligned}$$

as the primal and dual step-size, respectively.

(S4) Put

$$\begin{aligned} x^{k+1} &:= x^k + t_k^{\text{prim}} \Delta x^k, \\ y^{k+1} &:= y^k + t_k^{\text{dual}} \Delta \lambda^k, \\ s^{k+1} &:= s^k + t_k^{\text{dual}} \Delta s^k, \end{aligned}$$

$k \leftarrow k+1$, and go to **(S1)**.

Algorithm 7.5.1 we present a method frequently used in practice which goes back to Mehrotra in [7].
We will not prove any convergence results for Algorithm 7.5.1. Instead we now want to provide some comments on certain aspects of the algorithm which partly rely numerical experience and heuristic arguments.

Remark 7.5.1 (Mehrotra's predictor-corrector method)

a) As for step **(S0)***: We choose the starting point* (x^0, y^0, s^0) *to only satisfiy* $x^0 > 0$ *and* $s^0 > 0$*. In contrast to the foregoing interior-point methods there is hence a higher certain degree of freedom in the choice of the intitial vector, but refined techniques exist which can be numerically benificial, cf. [7]. The initial vector does not have to be feasible, in fact the linear equations* $Ax = b$ *and* $A^T y + s = c$ *are not necessarily fulfilled. Therefore Algorithm 7.5.1 falls into the class of* infeasible *interior-point methods.*

The parameter $\eta > 0$ *determines to which fraction of the maximum step size (which still guarantees* $x^{k+1}, s^{k+1} > 0$*) we take in* **(S3)***. In practice we will choose* η *very close to* 1*, e.g.* $\eta = 0.9995$*.*

b) As for step **(S1)***: The stopping criterion in* **(S1)** *has not been specified yet. It should be clear however that the condition* $\mu_k = \frac{(x^k)^T s^k}{n} \leq \varepsilon$ *is no longer applicable, since it is not clear whether the* x^k *and* (y^k, s^k) *are going to be feasible for the primal and dual linear program, respectively, cf. a). In practice, one will aim for a combination of a reduced "duality gap" and sufficient feasibility, for instance in the form of*

$$\frac{\|Ax^k - b\|}{\max\{1, \|b\|\}} + \frac{\|A^T y^k + s^k - c\|}{\max\{1, \|c\|\}} + \frac{|c^T x^k - b^T y^k|}{\max\{1, |c^T x^k|, |b^T y^k|\}} \leq \varepsilon,$$

where $\varepsilon = 10^{-8}$ *for example.*

c) As for **(S2)***: In this so-called* predictor step *one first determines a search direction* $\left(\Delta x^{k,P}, \Delta y^{k,P}, \Delta s^{k,P}\right)$ *as a solution of the linear equation* (7.22)*. This linear system is exactly the Newton equation* (7.9) *for* $\tau = 0$*, which corresponds to applying a Newton step to the actual optimality conditions from Theorem 7.2.1 which are of main interest. Note that, as opposed to the feasible methods above, we cannot set the first two block components of the right-hand side to zero, since already the starting vector does not necessarily satisfy the linear constraints* $Ax = b$ *and* $A^T y + s = c$*, see a).*

The step-sizes $t_{k,P}^{prim}$ *and* $t_{k,P}^{dual}$ *computed in* **(S2)** *at that are the largest possible (but less than 1) which still guarantee that*

$$x^k + t_{k,P}^{prim} \Delta x^{k,P} \geq 0 \quad \textit{und} \quad s^k + t_{k,P}^{dual} \Delta s^{k,P} \geq 0,$$

since the condtion

$$x_i^k + t\Delta x_i^{k,P} \geq 0 \quad (i = 1, \ldots, n)$$

comes directly out of

$$t \leq \frac{-x_i^k}{\Delta x_i^{k,P}} \quad (i : \ \Delta x_i^{k,P} < 0)$$

which explains the definition of $t_{k,P}^{prim}$. *Similarly we can justify the formula that determines* $t_{k,P}^{dual}$. *The fact that we discriminate between a primal and a dual step-size is the result of numerous test runs.*

The number $\mu_{k,P}$ *is the weighted duality gap for the data computed up to that point. If it is small the search direction* $(\Delta x^{k,P}, \Delta y^{k,P}, \Delta s^{k,P})$ *seems to be reasonably good and a small value of the centering parameter* σ_k *is potentialy advantageous. This is realized by the choice* $\sigma_k = \big(\mu_{k,P}/\mu_k\big)^3$. *If, in turn,* $\mu_{k,P}$ *is not small, i.e. the search direction* $(\Delta x^{k,P}, \Delta y^{k,P}, \Delta s^{k,P})$ *seems to be bad, the choice* $\sigma_k = \big(\mu_{k,P}/\mu_k\big)^3$ *yields a significantly larger centering parameter* σ_k, *which simply means that the soon to come* coorector step *become more of a centering step towards the central path.*

d) *As for* **(S3)***: In the corrector step one first computes a search direction* $(\Delta x^k, \Delta y^k, \Delta s^k)$ *as a solution of the linear system* (7.23) *which is essentially the Newton equation for the central path conditions with the centering parameter* σ_k *from* **(S2)**, *which has the same coefficient matrix as* (7.22) *(which can be exploited to save computations) but where in the third block row of the right-hand side we have the additional term* $-\Delta X_{k,P}\Delta S_{k,P}e$ *auf. This term is basically responsible for the name* corrector step *which we explain now: We observe that a full step from* (x^k, y^k, s^k) *in the direction of the predictor search direction* $(\Delta x^{k,P}, \Delta y^{k,P}, \Delta s^{k,P})$ *yields the following identity for the* i*-th component of the duality gap:*

$$\begin{aligned} \big(x_i^k + \Delta x_i^{k,P}\big)\big(s_i^k + \Delta s_i^{k,P}\big) &= x_i^k s_i^k + x_i^k \Delta s_i^{k,P} + s_i^k \Delta x_i^{k,P} + \Delta x_i^{k,P}\Delta s_i^{k,P} \\ &= \Delta x_i^{k,P}\Delta s_i^{k,P}, \end{aligned} \tag{7.24}$$

where the second equality follows from the last block row of (7.22). *Theoretically, in the end we should have* 0 *in* (7.24) *which is the reason why the right-hand side is corrected by substracting* $-\Delta x_i^{k,P}\Delta s_i^{k,P}$ *in* (7.23).

Alternatively, the computation of the search direction $\big(\Delta x^k, \Delta y^k, \Delta s^k\big)$ *can be described as follows: First determine the solution* $\big(\Delta x^{k,C}, \Delta y^{k,C}, \Delta s^{k,C}\big)$ *of the linear system*

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -\Delta X^{k,P}\Delta S^{k,P}e + \sigma_k\mu_k e \end{pmatrix}.$$

Then put

$$\left(\Delta x^k, \Delta y^k, \Delta s^k\right) := \left(\Delta x^{k,P}, \Delta y^{k,P}, \Delta s^{k,P}\right) + \left(\Delta x^{k,C}, \Delta y^{k,C}, \Delta s^{k,C}\right), \qquad (7.25)$$

where the triple $\left(\Delta x^{k,P}, \Delta y^{k,P}, \Delta s^{k,P}\right)$ *is the solution of* (7.22)*. the search direction* $\left(\Delta x^k, \Delta y^k, \Delta s^k\right)$ *coincides with the one generated in the corrector step* **(S3)** *of Algorithm 7.5.1. The representation in* (7.25) *has the advantage that one can explicitly see the share that that both the corrector and the predictor step in the eventual search direction* $\left(\Delta x^k, y^k, \Delta s^k\right)$*.*

After the computation of the corrector step, analogous to the predictor step, the maximal step-sizes $t_{\max}^{k,prim}$ *und* $t_{\max}^{k,dual}$ *are computed and those are the multplied by* $\eta \approx 1$ *in order to compute the ultimate step-sizes* t_k^{prim} *and* t_k^{dual}*, respectively.*

Exercises to Chapter 7

1. **(Solvability of central path conditions I)** Consider the linear program

$$\min\ x_1 + x_2 \quad \text{s.t.} \quad x_1 + x_2 \geq 1,\ x_1, x_2 \geq 0. \qquad (7.26)$$

   a) Determine all solutions of (7.26).

   b) Transform (7.26) to standard form.

   c) For $\tau > 0$ determine the solution(s) (x^τ, y^τ, s^τ) of the central path conditions associated with the reformulation from b).

   d) What can you say about $\lim_{\tau \to 0}(x_\tau, y_\tau, s_\tau)$?

2. **(Solvability of central path conditions II)** Determine a solution of the linear program

$$\min\ x_1 + x_2 \quad \text{s.t.} \quad x_1 + x_2 = 0,\ x_1, x_2 \geq 0$$

   and check the central path conditions for solvability.

3. **(Strictly convex functions)** A convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called *strictly convex* on the convex set $C \subset \operatorname{dom} f$ if

$$f(\lambda x + (1-\lambda)) < \lambda f(x) + (1-\lambda) f(y) \quad (x, y \in C, \lambda \in (0,1)).$$

   For $C = \operatorname{dom} f$ we simply call f strictly convex. Show the following:

   a) If f is strictly convex on $\operatorname{dom} f$ then *argmin* f is either empty or a singleton, i.e. f has a unique minimizer if any.

   b) Suppose C is open and f is continuously differentiable on C. Then f is strictly convex on C if and only if

$$f(x) > f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) \quad (x, \bar{x} \in C). \qquad (7.27)$$

4. **(Log-barrier function)** Let $\mathrm{lb} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be the log-barrier function

$$\mathrm{lb}(x) = \begin{cases} -\sum_{i=1}^n \log(x_i) & \text{if} \quad x > 0, \\ +\infty & \text{else.} \end{cases}$$

Show the following:

a) lb is continuously differentiable on $\operatorname{dom} f$ with

$$\nabla \mathrm{lb}(x) = \begin{pmatrix} x_1^{-1} \\ \vdots \\ x_n^{-1} \end{pmatrix} \quad (x > 0).$$

b) lb is strictly convex.

*c) f is continuous on the boundary of the domain, i.e. for all $\bar{x} \in \operatorname{bd}(\operatorname{dom} f)$ and $\{x_k \in \operatorname{dom} f\} \to \bar{x}$ we have

$$\lim_{k \to \infty} f(x_k) = f(\bar{x}).$$

5. **(Optimality conditions)** For the standard linear program defined by the data $(A, b, c) \in \mathbb{R}^{m \times n} \times \mathbb{R}^m \times \mathbb{R}^n$ consider the optimality conditions (cf. Theorem 7.2.1)

$$\begin{array}{rcll} A^T y + s & = & c, & \\ Ax & = & b, & \\ x_i s_i & = & 0 & (i = 1, \ldots, n), \\ x, s & \geq & 0. & \end{array} \tag{7.28}$$

Show that the solution set of (7.28) is convex.

6. **(Linear equation in interior-point methods)** For $A \in \mathbb{R}^{m \times n}$ and two positive definite diagonal matrices $S, X \in \mathbb{R}^{n \times n}$ consider the matrix

$$M := \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix}.$$

Show that if

$$M \Delta w = b$$

has a solution $\Delta w = (\Delta x, \Delta y, \Delta s) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ then the components Δx and Δs are uniquely determined.

Warning: You cannot assume that $\operatorname{rank} A = m$.

7. **(Central path conditions)** For $\tau > 0$ let $\varphi_\tau : \mathbb{R}^2 \to \mathbb{R}$ by given by

$$\varphi_\tau(a, b) := a + b - \sqrt{(a-b)^2 + 4\tau}.$$

a) Show that

$$\varphi_\tau(a, b) = 0 \quad \Leftrightarrow \quad ab = \tau,\ a, b > 0. \tag{7.29}$$

b) Use φ_τ to rewrite the *central path conditions* (see equation (7.4) in the notes) as a an equation of the form

$$\Phi_\tau(x, y, s) = 0,$$

where $\Phi_\tau : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ has to be defined appropriately.

8. **(Termination of interior-point algorithm)** Let x be feasible for the primal linear program

$$\min\ c^T x \quad \text{s.t.} \quad Ax = b,$$

and (y, s) be feasible for the dual linear program

$$\max\ b^T y \quad \text{s.t.} \quad A^T y + s = c$$

such that $x^T s \leq \varepsilon$ for some $\varepsilon > 0$. Show that

$$c^T \bar{x} \leq c^T x \leq c^T \bar{x} + \varepsilon \quad \text{and} \quad b^T \bar{y} - \varepsilon \leq b^T y \leq b^T \bar{y}$$

for any primal solution $\bar{x}$ and any dual solution $(\bar{y}, \bar{s})$.

8 Quadratic Programming

8.1 Optimality conditions

In this chapter we want to extend the framework of linear programming in the sense that we allow for a nonlinear objective in the form of a quadratic function. Concretely, we are going to study optimization problems of the form

$$\min \ \frac{1}{2}x^TQx + c^Tx + \gamma \quad \text{s.t.} \quad Ax = b,\ x \geq 0, \tag{8.1}$$

where $Q \in \mathbb{R}^{n\times n}$ is symmetric, $c \in \mathbb{R}^n$ and $\gamma \in \mathbb{R}$ while the constraints are described by $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^n$, i.e. the feasible set remains a polyhedron in standard form. This kind of problem is called a (standard) quadratic program.

In linear programming the standard form of the feasible set plays a central role as we know that solutions are to be found in the extreme points (basic feasible points) of the feasible polyhedron. This, however, is no longer true in quadratic programming. Therefore we also consider the more general form

$$\min \ \frac{1}{2}x^TQx + c^Tx + \gamma \quad \text{s.t.} \quad Bx = b,\ Ax \leq a \tag{8.2}$$

where the matrices B, A and the vectors b, a are chosen compatibly.

In linear programming there is no difference between local and global minima. This is no longer true in quadratic programming as the following example shows.

Example 8.1.1 *Consider the quadratic program*

$$\min \ (x_1 - 1)^2 - x_2^2 \quad \text{s.t.} \quad -1 \leq x_2 \leq 2.$$

It has the two local minima $(1, -1)^T$ *and* $(1, 2)^T$ *where only the latter is a global minimizer.*

In view of Proposition 4.1.6, since a quadratic program has a convex (even polyhedral) feasible set, this difference between local and global minima can only occur if the objective function is not convex. In fact, the objective in (8.1) (and (8.2)) is convex if (and actually only if) the symmetric matrix Q is positive semdefinite, cf. Exercise 4, Chapter 4, and in Example 8.1.1 we have $Q = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ which is clearly indefinite.

We now want to study optimality condtions for quadratic programs. As a preparation we prove the following result.

Lemma 8.1.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and $X \subset \mathbb{R}^n$ nonempty closed and convex. Then the following hold:*

a) If $\bar{x}$ is a local minimizer of

$$\min\ f(x) \quad \text{s.t.} \quad x \in X \tag{8.3}$$

then we have

$$\nabla f(\bar{x})^T(x - \bar{x}) \geq 0 \quad (x \in X). \tag{8.4}$$

b) If, in addition, f is convex and $\bar{x}$ satisfies (8.4) *then $\bar{x}$ solves* (8.3)*.*

Proof:

a) Suppose there were $x \in X$ such that $\nabla f(\bar{x})^T(x - \bar{x}) < 0$. Then by convexity of X we have $\bar{x} + t(x - \bar{x}) = tx + (1 - x)\bar{x} \in X$ for all $t \in (0, 1)$. By the mean vaue theorem, for all $t \in \mathbb{R}$ there exists ξ_t on the connection line between $\bar{x}$ and $\bar{x} + t(x - \bar{x})$ such that

$$f(\bar{x} + t(x - \bar{x})) - f(\bar{x}) = t\nabla f(\xi_t)^T(x - \bar{x}).$$

Hence, by continuity of ∇f (and the fact that $\xi_t \to \bar{x}$ as $t \downarrow 0$) we have

$$f(\bar{x} + t(x - \bar{x})) = f(\bar{x}) + t\nabla f(\xi_t)^T(x - \bar{x}) < f(\bar{x})$$

for all $t > 0$ sufficiently small. But as we already argued above, $\bar{x} + t(x - \bar{x}) \in X$ for these $t \in (0, 1)$, which contradicts the fact that $\bar{x}$ is a local minimizer of f on X.

b) From the assumptions and Proposition 4.2.1 (note that f is differentiable on $\mathbb{R}^n$) we obtain

$$f(\bar{x}) \leq f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x}) \leq f(x) \quad (x \in X)$$

which proves that $\bar{x}$ is a global minimizer of f on X.

□

We are now in a position to prove the main result of this section. The proof relies on combining the foregoing lemma with optimality conditions for linear programming.

Theorem 8.1.3 (Optimality conditions for quadratic programming) *Consider the quadratic program* (8.1) *with $Q \in \mathbb{R}^{n\times n}$ symmetric, $c \in \mathbb{R}^n$, $\gamma \in \mathbb{R}$ as well as $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Then the following hold:*

a) If $\bar{x}$ is a local minimizer of (8.1) *then there exist $\bar{y} \in \mathbb{R}^m$ and $\bar{s} \in \mathbb{R}^n$ such that the triple $(\bar{x}, \bar{y}, \bar{s})$ satisfies the following optimality conditions:*

$$\begin{aligned} Qx + c + A^T y - s &= 0, \\ Ax - b &= 0, \\ x_i \geq 0, s_i \geq 0, x_i s_i &= 0 \quad (i = 1, \ldots, n). \end{aligned} \tag{8.5}$$

b) In turn, if a triple $(\bar{x}, \bar{y}, \bar{s})$ satsifies (8.5) *and Q is positive semidefinite then $\bar{x}$ is a global minimizer of* (8.1).

Proof: Put

$$f(x) = \frac{1}{2} x^T Q x + c^T x + \gamma \quad \text{and} \quad X := \{x \in \mathbb{R}^n \mid Ax = b,\ x \geq 0\}.$$

a) If $\bar{x}$ is a local minimizer of (8.1) then by Lemma 8.1.2 we infer that $\bar{x}$ also solves the linear program

$$\min\ \nabla f(\bar{x})^T (x - \bar{x}) \quad \text{s.t.} \quad Ax = b,\ x \geq 0 \tag{8.6}$$

Observing that the fact that the constant term in the objective has no influence on optimality questions while the cost vector is $\nabla f(\bar{x})$ we infer from Theorem 7.2.1 that there exist $(\hat{y}, \bar{s})$ such that the triple $(\bar{x}, \hat{y}, \bar{s})$ satisfies

$$\begin{aligned} A^T y + s &= \nabla f(\bar{x}), \\ Ax &= b, \\ x_i s_i &= 0\ (i = 1, \ldots, n), \\ x, s &\geq 0. \end{aligned}$$

Putting $\bar{y} := -\hat{y}$ and observing that $\nabla f(\bar{x}) = Q\bar{x} + c$ (cf. Exercise 1) shows that $(\bar{x}, \bar{y}, \bar{s})$ satsifies (8.5).

b) If $(\bar{x}, \bar{y}, \bar{s})$ satisfies (8.5) then, by Theorem 7.2.1, $\bar{x}$ solves (8.6), i.e. $\nabla f(\bar{x})^T (x - \bar{x}) \geq 0$ for all $x \in X$. Thus, since f is convex as Q is positive semidefinite (see Exercise 4) and X is always closed and convex (since polyhedral), from Lemma 8.1.2 we infer that $\bar{x}$ solves (8.1).

□

The proof of the foregoing result relies heavily on the fact that we can identify (at least in the convex case) the solutions of the quadratic program (8.1) with those of a linear program (8.6) in standard form and then exploit the duality theory from linear programming. Using our usual techniques for reformulating arbitrary polyhedra in standard form, we can infer the following result from Theorem 8.1.3.

Corollary 8.1.4 *Consider the general quadratic program* (8.2)*. Then the following hold:*

a) If $\bar{x}$ *is a local minimizer of* (8.2) *then ther exist* $\bar{y} \in \mathbb{R}^m$ *and* $\bar{z} \in \mathbb{R}^p$ *such that* $(\bar{x}, \bar{y}, \bar{z})$ *solves the optimality conditions:*

$$\begin{aligned} Qx + c + A^T y + B^T z &= 0, \\ Bx - b &= 0, \\ y_i \geq 0,\ [a - Ax]_i \geq 0,\ y_i[a - Ax]_i &= 0 \quad (i = 1, \ldots, m). \end{aligned} \tag{8.7}$$

b) If the triple $(\bar{x}, \bar{y}, \bar{z})$ *solves the optimality conditions* (8.13) *and* Q *is positive semidefinite then* $\bar{x}$ *solves the general quadratic program* (8.2) *(i.e. is a global minimizer).*

Proof: See Exercise 2. □

The optimality conditions (8.13) are called the *KKT*[1] *conditions* of the optimization problem (8.2). They are necessary optimality conditions and in the convex case, i.e. when Q is positive semidefinite, they are also sufficient. Hence, in the latter case the KKT conditions are equivalent to the actual quadratic program. The triple $(\bar{x}, \bar{y}, \bar{z})$ is called a *KKT point* of the optimization problem 8.2.

8.2 The active-set method

For the moment we focus on an *equality constrained* quadratic program of the form

$$\min\ f(x) := \frac{1}{2} x^T Q x + c^T x + \gamma \quad \text{s.t.} \quad Bx = b, \tag{8.8}$$

where $Q \in \mathbb{R}^{n \times n}$ is symmetric, $c \in \mathbb{R}^n$ and $\gamma \in \mathbb{R}$ while the constraints are described by $B \in \mathbb{R}^{p \times n}$ and $b \in \mathbb{R}^p$.

From Corollary 8.1.4 (with $A = 0,\ a = 0$) we know that if $\bar{x}$ is a local minimizer of (8.8) then there exists $\bar{z} \in \mathbb{R}^p$ such that $(\bar{x}, \bar{z})$ satisfies the KKT conditions

$$Qx + B^T z = -c, \quad Bx = b$$

of (8.8). Therefore we have proven the following result.

Proposition 8.2.1 *The tuple* $(\bar{x}, \bar{z}) \in \mathbb{R}^n \times \mathbb{R}^p$ *is a KKT point of* (8.8) *if and only if it solves the linear equation*

$$\begin{pmatrix} Q & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix}. \tag{8.9}$$

[1] KKT stands for Karush-Kuhn-Tucker

Corollary 8.2.2 tells us that finding a KKT point of (8.8) is equivalent to solving a linear equation (namely (8.9)). Moreover, we to recall that in case Q is positive semidefinite, the KKT conditions are equivalent to the quadratic program, see the discussion above. This means that, in particular, finding a solution of the equality constrained quadratic program (8.8) is equivalent to solving the linear equation (8.9).

For our subsequent study we reformulate the statement of Proposition 8.2.1: If we write $x = x^k + \Delta x$ in the linear system (8.9) where x^k be feasible for (8.8) and $\Delta x \in \mathbb{R}^n$ is some correction vector, we infer the following equivalences from (8.9):

$$\begin{aligned}
\begin{pmatrix} Q & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix} \quad &\Leftrightarrow \quad \begin{pmatrix} Q & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} x^k + \Delta x \\ z \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix} \\
&\Leftrightarrow \quad \begin{pmatrix} Q & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ z \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix} - \begin{pmatrix} Q & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} x^k \\ 0 \end{pmatrix} \\
&\Leftrightarrow \quad \begin{pmatrix} Q & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ z \end{pmatrix} = \begin{pmatrix} -c - Qx^k \\ b - Bx^k \end{pmatrix} \\
&\Leftrightarrow \quad \begin{pmatrix} Q & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ z \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ 0 \end{pmatrix},
\end{aligned}$$

since x^k is feasible for (8.8) and $\nabla f(x) = Qx + c$. Therefore we have justified the following reformulation of Proposition 8.2.1 which will turn out to be useful shortly when we tackle problems that also have inequality constraints.

Corollary 8.2.2 *Let x^k be feasible for the quadratic optimization problem* (8.8)*. Then $(\bar{x}, \bar{z})$ is a KKT point of* (8.8) *if and only if $\bar{x} = x^k + \Delta\bar{x}$ and $(\Delta\bar{x}, \bar{z})$ solves the linear system*

$$\begin{pmatrix} Q & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ z \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ 0 \end{pmatrix}.$$

We would like to point out that solving the linear equation in Corollary (8.2.2) is equivalent to computing a Newton step for the equation

$$F(x, y) := \begin{pmatrix} \nabla f(x) + B^T y \\ Bx \end{pmatrix} = 0$$

which are again the KKT conditions for (8.8). We encourage the reader to verify this.

We now turn our attention to the general quadratic optimization problem (8.2) where we use the following notation

$$A \quad := \quad \begin{pmatrix} \vdots \\ a_i^T \ (i = 1, \ldots, m) \\ \vdots \end{pmatrix} \in \mathbb{R}^{m \times n}$$

$$B \quad := \quad \begin{pmatrix} \vdots \\ b_j^T \ (j = 1, \ldots, p) \\ \vdots \end{pmatrix} \in \mathbb{R}^{p \times n}$$

as well as

$$a = (\alpha_1, \ldots, \alpha_m)^T \quad \text{and} \quad b = (\beta_1 \ldots, \beta_p)^T.$$

Note that, in particular, the entries of a and b are **not** denoted by a_i and b_j, respectively!

Using thsis notation we can rewrite (8.2) as

$$\min \ f(x) \quad \text{s.t.} \quad b_j^T x = \beta_j \ (j = 1, \ldots, p), \quad a_i^T x \leq \alpha_i \ (i = 1, \ldots, m) \tag{8.10}$$

where $f(x) = \frac{1}{2} x^T Q x + c^T x + \gamma$.

From Corollary 8.2.2 we learned that finding a KKT point of an equality constrained quadratic program is equivalent to solving a linear equation, which is a very desirable situation, since the machinery for solving linear systems is powerful and refined. In order to exploit this, even if inequalities come into play, we will try to identify the inequalities which, at a current iterate x^k, are active (i.e. hold as an equality), and treat them as equalities, while neglecting the inactive ones, since by continuity, they are inactive in a whole neighborhood of the current point. At this, the critical part is to find a suitable approximation $\mathcal{A}_k$ for the set

$$I(x^k) := \left\{ i \;\middle|\; a_i^T x^k = \alpha_i \right\} \subset \{1, \ldots, m\}$$

of indices of inequality constraints active at x^k. Using the index set $\mathcal{A}_k$ we define the matrix

$$A_k := \begin{pmatrix} \vdots \\ a_i^T \ (i \in \mathcal{A}_k) \\ \vdots \end{pmatrix} \in \mathbb{R}^{|\mathcal{A}_k| \times n}$$

and give a pseudo-code for the active-set method in Algorithm 8.2.1.
Instead of going in a in-depth theoretical analysis of Algorithm 8.2.1 we confine ourselves with some remarks that explain the various steps to :

The current iterate x^k is feasible for (8.10) (since x^0 is chosen feasible and the feasibility of x^k will follow inductively from the subsequent explanations). Therefore, Corollary 8.2.2 implies that $(x^k + \Delta x^k, y_{\mathcal{A}_k}^{k+1}, z^{k+1})$ with $(\Delta x^k, y_{\mathcal{A}_k}^{k+1}, z^{k+1})$ determined by equation (8.11) in **(S2)** is a KKT point for the equality constrained quadratic program

$$\min \ f(x) \quad \text{s.t.} \quad b_j^T x = \beta_j \ (j = 1, \ldots, p), \ a_i^T x = \alpha_i \ (i \in \mathcal{A}_k). \tag{8.12}$$

We consider first the case that $\Delta x^k = 0$: This means that the x-component will not change, so no improvement in the objective value will occur.

Algorithm 8.2.1 Active-set method for quadratic programming

(S0) Choose $x^0 \in \mathbb{R}^n$ feasible for (8.10), $y^0 \in \mathbb{R}^m, z^0 \in \mathbb{R}^p$ and put $\mathcal{A}_0 := \left\{ i \mid a_i^T x^0 = \alpha_i \right\}$ and $k := 0$.

(S1) If (x^k, y^k, z^k) is a KKT point of (8.10): STOP.

(S2) Set $y_i^{k+1} := 0$ for $i \notin \mathcal{A}_k$ and determine $(\Delta x^k, y_{\mathcal{A}_k}^{k+1}, z^{k+1})$ as a solution of

$$\begin{pmatrix} Q & A_k^T & B^T \\ A_k & 0 & 0 \\ B & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ y_{\mathcal{A}_k} \\ z \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ 0 \\ 0 \end{pmatrix}. \tag{8.11}$$

(S3) Distinguish the following cases:

a) If $\Delta x^k = 0$ and $y_{\mathcal{A}_k}^{k+1} \geq 0$: STOP.

b) If $\Delta x^k = 0$ and $y_q^{k+1} := \min \left\{ y_i^{k+1} \mid i \in \mathcal{A}_k \right\} < 0$ put $x^{k+1} := x^k, \mathcal{A}_k := \mathcal{A}_k \backslash \{q\}$, and go to **(S4)**.

c) If $\Delta x^k \neq 0$ and $x^k + \Delta x^k$ is feasible for (8.10) put

$$x^{k+1} := x^k + \Delta x^k \quad \text{and} \quad \mathcal{A}_{k+1} := \mathcal{A}_k,$$

and got to **(S4)**.

d) If $\Delta x^k \neq 0$ and $x^k + \Delta x^k$ is infeasible for (8.10) determine and index r with

$$r \in \mathit{argmin} \left\{ \frac{\alpha_i - a_i^T x^k}{a_i^T \Delta x^k} \;\middle|\; i \notin \mathcal{A}_k : \; a_i^T \Delta x^k > 0 \right\},$$

put

$$\begin{aligned} t_k &:= \frac{\alpha_r - a_r^T x^k}{a_r^T \Delta x^k}, \\ x^{k+1} &:= x^k + t_k \Delta x^k, \\ \mathcal{A}_{k+1} &:= \mathcal{A}_k \cup \{r\}, \end{aligned}$$

and got to **(S4)**.

(S4) Set $k \leftarrow k + 1$, and got to **(S1)**.

If $y^{k+1}_{\mathcal{A}_k} \geq 0$ one sees immediately that the triple (x^k, y^{k+1}, z^{k+1}) with $y^{k+1}_i := 0$ ($i \notin \mathcal{A}_k$) is a KKT point of (8.10), which explains the stopping criterion in ***(S3)*** a).

If, in turn, $y^{k+1}_i < 0$ for some $i \in \mathcal{A}_k$ then on the one hand, we are not in a KKT point of (8.10), while on the other we do not reduce the objective function subject to the current restrictions see above. Therefore we relax these constraints by eliminating an index $q \in \mathcal{A}_k$ (so-called *inactivation step*). We choose this index to be such that the corresponding component y^{k+1}_q is maximally negative, see **(S3)** b) in Algorithm 8.2.1.

We now turn our attention to the case $\Delta x^k \neq 0$: If $x^k + \Delta x^k$ is feasible for (8.10), we accept the point $(x^k + \Delta x^k, y^{k+1}_{\mathcal{A}_k}, z^{k+1})$ as a new iterate without changing the index set $\mathcal{A}_k$, cf. step **(S3)** c).

If, in turn, $x^k + \Delta x^k$ is infeasible for (8.10), then one of the currently strict inequalities is violated. Instead of a full step $x^k + \Delta x^k$ we hence execute a step

$$x^k + t_k \Delta x^k$$

with a step-size $t_k > 0$ which is chosen just so that the inequalities $i \notin \mathcal{A}_k$ are satisfied at x^{k+1}. This yields the condition

$$a_i^T x^{k+1} = a_i^T x^k + t_k a_i^T \Delta x^k \leq \alpha_i \quad (i \in \mathcal{A}_k).$$

Since $a_i^T x^k \leq \alpha_i$ (recall that, by induction, x^k is feasible for (8.10)), this is automatically satisfied for all $i \notin \mathcal{A}_i$ with $a_i^{\Delta} x^k \leq 0$. Otherwise the above condition yields

$$t_k \leq \frac{\alpha_i - a_i^T x^k}{a_i^T \Delta^k}$$

or equivalently

$$t_k = \min \left\{ \frac{\alpha_i - a_i^T x^k}{a_i^T \Delta^k} \;\middle|\; i \notin \mathcal{A}_k : \; a_i^T \Delta x^k > 0 \right\}.$$

We add a newly active constraint to the index set $\mathcal{A}_k$ (so-called *activation step*). All in all, this explains the commands in step **(S3)** d).

We point out that

Exercises to Chapter 8

1. **(Gradient of quadratic functions)** Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$f(x) = \frac{1}{2} x^T M x + c^T x$$

where $M \in \mathbb{R}^{n \times n}$ is a quadratic (not necessarily symmetric) matrix and $c \in \mathbb{R}^n$. Compute $\nabla f(x)$ ($x \in \mathbb{R}^n$).

Hint: Do not use partial derivatives!

2. **(KKT conditions for general quadratic programming)** Consider the general quadratic program (8.2). Show that the following hold:

   a) If $\bar{x}$ is a local minimizer of (8.2) then there exist $\bar{y} \in \mathbb{R}^m$ and $\bar{z} \in \mathbb{R}^p$ such that $(\bar{x}, \bar{y}, \bar{z})$ solves the optimality conditions:

$$\begin{aligned} Qx + c + A^T y + B^T z &= 0, \\ Bx - b &= 0, \\ y_i \geq 0,\ [a - Ax]_i \geq 0,\ y_i[a - Ax]_i &= 0 \quad (i = 1, \ldots, m). \end{aligned} \tag{8.13}$$

   b) If the triple $(\bar{x}, \bar{y}, \bar{z})$ solves the optimality conditions (8.13) and Q is positive semidefinite then $\bar{x}$ solves the general quadratic program (8.2) (i.e. is a global minimizer).

3. **(Linear least squares problem)** For $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ consider

$$\min_{x \in \mathbb{R}^n} \ f(x) := \frac{1}{2} \|Ax - b\|_2^2. \tag{8.14}$$

   a) Show that $\operatorname{im} A^T A = \operatorname{im} A^T$.

   b) Prove that f is convex.

   c) Show that (8.14) always has a solution and that it is unique if $\operatorname{rank} A = n$.

9 Strategic games

9.1 Definition and examples of strategic games

We start with the central definition of this chapter.

Definition 9.1.1 *A (strategic) game (in normal form) is described by*

(a) a set $\{1, \ldots, N\}$ *of (finitely) many* players*;*

(b) strategy sets X_ν *for each player* $\nu = 1, \ldots, N$*;*

(c) payoff *or* utility functions $\theta_\nu : X \to \mathbb{R}$ *for each player* $\nu = 1, \ldots, N$ *where* $X := X_1 \times \ldots \times X_N$ *is the Cartesian product of all strategy sets.*

We will abbreviate such a game by $\Gamma = \{\theta_\nu, X_\nu\}_{\nu=1}^N$ *and call* Γ *an* N-person game*.*

We illustrate this concept by means of some very popular examples.

Example 9.1.2 (Prisoner's dilemma) *After a robbery two persons, 1 and 2, are arrested since they are under strong suspicion of having committed the crime. The police separate them to prevent them from communicating with each other and make the following offer to each one:*

- *If one of them confesses and the other one does not, the one the pleads guilty will be sentenced to only 1 year of prison (due to a key witness rule), while the other gets 10 years.*
- *If both confess they get 5 years each.*
- *If none of them confesses both of them will be sentenced to 2 years in prison due to illegal possession of firearms.*

This constitutes a game in the sense of Definition 9.1.1 with the set $\{1, 2\}$ *of 2 players, the strategy sets*

$$X_1 := X_2 := \{S, C\} \quad mit \quad S{:=}remain\ silent, \quad C{:=}confess$$

and the payoff functions

$$\theta_1 : X_1 \times X_2 \to \mathbb{R}, \quad \theta_2 : X_1 \times X_2 \to \mathbb{R},$$

defined elementwise by

$$\theta_1(S,S) := 2, \quad \theta_1(S,C) := 10, \quad \theta_1(C,S) := 1, \quad \theta_1(C,C) := 5$$

und

$$\theta_2(S,S) := 2, \quad \theta_2(S,C) := 1, \quad \theta_2(C,S) := 10, \quad \theta_2(C,C) := 5.$$

This is the formal description of the prisoner's dilemma as a strategic game. ◊

In Example 9.1.2 there were only 2 players and the corresponding strategy sets were finite. In this situation, the payoff functions are more handily described by a *payoff matrix* where, depending on the respective game, we speak of a *win* or *loss matrix*. In Example 9.1.2 the payoff matrices of both players read as follows:

Payoff matrix for player1

		Player 2	
		S	C
Player 1	S	2	10
	C	1	5

Payoff matrix for player 2

		Player 2	
		S	C
Player1	S	2	1
	C	10	5

Obviously, from an overall perspective, (S,S) is the best solution since then both players go to prison for 'only' 2 years. But unfortunately, the players are not allowed to communicate their answers and hence a player that makes the decision to reamain silent, runs a high risk of finally getting 10 years of prison if the other player decides to confess. To prevent this scenario, a strategically thinking player will choose to confess, since then he gets 5 years in the worst case (if the other player also confesses) and one year in the best case (if the other player remains silent). Hence, if both a players act strategically they will both choose to confess and end up with 5 years in prison. This is a 'dilemma' in as much they could get away with 2 years each if they could coordinate their response.

Example 9.1.3 (Battle of the sexes (gender neutral version)) *A couple wans to go to a concert. We call the Player 1and Player 2. Player 1 wants listen to concert of Bach's music where partner 2 (inexplicably) prefers Stravinsky. Thus, both players have the strategy sets*

$$X_1 := X_2 := \{\textit{Bach, Stravinsky}\}.$$

The payoff functions θ_1 and θ_2 can be represented by the following payoff matrix

		Player 2	
		Bach	*Stravinsky*
Player 1	*Bach*	$2,1$	$0,0$
	Stravinsky	$0,0$	$1,2,$

where the first entry in a tuple stands for the payoff of Player 1 and the second one for the payoff of Player 2 representing the respective value of the the occuring situation. Obviously, neither of them wants to go alone, so the value of different choices is set to 0.
The situation in this game is fundamentally different then in Example 9.1.2, since the players are able (and willing) to communicate. Hence, the tuple (Stravinsy, Bach) and (Bach, Stravinsky) will not occur. The homegeneous tuples (Bach, Bach) and (Strawinsky, Strawinsky) have equal accumulated value.

$\Diamond$

Example 9.1.4 (Cournot oligopoly) *A certain product is manufactured by N companies (where the economist will only speak of an oligopoly if N is small, but from a mathematical perspective we do not need this restriction). Let x_ν the amount produced by company μ, and let $c_\nu : x_\nu \mapsto c_\nu(x_\nu) \in \mathbb{R}$ be its cost function. We set $\xi := \sum_{\nu=1}^{N} x_\nu$ as the overall amount of the product and $p : \xi \mapsto p(\xi) \in \mathbb{R}$ the* inverse demand function[1] *which assigns to each amount ξ of the product the price per unit that the consumers are willing to pay when buying ξ units. Every company will aim at maximizing their profit and is hence facing the optimization problem*

$$\max_{x_\nu} \; x_\nu p\left(x_\nu + \sum_{\mu\neq\nu} x_\mu\right) - c_\nu(x_\nu) \quad \text{s.t.} \quad x_\nu \geq 0,$$

where x_μ ($\mu \neq \nu$) are the amounts of the product manufactured by the competitors of company ν. Company μ therefore has the strategy set $X_\nu := [0, +\infty)$ and the payoff function

$$\theta_\nu(x) := x_\nu p\left(x_\nu + \sum_{\mu\neq\nu} x_\mu\right) - c_\nu(x_\nu).$$

Example 9.1.4 differs from the foregoing in that it has infinite strategy sets (here even uncountable). These kinds of games are called *continuous games*.

9.2 Nash equilibria

In the foregoing section we studied some strategic games and sought for a solution in an intuive (somewhat heuristic) way. In this section we want to establish a formal notion

[1] The *demand function* $f : p \in \mathbb{R} \mapsto \xi \in \mathbb{R}$ assigns to every price p the overall demand $\xi = f(p)$ of the product at that price. Assuming that f be montonically decreasing, which from an economic perspective seems reasonable, f has an inverse function $p(\xi) := f^{-1}(\xi)$. This gives the inverse demand funtion.

of a solution for strategic games, namely the concept of a *Nash equilibrium*. We would like to point that there are different notions of a solution of a strategic game, but the Nash equilibrium is probably the most significant and prominent.

Here we assume w.l.o.g that the game we consider is a minimization problem.

Definition 9.2.1 *Let* $\Gamma = \{\theta_\nu, X_\nu\}_{\nu=1}^N$ *be a strategic game. A vector* $\bar{x} = (\bar{x}^\nu)_{\nu=1}^N$ *is called* Nash equilibrium *of* Γ *if*

$$\bar{x}^\nu \in X_\nu \quad \textit{and} \quad \theta_\nu(\bar{x}) \leq \theta_\nu(\bar{x}^1, \ldots, \bar{x}^{\nu-1}, x^\nu, \bar{x}^{\nu+1}, \ldots, \bar{x}^N) \quad (x^\nu \in X_\nu,\ \nu = 1, \ldots, N). \tag{9.1}$$

The interpretation of $\bar{x}$ being a Nash equilibrium of the game Γ is the folllowing: Given that all players $\mu \in \{1, \ldots, N\} \setminus \nu$ play the strategy $\bar{x}^\mu$, the strategy $\bar{x}^\nu$ is among the optimal ones for player ν. In other words, in a Nash equilibrium no player has an incentive to switch strategy provided that the other players play the Nash equilibrium strategy.

Clearly, if the game Γ we look at is a maximization problem we can either transform it with the usual techniques into a minimization problem and apply Definition 9.2.1 or we simply reverse the inequality in in the definition of a Nash equilibrium.

We introduce a very handy standard notation frequently used in game theory:

If for a vector $x = (x^1, \ldots, x^N)^T$ with the block components $x^\nu \in X_\nu, \nu = 1, \ldots, N$, we want to emphasize the ν-th component x^ν we write $x = (x^\nu, x^{-\nu})^T$, where $x^{-\nu}$ contains all block components x^μ with $\mu \neq \nu$. Therefore $(x^\nu, \bar{x}^{-\nu})$ denotes the vector $(\bar{x}^1, \ldots, \bar{x}^{\nu-1}, x^\nu, \bar{x}^{\nu+1}, \ldots, \bar{x}^N)^T$ where we substituted $\bar{x}^\nu$ is for x^ν. Using this notation $\bar{x} = (\bar{x}^\nu)_{\nu=1}^N$ is a Nash equilibrium of $\Gamma = \{\theta_\nu, X_\nu\}_{\nu=1}^N$ if

$$\bar{x}^\nu \in X_\nu \quad \text{and} \quad \theta_\nu(\bar{x}) \leq \theta_\nu(x^\nu, \bar{x}^{-\nu}) \quad (x^\nu \in X_\nu,\ \nu = 1, \ldots, N).$$

This is equivalent to saying that, for all $\nu = 1, \ldots, N$, the vector $\bar{x}^\nu$ solves the optimization problem

$$\min_{x^\nu} \ \theta_\nu(x^\nu, \bar{x}^{-\nu}) \quad \text{s.t.} \quad x^\nu \in X_\nu. \tag{9.2}$$

We point out that the obejctive function $\theta_\nu(\cdot, \bar{x}^{-\nu})$ in (9.2) explicitly assumes knowledge of the 'optimal' strategies $\bar{x}^{-\nu}$ which player ν does usually not have.

The notion of a Nash equilibrium can be reformulated in yet another way: For these purposes, the following notion of a *best-response (BR) function* is key.

Definition 9.2.2 (Best-response function) *Let* $\Gamma = \{\theta_\nu, X_\nu\}_{\nu=1}^N$ *be a strategic game and let* $x = (x^1, \ldots, x^N)^T \in X$. *The (set-valued) mapping*

$$x^{-\nu} \in X_{-\nu} \mapsto \mathcal{S}_\nu(x^{-\nu}) := \operatorname*{argmin}_{x^\nu} \left\{ \theta_\nu(x^\nu, x^{-\nu}) \mid x^\nu \in X_\nu \right\} \subset X_\nu \tag{9.3}$$

is called the best-response (BR) function *of player* $\nu = 1, \ldots, N$. *The (set-valued) mapping*

$$x \in X \mapsto S_1(x^{-1}) \times \cdots \times S(x^{-N}) \subset X$$

is called the best-response function *of* Γ.

Using the concept of best-response function we immediately obtain another characterization of a Nash equilibrium.

Theorem 9.2.3 *Let* $\Gamma = \{\theta_\nu, X_\nu\}_{\nu=1}^N$ *be a strategic game. Then* $\bar{x}$ *is a Nash equilibrium of* Γ *if and only if* $\bar{x}^\nu \in \mathcal{S}_\nu(\bar{x}^{-\nu})$ *for all* $\nu = 1, \ldots, N$, *i.e. if* $\bar{x} \in S(\bar{x})$.

Theorem 9.2.3 says that $\bar{x} \in X$ is a Nash equlibrium of the game Γ if and only if it is a (set-valued) fixed point of the corresponding BR function.

We will now apply the concept of a Nash equilibrium to the examples of the foregoing section.

Example 9.2.4 *a) The prisoner's dilemma from Example 9.1.2 has exactly one Nash equilibrium, namely (C,C). This coincides with what we have intuitively infered assuming that both players act strategically.*

b) The battle of the sexes from Example 9.1.3 has two Nash equilibria, namely (Bach, Bach) and (Stravinsky, Stravinsky). None of these two points is preferable with respect to the Nash equilibrium concept.

We now want to apply the concept of a Nash equilibrium to the Cournot oligopoly from Example 9.1.4, where for simplicity we consider only two companies, i.e. a *duopoly*.

Example 9.2.5 (Cournot duopoly) *Consider the oligopoly model from Example 9.1.4 for* $N = 2$ *companies with linear cost functions*

$$c_\nu : x_\nu \mapsto \alpha x_\nu \quad (\nu = 1, 2)$$

for some constant $\alpha > 0$ *and the inverse demand function*

$$p : \xi \mapsto \begin{cases} \gamma - \xi, & \text{falls } \xi \leq \gamma, \\ 0, & \text{falls } \xi \geq \gamma, \end{cases}$$

for another constant $\gamma > \alpha$, *see also Figure 9.1.*
Therefore the the objective function (to be maximized *w.r.t.* x_1*) for company 1 reads*

$$\theta_1 : (x_1, x_2) \mapsto x_1 p(x_1 + x_2) - c_1(x_1) = \begin{cases} x_1(\gamma - x_1 - x_2) - \alpha x_1 & \text{if} \quad x_1 + x_2 \leq \gamma, \\ -\alpha x_1, & \text{if} \quad x_1 + x_2 \geq \gamma. \end{cases}$$

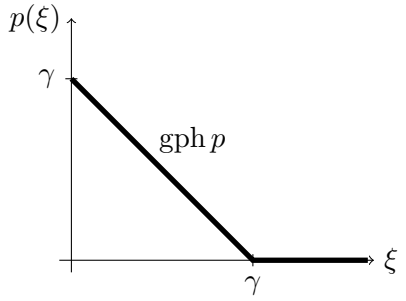

Figure 9.1: Inverse demand function p

Analogously, for company 2 we have the objective function

$$\theta_2(x_1,x_2) := x_2 p(x_1+x_2) - c_2(x_2) = \begin{cases} x_2(\gamma - x_1 - x_2) - \alpha x_2, & \text{if} \quad x_1+x_2 \le \gamma, \\ -\alpha x_2, & \text{if} \quad x_1+x_2 \ge \gamma. \end{cases}$$

Elementary considerations yield that given $x_1 \ge 0$ and $x_2 \ge 0$, respectively, the BR functions for either player read

$$\mathcal{S}_1(x_2) = \begin{cases} \frac{1}{2}(\gamma - \alpha - x_2), & \text{falls } x_2 \le \gamma - \alpha, \\ 0, & \text{falls } x_2 \ge \gamma - \alpha \end{cases}$$

and

$$\mathcal{S}_2(x_1) = \begin{cases} \frac{1}{2}(\gamma - \alpha - x_1) & \text{if} \quad x_1 \le \gamma - \alpha, \\ 0, & \text{if} \quad x_1 \ge \gamma - \alpha. \end{cases}$$

Now, $\bar{x} = (\bar{x}_1, \bar{x}_2)$ is a Nash equilibrium if and only if $\bar{x}_1 \in \mathcal{S}_1(\bar{x}_2)$ and $\bar{x}_2 \in \mathcal{S}_2(\bar{x}_1)$ holds. Since both BR functions are actually single-valued, this is equivalent to saying that

$$\bar{x}_1 = \begin{cases} \frac{1}{2}(\gamma - \alpha - \bar{x}_2) & \text{if} \quad \bar{x}_2 \le \gamma - \alpha, \\ 0, & \text{if} \quad \bar{x}_2 \ge \gamma - \alpha \end{cases}$$

and

$$\bar{x}_2 = \begin{cases} \frac{1}{2}(\gamma - \alpha - \bar{x}_1) & \text{if} \quad \bar{x}_1 \le \gamma - \alpha, \\ 0 & \text{if} \quad \bar{x}_1 \ge \gamma - \alpha. \end{cases}$$

An elementary geometric consideration shows (cf. Figure 9.2) that $\bar{x} := \big(\frac{1}{3}(\gamma-\alpha), \frac{1}{3}(\gamma-\alpha)\big)$ is the only Nash equilibrium.

$\Diamond$

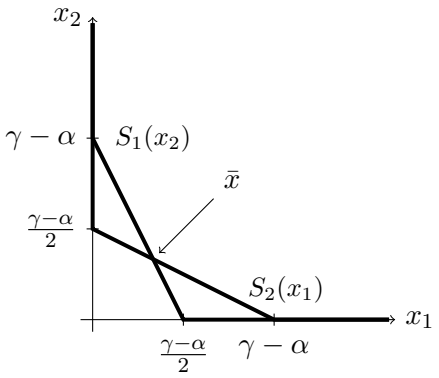

Figure 9.2: Nash equilibrium in the Cournot duopoly from Example 9.2.5

9.3 Matrix games

In this section we consider *matrix games* in maximization form. These are finite 2-person zero-sum games of the form

$$\begin{array}{ll|ll} \max_{x^1} & \theta_1(x) & \max_{x^2} & \theta_2(x) \\ \text{s.t.} & x^1 \in X_1 & \text{s.t.} & x^2 \in X_2 \end{array}$$

with

$$\theta_1(x) = -\theta_2(x) \quad (x \in X := X_1 \times X_2) \tag{9.4}$$

As the strategy sets X_1 and X_2 are finite we may write $X_1 = \{1, \ldots, m\}$ and $X_2 = \{1, \ldots, n\}$. The payoff functions can be conveniently represented by payoff matrices. Let $A \in \mathbb{R}^{m \times n}$ with

$$a_{ij} := \theta_1(i, j) \quad (i = 1, \ldots, m,\ j = 1, \ldots, n)$$

be the payoff matrix for player 1. Then $B := -A$ is the payoff matrix for player 2.

Before we start our general study of matrix games we consider an example.

Example 9.3.1 *We consider a matrix game with $m = 2, n = 3$ and a payoff matrix*

$$A = \begin{pmatrix} 3 & 1 & 8 \\ 4 & 10 & 0 \end{pmatrix}.$$

Which strategy will player 1 (row player) choose? If he plays his first strategy (first row, i=1) then his minimum payoff is 1, while for the second strategy (second row, i=2) his

minumum payoff is 0. Taking the maximum of both minimum payoffs, his minimum payoff is

$$\max_i \min_j a_{ij} = 1.$$

Player 2 (column player) acts analogously: If he chooses his first strategy (firs column, j=1), he has a maximum loss (=minimum payoff) of 4, for the second strategy (second column, j=2) this amounts to 10, while for the third strategy (third row, j=3) his maximum loss is 8*. Minimizing his maximum loss results in*

$$\min_j \max_i a_{ij} = 4.$$

The approach of maximizing the minimum payoff, called *minmax strategy*, is called *pessimistic* in the literature, where minimizing one's maximum loss *maxmin strategy* is called *opportunistic*.

We will now show that, in a matrix game, the minimum payoff is always a lower bound for the maximum loss.

Lemma 9.3.2 *For every matrix game with the payoff matrix $A \in \mathbb{R}^{m\times n}$ we have*

$$\max_i \min_j a_{ij} \leq \min_j \max_i a_{ij}. \tag{9.5}$$

Proof: For every row index $\ell \in \{1, \ldots, n\}$ and all $i \in \{1, \ldots, m\}$we have

$$\min_j a_{ij} \leq a_{i\ell}.$$

T his implies

$$\max_i \min_j a_{ij} \leq \max_i a_{i\ell}.$$

Since this holds for every $\ell \in \{1, \ldots, n\}$, writing j instead of ℓ, we obtain

$$\max_i \min_j a_{ij} \leq \min_j \max_i a_{ij}.$$

□

The numbers

$$\underline{v} := \max_i \min_j a_{ij} \quad \text{and} \quad \bar{v} := \min_j \max_i a_{ij}$$

are also called *lower game value* or *upper game value*, respectively. Using this new notation, Lemma 9.3.2 simply reads $\underline{v} \leq \bar{v}$. In what follows we are particularly interested in the case where equality holds.

We make some preliminary considerations. For these purposes, let $(\bar{i}, \bar{j})$ be a Nash equilibrium of a matrix game given by $A = (a_{ij}) \in \mathbb{R}^{m\times n}$. Then

$$\begin{aligned}
&\theta_1(\bar{i},\bar{j}) \geq \theta_1(i,\bar{j}) \ (i=1,\dots,m) \quad \text{and} \quad \theta_2(\bar{i},\bar{j}) \geq \theta_2(\bar{i},j) \ (j=1,\dots,n)\\
&\overset{\theta_1=-\theta_2}{\iff} \quad \theta_1(\bar{i},\bar{j}) \geq \theta_1(i,\bar{j}) \ (i=1,\dots,m) \quad \text{and} \quad \theta_1(\bar{i},\bar{j}) \leq \theta_1(\bar{i},j) \ (j=1,\dots,n)\\
&\iff \quad \theta_1(i,\bar{j}) \leq \theta_1(\bar{i},\bar{j}) \leq \theta_1(\bar{i},j) \quad (i=1,\dots,m, j=1,\dots,n)\\
&\iff \quad a_{i\bar{j}} \leq a_{\bar{i}\bar{j}} \leq a_{\bar{i}j} \quad (i=1,\dots,m, j=1,\dots,n).
\end{aligned}$$

A strategy pair $(\bar{i}, \bar{j})$ with the property

$$a_{i\bar{j}} \leq a_{\bar{i}\bar{j}} \leq a_{\bar{i}j} \quad (i=1,\dots,m, j=1,\dots,n).$$

is called a *saddle point* of the matrix game represented by A. We have thus proven the following result.

Theorem 9.3.3 *Let a matrix game be given by the matrix $A \in \mathbb{R}^{m\times n}$. Then a strategy pair $(\bar{i}, \bar{j})$ is a Nash equlibrium if and only if it is a saddle point.*

The next result characterizes the existence of a saddle point (i.e. Nash equilibrium) of a matrix game.

Theorem 9.3.4 *Let a matrix game be given by the matrix $A \in \mathbb{R}^{m\times n}$. Then there exists a saddle point (i.e. Nash equilibrium) if and only if equality holds in* (9.5).

Proof: First let $(\bar{i}, \bar{j})$ be a saddle point of the given matrix game, i.e.

$$a_{i\bar{j}} \leq a_{\bar{i}\bar{j}} \leq a_{\bar{i}j} \quad (i=1,\dots,m, \ \ j=1,\dots,n).$$

This implies

$$\max_i a_{i\bar{j}} \leq a_{\bar{i}\bar{j}} \leq \min_j a_{\bar{i}j},$$

which, in turn, yields

$$\overline{v} = \min_j \max_i a_{ij} \leq a_{\bar{i}\bar{j}} \leq \max_i \min_j a_{ij} = \underline{v}.$$

By Lemma 9.3.2 we thus have $\overline{v} = \underline{v}$.

In turn, let

$$\underline{v} = \overline{v} \iff \max_i \min_j a_{ij} = \min_j \max_i a_{ij}. \tag{9.6}$$

By the definiton of $\underline{v}$ and $\overline{v}$, respectively, there exist $\bar{i}$ and $\bar{j}$ such that

$$\underline{v} = \min_j a_{\bar{i}j} \quad \text{and} \quad \overline{v} = \max_i a_{i\bar{j}}.$$

From the inequality

$$\underline{v} = \min_j a_{\bar{i}j} \leq a_{\bar{i}\bar{j}} \leq \max_i a_{i\bar{j}} = \overline{v}$$

and the assumption that $\underline{v} = \overline{v}$ we therefore infer

$$\min_j a_{\bar{i}j} = a_{\bar{i}\bar{j}} = \max_i a_{i\bar{j}}.$$

But this implies

$$a_{\bar{i}\bar{j}} \geq a_{i\bar{j}} \ \ (i = 1, \ldots, m) \quad \text{and} \quad a_{\bar{i}\bar{j}} \leq a_{\bar{i}j} \ \ \forall j = 1, \ldots, n,$$

i.e. $(\bar{i}, \bar{j})$ is a saddle point of the matrix game. □

Due to Theorem 9.3.4, for a matrix game with payoff matrix $A = (a_{ij})$ that has a Nash equlibrium, the following number

$$v := \max_i \min_j a_{ij} = \max_j \min_i a_{ij} \quad \Longleftrightarrow \quad v := \underline{v} = \overline{v}$$

exists uniquely, and is called the *value* of the matrix game.

Next we consider a special kind of matrix game.

Definition 9.3.5 *A matrix game with payoff matrix $A \in \mathbb{R}^{m\times n}$ is called* symmetric *if $A = -A^T$ (i.e. when A is skew symmetric).*

Note that for a skew symmetric matrix A the diagonal elements are necessarily 0.

For a symmetric matrix game we have the following characterization of a saddle point.

Theorem 9.3.6 *A symmetric matrix game defined by the (skew symmetric matrix) $A \in \mathbb{R}^{n\times n}$ has a saddle point (and hence Nash-equilibrium) if and only if there exists and index $\bar{i} \in \{1, \ldots, n\}$ such that*

$$a_{\bar{i}j} \geq 0 \quad (j = 1, \ldots, n). \tag{9.7}$$

In this case $(\bar{i}, \bar{i})$ is a saddle point and $v = 0$ is the value of the matrix game.

Proof: Let $(\bar{i}, \bar{j})$ be a saddle point, i.e.

$$a_{\bar{i}j} \geq a_{\bar{i}\bar{j}} \geq a_{i\bar{j}} \quad (i, j = 1, \ldots, n).$$

In particular, for $i = \bar{j}$ this implies $a_{\bar{i}j} \geq a_{\bar{j}\bar{j}} = 0$ for all $j = 1, \ldots, n$.

In turn let $\bar{i} \in \{1, \ldots, n\}$ be an index with the property (9.7). Since $A = -A^T$, i.e. $a_{ij} = -a_{ji}$ for all $i, j = 1, \ldots, n$, this implies

$$a_{j\bar{i}} = -a_{\bar{i}j} \leq 0 \quad (j = 1, \ldots, n).$$

Hence

$$a_{\bar{i}j} \geq 0 \geq a_{j\bar{i}} \quad (j = 1, \ldots, n).$$

As $a_{\bar{i}\bar{i}} = 0$ the index pair $(\bar{i}, \bar{i})$ is a saddle point of the matrix game. □

The criterion (9.7) for symmetric matrix games is obviously very handy: A saddle point (i.e. a Nash equilibrium) exists if and only if there is a row of the payoff matrix that contains only nonnegative entries. For instance, the matrix

$$A := \begin{pmatrix} 0 & -1 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$$

defines symmetric matrix game where $\bar{i} := 2$ satisfies the criterion from Theorem 9.3.6 so that $(2, 2)$ is saddle pointe, thus a Nash equilibrium. On the other hand, Theorem 9.3.6 already shows that a matrix game need not have a Nash equlibrium. We will pursue this question now in more detail.

Thus far, we have only considered matrix games in *pure strategies*, i.e. each player decides for exactly one strategy from a given strategy set. If a game is played multiple times, it is reasonable to assume that each players chooses a strategy with a certain probability.

Player 1 chooses $x \in \mathbb{R}^m$ with

$$x = (x_1, \ldots, x_m)^T, \quad \sum_{i=1}^{m} x_i = 1, \quad x_i \geq 0 \quad (i = 1, \ldots, m),$$

Spieler 2 chooses $y \in \mathbb{R}^n$ with

$$y = (y_1, \ldots, y_n)^T, \quad \sum_{j=1}^{n} y_j = 1, \quad y_j \geq 0 \quad (j = 1, \ldots, n).$$

Here x_i and y_j are the probability with which player 1 and 2 play the strategy i and j, respectively. We call this a matrix game in *mixed strategies*. If x or y is a standard unit vector this corresponds to a pure strategy.

For a matrix game in mixed strategies the strategy sets for player 1 and player 2 are given by

$$\hat{X} := \left\{ x \in \mathbb{R}^m \;\middle|\; \sum_{i=1}^{m} x_i = 1,\ x_i \geq 0\ (i = 1, \ldots, m) \right\}$$

and

$$\hat{Y} := \left\{ y \in \mathbb{R}^n \;\middle|\; \sum_{j=1}^{n} y_j = 1,\ y_j \geq 0\ (j = 1, \ldots, n) \right\},$$

respectively. If player 1 chooses the mixed strategy $x \in \hat{X}$ and player 2 chooses $y \in \hat{Y}$ the expected payoff for player 1 is given by

$$\theta(x,y) := \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j a_{ij} = x^T A y.$$

We can now extend the notion of a Nash equibrium to matrix games in mixed strategies.

Definition 9.3.7 *Let $A \in \mathbb{R}^{m\times n}$ and let $\hat{X}, \hat{Y}$ and θ be defined as above. Then $(\bar{x},\bar{y}) \in \hat{X}\times\hat{Y}$ is said to be a* Nash equilibrium in mixed strategies *for the matrix game defined by A if*

$$\theta(\bar{x},\bar{y}) \geq \theta(x,\bar{y}) \quad (x \in \hat{X}) \quad \textit{and} \quad \theta(\bar{x},\bar{y}) \leq \theta(\bar{x},y) \quad (y \in \hat{Y}),$$

i.e. if $(\bar{x},\bar{y})$ is a saddle point of θ.

Apparently, Definition 9.3.7 is exactly what results in applying the standard definition of a Nash equilibrium (Definition 9.2.1) to the 2-person game with the payoff functions $\theta_1 := \theta$ and $\theta_2 = -\theta$ and the strategy sets $X_1 := \hat{X}$ and $X_2 := \hat{Y}$ for player 1 and 2, respectively.

The next result is of practical and theoretical significance since, on the one hand it

Theorem 9.3.8 *Consider a matrix game given by $A \in \mathbb{R}^{m\times n}$. Moreoever, let $\theta, \hat{X}$ and $\hat{Y}$ be defined as above . Then $(\bar{x},\bar{y})$ is a Nash equilibrium in mixed strategies if and only if the following conditions hold:*

(a) $\bar{x}$ (together with $\bar{v} \in \mathbb{R}$) solves the (primal) linear program

$$\max_{v,x}\ v \quad \text{s.t.} \quad x^T A \geq v 1_n^T,\ \sum_{i=1}^{m} x_i = 1, x \geq 0. \tag{9.8}$$

(b) $\bar{y}$ (together $\bar{w} \in \mathbb{R}$) solves the (dual) linear program

$$\min_{w,y}\ w \quad \text{s.t.} \quad Ay \leq w 1_m,\ \sum_{j=1}^{n} y_j = 1, y \geq 0. \tag{9.9}$$

Here $1_k := (1,\ldots,1)^T$ denotes the vector of all ones in $\mathbb{R}^k$.

Proof: First let $(\bar{x},\bar{y})$ be a Nash equilibrium in mixed strategies, i.e.

$$\theta(x,\bar{y}) \leq \theta(\bar{x},\bar{y}) \leq \theta(\bar{x},y) \quad (x \in \hat{X},\ y \in \hat{Y}). \tag{9.10}$$

Put

$$\bar{v} := \bar{x}^T A \bar{y} \quad \text{and} \quad \bar{w} := \bar{x}^T A \bar{y}.$$

We claim that $(\bar{v}, \bar{x})$ solves the linear program (9.8) and $(\bar{w}, \bar{y})$ solves the linear program (9.9). First consider $(\bar{v}, \bar{x})$. From (9.10) we infer that $\bar{v} \leq \bar{x}^T A y$ for all $y \in \hat{Y}$. In particular, for $y = e_j \in \hat{Y}$ we obtain $\bar{v} \leq [\bar{x}^T A]_j$ for all $j = 1, \ldots, n$. This yields that $(\bar{v}, \bar{x})$ is feasible for (9.8). Assume $(\bar{v}, \bar{x})$ were not optimal. Then there exists $(\hat{v}, \hat{x})$ feasible for (9.8) with $\hat{v} > \bar{v}$. Multiplying $[\hat{x}^T A]_j \geq \bar{v}$ by $\bar{y}_j \geq 0$ and and summing up over $j = 1, \ldots, n$ yields

$$\hat{x}^T A \bar{y} = \sum_{j=1}^{n} [\hat{x}^T A]_j \bar{y}_j \geq \hat{v} \sum_{j=1}^{n} \bar{y}_j = \hat{v} > \bar{v} = \bar{x}^T A \bar{y},$$

which contradicts (9.10). Hence, $(\bar{v}, \bar{x})$ indeed solves (9.8).

Now conisder $(\bar{w}, \bar{y})$. From (9.10) we infer that $x^T A \bar{y} \leq \bar{x}^T A \bar{y} = \bar{w}$ for all $x \in \hat{X}$, in particular, for $x = e_i \in \mathbb{R}^m$ we have $[A\bar{y}]_i \leq \bar{w}$ and hence $(\bar{w}, \bar{y})$ is feasible for (9.9). Assume there were $(\hat{w}, \hat{y})$ feasible for (9.9) such that $\hat{w} < \bar{w}$. Multiplying the inequality $[A\hat{y}]_i \leq \hat{w}$ by $\bar{x}_i \geq 0$ and summing over all $i = 1, \ldots, m$ yields

$$(\bar{x})^T A \hat{y} = \sum_{i=1}^{m} \bar{x}_i [A\hat{y}]_i \leq \sum_{i=1}^{m} \bar{x}_i \hat{w} = \hat{w} < \bar{w} = \bar{x}^T A \bar{y}$$

which, again, contradicts (9.10). Hence $(\bar{w}, \bar{y})$ is a solution of (9.9).

In order to prove the converse implication, first observe that (9.8) and (9.9) are dual to each other, cf. Exercise 2.

Now let $(\bar{v}, \bar{x})$ be a solution of (9.8) and $(\bar{w}, \bar{y})$ a solution of (9.9). From the constraints of (9.8) and (9.9), respectively, as well as the strong duality theorem (Theorem 2.4.6) for all $x \in \hat{X}$ and all $y \in \hat{Y}$ we obtain

$$x^T A \bar{y} \leq \bar{w} 1_m^T x = \bar{w} = \bar{v} = \bar{v} 1_n^T y \leq \bar{x}^T A y.$$

Analogously, from

$$\bar{w} = \bar{w} 1_m^T \bar{x} = (\bar{w} 1_m)^T \bar{x} \geq (A\bar{y})^T \bar{x} = \big(\bar{x}^T A\big)\bar{y} \geq \bar{v} 1_n^T \bar{y} = \bar{v} = \bar{w}$$

we infer $\bar{w} = \bar{x}^T A \bar{y} = \bar{v}$. Consequently,

$$\theta(x, \bar{y}) \leq \theta(\bar{x}, \bar{y}) \leq \theta(\bar{x}, y) \quad (x \in \hat{X},\ y \in \hat{Y}),$$

thus $(\bar{x}, \bar{y})$ is a Nash equilibrium in mixed strategies. □

We consider a small example to illustrate the findings of Theorem 9.3.8.

Example 9.3.9 *Consider a matrix game with the payoff matrix*

$$A := \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix}.$$

Obviously there does not exist a Nash equilibrium in pure strategies. Solving the linear programs from Theorem 9.3.8 yields the solutions

$$\bar{x} := \left(\frac{3}{4}, \frac{1}{4}\right)^T \quad \textit{and} \quad \bar{y} := \left(\frac{1}{2}, \frac{1}{2}\right)^T$$

with the value $\bar{v} := \frac{1}{2}$*. Hence, in the long run, player 1 can expect a payoff of* $\frac{1}{2}$ *if he plays srategy* 1 *with a probability of* 75%*and strategy with* 25%*. Player 2, in turn, will choose both of his strategies with a probability of* 50% *to contain his loss at a level of* $\frac{1}{2}$ *per game.* ♢

As a consequence of Theorem 9.3.8 we obtain the following result.

Theorem 9.3.10 *Every matrix game has a Nash equilibrium in mixed strategies.*

Proof: The linear programs in Theorem 9.3.8 are dual to each other. Since both are feasible, by weak duality they are both bounded, hence by Corollary 2.4.8 they both have a solution. Therefore Theorem 9.3.8 gives the assertion. □

Exercises to Chapter 9

1. **(Rock-paper-scissors)** Formulate the (in)famous rock-papers-scissors as a strategic game and check for Nash equilibria.

2. Show that the linear programs (9.8) and (9.9) are dual to each other.

Bibliography

[1] D. Bertsimas and J.N Tsitsiklis: *Linear Optimization.* Athena Scientific, Belmont, Massachusetts, 1997.

[2] G. Dantzig: *Linear Programming an Extensions.* Princeton University Press, Princeton, NJ.

[3] F. Jarre und J. Stoer: *Optimierung.* Springer, 2004.

[4] N. Karmarkar: *A new polynomial-time algorithm for linear programming.* Combinatorica 4, 1984, pp. 373–395.

[5] V. Klee und G.J. Minty: *How good is the simplex algorithm ?* In: O. Shisha (ed.) Inequalities. Academic Press, New York, pp. 159–175.

[6] J. Matoušek and B. Gärtner: *Understanding and Using Linear Programming.* Springer, 2007.

[7] S. Mehrotra: *On the Implementation of a Primal-Dual Interior Point Method.* SIAM Journal on Optimization, 2, 1992, pp. 575-601.

[8] S.J. Wright: *Primal-Dual Interior-Point Methods.* SIAM, 1997.

[9] M.H. Wright: *The interior-point revolution in optimization: History, recent developments, and lasting consequences.* Bulletin of the American Mathematical Society 42, 2005, pp. 39–56.

Index