Johannes Jahn

Introduction to the Theory of Nonlinear Optimization

Fourth Edition

Springer

Introduction to the Theory of Nonlinear Optimization

Johannes Jahn

Introduction to the Theory of Nonlinear Optimization

Fourth edition

Springer

Johannes Jahn
Department of Mathematics
University of Erlangen-Nürnberg
Erlangen, Germany

ISBN 978-3-030-42759-7 ISBN 978-3-030-42760-3 (eBook)
https://doi.org/10.1007/978-3-030-42760-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Claudia and Martin

Preface

This book presents an application-oriented introduction to the theory of nonlinear optimization. It describes basic notions and conceptions of optimization in the setting of normed or even Banach spaces. Various theorems are applied to problems in related mathematical areas. For instance, the Euler–Lagrange equation in the calculus of variations, the generalized Kolmogorov condition, and the alternation theorem in approximation theory as well as the Pontryagin's maximum principle in optimal control theory are derived from general results of optimization.

Because of the introductory character of this text, it is not intended to give a complete description of all the approaches in optimization. For instance, investigations on conjugate duality, sensitivity, stability, recession cones, and other concepts are not included in this book.

The bibliography gives a survey of books in the area of nonlinear optimization and related areas like approximation theory and optimal control theory. Important papers are cited as footnotes in the text.

This third edition is an enlarged and revised version containing an additional chapter on extended semidefinite optimization and an updated bibliography.

I am grateful to S. Geuß, S. Gmeiner, S. Keck, Prof. Dr. E.W. Sachs, and H. Winkler for their support, and I am especially indebted to D.G. Cunningham, Dr. G. Eichfelder, Dr. F. Hettlich, Dr. J. Klose, Prof. Dr. E.W. Sachs, Dr. T. Staib, and Dr. M. Stingl for fruitful discussions.

Erlangen, Germany Johannes Jahn
September 2006

Remarks to the Fourth Edition

This fourth edition extends the previous one by a recent theory on discrete-continuous optimization problems together with special separation results. I thank Dr. M. Knossalla for joint investigations of this subject. This edition uses a special layout designed for e-books. Some figures are added and all the figures are adapted to the e-book design.

Erlangen, Germany Johannes Jahn
November 2019

Contents

1 Introduction and Problem Formulation

In optimization one investigates problems of the determination of a minimal point of a functional on a nonempty subset of a real linear space. To be more specific this means: Let X be a real linear space, let S be a nonempty subset of X, and let $f : S \to \mathbb{R}$ be a given functional. We ask for the minimal points of f on S. An element $\bar{x} \in S$ is called a *minimal point* of f on S if

$$f(\bar{x}) \leq f(x) \text{ for all } x \in S,$$

and the functional value $f(\bar{x})$ is called *minimal value* of the optimization problem. The set S is also called *constraint set*, its elements are called *feasible elements*, and the functional f is called *objective functional*.

In order to introduce optimization we present various typical optimization problems from Applied Mathematics. First we discuss a design problem from structural engineering.

Example 1.1 (structural engineering).

As a simple example consider the design of a beam with a rectangular cross-section and a given length l (see Figs. 1.1 and 1.2). The height x_1 and the width x_2 have to be determined.

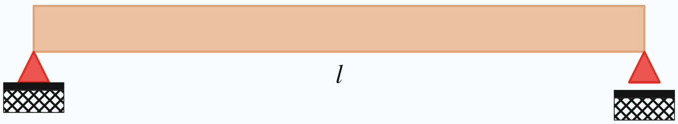

Fig. 1.1 Longitudinal section of the beam

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3_1

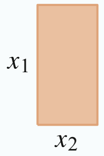

Fig. 1.2 Cross-section of the beam

The design variables x_1 and x_2 have to be chosen in an area which makes sense in practice. A certain stress condition must be satisfied, i.e. the arising stresses cannot exceed a feasible stress. This leads to the inequality

$$2000 \leq x_1^2 x_2. \tag{1.1}$$

Moreover, a certain stability of the beam must be guaranteed. In order to avoid a beam which is too slim we require

$$x_1 \leq 4x_2 \tag{1.2}$$

and

$$x_2 \leq x_1. \tag{1.3}$$

Finally, the design variables should be nonnegative which means

$$x_1 \geq 0 \tag{1.4}$$

and

$$x_2 \geq 0. \tag{1.5}$$

Among all feasible values for x_1 and x_2 we are interested in those which lead to a light construction. Instead of the weight we can also take the volume of the beam given as lx_1x_2 as a possible criterion (where we assume that the material is homogeneous). Consequently, we minimize lx_1x_2 subject to the constraints (1.1), ...,(1.5). This problem can be formalized as

$$\begin{gathered}
\min \ lx_1x_2 \\
\text{subject to the constraints} \\
2000 - x_1^2 x_2 \leq 0 \\
x_1 - 4x_2 \leq 0 \\
-x_1 + x_2 \leq 0 \\
-x_1 \leq 0 \\
-x_2 \leq 0 \\
x_1,\ x_2 \in \mathbb{R}.
\end{gathered}$$

Many optimization problems do not only have continuous variables and a few number of optimization problems may have constraints, which are only valid under a certain condition. The following very simple example describes such a problem with discrete-continuous variables and a conditional constraint.

Example 1.2 (discrete-continuous variables and vanishing constraints).

We investigate the simple optimization problem

$$\begin{gathered} \min\ x_2^2 + \sin x_1 \\ \text{subject to the constraints} \\ 1 \le x_1 \le 2\pi \\ \cos x_1 \ge 0.4,\ \text{if} x_1 \in (\pi, 2\pi] \\ x_1 \in \mathbb{Z},\ x_2 \in \mathbb{R}. \end{gathered} \tag{1.6}$$

We have a discrete and a continuous variable, and one inequality constraint is only satisfied, if $x_1 \in (\pi, 2\pi]$. For $x_1 \notin (\pi, 2\pi]$ the constraint $\cos x_1 \ge 0.4$ is void, i.e. it vanishes. Therefore, such a constraint is also called *vanishing constraint*.[1] This class of problems with vanishing constraints, which appears in structural optimization, is difficult to treat.
Figure 1.3 illustrates the feasible integers of this problem. It is obvious that (6, 0) is the unique minimal solution of problem (1.6).

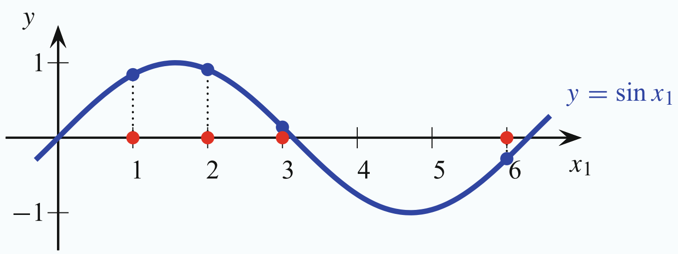

Fig. 1.3 Illustration of the sine function values at the four feasible points x_1

If we introduce a superset $\hat{S}$ of the feasible set S, we can collect all constraints, which cannot be written in the form of inequalities or equalities.

[1] Achtziger, W. and Kanzow, C.: "Mathematical programs with vanishing constraints: optimality conditions and constraint qualifications", *Math. Program., Ser. A* 114 (2008) 69–99.

Hence, we rewrite problem (1.6) as

$$\begin{aligned} &\min\ x_2^2 + \sin x_1 \\ &\text{subject to the constraints} \\ &1 - x_1 \leq 0 \\ &x_1 - 2\pi \leq 0 \\ &(x_1, x_2) \in \hat{S} \end{aligned}$$

with

$$\hat{S} := \left\{(x_1, x_2) \in \mathbb{Z} \times \mathbb{R} \mid x_1 \in (\pi, 2\pi] \Rightarrow \cos x_1 \geq 0.4\right\}.$$

So, if there are not only constraints in the form of inequalities or equalities, it makes sense to consider a superset $\hat{S}$.

With the next example we present a simple optimization problem from the calculus of variations.

Example 1.3 (calculus of variations).

In the calculus of variations one investigates, for instance, problems of minimizing a functional f given as

$$f(x) = \int_a^b l(x(t), \dot{x}(t), t)\, dt$$

where $-\infty < a < b < \infty$ and l is argumentwise continuous and continuously differentiable with respect to x and $\dot{x}$. A simple problem

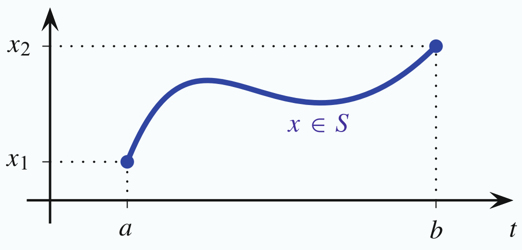

Fig. 1.4 Illustration of a feasible element $x \in S$

of the calculus of variations is the following: Minimize f subject to the class of curves from

$$S := \{x \in C^1[a, b] \mid x(a) = x_1 \text{ and } x(b) = x_2\}$$

where x_1 and x_2 are fixed endpoints (see Fig. 1.4).

In control theory there are also many problems which can be formulated as optimization problems. A simple problem of this type is given in the following example.

Example 1.4 (optimal control).

On the fixed time interval $[0, 1]$ we investigate the linear system of differential equations

$$\begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t) , \quad t \in (0, 1)$$

with the initial condition

$$\begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} = \begin{pmatrix} -2\sqrt{2} \\ 5\sqrt{2} \end{pmatrix} .$$

With the aid of an appropriate control function $u \in C[0, 1]$ this dynamical system should be steered from the given initial state to a terminal state in the set

$$M := \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\}.$$

In addition to this constraint a control function u minimizing the cost functional

$$f(u) = \int_0^1 (u(t))^2 dt$$

has to be determined.

Finally we discuss a simple problem from approximation theory.

Example 1.5 (approximation theory).

We consider the problem of the determination of a linear function which approximates the hyperbolic sine function on the interval [0, 2] with respect to the maximum norm in a best way (see Fig. 1.5). So,

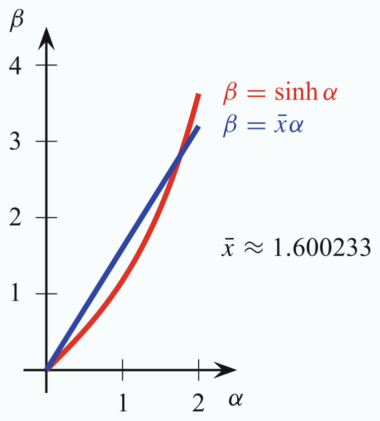

Fig. 1.5 Best approximation of sinh on [0, 2]

we minimize

$$\max_{\alpha\in[0,2]} |\alpha x - \sinh\,\alpha|.$$

This optimization problem can also be written as

$$\begin{gathered}\min\ \lambda\\ \text{subject to the constraints}\\ \lambda = \max_{\alpha\in[0,2]} |\alpha x - \sinh\,\alpha|\\ (x,\lambda)\in\mathbb{R}^2.\end{gathered}$$

The preceding problem is equivalent to the following optimization problem which has infinitely many constraints:

$$\begin{gathered}\min\ \lambda\\ \text{subject to the constraints}\\ \left.\begin{aligned}\alpha x - \sinh\,\alpha &\le \lambda\\ \alpha x - \sinh\,\alpha &\ge -\lambda\end{aligned}\right\}\ \text{for all } \alpha\in[0,2]\\ (x,\lambda)\in\mathbb{R}^2.\end{gathered}$$

Since this problem has finitely many variables and infinitely many constraints, it belongs to the class of so-called *semi-infinite* optimization problems.

In the following chapters most of the examples presented above will be investigated again. The solvability of the design problem (in Example 1.1) is discussed in Example 5.10 where the Karush-Kuhn-Tucker conditions are used as necessary optimality conditions. Discrete-continuous optimization problems (compare Example 1.2) are treated in Chap. 8. Theorem 3.21 presents a necessary optimality condition known as Euler-Lagrange equation for a minimal solution of the problem in Example 1.3. The Pontryagin maximum principle is the essential tool for the solution of the optimal control problem formulated in Example 1.4; an optimal control is determined in the Examples 5.21 and 5.23. An application of the alternation theorem leads to a solution of the linear Chebyshev approximation problem (given in Example 1.5) which is obtained in Example 6.19.

We complete this introduction with a short compendium of the structure of this textbook. Of course, the question of the solvability of a concrete nonlinear optimization problem is of primary interest and, therefore, existence theorems are presented in Chap. 2. Subsequently the question about characterizations of minimal points runs like a red thread through this book. For the formulation of such characterizations one has to approximate the objective functional (for that reason we discuss various concepts of a derivative in Chap. 3) and the constraint set (this is done with tangent cones in Chap. 4). Both approximations combined result in the optimality conditions of Chap. 5. The duality theory in Chap. 6 is closely related to optimality conditions as well; minimal points are characterized by another optimization problem being dual to the original problem. An application of optimality conditions and duality theory to semidefinite optimization being a topical field of research in optimization, is described in Chap. 7. The theory of this book is extended to discrete-continuous optimization problems in Chap. 8. The required separation theorems for this extension are also presented. The results in the last chapter show that solutions or characterizations of solutions of special optimization problems with a rich mathematical structure can be derived sometimes in a direct way.

It is interesting to note that the Hahn-Banach theorem (often in the version of a separation theorem like the Eidelheit separation theorem) proves itself to be the key for central characterization theorems.

2 Existence Theorems for Minimal Points

In this chapter we investigate a general optimization problem in a real normed space. For such a problem we present assumptions under which at least one minimal point exists. Moreover, we formulate simple statements on the set of minimal points. Finally the existence theorems obtained are applied to approximation and optimal control problems.

2.1 Problem Formulation

The standard assumption of this chapter reads as follows:

$$\left.\begin{array}{l}\text{Let } (X, \|\cdot\|) \text{ be a real normed space;}\\ \text{let } S \text{ be a nonempty subset of } X;\\ \text{and let } f : S \to \mathbb{R} \text{ be a given functional.}\end{array}\right\} \tag{2.1}$$

Under this assumption we investigate the optimization problem

$$\min_{x \in S} f(x), \tag{2.2}$$

i.e., we are looking for minimal points of f on S.

In general one does not know if the problem (2.2) makes sense because f does not need to have a minimal point on S. For instance, for $X = S = \mathbb{R}$ and $f(x) = e^x$ the optimization problem (2.2) is not solvable. In the next section we present conditions concerning f and S which ensure the solvability of the problem (2.2).

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3_2

2.2 Existence Theorems

A known existence theorem is the Weierstraß theorem which says that every continuous function attains its minimum on a compact set. This statement is modified in such a way that useful existence theorems can be obtained for the general optimization problem (2.2).

Definition 2.1 (weakly lower semicontinuous functional).

Let the assumption (2.1) be satisfied. The functional f is called *weakly lower semicontinuous* if for every sequence $(x_n)_{n\in\mathbb{N}}$ in S converging weakly to some $\bar{x} \in S$ we have:

$$\liminf_{n\to\infty} f(x_n) \geq f(\bar{x})$$

(see Appendix A for the definition of the weak convergence).

Example 2.2 (weakly lower semicontinuous functional).

The functional $f : \mathbb{R} \to \mathbb{R}$ with

$$f(x) = \left\{ \begin{matrix} 0 \text{ if } x = 0 \\ 1 \text{ otherwise} \end{matrix} \right\}$$

(see Fig. 2.1) is weakly lower semicontinuous (but not continuous at 0).

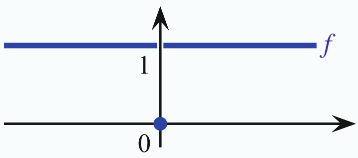

Fig. 2.1 Illustration of the functional f

Now we present the announced modification of the Weierstraß theorem.

Theorem 2.3 (solvability of problem (2.2)).

Let the assumption (2.1) be satisfied. If the set S is weakly sequentially compact and the functional f is weakly lower semicontinuous, then there is at least one $\bar{x} \in S$ with

$$f(\bar{x}) \leq f(x) \text{ for all } x \in S,$$

i.e., the optimization problem (2.2) has at least one solution.

Proof Let $(x_n)_{n\in\mathbb{N}}$ be a so-called infimal sequence in S, i.e., a sequence with

$$\lim_{n\to\infty} f(x_n) = \inf_{x\in S} f(x).$$

Since the set S is weakly sequentially compact, there is a subsequence $(x_{n_i})_{i\in\mathbb{N}}$ converging weakly to some $\bar{x} \in S$. Because of the weak lower semicontinuity of f it follows

$$f(\bar{x}) \leq \liminf_{i\to\infty} f(x_{n_i}) = \inf_{x\in S} f(x),$$

and the theorem is proved. □

Now we proceed to specialize the statement of Theorem 2.3 in order to get a version which is useful for applications. Using the concept of the epigraph we characterize weakly lower semicontinuous functionals.

Definition 2.4 (epigraph of a functional).

Let the assumption (2.1) be satisfied. The set

$$E(f) := \{(x, \alpha) \in S \times \mathbb{R} \mid f(x) \leq \alpha\}$$

is called *epigraph* of the functional f (see Fig. 2.2).

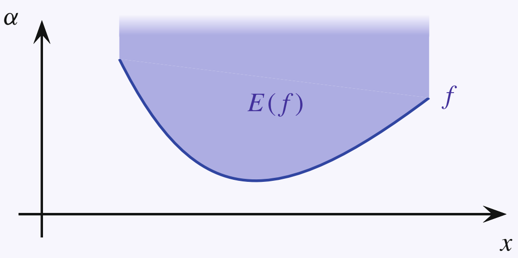

Fig. 2.2 Epigraph of a functional

Theorem 2.5 (characterizations of weakly lower semicontinuity).

Let the assumption (2.1) be satisfied, and let the set S be weakly sequentially closed. Then it follows:

$$\begin{aligned} & f \text{ is weakly lower semicontinuous} \\ \Longleftrightarrow\ & E(f) \text{ is weakly sequentially closed} \\ \Longleftrightarrow\ & \text{If for any } \alpha \in \mathbb{R} \text{ the set } S_\alpha := \{x \in S \mid f(x) \leq \alpha\} \text{ is} \\ & \text{nonempty, then } S_\alpha \text{ is weakly sequentially closed.} \end{aligned}$$

Proof

(a) Let f be weakly lower semicontinuous. If $(x_n, \alpha_n)_{n\in\mathbb{N}}$ is any sequence in $E(f)$ with a weak limit $(\bar{x}, \bar{\alpha}) \in X \times \mathbb{R}$, then $(x_n)_{n\in\mathbb{N}}$ converges weakly to $\bar{x}$ and $(\alpha_n)_{n\in\mathbb{N}}$ converges to $\bar{\alpha}$. Since S is weakly sequentially closed, we obtain $\bar{x} \in S$. Next we choose an arbitrary $\varepsilon > 0$. Then there is a number $n_0 \in \mathbb{N}$ with

$$f(x_n) \le \alpha_n < \bar{\alpha} + \varepsilon \text{ for all natural numbers } n \ge n_0.$$

Since f is weakly lower semicontinuous, it follows

$$f(\bar{x}) \le \liminf_{n\to\infty} f(x_n) < \bar{\alpha} + \varepsilon.$$

This inequality holds for an arbitrary $\varepsilon > 0$, and therefore we get $(\bar{x}, \bar{\alpha}) \in E(f)$. Consequently the set $E(f)$ is weakly sequentially closed.

(b) Now we assume that $E(f)$ is weakly sequentially closed, and we fix an arbitrary $\alpha \in \mathbb{R}$ for which the level set S_α is nonempty. Since the set $S \times \{\alpha\}$ is weakly sequentially closed, the set

$$S_\alpha \times \{\alpha\} = E(f) \cap (S \times \{\alpha\})$$

is also weakly sequentially closed. But then the set S_α is weakly sequentially closed as well.

(c) Finally we assume that the functional f is not weakly lower semicontinuous. Then there is a sequence $(x_n)_{n\in\mathbb{N}}$ in S converging weakly to some $\bar{x} \in S$ and for which

$$\liminf_{n\to\infty} f(x_n) < f(\bar{x}).$$

If one chooses any $\alpha \in \mathbb{R}$ with

$$\liminf_{n\to\infty} f(x_n) < \alpha < f(\bar{x}),$$

then there is a subsequence $(x_{n_i})_{i\in\mathbb{N}}$ converging weakly to $\bar{x} \in S$ and for which

$$x_{n_i} \in S_\alpha \text{ for all } i \in \mathbb{N}.$$

Because of $f(\bar{x}) > \alpha$ the set S_α is not weakly sequentially closed. □

Since not every continuous functional is weakly lower semicontinuous, we turn our attention to a class of functionals for which every continuous functional with a closed domain is weakly lower semicontinuous.

Definition 2.6 (convex set and convex functional).

Let S be a subset of a real linear space.

(a) The set S is called *convex* if for all $x, y \in S$

$$\lambda x + (1 - \lambda)y \in S \text{ for all } \lambda \in [0, 1]$$

(see Figs. 2.3 and 2.4).

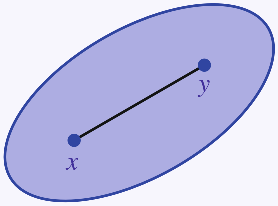

Fig. 2.3 Convex set

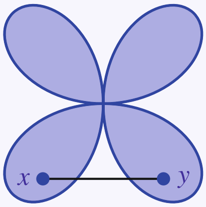

Fig. 2.4 Non-convex set

(b) Let the set S be nonempty and convex. A functional $f : S \to \mathbb{R}$ is called *convex* if for all $x, y \in S$

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

(see Figs. 2.5 and 2.6).

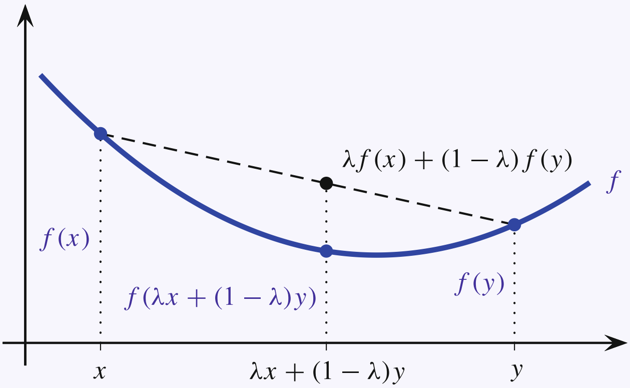

Fig. 2.5 Convex functional

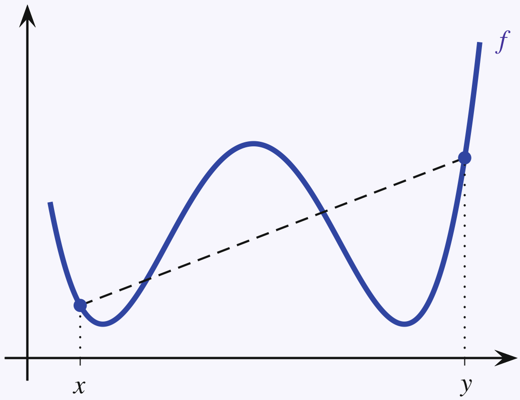

Fig. 2.6 Non-convex functional

(c) Let the set S be nonempty and convex. A functional $f : S \to \mathbb{R}$ is called *concave* if the functional $-f$ is convex (see Fig. 2.7).

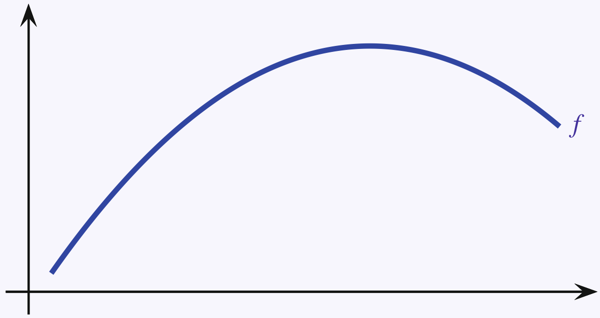

Fig. 2.7 Concave functional

Example 2.7 (convex sets and convex functionals).

(a) The empty set is always convex.
(b) The unit ball of a real normed space is a convex set.
(c) For $X = S = \mathbb{R}$ the function f with $f(x) = x^2$ for all $x \in \mathbb{R}$ is convex.
(d) Every norm on a real linear space is a convex functional.

The convexity of a functional can also be characterized with the aid of the epigraph.

Theorem 2.8 (characterization of a convex functional).

Let the assumption (2.1) be satisfied, and let the set S be convex. Then it follows:

$$\begin{aligned} & f \text{ is convex} \\ \Longleftrightarrow\ & E(f) \text{ is convex} \\ \Longrightarrow\ & \text{For every } \alpha \in \mathbb{R} \text{ the set } S_\alpha := \{x \in S \mid f(x) \le \alpha\} \text{ is convex.} \end{aligned}$$

Proof

(a) If f is convex, then it follows for arbitrary $(x, \alpha), (y, \beta) \in E(f)$ and an arbitrary $\lambda \in [0, 1]$

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &\le \lambda f(x) + (1-\lambda) f(y) \\ &\le \lambda\alpha + (1-\lambda)\beta \end{aligned}$$

resulting in

$$\lambda(x, \alpha) + (1 - \lambda)(y, \beta) \in E(f).$$

Consequently the epigraph of f is convex.

(b) Next we assume that $E(f)$ is convex and we choose any $\alpha \in \mathbb{R}$ for which the set S_α is nonempty (the case $S_\alpha = \emptyset$ is trivial). For arbitrary $x, y \in S_\alpha$ we have $(x, \alpha) \in E(f)$ and $(y, \alpha) \in E(f)$, and then we get for an arbitrary $\lambda \in [0, 1]$

$$\lambda(x, \alpha) + (1 - \lambda)(y, \alpha) \in E(f).$$

This means especially

$$f(\lambda x + (1 - \lambda)y) \leq \lambda\alpha + (1 - \lambda)\alpha = \alpha$$

and

$$\lambda x + (1 - \lambda)y \in S_\alpha.$$

Hence the set S_α is convex.

(c) Finally we assume that the epigraph $E(f)$ is convex and we show the convexity of f. For arbitrary $x, y \in S$ and an arbitrary $\lambda \in [0, 1]$ it follows

$$\lambda(x, f(x)) + (1 - \lambda)(y, f(y)) \in E(f)$$

which implies

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

Consequently the functional f is convex. □

In general the convexity of the level sets S_α does not imply the convexity of the functional f: this fact motivates the definition of the concept of quasiconvexity.

Definition 2.9 (quasiconvex functional).

Let the assumption (2.1) be satisfied, and let the set S be convex. If for every $\alpha \in \mathbb{R}$ the set $S_\alpha := \{x \in S \mid f(x) \leq \alpha\}$ is convex, then the functional f is called *quasiconvex*.

Example 2.10 (quasiconvex functionals).

(a) Every convex functional is also quasiconvex (see Theorem 2.8).
(b) For $X = S = \mathbb{R}$ the function f with $f(x) = x^3$ for all $x \in \mathbb{R}$ is quasiconvex but it is not convex. The quasiconvexity results from the convexity of the set

$$\{x \in S \mid f(x) \leq \alpha\} = \{x \in \mathbb{R} \mid x^3 \leq \alpha\} = \left(-\infty, \operatorname{sgn}(\alpha)\sqrt[3]{|\alpha|}\right]$$

for every $\alpha \in \mathbb{R}$ (compare Fig. 2.8).

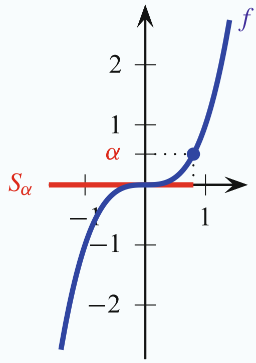

Fig. 2.8 Illustration of the function f

Now we are able to give assumptions under which every continuous functional is also weakly lower semicontinuous.

Lemma 2.11 (weakly lower semicontinuity).

Let the assumption (2.1) be satisfied, and let the set S be convex and closed. If the functional f is continuous and quasiconvex, then f is weakly lower semicontinuous.

Proof We choose an arbitrary $\alpha \in \mathbb{R}$ for which the set $S_\alpha := \{x \in S \mid f(x) \leq \alpha\}$ is nonempty. Since f is continuous and S is closed, the set S_α is also closed. Because of the quasiconvexity of f the set S_α is convex and therefore it is also weakly sequentially closed (see Appendix A). Then it follows from Theorem 2.5 that f is weakly lower semicontinuous. □

Using this lemma we obtain the following existence theorem which is useful for applications.

Theorem 2.12 (solvability in reflexive Banach spaces).

Let S be a nonempty, convex, closed and bounded subset of a reflexive real Banach space, and let $f : S \to \mathbb{R}$ be a continuous quasiconvex functional. Then f has at least one minimal point on S.

Proof With Theorem B.4 the set S is weakly sequentially compact and with Lemma 2.11 f is weakly lower semicontinuous. Then the assertion follows from Theorem 2.3. □

At the end of this section we investigate the question under which conditions a convex functional is also continuous. With the following lemma which may be helpful in connection with the previous theorem we show that every convex function which is defined on an open convex set and continuous at some point is also continuous on the whole set.

Lemma 2.13 (continuity of a convex functional).

Let the assumption (2.1) be satisfied, and let the set S be open and convex. If the functional f is convex and continuous at some $\bar{x} \in S$, then f is continuous on S.

Proof We show that f is continuous at any point of S. For that purpose we choose an arbitrary $\tilde{x} \in S$. Since f is continuous at $\bar{x}$ and S is open, there is a closed ball $B(\bar{x}, \varrho)$ around $\bar{x}$ with the radius ϱ so that f is bounded from above on $B(\bar{x}, \varrho)$ by some $\alpha \in \mathbb{R}$. Because S is convex and open there is a $\lambda > 1$ so that $\bar{x} + \lambda(\tilde{x} - \bar{x}) \in S$ and the closed ball $B(\tilde{x}, (1 - \frac{1}{\lambda})\varrho)$ around $\tilde{x}$ with the radius $(1 - \frac{1}{\lambda})\varrho$ is contained in S. Then for every $x \in B(\tilde{x}, (1 - \frac{1}{\lambda})\varrho)$ there is some $y \in B(0_X, \varrho)$ (closed ball around 0_X with the radius ϱ) so that because of the convexity of f

$$\begin{aligned}
f(x) &= f(\tilde{x} + (1 - \frac{1}{\lambda})y) \\
&= f(\tilde{x} - (1 - \frac{1}{\lambda})\bar{x} + (1 - \frac{1}{\lambda})(\bar{x} + y)) \\
&= f(\frac{1}{\lambda}(\bar{x} + \lambda(\tilde{x} - \bar{x})) + (1 - \frac{1}{\lambda})(\bar{x} + y)) \\
&\leq \frac{1}{\lambda} f(\bar{x} + \lambda(\tilde{x} - \bar{x})) + (1 - \frac{1}{\lambda}) f(\bar{x} + y) \\
&\leq \frac{1}{\lambda} f(\bar{x} + \lambda(\tilde{x} - \bar{x})) + (1 - \frac{1}{\lambda})\alpha \\
&=: \beta.
\end{aligned}$$

This means that f is bounded from above on $B(\tilde{x}, (1-\frac{1}{\lambda})\varrho)$ by β. For the proof of the continuity of f at $\tilde{x}$ we take any $\varepsilon \in (0, 1)$. Then we choose an arbitrary element x of the closed ball $B(\tilde{x}, \varepsilon(1-\frac{1}{\lambda})\varrho)$. Because of the convexity of f we get for some $y \in B(0_X, (1-\frac{1}{\lambda})\varrho)$

$$\begin{aligned} f(x) &= f(\tilde{x} + \varepsilon y) \\ &= f((1-\varepsilon)\tilde{x} + \varepsilon(\tilde{x} + y)) \\ &\leq (1-\varepsilon) f(\tilde{x}) + \varepsilon f(\tilde{x} + y) \\ &\leq (1-\varepsilon) f(\tilde{x}) + \varepsilon\beta \end{aligned}$$

which implies

$$f(x) - f(\tilde{x}) \leq \varepsilon(\beta - f(\tilde{x})). \tag{2.3}$$

Moreover we obtain

$$\begin{aligned} f(\tilde{x}) &= f\Big(\frac{1}{1+\varepsilon}(\tilde{x} + \varepsilon y) + (1 - \frac{1}{1+\varepsilon})(\tilde{x} - y)\Big) \\ &\leq \frac{1}{1+\varepsilon} f(\tilde{x} + \varepsilon y) + (1 - \frac{1}{1+\varepsilon}) f(\tilde{x} - y) \\ &\leq \frac{1}{1+\varepsilon} f(x) + (1 - \frac{1}{1+\varepsilon})\beta \\ &= \frac{1}{1+\varepsilon}(f(x) + \varepsilon\beta) \end{aligned}$$

which leads to

$$(1+\varepsilon) f(\tilde{x}) \leq f(x) + \varepsilon\beta$$

and

$$-(f(x) - f(\tilde{x})) \leq \varepsilon(\beta - f(\tilde{x})). \tag{2.4}$$

The inequalities (2.3) and (2.4) imply

$$|f(x) - f(\tilde{x})| \leq \varepsilon(\beta - f(\tilde{x})) \text{ for all } x \in B(\tilde{x}, \varepsilon(1 - \frac{1}{\lambda})\varrho).$$

So, f is continuous at $\tilde{x}$, and the proof is complete. □

Under the assumptions of the preceding lemma it is shown in [71, Prop. 2.2.6] that f is even Lipschitz continuous at every $x \in S$ (see Definition 3.33).

2.3 Set of Minimal Points

After answering the question about the existence of a minimal solution of an optimization problem, in this section the set of all minimal points is investigated.

Theorem 2.14 (convexity of the set of minimal points).

Let S be a nonempty convex subset of a real linear space. For every quasiconvex functional $f : S \to \mathbb{R}$ the set of minimal points of f on S is convex.

Proof If f has no minimal point on S, then the assertion is evident. Therefore we assume that f has at least one minimal point $\bar{x}$ on S. Since f is quasiconvex, the set

$$\bar{S} := \{x \in S \mid f(x) \leq f(\bar{x})\}$$

is also convex. But this set equals the set of minimal points of f on S. □

With the following definition we introduce the concept of a local minimal point.

Definition 2.15 (local minimal point).

Let the assumption (2.1) be satisfied. An element $\bar{x} \in S$ is called a *local minimal point* of f on S if there is a ball $B(\bar{x}, \varepsilon) := \{x \in X \mid \|x - \bar{x}\| \leq \varepsilon\}$ around $\bar{x}$ with the radius $\varepsilon > 0$ so that

$$f(\bar{x}) \leq f(x) \text{ for all } x \in S \cap B(\bar{x}, \varepsilon).$$

The following theorem says that local minimal solutions of a convex optimization problem are also (global) minimal solutions.

Theorem 2.16 (local minimal points).

Let S be a nonempty convex subset of a real normed space. Every local minimal point of a convex functional $f : S \to \mathbb{R}$ is also a minimal point of f on S.

Proof Let $\bar{x} \in S$ be a local minimal point of a convex functional $f : S \to \mathbb{R}$. Then there are an $\varepsilon > 0$ and a ball $B(\bar{x}, \varepsilon)$ so that $\bar{x}$ is a minimal point of f on $S \cap B(\bar{x}, \varepsilon)$. Now we consider an arbitrary $x \in S$ with $x \notin B(\bar{x}, \varepsilon)$. Then it is $\|x - \bar{x}\| > \varepsilon$. For $\lambda := \frac{\varepsilon}{\|x-\bar{x}\|} \in (0, 1)$ we obtain $x_\lambda := \lambda x + (1 - \lambda)\bar{x} \in S$ and

$$\|x_\lambda - \bar{x}\| = \|\lambda x + (1 - \lambda)\bar{x} - \bar{x}\| = \lambda\|x - \bar{x}\| = \varepsilon,$$

i.e., it is $x_\lambda \in S \cap B(\bar{x}, \varepsilon)$ (see Fig. 2.9).

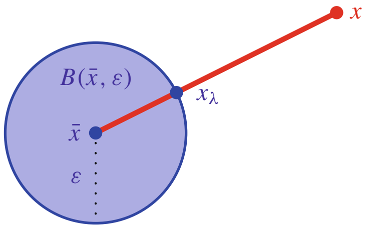

Fig. 2.9 Construction in the proof of Theorem 2.16

Therefore we get

$$f(\bar{x}) \leq f(x_\lambda) = f(\lambda x + (1-\lambda)\bar{x}) \leq \lambda f(x) + (1-\lambda) f(\bar{x})$$

resulting in

$$f(\bar{x}) \leq f(x).$$

Consequently $\bar{x}$ is a minimal point of f on S. □

It is also possible to formulate conditions ensuring that a minimal point is unique. This can be done under stronger convexity requirements, e.g., like "strict convexity" of the objective functional.

2.4 Application to Approximation Problems

Approximation problems can be formulated as special optimization problems. Therefore, existence theorems in approximation theory can be obtained with the aid of the results of Sect. 2.2. Such existence results are deduced for general approximation problems and especially also for a problem of Chebyshev approximation.

First we investigate a general problem of approximation theory. Let S be a nonempty subset of a real normed space $(X, \|\cdot\|)$, and let $\hat{x} \in X$ be a given element. Then we are looking for some $\bar{x} \in S$ for which the distance between $\hat{x}$ and S is minimal, i.e.,

$$\|\bar{x} - \hat{x}\| \leq \|x - \hat{x}\| \text{ for all } x \in S.$$

In this case $\bar{x}$ is a minimal solution of the optimization problem

$$\min_{x \in S} \|x - \hat{x}\|.$$

Definition 2.17 (best approximation).

Let S be a nonempty subset of a real normed space $(X, \|\cdot\|)$. The set S is called *proximinal* if for every $\hat{x} \in X$ there is a vector $\bar{x} \in S$ with the property

$$\|\bar{x} - \hat{x}\| \leq \|x - \hat{x}\| \text{ for all } x \in S.$$

In this case $\bar{x}$ is called *best approximation* to $\hat{x}$ from S (see Fig. 2.10).

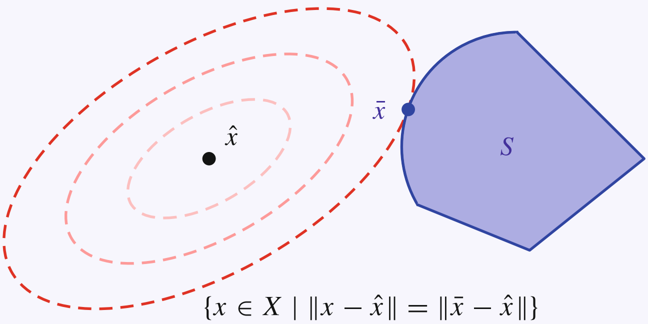

Fig. 2.10 Best approximation

So for a proximinal set the considered approximation problem is solvable for every arbitrary $\hat{x} \in X$. The following theorem gives a sufficient condition for the solvability of the general approximation problem.

Theorem 2.18 (proximal set).

Every nonempty convex closed subset of a reflexive real Banach space is proximinal.

Proof Let S be a nonempty convex closed subset of a reflexive Banach space $(X, \|\cdot\|)$, and let $\hat{x} \in X$ be an arbitrary element. Then we investigate the solvability of the optimization problem $\min\limits_{x \in S} \|x - \hat{x}\|$. For that purpose we define the objective functional $f : X \to \mathbb{R}$ with

$$f(x) = \|x - \hat{x}\| \text{ for all } x \in X.$$

The functional f is continuous because for arbitrary $x, y \in X$ we have

$$\begin{aligned}|f(x) - f(y)| &= \big|\, \|x - \hat{x}\| - \|y - \hat{x}\| \,\big| \\ &\leq \|x - \hat{x} - (y - \hat{x})\| \\ &= \|x - y\|.\end{aligned}$$

Next we show the convexity of the functional f. For arbitrary $x, y \in X$ and $\lambda \in [0, 1]$ we get

$$\begin{aligned}f(\lambda x + (1-\lambda)y) &= \|\lambda x + (1-\lambda)y - \hat{x}\| \\ &= \|\lambda(x - \hat{x}) + (1-\lambda)(y - \hat{x})\| \\ &\leq \lambda\|x - \hat{x}\| + (1-\lambda)\|y - \hat{x}\| \\ &= \lambda f(x) + (1-\lambda) f(y).\end{aligned}$$

Consequently f is continuous and quasiconvex. If we fix any $\tilde{x} \in S$ and we define

$$\tilde{S} := \{x \in S \mid f(x) \leq f(\tilde{x})\},$$

then $\tilde{S}$ is a convex subset of X. For every $x \in \tilde{S}$ we have

$$\|x\| = \|x - \hat{x} + \hat{x}\| \leq \underbrace{\|x - \hat{x}\|}_{=f(x)} + \|\hat{x}\| \leq f(\tilde{x}) + \|\hat{x}\|,$$

and therefore the set $\tilde{S}$ is bounded. Since the set S is closed and the functional f is continuous, the set $\tilde{S}$ is also closed. Then by the existence Theorem 2.12 f has at least one minimal point on $\tilde{S}$, i.e., there is a vector $\bar{x} \in \tilde{S}$ with

$$f(\bar{x}) \leq f(x) \text{ for all } x \in \tilde{S}.$$

The inclusion $\tilde{S} \subset S$ implies $\bar{x} \in S$ and for all $x \in S \setminus \tilde{S}$ we get

$$f(x) > f(\tilde{x}) \geq f(\bar{x}).$$

Consequently $\bar{x} \in S$ is a minimal point of f on S. □

The following theorem shows that, in general, the reflexivity of the Banach space plays an important role for the solvability of approximation problems. But notice also that under strong assumptions concerning the set S an approximation problem may be solvable in non-reflexive spaces.

Theorem 2.19 (characterization of reflexivity).

A real Banach space is reflexive if and only if every nonempty convex closed subset is proximinal.

Proof One direction of the assertion is already proved in the existence Theorem 2.18. Therefore we assume now that the considered real Banach space is not reflexive. Then the closed unit ball $B(0_X, 1) := \{x \in X \mid \|x\| \leq 1\}$ is not weakly sequentially compact and by a James theorem (Theorem B.2) there is a continuous linear functional l which does not attain its supremum on the set $B(0_X, 1)$, i.e.,

$$l(x) < \sup_{y \in B(0_X,1)} l(y) \quad \text{for all } x \in B(0_X, 1).$$

If one defines the convex closed set

$$S := \{x \in X \mid l(x) \geq \sup_{y \in B(0_X,1)} l(y)\},$$

then one obtains $S \cap B(0_X, 1) = \emptyset$. Hence, the set S is not proximinal. □

Now we turn our attention to a special problem, namely to a problem of uniform approximation of functions (problem of Chebyshev approximation). Let M be a compact metric space and let $C(M)$ be the real linear space of continuous real-valued functions on M equipped with the maximum norm $\| \cdot \|$ where

$$\|x\| = \max_{t \in M} |x(t)| \text{ for all } x \in C(M).$$

Moreover let S be a nonempty subset of $C(M)$, and let $\hat{x} \in C(M)$ be a given function. We are looking for a function $\bar{x} \in S$ with

$$\|\bar{x} - \hat{x}\| \leq \|x - \hat{x}\| \text{ for all } x \in S$$

(see Fig. 2.11).

Since $X = C(M)$ is not reflexive, Theorem 2.18 may not be applied directly to this special approximation problem. But the following result is true.

Theorem 2.20 (proximal set).

If S is a nonempty convex closed subset of the normed space $C(M)$ such that for any $\tilde{x} \in S$ the linear subspace spanned by $S - \{\tilde{x}\}$ is reflexive, then the set S is proximinal.

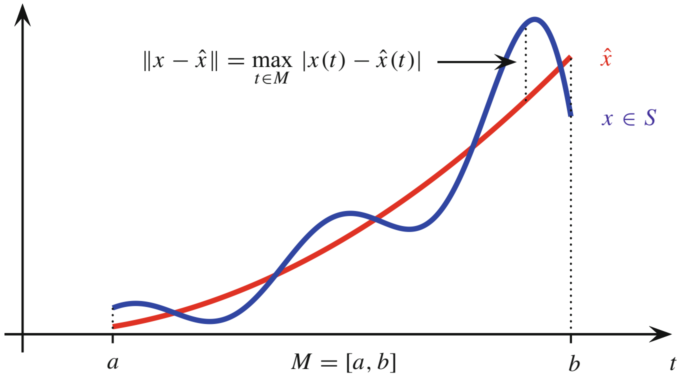

Fig. 2.11 Chebyshev approximation

Proof For $\tilde{x} \in S$ we have

$$\inf_{x \in S} \|x - \hat{x}\| = \inf_{x \in S} \|(x - \tilde{x}) - (\hat{x} - \tilde{x})\| = \inf_{x \in S - \{\tilde{x}\}} \|x - (\hat{x} - \tilde{x})\|.$$

If V denotes the linear subspace spanned by $\hat{x} - \tilde{x}$ and $S - \{\tilde{x}\}$, then V is reflexive and Theorem 2.18 can be applied to the reflexive real Banach space V. Consequently the set S is proximinal. □

In general, the linear subspace spanned by $S - \{\tilde{x}\}$ is finite dimensional and therefore reflexive, because S is very often a set of linear combinations of finitely many functions of $C(M)$ (for instance, monoms, i.e. functions of the form $x(t) = 1, t, t^2, \ldots, t^n$ with a fixed $n \in \mathbb{N}$). In this case a problem of Chebyshev approximation has at least one solution.

2.5 Application to Optimal Control Problems

In this section we apply the existence result of Theorem 2.12 to problems of optimal control. First we present a problem which does not have a minimal solution.

Example 2.21 (control problem with no solution).

We consider a dynamical system with the differential equation

$$\dot{x}(t) = -u(t)^2 \text{ almost everywhere on } [0,1], \tag{2.5}$$

the initial condition

$$x(0) = 1 \tag{2.6}$$

and the terminal condition

$$x(1) = 0. \tag{2.7}$$

Let the control u be an L_2-function, i.e. $u \in L_2[0, 1]$. A solution of the differential equation (2.5) is defined as

$$x(t) = c - \int_0^t u(s)^2 ds \text{ for all } t \in [0, 1]$$

with $c \in \mathbb{R}$. In view of the initial condition we get

$$x(t) = 1 - \int_0^t u(s)^2 ds \text{ for all } t \in [0, 1].$$

Then the terminal condition (2.7) is equivalent to

$$1 - \int_0^1 u(s)^2 ds = 0.$$

Question: Is there an optimal control minimizing $\int_0^1 t^2 u(t)^2 dt$?

For $X = L_2[0, 1]$ we define the constraint set

$$S := \left\{ u \in L_2[0, 1] \;\middle|\; \int_0^1 u(s)^2 ds = 1 \right\}$$

(S is exactly the unit sphere in $L_2[0, 1]$). The objective functional $f : S \to \mathbb{R}$ is given by

$$f(u) = \int_0^1 t^2 u(t)^2 dt \text{ for all } u \in S.$$

One can see immediately that

$$0 \le \inf_{u \in S} f(u).$$

Next we define a sequence of feasible controls $(u_n)_{n\in\mathbb{N}}$ by

$$u_n(t) = \begin{Bmatrix} n \text{ almost everywhere on } [0, \frac{1}{n^2}) \\ 0 \text{ almost everywhere on } [\frac{1}{n^2}, 1] \end{Bmatrix}.$$

Then we get for every $n \in \mathbb{N}$

$$\|u_n\|^2_{L_2[0,1]} = \int_0^1 |u_n(t)|^2 dt = \int_0^{\frac{1}{n^2}} n^2 dt = 1.$$

Hence we have

$$u_n \in S \text{ for all } n \in \mathbb{N}$$

(every u_n is an element of the unit sphere in $L_2[0, 1]$). Moreover we conclude for all $n \in \mathbb{N}$

$$f(u_n) = \int_0^1 t^2 u_n(t)^2 dt = \int_0^{\frac{1}{n^2}} t^2 n^2 dt = \frac{n^2}{3} t^3 \Big|_0^{\frac{1}{n^2}} = \frac{1}{3n^4}$$

and therefore we get

$$\lim_{n\to\infty} f(u_n) = 0 = \inf_{u \in S} f(u).$$

If we assume that f attains its infimal value 0 on S, then there is a control $\bar{u} \in S$ with $f(\bar{u}) = 0$, i.e.

$$\int_0^1 \underbrace{t^2 \bar{u}(t)^2}_{\ge 0} dt = 0.$$

But then we get

$$\bar{u}(t) = 0 \text{ almost everywhere on } [0,1]$$

and especially $\bar{u} \notin S$. Consequently f does *not* attain its infimum on S.

In the following we consider a special optimal control problem with a system of linear differential equations.

Problem 2.22 (optimal control problem).

Let A and B be given (n, n) and (n, m) matrices with real coefficients, respectively, and let the system of differential equations be given as

$$\dot{x}(t) = Ax(t) + Bu(t) \text{ almost everywhere on } [t_0, t_1] \tag{2.8}$$

with the initial condition

$$x(t_0) = x_0 \in \mathbb{R}^n \tag{2.9}$$

where $-\infty < t_0 < t_1 < \infty$. Let the control u be a $L_2^m[t_0, t_1]$ function. A solution x of the system (2.8) of differential equations with the initial condition (2.9) is defined as

$$x(t) = x_0 + \int_{t_0}^{t} e^{A(t-s)} Bu(s)\, ds \text{ for all } t \in [t_0, t_1].$$

The exponential function occurring in the above expression is the matrix exponential function, and the integral has to be understood in a component-wise sense. Let the constraint set $S \subset L_2^m[t_0, t_1]$ be given as

$$S := \{u \in L_2^m[t_0, t_1] \mid \|u(t)\| \leq 1 \text{ almost everywhere on } [t_0, t_1]\}$$

($\| \cdot \|$ denotes the l_2 norm on $\mathbb{R}^m$). The objective functional $f : S \to \mathbb{R}$ is defined by

$$f(u) = \int_{t_0}^{t_1} (g(x(t)) + h(u(t)))\, dt$$
$$= \int_{t_0}^{t_1} \Big(g\Big(x_0 + \int_{t_0}^{t} e^{A(t-s)} Bu(s)\, ds\Big) + h(u(t))\Big)\, dt \text{ for all } u \in S$$

where $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^m \to \mathbb{R}$ are real valued functions. Then we are looking for minimal points of f on S.

Theorem 2.23 (existence of an optimal control).

Let the Problem 2.22 be given. Let the functions g and h be convex and continuous, and let h be Lipschitz continuous on the closed unit ball. Then f has at least one minimal point on S.

Proof First notice that $X := L_2^m[t_0, t_1]$ is a reflexive Banach space. Since S is the closed unit ball in $L_2^m[t_0, t_1]$, the set S is closed, bounded and convex. Next we show the quasiconvexity of the objective functional f. For that purpose we define the linear mapping $L : S \to AC^n[t_0, t_1]$ (let $AC^n[t_0, t_1]$ denote the real linear space of absolutely continuous n vector functions equipped with the maximum norm) with

$$L(u)(t) = \int_{t_0}^{t_1} e^{A(t-s)} Bu(s)\, ds \text{ for all } u \in S \text{ and all } t \in [t_0, t_1].$$

If we choose arbitrary $u_1, u_2 \in S$ and $\lambda \in [0, 1]$, we get

$$g(x_0 + L(\lambda u_1 + (1 - \lambda)u_2)(t))$$
$$= g(x_0 + \lambda L(u_1)(t) + (1 - \lambda)L(u_2)(t))$$
$$= g(\lambda[x_0 + L(u_1)(t)] + (1 - \lambda)[x_0 + L(u_2)(t)])$$
$$\leq \lambda g(x_0 + L(u_1)(t)) + (1 - \lambda)g(x_0 + L(u_2)(t)) \text{ for all } t \in [t_0, t_1]$$

and

$$\begin{aligned}
& f(\lambda u_1 + (1-\lambda)u_2) \\
&= \int_{t_0}^{t_1} [g(x_0 + L(\lambda u_1 + (1-\lambda)u_2)(t)) \\
&\quad + h(\lambda u_1(t) + (1-\lambda)u_2(t))]\, dt \\
&\leq \int_{t_0}^{t_1} [\lambda g(x_0 + L(u_1)(t)) + (1-\lambda) g(x_0 + L(u_2)(t)) \\
&\quad + \lambda h(u_1(t)) + (1-\lambda) h(u_2(t))]\, dt \\
&= \lambda f(u_1) + (1-\lambda) f(u_2).
\end{aligned}$$

So, f is convex and, therefore, quasiconvex. Next we prove that the objective functional f is continuous. For all $u \in S$ we have

$$\begin{aligned}
\|L(u)\|_{AC^n[t_0,t_1]} &= \left\| \int_{t_0}^{\cdot} e^{A(\cdot - s)} B u(s)\, ds \right\|_{AC^n[t_0,t_1]} \\
&\leq c_1 \|u\|_{L_2^m[t_0,t_1]} \qquad (2.10)
\end{aligned}$$

where c_1 is a positive constant. Now we fix an arbitrary sequence $(u_n)_{n\in\mathbb{N}}$ in S converging to some $\bar{u} \in S$. Then we obtain

$$\begin{aligned}
f(u_n) - f(\bar{u}) &= \int_{t_0}^{t_1} [g(x_0 + L(u_n)(t)) + g(x_0 + L(\bar{u})(t))]\, dt \\
&\quad + \int_{t_0}^{t_1} [h(u_n(t)) - h(\bar{u}(t))]\, dt. \qquad (2.11)
\end{aligned}$$

Because of the inequality (2.10) and the continuity of g the following equation holds pointwise:

$$\lim_{n\to\infty} g(x_0 + L(u_n)(t)) = g(x_0 + L(\bar{u})(t)).$$

Since $\|u_n\|_{L_2^m[t_0,t_1]} \leq 1$ and $\|\bar{u}\|_{L_2^m[t_0,t_1]} \leq 1$, the convergence of the first integral in (2.11) to 0 follows from Lebesgue's theorem on the dominated convergence. The second integral expression in (2.11) converges to 0 as well because h is assumed to be Lipschitz continuous:

$$\int_{t_0}^{t_1} |h(u_n(t)) - h(\bar{u}(t))|dt \leq c_2 \int_{t_0}^{t_1} \|u_n(t) - \bar{u}(t)\|\, dt \leq c_2 \|u_n - \bar{u}\|_{L_2^m[t_0,t_1]}$$

(where $c_2 \in \mathbb{R}$ denotes the Lipschitz constant). Consequently f is continuous. We summarize our results: The objective functional f is quasiconvex and continuous, and the constraint set S is closed, bounded and convex. Hence the assertion follows from Theorem 2.12. □

Exercises

(2.1) Let S be a nonempty subset of a finite dimensional real normed space. Show that every continuous functional $f : S \to \mathbb{R}$ is also weakly lower semicontinuous.

(2.2) Show that the function $f : \mathbb{R} \to \mathbb{R}$ with

$$f(x) = xe^x \text{ for all } x \in \mathbb{R}$$

is quasiconvex.

(2.3) Let the assumption (2.1) be satisfied, and let the set S be convex. Prove that the functional f is quasiconvex if and only if for all $x, y \in S$

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

(2.4) Prove that every proximinal subset of a real normed space is closed.

(2.5) Show that the approximation problem from Example 1.5 is solvable.

(2.6) Let $C(M)$ denote the real linear space of continuous real valued functions on a compact metric space M equipped with the maximum norm. Prove that for every $n \in \mathbb{N}$ and every continuous function $\hat{x} \in C(M)$ there are real numbers $\bar{\alpha}_0, \ldots, \bar{\alpha}_n \in \mathbb{R}$ with the property

$$\max_{t \in M} \left| \sum_{i=0}^{n} \bar{\alpha}_i t^i - \hat{x}(t) \right| \leq \max_{t \in M} \left| \sum_{i=0}^{n} \alpha_i t^i - \hat{x}(t) \right| \text{ for all } \alpha_0, \ldots, \alpha_n \in \mathbb{R}.$$

(2.7) Which assumption of Theorem 2.12 is not satisfied for the optimization problem from Example 2.21?

(2.8) Let the optimal control problem given in Problem 2.22 be modified in such a way that we want to reach a given absolutely continuous state $\bar{x}$ as close as possible, i.e., we define the objective functional $f : S \to \mathbb{R}$ by

$$\begin{aligned} f(u) &= \max_{t \in [t_0, t_1]} |x(t) - \hat{x}(t)| \\ &= \max_{t \in [t_0, t_1]} \left| x_0 - \hat{x}(t) + \int_{t_0}^{t} e^{A(t-s)} Bu(s)\, ds \right| \text{ for all } u \in S. \end{aligned}$$

Show that f has at least one minimal point on S.

Generalized Derivatives

3

In this chapter various customary concepts of a derivative are presented and its properties are discussed. The following notions are investigated: directional derivatives, Gâteaux and Fréchet derivatives, subdifferentials, quasidifferentials and Clarke derivatives. Moreover, simple optimality conditions are given which can be deduced in connection with these generalized derivatives.

3.1 Directional Derivative

In this section we introduce the concept of a directional derivative and we present already a simple optimality condition.

Definition 3.1 (directional derivative).

Let X be a real linear space, let $(Y, \|\cdot\|)$ be a real normed space, let S be a nonempty subset of X and let $f : S \to Y$ be a given mapping.
If for two elements $\bar{x} \in S$ and $h \in X$ the limit

$$f'(\bar{x})(h) := \lim_{\lambda \to 0_+} \frac{1}{\lambda}(f(\bar{x} + \lambda h) - f(\bar{x}))$$

exists, then $f'(\bar{x})(h)$ is called the *directional derivative* of f at $\bar{x}$ in the direction h. If this limit exists for all $h \in X$, then f is called *directionally differentiable* at $\bar{x}$ (see Fig. 3.1).

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3_3

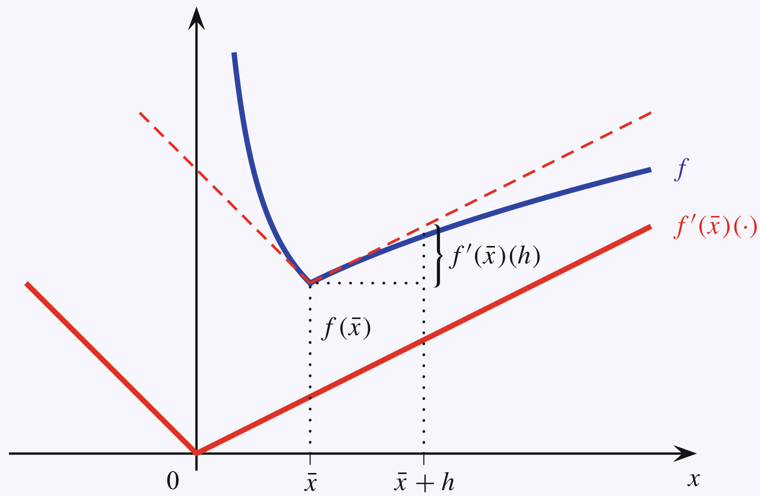

Fig. 3.1 A directionally differentiable function

Notice that for the limit defining the directional derivative one considers arbitrary sequences $(\lambda_n)_{n\in\mathbb{N}}$ converging to 0, $\lambda_n > 0$ for all $n \in \mathbb{N}$, with the additional property that $\bar{x} + \lambda_n h$ belongs to the domain S for all $n \in \mathbb{N}$. This restriction of the sequences converging to 0 can be dropped, for instance, if S equals the whole space X.

Example 3.2 (directionally differentiable function).

For the function $f : \mathbb{R}^2 \to \mathbb{R}$ with

$$f(x_1, x_2) = \begin{Bmatrix} x_1^2(1 + \frac{1}{x_2}) & \text{if } x_2 \neq 0 \\ 0 & \text{if } x_2 = 0 \end{Bmatrix} \text{ for all } (x_1, x_2) \in \mathbb{R}^2$$

which is not continuous at $0_{\mathbb{R}^2}$, we obtain the directional derivative

$$f'(0_{\mathbb{R}^2})(h_1, h_2) = \lim_{\lambda \to 0_+} \frac{1}{\lambda} f(\lambda(h_1, h_2)) = \begin{cases} \frac{h_1^2}{h_2} & \text{if } h_2 \neq 0 \\ 0 & \text{if } h_2 = 0 \end{cases}$$

in the direction $(h_1, h_2) \in \mathbb{R}^2$. Notice that $f'(0_{\mathbb{R}^2})$ is neither continuous nor linear.

As a first result on directional derivatives we show that every convex functional is directionally differentiable. For the proof we need the following lemma.

Lemma 3.3 (monotonicity of the difference quotient).

Let X be a real linear space, and let $f : X \to \mathbb{R}$ be a convex functional. Then for arbitrary $\bar{x}, h \in X$ the function $\varphi : \mathbb{R}_+ \setminus \{0\} \to \mathbb{R}$ with

$$\varphi(\lambda) = \frac{1}{\lambda}(f(\bar{x} + \lambda h) - f(\bar{x})) \text{ for all } \lambda > 0$$

is monotonically increasing (i.e., $0 < s \leq t$ implies $\varphi(s) \leq \varphi(t)$).

Proof For arbitrary $\bar{x}, h \in X$ we consider the function φ defined above. Because of the convexity of f we then get for arbitrary $0 < s \leq t$:

$$\begin{aligned} f(\bar{x} + sh) - f(\bar{x}) &= f\Big(\frac{s}{t}(\bar{x} + th) + \frac{t-s}{t}\bar{x}\Big) - f(\bar{x}) \\ &\leq \frac{s}{t} f(\bar{x} + th) + \frac{t-s}{t} f(\bar{x}) - f(\bar{x}) \\ &= \frac{s}{t}(f(\bar{x} + th) - f(\bar{x})) \end{aligned}$$

resulting in

$$\frac{1}{s}(f(\bar{x} + sh) - f(\bar{x})) \leq \frac{1}{t}(f(\bar{x} + th) - f(\bar{x})).$$

Consequently we have $\varphi(s) \leq \varphi(t)$. □

Theorem 3.4 (existence of the directional derivative).

Let X be a real linear space, and let $f : X \to \mathbb{R}$ be a convex functional. Then at every $\bar{x} \in X$ and in every direction $h \in X$ the directional derivative $f'(\bar{x})(h)$ exists.

Proof We choose arbitrary elements $\bar{x}, h \in X$ and define the function $\varphi : \mathbb{R} \to \mathbb{R}$ with

$$\varphi(\lambda) = \frac{1}{\lambda}(f(\bar{x} + \lambda h) - f(\bar{x})) \text{ for all } \lambda > 0.$$

Because of the convexity of f we get for all $\lambda > 0$

$$
\begin{aligned}
f(\bar{x}) &= f\Big(\frac{1}{1+\lambda}(\bar{x}+\lambda h) + \frac{\lambda}{1+\lambda}(\bar{x}-h)\Big) \\
&\leq \frac{1}{1+\lambda} f(\bar{x}+\lambda h) + \frac{\lambda}{1+\lambda} f(\bar{x}-h),
\end{aligned}
$$

and therefore, we have

$$(1+\lambda) f(\bar{x}) \leq f(\bar{x}+\lambda h) + \lambda f(\bar{x}-h)$$

implying

$$f(\bar{x}) - f(\bar{x}-h) \leq \frac{1}{\lambda}(f(\bar{x}+\lambda h) - f(\bar{x})) = \varphi(\lambda).$$

Hence the function φ is bounded from below. With Lemma 3.3 φ is also monotonically increasing. Consequently the limit

$$f'(\bar{x})(h) = \lim_{\lambda \to 0_+} \varphi(\lambda)$$

exists indeed. □

For the next assertion we need the concept of sublinearity.

Definition 3.5 (sublinear functional).

Let X be a real linear space. A functional $f : X \to \mathbb{R}$ is called *sublinear*, if

(a) $f(\alpha x) = \alpha f(x)$ for all $x \in X$ and all $\alpha \geq 0$ (positive homogenity) ,
(b) $f(x+y) \leq f(x) + f(y)$ for all $x, y \in X$ (subadditivity)
(compare Fig. 3.2).

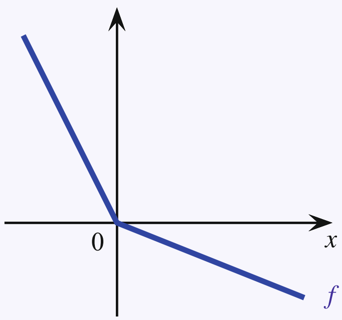

Fig. 3.2 Sublinear functional

Now we show that the directional derivative of a convex functional is sublinear with respect to the direction.

Theorem 3.6 (sublinearity of the directional derivative).

Let X be a real linear space, and let $f : X \to \mathbb{R}$ be a convex functional. Then for every $\bar{x} \in X$ the directional derivative $f'(\bar{x})(\cdot)$ is a sublinear functional.

Proof With Theorem 3.4 the directional derivative $f'(\bar{x})(\cdot)$ exists. First we notice that $f'(\bar{x})(0_X) = 0$. For arbitrary $h \in X$ and $\alpha > 0$ we obtain

$$f'(\bar{x})(\alpha h) = \lim_{\lambda \to 0_+} \frac{1}{\lambda}(f(\bar{x} + \lambda\alpha h) - f(\bar{x})) = \alpha f'(\bar{x})(h).$$

Consequently $f'(\bar{x})(\cdot)$ is positively homogeneous. For the proof of the subadditivity we fix arbitrary $h_1, h_2 \in X$. Then we obtain for an arbitrary $\lambda > 0$ because of the convexity of f

$$\begin{aligned} f(\bar{x} + \lambda(h_1 + h_2)) &= f\Big(\frac{1}{2}(\bar{x} + 2\lambda h_1) + \frac{1}{2}(\bar{x} + 2\lambda h_2)\Big) \\ &\le \frac{1}{2} f(\bar{x} + 2\lambda h_1) + \frac{1}{2} f(\bar{x} + 2\lambda h_2) \end{aligned}$$

and

$$\begin{aligned} \frac{1}{\lambda}[f(\bar{x} + \lambda(h_1 + h_2)) - f(\bar{x})] &\le \frac{1}{2\lambda}[f(\bar{x} + 2\lambda h_1) - f(\bar{x})] \\ &\quad + \frac{1}{2\lambda}[f(\bar{x} + 2\lambda h_2) - f(\bar{x})]. \end{aligned}$$

Hence we get for $\lambda \to 0_+$

$$f'(\bar{x})(h_1 + h_2) \le f'(\bar{x})(h_1) + f'(\bar{x})(h_2)$$

and the proof is complete. □

If a functional f is defined not on a whole real linear space X but on a nonempty subset S, the property that f has a directional derivative at $\bar{x}$ in any direction $x - \bar{x}$ with $x \in S$, requires necessarily

$$\bar{x} + \lambda(x - \bar{x}) = \lambda x + (1 - \lambda)\bar{x} \in S \text{ for sufficiently small } \lambda > 0.$$

This necessary condition is fulfilled, for instance, if S is starshaped with respect to $\bar{x}$ — a notion which is introduced next.

Definition 3.7 (starshaped set).

A nonempty subset S of a real linear space is called *starshaped* with respect to some $\bar{x} \in S$, if for all $x \in S$:

$$\lambda x + (1 - \lambda)\bar{x} \in S \text{ for all } \lambda \in [0, 1]$$

(see Fig. 3.3).

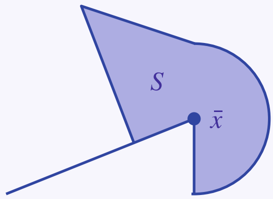

Fig. 3.3 A set S which is starshaped with respect to $\bar{x}$

Every nonempty convex subset of a real linear space is starshaped with respect to each of its elements. And conversely, every nonempty subset of a real linear space which is starshaped with respect to each of its elements is a convex set.

Using directional derivatives we obtain a simple necessary and sufficient optimality condition.

Theorem 3.8 (optimality condition).

Let S be a nonempty subset of a real linear space, and let $f : S \to \mathbb{R}$ be a given functional.

(a) Let $\bar{x} \in S$ be a minimal point of f on S. If the functional f has a directional derivative at $\bar{x}$ in every direction $x - \bar{x}$ with arbitrary $x \in S$, then

$$f'(\bar{x})(x - \bar{x}) \geq 0 \text{ for all } x \in S. \tag{3.1}$$

(b) Let the set S be convex and let the functional f be convex. If the functional f has a directional derivative at some $\bar{x} \in S$ in every direction $x - \bar{x}$ with arbitrary $x \in S$ and the inequality (3.1) is satisfied, then $\bar{x}$ is a minimal point of f on S.

Proof

(a) Take any $x \in S$. Since f has a directional derivative at $\bar{x}$ in the direction $x - \bar{x}$, we have

$$f'(\bar{x})(x - \bar{x}) = \lim_{\lambda \to 0_+} \frac{1}{\lambda}(f(\bar{x} + \lambda(x - \bar{x})) - f(\bar{x})).$$

$\bar{x}$ is assumed to be a minimal point of f on S, and therefore we get for sufficiently small $\lambda > 0$

$$f(\bar{x} + \lambda(x - \bar{x})) \geq f(\bar{x}).$$

Consequently we obtain

$$f'(\bar{x})(x - \bar{x}) \geq 0.$$

(b) Because of the convexity of f we have for an arbitrary $x \in S$ and all $\lambda \in (0, 1]$

$$f(\bar{x} + \lambda(x - \bar{x})) = f(\lambda x + (1 - \lambda)\bar{x}) \leq \lambda f(x) + (1 - \lambda) f(\bar{x})$$

and especially

$$f(x) \geq f(\bar{x}) + \frac{1}{\lambda}(f(\bar{x} + \lambda(x - \bar{x})) - f(\bar{x})).$$

Since f has a directional derivative at $\bar{x}$ in the direction $x - \bar{x}$, it follows

$$f(x) \geq f(\bar{x}) + f'(\bar{x})(x - \bar{x})$$

and with the inequality (3.1) we obtain

$$f(x) \geq f(\bar{x}).$$

Consequently $\bar{x}$ is a minimal point of f on S. □

In part (b) of the preceding theorem one can weaken the assumptions on f and S, if one assumes only that f is convex at $\bar{x}$. In this case S needs only to be starshaped with respect to $\bar{x}$.

3.2 Gâteaux and Fréchet Derivatives

In this section we turn our attention to stronger differentiability notions. We want to ensure especially that differentiable mappings are also continuous. Furthermore we investigate a known problem from the calculus of variations.

Definition 3.9 (Gâteaux derivative).

Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be real normed spaces, let S be a nonempty open subset of X, and let $f : S \to Y$ be a given mapping. If for some $\bar{x} \in S$ and all $h \in X$ the limit

$$f'(\bar{x})(h) := \lim_{\lambda \to 0} \frac{1}{\lambda}(f(\bar{x} + \lambda h) - f(\bar{x}))$$

exists and if $f'(\bar{x})$ is a continuous linear mapping from X to Y, then $f'(\bar{x})$ is called the *Gâteaux derivative* of f at $\bar{x}$ and f is called *Gâteaux differentiable* at $\bar{x}$.

Example 3.10 (Gâteaux derivatives).

(a) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a given function with continuous partial derivatives. Then for every $\bar{x} \in \mathbb{R}^n$ the Gâteaux derivative of f at $\bar{x}$ reads as

$$f'(\bar{x})(h) = \frac{d}{d\lambda} f(\bar{x} + \lambda h)\Big|_{\lambda=0} = \nabla f(\bar{x} + \lambda h)^T h\Big|_{\lambda=0} = \nabla f(\bar{x})^T h \quad \text{for all } h \in \mathbb{R}^n.$$

(b) Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be real normed spaces, and let $L : X \to Y$ be a continuous linear mapping. Then the Gâteaux derivative of L at every $\bar{x} \in X$ is given as

$$L'(\bar{x})(h) = L(h) \text{ for all } h \in X.$$

Sometimes the notion of a Gâteaux derivative does not suffice in optimization theory. Therefore we present now a stronger concept of a derivative.

Definition 3.11 (Fréchet derivative).

Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be real normed spaces, let S be a nonempty open subset of X, and let $f : S \to Y$ be a given mapping. Furthermore let an element $\bar{x} \in S$ be given. If there is a continuous linear mapping $f'(\bar{x}) : X \to Y$ with the property

$$\lim_{\|h\|_X \to 0} \frac{\|f(\bar{x} + h) - f(\bar{x}) - f'(\bar{x})(h)\|_Y}{\|h\|_X} = 0,$$

then $f'(\bar{x})$ is called the *Fréchet derivative* of f at $\bar{x}$ and f is called *Fréchet differentiable* at $\bar{x}$.

According to this definition we obtain for Fréchet derivatives with the notations used above

$$f(\bar{x}+h) = f(\bar{x}) + f'(\bar{x})(h) + o(\|h\|_X)$$

where the expression $o(\|h\|_X)$ of this Taylor series has the property

$$\lim_{\|h\|_X \to 0} \frac{o(\|h\|_X)}{\|h\|_X} = \lim_{\|h\|_X \to 0} \frac{f(\bar{x}+h) - f(\bar{x}) - f'(\bar{x})(h)}{\|h\|_X} = 0_Y.$$

Example 3.12 (Fréchet derivative).

We consider a function $l : \mathbb{R}^3 \to \mathbb{R}$ which is continuous with respect to each of its arguments and which has continuous partial derivatives with respect to the two first arguments. Moreover we consider a functional $f : C^1[a,b] \to \mathbb{R}$ (with $-\infty < a < b < \infty$) given by

$$f(x) = \int_a^b l(x(t), \dot{x}(t), t)\, dt \text{ for all } x \in C^1[a,b].$$

Then we obtain for arbitrary $\bar{x}, h \in C^1[a,b]$

$$\begin{aligned} & f(\bar{x}+h) - f(\bar{x}) \\ &= \int_a^b [l(\bar{x}(t)+h(t), \dot{\bar{x}}(t)+\dot{h}(t), t) - l(\bar{x}(t), \dot{\bar{x}}(t), t)]\, dt \\ &= \int_a^b [l_x(\bar{x}(t), \dot{\bar{x}}(t), t)h(t) + l_{\dot{x}}(\bar{x}(t), \dot{\bar{x}}(t), t)\dot{h}(t)]\, dt + o(\|h\|_{C^1[a,b]}). \end{aligned}$$

Consequently the Fréchet derivative of f at $\bar{x}$ can be written as

$$f'(\bar{x})(h) = \int_a^b [l_x(\bar{x}(t), \dot{\bar{x}}(t), t)h(t) + l_{\dot{x}}(\bar{x}(t), \dot{\bar{x}}(t), t)\dot{h}(t)]\, dt$$
$$\text{for all } h \in C^1[a,b].$$

Next we present some important properties of Fréchet derivatives.

Theorem 3.13 (Fréchet and Gâteaux derivative).

Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be real normed spaces, let S be a nonempty open subset of X, and let $f : S \to Y$ be a given mapping. If the Fréchet derivative of f at some $\bar{x} \in S$ exists, then the Gâteaux derivative of f at $\bar{x}$ exists as well and both are equal.

Proof Let $f'(\bar{x})$ denote the Fréchet derivative of f at $\bar{x}$. Then we have

$$\lim_{\lambda \to 0} \frac{\|f(\bar{x}+\lambda h) - f(\bar{x}) - f'(\bar{x})(\lambda h)\|_Y}{\|\lambda h\|_X} = 0 \text{ for all } h \in X \backslash \{0_X\}$$

implying

$$\lim_{\lambda \to 0} \frac{1}{|\lambda|} \|f(\bar{x}+\lambda h) - f(\bar{x}) - f'(\bar{x})(\lambda h)\|_Y = 0 \text{ for all } h \in X \backslash \{0_X\}.$$

Because of the linearity of $f'(\bar{x})$ we obtain

$$\lim_{\lambda \to 0} \frac{1}{\lambda} [f(\bar{x}+\lambda h) - f(\bar{x})] = f'(\bar{x})(h) \text{ for all } h \in X.$$

□

Corollary 3.14 (uniqueness of the Fréchet derivative).

Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be real normed spaces, let S be a nonempty open subset of X, and let $f : S \to Y$ be a given mapping. If f is Fréchet differentiable at some $\bar{x} \in S$, then the Fréchet derivative is uniquely determined.

Proof With Theorem 3.13 the Fréchet derivative coincides with the Gâteaux derivative. Since the Gâteaux derivative is as a limit uniquely determined, the Fréchet derivative is also uniquely determined. □

The following theorem says that Fréchet differentiability implies continuity as well.

Theorem 3.15 (continuity of a Fréchet differentiable mapping).

Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be real normed spaces, let S be a nonempty open subset of X, and let $f : S \to Y$ be a given mapping. If f is Fréchet differentiable at some $\bar{x} \in S$, then f is continuous at $\bar{x}$.

Proof To a sufficiently small $\varepsilon > 0$ there is a ball around $\bar{x}$ so that for all $\bar{x} + h$ of this ball

$$\|f(\bar{x}+h) - f(\bar{x}) - f'(\bar{x})(h)\|_Y \leq \varepsilon \|h\|_X.$$

Then we conclude for some $\alpha > 0$

$$\begin{aligned}
\|f(\bar{x}+h) - f(\bar{x})\|_Y &= \|f(\bar{x}+h) - f(\bar{x}) - f'(\bar{x})(h) + f'(\bar{x})(h)\|_Y \\
&\leq \|f(\bar{x}+h) - f(\bar{x}) - f'(\bar{x})(h)\|_Y + \|f'(\bar{x})(h)\|_Y \\
&\leq \varepsilon \|h\|_X + \alpha \|h\|_X \\
&= (\varepsilon + \alpha)\|h\|_X.
\end{aligned}$$

Consequently f is continuous at $\bar{x}$. □

One obtains an interesting characterization of a convex functional, if it is Gâteaux differentiable. This result is summarized in the following theorem.

Theorem 3.16 (characterization of a convex functional).

Let S be a nonempty convex open subset of a real normed space $(X, \|\cdot\|)$, and let $f : S \to \mathbb{R}$ be a given functional which is Gâteaux differentiable at every $\bar{x} \in S$. Then the functional f is convex if and only if

$$f(y) \geq f(x) + f'(x)(y-x) \text{ for all } x, y \in S. \tag{3.2}$$

Proof

(a) First let us assume that the functional f is convex. Then we get for all $x, y \in S$ and all $\lambda \in (0, 1]$

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

resulting in

$$f(y) \geq f(x) + \frac{1}{\lambda}(f(x + \lambda(y-x)) - f(x)).$$

Since f is Gâteaux differentiable at x, it follows with Theorem 3.13

$$f(y) \geq f(x) + f'(x)(y-x).$$

(b) Now we assume that the inequality (3.2) is satisfied. The set S is convex, and therefore we obtain for all $x, y \in S$ and all $\lambda \in [0, 1]$

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

and

$$f(y) \geq f(\lambda x + (1-\lambda)y) + f'(\lambda x + (1-\lambda)y)(-\lambda(x-y)).$$

Since Gâteaux derivatives are linear mappings, we conclude further

$$\begin{aligned}
&\lambda f(x) + (1-\lambda) f(y) \\
&\geq \lambda f(\lambda x + (1-\lambda)y) + \lambda(1-\lambda) f'(\lambda x + (1-\lambda)y)(x-y) \\
&\quad + (1-\lambda) f(\lambda x + (1-\lambda)y) \\
&\quad - \lambda(1-\lambda) f'(\lambda x + (1-\lambda)y)(x-y) \\
&= f(\lambda x + (1-\lambda)y).
\end{aligned}$$

Consequently, the functional f is convex. □

If S is a nonempty convex open subset of $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ is a continuously partially differentiable function, then the inequality (3.2) can also be written as

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) \text{ for all } x, y \in S.$$

If one considers for every $x \in S$ the tangent plane to f at $(x, f(x))$, this inequality means geometrically that the function is above all of these tangent planes (see Fig. 3.4).

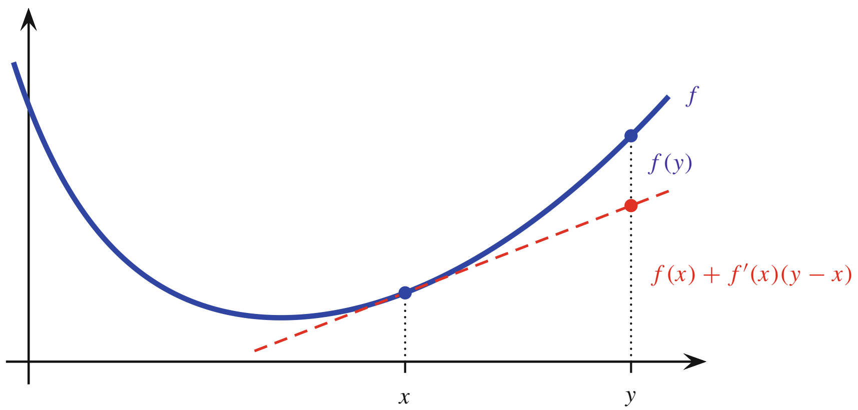

Fig. 3.4 Illustration of the result of Theorem 3.16

Next we formulate a necessary optimality condition for Gâteaux differentiable functionals.

Theorem 3.17 (necessary optimality condition).

Let $(X, \|\cdot\|)$ be a real normed space, and let $f : X \to \mathbb{R}$ be a given functional. If $\bar{x} \in X$ is a minimal point of f on X and f is Gâteaux differentiable at $\bar{x}$, then it follows

$$f'(\bar{x})(h) = 0 \text{ for all } h \in X.$$

Proof Let an element $h \in X$ be arbitrarily given. Then it follows for $x := h + \bar{x}$ with Theorem 3.8, (a)

$$f'(\bar{x})(h) \geq 0,$$

and for $x := -h + \bar{x}$ we get

$$f'(\bar{x})(-h) \geq 0.$$

Because of the linearity of the Gâteaux derivative the assertion follows immediately. □

Finally, we discuss an example from the calculus of variations. We proceed as in the proof of Theorem 3.17 which, in virtue of Theorem 3.13, holds also for Fréchet differentiable functionals.

Example 3.18 (calculus of variations).

We consider a function $l : \mathbb{R}^3 \to \mathbb{R}$ which is continuous with respect to all arguments and which has continuous partial derivatives with respect to the two first arguments. Moreover, let a functional $f : C^1[a, b] \to \mathbb{R}$ (with $-\infty < a < b < \infty$) with

$$f(x) = \int_a^b l(x(t), \dot{x}(t), t)\, dt \text{ for all } x \in C^1[a, b]$$

be given. But we are interested only in such functions x for which $x(a) = x_1$ and $x(b) = x_2$ where $x_1, x_2 \in \mathbb{R}$ are fixed endpoints. If we define the constraint set

$$S := \{x \in C^1[a, b] \mid x(a) = x_1 \text{ and } x(b) = x_2\},$$

then we ask for necessary optimality conditions for minimal points of f on S.
For the following we assume that $\bar{x} \in S$ is a minimal point of f on S. The constraint set S is convex and the objective functional f is Fréchet differentiable (compare Example 3.12). Then it follows from Theorem 3.8, (a) (in connection with Theorem 3.13) for the Fréchet derivative of f

$$f'(\bar{x})(x - \bar{x}) \geq 0 \text{ for all } x \in S$$

or

$$f'(\bar{x})(h) \geq 0 \text{ for all } h \in \tilde{S} := S - \{\bar{x}\}.$$

The set $\tilde{S}$ can also be written as

$$\tilde{S} = \{x \in C^1[a,b] \mid x(a) = x(b) = 0\}.$$

With $h \in \tilde{S}$ we have $-h \in \tilde{S}$ as well. Because of the linearity of the Fréchet derivative we obtain

$$f'(\bar{x})(h) = 0 \text{ for all } h \in \tilde{S}.$$

With Example 3.12 we have

$$f'(\bar{x})(h) = \int_a^b [l_x(\bar{x}(t), \dot{\bar{x}}(t), t)h(t) + l_{\dot{x}}(\bar{x}(t), \dot{\bar{x}}(t), t)\dot{h}(t)]\, dt$$

$$\text{for all } h \in \tilde{S}.$$

Hence our first result reads

$$\int_a^b [l_x(\bar{x}(t), \dot{\bar{x}}(t), t)h(t) + l_{\dot{x}}(\bar{x}(t), \dot{\bar{x}}(t), t)\dot{h}(t)]\, dt = 0 \text{ for all } h \in \tilde{S}. \tag{3.3}$$

For further conclusions in the previous example we need an important result which is prepared by the following lemma.

Lemma 3.19.

For $-\infty < a < b < \infty$ let

$$\tilde{S} = \{x \in C^1[a,b] \mid x(a) = x(b) = 0\}.$$

If for some function $x \in C[a, b]$

$$\int_a^b x(t)\dot{h}(t)\,dt = 0 \text{ for all } h \in \tilde{S},$$

then

$$x \equiv \text{ constant on } [a, b].$$

Proof We define

$$c := \frac{1}{b-a}\int_a^b x(t)\,dt$$

and choose especially $h \in \tilde{S}$ with

$$h(t) = \int_a^t (x(s) - c)\,ds \text{ for all } t \in [a, b].$$

Then we get

$$\begin{aligned}
\int_a^b (x(t) - c)^2 dt &= \int_a^b (x(t) - c)\dot{h}(t)\,dt \\
&= \int_a^b x(t)\dot{h}(t)\,dt - c[h(b) - h(a)] \\
&= -c\,h(b) \\
&= -c\left[\int_a^b x(s)\,ds - c(b-a)\right] \\
&= 0.
\end{aligned}$$

Hence it follows

$$x(t) = c \text{ for all } t \in [a, b].$$

□

Lemma 3.20 (fundamental lemma of calculus of variations).

For $-\infty < a < b < \infty$ *let*

$$\tilde{S} = \{x \in C^1[a,b] \mid x(a) = x(b) = 0\}.$$

If there are functions $x, y \in C[a,b]$ *with*

$$\int_a^b [x(t)h(t) + y(t)\dot{h}(t)]\,dt = 0 \text{ for all } h \in \tilde{S}, \tag{3.4}$$

then it follows $y \in C^1[a,b]$ *and* $\dot{y} = x$.

Proof We define a function $\varphi : [a,b] \to \mathbb{R}$ by

$$\varphi(t) = \int_a^t x(s)\,ds \text{ for all } t \in [a,b].$$

Then we obtain by integration by parts

$$\begin{aligned}\int_a^b x(t)h(t)\,dt &= \varphi(t)h(t)\Big|_a^b - \int_a^b \varphi(t)\dot{h}(t)\,dt \\ &= -\int_a^b \varphi(t)\dot{h}(t)\,dt \text{ for all } h \in \tilde{S},\end{aligned}$$

and from the equation (3.4) it follows

$$\int_a^b [-\varphi(t) + y(t)]\dot{h}(t)\,dt = 0 \text{ for all } h \in \tilde{S}.$$

With Lemma 3.19 we conclude for some constant $c \in \mathbb{R}$

$$y(t) = \varphi(t) + c \text{ for all } t \in [a,b].$$

Taking into consideration the definition of φ this equality leads to

$$\dot{y}(t) = x(t) \text{ for all } t \in [a,b],$$

and the assertion is shown. □

Using this last lemma we obtain the following theorem which is well known in the calculus of variations.

Theorem 3.21 (Euler-Lagrange equation).

Let the assumptions of Example 3.18 be satisfied. If $\bar{x} \in S$ is a minimal point of f on S, it follows

$$\frac{d}{dt} l_{\dot{x}}(\bar{x}(t), \dot{\bar{x}}(t), t) = l_x(\bar{x}(t), \dot{\bar{x}}(t), t) \text{ for all } t \in [a, b]. \tag{3.5}$$

Proof In Example 3.18 the equation (3.3) is already proved to be a necessary optimality condition. Then the application of Lemma 3.20 leads immediately to the assertion. □

In the calculus of variations the equation (3.5) is also called the *Euler-Lagrange equation*.

Example 3.22 (curve with smallest length).

Determine a curve $x \in C^1[a, b]$ (with $-\infty < a < b < \infty$) with smallest length which connects the two end points (a, x_1) and (b, x_2) (where $x_1, x_2 \in \mathbb{R}$). In other words: We are looking for a minimal point $\bar{x}$ of f on S with

$$S := \{x \in C^1[a, b] \mid x(a) = x_1 \text{ and } x(b) = x_2\}$$

and

$$f(x) = \int_a^b \sqrt{1 + \dot{x}(t)^2}\, dt \text{ for all } x \in S.$$

In this case the Euler-Lagrange equation (3.5) reads

$$\frac{d}{dt} \frac{\partial}{\partial \dot{x}} (\sqrt{1 + \dot{x}(t)^2}) \Big|_{x=\bar{x}} = 0.$$

This equation is equivalent to

$$\frac{d}{dt} \frac{2\dot{\bar{x}}(t)}{2\sqrt{1 + \dot{\bar{x}}(t)^2}} = 0.$$

Then we get for some constant $c \in \mathbb{R}$

$$\frac{\dot{x}(t)}{\sqrt{1+\dot{x}(t)^2}} = c \text{ for all } t \in [a,b]$$

and $\dot{x} \equiv$ constant. Hence we have the result that the optimal curve $\bar{x}$ is just the straight line connecting the points (a, x_1) and (b, x_2) (see Fig. 3.5). This result is certainly not surprising.

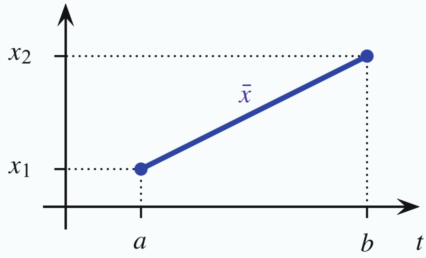

Fig. 3.5 Illustration of the result of Example 3.22

3.3 Subdifferential

In this section we present an additional concept of a derivative which is formulated especially for convex functionals. With the aid of this notion we derive the generalized Kolmogoroff condition known in approximation theory.

The characterization of convex Gâteaux differentiable functionals which is given in Theorem 3.16 proves to be very useful for the formulation of optimality conditions. This characterization motivates the following definition of a subgradient.

Definition 3.23 (subdifferential and subgradient).

Let $(X, \|\cdot\|)$ be a real normed space, and let $f : X \to \mathbb{R}$ be a convex functional. For an arbitrary $\bar{x} \in X$ the set $\partial f(\bar{x})$ of all continuous linear functionals l on X with

$$f(x) \geq f(\bar{x}) + l(x - \bar{x}) \text{ for all } x \in X$$

is called the *subdifferential* of f at $\bar{x}$. A continuous linear functional $l \in \partial f(\bar{x})$ is called a *subgradient* of f at $\bar{x}$ (see Fig. 3.6).

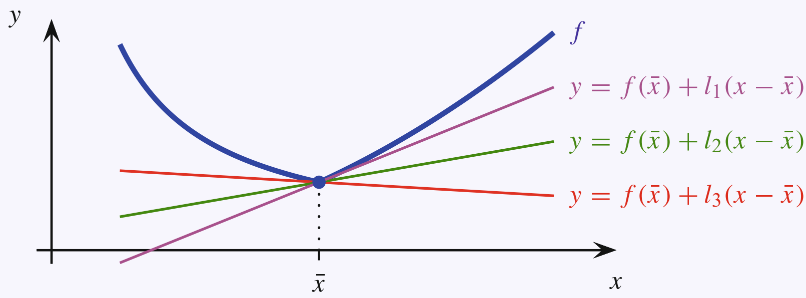

Fig. 3.6 Subgradients of a convex functional

Example 3.24 (subgradients).

(a) With Theorem 3.16 for every convex Gâteaux differentiable functional f defined on a real normed space the subdifferential $\partial f(\bar{x})$ at an arbitrary $\bar{x} \in X$ is nonempty. For every $\bar{x} \in X$ we have for the Gâteaux derivative $f'(\bar{x}) \in \partial f(\bar{x})$, i.e., $f'(\bar{x})$ is a subgradient of f at $\bar{x}$.

(b) Let $(X, \|\cdot\|)$ be a real normed space, and let $(X^*, \|\cdot\|_{X^*})$ denote the real normed space of continuous linear functionals on X (notice that $\|l\|_{X^*} = \sup\limits_{x \neq 0_X} \frac{|l(x)|}{\|x\|}$ for all $l \in X^*$).
Then for every $\bar{x} \in X$ the subdifferential of the norm at $\bar{x}$ is given as

$$\partial\|\bar{x}\| = \left\{ \begin{array}{ll} \{l \in X^* \mid l(\bar{x}) = \|\bar{x}\| \text{ and } \|l\|_{X^*} = 1\} & \text{if } \bar{x} \neq 0_X \\ \{l \in X^* \mid \|l\|_{X^*} \leq 1\} & \text{if } \bar{x} = 0_X \end{array} \right\}.$$

Proof of (b).

(i) For $\bar{x} = 0_X$ we obtain

$$\begin{aligned} \partial\|\bar{x}\| &= \{l \in X^* \mid \|x\| \geq l(x) \text{ for all } x \in X\} \\ &= \{l \in X^* \mid \frac{|l(x)|}{\|x\|} \leq 1 \text{ for all } x \in X \setminus \{0_X\}\} \\ &= \{l \in X^* \mid \|l\|_{X^*} \leq 1\}. \end{aligned}$$

(ii) Now let an arbitrary element $\bar{x} \neq 0_X$ be given. Then we obtain for every continuous linear functional $l \in X^*$ with $l(\bar{x}) = \|\bar{x}\|$ and $\|l\|_{X^*} = 1$ (see Theorem C.4 for the existence of such a functional)

$$l(x) \leq \|x\| \text{ for all } x \in X$$

which implies

$$\|\bar{x}\| + l(x - \bar{x}) = \|\bar{x}\| - l(\bar{x}) + l(x) \leq \|x\|.$$

Hence it follows $l \in \partial\|\bar{x}\|$.
Finally, we assume that l is a subgradient of the norm at $\bar{x} \neq 0_X$. Then we get

$$\|\bar{x}\| - l(\bar{x}) = \|2\bar{x}\| - \|\bar{x}\| - l(2\bar{x} - \bar{x}) \geq 0$$

and

$$-\|\bar{x}\| + l(\bar{x}) = \|0_X\| - \|\bar{x}\| - l(0_X - \bar{x}) \geq 0.$$

These two inequalities imply $l(\bar{x}) = \|\bar{x}\|$. Furthermore we obtain for all $x \in X$

$$\begin{aligned} \|x\| &\geq \|\bar{x}\| + l(x - \bar{x}) \\ &= \|\bar{x}\| + l(x) - \|\bar{x}\| \\ &= l(x). \end{aligned}$$

But then we conclude

$$\|l\|_{X^*} = \sup_{x \neq 0_X} \frac{|l(x)|}{\|x\|} \leq 1.$$

Because of $l(\bar{x}) = \|\bar{x}\|$ this leads to $\|l\|_{X^*} = 1$. So the assertion is proved. □

With the following lemma we also give an equivalent formulation of the subdifferential.

Lemma 3.25 (equivalent formulation of a subdifferential).

Let $(X, \|\cdot\|)$ be a real normed space, and let $f : X \to \mathbb{R}$ be a convex functional. Then we have for an arbitrary $\bar{x} \in X$

$$\partial f(\bar{x}) = \{l \in X^* \mid f'(\bar{x})(h) \geq l(h) \text{ for all } h \in X\}$$

(where $f'(\bar{x})(h)$ denotes the directional derivative of f at $\bar{x}$ in the direction h).

Proof

(a) For an arbitrary $l \in \partial f(\bar{x})$ we have

$$f'(\bar{x})(h) = \lim_{\lambda \to 0_+} \frac{1}{\lambda} \underbrace{(f(\bar{x} + \lambda h) - f(\bar{x}))}_{\geq l(\lambda h) = \lambda l(h)} \geq l(h) \text{ for all } h \in X.$$

Hence one set inclusion is shown.

(b) For the proof of the converse inclusion we assume that any $l \in X^*$ is given with

$$f'(\bar{x})(h) \geq l(h) \text{ for all } h \in X.$$

Then it follows with Lemma 3.3 (for $\lambda = 1$)

$$f(\bar{x} + h) - f(\bar{x}) \geq f'(\bar{x})(h) \geq l(h) \text{ for all } h \in X$$

which means that $l \in \partial f(\bar{x})$. □

Next we investigate the question under which assumption a convex functional already has a nonempty subdifferential.

Theorem 3.26 (existence of subgradients).

Let $(X, \|\cdot\|)$ be a real normed space, and let $f : X \to \mathbb{R}$ be a continuous convex functional. Then the subdifferential $\partial f(\bar{x})$ is nonempty for every $\bar{x} \in X$.

Proof Choose any point $\bar{x} \in X$. Since the functional f is continuous at $\bar{x}$, there is a ball around $\bar{x}$ on which the functional f is bounded from above by some $\bar{\alpha} \in \mathbb{R}$. Consequently, the epigraph $E(f)$ of f has a nonempty interior (e.g., $(\bar{x}, \bar{\alpha} + 1) \in \text{int}(E(f))$), and obviously we have $(\bar{x}, f(\bar{x})) \notin \text{int}(E(f))$. f is a convex functional, and therefore with Theorem 2.8 the epigraph $E(f)$ of f is convex. Hence the sets $E(f)$ and $\{(\bar{x}, f(\bar{x}))\}$ can be separated with the aid of the Eidelheit separation theorem (Theorem C.2). Then there are a number $\gamma \in \mathbb{R}$ and a continuous linear functional (l, β) on $X \times \mathbb{R}$ with $(l, \beta) \neq (0_{X^*}, 0)$ and

$$l(x) + \beta\alpha \leq \gamma \leq l(\bar{x}) + \beta f(\bar{x}) \text{ for all } (x, \alpha) \in E(f). \tag{3.6}$$

For $x = \bar{x}$ we obtain especially

$$\beta\alpha \leq \beta f(\bar{x}) \text{ for all } \alpha \geq f(\bar{x}).$$

Consequently we have $\beta \leq 0$. If we assume that $\beta = 0$, we obtain from the inequality (3.6)

$$l(x - \bar{x}) \leq 0 \text{ for all } x \in X$$

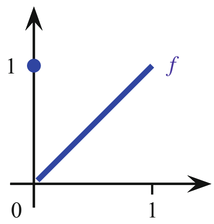

Fig. 3.7 Illustration of f

and therefore we conclude $l = 0_{X^*}$. But this is a contradiction to the condition $(l, \beta) \neq (0_{X^*}, 0)$. So we obtain $\beta < 0$, and the inequality (3.6) leads to

$$\frac{1}{\beta}l(x) + \alpha \geq \frac{1}{\beta}l(\bar{x}) + f(\bar{x}) \text{ for all } (x, \alpha) \in E(f)$$

which implies for $\alpha = f(x)$

$$f(x) \geq f(\bar{x}) - \frac{1}{\beta}l(x - \bar{x}) \text{ for all } x \in X.$$

Consequently, $-\frac{1}{\beta}l$ is an element of the subdifferential $\partial f(\bar{x})$. □

Under the assumptions of Theorem 3.26 it can be shown in addition that the subdifferential is a convex weak*-compact subset of X^*. Notice that with Lemma 2.13 the convex functional in the previous theorem is already continuous if it is continuous at some point.

Not every convex functional is already continuous. Figure 3.7 illustrates the function $f : [0, 1] \to \mathbb{R}$ with

$$f(x) = \left\{ \begin{array}{l} x \text{ if } 0 < x \leq 1 \\ 1 \text{ if } x = 0 \end{array} \right\},$$

which is convex but not continuous.

But it is shown in [74, p. 82–83] that every convex real-valued function defined on an open convex subset of $\mathbb{R}^n$ is continuous.

With the aid of subgradients we can immediately present a necessary and sufficient optimality condition. This theorem is formulated without proof because it is an obvious consequence of the definition of the subdifferential.

Theorem 3.27 (optimality condition).

Let $(X, \|\cdot\|)$ be a real normed space, and let $f : X \to \mathbb{R}$ be a convex functional. A point $\bar{x} \in X$ is a minimal point of f on X if and only if $0_{X^*} \in \partial f(\bar{x})$.

With the following theorem we investigate again the connection between the directional derivative and the subdifferential of a convex functional. We see that the directional derivative is the least upper bound of the subgradients (compare also Lemma 3.25).

Theorem 3.28 (directional derivative).

Let $(X, \|\cdot\|)$ be a real normed space, and let $f : X \to \mathbb{R}$ be a continuous convex functional. Then for every $\bar{x}, h \in X$ the directional derivative of f at $\bar{x}$ in the direction h is given as

$$f'(\bar{x})(h) = \max_{l \in \partial f(\bar{x})} l(h).$$

Proof Let $\bar{x} \in X$ be an arbitrary point and $h \in X$ be an arbitrary direction. With Theorem 3.4 the directional derivative $f'(\bar{x})(h)$ exists and with Theorem 3.26 the subdifferential $\partial f(\bar{x})$ is nonempty. With Lemma 3.25 we have

$$f'(\bar{x})(h) \geq l(h) \text{ for all } l \in \partial f(\bar{x}).$$

Hence it remains to show that there is a subgradient l with $f'(\bar{x})(h) = l(h)$. For that purpose we define the set

$$T := \{(\bar{x} + \lambda h, f(\bar{x}) + \lambda f'(\bar{x})(h)) \in X \times \mathbb{R} \mid \lambda \geq 0\}.$$

Because of Lemma 3.3 we have

$$f(\bar{x} + \lambda h) \geq f(\bar{x}) + \lambda f'(\bar{x})(h) \text{ for all } \lambda \geq 0.$$

Therefore we get

$$(\bar{x} + \lambda h, f(\bar{x}) + \lambda f'(\bar{x})(h)) \notin \text{int}(E(f)) \text{ for all } \lambda \geq 0$$

(as in the proof of Theorem 3.26 notice that the epigraph of f has a nonempty interior because f is continuous). Then it follows $\text{int}(E(f)) \cap T = \emptyset$. If we also notice that the sets $S := E(f)$ and T are convex, then the Eidelheit separation theorem is applicable (Theorem C.2). Consequently, there are a continuous linear functional l on X and real numbers β and γ with the property $(l, \beta) \neq (0_{X^*}, 0)$ and

$$l(x) + \beta\alpha \leq \gamma \leq l(\bar{x} + \lambda h) + \beta(f(\bar{x}) + \lambda f'(\bar{x})(h)) \tag{3.7}$$
$$\text{for all } (x, \alpha) \in E(f) \text{ and all } \lambda \geq 0.$$

For $x = \bar{x}$ and $\lambda = 0$ we obtain especially

$$\beta\alpha \leq \beta f(\bar{x}) \text{ for all } \alpha \geq f(\bar{x})$$

which leads to $\beta \leq 0$. If we assume that $\beta = 0$, then we obtain from the inequality (3.7) with $\lambda = 0$

$$l(x - \bar{x}) \leq 0 \text{ for all } x \in X$$

and therefore $l = 0_{X^*}$. But this is a contradiction to the condition $(l, \beta) \neq (0_{X^*}, 0)$. Consequently we get $\beta < 0$, and from the inequality (3.7) we conclude

$$\frac{1}{\beta} l(x - \bar{x} - \lambda h) + \alpha \geq f(\bar{x}) + \lambda f'(\bar{x})(h) \tag{3.8}$$
$$\text{for all } (x, \alpha) \in E(f) \text{ and all } \lambda \geq 0.$$

For $\alpha = f(x)$ and $\lambda = 0$ we obtain

$$f(x) \geq f(\bar{x}) - \frac{1}{\beta} l(x - \bar{x}) \text{ for all } x \in X,$$

i.e., $-\frac{1}{\beta}l$ is a subgradient of f at $\bar{x}$. For $x = \bar{x}$, $\alpha = f(\bar{x})$ and $\lambda = 1$ we also conclude from the inequality (3.8)

$$f'(\bar{x})(h) \leq -\frac{1}{\beta} l(h).$$

Because of $-\frac{1}{\beta}l \in \partial f(\bar{x})$ the assertion is shown. □

As a result of the previous theorem the following necessary and sufficient optimality condition can be given.

Corollary 3.29 (optimality condition).

Let S be a nonempty subset of a real normed space $(X, \|\cdot\|)$, and let $f : X \to \mathbb{R}$ be a continuous convex functional.

(a) *If* $\bar{x} \in S$ *is a minimal point of* f *on* S *and* S *is starshaped with respect to* $\bar{x}$*, then*

$$0 \leq \max_{l \in \partial f(\bar{x})} l(x - \bar{x}) \text{ for all } x \in S. \tag{3.9}$$

(b) *If for some* $\bar{x} \in S$ *the inequality (3.9) is satisfied, then* $\bar{x}$ *is a minimal point of* f *on* S*.*

Proof The part (a) of this theorem follows immediately from the Theorems 3.8,(a) and 3.28 (together with a remark on page 37). For the proof of the part (b) notice that with Theorem 3.28 and Lemma 3.3 it follows from the inequality (3.9)

$$\frac{1}{\lambda}(f(\bar{x} + \lambda(x - \bar{x})) - f(\bar{x})) \geq f'(\bar{x})(x - \bar{x}) \geq 0 \text{ for all } x \in S \text{ and all } \lambda > 0.$$

Hence we get for $\lambda = 1$

$$f(\bar{x}) \leq f(x) \text{ for all } x \in S.$$

Consequently, $\bar{x}$ is a minimal point of f on S. □

For the application of this corollary we turn our attention to approximation problems.

Corollary 3.30 (generalized Kolmogorov condition).

Let S *be a nonempty subset of a real normed space* $(X, \|\cdot\|)$*, and let* $\hat{x} \in X \setminus S$ *be a given element.*

(a) *If* $\bar{x} \in S$ *is a best approximation to* $\hat{x}$ *from* S *and* S *is starshaped with respect to* $\bar{x}$*, then*

$$\max\{l(x - \bar{x}) \mid l \in X^*,\ \ l(\bar{x} - \hat{x}) = \|\bar{x} - \hat{x}\| \text{ and } \|l\|_{X^*} = 1\} \geq 0 \text{ for all } x \in S. \tag{3.10}$$

(b) *If for some* $\bar{x} \in S$ *the inequality (3.10) is satisfied, then* $\bar{x}$ *is a best approximation to* $\hat{x}$ *from* S*.*

Proof $\bar{x} \in S$ is a best approximation to $\hat{x}$ from S if and only if $\bar{x} - \hat{x} \neq 0_X$ is a minimal point of the norm $\|\cdot\|$ on $S - \{\hat{x}\}$. With Example 3.24, (b) we have

$$\partial\|\bar{x} - \hat{x}\| = \{l \in X^* \mid l(\bar{x} - \hat{x}) = \|\bar{x} - \hat{x}\| \text{ and } \|l\|_{X^*} = 1\}.$$

Then the inequality (3.9) is equivalent to the inequality

$$\max\{l(x - \bar{x} + \hat{x}) \mid l \in X^*,\ l(\bar{x} - \hat{x}) = \|\bar{x} - \hat{x}\| \text{ and } \|l\|_{X^*} = 1\} \geq 0 \quad \text{for all } x \in S - \{\hat{x}\}$$

resulting in

$$\max\{l(x - \bar{x}) \mid l \in X^*,\ l(\bar{x} - \hat{x}) = \|\bar{x} - \hat{x}\| \text{ and } \|l\|_{X^*} = 1\} \geq 0 \quad \text{for all } x \in S.$$

Finally notice in part (a) that the set $S - \{\hat{x}\}$ is starshaped with respect to $\bar{x} - \hat{x}$ and the norm $\|\cdot\|$ is a continuous functional (compare page 23). So this theorem is proved using Corollary 3.29. □

The optimality condition for approximation problems given in Theorem 3.30 is also called *generalized Kolmogorov condition* in approximation theory.

3.4 Quasidifferential

The theory of subdifferentials may also be extended to certain nonconvex functionals. Such an extension was proposed by Dem'yanov and Rubinov[2] and is the subject of this section. We give only a short introduction to this theory of quasidifferentials.

Definition 3.31 (quasidifferential).

Let S be a nonempty open subset of a real normed space $(X, \|\cdot\|)$, let $f : S \to \mathbb{R}$ be a given functional, and let $\bar{x} \in S$ be a given element. The functional f is called *quasidifferentiable* at $\bar{x}$ if f is directionally differentiable at $\bar{x}$ and if there are two nonempty convex weak*-compact subsets $\underline{\partial} f(\bar{x})$ and $\overline{\partial} f(\bar{x})$ of the topological dual space X^* with the property

$$f'(\bar{x})(h) = \max_{\underline{l} \in \underline{\partial} f(\bar{x})} \underline{l}(h) + \min_{\overline{l} \in \overline{\partial} f(\bar{x})} \overline{l}(h) \text{ for all } h \in X.$$

The pair of sets $Df(\bar{x}) := (\underline{\partial} f(\bar{x}), \overline{\partial} f(\bar{x}))$ is called a *quasidifferential* of f at $\bar{x}$, and the sets $\underline{\partial} f(\bar{x})$ and $\overline{\partial} f(\bar{x})$ are called *subdifferential* and *superdifferential* of f at $\bar{x}$, respectively.

[2] V.F. Dem'yanov and A.M. Rubinov, "On quasidifferentiable functionals", *Soviet Math. Dokl.* 21 (1980) 14–17.

Quasidifferentials have interesting properties. But, in general, it is difficult to determine a quasidifferential to a given functional.

Notice in the preceding definition that the subdifferential and the superdifferential are not uniquely determined. For instance, for every ball $B(0_{X^*}, \varepsilon) := \{l \in X^* \mid \|l\|_{X^*} \leq \varepsilon\}$ with an arbitrary $\varepsilon > 0$ the pair of sets $(\underline{\partial} f(\bar{x}) + B(0_{X^*}, \varepsilon), \overline{\partial} f(\bar{x}) - B(0_{X^*}, \varepsilon))$ is a quasidifferential of f at $\bar{x}$ as well.

Example 3.32 (difference of convex functionals).

Let $(X, \|\cdot\|)$ be a real normed space, and let $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}$ be convex functionals. If f and g are continuous at some $\bar{x} \in X$, then the functional $\varphi := f - g$ is quasidifferentiable at $\bar{x}$. In this case $(\partial f(\bar{x}), -\partial g(\bar{x}))$ is a quasidifferential of φ at $\bar{x}$ where $\partial f(\bar{x})$ and $\partial g(\bar{x})$ denote the subdifferential of f and g at $\bar{x}$, respectively.

Proof By Theorem 3.4 f and g are directionally differentiable and therefore $\varphi = f - g$ is also directionally differentiable. If $\partial f(\bar{x})$ and $\partial g(\bar{x})$ denote the subdifferential of f and g at $\bar{x}$ (these two sets are nonempty, convex and weak*-compact), we define the sets $\underline{\partial}\varphi(\bar{x}) := \partial f(\bar{x})$ and $\overline{\partial}\varphi(\bar{x}) := -\partial g(\bar{x})$. By Theorem 3.28 the directional derivative of φ is given as

$$\begin{aligned}
\varphi'(\bar{x})(h) &= f'(\bar{x})(h) - g'(\bar{x})(h) \\
&= \max_{\underline{l} \in \partial f(\bar{x})} \underline{l}(h) - \max_{\overline{l} \in \partial g(\bar{x})} \overline{l}(h) \\
&= \max_{\underline{l} \in \underline{\partial}\varphi(\bar{x})} \underline{l}(h) + \min_{\overline{l} \in \overline{\partial}\varphi(\bar{x})} \overline{l}(h) \text{ for all } h \in X.
\end{aligned}$$

Hence $D\varphi(\bar{x}) := (\partial f(\bar{x}), -\partial g(\bar{x}))$ is a quasidifferential of φ at $\bar{x}$. □

This example shows that the concept of the quasidifferential is suitable for functionals which may be represented as the difference of two convex functionals. These functionals are also called *d.c. functionals*.

For locally Lipschitz continuous functionals we can present an interesting characterization of the notion of quasidifferentiability. We show the equivalence of the quasidifferentiability to a certain "Fréchet property" for locally Lipschitz continuous functionals on $\mathbb{R}^n$.

Definition 3.33 (Lipschitz continuity).

Let S be a nonempty subset of a real normed space $(X, \|\cdot\|)$, let $f : S \to \mathbb{R}$ be a given functional, and let $\bar{x} \in S$ be a given element. f is called *Lipschitz continuous* at $\bar{x}$ if there is a constant $k \geq 0$ and some $\varepsilon > 0$ with

$$|f(x) - f(y)| \leq k\|x - y\| \text{ for all } x, y \in S \cap B(\bar{x}, \varepsilon)$$

where

$$B(\bar{x}, \varepsilon) := \{x \in X \mid \|x - \bar{x}\| \leq \varepsilon\}.$$

f is called *Lipschitz continuous* if there is a constant $k \geq 0$ with

$$|f(x) - f(y)| \leq k\|x - y\| \text{ for all } x, y \in S.$$

The constant k is also called *Lipschitz constant*.

Definition 3.34 (Fréchet property).

Let S be a nonempty open subset of a real normed space $(X, \|\cdot\|)$, let $f : S \to \mathbb{R}$ be a given functional, let $\bar{f} : X \to \mathbb{R}$ be a positively homogeneous and Lipschitz continuous functional, and let $\bar{x} \in S$ be a given element. f is said to have the *Fréchet property* at $\bar{x}$ with the functional $\bar{f}$ if

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

If f is Fréchet differentiable at some $\bar{x} \in S$, then it has also the Fréchet property at $\bar{x}$ with $\bar{f} := f'(\bar{x})$ (Fréchet derivative of f at $\bar{x}$) because the Fréchet derivative $f'(\bar{x})$ is continuous and linear, and therefore it is also positively homogeneous and Lipschitz continuous. Hence the concept of the Fréchet property of a functional is closely related to the concept of the Fréchet differentiability.

The following theorem is due to Schade[3] and it is based on a result of Pallaschke, Recht and Urbański[4] stated in Theorem 3.36. It plays only the role of a lemma for Theorem 3.36, and it says that every directionally differentiable and locally Lipschitz continuous functional defined on $\mathbb{R}^n$ has already the Fréchet property.

Theorem 3.35 (Fréchet property).

Let S be a nonempty open subset of $\mathbb{R}^n$, and let $\bar{x} \in S$ be a given element. Every functional $f : S \to \mathbb{R}$ which is Lipschitz continuous at $\bar{x}$ and directionally differentiable at $\bar{x}$ has the Fréchet property at $\bar{x}$ with $\bar{f} := f'(\bar{x})$ (directional derivative of f at $\bar{x}$).

[3] R. Schade, *Quasidifferenzierbare Abbildungen* (diplom thesis, Technical University of Darmstadt, Germany, 1987).

[4] D. Pallaschke, P. Recht and R. Urbański, "On Locally-Lipschitz Quasi-Differentiable Functions in Banach-Spaces", *optimization* 17 (1986) 287–295.

Proof Let $f : S \to \mathbb{R}$ be Lipschitz continuous at $\bar{x}$ and directionally differentiable at $\bar{x}$. Since $f : S \to \mathbb{R}$ is Lipschitz continuous at $\bar{x}$, i.e., there are numbers $k \geq 0$ and $\varepsilon > 0$ with

$$|f(x) - f(y)| \leq k\|x - y\| \text{ for all } x, y \in S \cap B(\bar{x}, \varepsilon), \tag{3.11}$$

the directional derivative $f'(\bar{x}) : \mathbb{R}^n \to \mathbb{R}$ of f at $\bar{x}$ is also Lipschitz continuous because for every $x_1, x_2 \in \mathbb{R}^n$

$$\begin{aligned} |f'(\bar{x})(x_1) - f'(\bar{x})(x_2)| &= \left| \lim_{\lambda \to 0_+} \frac{1}{\lambda} (f(\bar{x} + \lambda x_1) - f(\bar{x})) \right. \\ &\quad \left. - \lim_{\lambda \to 0_+} \frac{1}{\lambda} (f(\bar{x} + \lambda x_2) - f(\bar{x})) \right| \\ &= \left| \lim_{\lambda \to 0_+} \frac{1}{\lambda} (f(\bar{x} + \lambda x_1) - f(\bar{x} + \lambda x_2)) \right| \\ &\leq \lim_{\lambda \to 0_+} \frac{1}{\lambda} k \|\lambda x_1 - \lambda x_2\| \\ &= k\|x_1 - x_2\|. \end{aligned} \tag{3.12}$$

So, $f'(\bar{x})$ is Lipschitz continuous and it is obvious that $f'(\bar{x})$ is also positively homogeneous.

Now assume that f does not have the Fréchet property at $\bar{x}$ with $\bar{f} := f'(\bar{x})$. Then we get for $\bar{f} := f'(\bar{x})$ which is positively homogeneous and Lipschitz continuous

$$\lim_{\|h\| \to 0} \frac{|f(\bar{x} + h) - f(\bar{x}) - f'(\bar{x})(h)|}{\|h\|} \neq 0.$$

Consequently, there is a $\beta > 0$ so that for all $i \in \mathbb{N}$ there is some $h_i \in \mathbb{R}^n$ with $0 \neq \|h_i\| \leq \frac{1}{i}$ and

$$|f(\bar{x} + h_i) - f(\bar{x}) - f'(\bar{x})(h_i)| \geq \beta \|h_i\|. \tag{3.13}$$

Next we set

$$g_i := \frac{\varepsilon h_i}{\|h_i\|} \text{ for all } i \in \mathbb{N}. \tag{3.14}$$

Obviously we have

$$\|g_i\| = \varepsilon \text{ for all } i \in \mathbb{N}, \tag{3.15}$$

i.e., g_i belongs to the sphere $\{x \in \mathbb{R}^n \mid \|x\| = \varepsilon\}$ which is compact. Therefore the sequence $(g_i)_{i\in\mathbb{N}}$ has a subsequence $(g_{i_j})_{j\in\mathbb{N}}$ converging to some g with $\|g\| = \varepsilon$. If we also set

$$\alpha_i := \frac{\|h_i\|}{\varepsilon} > 0 \text{ for all } i \in \mathbb{N},$$

we obtain $\lim\limits_{i\to\infty} \alpha_i = 0$ and with the equality (3.14)

$$h_i = \alpha_i g_i \text{ for all } i \in \mathbb{N}. \tag{3.16}$$

Finally we define for every $i \in \mathbb{N}$

$$\begin{aligned}
\phi_i &:= |f(\bar{x} + \alpha_i g) - f(\bar{x}) - f'(\bar{x})(\alpha_i g)| \\
&= |f(\bar{x} + h_i) - f(\bar{x}) - f'(\bar{x})(h_i) - f(\bar{x} + h_i) + f(\bar{x} + \alpha_i g) \\
&\quad + f'(\bar{x})(h_i) - f'(\bar{x})(\alpha_i g)| \\
&= |[f(\bar{x} + h_i) - f(\bar{x}) - f'(\bar{x})(h_i)] \\
&\quad - [(f(\bar{x} + h_i) - f(\bar{x} + \alpha_i g)) - (f'(\bar{x})(h_i) - f'(\bar{x})(\alpha_i g))]| \\
&\geq |f(\bar{x} + h_i) - f(\bar{x}) - f'(\bar{x})(h_i)| \\
&\quad - |(f(\bar{x} + h_i) - f(\bar{x} + \alpha_i g)) - (f'(\bar{x})(h_i) - f'(\bar{x})\alpha_i g))| \\
&\geq |f(\bar{x} + h_i) - f(\bar{x}) - f'(\bar{x})(h_i)| \\
&\quad - (|f(\bar{x} + h_i) - f(\bar{x} + \alpha_i g)| + |f'(\bar{x})(h_i) - f'(\bar{x})(\alpha_i g)|).
\end{aligned}$$

For sufficiently large $i \in \mathbb{N}$ we have

$$\bar{x} + h_i \in S \cap B(\bar{x}, \varepsilon)$$

and

$$\bar{x} + \alpha_i g \in S \cap B(\bar{x}, \varepsilon),$$

and therefore we get with the inequalities (3.13), (3.11), (3.12) and the equalities (3.16), (3.15)

$$\begin{aligned}
\phi_i &\geq \beta\|h_i\| - (k\|h_i - \alpha_i g\| + k\|h_i - \alpha_i g\|) \\
&= \beta\alpha_i\|g_i\| - 2k\alpha_i\|g_i - g\| \\
&= \alpha_i(\beta\varepsilon - 2k\|g_i - g\|).
\end{aligned}$$

Since the sequence $(g_{i_j})_{j\in\mathbb{N}}$ converges to g, we obtain for sufficiently large $j \in \mathbb{N}$

$$\|g_{i_j} - g\| \leq \frac{\beta\varepsilon}{4k}.$$

Hence we conclude for sufficiently large $j \in \mathbb{N}$

$$\phi_{i_j} \geq \alpha_{i_j}\left(\beta\varepsilon - \frac{\beta\varepsilon}{2}\right) = \alpha_{i_j}\frac{\beta\varepsilon}{2}$$

and because of the positive homogenity of $f'(\bar{x})$

$$\begin{aligned}
&\left|\frac{f(\bar{x} + \alpha_{i_j} g) - f(\bar{x})}{\alpha_{i_j}} - f'(\bar{x})(g)\right| \\
&\quad = \frac{|f(\bar{x} + \alpha_{i_j} g) - f(\bar{x}) - f'(\bar{x})(\alpha_{i_j} g)|}{\alpha_{i_j}} \\
&\quad = \frac{\phi_{i_j}}{\alpha_{i_j}} \geq \frac{\beta\varepsilon}{2} > 0.
\end{aligned}$$

From the preceding inequality it follows

$$f'(\bar{x})(g) \neq \lim_{j\to\infty} \frac{f(\bar{x} + \alpha_{i_j} g) - f(\bar{x})}{\alpha_{i_j}}$$

which is a contradiction to the definition of the directional derivative. □

The preceding theorem presents an interesting property of directionally differentiable and locally Lipschitz continuous functionals on $\mathbb{R}^n$. It is now used in order to prove the equivalence of the quasidifferentiability to the Fréchet property for locally Lipschitz continuous functionals on $\mathbb{R}^n$.

Theorem 3.36 (characterization of quasidifferentiability).

Let S be a nonempty open subset of $\mathbb{R}^n$, let $\bar{x} \in S$ be a given element, and let $f : S \to \mathbb{R}$ be a given functional which is Lipschitz continuous at $\bar{x}$. The functional f is quasidifferentiable at $\bar{x}$ if and only if f has the Fréchet property at $\bar{x}$ with some functional $\bar{f} : \mathbb{R}^n \to \mathbb{R}$ which can be represented as difference of two Lipschitz continuous sublinear functionals.

Proof

(i) First, assume that f is quasidifferentiable at $\bar{x}$. Then f is also directionally differentiable at $\bar{x}$, and by Theorem 3.35 it has the Fréchet property at $\bar{x}$ with the directional derivative of f at $\bar{x}$

$$\begin{aligned}\bar{f} := f'(\bar{x}) &= \max_{\underline{l}\in\underline{\partial} f(\bar{x})} \underline{l}(\cdot) + \min_{\bar{l}\in\bar{\partial} f(\bar{x})} \bar{l}(\cdot) \\ &= \max_{\underline{l}\in\underline{\partial} f(\bar{x})} \underline{l}(\cdot) - \max_{\bar{l}\in-\bar{\partial} f(\bar{x})} \bar{l}(\cdot). \end{aligned} \tag{3.17}$$

Next we define the functional $\varphi : \mathbb{R}^n \to \mathbb{R}$ by

$$\varphi(h) := \max_{\underline{l}\in\underline{\partial} f(\bar{x})} \underline{l}(h) \text{ for all } h \in \mathbb{R}^n.$$

φ is sublinear because for all $h_1, h_2 \in \mathbb{R}^n$ and all $\lambda \geq 0$ we have

$$\begin{aligned}\varphi(h_1 + h_2) &= \max_{\underline{l}\in\underline{\partial} f(\bar{x})} \underline{l}(h_1 + h_2) \\ &= \max_{\underline{l}\in\underline{\partial} f(\bar{x})} \underline{l}(h_1) + \underline{l}(h_2) \\ &\leq \max_{\underline{l}\in\underline{\partial} f(\bar{x})} \underline{l}(h_1) + \max_{\underline{l}\in\underline{\partial} f(\bar{x})} \underline{l}(h_2) \\ &= \varphi(h_1) + \varphi(h_2)\end{aligned}$$

and

$$\begin{aligned}\varphi(\lambda h_1) = \max_{\underline{l}\in\underline{\partial} f(\bar{x})} \underline{l}(\lambda h_1) &= \max_{\underline{l}\in\underline{\partial} f(\bar{x})} \lambda\underline{l}(h_1) \\ = \lambda \max_{\underline{l}\in\underline{\partial} f(\bar{x})} \underline{l}(h_1) &= \lambda\varphi(h_1).\end{aligned}$$

The functional φ is also continuous because for all $h \in \mathbb{R}^n$

$$\begin{aligned}|\varphi(h)| = \left| \max_{\underline{l}\in\underline{\partial} f(\bar{x})} \underline{l}(h) \right| &\leq \max_{\underline{l}\in\underline{\partial} f(\bar{x})} |\underline{l}(h)| \\ &\leq \max_{\underline{l}\in\underline{\partial} f(\bar{x})} \|\underline{l}\| \, \|h\| \\ &= \|h\| \max_{\underline{l}\in\underline{\partial} f(\bar{x})} \|\underline{l}\| \\ &= \|h\| L \end{aligned} \tag{3.18}$$

with

$$L := \max_{\underline{l} \in \underline{\partial} f(\bar{x})} \|\underline{l}\|$$

($L > 0$ exists because $\underline{\partial} f(\bar{x})$ is weak*-compact).

Now we show that the continuous sublinear functional φ is also Lipschitz continuous. For that proof take any $h_1, h_2 \in \mathbb{R}^n$. Then we get with the inequality (3.18)

$$\begin{aligned} \varphi(h_1) = \varphi(h_1 - h_2 + h_2) &\leq \varphi(h_1 - h_2) + \varphi(h_2) \\ &\leq L\|h_1 - h_2\| + \varphi(h_2) \end{aligned}$$

resulting in

$$\varphi(h_1) - \varphi(h_2) \leq L\|h_1 - h_2\|.$$

Similarly one obtains

$$\varphi(h_2) - \varphi(h_1) \leq L\|h_1 - h_2\|,$$

and so it follows

$$|\varphi(h_1) - \varphi(h_2)| \leq L\|h_1 - h_2\|.$$

Consequently we have shown that f has the Fréchet property at $\bar{x}$ with $\bar{f} := f'(\bar{x})$ which, by the equation (3.17), can be written as the difference of two Lipschitz continuous sublinear functionals.

(ii) Now we assume that f has the Fréchet property at $\bar{x}$ with some functional $\bar{f} : \mathbb{R}^n \to \mathbb{R}$ which can be represented as difference of two Lipschitz continuous sublinear functionals. First we prove that $\bar{f}$ is the directional derivative $f'(\bar{x})$ of f at $\bar{x}$. Because of the positive homogenity of $f'(\bar{x})$ and $\bar{f}$ we have

$$f'(\bar{x})(0_{\mathbb{R}^n}) = \bar{f}(0_{\mathbb{R}^n}) = 0.$$

Since f has the Fréchet property at $\bar{x}$ with $\bar{f}$, we get for every $h \in \mathbb{R}^n$, $h \neq 0_{\mathbb{R}^n}$,

$$\lim_{\lambda \to 0_+} \frac{|f(\bar{x} + \lambda h) - f(\bar{x}) - \bar{f}(\lambda h)|}{\|\lambda h\|} = 0$$

and

$$\lim_{\lambda \to 0_+} \frac{|f(\bar{x} + \lambda h) - f(\bar{x}) - \bar{f}(\lambda h)|}{\lambda} = 0.$$

Because $\bar{f}$ is positively homogeneous, we obtain

$$\lim_{\lambda \to 0_+} \left| \frac{f(\bar{x} + \lambda h) - f(\bar{x})}{\lambda} - \bar{f}(h) \right| = 0.$$

Hence f is directionally differentiable at $\bar{x}$ with $\bar{f} = f'(\bar{x})$, and the directional derivative $f'(\bar{x})$ can be written as difference of two Lipschitz continuous sublinear functionals $\varphi_1, \varphi_2 : \mathbb{R}^n \to \mathbb{R}$, i.e.

$$f'(\bar{x}) = \varphi_1 - \varphi_2. \tag{3.19}$$

Now fix an arbitrary $i \in \{1, 2\}$ and define the set

$$A_i := \{\varphi \in \mathbb{R}^n \mid \varphi^T x \leq \varphi_i(x) \text{ for all } x \in \mathbb{R}^n\}$$

which is nonempty convex and weak*-compact (in fact, it is a compact subset of $\mathbb{R}^n$). Then we have for all $x \in \mathbb{R}^n$

$$\varphi_i(x) \geq \max_{\varphi \in A_i} \varphi^T x. \tag{3.20}$$

Next, fix any $\bar{x} \in \mathbb{R}^n$ and consider the set $\{(\bar{x}, \varphi_i(\bar{x}))\}$ and the epigraph $E(\varphi_i)$. Notice that this epigraph is convex and it has a nonempty interior because φ_i is a Lipschitz continuous sublinear functional. Then by the application of the Eidelheit separation theorem (Theorem C.2) there are a number $\gamma \in \mathbb{R}$ and a vector $(l, \beta) \in \mathbb{R}^{n+1}$ with $(l, \beta) \neq 0_{\mathbb{R}^{n+1}}$ and

$$l^T x + \beta\alpha \leq \gamma \leq l^T \bar{x} + \beta\varphi_i(\bar{x}) \text{ for all } (x, \alpha) \in E(\varphi_i). \tag{3.21}$$

With the same arguments used in the proof of Theorem 3.26 we get $\beta < 0$. If we set $\bar{\varphi} := -\frac{1}{\beta} l$, we get for $x = 0_{\mathbb{R}^n}$ and $\alpha = \varphi_i(0_{\mathbb{R}^n}) = 0$ from the inequality (3.21)

$$\varphi_i(\bar{x}) \leq \bar{\varphi}^T \bar{x}. \tag{3.22}$$

It follows from the inequality (3.21) that

$$l^T x + \beta\varphi_i(x) \leq 0 \text{ for all } x \in \mathbb{R}^n \tag{3.23}$$

(otherwise we get for some $x \in \mathbb{R}^n$ with $l^T x + \beta\varphi_i(x) > 0$

$$l^T(\delta x) + \beta\varphi_i(\delta x) = \delta(l^T x + \beta\varphi_i(x)) \longrightarrow \infty \text{ for } \delta \to \infty$$

which contradicts the inequality (3.21)). From the inequality (3.23) we conclude

$$\bar{\varphi}^T x - \varphi_i(x) \leq 0 \text{ for all } x \in \mathbb{R}^n,$$

i.e. $\bar{\varphi} \in A_i$. Then it follows from the inequalities (3.20) and (3.22) that

$$\varphi_i(x) = \max_{\varphi \in A_i} \varphi^T x,$$

and so we have with the equality (3.19)

$$\begin{aligned} f'(\bar{x})(x) &= \max_{\varphi \in A_1} \varphi^T x - \max_{\varphi \in A_2} \varphi^T x \\ &= \max_{\varphi \in A_1} \varphi^T x + \min_{\varphi \in -A_2} \varphi^T x \text{ for all } x \in \mathbb{R}^n. \end{aligned}$$

Consequently, the functional f is quasidifferentiable at $\bar{x}$. □

Finally, we also present a necessary optimality condition for quasidifferentiable functionals.

Theorem 3.37 (necessary optimality condition).

Let $(X, \|\cdot\|)$ be a real normed space, and let $f : X \to \mathbb{R}$ be a given functional. If $\bar{x} \in X$ is a minimal point of f on X and if f is quasidifferentiable at $\bar{x}$ with a quasidifferential $(\underline{\partial} f(\bar{x}), \overline{\partial} f(\bar{x}))$, then it follows

$$-\overline{\partial} f(\bar{x}) \subset \underline{\partial} f(\bar{x}).$$

Proof Using Theorem 3.8,(a) we obtain the following necessary optimality condition for the directional derivative:

$$f'(\bar{x})(h) \geq 0 \text{ for all } h \in X.$$

Then, by Definition 3.31, we get for a quasidifferential $(\underline{\partial} f(\bar{x}), \overline{\partial} f(\bar{x}))$

$$\begin{aligned} \max_{\underline{l} \in \underline{\partial} f(\bar{x})} \underline{l}(h) &\geq - \min_{\overline{l} \in \overline{\partial} f(\bar{x})} \overline{l}(h) \\ &= \max_{\overline{l} \in -\overline{\partial} f(\bar{x})} \overline{l}(h) \text{ for all } h \in X. \end{aligned} \tag{3.24}$$

Now assume that there is some $l \in -\overline{\partial} f(\bar{x})$ with the property $l \notin \underline{\partial} f(\bar{x})$. Since the subdifferential $\underline{\partial} f(\bar{x})$ is convex and weak*-compact, by a separation theorem (Theorem C.3) there is a weak*-continuous linear functional x^{**} on X^* with

$$x^{**}(l) > \sup_{\underline{l} \in \underline{\partial} f(\bar{x})} x^{**}(\underline{l}). \tag{3.25}$$

Every weak*-continuous linear functional on X^* is a point functional. In our special case this means that there is some $h \in X$ with

$$x^{**}(\tilde{l}) = \tilde{l}(h) \text{ for all } \tilde{l} \in X^*.$$

Then it follows from the inequality (3.25)

$$\max_{\bar{l} \in -\overline{\partial} f(\bar{x})} \bar{l}(h) \geq l(h) > \max_{\underline{l} \in \underline{\partial} f(\bar{x})} \underline{l}(h)$$

which is a contradiction to the inequality (3.24). Hence our assumption is not true and we have $-\overline{\partial} f(\bar{x}) \subset \underline{\partial} f(\bar{x})$. □

3.5 Clarke Derivative

An interesting extension of the concept of the directional derivative for real-valued mappings was introduced by Clarke[5]. This section presents a short discussion of this notion of a derivative. A simple necessary optimality condition is also given.

Definition 3.38 (Clarke derivative).

Let S be a nonempty subset of a real normed space $(X, \|\cdot\|)$, let $f : S \to \mathbb{R}$ be a given functional, and let two elements $\bar{x} \in S$ and $h \in X$ be given. If the limit superior

$$f'(\bar{x})(h) = \limsup_{\substack{x \to \bar{x} \\ \lambda \to 0_+}} \frac{1}{\lambda} (f(x + \lambda h) - f(x))$$

exists, then $f'(\bar{x})(h)$ is called the *Clarke derivative* of f at $\bar{x}$ in the direction h. If this limit superior exists for all $h \in X$, then f is called *Clarke differentiable* at $\bar{x}$.

The difference between the Clarke derivative and the directional derivative is based on the fact that for the Clarke derivative the limit superior has to be determined and the base element x of the difference quotient has to be varied.

In this section we see that the Clarke derivative has interesting properties. But it has also the disadvantage that this derivative describes a functional only "cumulatively".

[5]F.H. Clarke, "Generalized gradients and applications", *Trans. Amer. Math. Soc.* 205 (1975) 247–262.

Notice that for the Clarke derivative the limit superior is considered only for those $x \in X$ and $\lambda > 0$ for which $x \in S$ and $x + \lambda h \in S$. There are no difficulties, for instance, if $\bar{x}$ belongs to the interior of the set S. But other types of sets are possible, too.

Example 3.39 (Clarke derivative).

For the absolute value function $f : \mathbb{R} \to \mathbb{R}$ with

$$f(x) = |x| \text{ for all } x \in \mathbb{R}$$

the Clarke derivative at 0 reads for every $h \in \mathbb{R}$

$$f'(0)(h) = \limsup_{\substack{x \to 0 \\ \lambda \to 0_+}} \frac{1}{\lambda} (|x + \lambda h| - |x|) = |h|.$$

In order to see this result, notice that we get with the aid of the triangle inequality

$$\begin{aligned} f'(0)(h) &= \limsup_{\substack{x \to 0 \\ \lambda \to 0_+}} \frac{1}{\lambda} (|x + \lambda h| - |x|) \\ &\leq \limsup_{\substack{x \to 0 \\ \lambda \to 0_+}} \frac{1}{\lambda} (|x| + \lambda |h| - |x|) \\ &= |h|. \end{aligned}$$

For $x = \lambda h$ we obtain

$$\begin{aligned} f'(0)(h) &= \limsup_{\substack{x \to 0 \\ \lambda \to 0_+}} \frac{1}{\lambda} (|x + \lambda h| - |x|) \\ &\geq \limsup_{\lambda \to 0_+} \frac{1}{\lambda} (2\lambda |h| - \lambda |h|) \\ &= |h|. \end{aligned}$$

Hence we have $f'(0)(h) = |h|$.

The class of locally Lipschitz continuous functionals is already differentiable in the sense of Clarke.

Theorem 3.40 (Clarke differentiable functional).

Let S be a subset of a real normed space $(X, \|\cdot\|)$ with nonempty interior, let $\bar{x} \in int(S)$ be a given element, and let $f : S \to \mathbb{R}$ be a functional which is Lipschitz continuous at $\bar{x}$ with a Lipschitz constant k. Then f is Clarke differentiable at $\bar{x}$ and

$$|f'(\bar{x})(h)| \le k\|h\| \text{ for all } h \in X.$$

Proof For an arbitrary $h \in X$ we obtain for the absolute value of the difference quotient in the expression for $f'(\bar{x})(h)$

$$\left|\frac{1}{\lambda}\left(f(x+\lambda h) - f(x)\right)\right| \le \frac{1}{\lambda} k \|x + \lambda h - x\| = k\|h\|,$$

if x is sufficiently close to $\bar{x}$ and λ is sufficiently close to 0. Because of this boundedness the limit superior $f'(\bar{x})(h)$ exists. Furthermore we have

$$\begin{aligned} |f'(\bar{x})(h)| &= \left|\limsup_{\substack{x \to \bar{x} \\ \lambda \to 0_+}} \frac{1}{\lambda}\left(f(x+\lambda h) - f(x)\right)\right| \\ &\le \limsup_{\substack{x \to \bar{x} \\ \lambda \to 0_+}} \left|\frac{1}{\lambda}\left(f(x+\lambda h) - f(x)\right)\right| \\ &\le k\|h\| \end{aligned}$$

which is to prove. □

The assumption in the preceding theorem that $\bar{x}$ belongs to the interior of the set S can be weakened essentially. But then Theorem 3.40 becomes more technical.

Clarke derivatives have the interesting property to be sublinear with respect to the direction h.

Theorem 3.41 (sublinearity of the Clarke derivative).

Let S be a subset of a real normed space $(X, \|\cdot\|)$ with nonempty interior, let $\bar{x} \in int(S)$ be a given element, and let $f : S \to \mathbb{R}$ be a functional which is Clarke differentiable at $\bar{x}$. Then the Clarke derivative $f'(\bar{x})$ is a sublinear functional.

Proof For the proof of the positive homogenity of $f'(\bar{x})$ notice that $f'(\bar{x})(0_X) = 0$ and that for arbitrary $h \in X$ and $\alpha > 0$

$$\begin{aligned} f'(\bar{x})(\alpha h) &= \limsup_{\substack{x \to \bar{x} \\ \lambda \to 0_+}} \frac{1}{\lambda} \left(f(x + \lambda \alpha h) - f(x) \right) \\ &= \alpha \limsup_{\substack{x \to \bar{x} \\ \lambda \to 0_+}} \frac{1}{\lambda \alpha} \left(f(x + \lambda \alpha h) - f(x) \right) \\ &= \alpha f'(\bar{x})(h). \end{aligned}$$

Next we prove the subadditivity of $f'(\bar{x})$. For arbitrary $h_1, h_2 \in X$ we get

$$\begin{aligned} & f'(\bar{x})(h_1 + h_2) \\ &= \limsup_{\substack{x \to \bar{x} \\ \lambda \to 0_+}} \frac{1}{\lambda} \left(f(x + \lambda(h_1 + h_2)) - f(x) \right) \\ &= \limsup_{\substack{x \to \bar{x} \\ \lambda \to 0_+}} \frac{1}{\lambda} \left(f(x + \lambda h_1 + \lambda h_2) - f(x + \lambda h_2) + f(x + \lambda h_2) - f(x) \right) \\ &\leq \limsup_{\substack{x \to \bar{x} \\ \lambda \to 0_+}} \frac{1}{\lambda} \left(f(x + \lambda h_2 + \lambda h_1) - f(x + \lambda h_2) \right) \\ &\quad + \limsup_{\substack{x \to \bar{x} \\ \lambda \to 0_+}} \frac{1}{\lambda} \left(f(x + \lambda h_2) - f(x) \right) \\ &= f'(\bar{x})(h_1) + f'(\bar{x})(h_2). \end{aligned}$$

Consequently, $f'(\bar{x})$ is sublinear. □

In the case of a locally Lipschitz continuous convex functional the directional derivative and the Clarke derivative coincide.

Theorem 3.42 (directional and Clarke derivative).

Let $(X, \|\cdot\|)$ be a real normed space, and let $f : X \to \mathbb{R}$ be a convex functional which is Lipschitz continuous at some $\bar{x} \in X$. Then the directional derivative of f at $\bar{x}$ coincides with the Clarke derivative of f at $\bar{x}$.

Proof Let $h \in X$ denote an arbitrary direction. By Theorems 3.4 and 3.40 the directional derivative $f'(\bar{x})(h)$ and the Clarke derivative $f^0(\bar{x})(h)$ of f at $\bar{x}$ in the direction h exist. By the definition of these derivatives it follows immediately

$$f'(\bar{x})(h) \leq f^0(\bar{x})(h).$$

For the proof of the converse inequality we write

$$\begin{aligned} f^0(\bar{x})(h) &= \limsup_{\substack{x \to \bar{x} \\ \lambda \to 0_+}} \frac{1}{\lambda}\,(f(x+\lambda h) - f(x)) \\ &= \lim_{\substack{\delta \to 0_+ \\ \varepsilon \to 0_+}} \sup_{\|x-\bar{x}\|<\delta} \sup_{0<\lambda<\varepsilon} \frac{1}{\lambda}\,(f(x+\lambda h) - f(x)) . \end{aligned}$$

Since f is convex, Lemma 3.3 leads to the equality

$$f^0(\bar{x})(h) = \lim_{\substack{\delta \to 0_+ \\ \varepsilon \to 0_+}} \sup_{\|x-\bar{x}\|<\delta} \frac{1}{\varepsilon}\,(f(x+\varepsilon h) - f(x)) ,$$

and (since this limit exists, one can choose a special sequence, i.e.) for an arbitrary $\alpha > 0$ we obtain

$$f^0(\bar{x})(h) = \lim_{\varepsilon \to 0_+} \sup_{\|x-\bar{x}\|<\varepsilon\alpha} \frac{1}{\varepsilon}\,(f(x+\varepsilon h) - f(x)) .$$

If we notice that because of the local Lipschitz continuity of f we have for sufficiently small $\varepsilon > 0$

$$\begin{aligned} &\left| \frac{1}{\varepsilon}\,(f(x+\varepsilon h) - f(x)) - \frac{1}{\varepsilon}\,(f(\bar{x}+\varepsilon h) - f(\bar{x})) \right| \\ &\quad \leq \frac{1}{\varepsilon}\,|f(x+\varepsilon h) - f(\bar{x}+\varepsilon h)| + \frac{1}{\varepsilon}\,|f(x) - f(\bar{x})| \\ &\quad \leq \frac{k}{\varepsilon}\,\|x-\bar{x}\| + \frac{k}{\varepsilon}\,\|x-\bar{x}\| \\ &\quad \leq 2k\alpha \end{aligned}$$

($k \geq 0$ denotes a Lipschitz constant), then it follows

$$f^0(\bar{x})(h) \leq \lim_{\varepsilon \to 0_+} \frac{1}{\varepsilon}\,(f(\bar{x}+\varepsilon h) - f(\bar{x})) + 2k\alpha = f'(\bar{x})(h) + 2k\alpha.$$

Since $\alpha > 0$ has been chosen arbitrarily, we obtain

$$f^0(\bar{x})(h) \leq f'(\bar{x})(h).$$

This completes the proof. □

With the aid of the Clarke derivative it is possible to introduce a so-called generalized gradient for locally Lipschitz continuous functionals.

Definition 3.43 (generalized gradient).

Let S be a subset of a real normed space $(X, \|\cdot\|)$ with nonempty interior, and let $f : S \to \mathbb{R}$ be a functional which is Lipschitz continuous at some $\bar{x} \in \text{int}(S)$. Then the set $\partial_{Cl} f(\bar{x})$ of all continuous linear functionals l on X with

$$f'(\bar{x})(h) \geq l(h) \text{ for all } h \in X$$

is called the *generalized gradient* of f at $\bar{x}$ (where $f'(\bar{x})(h)$ denotes the Clarke derivative of f at $\bar{x}$ in the direction h).

For functionals defined on the whole space, notice the formal analogy of the definition of the generalized gradient and the equivalent definition of the subdifferential from Lemma 3.25. The formal difference lies in the fact that one uses the directional derivative for the subdifferential whereas one works with the Clarke derivative for the generalized gradient.

The next result follows immediately from Theorem 3.42 and Lemma 3.25.

Corollary 3.44 (subdifferential and generalized gradient).

Let $(X, \|\cdot\|)$ be a real normed space, and let $f : X \to \mathbb{R}$ be a convex functional which is Lipschitz continuous at some $\bar{x} \in X$. Then the subdifferential $\partial f(\bar{x})$ of f at $\bar{x}$ coincides with the generalized gradient $\partial_{Cl} f(\bar{x})$ of f at $\bar{x}$.

With the following theorem we show that locally Lipschitz continuous functionals have a nonempty generalized gradient.

Theorem 3.45 (nonemptiness of the generalized gradient).

Let S be a subset of a real normed space $(X, \|\cdot\|)$ with nonempty interior, and let $f : S \to \mathbb{R}$ be a given functional. If f is Lipschitz continuous at some $\bar{x} \in \text{int}(S)$, then the generalized gradient $\partial_{Cl} f(\bar{x})$ of f at $\bar{x}$ is nonempty.

Proof By Theorem 3.40 the Clarke derivative exists and by Theorem 3.41 it is sublinear. Consequently, by the basic version of the Hahn-Banach theorem (compare Theorem C.1) there is a linear functional l on X which satisfies the inequality

$$f'(\bar{x})(h) \geq l(h) \text{ for all } h \in X. \tag{3.26}$$

For the proof of the continuity of l we choose an arbitrary $h \in X$. Then it follows from the inequality (3.26) and Theorem 3.40

$$l(h) \leq f'(\bar{x})(h) \leq |f'(\bar{x})(h)| \leq k\|h\|$$

(where $k \geq 0$ denotes a Lipschitz constant) and

$$-l(h) = l(-h) \leq f'(\bar{x})(-h) \leq |f'(\bar{x})(-h)| \leq k\| - h\| = k\|h\|.$$

This leads to the inequality

$$|l(h)| \leq k\|h\|.$$

Hence l is continuous at 0_X. Because of the linearity of l the functional l is also continuous on X. This completes the proof. □

It is also possible to derive a necessary optimality condition for Clarke differentiable functionals. This condition is given in the next theorem.

Theorem 3.46 (necessary optimality condition).

Let T be a superset of a nonempty subset S of a real normed space $(X, \|\cdot\|)$, let $f : T \to \mathbb{R}$ be a given functional, and let T have a nonempty interior. If $\bar{x} \in S \cap \text{int}(T)$ is a minimal point of f on S, the set S is starshaped with respect to $\bar{x}$ and the functional f is Lipschitz continuous at $\bar{x}$, then the following inequality holds for the Clarke derivative

$$f'(\bar{x})(x - \bar{x}) \geq 0 \textit{ for all } x \in S.$$

Proof Let $\bar{x} \in S$ be a minimal point of f on S. Since $\bar{x} \in \text{int}(T)$ and f is Lipschitz continuous at $\bar{x}$, we have for an arbitrary $x \in S$

$$\begin{aligned}\left|\frac{1}{\lambda}(f(\bar{x} + \lambda(x - \bar{x})) - f(\bar{x}))\right| &\leq \frac{k}{\lambda}\|\lambda(x - \bar{x})\| \\ &= k\|x - \bar{x}\| \quad \text{for sufficiently small } \lambda > 0.\end{aligned}$$

Consequently the expression

$$\limsup_{\lambda \to 0_+} \frac{1}{\lambda}(f(\bar{x} + \lambda(x - \bar{x})) - f(\bar{x}))$$

exists. Because of the minimality of f at $\bar{x}$ and the starshapedness of S with respect to $\bar{x}$ this limit superior is nonnegative. Then we conclude

$$\begin{aligned} 0 &\leq \limsup_{\lambda \to 0_+} \frac{1}{\lambda}(f(\bar{x} + \lambda(x - \bar{x})) - f(\bar{x})) \\ &\leq \limsup_{\substack{y \to \bar{x} \\ \lambda \to 0_+}} \frac{1}{\lambda}(f(y + \lambda(x - \bar{x})) - f(y)) \\ &= f'(\bar{x})(x - \bar{x}) \end{aligned}$$

which completes the proof. □

If $(X, \|\cdot\|)$ is a real normed space and $f : X \to \mathbb{R}$ is a given functional, then in the case of $S = X$ the assertion of Theorem 3.46 can also be interpreted as follows: If $\bar{x} \in X$ is a minimal point of f on X, then the functional 0_{X^*} is an element of the generalized gradient of f at $\bar{x}$.

Exercises

(3.1) For the function $f : \mathbb{R} \to \mathbb{R}$ with

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

determine the directional derivative at $\bar{x} = 0$.

(3.2) Let M be a compact subset of $\mathbb{R}^n$, and let $C(M)$ denote the linear space of continuous real-valued functions on M equipped with the maximum norm $\|\cdot\|$ where

$$\|x\| = \max_{t \in M} |x(t)| \text{ for all } x \in C(M).$$

To a given function $\hat{x} \in C(M)$ we consider a functional $f : C(M) \to \mathbb{R}$ with

$$f(x) = \|x - \hat{x}\| \text{ for all } x \in C(M).$$

Show that the directional derivative of f at an arbitrary $\bar{x} \in C(M)$ is given as

$$f'(\bar{x})(h) = \begin{cases} \max\limits_{t \in M(\bar{x})} \operatorname{sgn}(\bar{x}(t) - \hat{x}(t))h(t) & \text{if } \bar{x} \neq \hat{x} \\ \max\limits_{t \in M(\bar{x})} |h(t)| & \text{if } \bar{x} = \hat{x} \end{cases}$$

with

$$M(\bar{x}) := \{t \in M \mid |\bar{x}(t) - \hat{x}(t)| = f(\bar{x})\}.$$

(3.3) Let $(X, \|\cdot\|)$ be a real normed space, and let $f : X \to \mathbb{R}$ be a convex functional which is Gâteaux differentiable at some $\bar{x} \in X$. Prove that $\bar{x}$ is a minimal point of f on X if and only if $f'(\bar{x}) = 0_{X^*}$.

(3.4) For the function $f : \mathbb{R} \to \mathbb{R}$ with

$$f(x) = |x| \text{ for all } x \in \mathbb{R}$$

determine the subdifferential $\partial f(0)$.

(3.5) Let $(X, \|\cdot\|)$ be a real normed space, and let $f : X \to \mathbb{R}$ be a convex functional. Show: For an arbitrary $\bar{x} \in X$ the subdifferential $\partial f(\bar{x})$ is a convex set.

(3.6) Prove: For every convex function $f : \mathbb{R}^n \to \mathbb{R}$ which is differentiable at some $\bar{x} \in \mathbb{R}^n$ it follows $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$.

(3.7) Let the function $f : \mathbb{R}^2 \to \mathbb{R}$ with

$$f(x_1, x_2) = |x_1 x_2| \text{ for all } (x_1, x_2) \in \mathbb{R}^2$$

be given. Determine a quasidifferential of f at an arbitrary point $(x_1, x_2) \in \mathbb{R}^2$.

(3.8) Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ with

$$f(x_1, x_2) = \begin{cases} |x_1| - |x_2| + \frac{|x_1^3 x_2|}{x_1^2 + x_2^2} & \text{if } (x_1, x_2) \neq (0, 0) \\ 0 & \text{if } (x_1, x_2) = (0, 0) \end{cases}.$$

Show that f is quasidifferentiable at $\bar{x} := (0, 0)$.

(3.9) Let the function $f : \mathbb{R}^n \to \mathbb{R}$ with

$$f(x_1, \ldots, x_n) = \max\{x_1, \ldots, x_n\} \text{ for all } x_1, \ldots, x_n \in \mathbb{R}$$

be given. For an arbitrary $\bar{x} \in \mathbb{R}^n$ let

$$I(\bar{x}) := \{i \in \{1, \ldots, n\} \mid f(\bar{x}) = \bar{x}_i\}.$$

Show that the Clarke derivative of f at $\bar{x}$ in an arbitrary direction $h \in \mathbb{R}^n$ is given as

$$f'(\bar{x})(h) = \max_{i \in I(\bar{x})} \{h_i\}.$$

Tangent Cones

4

In this chapter certain approximations of sets are considered which are very useful for the formulation of optimality conditions. We investigate so-called tangent cones which approximate a given set in a local sense. First, we discuss several basic properties of tangent cones, and then we present optimality conditions with the aid of these cones. Finally, we formulate a Lyusternik theorem.

4.1 Definition and Properties

In this section we turn our attention to the sequential Bouligand tangent cone which is also called the contingent cone. For this tangent cone we prove several basic properties.

First, we introduce the concept of a cone.

Definition 4.1 (cone).

Let C be a nonempty subset of a real linear space X.

(a) The set C is called a *cone* if

$$x \in C,\ \lambda \geq 0 \implies \lambda x \in C$$

(compare Fig. 4.1).

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3_4

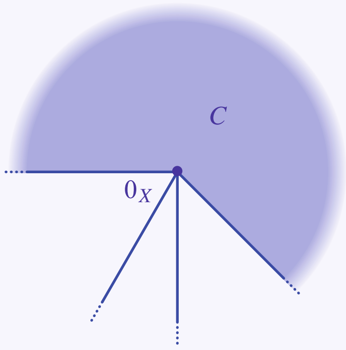

Fig. 4.1 Cone

(b) A cone C is called *pointed* if

$$x \in C,\ -x \in C \implies x = 0_X$$

(compare Fig. 4.2).

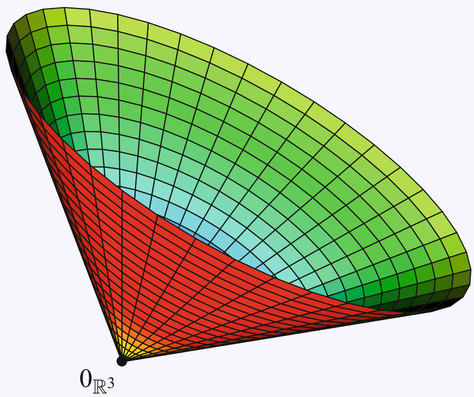

Fig. 4.2 Pointed cone

Example 4.2 (pointed cone).

(a) The set

$$\mathbb{R}^n_+ := \{x \in \mathbb{R}^n \mid x_i \geq 0 \text{ for all } i \in \{1, \ldots, n\}\}$$

is a pointed cone.

(b) The set

$$C := \{x \in C[0, 1] \mid x(t) \geq 0 \text{ for all } t \in [0, 1]\}$$

is a pointed cone.

In order theory and optimization theory convex cones are of special interest. Such cones may be characterized as follows:

Theorem 4.3 (convex cone).

A cone C in a real linear space is convex if and only if for all $x, y \in C$

$$x + y \in C. \tag{4.1}$$

Proof

(a) Let C be a convex cone. Then it follows for all $x, y \in C$

$$\frac{1}{2}(x + y) = \frac{1}{2}x + \frac{1}{2}y \in C$$

which implies $x + y \in C$.

(b) For arbitrary $x, y \in C$ and $\lambda \in [0, 1]$ we have $\lambda x \in C$ and $(1 - \lambda)y \in C$. Then we get with the condition (4.1) $\lambda x + (1 - \lambda)y \in C$. Consequently, the cone C is convex. □

In the sequel we also define cones generated by sets.

Definition 4.4 (cone generated by a set).

Let S be a nonempty subset of a real linear space. The set

$$\text{cone}(S) := \{\lambda s \mid \lambda \geq 0 \text{ and } s \in S\}$$

is called the *cone generated* by S (compare Fig. 4.3).

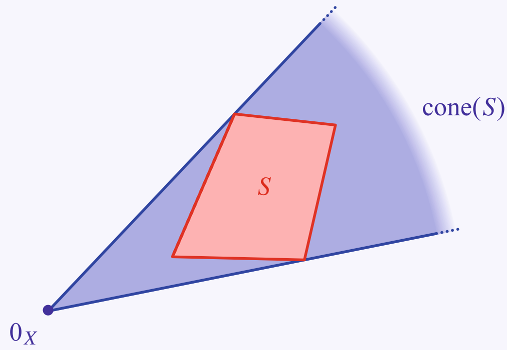

Fig. 4.3 Cone generated by S

Example 4.5 (cone generated by a set).

(a) Let $B(0_X, 1)$ denote the closed unit ball in a real normed space $(X, \|\cdot\|)$. Then the cone generated by $B(0_X, 1)$ equals the linear space X.

(b) Let S denote the graph of the function $f : \mathbb{R} \to \mathbb{R}$ with

$$f(x) = \begin{cases} x \sin \frac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

Then the cone generated by S is given as

$$\text{cone}(S) = \{(x, y) \in \mathbb{R}^2 \mid |y| \leq |x|\}$$

(see Fig. 4.4).

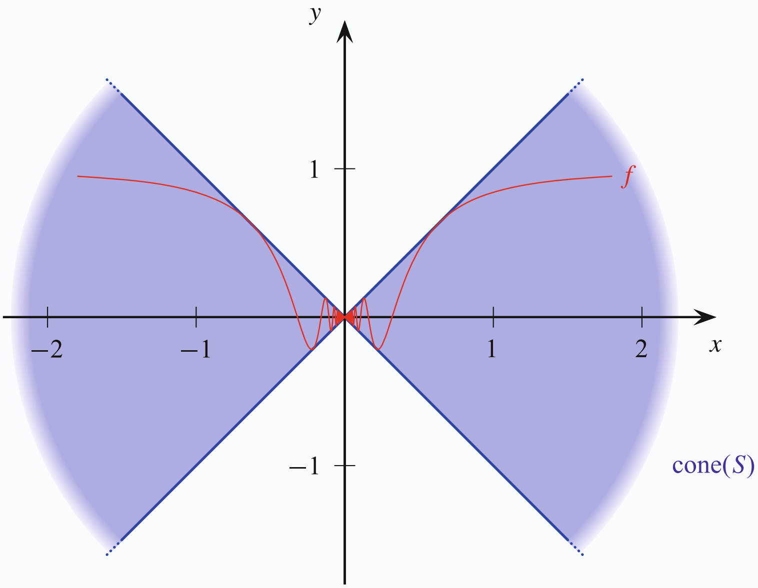

Fig. 4.4 Illustration of cone(S)

Now we turn our attention to tangent cones.

Definition 4.6 (contingent cone).

Let S be a nonempty subset of a real normed space $(X, \|\cdot\|)$.

(a) Let $\bar{x} \in \text{cl}(S)$ be a given element. A vector $h \in X$ is called a *tangent vector* to S at $\bar{x}$, if there are a sequence $(x_n)_{n\in\mathbb{N}}$ of elements in S and a sequence $(\lambda_n)_{n\in\mathbb{N}}$ of positive real numbers with

$$\bar{x} = \lim_{n\to\infty} x_n$$

and

$$h = \lim_{n\to\infty} \lambda_n (x_n - \bar{x}).$$

(b) The set $T(S, \bar{x})$ of all tangent vectors to S at $\bar{x}$ is called *sequential Bouligand tangent cone* to S at $\bar{x}$ or *contingent cone* to S at $\bar{x}$ (compare Fig. 4.5).

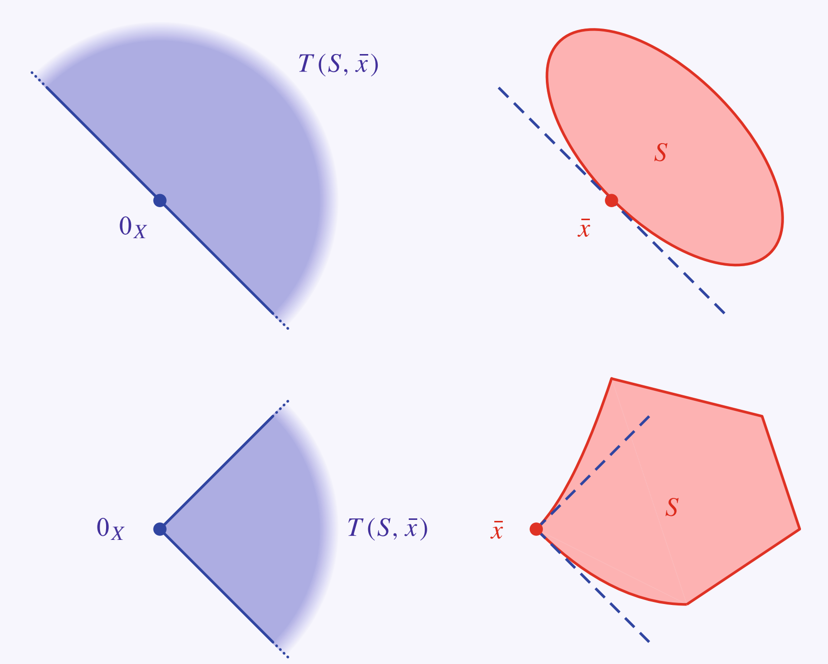

Fig. 4.5 Two examples of contingent cones

The contingent cone is due to Bouligand[6] and Severi. Notice that $\bar{x}$ needs only to belong to the closure of the set S in the definition of $T(S,\bar{x})$. But later we will assume that $\bar{x} \in S$.

By the definition of tangent vectors it follows immediately that the contingent cone is in fact a cone.

Before investigating the contingent cone we briefly present the definition of the Clarke tangent cone which is not used any further in this chapter.

Remark 4.7 (Clarke tangent cone).

Let $\bar{x}$ be an element of the closure of a nonempty subset S of a real normed space $(X, \|\cdot\|)$.

[6]M.G. Bouligand, "Sur les surfaces dépourvues de points hyperlimites (ou: un thèorème d'existence du plan tangent)", *Ann. Soc. Polon. Math.* 9 (1930) 32–41.

F. Severi remarked that he has independently introduced this notion (F. Severi, "Su alcune questioni di topologia infinitesimale", *Ann. Soc. Polon. Math.* 9 (1930) 97–108).

(a) The set

$$T_{Cl}(S,\bar{x}) := \{h \in X \mid \text{ for every sequence } (x_n)_{n\in\mathbb{N}} \text{ of elements of } S \text{ with } \bar{x} = \lim_{n\to\infty} x_n \text{ and for every sequence } (\lambda_n)_{n\in\mathbb{N}} \text{ of positive real numbers converging to 0 there is a sequence } (h_n)_{n\in\mathbb{N}} \text{ with } h = \lim_{n\to\infty} h_n \text{ and } x_n + \lambda_n h_n \in S \text{ for all } n \in \mathbb{N}\}$$

is called *(sequential) Clarke tangent cone* to S at $\bar{x}$.

(b) It is evident that the Clarke tangent cone $T_{Cl}(S,\bar{x})$ is always a cone.

(c) If $\bar{x} \in S$, then the Clarke tangent cone $T_{Cl}(S,\bar{x})$ is contained in the contingent cone $T(S,\bar{x})$.
For the proof of this assertion let some $h \in T_{Cl}(S,\bar{x})$ be given arbitrarily. Then we choose the special sequence $(\bar{x})_{n\in\mathbb{N}}$ and an arbitrary sequence $(\lambda_n)_{n\in\mathbb{N}}$ of positive real numbers converging to 0. Consequently, there is a sequence $(h_n)_{n\in\mathbb{N}}$ with $h = \lim_{n\to\infty} h_n$ and $\bar{x}+\lambda_n h_n \in S$ for all $n \in \mathbb{N}$. Now we set

$$y_n := \bar{x} + \lambda_n h_n \text{ for all } n \in \mathbb{N}$$

and

$$t_n := \frac{1}{\lambda_n} \text{ for all } n \in \mathbb{N}.$$

Then it follows

$$y_n \in S \text{ for all } n \in \mathbb{N},$$

$$\lim_{n\to\infty} y_n = \lim_{n\to\infty} (\bar{x} + \lambda_n h_n) = \bar{x}$$

and

$$\lim_{n\to\infty} t_n(y_n - \bar{x}) = \lim_{n\to\infty} \frac{1}{\lambda_n}(\bar{x} + \lambda_n h_n - \bar{x}) = \lim_{n\to\infty} h_n = h.$$

Consequently, h is a tangent vector. □

This result is illustrated in Fig. 4.6.

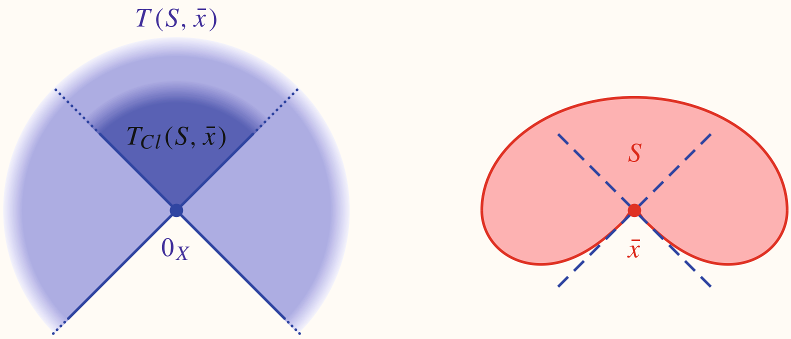

Fig. 4.6 Illustration of the result in Remark 4.7,(c)

(d) The Clarke tangent cone $T_{Cl}(S, \bar{x})$ is always a closed convex cone. We mention this result without proof. Notice that this assertion is true without any assumption on the set S.

Next, we come back to the contingent cone and we investigate the relationship between the contingent cone $T(S, \bar{x})$ and the cone generated by $S - \{\bar{x}\}$.

Theorem 4.8 (subset of a contingent cone).

Let S be a nonempty subset of a real normed space. If S is starshaped with respect to some $\bar{x} \in S$, then it follows

$$cone(S - \{\bar{x}\}) \subset T(S, \bar{x}).$$

Proof Let the set S be starshaped with respect to some $\bar{x} \in S$, and let an arbitrary element $x \in S$ be given. Then we define a sequence $(x_n)_{n\in\mathbb{N}}$ with

$$x_n := \bar{x} + \frac{1}{n}(x - \bar{x}) = \frac{1}{n}x + \left(1 - \frac{1}{n}\right)\bar{x} \in S \text{ for all } n \in \mathbb{N}.$$

For this sequence we have

$$\lim_{n\to\infty} x_n = \bar{x}$$

and

$$\lim_{n\to\infty} n(x_n - \bar{x}) = x - \bar{x}.$$

Consequently, $x - \bar{x}$ is a tangent vector, and we obtain

$$S - \{\bar{x}\} \subset T(S, \bar{x}).$$

Since $T(S, \bar{x})$ is a cone, we conclude

$$\text{cone}(S - \{\bar{x}\}) \subset \text{cone}(T(S, \bar{x})) = T(S, \bar{x}). \qquad \square$$

Theorem 4.9 (superset of a contingent cone).

Let S be a nonempty subset of a real normed space. For every $\bar{x} \in S$ it follows

$$T(S, \bar{x}) \subset cl(cone(S - \{\bar{x}\})).$$

Proof We fix an arbitrary $\bar{x} \in S$ and we choose any $h \in T(S, \bar{x})$. Then there are a sequence $(x_n)_{n \in \mathbb{N}}$ of elements in S and a sequence $(\lambda_n)_{n \in \mathbb{N}}$ of positive real numbers with $\bar{x} = \lim_{n \to \infty} x_n$ and $h = \lim_{n \to \infty} \lambda_n (x_n - \bar{x})$. The last equation implies

$$h \in \text{cl}(\text{cone}(S - \{\bar{x}\}))$$

which has to be shown. $\square$

By the two preceding theorems we obtain the following inclusion chain for a set S which is starshaped with respect to some $\bar{x} \in S$:

$$\text{cone}(S - \{\bar{x}\}) \;\subset\; T(S, \bar{x}) \;\subset\; \text{cl}(\text{cone}(S - \{\bar{x}\})). \tag{4.2}$$

The next theorem says that the contingent cone is always closed.

Theorem 4.10 (closedness of a contingent cone).

Let S be a nonempty subset of a real normed space $(X, \|\cdot\|)$. For every $\bar{x} \in S$ the contingent cone $T(S, \bar{x})$ is closed.

Proof Let $\bar{x} \in S$ be arbitrarily chosen, and let $(h_n)_{n \in \mathbb{N}}$ be an arbitrary sequence of tangent vectors to S at $\bar{x}$ with $\lim_{n \to \infty} h_n = h \in X$. For every tangent vector h_n there are a sequence $(x_{n_i})_{i \in \mathbb{N}}$ of elements in S and a sequence $(\lambda_{n_i})_{i \in \mathbb{N}}$ of positive real numbers with $\bar{x} = \lim_{i \to \infty} x_{n_i}$ and $h_n = \lim_{i \to \infty} \lambda_{n_i} (x_{n_i} - \bar{x})$. Consequently, for every

$n \in \mathbb{N}$ there is a number $i(n) \in \mathbb{N}$ with

$$\|x_{n_i} - \bar{x}\| \leq \frac{1}{n} \text{ for all } i \in \mathbb{N} \text{ with } i \geq i(n)$$

and

$$\|\lambda_{n_i}(x_{n_i} - \bar{x}) - h_n\| \leq \frac{1}{n} \text{ for all } i \in \mathbb{N} \text{ with } i \geq i(n).$$

If we define the sequences $(y_n)_{n\in\mathbb{N}}$ and $(\mu_n)_{n\in\mathbb{N}}$ by

$$y_n := x_{n_{i(n)}} \in S \text{ for all } n \in \mathbb{N}$$

and

$$\mu_n := \lambda_{n_{i(n)}} > 0 \text{ for all } n \in \mathbb{N},$$

then we obtain $\lim\limits_{n\to\infty} y_n = \bar{x}$ and

$$\begin{aligned}\|\mu_n(y_n - \bar{x}) - h\| &= \|\lambda_{n_{i(n)}}(x_{n_{i(n)}} - \bar{x}) - h_n + h_n - h\| \\ &\leq \frac{1}{n} + \|h_n - h\| \text{ for all } n \in \mathbb{N}.\end{aligned}$$

Hence we have

$$h = \lim_{n\to\infty} \mu_n(y_n - \bar{x})$$

and h is a tangent vector to S at $\bar{x}$. □

Since the inclusion chain (4.2) is also valid for the corresponding closed sets, it follows immediately with the aid of Theorem 4.10:

Corollary 4.11 (characterization of a contingent cone).

Let S be a nonempty subset of a real normed space. If the set S is starshaped with respect to some $\bar{x} \in S$, then it is

$$T(S, \bar{x}) = cl(cone(S - \{\bar{x}\})).$$

If the set S is starshaped with respect to some $\bar{x} \in S$, then Corollary 4.11 says essentially that for the determination of the contingent cone to S at $\bar{x}$ we have to consider only rays emanating from $\bar{x}$ and passing through S.

Finally, we show that the contingent cone to a nonempty convex set is also convex.

Theorem 4.12 (convexity of a contingent cone).

If S is a nonempty convex subset of a real normed space $(X, \|\cdot\|)$, then the contingent cone $T(S,\bar{x})$ is convex for all $\bar{x} \in S$.

Proof We choose an arbitrary $\bar{x} \in S$ and we fix two arbitrary tangent vectors $h_1, h_2 \in T(S,\bar{x})$ with $h_1, h_2 \neq 0_X$. Then there are sequences $(x_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$ of elements in S and sequences $(\lambda_n)_{n\in\mathbb{N}}$, $(\mu_n)_{n\in\mathbb{N}}$ of positive real numbers with

$$\bar{x} = \lim_{n\to\infty} x_n, \;\; h_1 = \lim_{n\to\infty} \lambda_n (x_n - \bar{x})$$

and

$$\bar{x} = \lim_{n\to\infty} y_n, \;\; h_2 = \lim_{n\to\infty} \mu_n (y_n - \bar{x}).$$

Next, we define additional sequences $(\nu_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ with

$$\nu_n := \lambda_n + \mu_n \text{ for all } n \in \mathbb{N}$$

and

$$z_n := \frac{1}{\nu_n}(\lambda_n x_n + \mu_n y_n) \text{ for all } n \in \mathbb{N}.$$

Because of the convexity of S we have

$$z_n = \frac{\lambda_n}{\lambda_n + \mu_n} x_n + \frac{\mu_n}{\lambda_n + \mu_n} y_n \in S \text{ for all } n \in \mathbb{N},$$

and we conclude

$$\begin{aligned}
\lim_{n\to\infty} z_n &= \lim_{n\to\infty} \frac{1}{\nu_n}(\lambda_n x_n + \mu_n y_n) \\
&= \lim_{n\to\infty} \frac{1}{\nu_n}(\lambda_n x_n - \lambda_n \bar{x} + \mu_n y_n - \mu_n \bar{x} + \lambda_n \bar{x} + \mu_n \bar{x}) \\
&= \lim_{n\to\infty} \left(\frac{\lambda_n}{\nu_n}(x_n - \bar{x}) + \frac{\mu_n}{\nu_n}(y_n - \bar{x}) + \bar{x} \right) \\
&= \bar{x}
\end{aligned}$$

and

$$\begin{aligned}\lim_{n\to\infty} \nu_n(z_n - \bar{x}) &= \lim_{n\to\infty} (\lambda_n x_n + \mu_n y_n - \nu_n \bar{x}) \\ &= \lim_{n\to\infty} (\lambda_n (x_n - \bar{x}) + \mu_n (y_n - \bar{x})) \\ &= h_1 + h_2.\end{aligned}$$

Hence it follows $h_1 + h_2 \in T(S, \bar{x})$. Since $T(S, \bar{x})$ is a cone, Theorem 4.3 leads to the assertion. □

Notice that the Clarke tangent cone to an arbitrary nonempty set S is already a convex cone, while we have shown the convexity of the contingent cone only under the assumption of the convexity of S.

4.2 Optimality Conditions

In this section we present several optimality conditions which result from the theory on contingent cones.

First, we show, for example, for convex optimization problems with a continuous objective functional that every minimal point $\bar{x}$ of f on S can be characterized as a minimal point of f on $\{\bar{x}\} + T(S, \bar{x})$.

Theorem 4.13 (optimality condition).

Let S be a nonempty subset of a real normed space $(X, \|\cdot\|)$, and let $f : X \to \mathbb{R}$ be a given functional.

(a) If the functional f is continuous and convex, then for every minimal point $\bar{x} \in S$ of f on S it follows:

$$f(\bar{x}) \leq f(\bar{x} + h) \text{ for all } h \in T(S, \bar{x}), \tag{4.3}$$

i.e., 0_X is a minimal solution of the optimization problem
$$\min_{h \in T(S,\bar{x})} f(\bar{x} + h) \quad \text{(compare Fig. 4.7).}$$

(b) If the set S is starshaped with respect to some $\bar{x} \in S$ and if the inequality (4.3) is satisfied, then $\bar{x}$ is a minimal point of f on S.

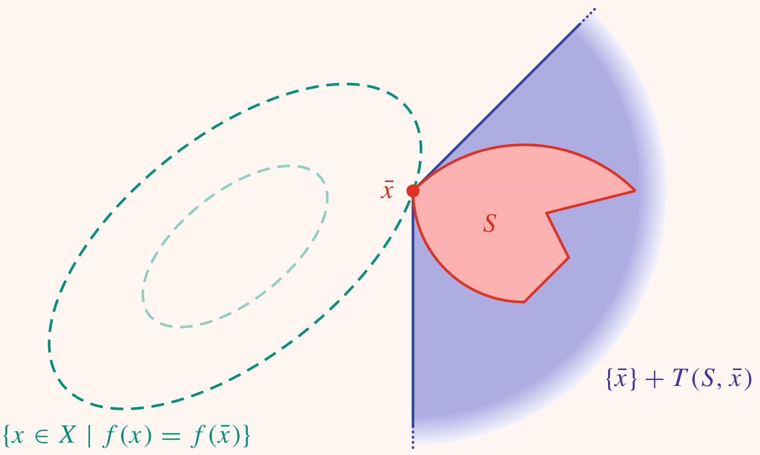

Fig. 4.7 Geometric illustration of the result of Theorem 4.13

Proof

(a) We fix an arbitrary $\bar{x} \in S$ and assume that the inequality (4.3) does not hold. Then there are a vector $h \in T(S, \bar{x}) \setminus \{0_X\}$ and a number $\alpha > 0$ with

$$f(\bar{x}) - f(\bar{x} + h) > \alpha > 0. \tag{4.4}$$

By the definition of h there are a sequence $(x_n)_{n\in\mathbb{N}}$ of elements in S and a sequence $(\lambda_n)_{n\in\mathbb{N}}$ of positive real numbers with

$$\bar{x} = \lim_{n\to\infty} x_n$$

and

$$h = \lim_{n\to\infty} h_n$$

where

$$h_n := \lambda_n(x_n - \bar{x}) \text{ for all } n \in \mathbb{N}.$$

Because of $h \neq 0_X$ we have $\lim_{n\to\infty} \frac{1}{\lambda_n} = 0$. Since f is convex and continuous, we obtain with the inequality (4.4) for sufficiently large $n \in \mathbb{N}$:

$$\begin{aligned} f(x_n) &= f\Big(\frac{1}{\lambda_n}\bar{x} + x_n - \bar{x} + \bar{x} - \frac{1}{\lambda_n}\bar{x}\Big) \\ &= f\Big(\frac{1}{\lambda_n}(\bar{x} + h_n) + \Big(1 - \frac{1}{\lambda_n}\Big)\bar{x}\Big) \\ &\leq \frac{1}{\lambda_n} f(\bar{x} + h_n) + \Big(1 - \frac{1}{\lambda_n}\Big) f(\bar{x}) \\ &\leq \frac{1}{\lambda_n}(f(\bar{x} + h) + \alpha) + \Big(1 - \frac{1}{\lambda_n}\Big) f(\bar{x}) \\ &< \frac{1}{\lambda_n} f(\bar{x}) + \Big(1 - \frac{1}{\lambda_n}\Big) f(\bar{x}) \\ &= f(\bar{x}). \end{aligned}$$

Consequently, $\bar{x}$ is not a minimal point of f on S.

(b) If the set S is starshaped with respect to some $\bar{x} \in S$, then it follows by Theorem 4.8

$$S - \{\bar{x}\} \subset T(S, \bar{x}).$$

Hence we get with the inequality (4.3)

$$f(\bar{x}) \leq f(\bar{x} + h) \text{ for all } h \in S - \{\bar{x}\},$$

i.e., $\bar{x}$ is a minimal point of f on S. □

Using Fréchet derivatives the following necessary optimality condition can be formulated.

Theorem 4.14 (necessary optimality condition).

Let S be a nonempty subset of a real normed space $(X, \|\cdot\|)$, and let f be a functional defined on an open superset of S. If $\bar{x} \in S$ is a minimal point of f on S and if f is Fréchet differentiable at $\bar{x}$, then it follows

$$f'(\bar{x})(h) \geq 0 \text{ for all } h \in T(S, \bar{x}),$$

i.e., 0_X is a minimal solution of the optimization problem
$$\min_{h \in T(S,\bar{x})} f'(\bar{x})(h).$$

Proof Let $\bar{x} \in S$ be a minimal point of f on S, and let some $h \in T(S, \bar{x}) \setminus \{0_X\}$ be arbitrarily given (for $h = 0_X$ the assertion is trivial). Then there are a sequence $(x_n)_{n\in\mathbb{N}}$ of elements in S and a sequence $(\lambda_n)_{n\in\mathbb{N}}$ of positive real numbers with $\bar{x} = \lim_{n\to\infty} x_n$ and $h = \lim_{n\to\infty} h_n$ where

$$h_n := \lambda_n(x_n - \bar{x}) \text{ for all } n \in \mathbb{N}.$$

By the definition of the Fréchet derivative and because of the minimality of f at $\bar{x}$ it follows:

$$\begin{aligned}
f'(\bar{x})(h) &= f'(\bar{x})\Big(\lim_{n\to\infty} \lambda_n(x_n - \bar{x})\Big) \\
&= \lim_{n\to\infty} \lambda_n f'(\bar{x})(x_n - \bar{x}) \\
&= \lim_{n\to\infty} \lambda_n[f(x_n) - f(\bar{x}) - (f(x_n) - f(\bar{x}) - f'(\bar{x})(x_n - \bar{x}))] \\
&\geq -\lim_{n\to\infty} \lambda_n(f(x_n) - f(\bar{x}) - f'(\bar{x})(x_n - \bar{x})) \\
&= -\lim_{n\to\infty} \|h_n\| \frac{f(x_n) - f(\bar{x}) - f'(\bar{x})(x_n - \bar{x})}{\|x_n - \bar{x}\|} \\
&= 0.
\end{aligned}$$

Hence, the assertion is proved. □

Next, we investigate under which assumptions the condition in Theorem 4.14 is a sufficient optimality condition. For this purpose we define pseudoconvex functionals.

Definition 4.15 (pseudoconvex functional).

Let S be a nonempty subset of a real linear space, and let $f : S \to \mathbb{R}$ be a given functional which has a directional derivative at some $\bar{x} \in S$ in every direction $x - \bar{x}$ with arbitrary $x \in S$. The functional f is called *pseudoconvex* at $\bar{x}$ if for all $x \in S$

$$f'(\bar{x})(x - \bar{x}) \geq 0 \implies f(x) - f(\bar{x}) \geq 0.$$

Example 4.16 (pseudoconvex functions).

The functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ with

$$f(x) = xe^x \text{ for all } x \in \mathbb{R}$$

and

$$g(x) = -\frac{1}{1+x^2} \text{ for all } x \in \mathbb{R}$$

are pseudoconvex at every $\bar{x} \in \mathbb{R}$. But the two functions are not convex (see Fig. 4.8).

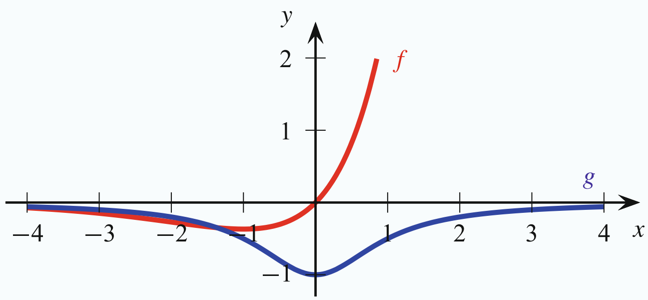

Fig. 4.8 Functions f and g in Example 4.16

A relationship between convex and pseudoconvex functionals is given by the next theorem.

Theorem 4.17 (pseudoconvexity).

Let S be a nonempty convex subset of a real linear space, and let $f : S \to \mathbb{R}$ be a convex functional which has a directional derivative at some $\bar{x} \in S$ in every direction $x - \bar{x}$ with arbitrary $x \in S$. Then f is pseudoconvex at $\bar{x}$.

Proof We fix an arbitrary $x \in S$. Because of the convexity of f we get for all $\lambda \in (0, 1]$

$$f(\lambda x + (1-\lambda)\bar{x}) \leq \lambda f(x) + (1-\lambda) f(\bar{x})$$

and

$$\begin{aligned} f(x) &\geq f(\bar{x}) + \frac{1}{\lambda}(f(\lambda x + (1-\lambda)\bar{x}) - f(\bar{x})) \\ &= f(\bar{x}) + \frac{1}{\lambda}(f(\bar{x} + \lambda(x - \bar{x})) - f(\bar{x})). \end{aligned}$$

Since f has a directional derivative at $\bar{x}$ in the direction $x - \bar{x}$, we conclude

$$f(x) - f(\bar{x}) \geq f'(\bar{x})(x - \bar{x}).$$

Consequently, if $f'(\bar{x})(x - \bar{x}) \geq 0$, then

$$f(x) - f(\bar{x}) \geq 0.$$

Hence f is pseudoconvex at $\bar{x}$. □

It is also possible to formulate a relationship between quasiconvex and pseudoconvex functionals.

Theorem 4.18 (pseudoconvexity and quasiconvexity).

Let S be a nonempty convex subset of a real normed space, and let f be a functional which is defined on an open superset of S. If f is Fréchet differentiable at every $\bar{x} \in S$ and pseudoconvex at every $\bar{x} \in S$, then f is also quasiconvex on S.

Proof Under the given assumptions we prove that for every $\alpha \in \mathbb{R}$ the level set

$$S_\alpha := \{x \in S \mid f(x) \leq \alpha\}$$

is a convex set. For this purpose we fix an arbitrary $\alpha \in \mathbb{R}$ so that S_α is a nonempty set. Furthermore we choose two arbitrary elements $x, y \in S_\alpha$. In the following we assume that there is a $\hat{\lambda} \in [0, 1]$ with

$$f(\hat{\lambda}x + (1 - \hat{\lambda})y) > \alpha \geq \max\{f(x), f(y)\}.$$

Then it follows $\hat{\lambda} \in (0, 1)$. Since f is Fréchet differentiable on S, by Theorem 3.15 f is also continuous on S. Consequently, there is a $\bar{\lambda} \in (0, 1)$ with

$$f(\bar{\lambda}x + (1 - \bar{\lambda})y) \geq f(\lambda x + (1 - \lambda)y) \text{ for all } \lambda \in (0, 1).$$

Using Theorems 3.13 and 3.8, (a) (which is now applied to a maximum problem) it follows for $\bar{x} := \bar{\lambda}x + (1 - \bar{\lambda})y$

$$f'(\bar{x})(x - \bar{x}) \leq 0$$

and

$$f'(\bar{x})(y - \bar{x}) \leq 0.$$

With

$$x - \bar{x} = x - \bar{\lambda}x - (1 - \bar{\lambda})y = (1 - \bar{\lambda})(x - y),$$

$$y - \bar{x} = y - \bar{\lambda}x - (1 - \bar{\lambda})y = -\bar{\lambda}(x - y) \tag{4.5}$$

and the linearity of $f'(\bar{x})$ we obtain

$$0 \geq f'(\bar{x})(x - \bar{x}) = (1 - \bar{\lambda})f'(\bar{x})(x - y)$$

and

$$0 \geq f'(\bar{x})(y - \bar{x}) = -\bar{\lambda}f'(\bar{x})(x - y).$$

Hence we have $0 = f'(\bar{x})(x - y)$, and with the equality (4.5) it also follows $f'(\bar{x})(y-\bar{x}) = 0$. By assumption f is pseudoconvex at $\bar{x}$ and therefore we conclude

$$f(y) - f(\bar{x}) \geq 0.$$

But this inequality contradicts the following inequality:

$$\begin{aligned} f(y) - f(\bar{x}) &= f(y) - f(\bar{\lambda}x + (1 - \bar{\lambda})y) \\ &\leq f(y) - f(\hat{\lambda}x + (1 - \hat{\lambda})y) \\ &< f(y) - \alpha \\ &\leq 0. \end{aligned}$$

□

Using Theorem 3.13 the result of the Theorems 4.17 and 4.18 can be specialized in the following way: If $(X, \|\cdot\|)$ is a real normed space and if $f : X \to \mathbb{R}$ is a functional which is Fréchet differentiable at every $\bar{x} \in X$, then the following implications are satisfied:

$$\begin{aligned} f \text{ convex} &\Longrightarrow f \text{ pseudoconvex at every } \bar{x} \in X \\ &\Longrightarrow f \text{ quasiconvex.} \end{aligned}$$

After these investigations we come back to the question leading to the introduction of pseudoconvex functionals. With the next theorem we present now assumptions under which the condition in Theorem 4.14 is a sufficient optimality condition.

Theorem 4.19 (sufficient optimality condition).

Let S be a nonempty subset of a real normed space, and let f be a functional defined on an open superset of S. If S is starshaped with respect to some $\bar{x} \in S$, if f is directionally differentiable at $\bar{x}$ and pseudoconvex at $\bar{x}$, and if

$$f'(\bar{x})(h) \geq 0 \text{ for all } h \in T(S, \bar{x}),$$

then $\bar{x}$ is a minimal point of f on S.

Proof Because of the starshapedness of S with respect to $\bar{x} \in S$ it follows by Theorem 4.8 $S - \{\bar{x}\} \subset T(S, \bar{x})$, and therefore we have

$$f'(\bar{x})(x - \bar{x}) \geq 0 \text{ for all } x \in S.$$

Since f is pseudoconvex at $\bar{x}$, we conclude

$$f(x) - f(\bar{x}) \geq 0 \text{ for all } x \in S,$$

i.e., $\bar{x}$ is a minimal point of f on S. □

Notice that the assumption in Theorem 3.8,(b) under which the inequality (3.1) is a sufficient condition, can be weakened with the aid of the pseudoconvexity assumption. This result is summarized with Theorem 3.8 in the next corollary.

Corollary 4.20 (necessary and sufficient optimality condition).

Let S be a nonempty subset of a real linear space, and let $f : S \to \mathbb{R}$ be a given functional. Moreover, let the functional f have a directional derivative at some $\bar{x} \in S$ in every direction $x - \bar{x}$ with arbitrary $x \in S$ and let f be pseudoconvex at $\bar{x}$. Then $\bar{x}$ is a minimal point of f on S if and only if

$$f'(\bar{x})(x - \bar{x}) \geq 0 \text{ for all } x \in S.$$

4.3 A Lyusternik Theorem

For the application of the necessary optimality condition given in Theorem 4.14 to optimization problems with equality constraints we need a profound theorem which generalizes a result given by Lyusternik[7] published in 1934. This theorem

[7] L.A. Lyusternik, "Conditional extrema of functionals", *Mat. Sb.* 41 (1934) 390–401.

says under appropriate assumptions that the contingent cone to a set described by equality constraints is a superset of the set of the linearized constraints. Moreover, it can be shown under these assumptions that both sets are equal.

Theorem 4.21 (Lyusternik theorem).

Let $(X, \|\cdot\|_X)$ and $(Z, \|\cdot\|_Z)$ be real Banach spaces, and let $h : X \to Z$ be a given mapping. Furthermore, let some $\bar{x} \in S$ with

$$S := \{x \in X \mid h(x) = 0_Z\}$$

be given. Let h be Fréchet differentiable on a neighborhood of $\bar{x}$, let $h'(\cdot)$ be continuous at $\bar{x}$, and let $h'(\bar{x})$ be surjective. Then it follows

$$\{x \in X \mid h'(\bar{x})(x) = 0_Z\} \subset T(S, \bar{x}). \tag{4.6}$$

Proof We present a proof of Lyusternik's theorem which is put forward by Werner [365]. This proof can be carried out in several steps. First we apply an open mapping theorem and then we prove the technical inequality (4.14). In the third part we show the equations (4.26) and (4.27) with the aid of a construction of special sequences, and based on these equations we get the inclusion (4.6) in the last part.

(1) Since $h'(\bar{x})$ is continuous, linear and surjective by the open mapping theorem the mapping $h'(\bar{x})$ is open, i.e. the image of every open set is open. Therefore, if $B(0_X, 1)$ denotes the open unit ball in X, there is some $\varrho > 0$ such that

$$B(0_Z, \varrho) \subset h'(\bar{x})\, B(0_X, 1) \tag{4.7}$$

where $B(0_Z, \varrho)$ denotes the open ball around 0_Z with radius ϱ. Because of the continuity of $h'(\bar{x})$ there is a

$$\varrho_0 := \sup\{\varrho > 0 \mid B(0_Z, \varrho) \subset h'(\bar{x})\, B(0_X, 1)\}.$$

(2) Next we choose an arbitrary $\varepsilon \in (0, \frac{\varrho_0}{2})$. $h'(\cdot)$ is assumed to be continuous at $\bar{x}$, and therefore there is a $\delta > 0$ with

$$\|h'(\tilde{x}) - h'(\bar{x})\|_{L(X,Z)} \leq \varepsilon \text{ for all } \tilde{x} \in B(\bar{x}, 2\delta). \tag{4.8}$$

Now we fix arbitrary elements $\tilde{x}, \tilde{\tilde{x}} \in B(\bar{x}, 2\delta)$. By a Hahn-Banach theorem (Theorem C.4) there is a continuous linear functional $l \in Z^*$ with

$$\|l\|_{Z^*} = 1 \tag{4.9}$$

and

$$l(h(\tilde{\tilde{x}}) - h(\tilde{x}) - h'(\bar{x})(\tilde{\tilde{x}} - \tilde{x})) = \|h(\tilde{\tilde{x}}) - h(\tilde{x}) - h'(\bar{x})(\tilde{\tilde{x}} - \tilde{x})\|_Z. \tag{4.10}$$

Next we define a functional $\varphi : [0, 1] \to \mathbb{R}$ by

$$\varphi(t) = l(h(\tilde{x} + t(\tilde{\tilde{x}} - \tilde{x})) - th'(\bar{x})(\tilde{\tilde{x}} - \tilde{x})) \text{ for all } t \in [0, 1]. \tag{4.11}$$

φ is differentiable on $[0, 1]$ and we get

$$\varphi'(t) = l(h'(\tilde{x} + t(\tilde{\tilde{x}} - \tilde{x}))(\tilde{\tilde{x}} - \tilde{x}) - h'(\bar{x})(\tilde{\tilde{x}} - \tilde{x})). \tag{4.12}$$

By the mean value theorem there is a $\bar{t} \in (0, 1)$ with

$$\varphi(1) - \varphi(0) = \varphi'(\bar{t}). \tag{4.13}$$

Then we obtain with (4.10), (4.11), (4.13), (4.12), (4.9) and (4.8)

$$\begin{aligned}
&\|h(\tilde{\tilde{x}}) - h(\tilde{x}) - h'(\bar{x})(\tilde{\tilde{x}} - \tilde{x})\|_Z \\
&\quad = l(h(\tilde{\tilde{x}}) - h(\tilde{x}) - h'(\bar{x})(\tilde{\tilde{x}} - \tilde{x})) \\
&\quad = \varphi(1) - \varphi(0) \\
&\quad = \varphi'(\bar{t}) \\
&\quad = l(h'(\tilde{x} + \bar{t}(\tilde{\tilde{x}} - \tilde{x}))(\tilde{\tilde{x}} - \tilde{x}) - h'(\bar{x})(\tilde{\tilde{x}} - \tilde{x})) \\
&\quad \leq \|h'(\tilde{x} + \bar{t}(\tilde{\tilde{x}} - \tilde{x})) - h'(\bar{x})\|_{L(X,Z)} \, \|\tilde{\tilde{x}} - \tilde{x}\|_X \\
&\quad \leq \varepsilon \|\tilde{\tilde{x}} - \tilde{x}\|_X.
\end{aligned}$$

Hence we conclude

$$\|h(\tilde{\tilde{x}}) - h(\tilde{x}) - h'(\bar{x})(\tilde{\tilde{x}} - \tilde{x})\|_Z \leq \varepsilon \|\tilde{\tilde{x}} - \tilde{x}\|_X \text{ for all } \tilde{x}, \tilde{\tilde{x}} \in B(\bar{x}, 2\delta). \tag{4.14}$$

(3) Now we choose an arbitrary $\alpha > 1$ so that $\alpha(\frac{1}{2} + \frac{\varepsilon}{\varrho_0}) \leq 1$ (notice that $\frac{\varepsilon}{\varrho_0} < \frac{1}{2}$). For the proof of the inclusion (4.6) we take an arbitrary $x \in X$ with $h'(\bar{x})(x) = 0_Z$. For $x = 0_X$ the assertion is trivial, therefore we assume that $x \neq 0_X$. We set $\hat{\lambda} := \frac{\delta}{\|x\|_X}$ and fix an arbitrary $\lambda \in (0, \hat{\lambda}]$. Now we define sequences $(r_n)_{n\in\mathbb{N}}$ and $(u_n)_{n\in\mathbb{N}}$ as follows:

$$r_1 = 0_X,$$

$$h'(\bar{x})(u_n) = h(\bar{x} + \lambda x + r_n) \text{ for all } n \in \mathbb{N}, \tag{4.15}$$

$$r_{n+1} = r_n - u_n \text{ for all } n \in \mathbb{N}. \tag{4.16}$$

Since $h'(\bar{x})$ is assumed to be surjective, for a given $r_n \in X$ there is always a vector $u_n \in X$ with the property (4.15). Consequently, sequences $(r_n)_{n\in\mathbb{N}}$ and $(u_n)_{n\in\mathbb{N}}$ are well-defined (although they do not need to be unique). From the inclusion (4.7) which holds for $\varrho = \frac{\varrho_0}{\alpha}$ and the equation (4.15) we conclude for every $n \in \mathbb{N}$

$$\|u_n\|_X \leq \frac{\alpha}{\varrho_0}\|h(\bar{x} + \lambda x + r_n)\|_Z. \tag{4.17}$$

For simplicity we set

$$d(\lambda) := \|h(\bar{x} + \lambda x)\|_Z$$

and

$$q := \frac{\varepsilon\alpha}{\varrho_0}.$$

Since $\|\lambda x\|_X \leq \delta$ we get from the inequality (4.14)

$$\begin{aligned} d(\lambda) &= \|h(\bar{x} + \lambda x) - h(\bar{x}) - h'(\bar{x})(\lambda x)\|_Z \\ &\leq \varepsilon\|\lambda x\|_X \\ &\leq \varepsilon\delta, \end{aligned} \tag{4.18}$$

and moreover, because of $\alpha > 1$ we have

$$q \leq 1 - \frac{\alpha}{2} < \frac{1}{2}. \tag{4.19}$$

Then we assert for all $n \in \mathbb{N}$:

$$\|r_n\|_X \leq \frac{\alpha}{\varrho_0}d(\lambda)\frac{1-q^{n-1}}{1-q}, \tag{4.20}$$

$$\|h(\bar{x} + \lambda x + r_n)\|_Z \leq d(\lambda)q^{n-1} \tag{4.21}$$

and

$$\|u_n\|_X \leq \frac{\alpha}{\varrho_0}d(\lambda)q^{n-1}. \tag{4.22}$$

We prove the preceding three inequalities by induction. For $n = 1$ we get

$$\|r_1\|_X = 0,$$

$$\|h(\bar{x} + \lambda x + r_1)\|_Z = d(\lambda)$$

and by the inequality (4.17)

$$\begin{aligned}\|u_1\|_X &\leq \frac{\alpha}{\varrho_0}\|h(\bar{x}+\lambda x+r_1)\|_Z \\ &= \frac{\alpha}{\varrho_0}d(\lambda).\end{aligned}$$

Hence the inequalities (4.20), (4.21) and (4.22) are fulfilled for $n = 1$. Next assume that they are also fulfilled for any $n \in \mathbb{N}$. Then we get with (4.16), (4.20) and (4.22)

$$\begin{aligned}\|r_{n+1}\|_X &= \|r_n - u_n\|_X \\ &\leq \|r_n\|_X + \|u_n\|_X \\ &\leq \frac{\alpha}{\varrho_0}d(\lambda)\left(\frac{1-q^{n-1}}{1-q} + q^{n-1}\right) \\ &= \frac{\alpha}{\varrho_0}d(\lambda)\frac{1-q^n}{1-q}.\end{aligned}$$

Hence the inequality (4.20) is proved. For the proof of the following inequalities notice that from (4.20), (4.18) and (4.19)

$$\begin{aligned}\|\lambda x + r_n\|_X &\leq \|\lambda x\|_X + \|r_n\|_X \\ &\leq \delta + \frac{\alpha}{\varrho_0}d(\lambda)\frac{1-q^{n-1}}{1-q} \\ &\leq \delta + \frac{\alpha\varepsilon\delta}{\varrho_0}\,\frac{1-q^{n-1}}{1-q} \\ &= \delta\Big(1 + \underbrace{\frac{q}{1-q}}_{<1}\underbrace{(1-q^{n-1})}_{<1}\Big) \\ &< 2\delta \end{aligned} \tag{4.23}$$

and from (4.16), (4.20), (4.18) and (4.19)

$$\begin{aligned}\|\lambda x + r_n - u_n\|_X &\leq \|\lambda x\|_X + \|r_{n+1}\|_X \\ &\leq \delta + \frac{\alpha}{\varrho_0}d(\lambda)\frac{1-q^n}{1-q} \\ &\leq \delta\Big(1 + \underbrace{\frac{q}{1-q}}_{<1}\underbrace{(1-q^{n})}_{<1}\Big) \\ &< 2\delta. \end{aligned} \tag{4.24}$$

Next with (4.16), (4.15), (4.23), (4.24), (4.14) and (4.22) we conclude

$$\begin{aligned}
&\|h(\bar{x}+\lambda x+r_{n+1})\|_Z \\
&\quad =\|h(\bar{x}+\lambda x+r_n-u_n)\|_Z \\
&\quad =\|-h'(\bar{x})(-u_n)-h(\bar{x}+\lambda x+r_n)+h(\bar{x}+\lambda x+r_n-u_n)\|_Z \\
&\quad \leq \varepsilon\,\|-u_n\|_X \\
&\quad \leq \varepsilon\,\frac{\alpha}{\varrho_0}d(\lambda)q^{n-1} \\
&\quad = d(\lambda)q^n, \qquad (4.25)
\end{aligned}$$

and with (4.17) and (4.25) we obtain

$$\begin{aligned}
\|u_{n+1}\|_X &\leq \frac{\alpha}{\varrho_0}\|h(\bar{x}+\lambda x+r_{n+1})\|_Z \\
&\leq \frac{\alpha}{\varrho_0}d(\lambda)q^n.
\end{aligned}$$

Consequently, the inequalities (4.21) and (4.22) are fulfilled. From the inequalities (4.22) and (4.18) we get

$$\begin{aligned}
\|u_n\|_X &\leq \frac{\alpha}{\varrho_0}d(\lambda)q^{n-1} \\
&\leq \frac{\alpha\varepsilon\delta}{\varrho_0}q^{n-1} \\
&= \delta q^n \text{ for all } n\in\mathbb{N},
\end{aligned}$$

and because of the inequality (4.19) it follows $\lim\limits_{n\to\infty} u_n = 0_X$. With the equation (4.16) and the inequalities (4.22) and (4.19) we see for all $n,k\in\mathbb{N}$

$$\begin{aligned}
\|r_{n+k}-r_n\|_X &= \|r_n-u_{n+k-1}-u_{n+k-2}-\cdots-u_n-r_n\|_X \\
&\leq \|u_n\|_X+\|u_{n+1}\|_X+\cdots+\|u_{n+k-1}\|_X \\
&\leq \frac{\alpha}{\varrho_0}d(\lambda)(q^{n-1}+q^n+\cdots+q^{n+k-2}) \\
&= \frac{\alpha}{\varrho_0}d(\lambda)q^{n-1}(1+q+\cdots+q^{k-1}) \\
&= \frac{\alpha}{\varrho_0}d(\lambda)q^{n-1}\frac{1-q^k}{1-q} \\
&< \frac{\alpha d(\lambda)}{\varrho_0(1-q)}q^{n-1},
\end{aligned}$$

and therefore $(r_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. So, there is a vector $r(\lambda) \in X$ with $\lim_{n\to\infty} r_n = r(\lambda)$. Furthermore, we obtain from the equation (4.15) in the limit

$$h(\bar{x} + \lambda x + r(\lambda)) = 0_Z. \tag{4.26}$$

From (4.20) we conclude

$$\begin{aligned} \frac{\|r(\lambda)\|_X}{\lambda} &\leq \frac{\alpha}{\lambda \varrho_0} d(\lambda) \frac{1}{1-q} \\ &= \frac{\alpha}{\varrho_0(1-q)} \, \frac{\|h(\bar{x} + \lambda x) - h(\bar{x}) - \lambda h'(\bar{x})(x)\|_Z}{\lambda}, \end{aligned}$$

and therefore we have

$$\lim_{\lambda\to 0_+} \frac{r(\lambda)}{\lambda} = 0_X. \tag{4.27}$$

(4) Finally we show that x belongs to the contingent cone $T(S, \bar{x})$. Take any sequence $(\lambda_n)_{n\in\mathbb{N}}$ with $\lambda_n \in (0, \hat{\lambda}]$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} \lambda_n = 0$, and define the sequences $(\mu_n)_{n\in\mathbb{N}}$ and $(x_n)_{n\in\mathbb{N}}$ with

$$\mu_n := \frac{1}{\lambda_n} > 0 \text{ for all } n \in \mathbb{N}$$

and

$$x_n := \bar{x} + \lambda_n x + r(\lambda_n) \text{ for all } n \in \mathbb{N}.$$

From the equation (4.26) we get

$$x_n \in S \text{ for all } n \in \mathbb{N}.$$

Moreover, we have with (4.27)

$$\begin{aligned} \lim_{n\to\infty} x_n &= \lim_{n\to\infty} \bar{x} + \lambda_n x + r(\lambda_n) \\ &= \lim_{n\to\infty} \bar{x} + \lambda_n \left(x + \frac{r(\lambda_n)}{\lambda_n} \right) \\ &= \bar{x}, \end{aligned}$$

and we conclude

$$\begin{aligned}\lim_{n\to\infty} \mu_n(x_n - \bar{x}) &= \lim_{n\to\infty} \frac{1}{\lambda_n}(\lambda_n x + r(\lambda_n)) \\ &= \lim_{n\to\infty} x + \frac{r(\lambda_n)}{\lambda_n} \\ &= x.\end{aligned}$$

Consequently, we obtain $x \in T(S, \bar{x})$ which completes the proof. □

With the following theorem we show that the inclusion (4.6) also holds in the opposite direction.

Theorem 4.22 (converse of Lyusternik's inclusion).

Let $(X, \|\cdot\|_X)$ and $(Z, \|\cdot\|_Z)$ be real normed spaces, and let $h : X \to Z$ be a given mapping. Furthermore, let some $\bar{x} \in S$ with

$$S := \{x \in X \mid h(x) = 0_Z\}$$

be given. If h is Fréchet differentiable at $\bar{x}$, then it follows

$$T(S, \bar{x}) \subset \{x \in X \mid h'(\bar{x})(x) = 0_Z\}.$$

Proof Let $y \in T(S, \bar{x}) \setminus \{0_X\}$ be an arbitrary tangent vector (the assertion is evident for $y = 0_X$). Then there are a sequence $(x_n)_{n\in\mathbb{N}}$ of elements in S and a sequence $(\lambda_n)_{n\in\mathbb{N}}$ of positive real numbers with

$$\bar{x} = \lim_{n\to\infty} x_n$$

and

$$y = \lim_{n\to\infty} y_n$$

where

$$y_n := \lambda_n(x_n - \bar{x}) \text{ for all } n \in \mathbb{N}.$$

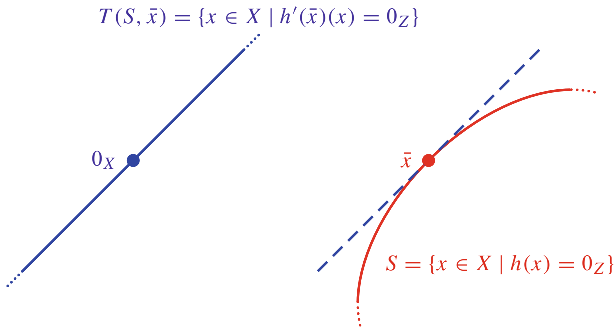

Fig. 4.9 Contingent cone with respect to an equality constraint

Consequently, by the definition of the Fréchet derivative we obtain:

$$\begin{aligned} h'(\bar{x})(y) &= h'(\bar{x})\Big(\lim_{n\to\infty} \lambda_n (x_n - \bar{x})\Big) \\ &= \lim_{n\to\infty} \lambda_n h'(\bar{x})(x_n - \bar{x}) \\ &= \lim_{n\to\infty} -\lambda_n [h(x_n) - h(\bar{x}) - h'(\bar{x})(x_n - \bar{x})] \\ &= -\lim_{n\to\infty} \|y_n\|_X \frac{h(x_n) - h(\bar{x}) - h'(\bar{x})(x_n - \bar{x})}{\|x_n - \bar{x}\|_X} \\ &= 0_Z. \end{aligned}$$

□

The proof of the preceding theorem is similar to the proof of Theorem 4.14. Since the assumptions of Theorem 4.22 are weaker than those of Theorem 4.21, we summarize the results of the two preceding theorems as follows: Under the assumptions of Theorem 4.21 we conclude

$$T(S, \bar{x}) = \{x \in X \mid h'(\bar{x})(x) = 0_Z\}.$$

(compare Fig. 4.9).

Exercises

(4.1) Let C be a convex cone in a real normed space with nonempty interior int(C). Show: int(C)= int(C)+C.

(4.2) Let X be a real linear space. Prove that a functional $f : X \to \mathbb{R}$ is sublinear if and only if its epigraph is a convex cone.

(4.3) Let S be a nonempty convex subset of a real linear space. Show that the cone generated by S is convex.

(4.4) In $\mathbb{R}^2$ let the set $S := \{(x, y) \in \mathbb{R}^2 \mid -x + y \leq 1,\ 2x + y \leq 4,\ 0 \leq x \leq \frac{3}{2},\ y \geq 0\}$ be given. Determine the cone generated by S.

(4.5) Let the set S be given as in Exercise (4.4). Determine the contingent cone to S at $(1, 2)$.

(4.6) Let S be a subset of a real normed space $(X, \|\cdot\|)$ with nonempty interior $\text{int}(S)$. For every $\bar{x} \in \text{int}(S)$ show $T(S, \bar{x}) = X$.

(4.7) Let S_1 and S_2 be two nonempty subsets of a real normed space. Prove the following implications:
(a) $\bar{x} \in S_1 \subset S_2 \ \Rightarrow\ T(S_1, \bar{x}) \subset T(S_2, \bar{x})$,
(b) $\bar{x} \in S_1 \cap S_2 \ \Rightarrow\ T(S_1 \cap S_2, \bar{x}) \subset T(S_1, \bar{x}) \cap T(S_2, \bar{x})$.

(4.8) Let S be a nonempty subset of a real normed space $(X, \|\cdot\|)$, and let some $\bar{x} \in S$ be arbitrarily given. Show:
$T(S, \bar{x}) = \{h \in X \mid$ there are a number $\sigma > 0$ and a mapping $r : (0, \sigma] \to X$ with $\lim_{t \to 0_+} \frac{1}{t} r(t) = 0_X$, and there is a sequence $(t_n)_{n \in \mathbb{N}}$ of positive real numbers converging to 0 so that $\bar{x} + t_n h + r(t_n) \in S$ for all $n \in \mathbb{N}\}$.

(4.9) Let $\bar{x}$ be an element of a subset S of a real normed space. Prove that the Clarke tangent cone $T_{Cl}(S, \bar{x})$ is closed and convex.

(4.10) Is the function $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^3$ for all $x \in \mathbb{R}$ pseudoconvex at an arbitrary $\bar{x} \in \mathbb{R}$?

5 Generalized Lagrange Multiplier Rule

In this chapter we investigate optimization problems with constraints in the form of inequalities and equalities. For such constrained problems we formulate a multiplier rule as a necessary optimality condition and we give assumptions under which this multiplier rule is also a sufficient optimality condition. The optimality condition presented generalizes the known multiplier rule published by Lagrange in 1797. With the aid of this optimality condition we deduce then the Pontryagin maximum principle known from control theory.

The classical Lagrange multiplier rule is a generalization of a Fermat theorem (given in 1629) to optimization problems with constraints in the form of equalities. Lagrange developed this rule in connection with problems from mechanics. First he applied this principle to infinite dimensional problems of the classical calculus of variations (where it led to the Euler-Lagrange equation) and later he extended it also to finite dimensional problems.

5.1 Problem Formulation

First, we present the class of optimization problems for which we formulate the generalized Lagrange multiplier rule as an optimality condition. Furthermore, we discuss several examples.

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3_5

The standard assumption of this chapter reads as follows:

$$\left.\begin{array}{l}
\text{Let } (X, \|\cdot\|_X) \text{ and } (Z, \|\cdot\|_Z) \text{be real Banach spaces;}\\
\text{let } (Y, \|\cdot\|_Y) \text{ be a partially ordered real normed space}\\
\text{with ordering cone } C \text{ with a nonempty interior int}(C);\\
\text{let } \hat{S} \text{ be a convex subset of } X \text{ with nonempty interior}\\
\text{int}(\hat{S});\\
\text{let } f : X \to \mathbb{R} \text{ be a given functional, and}\\
\text{let } g : X \to Y, h : X \to Z \text{ be given mappings;}\\
\text{furthermore let the constraint set}\\
S := \{x \in \hat{S} \mid g(x) \in -C,\ h(x) = 0_Z\}\\
\text{be nonempty.}
\end{array}\right\} \tag{5.1}$$

Under this assumption we consider the optimization problem

$$\min_{x \in S} f(x),$$

i.e., we are looking for minimal points of f on S. Since the superset $\hat{S}$ is assumed to be convex, discrete variables cannot be described with this set. Therefore, a different Lagrange theory of the discrete-continuous case is developed in Sect. 8.3.

The following examples illustrate why the considered class of constraint sets is important for applications.

Example 5.1 (finite and infinite dimensional problems).

(a) We consider again the design problem in Example 1.1. For this optimization problem the constraint set reads as follows:

$$S := \{x \in \mathbb{R}^2 \mid 2000 \le x_1^2 x_2,\ x_1 \le 4x_2,\ x_2 \le x_1,\ x_1 \ge 0,\ x_2 \ge 0\}.$$

This set can be obtained, for instance, if we choose in the standard assumption (5.1): $X = \mathbb{R}^2$, $Y = \mathbb{R}^3$, $C = \mathbb{R}^3_+$, $\hat{S} = \mathbb{R}^2_+$ and $g : \mathbb{R}^2 \to \mathbb{R}^3$ with

$$g(x_1, x_2) = \begin{pmatrix} 2000 - x_1^2 x_2 \\ x_1 - 4x_2 \\ -x_1 + x_2 \end{pmatrix} \text{ for all } (x_1, x_2) \in \mathbb{R}^2.$$

Notice that the mapping h does not appear explicitly (formally, one can also choose the mapping being zero).

(b) In Example 1.5 an optimization problem is given which has the constraint set

$$S := \{(x, \lambda) \in \mathbb{R}^2 \mid \alpha x - \sinh \alpha \leq \lambda \text{ for all } \alpha \in [0, 2],$$
$$\alpha x - \sinh \alpha \geq -\lambda \text{ for all } \alpha \in [0, 2]\}.$$

For the description of this set we choose especially in the standard assumption (5.1): $X = \mathbb{R}^2$, $Y = C[0, 2]^2$,
$C = \{(\varphi_1, \varphi_2) \in C[0, 2]^2 \mid \varphi_1(t) \geq 0 \text{ and } \varphi_2(t) \geq 0 \text{ for all } t \in [0, 2]\}$,
$\hat{S} = \mathbb{R}^2$ and $g : \mathbb{R}^2 \to C[0, 2]^2$ with

$$g(x, \lambda) = \begin{pmatrix} x \,\mathrm{id} - \sinh - \lambda \mathbf{1} \\ -x \,\mathrm{id} + \sinh - \lambda \mathbf{1} \end{pmatrix} \text{ for all } (x, \lambda) \in \mathbb{R}^2.$$

Let id denote the identity on $[0, 2]$, and let $\mathbf{1}$ denote the $C[0, 2]$ function which equals 1 on $[0, 2]$. The mapping h does not appear explicitly.

(c) In nonlinear control theory one considers, among other things, the following dynamical system with additional conditions:

$$\dot{x}(t) = f(x(t), u(t)) \text{ almost everywhere on } [t_0, t_1],$$
$$x(t_0) = x_0,$$
$$\tilde{g}(x(t_1)) = 0_{\mathbb{R}^r},$$
$$u(t) \in \Omega \text{ almost everywhere on } [t_0, t_1].$$

Next, we discuss the used notations and the necessary assumptions. The control process is considered on the fixed time interval $[t_0, t_1]$ where $-\infty < t_0 < t_1 < \infty$. Let the control u be an L_∞^m function, i.e., $u \in L_\infty^m[t_0, t_1]$. The dynamical system is described by a system of ordinary differential equations of first order. Let the function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ be continuously partially differentiable. If we define

$$W_{1,\infty}^n[t_0, t_1] := \{x : [t_0, t_1] \to \mathbb{R}^n \text{ absolutely continuous } \mid$$
$$\dot{x} \in L_\infty^n[t_0, t_1]\},$$

then the space $W_{1,\infty}^n[t_0, t_1]$ equipped with the norm $\|\cdot\|$ defined by

$$\|x\| = \max\{\|x\|_{L_\infty^n[t_0,t_1]}, \|\dot{x}\|_{L_\infty^n[t_0,t_1]}\} \text{ for all } x \in W_{1,\infty}^n[t_0, t_1]$$

is a Banach space. A solution x of the differential equation

$$\dot{x} = f(x, u)$$

for a fixed $u \in L_\infty^m[t_0, t_1]$ is defined as a function $x \in W_{1,\infty}^n[t_0, t_1]$ with

$$\dot{x}(t) = f(x(t), u(t)) \text{ almost everywhere on } [t_0, t_1].$$

Then we conclude with the initial condition $x(t_0) = x_0$ (where $x_0 \in \mathbb{R}^n$ is a given vector)

$$x(t) = x_0 + \int_{t_0}^{t} f(x(s), u(s))\, ds \text{ for all } t \in [t_0, t_1].$$

For the terminal state $x(t_1)$ we require that

$$\tilde{g}(x(t_1)) = 0_{\mathbb{R}^r}$$

where $\tilde{g} : \mathbb{R}^n \to \mathbb{R}^r$ is a continuously partially differentiable vector function. Let Ω be a convex subset of $\mathbb{R}^m$ with nonempty interior.
Among all feasible controls one tries to determine such a control for which a given functional becomes minimal. For the description of the constraint set S of this optimization problem we use the following notations in the standard assumption (5.1): $X = W_{1,\infty}^n[t_0, t_1] \times L_\infty^m[t_0, t_1]$, $Z = W_{1,\infty}^n[t_0, t_1] \times \mathbb{R}^r$, $\hat{S} = \{(x, u) \in X \mid u(t) \in \Omega$ almost everywhere on $[t_0, t_1]\}$, and $h : X \to Z$ with

$$h(x, u) = \begin{pmatrix} x(\cdot) - x_0 - \int_{t_0}^{\cdot} f(x(s), u(s))\, ds \\ \tilde{g}(x(t_1)) \end{pmatrix} \text{ for all } (x, u) \in X.$$

The constraint g does not appear explicitly in (5.1).

5.2 Necessary Optimality Conditions

In this section we present, under the assumption (5.1), a necessary condition for minimal points of f on S. This optimality condition generalizes the known Lagrange multiplier rule.

As an essential tool for the proof of the multiplier rule we need the following lemma which is obtained with the aid of the necessary optimality condition of Theorem 4.14 and the Lyusternik theorem.

Lemma 5.2 (multiplier-free optimality condition).

Let the assumption (5.1) be satisfied, and let $\bar{x}$ be a minimal point of f on S. Let the functional f and the mapping g be Fréchet differentiable at $\bar{x}$. Let the mapping h be Fréchet differentiable in a neighborhood of $\bar{x}$, and let $h'(\cdot)$ be continuous at $\bar{x}$. Moreover, let the mapping $h'(\bar{x})$ be surjective. Then there is no $x \in int(\hat{S})$ with $f'(\bar{x})(x-\bar{x}) < 0,\ g(\bar{x}) + g'(\bar{x})(x-\bar{x}) \in -\,int(C)$ and $h'(\bar{x})(x-\bar{x}) = 0_Z$.

Proof Let $\bar{x} \in S$ be a minimal point of f on S. We fix an arbitrary $x \in \text{int}(\hat{S})$ with $x \neq \bar{x},\ g(\bar{x}) + g'(\bar{x})(x-\bar{x}) \in -\,\text{int}(C)$ and $h'(\bar{x})(x-\bar{x}) = 0_Z$ (if such an x does not exist, the assertion is evident). By the Lyusternik Theorem 4.21 we get $x - \bar{x} \in T(\tilde{S}, \bar{x})$ with

$$\tilde{S} := \{x \in X \mid h(x) = 0_Z\}.$$

Consequently, there are a sequence $(x_n)_{n\in\mathbb{N}}$ of elements in $\tilde{S}$ and a sequence $(\lambda_n)_{n\in\mathbb{N}}$ of positive real numbers with

$$\bar{x} = \lim_{n\to\infty} x_n$$

and

$$x - \bar{x} = \lim_{n\to\infty} y_n \tag{5.2}$$

where

$$y_n := \lambda_n(x_n - \bar{x}) \text{ for all } n \in \mathbb{N}.$$

Because of $x \in \text{int}(\hat{S})$ we obtain with the equation (5.2)

$$\bar{x} + y_n \in \hat{S} \text{ for sufficiently large } n \in \mathbb{N}.$$

Then we get with the convexity of $\hat{S}$ for sufficiently large $n \in \mathbb{N}$

$$\begin{aligned} x_n &= \bar{x} + \frac{1}{\lambda_n} y_n \\ &= \bar{x} + \frac{1}{\lambda_n}(y_n + \bar{x} - \bar{x}) \\ &= \Big(1 - \frac{1}{\lambda_n}\Big)\bar{x} + \frac{1}{\lambda_n}(y_n + \bar{x}) \in \hat{S}, \end{aligned}$$

and therefore we have

$$x_n \in \hat{S} \cap \tilde{S} \text{ for sufficiently large } n \in \mathbb{N}. \tag{5.3}$$

For the constraint g we obtain

$$\begin{aligned} g(x_n) &= g(x_n) - g'(\bar{x})(x_n - \bar{x}) + \frac{1}{\lambda_n} g'(\bar{x})(y_n) \\ &= \frac{1}{\lambda_n}[\lambda_n(g(x_n) - g(\bar{x}) - g'(\bar{x})(x_n - \bar{x})) + g'(\bar{x})(y_n - (x - \bar{x})) \\ &\quad + g(\bar{x}) + g'(\bar{x})(x - \bar{x})] + \Big(1 - \frac{1}{\lambda_n}\Big) g(\bar{x}) \text{ for all } n \in \mathbb{N}. \end{aligned} \tag{5.4}$$

For $n \to \infty$ we conclude with $\lambda_n = \frac{\|y_n\|_X}{\|x_n - \bar{x}\|_X}$ for sufficiently large $n \in \mathbb{N}$ and the Fréchet differentiability of g

$$\lambda_n(g(x_n) - g(\bar{x}) - g'(\bar{x})(x_n - \bar{x})) + g'(\bar{x})(y_n - (x - \bar{x})) \to 0. \tag{5.5}$$

Because of

$$g(\bar{x}) + g'(\bar{x})(x - \bar{x}) \in -\operatorname{int}(C)$$

it follows with (5.5) for sufficiently large $n \in \mathbb{N}$

$$\begin{aligned} &\lambda_n(g(x_n) - g(\bar{x}) - g'(\bar{x})(x_n - \bar{x})) + g'(\bar{x})(y_n - (x - \bar{x})) \\ &+ g(\bar{x}) + g'(\bar{x})(x - \bar{x}) \in -C. \end{aligned} \tag{5.6}$$

Since $g(\bar{x}) \in -C$, we get from (5.4) and (5.6) with the convexity of C

$$g(x_n) \in -C \text{ for sufficiently large } n \in \mathbb{N}.$$

Hence we obtain with (5.3) for sufficiently large $n \in \mathbb{N}$

$$x_n \in S = \{x \in \hat{S} \mid g(x) \in -C,\ h(x) = 0_Z\},$$

and it follows

$$x - \bar{x} \in T(S, \bar{x}).$$

Then we conclude with Theorem 4.14

$$f'(\bar{x})(x - \bar{x}) \geq 0.$$

This leads to the assertion. □

Now we are able to present the *generalized Lagrange multiplier rule*.

Theorem 5.3 (generalized Lagrange multiplier rule).

Let the assumption (5.1) be satisfied, and let $\bar{x}$ be a minimal point of f on S. Let the functional f and the mapping g be Fréchet differentiable at $\bar{x}$. Let the mapping h be Fréchet differentiable in a neighborhood of $\bar{x}$, let $h'(\cdot)$ be continuous at $\bar{x}$, and let the image set $h'(\bar{x})(X)$ be closed. Then there are a real number $\mu \geq 0$ and continuous linear functionals $l_1 \in C^*$ and $l_2 \in Z^*$ with $(\mu, l_1, l_2) \neq (0, 0_{Y^*}, 0_{Z^*})$,

$$(\mu f'(\bar{x}) + l_1 \circ g'(\bar{x}) + l_2 \circ h'(\bar{x}))(x - \bar{x}) \geq 0 \text{ for all } x \in \hat{S} \tag{5.7}$$

and

$$l_1(g(\bar{x})) = 0. \tag{5.8}$$

If, in addition to the above assumptions,

$$\begin{pmatrix} g'(\bar{x}) \\ h'(\bar{x}) \end{pmatrix} \text{cone}\,(\hat{S} - \{\bar{x}\}) \;+\; \text{cone}\begin{pmatrix} C + \{g(\bar{x})\} \\ \{0_Z\} \end{pmatrix} = Y \times Z, \tag{5.9}$$

then it follows $\mu > 0$.

Proof For the proof of this theorem we consider the two cases that $h'(\bar{x})$ is not surjective or alternatively that $h'(\bar{x})$ is surjective. First, we assume that $h'(\bar{x})$ is not surjective. Then there is a $\bar{z} \in Z$ with $\bar{z} \notin h'(\bar{x})(X) = \text{cl}(h'(\bar{x})(X))$, and by a separation theorem (Theorem C.3) there is a continuous linear functional $l_2 \in Z^* \backslash \{0_{Z^*}\}$ with

$$l_2(\bar{z}) < \inf_{z \in h'(\bar{x})(X)} l_2(z).$$

Because of the linearity of $h'(\bar{x})$ it follows for every $z \in h'(\bar{x})(X)$

$$l_2(\bar{z}) < l_2(\lambda z) = \lambda l_2(z) \text{ for all } \lambda \in \mathbb{R},$$

and so we get

$$l_2(z) = 0 \text{ for all } z \in h'(\bar{x})(X)$$

resulting in

$$l_2 \circ h'(\bar{x}) = 0_{X^*}.$$

If we set $\mu = 0$ and $l_1 = 0_{Y^*}$, then the inequality (5.7) and the equation (5.8) are fulfilled, and the first part of the assertion is proved.

For the following we assume the surjectivity of $h'(\bar{x})$. In the product space $\mathbb{R} \times Y \times Z$ we define the nonempty set

$$M := \{(f'(\bar{x})(x - \bar{x}) + \alpha,\ g(\bar{x}) + g'(\bar{x})(x - \bar{x}) + y,\ h'(\bar{x})(x - \bar{x})) \\ \in \mathbb{R} \times Y \times Z \mid x \in \mathrm{int}(\hat{S}),\ \alpha > 0,\ y \in \mathrm{int}(C)\},$$

and we show several properties of this set.

First, we prove that M is a convex set. For this proof we fix two arbitrary elements $(a_1, b_1, c_1), (a_2, b_2, c_2) \in M$ and an arbitrary $\lambda \in [0, 1]$. By definition there are elements x_1, α_1, y_1 and x_2, α_2, y_2 with the properties

$$a_1 = f'(\bar{x})(x_1 - \bar{x}) + \alpha_1,\ a_2 = f'(\bar{x})(x_2 - \bar{x}) + \alpha_2,$$

$$b_1 = g(\bar{x}) + g'(\bar{x})(x_1 - \bar{x}) + y_1,\ b_2 = g(\bar{x}) + g'(\bar{x})(x_2 - \bar{x}) + y_2,$$

$$c_1 = h'(\bar{x})(x_1 - \bar{x}),\ c_2 = h'(\bar{x})(x_2 - \bar{x}).$$

Consequently, we obtain

$$\lambda a_1 + (1 - \lambda)a_2 = f'(\bar{x})(\lambda x_1 + (1 - \lambda)x_2 - \bar{x}) + \lambda\alpha_1 + (1 - \lambda)\alpha_2,$$

$$\lambda b_1 + (1 - \lambda)b_2 = g(\bar{x}) + g'(\bar{x})(\lambda x_1 + (1 - \lambda)x_2 - \bar{x}) + \lambda y_1 + (1 - \lambda)y_2,$$

$$\lambda c_1 + (1 - \lambda)c_2 = h'(\bar{x})(\lambda x_1 + (1 - \lambda)x_2 - \bar{x})$$

which implies

$$\lambda(a_1, b_1, c_1) + (1 - \lambda)(a_2, b_2, c_2) \in M.$$

Next, we show that M is an open set (i.e. $M = \mathrm{int}(M)$). Since $\mathrm{int}\,(M) \subset M$ by definition, we prove the inclusion $M \subset \mathrm{int}(M)$. We choose an arbitrary triple $(a, b, c) \in M$. Then there are elements $x \in \mathrm{int}(\hat{S})$, $\alpha > 0$ and $y \in \mathrm{int}(C)$ with

$$a = f'(\bar{x})(x - \bar{x}) + \alpha,$$

$$b = g(\bar{x}) + g'(\bar{x})(x - \bar{x}) + y$$

and

$$c = h'(\bar{x})(x - \bar{x}).$$

The mapping $h'(\bar{x})$ is continuous, linear and surjective. By the open mapping theorem the image of every open set is open under the mapping $h'(\bar{x})$. If we notice furthermore that $\alpha > 0$, $y \in \text{int}(C)$ and that Fréchet derivatives are continuous and linear, it follows $(a, b, c) \in \text{int}\,(M)$.
By Lemma 5.2 we have

$$(0, 0_Y, 0_Z) \notin M,$$

i.e.

$$M \cap \{(0, 0_Y, 0_Z)\} = \emptyset.$$

By the Eidelheit separation theorem (Theorem C.2) there are a real number μ, continuous linear functionals $l_1 \in Y^*$ and $l_2 \in Z^*$ and a real number γ with $(\mu, l_1, l_2) \neq (0, 0_{Y^*}, 0_{Z^*})$ and

$$\begin{aligned} &\mu(f'(\bar{x})(x - \bar{x}) + \alpha) + l_1(g(\bar{x}) + g'(\bar{x})(x - \bar{x}) + y) \\ &+ l_2(h'(\bar{x})(x - \bar{x})) > \gamma \geq 0 \\ &\text{for all } x \in \text{int}(\hat{S}),\ \alpha > 0 \text{ and } y \in \text{int}(C). \end{aligned} \tag{5.10}$$

If we notice that every convex subset of a real normed space with nonempty interior is contained in the closure of the interior of this set, then we get from the inequality (5.10)

$$\begin{aligned} &\mu(f'(\bar{x})(x - \bar{x}) + \alpha) + l_1(g(\bar{x}) + g'(\bar{x})(x - \bar{x}) + y) \\ &+ l_2(h'(\bar{x})(x - \bar{x})) \geq \gamma \geq 0 \\ &\text{for all } x \in \hat{S},\ \alpha \geq 0 \text{ and } y \in C. \end{aligned} \tag{5.11}$$

From the inequality (5.11) we obtain for $x = \bar{x}$

$$\mu\alpha + l_1(g(\bar{x}) + y) \geq 0 \text{ for all } \alpha \geq 0 \text{ and } y \in C. \tag{5.12}$$

With $\alpha = 1$ and $y = -g(\bar{x})$ we get $\mu \geq 0$. From the inequality (5.12) it follows for $\alpha = 0$

$$l_1(g(\bar{x})) \geq -l_1(y) \text{ for all } y \in C. \tag{5.13}$$

Assume that for some $y \in C$ it is $l_1(y) < 0$, then with $\lambda y \in C$ for some sufficiently large $\lambda > 0$ one gets a contradiction to the inequality (5.13). Therefore we have

$$l_1(y) \geq 0 \text{ for all } y \in C, \tag{5.14}$$

i.e., l_1 is an element of the dual cone C^* of C. Moreover, the inequality (5.13) implies $l_1(g(\bar{x})) \geq 0$. Since $\bar{x}$ satisfies the inequality constraint, i.e., it is $g(\bar{x}) \in -C$, we also conclude with the inequality (5.14) $l_1(g(\bar{x})) \leq 0$. Hence we get $l_1(g(\bar{x})) = 0$ and the equation (5.8) is proved.
Now, we show the equation (5.7). For $\alpha = 0$ and $y = -g(\bar{x})$ we obtain from the inequality (5.11)

$$\mu f'(\bar{x})(x-\bar{x}) + l_1(g'(\bar{x})(x-\bar{x})) + l_2(h'(\bar{x})(x-\bar{x})) \geq 0 \text{ for all } x \in \hat{S}$$

and

$$(\mu f'(\bar{x}) + l_1 \circ g'(\bar{x}) + l_2 \circ h'(\bar{x}))(x-\bar{x}) \geq 0 \text{ for all } x \in \hat{S}.$$

Finally, we consider the case that in addition to the given assumptions

$$\begin{pmatrix} g'(\bar{x}) \\ h'(\bar{x}) \end{pmatrix} \text{cone}\,(\hat{S} - \{\bar{x}\}) \; + \; \text{cone} \begin{pmatrix} C + \{g(\bar{x})\} \\ \{0_Z\} \end{pmatrix} = Y \times Z.$$

For arbitrary elements $y \in Y$ and $z \in Z$ there are nonnegative real numbers α and β and vectors $x \in \hat{S}$ and $c \in C$ with

$$y = g'(\bar{x})(\alpha(x-\bar{x})) + \beta(c + g(\bar{x}))$$

and

$$z = h'(\bar{x})(\alpha(x-\bar{x})).$$

Assume that $\mu = 0$. Then we obtain with the inequality (5.7), the equation (5.8) and the positivity of l_1

$$\begin{aligned} & l_1(y) + l_2(z) \\ & = (l_1 \circ g'(\bar{x}))(\alpha(x-\bar{x})) + \beta l_1(c + g(\bar{x})) + (l_2 \circ h'(\bar{x}))(\alpha(x-\bar{x})) \\ & \geq 0. \end{aligned}$$

Consequently, we have $l_1 = 0_{Y^*}$ and $l_2 = 0_{Z^*}$. But this contradicts the assertion that $(\mu, l_1, l_2) \neq (0, 0_{Y^*}, 0_{Z^*})$. □

Every assumption which ensures that the multiplier μ is positive is also called a *regularity assumption* or a *constraint qualification* (*CQ*). We call the additional assumption (5.9) given in Theorem 5.3 the *Kurcyusz-Robinson-Zowe*[8] *regularity condition*. Notice that a regularity assumption is only a condition on the constraint set S and not a condition on the objective functional f. For $\mu = 0$ the inequality (5.7) reads

$$(l_1 \circ g'(\bar{x}) + l_2 \circ h'(\bar{x}))\,(x - \bar{x}) \geq 0 \text{ for all } x \in \hat{S},$$

and in this case the generalized Lagrange multiplier rule does not contain any information on the objective functional f — this is not desirable. Therefore, in general, one is interested in a necessary optimality condition with $\mu > 0$. For $\mu > 0$ the inequality (5.7) leads to

$$\Big(f'(\bar{x}) + \frac{1}{\mu} l_1 \circ g'(\bar{x}) + \frac{1}{\mu} l_2 \circ h'(\bar{x})\Big)(x - \bar{x}) \geq 0 \text{ for all } x \in \hat{S},$$

and from the equation (5.8) it follows

$$\frac{1}{\mu} l_1(g(\bar{x})) = 0.$$

If we define the continuous linear functionals $u := \frac{1}{\mu} l_1 \in C^*$ and $v := \frac{1}{\mu} l_2 \in Z^*$, then we obtain

$$(f'(\bar{x}) + u \circ g'(\bar{x}) + v \circ h'(\bar{x}))(x - \bar{x}) \geq 0 \text{ for all } x \in \hat{S} \tag{5.15}$$

and

$$u(g(\bar{x})) = 0.$$

The functional

$$L := f + u \circ g + v \circ h$$

is also called *Lagrange functional*. Then the inequality (5.15) can also be written as

$$L'(\bar{x})(x - \bar{x}) \geq 0 \text{ for all } x \in \hat{S}$$

where $L'(\bar{x})$ denotes the Fréchet derivative of the Lagrange functional at $\bar{x}$.

[8] S.M. Robinson, "Stability theory for systems of inequalities in nonlinear programming, part II: differentiable nonlinear systems", *SIAM J. Numer. Anal.* 13 (1976) 497–513.

J. Zowe and S. Kurcyusz, "Regularity and stability for the mathematical programming problem in Banach spaces", *Appl. Math. Optim.* 5 (1979) 49–62.

If the superset $\hat{S}$ of the constraint set S equals the whole space X, then the generalized Lagrange multiplier rule can be specialized as follows:

Corollary 5.4 (specialized multiplier rule).

Let the assumption (5.1) with $\hat{S} = X$ be satisfied, and let $\bar{x}$ be a minimal point of f on S. Let the functional f and the mapping g be Fréchet differentiable at $\bar{x}$. Let the mapping h be Fréchet differentiable in a neighborhood of $\bar{x}$, let $h'(\cdot)$ be continuous at $\bar{x}$ and let $h'(\bar{x})(X)$ be closed. Then there are a real number $\mu \geq 0$ and continuous linear functionals $l_1 \in C^*$ and $l_2 \in Z^*$ with $(\mu, l_1, l_2) \neq (0, 0_{Y^*}, 0_{Z^*})$,

$$\mu f'(\bar{x}) + l_1 \circ g'(\bar{x}) + l_2 \circ h'(\bar{x}) = 0_{X^*}$$

and

$$l_1(g(\bar{x})) = 0.$$

If, in addition to the above assumptions, the Kurcyusz-Robinson-Zowe regularity assumption (5.9) is satisfied, then it follows $\mu > 0$.

Proof In this special setting the inequality (5.7) reads

$$(\mu f'(\bar{x}) + l_1 \circ g'(\bar{x}) + l_2 \circ h'(\bar{x}))\,(x - \bar{x}) \geq 0 \text{ for all } x \in X$$

which implies because of the linearity of the considered mappings

$$\mu f'(\bar{x}) + l_1 \circ g'(\bar{x}) + l_2 \circ h'(\bar{x}) = 0_{X^*}.$$

Then the assertion follows from Theorem 5.3. □

The assumptions of Theorem 5.3 (and also those of Corollary 5.4) can be weakened considerably: Instead of the assumption that $\text{int}(C)$ is nonempty and $h'(\bar{x})(X)$ is closed, Theorem 5.3 can also be proved under the assumption that either the set

$$\begin{pmatrix} g'(\bar{x}) \\ h'(\bar{x}) \end{pmatrix} \text{cone}\,(\hat{S} - \{\bar{x}\}) \;+\; \text{cone}\begin{pmatrix} C + \{g(\bar{x})\} \\ \{0_Z\} \end{pmatrix}$$

is closed or the product space $Y \times Z$ is finite dimensional (compare Theorem 5.3.6 in the book [365] by Werner).

In the proof of Theorem 5.3 we have shown the following implication: If the Kurcyusz-Robinson-Zowe condition is satisfied at some $\bar{x} \in S$, then the generalized Lagrange multiplier rule is not fulfilled with $\mu = 0$ at $\bar{x}$. Conversely we prove now: If the generalized Lagrange multiplier rule does not hold with $\mu = 0$ at some $\bar{x} \in S$, then a condition is satisfied at $\bar{x}$ which is in a certain sense a modified Kurcyusz-Robinson-Zowe condition (condition (5.16)). This result shows that the Kurcyusz-Robinson-Zowe condition is a very weak regularity assumption.

Theorem 5.5 (modified Kurcyusz-Robinson-Zowe condition).

Let the assumption (5.1) be satisfied (without the assumption $int(C) \neq \emptyset$), and let some $\bar{x} \in S$ be given. Let the mappings g and h be Fréchet differentiable at $\bar{x}$. If there are no continuous linear functionals $l_1 \in C^*$ and $l_2 \in Z^*$ with $(l_1, l_2) \neq (0_{Y^*}, 0_{Z^*})$,

$$(l_1 \circ g'(\bar{x}) + l_2 \circ h'(\bar{x}))\,(x - \bar{x}) \geq 0 \text{ for all } x \in \hat{S}$$

and

$$l_1(g(\bar{x})) = 0,$$

then it follows

$$cl\left(\begin{pmatrix} g'(\bar{x}) \\ h'(\bar{x}) \end{pmatrix} cone\,(\hat{S} - \{\bar{x}\}) \;+\; cone\begin{pmatrix} C + \{g(\bar{x})\} \\ \{0_Z\} \end{pmatrix}\right) = Y \times Z. \tag{5.16}$$

Proof We prove the assertion by contraposition and assume that there is a pair $(\hat{y}, \hat{z}) \in Y \times Z$ with

$$\begin{pmatrix} \hat{y} \\ \hat{z} \end{pmatrix} \notin \text{cl}\left(\begin{pmatrix} g'(\bar{x}) \\ h'(\bar{x}) \end{pmatrix} \text{cone}\,(\hat{S} - \{\bar{x}\}) \;+\; \text{cone}\begin{pmatrix} C + \{g(\bar{x})\} \\ \{0_Z\} \end{pmatrix}\right).$$

The set appearing in the right hand side of this condition is nonempty, closed and convex. By a separation theorem (Theorem C.3) there is then a continuous linear functional $(l_1, l_2) \in Y^* \times Z^*$ with $(l_1, l_2) \neq (0_{Y^*}, 0_{Z^*})$ and

$$\begin{aligned} l_1(\hat{y}) + l_2(\hat{z}) < \; & (l_1 \circ g'(\bar{x}))(\alpha(x - \bar{x})) + \beta l_1(c + g(\bar{x})) \\ & +(l_2 \circ h'(\bar{x}))\,(\alpha(x - \bar{x})) \\ & \text{for all } \alpha \geq 0,\; \beta \geq 0,\; x \in \hat{S},\; c \in C. \end{aligned}$$

With standard arguments it follows

$$\begin{aligned} \alpha(l_1 \circ g'(\bar{x}) + l_2 \circ h'(\bar{x}))(x - \bar{x}) + \beta l_1(c + g(\bar{x})) \geq 0 & \\ \text{for all } \alpha \geq 0,\; \beta \geq 0,\; x \in \hat{S},\; c \in C. & \end{aligned} \tag{5.17}$$

Then we get with $\alpha = 0$ and $\beta = 1$

$$l_1(c) \geq -l_1(g(\bar{x})) \text{ for all } c \in C$$

which leads to $l_1 \in C^*$ and $l_1(g(\bar{x})) = 0$. From the inequality (5.17) we obtain with $\alpha = 1$ and $\beta = 0$

$$(l_1 \circ g'(\bar{x}) + l_2 \circ h'(\bar{x}))\,(x - \bar{x}) \geq 0 \text{ for all } x \in \hat{S}.$$

Hence the generalized Lagrange multiplier rule is fulfilled with $\mu = 0$ at $\bar{x}$. □

The Kurcyusz-Robinson-Zowe regularity assumption may seem to be unwieldy. In the following we see that there are simpler (and therefore more restrictive) conditions implying this regularity assumption.

Theorem 5.6 (generalized MFCQ).

Let the assumption (5.1) be satisfied, and let some $\bar{x} \in S$ be given. Let the mappings g and h be Fréchet differentiable at $\bar{x}$. If the mapping $h'(\bar{x})$ is surjective and if there is a vector $\hat{x} \in int(\hat{S})$ with $g(\bar{x}) + g'(\bar{x})(\hat{x} - \bar{x}) \in -int(C)$ and $h'(\bar{x})(\hat{x} - \bar{x}) = 0_Z$, then the Kurcyusz-Robinson-Zowe regularity assumption (5.9) is satisfied.

Proof Let $y \in Y$ and $z \in Z$ be arbitrarily given elements. Because of the surjectivity of $h'(\bar{x})$ there is a vector $x \in X$ with $h'(\bar{x})(x) = z$. Then we have

$$z = h'(\bar{x})(x + \lambda(\hat{x} - \bar{x})) \text{ for all } \lambda > 0.$$

Since $\hat{x} \in \text{int}(\hat{S})$, it follows for sufficiently large $\lambda > 0$

$$x + \lambda(\hat{x} - \bar{x}) = \lambda\Big(\hat{x} + \frac{1}{\lambda}x - \bar{x}\Big) \in \text{cone}(\hat{S} - \{\bar{x}\}).$$

Because of $g(\bar{x}) + g'(\bar{x})(\hat{x} - \bar{x}) \in -\text{int}(C)$ we also get for sufficiently large $\lambda > 0$

$$-g(\bar{x}) - g'(\bar{x})(\hat{x} - \bar{x}) + \frac{1}{\lambda}(y - g'(\bar{x})(x)) \in C.$$

If we notice that

$$y = g'(\bar{x})(x + \lambda(\hat{x} - \bar{x})) + \lambda\Big(- g(\bar{x}) + g(\bar{x}) + \frac{1}{\lambda}(y - g'(\bar{x})(x + \lambda(\hat{x} - \bar{x})))\Big)$$

$$= g'(\bar{x})(x + \lambda(\hat{x} - \bar{x})) + \lambda\Big(- g(\bar{x}) - g'(\bar{x})(\hat{x} - \bar{x})$$
$$+ \frac{1}{\lambda}(y - g'(\bar{x})(x)) + g(\bar{x})\Big) \text{ for all } \lambda > 0,$$

then we conclude

$$\begin{pmatrix} y \\ z \end{pmatrix} \in \begin{pmatrix} g'(\bar{x}) \\ h'(\bar{x}) \end{pmatrix} \text{cone}\,(\hat{S} - \{\bar{x}\}) \;+\; \text{cone}\begin{pmatrix} C + \{g(\bar{x})\} \\ \{0_Z\} \end{pmatrix}.$$

Consequently, the Kurcyusz-Robinson-Zowe regularity assumption (5.9) is satisfied. □

The regularity assumption given in Theorem 5.6 is called *generalized Mangasarian-Fromovitz constraint qualification* or *generalized MFCQ*.

For the proof of the generalized Lagrange multiplier rule we have assumed that the ordering cone C has a nonempty interior. If we drop this restrictive assumption, the following example shows that the Kurcyusz-Robinson-Zowe regularity condition can be satisfied although the generalized Mangasarian-Fromovitz constraint qualification of Theorem 5.6 is not fulfilled.

Example 5.7 (Kurcyusz-Robinson-Zowe regularity condition).

We consider especially $X = Y = L_2[0, 1]$ with the natural ordering cone

$$C := \{x \in L_2[0, 1] \mid \; x(t) \geq 0 \; \text{ almost everywhere on } [0, 1]\}$$

(notice that $\text{int}(C) = \emptyset$). For an arbitrary $a \in L_2[0, 1]$ we investigate the optimization problem

$$\begin{gathered} \min < x, x > \\ \text{subject to the constraints} \\ x - a \in C \\ x \in C. \end{gathered}$$

Let $< \cdot, \cdot >$ denote the scalar product in the Hilbert space $L_2[0, 1]$. Since the ordering cone C is closed and convex, this optimization problem has at least one minimal solution $\bar{x}$ (by Theorem 2.18). If we define the set $\hat{S} := C$ and the constraint mapping $g : X \to Y$ by

$$g(x) = -x + a \text{ for all } x \in X,$$

then we obtain for this minimal solution $\bar{x}$

$$\begin{aligned} & g'(\bar{x})\,\mathrm{cone}(\hat{S} - \{\bar{x}\}) + \mathrm{cone}(C + \{g(\bar{x})\}) \\ &= g'(\bar{x})\,\mathrm{cone}(C - \{\bar{x}\}) + \mathrm{cone}(C + \{g(\bar{x})\}) \\ &= -C + \mathrm{cone}(\{\bar{x}\}) + C + \mathrm{cone}(\{g(\bar{x})\}) \\ &= X \end{aligned}$$

because we have $X = C - C$. Hence this optimization problem satisfies the Kurcyusz-Robinson-Zowe regularity condition.

In the following we turn our attention to finite dimensional problems. We specialize Corollary 5.4 for such problems. In this finite dimensional setting one speaks of the so-called *F. John conditions*[9] and in the case of $\mu > 0$ one speaks of the *Karush-Kuhn-Tucker conditions*[10] or *KKT conditions*.

Theorem 5.8 (KKT conditions).

Let the objective function $f : \mathbb{R}^n \to \mathbb{R}$ *and the constraint functions* $g : \mathbb{R}^n \to \mathbb{R}^m$ *and* $h : \mathbb{R}^n \to \mathbb{R}^p$ *be given. Let the constraint set* S *which is assumed to be nonempty be given as*

$$\begin{aligned} S := \{x \in \mathbb{R}^n \mid\ & g_i(x) \le 0 \text{ for all } i \in \{1, \ldots, m\} \text{ and} \\ & h_i(x) = 0 \text{ for all } i \in \{1, \ldots, p\}\}. \end{aligned}$$

Let $\bar{x} \in S$ *be a minimal point of* f *on* S. *Let* f *and* g *be differentiable at* $\bar{x}$ *and let* h *be continuously differentiable at* $\bar{x}$. *Moreover, let the following regularity assumption be satisfied: Assume that there is a vector* $x \in \mathbb{R}^n$ *with*

$$\nabla g_i(\bar{x})^T x < 0 \text{ for all } i \in I(\bar{x})$$

[9] F. John, "Extremum problems with inequalities as side conditions", in: K.O. Friedrichs, O.E. Neugebauer and J.J. Stoker (eds.), *Studies and Essays*, Courant Anniversary Volume (Interscience, New York, 1948).

[10] W.E. Karush, *Minima of functions of several variables with inequalities as side conditions* (Master's Dissertation, University of Chicago, 1939).

H.W. Kuhn and A.W. Tucker, "Nonlinear programming", in: J. Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, 1951), p. 481–492.

and

$$\nabla h_i(\bar{x})^T x = 0 \text{ for all } i \in \{1, \ldots, p\},$$

and that the vectors $\nabla h_1(\bar{x}), \ldots, \nabla h_p(\bar{x})$ *are linearly independent. Here let*

$$I(\bar{x}) := \{i \in 1, \ldots, m\} \mid g_i(\bar{x}) = 0\}$$

denote the index set of the inequality constraints which are "active" at $\bar{x}$*. Then there are multipliers* $u_i \geq 0$ $(i \in I(\bar{x}))$ *and* $v_i \in \mathbb{R}$ $(i \in \{1, \ldots, p\})$ *with the property*

$$\nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{p} v_i \nabla h_i(\bar{x}) = 0_{\mathbb{R}^n}.$$

Proof We verify the assumptions of Corollary 5.4. $h'(\bar{x})$ is surjective because the vectors $\nabla h_1(\bar{x}), \ldots, \nabla h_p(\bar{x})$ are linearly independent. The ordering cone C in $\mathbb{R}^m$ is given as $C = \mathbb{R}^m_+$. Then we have

$$\text{int}(C) = \{y \in \mathbb{R}^m \mid y_i > 0 \text{ for all } i \in \{1, \ldots, m\}\}$$

and $C^* = \mathbb{R}^m_+$. Consequently, we obtain for some sufficiently small $\lambda > 0$

$$g(\bar{x}) + g'(\bar{x})(\lambda x) = \begin{pmatrix} g_1(\bar{x}) + \lambda \nabla g_1(\bar{x})^T x \\ \vdots \\ g_m(\bar{x}) + \lambda \nabla g_m(\bar{x})^T x \end{pmatrix} \in -\text{int}(C)$$

and

$$h'(\bar{x})(\lambda x) = \begin{pmatrix} \lambda \nabla h_1(\bar{x})^T x \\ \vdots \\ \lambda \nabla h_p(\bar{x})^T x \end{pmatrix} = 0_{\mathbb{R}^p}.$$

Because of Theorem 5.6 the Kurcyusz-Robinson-Zowe regularity assumption is then also satisfied. By Corollary 5.4 there are elements $\mu > 0$, $l_1 \in \mathbb{R}^m_+$ and $l_2 \in \mathbb{R}^p$ with

$$\mu \nabla f(\bar{x}) + \sum_{i=1}^{m} l_{1_i} \nabla g_i(\bar{x}) + \sum_{i=1}^{p} l_{2_i} \nabla h_i(\bar{x}) = 0_{\mathbb{R}^n}$$

and

$$\sum_{i=1}^{m} l_{1_i} g_i(\bar{x}) = 0.$$

For $u := \frac{1}{\mu} l_1 \in \mathbb{R}^m_+$ and $v := \frac{1}{\mu} l_2 \in \mathbb{R}^p$ it follows

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{p} v_i \nabla h_i(\bar{x}) = 0_{\mathbb{R}^n} \tag{5.18}$$

and

$$\sum_{i=1}^{m} u_i g_i(\bar{x}) = 0. \tag{5.19}$$

Because of the inequalities

$$g_i(\bar{x}) \leq 0 \text{ for all } i \in \{1, \ldots, m\},$$

$$u_i \geq 0 \text{ for all } i \in \{1, \ldots, m\}$$

and the equation (5.19) we obtain

$$u_i g_i(\bar{x}) = 0 \text{ for all } i \in \{1, \ldots, m\}. \tag{5.20}$$

For every $i \in \{1, \ldots, m\} \backslash I(\bar{x})$ we get $g_i(\bar{x}) < 0$, and therefore we conclude with (5.20) $u_i = 0$. Hence the equation (5.18) can also be written as

$$\nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{p} v_i \nabla h_i(\bar{x}) = 0_{\mathbb{R}^n}.$$

This completes the proof. □

The regularity assumption given in the previous theorem is also called *Mangasarian-Fromovitz constraint qualification* or *MFCQ* (sometimes it is also called *Arrow-Hurwicz-Uzawa condition*). Figure 5.1 illustrates the result of Theorem 5.8. In this figure we see at a minimal solution $\bar{x}$ for some $u_1, u_2 \geq 0$

$$-\nabla f(\bar{x}) = u_2 \nabla g_2(\bar{x}) + u_3 \nabla g_3(\bar{x})$$

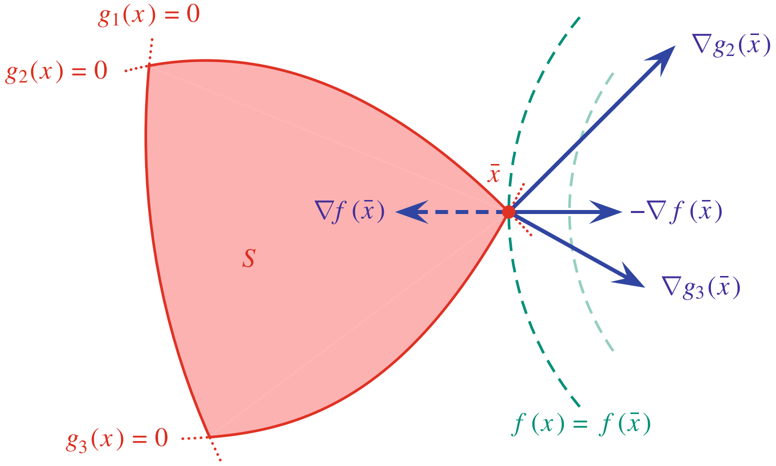

Fig. 5.1 Geometric interpretation of Theorem 5.8

or

$$\nabla f(\bar{x}) + u_2 \nabla g_2(\bar{x}) + u_3 \nabla g_3(\bar{x}) = 0_{\mathbb{R}^2}.$$

If a finite optimization problem has only constraints in the form of inequalities, then a simple sufficient condition for the Mangasarian-Fromovitz constraint qualification can be given. This condition presented in the next lemma is also called *Slater condition*.

Lemma 5.9 (Slater condition).

Let $g : \mathbb{R}^n \to \mathbb{R}^m$ be a given vector function, and let the constraint set

$$S := \{x \in \mathbb{R}^n \mid g_i(x) \leq 0 \text{ for all } i \in \{1, \ldots, m\}\}$$

be nonempty. If the functions $g_1, \ldots, g_m$ are differentiable and convex, and if there is a vector $x \in \mathbb{R}^n$ with

$$g_i(x) < 0 \text{ for all } i \in \{1, \ldots, m\},$$

then the regularity assumption of Theorem 5.8 is satisfied.

Proof With Theorem 3.16 we get for every $\bar{x} \in S$

$$g_i(x) \geq g_i(\bar{x}) + \nabla g_i(\bar{x})^T (x - \bar{x}) \text{ for all } i \in \{1, \ldots, m\},$$

and then by the assumptions we conclude

$$\begin{aligned} \nabla g_i(\bar{x})^T (x - \bar{x}) &\leq g_i(x) - g_i(\bar{x}) \\ &< 0 \ \text{ for all } i \in I(\bar{x}) \end{aligned}$$

where $I(\bar{x})$ denotes the index set of the inequality constraints being "active" at $\bar{x}$. □

The Slater condition can be checked with the aid of the constraint functions $g_1, \ldots, g_m$ without the knowledge of the minimal point $\bar{x}$. But this condition is also very restrictive since, in general, one has to assume that the functions $g_1, \ldots, g_m$ are convex.

Example 5.10 (KKT conditions).

We consider again Example 1.1. For $X = \mathbb{R}^2$ we define the objective function f by

$$f(x) = l x_1 x_2 \text{ for all } x \in \mathbb{R}^2.$$

The constraint functions $g_1, \ldots, g_5$ are given as

$$\left.\begin{aligned} g_1(x) &= 2000 - x_1^2 x_2 \\ g_2(x) &= x_1 - 4x_2 \\ g_3(x) &= -x_1 + x_2 \\ g_4(x) &= -x_1 \\ g_5(x) &= -x_2 \end{aligned}\right\} \text{ for all } x \in \mathbb{R}^2.$$

The constraint set S reads as follows

$$S := \{x \in \mathbb{R}^2 \mid g_i(x) \leq 0 \text{ for all } i \in \{1, \ldots, 5\}\}.$$

In this example there are no equality constraints. Figure 5.2 illustrates the constraint set. One can see immediately that the constraints described by g_4 and g_5 do not become active at any $x \in S$. These constraints are therefore called *redundant*.

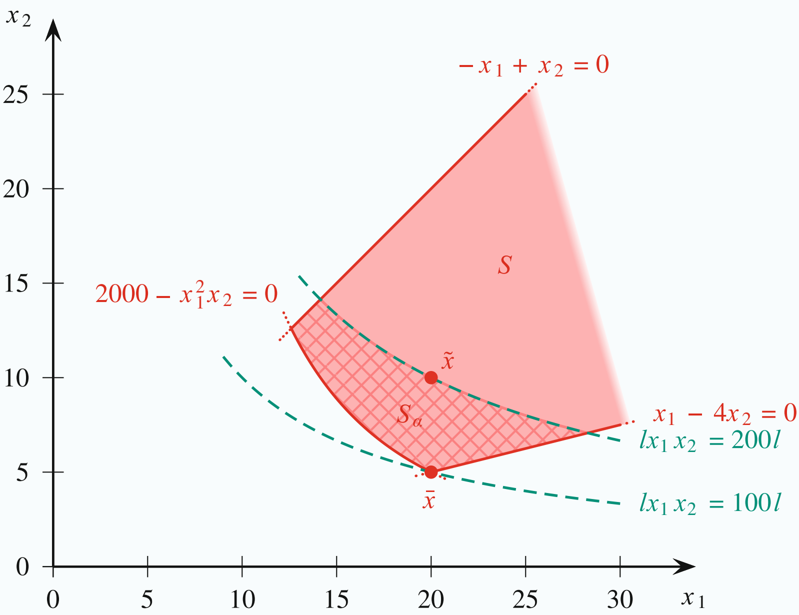

Fig. 5.2 Illustration of the constraint set S

For $\tilde{x} := (20, 10)$ the set

$$S_\alpha := \{x \in S \mid f(x) \leq \alpha\}$$

with $\alpha := f(\tilde{x}) = 200l$ is certainly compact because of the continuity of f. Hence f has at least one minimal point $\bar{x}$ on S_α. Then $\bar{x}$ is also a minimal point of f on S. If we notice that the assumptions of Theorem 5.8 are satisfied (e.g., the regularity assumption is satisfied for $x = \tilde{x} - \bar{x}$), then there are multipliers $u_1, u_2, u_3 \geq 0$ (g_4 and g_5 do not become active) with the property

$$\nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} u_i \nabla g_i(\bar{x}) = 0_{\mathbb{R}^2}.$$

For the calculation of $\bar{x}$, u_1, u_2 and u_3 one can investigate all possible cases of no, one or two constraints being active at $\bar{x}$. For the following we assume that g_1 and g_2 are active. Then we get

$$\begin{pmatrix} l\bar{x}_2 \\ l\bar{x}_1 \end{pmatrix} + u_1 \begin{pmatrix} -2\bar{x}_1\bar{x}_2 \\ -\bar{x}_1^2 \end{pmatrix} + u_2 \begin{pmatrix} 1 \\ -4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$2000 = \bar{x}_1^2 \bar{x}_2,$$

$$\bar{x}_1 = 4\bar{x}_2,$$

$$u_1 \geq 0, u_2 \geq 0.$$

A solution of this nonlinear system reads $\bar{x}_1 = 20$, $\bar{x}_2 = 5$, $u_1 = \frac{1}{30}l$, $u_2 = \frac{5}{3}l$. Consequently, $\bar{x} = (20, 5) \in S$ satisfies the Karush-Kuhn-Tucker conditions.

5.3 Sufficient Optimality Conditions

The necessary optimality conditions formulated in the preceding section are, in general, not sufficient optimality conditions if we do not consider additional assumptions. Therefore we introduce first so-called $\tilde{C}$-quasiconvex mappings and we show the equivalence of the $\tilde{C}$-quasiconvexity of a certain mapping with the sufficiency of the generalized multiplier rule as optimality condition for a modified problem. Moreover, we present a special sufficient optimality condition for finite-dimensional optimization problems.

In Definition 2.9 we already introduced quasiconvex functionals. With the following theorem we give a necessary condition for a quasiconvex directionally differentiable functional.

Theorem 5.11 (quasiconvex functional).

Let S be a nonempty convex subset of a real linear space X, and let $f : S \to \mathbb{R}$ be a quasiconvex functional having a directional derivative at some $\bar{x} \in S$ in every direction $x - \bar{x}$ with arbitrary $x \in S$. Then the following implication is satisfied for all $x \in S$

$$f(x) - f(\bar{x}) \leq 0 \Longrightarrow f'(\bar{x})(x - \bar{x}) \leq 0.$$

Proof For an arbitrary $x \in S$ we assume that

$$f(x) - f(\bar{x}) \leq 0.$$

Because of the quasiconvexity of f the level set

$$S_{f(\bar{x})} := \{\tilde{x} \in S \mid f(\tilde{x}) \leq f(\bar{x})\}$$

is then convex. Since $x, \bar{x} \in S_{f(\bar{x})}$ we obtain

$$\lambda x + (1-\lambda)\bar{x} \in S_{f(\bar{x})} \text{ for all } \lambda \in [0,1]$$

and especially

$$f(\lambda x + (1-\lambda)\bar{x}) \leq f(\bar{x}) \text{ for all } \lambda \in [0,1].$$

Then it follows

$$\frac{1}{\lambda}(f(\bar{x} + \lambda(x-\bar{x})) - f(\bar{x})) \leq 0 \text{ for all } \lambda \in (0,1].$$

Finally we conclude because of the directional differentiability of f at $\bar{x}$

$$f'(\bar{x})(x-\bar{x}) = \lim_{\lambda \to 0_+} \frac{1}{\lambda}(f(\bar{x} + \lambda(x-\bar{x})) - f(\bar{x})) \leq 0. \qquad \square$$

The previous theorem motivates the following definition of $\tilde{C}$-quasiconvex mappings.

Definition 5.12 ($\tilde{C}$-quasiconvex mapping).

Let S be a nonempty subset of a real linear space X, and let $\tilde{C}$ be a nonempty subset of a real normed space $(Y, \|\cdot\|)$. Let $f : S \to Y$ be a given mapping having a directional derivative at some $\bar{x} \in S$ in every direction $x - \bar{x}$ with arbitrary $x \in S$. The mapping f is called *$\tilde{C}$-quasiconvex* at $\bar{x}$, if for all $x \in S$:

$$f(x) - f(\bar{x}) \in \tilde{C} \implies f'(\bar{x})(x-\bar{x}) \in \tilde{C}.$$

Example 5.13 ($\tilde{C}$-quasiconvex functionals).

(a) Let S be a nonempty convex subset of a real linear space, and let $f : S \to \mathbb{R}$ be a quasiconvex functional having a directional derivative at some $\bar{x} \in S$ in every direction $x - \bar{x}$ with arbitrary $x \in S$. Then f is $\mathbb{R}_-$-quasiconvex at $\bar{x}$.

Proof We choose an arbitrary $x \in S$ with $f(x) - f(\bar{x}) \leq 0$. Then it follows with Theorem 5.11 $f'(\bar{x})(x-\bar{x}) \leq 0$, and the assertion is proved. $\square$

(b) Let S be a nonempty subset of a real linear space, and let $f : S \to \mathbb{R}$ be a given functional having a directional derivative at some $\bar{x} \in S$ in every direction $x - \bar{x}$ with arbitrary $x \in S$ and let f be pseudoconvex at $\bar{x} \in S$. Then f is also $(\mathbb{R}_- \backslash \{0\})$-quasiconvex at $\bar{x}$.

Proof For an arbitrary $x \in S$ with $f(x) - f(\bar{x}) < 0$ it follows $f'(\bar{x})(x - \bar{x}) < 0$ because of the pseudoconvexity of f at $\bar{x}$. □

With the aid of the $\tilde{C}$-quasiconvexity it is now possible to characterize the sufficiency of the generalized multiplier rule as an optimality condition for a modified optimization problem. For that purpose we need the following assumption:

$$\left.\begin{array}{l}\text{Let } \hat{S} \text{ be a nonempty subset of a real linear space } X;\\ \text{let } (Y, \|\cdot\|_Y) \text{ be a partially ordered real normed space}\\ \text{with an ordering cone } C;\\ \text{let } (Z, \|\cdot\|_Z) \text{ be a real normed space;}\\ \text{let } f : \hat{S} \to \mathbb{R} \text{ be a given functional;}\\ \text{let } g : \hat{S} \to Y \text{ and } h : \hat{S} \to Z \text{ be given mappings;}\\ \text{moreover, let the constraint set}\\ S := \{x \in \hat{S} \mid g(x) \in -C,\ h(x) = 0_Z\}\\ \text{be nonempty.}\end{array}\right\} \tag{5.21}$$

Theorem 5.14 (sufficient optimality condition).

Let the assumption (5.21) be satisfied, and let f, g, h have a directional derivative at some $\bar{x} \in S$ in every direction $x - \bar{x}$ with arbitrary $x \in \hat{S}$. Moreover, assume that there are linear functionals $u \in C'$ and $v \in Z'$ with

$$(f'(\bar{x}) + u \circ g'(\bar{x}) + v \circ h'(\bar{x}))\,(x - \bar{x}) \geq 0 \text{ for all } x \in \hat{S} \tag{5.22}$$

and

$$u(g(\bar{x})) = 0. \tag{5.23}$$

Then $\bar{x}$ is a minimal point of f on

$$\tilde{S} := \{x \in \hat{S} \mid g(x) \in -C + cone(\{g(\bar{x})\}) - cone(\{g(\bar{x})\}), h(x) = 0_Z\}$$

if and only if the mapping

$$(f, g, h) : \hat{S} \to \mathbb{R} \times Y \times Z$$

is $\tilde{C}$-quasiconvex at $\bar{x}$ with

$$\tilde{C} := (\mathbb{R}_- \backslash \{0\}) \times (-C + cone(\{g(\bar{x})\}) - cone(\{g(\bar{x})\})) \times \{0_Z\}.$$

Proof First we show under the given assumptions

$$(f'(\bar{x})(x - \bar{x}), g'(\bar{x})(x - \bar{x}), h'(\bar{x})(x - \bar{x})) \notin \tilde{C} \text{ for all } x \in \hat{S}. \tag{5.24}$$

For the proof of this assertion assume that there is a vector $x \in \hat{S}$ with

$$(f'(\bar{x})(x - \bar{x}), g'(\bar{x})(x - \bar{x}), h'(\bar{x})(x - \bar{x})) \in \tilde{C},$$

i.e.

$$\begin{aligned} f'(\bar{x})(x - \bar{x}) &< 0, \\ g'(\bar{x})(x - \bar{x}) &\in -C + \text{cone}(\{g(\bar{x})\}) - \text{cone}(\{g(\bar{x})\}), \\ h'(\bar{x})(x - \bar{x}) &= 0_Z. \end{aligned}$$

Hence we get with the equation (5.23) for some $\alpha, \beta \geq 0$

$$\begin{aligned} (f'(\bar{x}) + u \circ g'(\bar{x}) + v \circ h'(\bar{x}))(x - \bar{x}) &< u(g'(\bar{x})\,(x - \bar{x})) \\ &\leq \alpha u(g(\bar{x})) - \beta u(g(\bar{x})) \\ &= 0. \end{aligned}$$

But this inequality contradicts the inequality (5.22). Consequently, we have shown that the condition (5.24) is satisfied.

If the mapping (f, g, h) is $\tilde{C}$-quasiconvex at $\bar{x}$, then it follows from (5.24)

$$(f(x) - f(\bar{x}), g(x) - g(\bar{x}), h(x) - h(\bar{x})) \notin \tilde{C} \text{ for all } x \in \hat{S},$$

i.e. there is no $x \in \hat{S}$ with

$$\begin{aligned} f(x) &< f(\bar{x}), \\ g(x) &\in \{g(\bar{x})\} - C + \text{cone}(\{g(\bar{x})\}) - \text{cone}(\{g(\bar{x})\}) \\ &= -C + \text{cone}(\{g(\bar{x})\}) - \text{cone}(\{g(\bar{x})\}), \\ h(x) &= 0_Z. \end{aligned}$$

If we notice that with

$$g(\bar{x}) \in -C \subset -C + \text{cone}(\{g(\bar{x})\}) - \text{cone}(\{g(\bar{x})\})$$

it also follows $\bar{x} \in \tilde{S}$, then $\bar{x}$ is a minimal point of f on $\tilde{S}$.

Now we assume in the converse case that $\bar{x}$ is a minimal point of f on $\tilde{S}$, then there is no $x \in \tilde{S}$ with

$$\begin{aligned} f(x) &< f(\bar{x}), \\ g(x) &\in -C + \text{cone}(\{g(\bar{x})\}) - \text{cone}(\{g(\bar{x})\}) \\ &\quad = \{g(\bar{x})\} - C + \text{cone}(\{g(\bar{x})\}) - \text{cone}(\{g(\bar{x})\}), \\ h(x) &= 0_Z, \end{aligned}$$

i.e.

$$(f(x) - f(\bar{x}), g(x) - g(\bar{x}), h(x) - h(\bar{x})) \notin \tilde{C} \text{ for all } x \in \hat{S}.$$

Consequently, with the condition (5.24) we conclude that the mapping (f, g, h) is $\tilde{C}$-quasiconvex at $\bar{x}$. □

By Theorem 5.14 the $\tilde{C}$-quasiconvexity of the mapping (f, g, h) is characteristic of the sufficiency of the generalized Lagrange multiplier rule as an optimality condition for the optimization problem

$$\min_{x \in \tilde{S}} f(x)$$

with

$$\tilde{S} := \{x \in \hat{S} \mid g(x) \in -C + \text{cone}(\{g(\bar{x})\}) - \text{cone}(\{g(\bar{x})\}), h(x) = 0_Z\}.$$

The set $\text{cone}(\{g(\bar{x})\}) - \text{cone}(\{g(\bar{x})\})$ equals the one dimensional subspace of Y spanned by $g(\bar{x})$. Figure 5.3 illustrates the modified constraint set $\tilde{S}$.

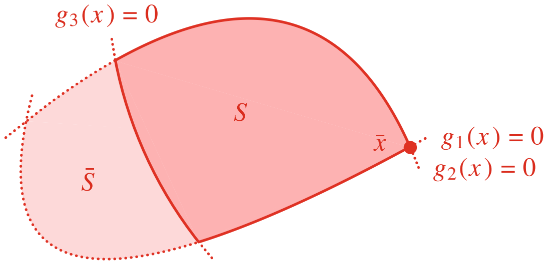

Fig. 5.3 Illustration of the set $\tilde{S} := \bar{S} \cup S$

For the original problem

$$\min_{x \in S} f(x)$$

we obtain the following result.

Corollary 5.15 (sufficient optimality condition).

Let the assumption (5.21) be satisfied, and let f, g, h have a directional derivative at some $\bar{x} \in S$ in every direction $x - \bar{x}$ with arbitrary $x \in \hat{S}$. If there are linear functionals $u \in C'$ and $v \in Z'$ with

$$(f'(\bar{x}) + u \circ g'(\bar{x}) + v \circ h'(\bar{x}))\,(x - \bar{x}) \geq 0 \text{ for all } x \in \hat{S}$$

and

$$u(g(\bar{x})) = 0,$$

and if the mapping

$$(f, g, h) : \hat{S} \to \mathbb{R} \times Y \times Z$$

is $\tilde{C}$-quasiconvex at $\bar{x}$ with

$$\tilde{C} := (\mathbb{R}_{-}\backslash\{0\}) \times (-C + cone(\{g(\bar{x})\}) - cone(\{g(\bar{x})\})) \times \{0_Z\},$$

then $\bar{x}$ is a minimal point of f on S.

Proof By Theorem 5.14 $\bar{x}$ is a minimal point of f on $\tilde{S}$. For every $x \in S$ we have

$$\begin{aligned} g(x) &\in -C \\ &\subset -C + \text{cone}(\{g(\bar{x})\}) - \text{cone}(\{g(\bar{x})\}). \end{aligned}$$

Consequently we get $S \subset \tilde{S}$, and therefore $\bar{x}$ is also a minimal point of f on S. □

With the following lemma we present conditions on f, g and h which ensure that the composite mapping (f, g, h) is $\tilde{C}$-quasiconvex.

Lemma 5.16 ($\tilde{C}$-quasiconvex mapping).

Let the assumption (5.21) be satisfied, and let f, g, h have a directional derivative at some $\bar{x} \in S$ in every direction $x - \bar{x}$ with arbitrary $x \in \hat{S}$. If the *functional f is pseudoconvex at $\bar{x}$, the mapping g is $(-C + cone(\{g(\bar{x})\}) - cone(\{g(\bar{x})\}))$-quasiconvex at $\bar{x}$ and the mapping h is $\{0_Z\}$-quasiconvex at $\bar{x}$, then the composite mapping (f, g, h) is $\tilde{C}$-quasiconvex at $\bar{x}$ with*

$$\tilde{C} := (\mathbb{R}_-\backslash\{0\}) \times (-C + cone(\{g(\bar{x})\}) - cone(\{g(\bar{x})\})) \times \{0_Z\}.$$

Proof Choose an arbitrary $x \in \hat{S}$ with

$$(f, g, h)(x) - (f, g, h)(\bar{x}) \in \tilde{C},$$

i.e.

$$\begin{aligned} f(x) - f(\bar{x}) &< 0, \\ g(x) - g(\bar{x}) &\in -C + \text{cone}(\{g(\bar{x})\}) - \text{cone}(\{g(\bar{x})\}), \\ h(x) - h(\bar{x}) &= 0_Z. \end{aligned}$$

Because of the pseudoconvexity of f it follows

$$f'(\bar{x})(x - \bar{x}) < 0,$$

the $(-C + \text{cone}(\{g(\bar{x})\}) - \text{cone}(\{g(\bar{x})\}))$-quasiconvexity of g leads to

$$g'(\bar{x})(x - \bar{x}) \in -C + \text{cone}(\{g(\bar{x})\}) - \text{cone}(\{g(\bar{x})\}),$$

and with the $\{0_Z\}$-quasiconvexity of h we obtain

$$h'(\bar{x})(x - \bar{x}) = 0_Z.$$

This completes the proof. □

Notice that the assumption of $\{0_Z\}$-quasiconvexity of the mapping h at $\bar{x}$ is very restrictive. In this case the following implication is satisfied for all $x \in \hat{S}$:

$$h(x) - h(\bar{x}) = 0_Z \Rightarrow h'(\bar{x})(x - \bar{x}) = 0_Z. \tag{5.25}$$

Such a mapping is also called *quasilinear* at $\bar{x}$. For instance, every affine linear mapping h satisfies the implication (5.25), but also the nonlinear function $h : \mathbb{R} \to \mathbb{R}$ with $h(x) = x^3$ for all $x \in \mathbb{R}$ is quasilinear at every $\bar{x} \in \mathbb{R}$.

Now we turn our attention to finite dimensional optimization problems and we give assumptions on f, g and h under which the Karush-Kuhn-Tucker conditions are sufficient optimality conditions.

Theorem 5.17 (KKT conditions).

Let an objective function $f : \mathbb{R}^n \to \mathbb{R}$ as well as constraint functions $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ be given. Let the constraint set S which is assumed to be nonempty be given as

$$S := \{x \in \mathbb{R}^n \mid g_i(x) \leq 0 \text{ for all } i \in \{1, \ldots, m\} \text{ and } h_i(x) = 0 \text{ for all } i \in \{1, \ldots, p\}\}.$$

Let the functions $f, g_1, \ldots, g_m, h_1, \ldots, h_p$ be differentiable at some $\bar{x} \in S$. Let the set

$$I(\bar{x}) := \{i \in \{1, \ldots, m\} \mid g_i(\bar{x}) = 0\}$$

denote the index set of the inequality constraints being "active" at $\bar{x}$. Assume that the objective function f is pseudoconvex at $\bar{x}$, the constraint functions g_i $(i \in I(\bar{x}))$ are quasiconvex at $\bar{x}$, and the constraint functions $h_1, \ldots, h_p$ are quasilinear at $\bar{x}$. If there are multipliers $u_i \geq 0$ $(i \in I(\bar{x}))$ and $v_i \in \mathbb{R}$ $(i \in \{1, \ldots, p\})$ with

$$\nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{p} v_i \nabla h_i(\bar{x}) = 0_{\mathbb{R}^n}, \tag{5.26}$$

then $\bar{x}$ is a minimal point of f on S.

Proof If we define additional multipliers

$$u_i := 0 \text{ for all } i \in \{1, \ldots, m\} \backslash I(\bar{x}),$$

then it follows from the equation (5.26)

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{p} v_i \nabla h_i(\bar{x}) = 0_{\mathbb{R}^n}$$

and

$$\sum_{i=1}^{m} u_i g_i(\bar{x}) = 0.$$

Then the assertion results from Corollary 5.15 in connection with Lemma 5.16. One interesting point is only the assumption of the $(-\mathbb{R}^m_+ + \text{cone}(\{g(\bar{x})\})$

$-$ cone$(\{g(\bar{x})\}))$-quasiconvexity of g at $\bar{x}$. For the verification of this assumption we choose an arbitrary $x \in \mathbb{R}^n$ with

$$g_i(x) - g_i(\bar{x}) \leq \alpha g_i(\bar{x}) - \beta g_i(\bar{x}) \text{ for all } i \in \{1, \ldots, m\} \text{ and some } \alpha, \beta \geq 0. \tag{5.27}$$

The inequality (5.27) implies

$$g_i(x) - g_i(\bar{x}) \leq 0 \text{ for all } i \in I(\bar{x}).$$

Because of the quasiconvexity of the g_i ($i \in I(\bar{x})$) it then follows

$$\nabla g_i(\bar{x})^T(x - \bar{x}) \leq 0 \text{ for all } i \in I(\bar{x}).$$

Moreover, there are numbers $\nu, \mu \geq 0$ with

$$\nabla g_i(\bar{x})^T(x - \bar{x}) \leq (\nu - \mu)g_i(\bar{x}) \text{ for all } i \in \{1, \ldots, m\}.$$

Consequently, the vector function g is $(-\mathbb{R}^m_+ + \text{cone}(\{g(\bar{x})\}) - \text{cone }(\{g(\bar{x})\}))$-quasiconvex at $\bar{x}$. This completes the proof. □

Example 5.18 (KKT conditions).

We investigate the following optimization problem:

$$\begin{gathered}\min\ 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \\ \text{subject to the constraints} \\ x_1^2 + x_2^2 - 5 \leq 0, \\ 3x_1 + x_2 - 6 \leq 0, \\ x_1, x_2 \in \mathbb{R}.\end{gathered}$$

The objective function $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$f(x_1, x_2) = 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \text{ for all } (x_1, x_2) \in \mathbb{R}^2,$$

and the constraint functions $g_1, g_2 : \mathbb{R}^2 \to \mathbb{R}$ are given by

$$g_1(x_1, x_2) = x_1^2 + x_2^2 - 5 \text{ for all } (x_1, x_2) \in \mathbb{R}^2$$

and

$$g_2(x_1, x_2) = 3x_1 + x_2 - 6 \text{ for all } (x_1, x_2) \in \mathbb{R}^2.$$

Then the Karush-Kuhn-Tucker conditions for some $\bar{x} \in \mathbb{R}^2$ read as follows:

$$\begin{pmatrix} 4\bar{x}_1 + 2\bar{x}_2 - 10 \\ 2\bar{x}_1 + 2\bar{x}_2 - 10 \end{pmatrix} + u_1 \begin{pmatrix} 2\bar{x}_1 \\ 2\bar{x}_2 \end{pmatrix} + u_2 \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$u_1(\bar{x}_1^2 + \bar{x}_2^2 - 5) = 0,$$

$$u_2(3\bar{x}_1 + \bar{x}_2 - 6) = 0.$$

Notice that $\bar{x}_1$, $\bar{x}_2$, u_1 and u_2 must also fulfill the following inequalities:

$$\begin{aligned} \bar{x}_1^2 + \bar{x}_2^2 - 5 &\leq 0, \\ 3\bar{x}_1 + \bar{x}_2 - 6 &\leq 0, \\ u_1 &\geq 0, \\ u_2 &\geq 0. \end{aligned}$$

For the determination of solutions $\bar{x}_1, \bar{x}_2, u_1, u_2$ we can consider all possible cases of no, one or two active constraints. Under the assumption that only the constraint function g_1 is active ($\Rightarrow u_2 = 0$) it follows

$$\begin{aligned} 4\bar{x}_1 + 2\bar{x}_2 - 10 + 2u_1\bar{x}_1 &= 0, \\ 2\bar{x}_1 + 2\bar{x}_2 - 10 + 2u_1\bar{x}_2 &= 0, \\ \bar{x}_1^2 + \bar{x}_2^2 - 5 &= 0, \\ 3\bar{x}_1 + \bar{x}_2 - 6 &< 0, \\ u_1 &\geq 0. \end{aligned}$$

$\bar{x}_1 = 1$, $\bar{x}_2 = 2$ and $u_1 = 1$ are a solution of this system. Hence $\bar{x} = (1, 2)$ satisfies the Karush-Kuhn-Tucker conditions.
Question: Is $\bar{x}$ also a minimal point of f on the constraint set?
In order to answer this question we use the result of Theorem 5.17. The Hessian matrix H of the objective function f reads

$$H = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$$

and is positive definite (the eigenvalues are $3 \pm \sqrt{5}$). Consequently we have

$$\begin{aligned} f(y) &= f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T H(y - x) \\ &\geq f(x) + \nabla f(x)^T (y - x) \text{ for all } x, y \in \mathbb{R}^2. \end{aligned}$$

Then we know with Theorem 3.16 that f is convex. Since the Hessian matrix of the constraint function g_1 is also positive definite, we conclude with the same arguments as for f that g_1 is convex. Consequently, by Theorem 5.17 $\bar{x} = (1, 2)$ is a minimal point of f on the constraint set.

5.4 Application to Optimal Control Problems

It is the aim of this section to apply the generalized Lagrange multiplier rule to an optimal control problem which was already described in Example 5.1,(c). For this optimal control problem we deduce the Pontryagin maximum principle as a necessary optimality condition, and moreover we give assumptions under which this maximum principle is a sufficient optimality condition.

In the following we consider the optimal control problem in Example 5.1,(c) with a special objective functional and g instead of $\tilde{g}$. Let $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be continuously partially differentiable functions. Then the investigated optimal control problem reads as follows:

$$\begin{aligned} &\min\ f_1(x(t_1)) + \int_{t_0}^{t_1} f_2(x(t), u(t))\, dt \\ &\text{subject to the constraints} \\ &\dot{x}(t) = f(x(t), u(t)) \ \text{ almost everywhere on } [t_0, t_1], \\ &x(t_0) = x_0, \\ &g(x(t_1)) = 0_{\mathbb{R}^r}, \\ &u(t) \in \Omega \ \text{ almost everywhere on } [t_0, t_1]. \end{aligned} \tag{5.28}$$

The assumptions were already given in Example 5.1,(c) (for $\tilde{g}$ instead of g).

With the following theorem we present a necessary optimality condition for an optimal control of this control problem. This optimality condition is also called the *Pontryagin maximum principle*.

Theorem 5.19 (Pontryagin maximum principle).

Let the optimal control problem (5.28) be given. Let the functions $f_1 : \mathbb{R}^n \to \mathbb{R}$, $f_2 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^r$ be continuously partially differentiable. Let Ω be a convex subset of $\mathbb{R}^m$ with nonempty interior. Let $\bar{u} \in L_\infty^m[t_0, t_1]$ be an optimal control and let $\bar{x} \in$

$W^n_{1,\infty}[t_0, t_1]$ be the resulting state. Let the matrix $\frac{\partial g}{\partial x}(\bar{x}(t_1))$ be row regular. Moreover, let the linearized system

$$\begin{aligned} \dot{x}(t) &= \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t))\, x(t) + \frac{\partial f}{\partial u}(\bar{x}(t), \bar{u}(t))\, u(t) \\ &\qquad \text{almost everywhere on } [t_0, t_1], \\ x(t_0) &= 0_{\mathbb{R}^n} \end{aligned}$$

be controllable (i.e., for every $x_1 \in \mathbb{R}^n$ there are a control $u \in L^m_\infty[t_0, t_1]$ and a resulting trajectory $x \in W^n_{1,\infty}[t_0, t_1]$ satisfying this linearized system and for which $x(t_1) = x_1$ is fulfilled).
Then there are a function $p \in W^n_{1,\infty}[t_0, t_1]$ and a vector $a \in \mathbb{R}^r$ so that

(a) $-\dot{p}(t)^T = p(t)^T \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t)) - \frac{\partial f_2}{\partial x}(\bar{x}(t), \bar{u}(t))$

almost everywhere on $[t_0, t_1]$ (adjoint equation),

(b) $-p(t_1)^T = a^T \frac{\partial g}{\partial x}(\bar{x}(t_1)) + \frac{\partial f_1}{\partial x}(\bar{x}(t_1))$

(transversality condition),

(c) *for every control $u \in L^m_\infty[t_0, t_1]$ with*

$u(t) \in \Omega$ *almost everywhere on $[t_0, t_1]$*

the following inequality is satisfied:

$$\left[p(t)^T \frac{\partial f}{\partial u}(\bar{x}(t), \bar{u}(t)) - \frac{\partial f_2}{\partial u}(\bar{x}(t), \bar{u}(t)) \right] (u(t) - \bar{u}(t)) \leq 0$$

almost everywhere on $[t_0, t_1]$ (local Pontryagin maximum principle).

Proof It is our aim to derive the given necessary optimality conditions from the generalized Lagrange multiplier rule (Theorem 5.3).

The control problem (5.28) can be treated as an optimization problem with respect to the variables (x, u). Then we define the product spaces $X := W^n_{1,\infty}[t_0, t_1] \times L^m_\infty[t_0, t_1]$ and $Z := W^n_{1,\infty}[t_0, t_1] \times \mathbb{R}^r$. The objective functional $\varphi : X \to \mathbb{R}$ is defined by

$$\varphi(x, u) = f_1(x(t_1)) + \int_{t_0}^{t_1} f_2(x(t), u(t))\, dt \text{ for all } (x, u) \in X.$$

The constraint mapping $h : X \to Z$ is given by

$$h(x,u) = \begin{pmatrix} x(\cdot) - x_0 - \int\limits_{t_0}^{\cdot} f(x(s),u(s))\,ds \\ g(x(t_1)) \end{pmatrix} \text{ for all } (x,u) \in X.$$

Furthermore, we define the set

$$\hat{S} := \{(x,u) \in X \mid u(t) \in \Omega \text{ almost everywhere on } [t_0,t_1]\}.$$

Then the optimization problem which has to be investigated reads as follows:

$$\begin{array}{c} \min \ \varphi(x,u) \\ \text{subject to the constraints} \\ h(x,u) = 0_Z \\ (x,u) \in \hat{S}. \end{array} \tag{5.29}$$

By the assumption $(\bar{x},\bar{u})$ is a minimal solution of the optimization problem (5.29). For the formulation of the generalized Lagrange multiplier rule for the problem (5.29) we need the Fréchet derivatives of φ and h at $(\bar{x},\bar{u})$. One can show that these derivatives are given as follows:

$$\varphi'(\bar{x},\bar{u})(x,u) = \frac{\partial f_1}{\partial x}(\bar{x}(t_1))\,x(t_1)$$
$$+ \int\limits_{t_0}^{t_1} \left[\frac{\partial f_2}{\partial x}(\bar{x}(s),\bar{u}(s))\,x(s) + \frac{\partial f_2}{\partial u}(\bar{x}(s),\bar{u}(s))\,u(s)\right] ds \text{ for all } (x,u) \in X$$

and

$$h'(\bar{x},\bar{u})\,(x,u) =$$
$$\Bigg(x(\cdot) - \int\limits_{t_0}^{\cdot} \left[\frac{\partial f}{\partial x}(\bar{x}(s),\bar{u}(s))\,x(s) + \frac{\partial f}{\partial u}(\bar{x}(s),\bar{u}(s))\,u(s)\right] ds,$$
$$\frac{\partial g}{\partial x}(\bar{x}(t_1))\,x(t_1)\Bigg)^T \text{ for all } (x,u) \in X.$$

Notice that h is continuously Fréchet differentiable at $(\bar{x},\bar{u})$.

Next, we show that the optimization problem (5.29) satisfies a regularity condition. By Theorem 5.6 the problem is regular, if the mapping $h'(\bar{x},\bar{u})$ is surjective (notice that we do not have inequality constraints). For the proof of the surjectivity of $h'(\bar{x}, \bar{u})$ we fix arbitrary elements $w \in W^n_{1,\infty}[t_0, t_1]$ and $y \in \mathbb{R}^r$. Since the matrix $\frac{\partial g}{\partial x}(\bar{x}(t_1))$ is row regular, there is a vector $\tilde{y} \in \mathbb{R}^n$ with

$$\frac{\partial g}{\partial x}(\bar{x}(t_1))\, \tilde{y} = y.$$

The integral equation

$$x(t) = w(t) + \int_{t_0}^{t} \frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\, x(s)\, ds \text{ for all } t \in [t_0, t_1]$$

is a linear Volterra equation of the second kind and therefore it has a solution $x := z \in W^n_{1,\infty}[t_0, t_1]$. With this solution z we then consider the linearized system of differential equations

$$\dot{x}(t) = \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t))\, x(t) + \frac{\partial f}{\partial u}(\bar{x}(t), \bar{u}(t))\, u(t) \text{ almost everywhere on } [t_0, t_1]$$

with the initial condition

$$x(t_0) = 0_{\mathbb{R}^n}$$

and the terminal condition

$$x(t_1) = \tilde{y} - z(t_1).$$

Because of the controllability of this linearized system there are a control $\tilde{u} \in L^m_\infty[t_0, t_1]$ and a resulting trajectory $\tilde{x} \in W^n_{1,\infty}[t_0, t_1]$ satisfying the initial and terminal condition. Then we obtain

$$\begin{aligned}
&h'(\bar{x}, \bar{u})\, (\tilde{x} + z, \tilde{u}) \\
&= \Bigg(\tilde{x}(\cdot) + z(\cdot) - \int_{t_0}^{\cdot} \bigg[\frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\, (\tilde{x}(s) + z(s)) \\
&\quad + \frac{\partial f}{\partial u}(\bar{x}(s), \bar{u}(s))\, \tilde{u}(s)\bigg] ds,\ \ \frac{\partial g}{\partial x}(\bar{x}(t_1))\, (\tilde{x}(t_1) + z(t_1))\Bigg)^T \\
&= \Bigg(w(\cdot) + \tilde{x}(\cdot) - \int_{t_0}^{\cdot} \bigg[\frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\tilde{x}(s) + \frac{\partial f}{\partial u}(\bar{x}(s), \bar{u}(s))\tilde{u}(s)\bigg] ds, \\
&\quad \frac{\partial g}{\partial x}(\bar{x}(t_1))\tilde{y}\Bigg)^T \\
&= (w, y)^T.
\end{aligned}$$

Consequently, the mapping $h'(\bar{x}, \bar{u})$ is surjective, and we can choose the parameter μ as 1 in the generalized Lagrange multiplier rule.

Since all assumptions of Theorem 5.3 are satisfied, there are a continuous linear functional $l \in (W^n_{1,\infty}[t_0, t_1])^*$ and a vector $a \in \mathbb{R}^r$ with

$$(\varphi'(\bar{x}, \bar{u}) + (l, a) \circ h'(\bar{x}, \bar{u}))\,(x - \bar{x}, u - \bar{u}) \geq 0 \text{ for all } x \in \hat{S}.$$

Then it follows

$$\begin{aligned}
&\frac{\partial f_1}{\partial x}(\bar{x}(t_1))\,(x(t_1) - \bar{x}(t_1)) + \int_{t_0}^{t_1} \Big[\frac{\partial f_2}{\partial x}(\bar{x}(s), \bar{u}(s))\,(x(s) - \bar{x}(s)) \\
&\quad + \frac{\partial f_2}{\partial u}(\bar{x}(s), \bar{u}(s))\,(u(s) - \bar{u}(s))\Big]\,ds \\
&\quad + l\Big(x(\cdot) - \bar{x}(\cdot) - \int_{t_0}^{\cdot} \Big[\frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\,(x(s) - \bar{x}(s)) \\
&\quad + \frac{\partial f}{\partial u}(\bar{x}(s), \bar{u}(s))\,(u(s) - \bar{u}(s))\Big]\,ds\Big) \\
&\quad + a^T \frac{\partial g}{\partial x}(\bar{x}(t_1))\,(x(t_1) - \bar{x}(t_1)) \\
&\quad \geq 0 \text{ for all } (x, u) \in \hat{S}.
\end{aligned} \tag{5.30}$$

If we plug $u = \bar{u}$ in the inequality (5.30), then we get

$$\begin{aligned}
&\frac{\partial f_1}{\partial x}(\bar{x}(t_1))\,(x(t_1) - \bar{x}(t_1)) + \int_{t_0}^{t_1} \frac{\partial f_2}{\partial x}(\bar{x}(s), \bar{u}(s))\,(x(s) - \bar{x}(s))\,ds \\
&+ l\left(x(\cdot) - \bar{x}(\cdot) - \int_{t_0}^{\cdot} \frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\,(x(s) - \bar{x}(s))\,ds\right) \\
&+ a^T \frac{\partial g}{\partial x}(\bar{x}(t_1))\,(x(t_1) - \bar{x}(t_1)) \geq 0 \text{ for all } x \in W^n_{1,\infty}[t_0, t_1]
\end{aligned}$$

and

$$\frac{\partial f_1}{\partial x}(\bar{x}(t_1))\,x(t_1) + \int_{t_0}^{t_1} \frac{\partial f_2}{\partial x}(\bar{x}(s),\bar{u}(s))\,x(s)\,ds$$

$$+\, l\left(x(\cdot) - \int_{t_0}^{\cdot} \frac{\partial f}{\partial x}(\bar{x}(s),\bar{u}(s))\,x(s)\,ds\right) + a^T \frac{\partial g}{\partial x}(\bar{x}(t_1))\,x(t_1)$$

$$= 0 \text{ for all } x \in W^n_{1,\infty}[t_0,t_1]; \tag{5.31}$$

for $x = \bar{x}$ it follows

$$\int_{t_0}^{t_1} \frac{\partial f_2}{\partial u}(\bar{x}(s),\bar{u}(s))\,(u(s)-\bar{u}(s))\,ds$$

$$+\, l\left(-\int_{t_0}^{\cdot} \frac{\partial f}{\partial u}(\bar{x}(s),\bar{u}(s))\,(u(s)-\bar{u}(s))\,ds\right)$$

$$\geq 0 \text{ for all } u \in L^m_\infty[t_0,t_1] \text{ with } u(t) \in \Omega \text{ almost everywhere on } [t_0,t_1]. \tag{5.32}$$

Next, we consider the equation (5.31) and we try to characterize the continuous linear functional l. For this characterization we need the following assertion:

If Φ is the unique solution of

$$\dot{\Phi}(t) = \frac{\partial f}{\partial x}(\bar{x}(t),\bar{u}(t))\,\Phi(t) \text{ almost everywhere on } [t_0,t_1] \tag{5.33}$$

$$\Phi(t_0) = I,$$

then for an arbitrary $y \in W^n_{1,\infty}[t_0,t_1]$ the function

$$x(\cdot) = y(\cdot) + \Phi(\cdot) \int_{t_0}^{\cdot} \Phi^{-1}(s) \frac{\partial f}{\partial x}(\bar{x}(s),\bar{u}(s))\,y(s)\,ds \tag{5.34}$$

satisfies the integral equation

$$x(\cdot) - \int_{t_0}^{\cdot} \frac{\partial f}{\partial x}(\bar{x}(s),\bar{u}(s))\,x(s)\,ds = y(\cdot). \tag{5.35}$$

For the proof of this assertion we plug x (as given in (5.34)) in the left hand side of the equation (5.35) and we obtain by integration by parts

$$
\begin{aligned}
&x(\cdot) - \int_{t_0}^{\cdot} \frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\, x(s)\, ds \\
&= y(\cdot) + \Phi(\cdot) \int_{t_0}^{\cdot} \Phi^{-1}(s) \frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\, y(s)\, ds \\
&\quad - \int_{t_0}^{\cdot} \frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s)) \Big[y(s) \\
&\quad + \Phi(s) \int_{t_0}^{s} \Phi^{-1}(\sigma) \frac{\partial f}{\partial x}(\bar{x}(\sigma), \bar{u}(\sigma))\, y(\sigma)\, d\sigma \Big]\, ds \\
&= y(\cdot) + \Phi(\cdot) \int_{t_0}^{\cdot} \Phi^{-1}(s) \frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\, y(s)\, ds \\
&\quad - \int_{t_0}^{\cdot} \frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\, y(s)\, ds \\
&\quad - \int_{t_0}^{\cdot} \dot{\Phi}(s) \int_{t_0}^{s} \Phi^{-1}(\sigma) \frac{\partial f}{\partial x}(\bar{x}(\sigma), \bar{u}(\sigma))\, y(\sigma)\, d\sigma\, ds \\
&= y(\cdot) + \Phi(\cdot) \int_{t_0}^{\cdot} \Phi^{-1}(s) \frac{\partial f}{\partial x}(\bar{x}(\sigma), \bar{u}(\sigma))\, y(\sigma)\, d\sigma\, ds \\
&\quad - \int_{t_0}^{\cdot} \frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\, y(s)\, ds \\
&\quad - \Phi(\cdot) \int_{t_0}^{\cdot} \Phi^{-1}(s) \frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\, y(s)\, ds \\
&\quad + \int_{t_0}^{\cdot} \Phi(s) \Phi^{-1}(s) \frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\, y(s)\, ds \\
&= y(\cdot).
\end{aligned}
$$

Hence the equation (5.35) is proved.

For an arbitrary function $y \in W^n_{1,\infty}[t_0, t_1]$ we conclude from the equation (5.31) with the aid of the equation (5.34)

$$l(y) = -\left(\frac{\partial f_1}{\partial x}(\bar{x}(t_1)) + a^T \frac{\partial g}{\partial x}(\bar{x}(t_1))\right)$$
$$\left(y(t_1) + \Phi(t_1) \int_{t_0}^{t_1} \Phi^{-1}(s) \frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\, y(s)\, ds\right)$$
$$- \int_{t_0}^{t_1} \frac{\partial f_2}{\partial x}(\bar{x}(s), \bar{u}(s)) \left(y(s) + \Phi(s) \int_{t_0}^{s} \Phi^{-1}(\sigma)\right.$$
$$\left.\frac{\partial f}{\partial x}(\bar{x}(\sigma), \bar{u}(\sigma))\, y(\sigma)\, d\sigma\right) ds.$$

Integration by parts leads to

$$l(y) = -\left(\frac{\partial f_1}{\partial x}(\bar{x}(t_1)) + a^T \frac{\partial g}{\partial x}(\bar{x}(t_1))\right) \left(y(t_1) + \Phi(t_1) \int_{t_0}^{t_1} \Phi^{-1}(s)\right.$$
$$\left.\frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\, y(s)\, ds\right) - \int_{t_0}^{t_1} \frac{\partial f_2}{\partial x}(\bar{x}(s), \bar{u}(s))\, y(s)\, ds$$
$$- \int_{t_0}^{t_1} \frac{\partial f_2}{\partial x}(\bar{x}(s), \bar{u}(s))\, \Phi(s)\, ds$$
$$\int_{t_0}^{t} \Phi^{-1}(s) \frac{\partial f}{\partial x}(\bar{x}(s), \bar{u}(s))\, y(s)\, ds \Bigg|_{t_0}^{t_1}$$
$$+ \int_{t_0}^{t_1} \int_{t_0}^{t} \frac{\partial f_2}{\partial x}(\bar{x}(s), \bar{u}(s))\, \Phi(s)\, ds\; \Phi^{-1}(t) \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t))\, y(t)\, dt$$

$$= -\left(\frac{\partial f_1}{\partial x}(\bar{x}(t_1)) + a^T \frac{\partial g}{\partial x}(\bar{x}(t_1))\right) y(t_1)$$

$$
\begin{aligned}
& + \int_{t_0}^{t_1} \Bigg[- \left(\frac{\partial f_1}{\partial x}(\bar{x}(t_1)) + a^T \frac{\partial g}{\partial x}(\bar{x}(t_1)) \right) \Phi(t_1)\, \Phi^{-1}(t) \\
& \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t)) - \frac{\partial f_2}{\partial x}(\bar{x}(t), \bar{u}(t)) \\
& - \int_{t_0}^{t_1} \frac{\partial f_2}{\partial x}(\bar{x}(s), \bar{u}(s))\, \Phi(s)\, ds\ \Phi^{-1}(t) \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t)) \\
& + \int_{t_0}^{t} \frac{\partial f_2}{\partial x}(\bar{x}(s), \bar{u}(s))\, \Phi(s)\, ds\ \Phi^{-1}(t) \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t)) \Bigg] y(t)\, dt \\
= & - \left(\frac{\partial f_1}{\partial x}(\bar{x}(t_1)) + a^T \frac{\partial g}{\partial x}(\bar{x}(t_1)) \right) y(t_1) \\
& + \int_{t_0}^{t_1} \Bigg[- \left(\frac{\partial f_1}{\partial x}(\bar{x}(t_1)) + a^T \frac{\partial g}{\partial x}(\bar{x}(t_1)) \right) \Phi(t_1)\, \Phi^{-1}(t) \\
& \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t)) - \frac{\partial f_2}{\partial x}(\bar{x}(t), \bar{u}(t)) \\
& - \int_{t}^{t_1} \frac{\partial f_2}{\partial x}(\bar{x}(s), \bar{u}(s))\, \Phi(s)\, ds\ \Phi^{-1}(t) \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t)) \Bigg] y(t)\, dt \\
& \qquad \text{for all } y \in W_{1,\infty}^n[t_0, t_1].
\end{aligned}
$$

For the expression in brackets we introduce the notation $r(t)^T$, i.e.

$$
\begin{aligned}
r(t)^T := & - \left(\frac{\partial f_1}{\partial x}(\bar{x}(t_1)) + a^T \frac{\partial g}{\partial x}(\bar{x}(t_1)) \right) \Phi(t_1)\, \Phi^{-1}(t) \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t)) \\
& - \frac{\partial f_2}{\partial x}(\bar{x}(t), \bar{u}(t)) \\
& - \int_{t}^{t_1} \frac{\partial f_2}{\partial x}(\bar{x}(s), \bar{u}(s))\, \Phi(s)\, ds\, \Phi^{-1}(t) \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t)) \\
& \qquad \text{almost everywhere on } [t_0, t_1].
\end{aligned}
$$

With the equation (5.33) it follows (compare page 256)

$$
\begin{aligned}
\left(\overset{\bullet}{\Phi^{-1}}(t)\right) &= -\Phi^{-1}(t)\,\dot{\Phi}(t)\,\Phi^{-1}(t)\\
&= -\Phi^{-1}(t)\,\frac{\partial f}{\partial x}(\bar{x}(t),\bar{u}(t))\Phi(t)\,\Phi^{-1}(t)\\
&= -\Phi^{-1}(t)\,\frac{\partial f}{\partial x}(\bar{x}(t),\bar{u}(t)) \;\text{ almost everywhere on } [t_0,t_1].
\end{aligned}
$$

Then we obtain

$$
\begin{aligned}
r(t)^T = &\left(\frac{\partial f_1}{\partial x}(\bar{x}(t_1)) + a^T\frac{\partial g}{\partial x}(\bar{x}(t_1))\right)\Phi(t_1)\left(\overset{\bullet}{\Phi^{-1}}(t)\right)\\
&-\frac{\partial f_2}{\partial x}(\bar{x}(t),\bar{u}(t)) + \int_t^{t_1}\frac{\partial f_2}{\partial x}(\bar{x}(s),\bar{u}(s))\,\Phi(s)\,ds\left(\overset{\bullet}{\Phi^{-1}}(t)\right)\\
&\qquad\text{almost everywhere on } [t_0,t_1].
\end{aligned}
$$

For

$$
\begin{aligned}
p(t)^T := &-\left(\frac{\partial f_1}{\partial x}(\bar{x}(t_1)) + a^T\frac{\partial g}{\partial x}(\bar{x}(t_1))\right)\Phi(t_1)\,\Phi^{-1}(t)\\
&-\int_t^{t_1}\frac{\partial f_2}{\partial x}(\bar{x}(s),\bar{u}(s))\,\Phi(s)\,ds\,\Phi^{-1}(t) \text{ for all } t\in[t_0,t_1]
\end{aligned}
$$

we get

$$
\dot{p}(t) = -r(t) \;\text{ almost everywhere on } [t_0,t_1].
$$

Then it follows

$$
-p(t_1)^T = a^T\frac{\partial g}{\partial x}(\bar{x}(t_1)) + \frac{\partial f_1}{\partial x}(\bar{x}(t_1)),
$$

i.e., the transversality condition is satisfied. Moreover, we conclude

$$
\begin{aligned}
&p(t)^T\frac{\partial f}{\partial x}(\bar{x}(t),\bar{u}(t)) - \frac{\partial f_2}{\partial x}(\bar{x}(t),\bar{u}(t))\\
&= -\left(\frac{\partial f_1}{\partial x}(\bar{x}(t_1)) + a^T\frac{\partial g}{\partial x}(\bar{x}(t_1))\right)\Phi(t_1)\,\Phi^{-1}(t)\frac{\partial f}{\partial x}(\bar{x}(t),\bar{u}(t))
\end{aligned}
$$

$$- \int_{t}^{t_1} \frac{\partial f_2}{\partial x}(\bar{x}(s), \bar{u}(s))\, \Phi(s)\, ds\, \Phi^{-1}(t) \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t)) - \frac{\partial f_2}{\partial x}(\bar{x}(t), \bar{u}(t))$$

$$= r(t)^T$$

$$= -\dot{p}(t)^T \quad \text{almost everywhere on } [t_0, t_1].$$

Hence p satisfies the adjoint equation

$$-\dot{p}(t)^T = p(t)^T \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t)) - \frac{\partial f_2}{\partial x}(\bar{x}(t), \bar{u}(t)) \text{ almost everywhere on } [t_0, t_1].$$

Then the continuous linear functional l can be written as

$$l(y) = p(t_1)^T y(t_1) - \int_{t_0}^{t_1} \dot{p}(t)^T y(t)\, dt \text{ for all } y \in W^n_{1,\infty}[t_0, t_1].$$

Now we turn our attention to the inequality (5.32). From this inequality we obtain by integration by parts

$$0 \leq \int_{t_0}^{t_1} \frac{\partial f_2}{\partial u}(\bar{x}(s), \bar{u}(s))\, (u(s) - \bar{u}(s))\, ds$$

$$-l\left(\int_{t_0}^{\cdot} \frac{\partial f}{\partial u}(\bar{x}(s), \bar{u}(s))\, (u(s) - \bar{u}(s))\, ds \right)$$

$$= \int_{t_0}^{t_1} \frac{\partial f_2}{\partial u}(\bar{x}(s), \bar{u}(s))\, (u(s) - \bar{u}(s))\, ds$$

$$-p(t_1)^T \int_{t_0}^{t_1} \frac{\partial f}{\partial u}(\bar{x}(s), \bar{u}(s))\, (u(s) - \bar{u}(s))\, ds$$

$$+ \int_{t_0}^{t_1} \dot{p}(t)^T \int_{t_0}^{t} \frac{\partial f}{\partial u}(\bar{x}(s), \bar{u}(s))\, (u(s) - \bar{u}(s))\, ds\, dt$$

$$= \int_{t_0}^{t_1} \frac{\partial f_2}{\partial u}(\bar{x}(s), \bar{u}(s))\, (u(s) - \bar{u}(s))\, ds$$

$$-p(t_1)^T \int_{t_0}^{t_1} \frac{\partial f}{\partial u}(\bar{x}(s), \bar{u}(s))\,(u(s) - \bar{u}(s))\,ds$$

$$+\, p(t)^T \int_{t_0}^{t} \frac{\partial f}{\partial u}(\bar{x}(s), \bar{u}(s))\,(u(s) - \bar{u}(s))\,ds \Bigg|_{t_0}^{t_1}$$

$$-\int_{t_0}^{t_1} p(t)^T \frac{\partial f}{\partial u}(\bar{x}(t), \bar{u}(t))\,(u(t) - \bar{u}(t))\,dt$$

$$= \int_{t_0}^{t_1} \frac{\partial f_2}{\partial u}(\bar{x}(s), \bar{u}(s))\,(u(s) - \bar{u}(s))\,ds$$

$$-\int_{t_0}^{t_1} p(t)^T \frac{\partial f}{\partial u}(\bar{x}(s), \bar{u}(s))\,(u(s) - \bar{u}(s))\,ds$$

$$= \int_{t_0}^{t_1} \left[\frac{\partial f_2}{\partial u}(\bar{x}(t), \bar{u}(t)) - p(t)^T \frac{\partial f}{\partial u}(\bar{x}(t), \bar{u}(t)) \right] (u(t) - \bar{u}(t))\,dt$$

for all $u \in L_\infty^m[t_0, t_1]$ with $u(t) \in \Omega$ almost everywhere on $[t_0, t_1]$.

Then we get for every control $u \in L_\infty^m[t_0, t_1]$ with $u(t) \in \Omega$ almost everywhere on $[t_0, t_1]$

$$\left[p(t)^T \frac{\partial f}{\partial u}(\bar{x}(t), \bar{u}(t)) - \frac{\partial f_2}{\partial u}(\bar{x}(t), \bar{u}(t)) \right] (u(t) - \bar{u}(t)) \leq 0$$

almost everywhere on $[t_0, t_1]$.

Hence the local Pontryagin maximum principle is also shown, and the proof of Theorem 5.19 is completed. □

Remark 5.20 (Hamilton function and Kalman condition).

(a) If one defines the so-called *Hamilton function*

$$H : W_{1,\infty}^n[t_0, t_1] \times L_\infty^m[t_0, t_1] \times W_{1,\infty}^n[t_0, t_1] \to W_{1,\infty}^n[t_0, t_1]$$

pointwise by

$$H(x, u, p)(t) = p(t)^T f(x(t), u(t)) - f_2(x(t), u(t)) \text{ for all } t \in [t_0, t_1],$$

then the adjoint equation reads

$$-\dot{p}(t)^T = \frac{\partial H}{\partial x}(\bar{x}, \bar{u}, p)(t) \text{ almost everywhere on } [t_0, t_1],$$

and the local Pontryagin maximum principle can be written as

$$\frac{\partial H}{\partial u}(\bar{x}, \bar{u}, p)(t)\ (u(t) - \bar{u}(t)) \leq 0 \text{ almost everywhere on } [t_0, t_1],$$

(for all $u \in L_\infty^m[t_0, t_1]$ with $u(t) \in \Omega$ almost everywhere on $[t_0, t_1]$).

(b) In Theorem 5.19 it is assumed among other things that the linearized system

$$\dot{x}(t) = A(t)\,x(t) + B(t)\,u(t) \text{ almost everywhere on } [t_0, t_1],$$
$$x(t_0) = 0_{\mathbb{R}^n}$$

with $A(t) := \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t))$ and $B(t) := \frac{\partial f}{\partial u}(\bar{x}(t), \bar{u}(t))$ is controllable. If the matrix functions A and B are independent of time, i.e. $A := A(t)$ almost everywhere on $[t_0, t_1]$ and $B := B(t)$ almost everywhere on $[t_0, t_1]$, then, by a known result of control theory, this system is controllable, if the so-called *Kalman condition* is satisfied, i.e.

$$\operatorname{rank}(B, AB, A^2B, \ldots, A^{n-1}B) = n.$$

(c) If the set Ω in the considered control problem is of the special form $\Omega = \mathbb{R}^m$, then the local Pontryagin maximum principle can be formulated in the special form:

For all $u \in L_\infty^m[t_0, t_1]$ it follows

$$p(t)^T \frac{\partial f}{\partial u}(\bar{x}(t), \bar{u}(t)) - \frac{\partial f_2}{\partial u}(\bar{x}(t), \bar{u}(t)) = 0$$
$$\text{almost everywhere on } [t_0, t_1].$$

Example 5.21 (Pontryagin maximum principle).

We consider Example 1.4 and investigate the following optimal control problem:
Determine a control $u \in L_\infty[0, 1]$ which minimizes

$$\int_0^1 (u(t))^2 \, dt$$

subject to the constraints

$$\begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t) \;\text{ almost everywhere on } [0, 1],$$

$$\begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} = \begin{pmatrix} -2\sqrt{2} \\ 5\sqrt{2} \end{pmatrix},$$

$$(x_1(1))^2 + (x_2(1))^2 - 1 = 0.$$

The system of linear differential equations of this problem satisfies the Kalman condition. According to Remark 5.20,(b) this system is controllable. We assume that there is an optimal control $\bar{u} \in L_\infty[0, 1]$ for this problem. Then the adjoint equation reads as follows

$$\begin{aligned}(-\dot{p}_1(t), -\dot{p}_2(t)) &= (p_1(t), p_2(t)) \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \\ &= (0, p_1(t)) \;\text{ almost everywhere on } [0, 1],\end{aligned}$$

i.e. we have

$$\dot{p}_1(t) = 0 \;\text{ almost everywhere on } [0, 1]$$

and

$$\dot{p}_2(t) = -p_1(t) \;\text{ almost everywhere on } [0, 1].$$

This leads to the general solution

$$p(t) = \begin{pmatrix} c_1 \\ -c_1 t + c_2 \end{pmatrix} \;\text{ for all } t \in [0, 1]$$

with real numbers c_1 and c_2. The transversality condition can be written as

$$(-p_1(1), -p_2(1)) = a(2\bar{x}_1(1), 2\bar{x}_2(1))$$

or

$$\begin{pmatrix} p_1(1) \\ p_2(1) \end{pmatrix} = -2a \begin{pmatrix} \bar{x}_1(1) \\ \bar{x}_2(1) \end{pmatrix}.$$

Hence it follows

$$\begin{pmatrix} c_1 \\ -c_1 + c_2 \end{pmatrix} = -2a \begin{pmatrix} \bar{x}_1(1) \\ \bar{x}_2(1) \end{pmatrix}.$$

Next, we consider the local Pontryagin maximum principle as given in Remark 5.20,(c):

$$(p_1(t), p_2(t)) \begin{pmatrix} 0 \\ 1 \end{pmatrix} - 2\bar{u}(t) = 0 \text{ almost everywhere on } [0, 1].$$

Consequently we get

$$\bar{u}(t) = \frac{1}{2} p_2(t) = \frac{1}{2}(-c_1 t + c_2) \text{ almost everywhere on } [0, 1].$$

Moreover, we have with the second linear differential equation as constraint

$$\dot{\bar{x}}_2(t) = \bar{u}(t) = \frac{1}{2}(-c_1 t + c_2) \text{ almost everywhere on } [0, 1]$$

and

$$\bar{x}_2(t) = -\frac{c_1}{4} t^2 + \frac{c_2}{2} t + 5\sqrt{2} \text{ for all } t \in [0, 1].$$

With this equation and the first linear differential equation as constraint we obtain

$$\bar{x}_1(t) = -\frac{c_1}{12} t^3 + \frac{c_2}{4} t^2 + 5\sqrt{2}\, t - 2\sqrt{2} \text{ for all } t \in [0, 1].$$

With the terminal condition

$$(\bar{x}_1(1))^2 + (\bar{x}_2(1))^2 = 1$$

we then conclude

$$(-\frac{c_1}{12}+\frac{c_2}{4}+3\sqrt{2})^2+(-\frac{c_1}{4}+\frac{c_2}{2}+5\sqrt{2})^2=1.$$

We summarize our results as follows: For an optimal control $\bar{u}\in L_\infty[0,1]$ there are real numbers α, β and γ with the property

$$\bar{u}(t)=\alpha t+\beta \ \text{ almost everywhere on } [0,1],$$
$$(\frac{\alpha}{6}+\frac{\beta}{2}+3\sqrt{2})^2+(\frac{\alpha}{2}+\beta+5\sqrt{2})^2=1,$$
$$-\alpha=\gamma(\frac{\alpha}{6}+\frac{\beta}{2}+3\sqrt{2}),$$
$$\alpha+\beta=\gamma(\frac{\alpha}{2}+\beta+5\sqrt{2}).$$

$(\alpha,\beta,\gamma)=(3\sqrt{2},-6\sqrt{2},-6)$ is a solution of these nonlinear equations. Then the resulting control satisfies the necessary optimality conditions of Theorem 5.19.

At the end of this section we investigate the question under which assumptions the conditions (a), (b) and (c) of Theorem 5.19 are sufficient optimality conditions.

Theorem 5.22 (sufficient optimality conditions).

Let the optimal control problem (5.28) be given. Furthermore, let a control $\bar{u}\in L_\infty^m[t_0,t_1]$ *with*

$$\bar{u}(t)\in\Omega \ \textit{ almost everywhere on } [t_0,t_1]$$

and a resulting state $\bar{x}\in W_{1,\infty}^n[t_0,t_1]$ *be given where*

$$\dot{\bar{x}}(t)=f(\bar{x}(t),\bar{u}(t)) \ \textit{ almost everywhere on } [t_0,t_1],$$
$$\bar{x}(t_0)=x_0,$$
$$g(\bar{x}(t_1))=0_{\mathbb{R}^r}.$$

Let the function f_1 *be convex (at* $\bar{x}(t_1)$*) and differentiable at* $\bar{x}(t_1)$*. Let the function* f_2 *be convex and differentiable. Let the function* f *be differentiable. Let the function* g *be differentiable at* $\bar{x}(t_1)$*. Moreover, let there are a function* $p\in W_{1,\infty}^n[t_0,t_1]$ *and a vector* $a\in\mathbb{R}^r$ *so that*

$$(a) \quad -\dot{p}(t)^T = p(t)^T \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t)) - \frac{\partial f_2}{\partial x}(\bar{x}(t), \bar{u}(t))$$

$$\text{almost everywhere on } [t_0, t_1], \tag{5.36}$$

$$(b) \quad -p(t_1)^T = a^T \frac{\partial g}{\partial x}(\bar{x}(t_1)) + \frac{\partial f_1}{\partial x}(\bar{x}(t_1)), \tag{5.37}$$

(c) *for every control* $u \in L_\infty^m[t_0, t_1]$ *with*

$u(t) \in \Omega$ *almost everywhere on* $[t_0, t_1]$

we have

$$\Big[p(t)^T \frac{\partial f}{\partial u}(\bar{x}(t), \bar{u}(t)) - \frac{\partial f_2}{\partial u}(\bar{x}(t), \bar{u}(t))\Big](u(t) - \bar{u}(t)) \leq 0$$

$$\text{almost everywhere on } [t_0, t_1]. \tag{5.38}$$

Let the function $a^T g(\cdot)$ *be quasiconvex at* $\bar{x}(t_1)$ *and almost everywhere on* $[t_0, t_1]$ *let the functional defined by* $-p(t)^T f(x(t), u(t))$ *be convex (at* $(\bar{x}(t), \bar{u}(t))$*). Then* $\bar{u}$ *is an optimal control for the control problem (5.28).*

Proof Let $u \in L_\infty^m[t_0, t_1]$ be an arbitrary control with the resulting state $x \in W_{1,\infty}^n[t_0, t_1]$ such that (x, u) satisfies the constraints of the problem (5.28). Then we get with the adjoint equation (5.36)

$$\begin{aligned}
&-\frac{d}{dt}(p(t)^T(x(t) - \bar{x}(t))) \\
&= -\dot{p}(t)^T(x(t) - \bar{x}(t)) - p(t)^T(\dot{x}(t) - \dot{\bar{x}}(t)) \\
&= \Big[p(t)^T \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t)) - \frac{\partial f_2}{\partial x}(\bar{x}(t), \bar{u}(t))\Big](x(t) - \bar{x}(t)) \\
&\quad -p(t)^T[f(x(t), u(t)) - f(\bar{x}(t), \bar{u}(t))] \text{ almost everywhere on } [t_0, t_1].
\end{aligned}$$

With this relationship it follows

$$\begin{aligned}
&f_2(x(t), u(t)) - f_2(\bar{x}(t), \bar{u}(t)) - \frac{d}{dt}(p(t)^T(x(t) - \bar{x}(t))) \\
&= f_2(x(t), u(t)) - f_2(\bar{x}(t), \bar{u}(t)) - \frac{\partial f_2}{\partial x}(\bar{x}(t), \bar{u}(t))\,(x(t) - \bar{x}(t)) \\
&\quad - p(t)^T\Big[f(x(t), u(t)) - f(\bar{x}(t), \bar{u}(t)) \\
&\quad - \frac{\partial f}{\partial x}(\bar{x}(t), \bar{u}(t))\,(x(t) - \bar{x}(t))\Big] \text{ almost everywhere on } [t_0, t_1].
\end{aligned} \tag{5.39}$$

Since the function f_2 is convex and differentiable, we conclude

$$f_2(x(t),u(t)) - f_2(\bar{x}(t),\bar{u}(t)) - \frac{\partial f_2}{\partial x}(\bar{x}(t),\bar{u}(t))\,(x(t)-\bar{x}(t))$$
$$\geq \frac{\partial f_2}{\partial u}(\bar{x}(t),\bar{u}(t))\,(u(t)-\bar{u}(t)) \text{ almost everywhere on } [t_0,t_1].$$

Similarly we obtain because of the convexity of the functional defined by $-p(t)^T f(x(t),u(t))$ (at $(\bar{x}(t),\bar{u}(t))$)

$$-p(t)^T\Big[f(x(t),u(t)) - f(\bar{x}(t),\bar{u}(t)) - \frac{\partial f}{\partial x}(\bar{x}(t),\bar{u}(t))^T(x(t)-\bar{x}(t))\Big]$$
$$\geq -p(t)^T\frac{\partial f}{\partial u}(\bar{x}(t),\bar{u}(t))^T(u(t)-\bar{u}(t)) \text{ almost everywhere on } [t_0,t_1].$$

Then it results from the equation (5.39) and the inequality (5.38)

$$f_2(x(t),u(t)) - f_2(\bar{x}(t),\bar{u}(t)) - \frac{d}{dt}(p(t)^T(x(t)-\bar{x}(t)) \geq 0$$
$$\text{almost everywhere on } [t_0,t_1].$$

Because of $x(t_0)=\bar{x}(t_0)=x_0$ integration leads to

$$\int_{t_0}^{t_1}[f_2(x(t),u(t)) - f_2(\bar{x}(t),\bar{u}(t))]\,dt - p(t_1)^T(x(t_1)-\bar{x}(t_1)) \geq 0. \tag{5.40}$$

With the transversality condition (5.37) and the differentiability and convexity of f_1 (at $\bar{x}(t_1)$) we get

$$\begin{aligned}
&-p(t_1)^T(x(t_1)-\bar{x}(t_1))\\
&\quad= a^T\frac{\partial g}{\partial x}(\bar{x}(t_1))\,(x(t_1)-\bar{x}(t_1)) + \frac{\partial f_1}{\partial x}(\bar{x}(t_1))\,(x(t_1)-\bar{x}(t_1))\\
&\quad\leq a^T\frac{\partial g}{\partial x}(\bar{x}(t_1))\,(x(t_1)-\bar{x}(t_1)) + f_1(x(t_1)) - f_1(\bar{x}(t_1)).
\end{aligned} \tag{5.41}$$

Because of the differentiability and quasiconvexity of $a^T g(\cdot)$ at $\bar{x}(t_1)$ the equation

$$0 = a^T g(x(t_1)) - a^T g(\bar{x}(t_1))$$

implies the inequality

$$0 \geq a^T \frac{\partial g}{\partial x}(\bar{x}(t_1))\,(x(t_1) - \bar{x}(t_1)). \tag{5.42}$$

The inequalities (5.41) and (5.42) then lead to

$$-p(t_1)^T (x(t_1) - \bar{x}(t_1)) \leq f_1(x(t_1)) - f_1(\bar{x}(t_1))$$

which implies with the inequality (5.40)

$$f_1(x(t_1)) + \int_{t_0}^{t_1} f_2(x(t), u(t))\,dt \geq f_1(\bar{x}(t_1)) + \int_{t_0}^{t_1} f_2(\bar{x}(t), \bar{u}(t))\,dt.$$

Hence $\bar{u}$ is an optimal control for the control problem (5.28). □

For the proof of the preceding theorem we did not use the general result of Theorem 5.14. Therefore the given assumptions under which the optimality conditions are sufficient are certainly not the weakest assumptions on the arising functions.

Example 5.23 (Pontryagin maximum principle).

We consider again the control problem of Example 5.21. We have already shown that the control $\bar{u} \in L_\infty[0,1]$ with

$$\bar{u}(t) = 3\sqrt{2}\,t - 6\sqrt{2} \ \text{ almost everywhere on } [0,1]$$

satisfies the optimality conditions (5.36), (5.37) and (5.38) with $p \in W^2_{1,\infty}[0,1]$ defined by

$$p(t) = \begin{pmatrix} -6\sqrt{2} \\ 6\sqrt{2}\,t - 12\sqrt{2} \end{pmatrix} \ \text{ for all } t \in [0,1]$$

and

$$a := 6.$$

The functions g and f_2 are convex. The vector function f is linear and every component of p is negative. Consequently, all assumptions of Theorem 5.22 are satisfied. Then this theorem says that $\bar{u}$ is an optimal control for the control problem of Example 5.21.

Exercises

(5.1) Let S be a closed linear subspace of a real normed space $(X, \|\cdot\|)$. Prove: If there is a vector $x \in X \setminus S$, then there is a continuous linear functional $l \in X^* \setminus \{0_{X^*}\}$ with

$$l(s) = 0 \text{ for all } s \in S.$$

(5.2) Show: For every convex subset S of a real normed space with nonempty interior it follows $\mathrm{cl}\,(\mathrm{int}\,(S)) = \mathrm{cl}\,(S)$.

(5.3) Does the constraint set

$$S := \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1 \text{ and } (x_1 - 1)^2 + (x_2 - 1)^2 \leq 1\}$$

satisfy a regularity assumption?

(5.4) Let the optimization problem

$$\begin{gathered} \min\ x_1 + x_2 \\ \text{subject to the constraints} \\ x_2 \leq x_1^3 \\ x_1 \in \mathbb{R},\ x_2 \geq 0 \end{gathered}$$

be given.

(a) Show that $\bar{x} = (0, 0)$ is a solution of this optimization problem.
(b) Is the MFCQ satisfied at $\bar{x} = (0, 0)$?
(c) Are the KKT conditions satisfied at $\bar{x} = (0, 0)$?

(5.5) Determine a minimal solution of the optimization problems:

(a) $\min\ (x-3)^2 + (y-2)^2$
subject to the constraints
$x^2 + y^2 \leq 5$
$x + y \leq 3$
$x \geq 0,\ y \geq 0.$

(b) $\min\ (x - \frac{9}{4})^2 + (y-2)^2$
subject to the constraints
$x^2 - y \leq 0$
$x + y - 6 \leq 0$
$x \geq 0,\ y \geq 0.$

(c) $\max\ 3x - y - 4z^2$
subject to the constraints
$x + y + z \leq 0$
$-x + 2y + z^2 = 0$
$x, y, z \in \mathbb{R}.$

(5.6) Is every point on the straight line between (0, 0) and (6, 0) a minimal solution of the optimization problem

$$\begin{aligned} &\min \ \tfrac{x+3y+3}{2x+y+6} \\ &\text{subject to the constraints} \\ &2x + y \leq 12 \\ &-x + 2y \leq 4 \\ &x \geq 0, \ y \geq 0 \ \ ? \end{aligned}$$

(5.7) For given functions $f_1, \ldots, f_n : \mathbb{R} \to \mathbb{R}$ consider the optimization problem

$$\begin{aligned} &\min \sum_{i=1}^{n} f_i(x_i) \\ &\text{subject to the constraints} \\ &\sum_{i=1}^{n} x_i = 1 \\ &x_i \geq 0 \text{ for all } i \in \{1, \ldots, n\}. \end{aligned}$$

Prove: If $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)$ is a minimal solution of this problem and for every $i \in \{1, \ldots, n\}$ the function f_i is differentiable at $\bar{x}_i$, then there is a real number α with

$$\left. \begin{array}{l} f_i'(\bar{x}_i) \geq \alpha \\ (f_i'(\bar{x}_i) - \alpha)\, \bar{x}_i = 0 \end{array} \right\} \text{ for all } i \in \{1, \ldots, n\}.$$

(5.8) Let $\hat{S}$ be a nonempty subset of $\mathbb{R}^n$, and let $f : \hat{S} \to \mathbb{R}$, $g : \hat{S} \to \mathbb{R}^m$ and $h : \hat{S} \to \mathbb{R}^p$ be given functions. Let the constraint set

$$\begin{aligned} S := \{x \in \hat{S} \mid g_i(x) \leq 0 \text{ for all } i \in \{1, \ldots, m\} \text{ and} \\ h_i(x) = 0 \text{ for all } i \in \{1, \ldots, p\}\} \end{aligned}$$

be nonempty. Let the functions $f, g_1, \ldots, g_m, h_1, \ldots, h_p$ be differentiable at some $\bar{x} \in S$. Let there be multipliers $u_i \geq 0$ $(i \in I(\bar{x}))$ and $v_i \in \mathbb{R}$ $(i \in \{1, \ldots, p\})$ with

$$\left(\nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{p} v_i \nabla h_i(\bar{x}) \right)^T (x - \bar{x}) \geq 0 \text{ for all } x \in \hat{S}.$$

Let f be pseudoconvex at $\bar{x}$, for every $i \in I(\bar{x})$ let g_i be quasiconvex at $\bar{x}$, for every $i \in \{1, \ldots, p\}$ with $v_i > 0$ let h_i be quasiconvex at $\bar{x}$, and for every $i \in \{1, \ldots, p\}$ with $v_i < 0$ let $-h_i$ be quasiconvex at $\bar{x}$. Prove that $\bar{x}$ is a minimal point of f on S.

(5.9) Determine an optimal control $\bar{u} \in L_\infty^2[0,1]$ of the following problem:

$$\min \int_0^1 \left[u_1(t) - \frac{1}{3}x_1(t) + 2u_2(t) - \frac{2}{3}x_2(t)\right] dt$$

subject to the constraints

$$\left.\begin{aligned} \dot{x}_1(t) &= 12u_1(t) - 2u_1(t)^2 - x_1(t) - u_2(t) \\ \dot{x}_2(t) &= 12u_2(t) - 2u_2(t)^2 - x_2(t) - u_1(t) \end{aligned}\right\} \text{ a.e. on } [0,1]$$

$$x_1(0) = x_{0_1},\ x_2(0) = x_{0_2}$$

$$\left.\begin{aligned} u_1(t) &\geq 0 \\ u_2(t) &\geq 0 \end{aligned}\right\} \text{ almost everywhere on } [0,1]$$

where x_{0_1} and x_{0_2} are given real numbers.

6 Duality

The duality theory is also an additional important part of the optimization theory. A main question which is investigated in duality theory reads as follows: Under which assumptions is it possible to associate an equivalent maximization problem to a given (in general convex) minimization problem. This maximization problem is also called the optimization problem dual to the minimization problem. In this chapter we formulate the dual problem to a constrained minimization problem and we investigate the relationships between the both optimization problems. For a linear problem we transform the dual problem in such a way that we again obtain a linear optimization problem. Finally, we apply these results to a problem of linear Chebyshev approximation.

6.1 Problem Formulation

In this section we consider a constrained optimization problem. Let the constraints be given in the form of a general system of inequalities. Then we associate a so-called dual problem to this optimization problem, the so-called primal problem.

First, we formulate the standard assumption for the following investigations:

$$\left.\begin{array}{l}\text{Let } \hat{S} \text{ be a nonempty subset of a real linear space } X;\\ \text{let } (Y, \|\cdot\|) \text{ be a partially ordered real normed space with}\\ \text{the ordering cone } C;\\ \text{let } f : \hat{S} \to \mathbb{R} \text{ be a given objective functional;}\\ \text{let } g : \hat{S} \to Y \text{ be a given constraint mapping;}\\ \text{let the constraint set be given as } S := \{x \in \hat{S} \mid g(x) \in -C\}\\ \text{which is assumed to be nonempty.}\end{array}\right\} \quad (6.1)$$

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*, https://doi.org/10.1007/978-3-030-42760-3_6

Under the assumption (6.1) we investigate the constrained optimization problem

$$\begin{aligned} &\min f(x) \\ &\text{subject to the constraints} \\ &g(x) \in -C \\ &x \in \hat{S}. \end{aligned} \tag{6.2}$$

In this context the optimization problem (6.2) is also called *primal problem*. With the following lemma we see that, under the additional assumption of the ordering cone being closed, this problem is equivalent to the optimization problem

$$\min_{x \in \hat{S}} \sup_{u \in C^*} f(x) + u(g(x)) \tag{6.3}$$

where C^* denotes the dual cone of C.

Lemma 6.1 (equivalence of problems (6.2) and (6.3)).

Let the assumption (6.1) be satisfied and in addition let the ordering cone C be closed. Then $\bar{x}$ is a minimal solution of the problem (6.2) if and only if $\bar{x}$ is a minimal solution of the problem (6.3). In this case the extremal values of both problems are equal.

Proof We start this proof with two simple observations.

(a) For every $x \in \hat{S}$ with $g(x) \in -C$ we have

$$u(g(x)) \leq 0 \text{ for all } u \in C^*$$

and therefore we get

$$\sup_{u \in C^*} u(g(x)) = 0.$$

(b) For every $x \in \hat{S}$ with $g(x) \notin -C$ there is, by a separation theorem (Theorem C.3), a $\bar{u} \in C^* \setminus \{0_{X^*}\}$ with

$$\bar{u}(g(x)) > 0$$

which implies

$$\sup_{u \in C^*} u(g(x)) = \infty$$

(notice that the cone C is convex and closed).

Now we begin with the actual proof of this lemma. Let $\bar{x} \in S$ be a minimal point of f on S. Consequently, we obtain for every $x \in \hat{S}$

$$\begin{aligned}
\sup_{u \in C^*} f(\bar{x}) + u(g(\bar{x})) &= f(\bar{x}) + \sup_{u \in C^*} u(g(\bar{x})) \\
&= f(\bar{x}) \\
&\leq f(\bar{x}) + \sup_{u \in C^*} u(g(x)) \\
&\leq f(x) + \sup_{u \in C^*} u(g(x)) \\
&\leq \sup_{u \in C^*} f(x) + u(g(x)).
\end{aligned}$$

Hence $\bar{x} \in S$ is also a minimal solution of the optimization problem (6.3).

Finally, we assume that $\bar{x} \in \hat{S}$ is a minimal point of the functional $\varphi : \hat{S} \to \mathbb{R} \cup \{\infty\}$ with

$$\varphi(x) = \sup_{u \in C^*} f(x) + u(g(x)) \text{ for all } x \in \hat{S}$$

on $\hat{S}$. Assume that $g(\bar{x}) \notin -C$. Then with the arguments under (b) we get

$$\sup_{u \in C^*} u(g(\bar{x})) = \infty$$

which is a contradiction to the solvability of problem (6.3). Consequently, by (a) we have

$$\sup_{u \in C^*} u(g(\bar{x})) = 0.$$

Then we obtain for all $x \in S$

$$\begin{aligned}
f(\bar{x}) &= f(\bar{x}) + \sup_{u \in C^*} u(g(\bar{x})) \\
&= \sup_{u \in C^*} f(\bar{x}) + u(g(\bar{x})) \\
&\leq \sup_{u \in C^*} f(x) + u(g(x)) \\
&= f(x) + \sup_{u \in C^*} u(g(x)) \\
&= f(x).
\end{aligned}$$

Hence $\bar{x} \in S$ is a minimal point of f on S. □

Now we associate another problem to the primal problem (6.2). This new problem results from the problem (6.3) by exchanging "min" and "sup" and by replacing "min" by "inf" and "sup" by "max". This optimization problem then reads:

$$\max_{u \in C^*} \inf_{x \in \hat{S}} f(x) + u(g(x)). \tag{6.4}$$

The optimization problem (6.4) is called the *dual problem* associated to the primal problem (6.2). Obviously, this dual problem is equivalent to the optimization problem

$$\begin{gathered} \max \ \lambda \\ \text{subject to the constraints} \\ f(x) + u(g(x)) \geq \lambda \text{ for all } x \in \hat{S} \\ \lambda \in \mathbb{R}, u \in C^*. \end{gathered} \tag{6.5}$$

If $\bar{u} \in C^*$ is a maximal solution of the dual problem (6.4) with the maximal value $\bar{\lambda}$, then $(\bar{\lambda}, \bar{u})$ is a maximal solution of the problem (6.5). Conversely, for every maximal solution $(\bar{\lambda}, \bar{u})$ of the problem (6.5) $\bar{u}$ is a maximal solution of the dual problem with the maximal value $\bar{\lambda}$.

Example 6.2 (primal and dual problem).

We investigate the very simple primal problem

$$\begin{gathered} \min \ -2x_1 + x_2 \\ \text{subject to the constraints} \\ x_1 + x_2 - 3 \leq 0 \\ x \in \hat{S} := \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}. \end{gathered} \tag{6.6}$$

It is obvious that the constraint set S consists of the three points $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Then $\bar{x} := \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ is a minimal solution of problem (6.6) with the minimal value -3. According to (6.4) the dual problem associated to problem (6.6) can be written as

$$\max_{u \geq 0} \inf_{x \in \hat{S}} -2x_1 + x_2 + u(x_1 + x_2 - 3)$$

$$= \max_{u \geq 0} \min\{-3u,\ 4+u,\ -4+5u,\ -8+u,\ 0,\ -3\}$$

$$= \max_{u \geq 0} \left\{ \begin{array}{ll} -4+5u, & \text{if } u \leq -1 \\ -8+u, & \text{if } -1 \leq u \leq 2 \\ -3u, & \text{if } u \geq 2 \end{array} \right\}. \tag{6.7}$$

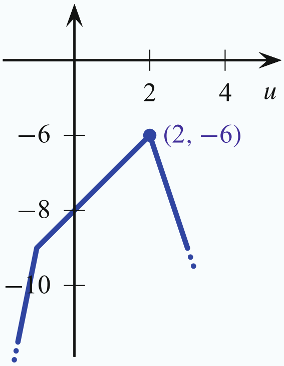

Fig. 6.1 Illustration of the objective of the dual problem (6.7)

Figure 6.1 illustrates the objective function of this maximization problem. It is evident that $\bar{u} := 2$ is a maximal solution of the dual problem with the maximal value -6. Notice that minimal and maximal value of the primal and dual problem, respectively, do not coincide.

For fundamental results in duality one needs some type of convexity. We start with the definition of convex mappings.

Definition 6.3 (convex mapping).

Let $\hat{S}$ be a nonempty convex subset of a real linear space, and let Y be a partially ordered real linear space with an ordering cone C. A mapping $g : \hat{S} \to Y$ is called *convex*, if for all $x, y \in \hat{S}$:

$$\lambda g(x) + (1-\lambda)\, g(y) - g(\lambda x + (1-\lambda) y) \in C \quad \text{for all } \lambda \in [0, 1].$$

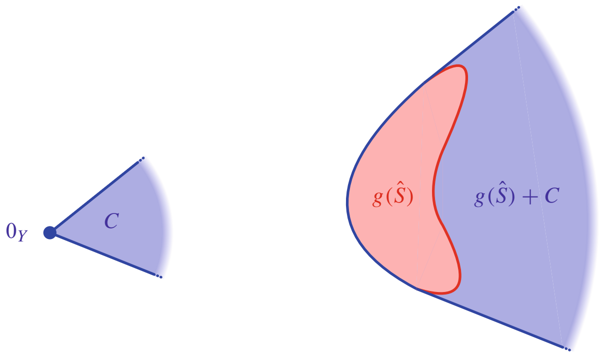

Fig. 6.2 Illustration of a convex-like mapping g

Example 6.4 (convex mapping).

Let $\hat{S}$ be a nonempty convex subset of a real linear space, and let $f_1, \ldots, f_n : \hat{S} \to \mathbb{R}$ be convex functionals. If the linear space $\mathbb{R}^n$ is supposed to be partially ordered in a natural way (i.e., $C := \mathbb{R}^n_+$), then the vector function $f = (f_1, \ldots, f_n) : \hat{S} \to \mathbb{R}^n$ is convex.

Now we turn our attention to a class of mappings which are slightly more general than convex ones.

Definition 6.5 (convex-like mapping).

Let $\hat{S}$ be a nonempty subset of a real linear space and let Y be a partially ordered real linear space with an ordering cone C. A mapping $g : \hat{S} \to Y$ is called *convex-like*, if the set $g(\hat{S}) + C$ is convex (Fig. 6.2 illustrates this notion).

The following example shows that the class of convex-like mappings includes the class of convex mappings, and, in fact, it goes beyond this class slightly.

Example 6.6 (convex-like mappings).

(a) Let $\hat{S}$ be a nonempty convex subset of a real linear space, and let Y be a partially ordered real linear space with an ordering cone C. Every convex mapping $g : \hat{S} \to Y$ is also convex-like.

Proof We have to show that the set $g(\hat{S}) + C$ is a convex set. For that purpose choose arbitrary elements $y_1, y_2 \in g(\hat{S}) + C$ and an arbitrary number $\lambda \in [0, 1]$. Then there are elements $x_1, x_2 \in \hat{S}$ and $c_1, c_2 \in C$ with

$$y_1 = g(x_1) + c_1$$

and

$$y_2 = g(x_2) + c_2.$$

Consequently, we get with the convexity of g

$$\begin{aligned}
&\lambda y_1 + (1 - \lambda)\, y_2 \\
&= \lambda\, g(x_1) + (1 - \lambda)\, g(x_2) + \lambda c_1 + (1 - \lambda)\, c_2 \\
&\in \{g(\lambda x_1 + (1 - \lambda)\, x_2)\} + C + \lambda C + (1 - \lambda)\, C \\
&= \{g(\lambda x_1 + (1 - \lambda)\, x_2)\} + C,
\end{aligned}$$

i.e.

$$\lambda y_1 + (1 - \lambda y_2) \in g(\hat{S}) + C.$$

Hence the set $g(\hat{S}) + C$ is convex, and the mapping g is convex-like. □

(b) We consider the mapping $g : \mathbb{R} \to \mathbb{R}^2$ with

$$g(x) = \begin{pmatrix} x \\ \sin x \end{pmatrix} \quad \text{for all } x \in \mathbb{R}.$$

Let the real linear space $\mathbb{R}^2$ be partially ordered in a natural way (i.e., $C := \mathbb{R}^2_+$). Then the mapping g is convex-like but it is certainly not convex (see Fig. 6.3).

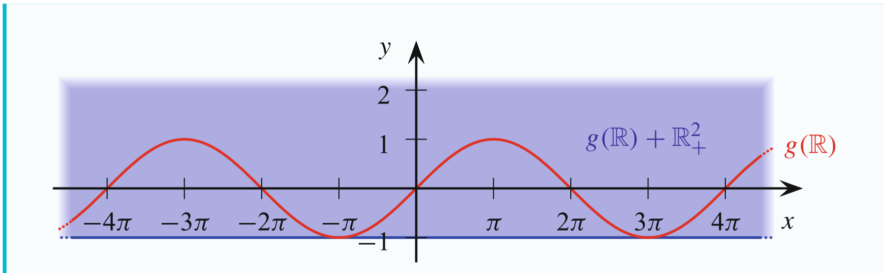

Fig. 6.3 Illustration of the result in Example 6.6,(b)

If under the assumption (6.1) the set $\hat{S}$ is convex, if the objective functional f is convex and if the constraint mapping g is convex, then the composite mapping $(f, g) : \hat{S} \to \mathbb{R} \times Y$ is convex-like (with respect to the product cone $\mathbb{R}_+ \times C$ in $\mathbb{R} \times Y$). With the assumption of the convex-likeness of (f, g) it is even possible to treat certain nonconvex optimization problems with this duality theory.

6.2 Duality Theorems

In this section the relationships between the primal problem (6.2) and the dual problem (6.4) are investigated. We present a so-called weak duality theorem and a so-called strong duality theorem which says in which sense the primal and dual problem are equivalent.

First we formulate a so-called *weak duality theorem*.

Theorem 6.7 (weak duality theorem).

Let the assumption (6.1) be satisfied. For every $\hat{x} \in S$ (i.e., for every feasible element of the primal problem (6.2)) and for every $\hat{u} \in C^*$ (i.e., for every feasible element of the dual problem (6.4)) the following inequality is satisfied:

$$\inf_{x \in \hat{S}} f(x) + \hat{u}(g(x)) \leq f(\hat{x}).$$

Proof For arbitrary elements $\hat{x} \in S$ and $\hat{u} \in C^*$ it follows

$$\inf_{x \in \hat{S}} f(x) + \hat{u}(g(x)) \leq f(\hat{x}) + \hat{u}(g(\hat{x})) \leq f(\hat{x})$$

because $g(\hat{x}) \in -C$. □

It follows immediately from the weak duality theorem that the maximal value of the dual problem is bounded from above by the minimal value of the primal problem (if these values exist and the assumption (6.1) is satisfied). In particular, one obtains a lower bound of the minimal value of the primal problem, if one determines the value of the objective functional of the dual problem at an arbitrary element of the constraint set of the dual problem.

If the primal and dual problem are solvable, then it is not guaranteed in general that the extremal values of these two problems are equal. If these two problems are solvable and the extremal values are not equal, then one speaks of a *duality gap*. In Example 6.2 an optimization problem is presented for which a duality gap arises.

Next, we come to an important result concerning the solvability of the dual problem and the obtained maximal value. With the aid of a generalized Slater condition it can be shown that a duality gap cannot arise. The following theorem is also called a *strong duality theorem*.

Theorem 6.8 (strong duality theorem).

Let the assumption (6.1) be satisfied, and in addition let the ordering cone C have a nonempty interior int(C) and let the composite mapping $(f, g) : \hat{S} \to \mathbb{R} \times Y$ be convex-like (with respect to the product cone $\mathbb{R}_+ \times C$ in $\mathbb{R} \times Y$). If the primal problem (6.2) is solvable and the generalized Slater condition is satisfied, i.e., there is a vector $\hat{x} \in \hat{S}$ with $g(\hat{x}) \in -$int(C), then the dual problem (6.4) is also solvable and the extremal values of the two problems are equal.

Proof In the following we investigate the set

$$
\begin{aligned}
M &:= \{\, (f(x) + \alpha,\ g(x) + y) \in \mathbb{R} \times Y \mid x \in \hat{S},\ \alpha \geq 0,\ y \in C \,\} \\
&= (f, g)\,(\hat{S}) + \mathbb{R}_+ \times C.
\end{aligned}
$$

By assumption the composite mapping $(f, g) : \hat{S} \to \mathbb{R} \times Y$ is convex-like, and therefore the set M is convex. Because of $\text{int}(C) \neq \emptyset$ the set M has a nonempty interior $\text{int}(M)$ as well. Since the primal problem is solvable there is a vector $\bar{x} \in S$ with

$$f(\bar{x}) \leq f(x) \text{ for all } x \in S.$$

Consequently we have

$$(f(\bar{x}), 0_Y) \notin \text{int}(M)$$

and

$$\text{int}(M) \cap \{\, (f(\bar{x}), 0_Y) \,\} = \emptyset.$$

By the Eidelheit separation theorem (Theorem C.2) there are real numbers μ and γ and a continuous linear functional $u \in Y^*$ with $(\mu, u) \neq (0, 0_{Y^*})$ and

$$\mu\beta + u(z) > \gamma \geq \mu f(\bar{x}) \text{ for all } (\beta, z) \in \text{int}(M). \tag{6.8}$$

Since every convex subset of a real normed space with nonempty interior is contained in the closure of the interior of this set, we conclude from the inequality (6.8)

$$\mu(f(x) + \alpha) + u(g(x) + y) \geq \gamma \geq \mu f(\bar{x}) \text{ for all } x \in \hat{S}, \alpha \geq 0, y \in C. \tag{6.9}$$

For $x = \bar{x}$ and $\alpha = 0$ it follows from the inequality (6.9)

$$u(y) \geq -u(g(\bar{x})) \text{ for all } y \in C. \tag{6.10}$$

With standard arguments we get immediately $u \in C^*$. For $y = 0_Y$ it follows from the inequality (6.10) $u(g(\bar{x})) \geq 0$. Because of $g(\bar{x}) \in -C$ and $u \in C^*$ we also have $u(g(\bar{x})) \leq 0$ which leads to

$$u(g(\bar{x})) = 0.$$

For $x = \bar{x}$ and $y = 0_Y$ we get from the inequality (6.9)

$$\mu\alpha \geq 0 \text{ for all } \alpha \geq 0$$

which implies $\mu \geq 0$. For the proof of $\mu > 0$ we assume that $\mu = 0$. Then it follows from the inequality (6.9) with $y = 0_Y$

$$u(g(x)) \geq 0 \text{ for all } x \in \hat{S}.$$

Because of the generalized Slater condition there is one $\hat{x} \in \hat{S}$ with $g(\hat{x}) \in -\text{int}(C)$, and then we have

$$u(g(\hat{x})) = 0.$$

Now we want to show that $u = 0_{Y^*}$. For that purpose we assume that $u \neq 0_{Y^*}$, i.e., there is one $y \in Y$ with $u(y) > 0$. Then we have

$$u(\lambda y + (1 - \lambda)\, g(\hat{x})) > 0 \text{ for all } \lambda \in (0, 1], \tag{6.11}$$

and because of $g(\hat{x}) \in -\text{int}(C)$ there is one $\bar{\lambda} \in (0, 1)$ with

$$\lambda y + (1 - \lambda)\, g(\hat{x}) \in -C \text{ for all } \lambda \in [0, \bar{\lambda}].$$

Then we get

$$u(\lambda y + (1-\lambda)\, g(\hat{x})) \le 0 \text{ for all } \lambda \in [0, \bar{\lambda}]$$

which contradicts the inequality (6.11). With the assumption $\mu = 0$ we also obtain $u = 0_{Y^*}$, a contradiction to $(\mu, u) \neq (0, 0_{Y^*})$. Consequently, we have $\mu \neq 0$ and therefore $\mu > 0$. Then we conclude from the inequality (6.9) with $\alpha = 0$ and $y = 0_Y$

$$\mu f(x) + u(g(x)) \ge \mu f(\bar{x}) \text{ for all } x \in \hat{S}$$

and

$$f(x) + \frac{1}{\mu} u(g(x)) \ge f(\bar{x}) \text{ for all } x \in \hat{S}.$$

If we define $\bar{u} := \frac{1}{\mu} u \in C^*$, we obtain with $\bar{u}(g(\bar{x})) = 0$

$$\inf_{x \in \hat{S}} f(x) + \bar{u}(g(x)) \ge f(\bar{x}) + \bar{u}(g(\bar{x})).$$

Hence we have

$$f(\bar{x}) + \bar{u}(g(\bar{x})) = \inf_{x \in \hat{S}} f(x) + \bar{u}(g(x)),$$

and with the weak duality theorem $\bar{u} \in C^*$ is a maximal solution of the dual problem (6.4). Obviously, the extremal values of the primal and dual problem are equal. □

In the following we discuss the practical importance of the strong duality theorem. If one wants to solve the primal problem and if one is interested in the minimal value in particular, then under suitable assumptions one can also solve the dual problem and determine the maximal value which is then equal to the minimal value of the primal problem. If the dual problem is simpler to solve than the primal problem, then this method is very useful.

6.3 Saddle Point Theorems

Relationships between the primal and the dual problem can also be described by a saddle point behavior of the Lagrange functional. These relationships will be investigated in this section.

First, we define the notion of the Lagrange functional which has already been mentioned in the context of the generalized Lagrange multiplier rule in Sect. 5.2.

Definition 6.9 (Lagrange functional).

Let the assumption (6.1) be satisfied. The functional $L : \hat{S} \times C^* \to \mathbb{R}$ with

$$L(x, u) = f(x) + u(g(x)) \text{ for all } x \in \hat{S} \text{ and all } u \in C^*$$

is called *Lagrange functional*.

Since we will investigate saddle points of the Lagrange functional L, we introduce the following notion.

Definition 6.10 (saddle point).

Let the assumption (6.1) be satisfied. A point $(\bar{x}, \bar{u}) \in \hat{S} \times C^*$ is called a *saddle point* of the Lagrange functional L if

$$L(\bar{x}, u) \leq L(\bar{x}, \bar{u}) \leq L(x, \bar{u}) \text{ for all } x \in \hat{S} \text{ and all } u \in C^*.$$

A saddle point of the Lagrange functional can be characterized by a "min sup = max inf" result which goes back to a known John von Neumann[11] saddle point theorem .

Theorem 6.11 (characterization of a saddle point).

Let the assumption (6.1) be satisfied. A point $(\bar{x}, \bar{u}) \in \hat{S} \times C^*$ is a saddle point of the Lagrange functional L if and only if

$$L(\bar{x}, \bar{u}) = \min_{x \in \hat{S}} \sup_{u \in C^*} L(x, u) = \max_{u \in C^*} \inf_{x \in \hat{S}} L(x, u). \tag{6.12}$$

Proof First we assume that the equation (6.12) is satisfied. Then we have with $\bar{x} \in \hat{S}$ and $\bar{u} \in C^*$

$$\sup_{u \in C^*} L(\bar{x}, u) = L(\bar{x}, \bar{u}) = \inf_{x \in \hat{S}} L(x, \bar{u}).$$

Hence $(\bar{x}, \bar{u})$ is a saddle point of the Lagrange functional L.

Next we assume that $(\bar{x}, \bar{u}) \in \hat{S} \times C^*$ is a saddle point of L. Then we obtain

$$\max_{u \in C^*} L(\bar{x}, u) = L(\bar{x}, \bar{u}) = \min_{x \in \hat{S}} L(x, \bar{u}). \tag{6.13}$$

[11] J. von Neumann, "Zur Theorie der Gesellschaftsspiele", *Math. Ann.* 100 (1928) 295–320.

For arbitrary $\hat{x} \in \hat{S}$ and $\hat{u} \in C^*$ we have

$$\inf_{x \in \hat{S}} L(x, \hat{u}) \leq L(\hat{x}, \hat{u}),$$

and therefore we conclude

$$\sup_{u \in C^*} \inf_{x \in \hat{S}} L(x, u) \leq \sup_{u \in C^*} L(\hat{x}, u)$$

and

$$\sup_{u \in C^*} \inf_{x \in \hat{S}} L(x, u) \leq \inf_{x \in \hat{S}} \sup_{u \in C^*} L(x, u).$$

With this inequality and the equation (6.13) it follows

$$\begin{aligned} L(\bar{x}, \bar{u}) &= \inf_{x \in \hat{S}} L(x, \bar{u}) \leq \sup_{u \in C^*} \inf_{x \in \hat{S}} L(x, u) \\ &\leq \inf_{x \in \hat{S}} \sup_{u \in C^*} L(x, u) \leq \sup_{u \in C^*} L(\bar{x}, u) \\ &= L(\bar{x}, \bar{u}). \end{aligned}$$

Consequently, we have

$$L(\bar{x}, \bar{u}) = \max_{u \in C^*} \inf_{x \in \hat{S}} L(x, u) = \min_{x \in \hat{S}} \sup_{u \in C^*} L(x, u)$$

which has to be shown. □

Using the preceding theorem we are able to present a relationship between a saddle point of the Lagrange functional and the solutions of the primal and dual problem.

Theorem 6.12 (characterization of a saddle point).

Let the assumption (6.1) be satisfied, and in addition, let the ordering cone C be closed. A point $(\bar{x}, \bar{u}) \in \hat{S} \times C^*$ is a saddle point of the Lagrange functional L if and only if $\bar{x}$ is a solution of the primal problem (6.2), $\bar{u}$ is a solution of the dual problem (6.4) and the extremal values of the two problems are equal.

Proof We assume that $(\bar{x}, \bar{u}) \in \hat{S} \times C^*$ is a saddle point of the Lagrange functional L. By Theorem 6.11 we then have

$$L(\bar{x}, \bar{u}) = \min_{x \in \hat{S}} \sup_{u \in C^*} L(x, u) = \max_{u \in C^*} \inf_{x \in \hat{S}} L(x, u).$$

Consequently, $\bar{x}$ is a minimal solution of the problem (6.3) and with Lemma 6.1 $\bar{x}$ is then also a minimal solution of the primal problem (6.2). Moreover, $\bar{u}$ is a maximal solution of the dual problem (6.4) and the extremal values of the primal and dual problem are equal.

Next, we assume that $\bar{x}$ is a minimal solution of the primal problem (6.2), $\bar{u}$ is a maximal solution of the dual problem (6.4) and the extremal values of the two problems are equal. Then we have

$$\lambda := \inf_{x \in \hat{S}} L(x, \bar{u}) = \max_{u \in C^*} \inf_{x \in \hat{S}} L(x, u),$$

and with Lemma 6.1 we get

$$f(\bar{x}) = \sup_{u \in C^*} L(\bar{x}, u) = \min_{x \in \hat{S}} \sup_{u \in C^*} L(x, u).$$

Because of $\lambda = f(\bar{x})$ we obtain

$$\bar{u}(g(\bar{x})) \geq -f(\bar{x}) + \inf_{x \in \hat{S}} f(x) + \bar{u}(g(x)) = -f(\bar{x}) + \lambda = 0$$

and because of $g(\bar{x}) \in -C$, $\bar{u} \in C^*$ we have $\bar{u}(g(\bar{x})) \leq 0$ resulting in $\bar{u}(g(\bar{x})) = 0$ which implies $f(\bar{x}) = L(\bar{x}, \bar{u})$. Then it follows

$$L(\bar{x}, \bar{u}) = \min_{x \in \hat{S}} \sup_{u \in C^*} L(x, u) = \max_{u \in C^*} \inf_{x \in \hat{S}} L(x, u),$$

and by Theorem 6.11 it follows that $(\bar{x}, \bar{u})$ is a saddle point of the Lagrange functional L. □

With the aid of the strong duality theorem we also present a sufficient condition for the existence of a saddle point of the Lagrange functional.

Corollary 6.13 (sufficient condition for the existence of a saddle point).

Let the assumption (6.1) be satisfied, and in addition, let the ordering cone C be closed, let C have a nonempty interior int(C) and let the composite mapping $(f, g) : \hat{S} \to \mathbb{R} \times Y$ be convex-like (with respect to the product cone $\mathbb{R}_+ \times C$ in $\mathbb{R} \times Y$). If $\bar{x} \in S$ is a minimal solution of the primal *problem (6.2) and the generalized Slater condition is satisfied, i.e., there is one $\hat{x} \in \hat{S}$ with $g(\hat{x}) \in -int(C)$, then there is a $\bar{u} \in C^*$ so that $(\bar{x}, \bar{u})$ is a saddle point of the Lagrange functional.*

Proof If $\bar{x} \in S$ is a minimal solution of the primal problem then, by Theorem 6.8, there is a maximal solution $\bar{u} \in C^*$ of the dual problem and the extremal values of the two problems are equal. Consequently, by Theorem 6.12, $(\bar{x}, \bar{u})$ is a saddle point of the Lagrange functional. □

The preceding corollary can also be proved directly without the assumption that the ordering cone is closed.

6.4 Linear Problems

An excellent application of the duality theory can be given for linear optimization problems because the dual problem of a linear minimization problem is equivalent to a linear maximization problem. It is the aim of this section to transform this dual problem in an appropriate way so that one gets a problem formulation which is useful from the point of view of the applications.

In the following we specialize the problem (6.2). For that purpose we need the following assumption:

Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be partially ordered real normed spaces with the ordering cones C_X and C_Y, respectively;
let $c \in X^*$ be a continuous linear functional;
let $A : X \to Y$ be a continuous linear mapping;
let $b \in Y$ be a given element;
let the constraint set $S := \{x \in C_X \mid A(x) - b \in C_Y\}$ be nonempty. (6.14)

Under this assumption we consider the primal problem

$$\begin{aligned} &\min \; c(x) \\ &\text{subject to the constraints} \\ &A(x) - b \in C_Y \\ &x \in C_X. \end{aligned} \tag{6.15}$$

In the problem formulation (6.2) we have replaced the objective functional f by the continuous linear functional c and the constraint mapping g by $b - A(\cdot)$. The set $\hat{S}$ equals the ordering cone C_X. Notice that under the assumption (6.14) the composite mapping $(c(\cdot), b - A(\cdot)) : C_X \to \mathbb{R} \times Y$ is also convex-like.

In this case the dual problem reads (by (6.4))

$$\max_{u \in C_Y^*} \inf_{x \in C_X} c(x) + u(b - A(x)).$$

This problem is equivalent to the problem (compare (6.5))

$$\begin{gathered} \max \ \lambda \\ \text{subject to the constraints} \\ c(x) + u(b - A(x)) \geq \lambda \text{ for all } x \in C_X \\ \lambda \in \mathbb{R}, \ u \in C_Y^*. \end{gathered} \tag{6.16}$$

If we define the constraint set of the problem (6.16) as

$$S^* := \{(\lambda, u) \in \mathbb{R} \times C_Y^* \mid c(x) + u(b - A(x)) \geq \lambda \text{ for all } x \in C_X\}, \tag{6.17}$$

then we can reformulate this constraint set using the following lemma.

Lemma 6.14 (reformulation of the constraint set S^*).

Let the assumption (6.14) be satisfied, and let the set S^* be given by (6.17). Then it follows

$$S^* = \{(\lambda, u) \in \mathbb{R} \times C_Y^* \mid c - A^*(u) \in C_X^* \text{ and } \lambda \leq u(b)\}$$

(C_X^* and C_Y^* denote the dual cone of C_X and C_Y, respectively; $A^* : Y^* \to X^*$ denotes the adjoint mapping of A).

Proof First we assume that a pair $(\lambda, u) \in S^*$ is given arbitrarily. Then it follows

$$c(x) + u(b - A(x)) \geq \lambda \text{ for all } x \in C_X$$

and

$$(c - u \circ A)(x) \geq \lambda - u(b) \text{ for all } x \in C_X. \tag{6.18}$$

For $x = 0_X$ we get especially

$$\lambda \leq u(b).$$

From the inequality (6.18) we also obtain

$$(c - u \circ A)(x) \geq 0 \text{ for all } x \in C_X$$

(because the assumption that $(c - u \circ A)(x) < 0$ for some $x \in C_X$ leads to a contradiction to the inequality (6.18)). Consequently we have

$$c - u \circ A \in C_X^*$$

resulting in

$$c - A^*(u) \in C_X^*.$$

This proves the first part of the assertion.

Next, we choose an arbitrary pair $(\lambda, u) \in \mathbb{R} \times C_Y^*$ with $c - A^*(u) \in C_X^*$ and $\lambda \leq u(b)$. Then we conclude

$$(c - u \circ A)(x) \geq 0 \geq \lambda - u(b) \text{ for all } x \in C_X,$$

and therefore it follows $(\lambda, u) \in S^*$. □

With Lemma 6.14 the equivalent dual problem (6.16) is also equivalent to the problem

$$\begin{gathered} \max\ \lambda \\ \text{subject to the constraints} \\ c - A^*(u) \in C_X^* \\ \lambda \leq u(b) \\ \lambda \in \mathbb{R},\ \ u \in C_Y^*. \end{gathered}$$

Because of the second constraint this problem is again equivalent to the problem

$$\begin{gathered} \max\ u(b) \\ \text{subject to the constraints} \\ c - A^*(u) \in C_X^* \\ u \in C_Y^*. \end{gathered} \tag{6.19}$$

The problem (6.19) generalizes the dual optimization problem known from linear programming.

Example 6.15 (duality in linear programming).

Let $A_{11} \in \mathbb{R}^{(m_1,n_1)}$, $A_{12} \in \mathbb{R}^{(m_1,n_2)}$, $A_{21} \in \mathbb{R}^{(m_2,n_1)}$ and $A_{22} \in \mathbb{R}^{(m_2,n_2)}$ with $n_1, n_2, m_1, m_2 \in \mathbb{N}$ be given matrices, and let $b_1 \in \mathbb{R}^{m_1}$, $b_2 \in \mathbb{R}^{m_2}$,

$c_1 \in \mathbb{R}^{n_1}$ and $c_2 \in \mathbb{R}^{n_2}$ be given vectors. Consider the finite dimensional linear optimization problem

$$\begin{gathered} \min\ c_1^T x_1 + c_2^T x_2 \\ \text{subject to the constraints} \\ A_{11}x_1 + A_{12}x_2 = b_1 \\ A_{21}x_1 + A_{22}x_2 \geq b_2 \\ x_1 \geq 0_{\mathbb{R}^{n_1}},\ x_2 \in \mathbb{R}^{n_2}. \end{gathered}$$

"$\geq$" (and "$\leq$" in the dual problem) has to be understood in a componentwise sense. If we set $X := \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$, $Y := \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$, $C_X := \mathbb{R}^{n_1}_+ \times \mathbb{R}^{n_2}$ and $C_Y := \{0_{\mathbb{R}^{m_1}}\} \times \mathbb{R}^{m_2}_+$, then it is evident that the dual problem (6.19) can be written as

$$\begin{gathered} \max\ b_1^T u_1 + b_2^T u_2 \\ \text{subject to the constraints} \\ A_{11}^T u_1 + A_{21}^T u_2 \leq c_1 \\ A_{12}^T u_1 + A_{22}^T u_2 = c_2 \\ u_1 \in \mathbb{R}^{m_1},\ u_2 \geq 0_{\mathbb{R}^{m_2}}. \end{gathered}$$

So, we obtain the dual problem known from linear programming.

Since the equivalent dual problem (6.19) is also a linear optimization problem, one can again formulate a dual problem of this dual one. If one assumes in addition that X is reflexive and the ordering cones C_X and C_Y are closed, one can show that by double dualization one comes back to the primal problem.

6.5 Application to Approximation Problems

In this section we investigate a special linear optimization problem. This is a problem of the linear Chebyshev approximation. For this approximation problem we formulate the dual problem which we transform in an appropriate way. Moreover, with the aid of the duality theory we prove an alternation theorem of the linear Chebyshev approximation.

First we formulate the assumptions of this section:

$$\left.\begin{array}{l}\text{Let } M \text{ be a compact metric space;}\\ \text{let } C(M) \text{ denote the linear space of continuous real-valued functions}\\ \text{on } M \text{ equipped with the maximum norm } \|\cdot\| \text{ where}\\ \|x\| = \max\limits_{t\in M} |x(t)| \text{ for all } x \in C(M);\\ \text{let } v_1, \ldots, v_n, \hat{v} \in C(M) \text{ be given functions.}\end{array}\right\} \tag{6.20}$$

Under this assumption we investigate the following problem of linear Chebyshev approximation :

$$\min_{x\in\mathbb{R}^n} \left\| \hat{v} - \sum_{i=1}^{n} x_i v_i \right\|. \tag{6.21}$$

Hence we are looking for a linear combination of the functions $v_1, \ldots, v_n$ which uniformly approximates the function $\hat{v}$ in the best possible way. The problem (6.21) is equivalent to the problem

$$\begin{gathered}\min\ \lambda\\ \text{subject to the constraints}\\ \left\| \hat{v} - \sum_{i=1}^{n} x_i v_i \right\| \le \lambda\\ \lambda \in \mathbb{R},\ x \in \mathbb{R}^n\end{gathered}$$

which can also be written as:

$$\begin{gathered}\min\ \lambda\\ \text{subject to the constraints}\\ \left.\begin{array}{l}\lambda + \sum\limits_{i=1}^{n} x_i v_i(t) \ge \hat{v}(t)\\[2ex] \lambda - \sum\limits_{i=1}^{n} x_i v_i(t) \ge -\hat{v}(t)\end{array}\right\} \text{ for all } t \in M\\ \lambda \in \mathbb{R},\ x \in \mathbb{R}^n.\end{gathered} \tag{6.22}$$

If M contains infinitely many elements, then the problem (6.22) is a semi-infinite optimization problem . A problem of this type is discussed in Example 1.5.

Question: What is the dual problem to (6.22)?

In order to answer this question we introduce some notations: $X := \mathbb{R}^{n+1}$; $C_X := \mathbb{R}^{n+1}$; let E denote the finite dimensional linear subspace of $C(M)$ spanned by the functions $v_1, \ldots, v_n, \hat{v}, e$ (where $e \in C(M)$ with $e(t) = 1$

for all $t \in M$); $Y := E \times E$; and $C_Y := \{(f_1, f_2) \in Y \mid f_1(t) \geq 0 \text{ and } f_2(t) \geq 0 \text{ for all } t \in M\}$. If we define $c := (1, 0, \ldots, 0) \in \mathbb{R}^{n+1}$, $b := (\hat{v}, -\hat{v}) \in Y$ and the mapping $A : X \to Y$ with

$$A(\lambda, x) = \begin{pmatrix} \lambda e + \sum_{i=1}^{n} x_i v_i \\ \\ \lambda e - \sum_{i=1}^{n} x_i v_i \end{pmatrix} \text{ for all } (\lambda, x) \in \mathbb{R}^{n+1},$$

then the problem (6.22) can also be written as follows:

$$\begin{gathered} \min \; c^T(\lambda, x) \\ \text{subject to the constraints} \\ A(\lambda, x) - b \in C_Y \\ (\lambda, x) \in C_X. \end{gathered} \tag{6.23}$$

This is a linear optimization problem which was already discussed in the preceding section. For the formulation of the equivalent dual problem (by (6.19)) we need the adjoint mapping A^* of A, among other things. The mapping $A^* : Y^* \to X^*$ ($= \mathbb{R}^{n+1}$) is defined by

$$\begin{aligned} A^*(u_1, u_2)\,(\lambda, x) &= (u_1, u_2)\,(A(\lambda, x)) \\ &= u_1\left(\lambda e + \sum_{i=1}^{n} x_i v_i\right) + u_2\left(\lambda e - \sum_{i=1}^{n} x_i v_i\right) \text{ for all } (\lambda, x) \in \mathbb{R}^{n+1}. \end{aligned}$$

The statement

$$c - A^*(u_1, u_2) \in C_X^*$$

is equivalent to

$$\lambda - u_1\left(\lambda e + \sum_{i=1}^{n} x_i v_i\right) - u_2\left(\lambda e - \sum_{i=1}^{n} x_i v_i\right) = 0 \text{ for all } (\lambda, x) \in \mathbb{R}^{n+1}$$

resulting in

$$\lambda(1 - u_1(e) - u_2(e)) + \sum_{i=1}^{n} x_i(u_2(v_i) - u_1(v_i)) = 0 \text{ for all } (\lambda, x) \in \mathbb{R}^{n+1}.$$

This equation is also equivalent to

$$u_1(v_i) - u_2(v_i) = 0 \text{ for all } i \in \{1, \ldots, n\}$$

and

$$u_1(e) + u_2(e) = 1.$$

Consequently, the equivalent dual problem (by (6.19)) which is associated to the problem (6.23) reads as follows:

$$\begin{array}{c} \max\ u_1(\hat{v}) - u_2(\hat{v}) \\ \text{subject to the constraints} \\ u_1(v_i) - u_2(v_i) = 0 \text{ for all } i \in \{1, \ldots, n\} \\ u_1(e) + u_2(e) = 1 \\ (u_1, u_2) \in C_Y^*. \end{array} \tag{6.24}$$

This problem is also a semi-infinite optimization problem which has finitely many constraints in the form of equalities. With the following representation theorem for positive linear forms on $C(M)$ the problem (6.24) can be simplified essentially. A proof of this representation theorem can be found in the book [211, p. 184] by Krabs.

Theorem 6.16 (representation theorem).

Let F be a finite dimensional linear subspace of $C(M)$ (compare (6.20)) spanned by functions $f_1, \ldots, f_m \in C(M)$. Let F be partially ordered in a natural way, and assume that there is a function $\tilde{f} \in F$ with

$$\tilde{f}(t) > 0 \text{ for all } t \in M.$$

Then every continuous linear functional $l \in C_F^*$ (dual cone in F^*) can be represented as

$$l(f) = \sum_{j=1}^{k} \lambda_j f(t_j) \text{ for all } f \in F$$

where $k \in \mathbb{N}$, $t_1, \ldots, t_k \in M$ are different points, and $\lambda_1, \ldots, \lambda_k$ are nonnegative real numbers.

Now we apply this theorem to the linear subspace E. Since $e \in E$ with

$$e(t) = 1 > 0 \text{ for all } t \in M,$$

all assumptions of Theorem 6.16 are fulfilled, and therefore we obtain the following representations for $u_1, u_2 \in C_E^*$ (dual cone in E^*)

$$u_1(v) = \sum_{j=1}^{k_1} \lambda_{1_j} v(t_{1_j}) \text{ for all } v \in E$$

and

$$u_2(v) = \sum_{j=1}^{k_2} \lambda_{2_j} v(t_{2_j}) \text{ for all } v \in E.$$

Here we have $k_1, k_2 \in \mathbb{N}$; $t_{1_1}, \dots, t_{1_{k_1}} \in M$ are different points; $t_{2_1}, \dots, t_{2_{k_2}} \in M$ are different points; and it is $\lambda_{1_1}, \dots, \lambda_{1_{k_1}}, \lambda_{2_1}, \dots, \lambda_{2_{k_2}} \geq 0$.

Consequently, the problem (6.24) is equivalent to the following problem:

$$\begin{gathered}
\max \sum_{j=1}^{k_1} \lambda_{1_j} \hat{v}(t_{1_j}) - \sum_{j=1}^{k_2} \lambda_{2_j} \hat{v}(t_{2_j}) \\
\text{subject to the constraints} \\
\sum_{j=1}^{k_1} \lambda_{1_j} v_i(t_{1_j}) - \sum_{j=1}^{k_2} \lambda_{2_j} v_i(t_{2_j}) = 0 \text{ for all } i \in \{1, \dots, n\} \\
\sum_{j=1}^{k_1} \lambda_{1_j} + \sum_{j=1}^{k_2} \lambda_{2_j} = 1 \\
\lambda_{1_1}, \dots, \lambda_{1_{k_1}}, \lambda_{2_1}, \dots, \lambda_{2_{k_2}} \geq 0 \\
t_{1_1}, \dots, t_{1_{k_1}} \in M \\
t_{2_1}, \dots, t_{2_{k_2}} \in M.
\end{gathered} \tag{6.25}$$

Before simplifying this problem we discuss the question of solvability.

Theorem 6.17 (solvability of problem (6.25)).

Let the assumption (6.20) be satisfied. Then the optimization problem (6.25) has at least one maximal solution $(\lambda_{1_1}, \dots, \lambda_{1_{k_1}}, \lambda_{2_1}, \dots, \lambda_{2_{k_2}}, t_{1_1}, \dots, t_{1_{k_1}}, t_{2_1}, \dots, t_{2_{k_2}})$, *and the extremal value of this problem equals the extremal value of the problem (6.21).*

Proof By Theorem 2.18 the problem (6.21) of linear Chebyshev approximation is solvable. Then the equivalent linear optimization problem (6.23) is also solvable.

The ordering cone C_Y has a nonempty interior; and the generalized Slater condition is satisfied, because for an arbitrary $\hat{x} \in \mathbb{R}^n$ we obtain with

$$\hat{\lambda} := \left\| \hat{v} - \sum_{i=1}^{n} \hat{x}_i v_i \right\| + 1$$

also $b - A(\hat{\lambda}, \hat{x}) \in -\text{int}(C)$. Then by Theorem 6.8 the problem (6.24) which is equivalent to the dual problem of (6.23) is also solvable. With the preceding remarks this problem is also equivalent to the maximization problem (6.25) which has therefore a solution. Finally, we conclude with Theorem 6.8 that the extremal values of the corresponding problems are equal. □

The maximization problem (6.25) is a finite optimization problem with finitely many variables and finitely many constraints. But it is unwieldy because k_1 and k_2 are not known. One can show that $k_1 + k_2 \leq n + 1$ (we refer to Krabs [211, p. 54]). But even if we restrict the number of variables in the maximization problem (6.25) in this way, this problem is a finite nonlinear optimization problem which, from a numerical point of view, is not easier to solve than the original problem of linear Chebyshev approximation.

Finally, we formulate a so-called *alternation theorem* for the investigated problem of linear Chebyshev approximation .

Theorem 6.18 (alternation theorem).

Let the assumption (6.20) be satisfied. A vector $\bar{x} \in \mathbb{R}^n$ is a solution of the problem (6.21) of linear Chebyshev approximation (i.e., $\sum_{i=1}^{n} \bar{x}_i v_i$ is a best approximation to $\hat{v}$ in E) if and only if there are $k \leq n + 1$ different points $t_1, \ldots, t_k \in M$ with

$$\left| \hat{v}(t_j) - \sum_{i=1}^{n} \bar{x}_i v_i(t_j) \right| = \left\| \hat{v} - \sum_{i=1}^{n} \bar{x}_i v_i \right\| \text{ for all } j = 1, \ldots, k \tag{6.26}$$

and there are numbers $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ *with*

$$\sum_{j=1}^{k} |\lambda_j| = 1, \tag{6.27}$$

$$\sum_{j=1}^{k} \lambda_j v_i(t_j) = 0 \text{ for all } i = 1, \ldots, n, \tag{6.28}$$

$$\lambda_j \neq 0 \textit{ for } j = 1, \ldots, k \Rightarrow \hat{v}(t_j) - \sum_{i=1}^{n} \bar{x}_i v_i(t_j) = \left\| \hat{v} - \sum_{i=1}^{n} \bar{x}_i v_i \right\| \textit{sgn}(\lambda_j). \tag{6.29}$$

Proof First we assume that for some $\bar{x} \in \mathbb{R}^n$ there are $k \leq n+1$ different points $t_1, \ldots, t_k \in M$ so that the conditions (6.26), (6.27), (6.28) and (6.29) are satisfied. Then we obtain for every $x \in \mathbb{R}^n$

$$\begin{aligned}
\left\| \hat{v} - \sum_{i=1}^{n} \bar{x}_i v_i \right\| &= \sum_{j=1}^{k} |\lambda_j| \left\| \hat{v} - \sum_{i=1}^{n} \bar{x}_i v_i \right\| && \text{(by (6.27))} \\
&= \sum_{j=1}^{k} |\lambda_j| \operatorname{sgn}(\lambda_j) \left(\hat{v}(t_j) - \sum_{i=1}^{n} \bar{x}_i v_i(t_j) \right) && \text{(by (6.29))} \\
&= \sum_{j=1}^{k} \lambda_j \hat{v}(t_j) - \sum_{i=1}^{n} \bar{x}_i \sum_{j=1}^{k} \lambda_j v_i(t_j) \\
&= \sum_{j=1}^{k} \lambda_j \hat{v}(t_j) && \text{(by (6.28))} \\
&= \sum_{j=1}^{k} \lambda_j \hat{v}(t_j) - \sum_{i=1}^{n} x_i \sum_{j=1}^{k} \lambda_j v_i(t_j) && \text{(by (6.28))} \\
&= \sum_{j=1}^{k} \lambda_j \left(\hat{v}(t_j) - \sum_{i=1}^{n} x_i v_i(t_j) \right) \\
&\leq \sum_{j=1}^{k} |\lambda_j| \left\| \hat{v} - \sum_{i=1}^{n} x_i v_i \right\| \\
&= \left\| \hat{v} - \sum_{i=1}^{n} x_i v_i \right\| && \text{(by (6.27))}.
\end{aligned}$$

Consequently, $\bar{x}$ is a solution of the problem (6.21) of the linear Chebyshev approximation.

Next, we assume that $\bar{x} \in \mathbb{R}^n$ solves the problem (6.21). By Theorem 6.17 the optimization problem (6.25) has a maximal solution $(\lambda_{1_1}, \ldots, \lambda_{1_{k_1}}, \lambda_{2_1}, \ldots, \lambda_{2_{k_2}}, t_{1_1}, \ldots, t_{1_{k_1}}, t_{2_1}, \ldots, t_{2_{k_2}})$ (with positive $\lambda_{1_1}, \ldots, \lambda_{1_{k_1}}, \lambda_{2_1}, \ldots, \lambda_{2_{k_2}}$, otherwise, if $\lambda_{i_j} = 0$ for some $i \in \{1, 2\}$ and some $j \in \{1, \ldots, k_i\}$, we can drop the variable λ_{i_j} together with the point t_{i_j} without changing the minimal value of the problem (6.25)), and the extremal values of the two problems are equal, i.e.

$$\beta := \left\| \hat{v} - \sum_{i=1}^{n} \bar{x}_i v_i \right\| = \sum_{j=1}^{k_1} \lambda_{1_j} \hat{v}(t_{1_j}) - \sum_{j=1}^{k_2} \lambda_{2_j} \hat{v}(t_{2_j}). \tag{6.30}$$

Because of the constraint

$$\sum_{j=1}^{k_1} \lambda_{1_j} v_i(t_{1_j}) - \sum_{j=1}^{k_2} \lambda_{2_j} v_i(t_{2_j}) = 0 \text{ for all } i \in \{1, \ldots, n\} \tag{6.31}$$

it follows

$$\sum_{j=1}^{k_1} \lambda_{1_j} \sum_{i=1}^{n} \bar{x}_i v_i(t_{1_j}) - \sum_{j=1}^{k_2} \lambda_{2_j} \sum_{i=1}^{n} \bar{x}_i v_i(t_{2_j}) = 0, \tag{6.32}$$

and with the constraint

$$\sum_{j=1}^{k_1} \lambda_{1_j} + \sum_{j=1}^{k_2} \lambda_{2_j} = 1 \tag{6.33}$$

and the equations (6.32) and (6.30) we conclude

$$\begin{aligned}
&\sum_{j=1}^{k_1} \lambda_{1_j} \left[-\hat{v}(t_{1_j}) + \sum_{i=1}^{n} \bar{x}_i v_i(t_{1_j}) + \beta \right] + \sum_{j=1}^{k_2} \lambda_{2_j} \left[\hat{v}(t_{2_j}) - \sum_{i=1}^{n} \bar{x}_i v_i(t_{2_j}) + \beta \right] \\
&= - \sum_{j=1}^{k_1} \lambda_{1_j} \hat{v}(t_{1_j}) + \sum_{j=1}^{k_2} \lambda_{2_j} \hat{v}(t_{2_j}) + \beta \left(\sum_{j=1}^{k_1} \lambda_{1_j} + \sum_{j=1}^{k_2} \lambda_{2_j} \right) \\
&= 0.
\end{aligned}$$

Then the following equations are satisfied:

$$\hat{v}(t_{1_j}) - \sum_{i=1}^{n} \bar{x}_i v_i(t_{1_j}) = \beta \text{ for all } j \in \{1, \ldots, k_1\},$$

$$\hat{v}(t_{2_j}) - \sum_{i=1}^{n} \bar{x}_i v_i(t_{2_j}) = -\beta \text{ for all } j \in \{1, \ldots, k_2\}.$$

If we define the variables

$$\mu_j := \lambda_{1_j} \text{ for } j = 1, \ldots, k_1,$$
$$s_j := t_{1_j} \text{ for } j = 1, \ldots, k_1$$

and

$$\mu_{k_1+j} := -\lambda_{2_j} \text{ for } j = 1, \ldots, k_2,$$
$$s_{k_1+j} := t_{2_j} \text{ for } j = 1, \ldots, k_2,$$

we get with the equation (6.33)

$$\sum_{j=1}^{k_1+k_2} |\mu_j| = 1,$$

and with the equation (6.31) it follows

$$\sum_{j=1}^{k_1+k_2} \mu_j v_i(s_j) = 0 \text{ for all } i = 1, \ldots, n.$$

Moreover, the following equation is satisfied:

$$\hat{v}(s_j) - \sum_{i=1}^{n} \bar{x}_i v_i(s_j) = \left\| \hat{v} - \sum_{i=1}^{n} \bar{x}_i v_i \right\| \operatorname{sgn}(\mu_j) \text{ for all } j \in \{1, \ldots, k_1 + k_2\}.$$

If we notice that $k_1 + k_2 \leq n + 1$, then the assertion follows immediately. □

Example 6.19 (determination of a best approximation).

We consider again Example 1.5 and ask for a solution of the problem

$$\min_{x\in\mathbb{R}} \max_{t\in[0,2]} |\sinh t - xt|.$$

By the alternation theorem the necessary and sufficient conditions for a minimal solution $\bar{x} \in \mathbb{R}$ of this problem read as follows:

$$\begin{aligned}
|\lambda_1| + |\lambda_2| &= 1\\
\lambda_1 t_1 + \lambda_2 t_2 &= 0\\
\lambda_1 \neq 0 \ \Rightarrow\ \sinh t_1 - \bar{x}t_1 &= \|\sinh - \bar{x}\,\mathrm{id}\|\,\mathrm{sgn}(\lambda_1)\\
\lambda_2 \neq 0 \ \Rightarrow\ \sinh t_2 - \bar{x}t_2 &= \|\sinh - \bar{x}\,\mathrm{id}\|\,\mathrm{sgn}(\lambda_2)\\
|\sinh t_1 - \bar{x}t_1| &= \|\sinh - \bar{x}\,\mathrm{id}\|\\
|\sinh t_2 - \bar{x}t_2| &= \|\sinh - \bar{x}\,\mathrm{id}\|\\
\lambda_1, \lambda_2 \in \mathbb{R};\ & t_1, t_2 \in [0, 2].
\end{aligned}$$

One obtains from these conditions that $\bar{x}$ is a minimal solution of the considered approximation problem if and only if $\bar{x} \approx 1.600233$ (see Fig. 1.5).

Exercises

(6.1) Determine a maximal solution of the dual problem associated to the primal problem

$$\begin{gathered}
\min\ x_1 + 2(x_2 - 1)^2\\
\text{subject to the constraint}\\
-x_1 - x_2 + 1 \leq 0\\
x_1, x_2 \in \mathbb{R}.
\end{gathered}$$

(6.2) Let the following primal minimization problem be given:

$$\min\ 2\alpha + \int_0^1 t\, x(t)\, dt$$

subject to the constraints

$$1 - \alpha - \int_t^1 x(s)\, ds \le 0 \ \text{ almost everywhere on } [0, 1]$$

$$x(t) \ge 0 \ \text{ almost everywhere on } [0, 1]$$

$$\alpha \ge 0$$

$$x \in L_2[0, 1],\ \alpha \in \mathbb{R}.$$

(a) Formulate the equivalent dual problem (6.5) of this minimization problem.
(b) Show that the minimal value of this problem is 2 and that the maximal value of the dual problem (6.4) is 1. Consequently, there is a duality gap.

(6.3) Consider the problem (6.21) of the linear Chebyshev approximation with $M = [0, 1]$, $\hat{v}(t) = t^2$ for all $t \in [0, 1]$, $n = 1$, $v_1(t) = t$ for all $t \in [0, 1]$. With the aid of the alternation theorem (Theorem 6.18) determine a solution of this problem.

Application to Extended Semidefinite Optimization

7

In semidefinite optimization one investigates nonlinear optimization problems in finite dimensions with a constraint requiring that a certain matrix-valued function is negative semidefinite. This type of problems arises in convex optimization, approximation theory, control theory, combinatorial optimization and engineering. In system and control theory so-called linear matrix inequalities (LMI's) and extensions like bilinear matrix inequalities (BMI's) fit into this class of constraints. Our investigations include various partial orderings for the description of the matrix constraint and in this way we extend the standard semidefinite case to other types of constraints. We apply the theory on optimality conditions developed in Chap. 5 and the duality theory of Chap. 6 to these extended semidefinite optimization problems.

7.1 Löwner Ordering Cone and Extensions

In the so-called *conic optimization* one investigates finite dimensional optimization problems with an inequality constraint with respect to a special matrix space. To be more specific, let $\mathcal{S}^n$ denote the real linear space of symmetric (n, n)-matrices. It is obvious that this space is a finite dimensional Hilbert space with the scalar product $\langle \cdot, \cdot \rangle$ defined by

$$\langle A, B \rangle = \text{trace}(A \cdot B) \;\text{ for all }\; A, B \in \mathcal{S}^n. \tag{7.1}$$

Recall that the trace of a matrix is defined as sum of all diagonal elements of the matrix. Let C be a convex cone in $\mathcal{S}^n$ inducing a partial ordering $\preccurlyeq$. Then we consider a matrix function $G : \mathbb{R}^m \to \mathcal{S}^n$ defining the inequality constraint

$$G(x) \preccurlyeq 0_{\mathcal{S}^n}. \tag{7.2}$$

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3_7

If $f : \mathbb{R}^m \to \mathbb{R}$ denotes a given objective function, then we obtain the *conic optimization problem*

$$\begin{gathered} \min f(x) \\ \text{subject to the constraints} \\ G(x) \preccurlyeq 0_{\mathcal{S}^n} \\ x \in \mathbb{R}^m. \end{gathered} \tag{7.3}$$

The name of this problem comes from the fact that the matrix inequality has to be interpreted using the ordering cone C. Obviously, the theory developed in this book is fully applicable to this problem structure.

In the special literature one often investigates problems of the form

$$\begin{gathered} \min \hat{f}(X) \\ \text{subject to the constraints} \\ \hat{G}(X) \preccurlyeq 0_{\mathcal{S}^n} \\ X \in \mathcal{S}^p \end{gathered} \tag{7.4}$$

with given functions $\hat{f} : \mathcal{S}^p \to \mathbb{R}$ and $\hat{G} : \mathcal{S}^p \to \mathcal{S}^n$. In this case the matrix $X \in \mathcal{S}^p$ can be transformed to a vector $x \in \mathbb{R}^{p \cdot p}$ where x consists of all columns of X by stacking up columns of X from the first to the p-th column. The dimension can be reduced because X is symmetric. Then we obtain $x \in \mathbb{R}^{\frac{p(p+1)}{2}}$. If φ denotes the transformation from the vector x to the matrix X, then the problem (7.4) can be written as

$$\begin{gathered} \min\ (\hat{f} \circ \varphi)(x) \\ \text{subject to the constraints} \\ (\hat{G} \circ \varphi)(x) \preccurlyeq 0_{\mathcal{S}^n} \\ x \in \mathbb{R}^{\frac{p(p+1)}{2}}. \end{gathered}$$

Hence, the optimization problem is of the form of problem (7.3) and it is not necessary to study the nonlinear optimization problem (7.4) separately.

In practice, one works with special ordering cones for the Hilbert space $\mathcal{S}^n$. The Löwner[12] ordering cone and further cones are discussed now.

[12] K. Löwner, "Über monotone Matrixfunktionen", *Mathematische Zeitschrift* 38 (1934) 177–216.

Remark 7.1 (ordering cones in $\mathcal{S}^n$).

Let $\mathcal{S}^n$ denote the real linear space of symmetric (n, n) matrices.

(a) The convex cone

$$\mathcal{S}^n_+ := \{X \in \mathcal{S}^n \mid X \text{ is positive semidefinite}\}$$

is called the *Löwner ordering cone*.
The partial ordering induced by the convex cone $\mathcal{S}^n_+$ is also called *Löwner partial ordering* $\preceq$ (notice that we use the special symbol $\preceq$ for this partial ordering). The problem (7.3) equipped with the Löwner partial ordering is then called a *semidefinite optimization problem*. The name of this problem is caused by the fact that the inequality constraint means that the matrix $G(x)$ has to be negative semidefinite.
Although the semidefinite optimization problem is only a finite dimensional problem, it is not a usual problem in $\mathbb{R}^m$ because the Löwner partial ordering makes the inequality constraint complicated. In fact, the inequality (7.2) is equivalent to infinitely many inequalities of the form

$$y^T G(x) y \leq 0 \text{ for all } y \in \mathbb{R}^n.$$

(b) The *K-copositive ordering cone* is defined by

$$C^n_K := \{X \in \mathcal{S}^n \mid y^T X y \geq 0 \text{ for all } y \in K\}$$

for a given convex cone $K \subset \mathbb{R}^n$, i.e., we consider only matrices for which the quadratic form is nonnegative on the convex cone K. If the partial ordering induced by this convex cone is used in problem (7.3), then we speak of a *K-copositive optimization problem*.
It is evident that $\mathcal{S}^n_+ \subset C^n_K$ for every convex cone K and $\mathcal{S}^n_+ = C^n_{\mathbb{R}^n}$. Therefore, we have for the dual cones $(C^n_K)^* \subset (\mathcal{S}^n_+)^*$.
If K equals the positive orthant $\mathbb{R}^n_+$, then $C^n_{\mathbb{R}^n_+}$ is simply called *copositive ordering cone* and the problem (7.3) is then called *copositive optimization problem*.

(c) The *nonnegative ordering cone* is defined by

$$N^n := \{X \in \mathcal{S}^n \mid X_{ij} \geq 0 \text{ for all } i, j \in \{1, \ldots, n\}\}.$$

In this case the optimization problem (7.3) with the partial ordering induced by the convex cone N^n reduces to a standard optimization problem of the form

$$\begin{gathered} \min f(x) \\ \text{subject to the constraints} \\ G_{ij}(x) \leq 0 \;\text{ for all }\; i, j \in \{1, \ldots, n\} \\ x \in \mathbb{R}^m. \end{gathered}$$

The number of constraints can actually be reduced to $\frac{n(n+1)}{2}$ because the matrix $G(x)$ is assumed to be symmetric. So, such a problem can be investigated with the standard theory of nonlinear optimization in finite dimensions.

(d) The *doubly nonnegative ordering cone* is defined by

$$\begin{aligned} D^n &:= \mathcal{S}^n_+ \cap N^n \\ &= \{X \in \mathcal{S}^n \mid X \text{ is positive semidefinite and} \\ &\qquad \text{elementwise nonnegative}\}. \end{aligned}$$

If we use the partial ordering induced by this convex cone in the constraint (7.2), then the optimization problem (7.3) can be written as

$$\begin{gathered} \min f(x) \\ \text{subject to the constraints} \\ G(x) \preceq 0_{\mathcal{S}^n} \\ G_{ij}(x) \leq 0 \;\text{ for all }\; i, j \in \{1, \ldots, n\} \\ x \in \mathbb{R}^m. \end{gathered}$$

So, we have a semidefinite optimization problem with additional finitely many nonlinear constraints. Obviously, for every convex cone K we have $D^n \subset C^n_K$ and $(C^n_K)^* \subset (D^n)^*$.

Before discussing some examples we need an important lemma on the *Schur complement*.

Lemma 7.2 (Schur complement).

Let $X = \begin{pmatrix} A & B^T \\ B & C \end{pmatrix} \in \mathcal{S}^{k+l}$ with $A \in \mathcal{S}^k$, $C \in \mathcal{S}^l$ and $B \in \mathbb{R}^{(l,k)}$ be given, and assume that A is positive definite. Then we have for the Löwner partial ordering $\preceq$

$$-X \preceq 0_{\mathcal{S}^{k+l}} \iff -(C - BA^{-1}B^T) \preceq 0_{\mathcal{S}^l}$$

(the matrix $C - BA^{-1}B^T$ is called the Schur complement of A in X).

Proof We have

$$
\begin{aligned}
-X \preceq 0_{\mathcal{S}^{k+l}} \iff 0 &\le (x^T, y^T) \begin{pmatrix} A & B^T \\ B & C \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= x^T A x + 2x^T B^T y + y^T C y \quad \text{for all } x \in \mathbb{R}^k \\
&\qquad \text{and all } y \in \mathbb{R}^l \\
\iff 0 &\le \min_{x \in \mathbb{R}^k} x^T A x + 2x^T B^T y + y^T C y \quad \text{for all } y \in \mathbb{R}^l.
\end{aligned}
$$

Since A is positive definite, for an arbitrarily chosen $y \in \mathbb{R}^l$ this optimization problem has the minimal solution $-A^{-1}B^T y$ with the minimal value

$$-y^T B A^{-1} B^T y + y^T C y = y^T (C - B A^{-1} B^T) y.$$

Consequently we get

$$
\begin{aligned}
-X \preceq 0_{\mathcal{S}^{k+l}} &\iff y^T (C - B A^{-1} B^T) y \ge 0 \quad \text{for all } y \in \mathbb{R}^l \\
&\iff -(C - B A^{-1} B^T) \preceq 0_{\mathcal{S}^l}.
\end{aligned}
$$

□

The following example illustrates the significance of semidefinite optimization.

Example 7.3 (semidefinite optimization).

(a) The problem of determining the smallest among the largest eigenvalues of a matrix-valued function $A : \mathbb{R}^m \to \mathcal{S}^n$ leads to the semidefinite optimization problem

$$
\begin{gathered}
\min \ \lambda \\
\text{subject to the constraints} \\
A(x) - \lambda I \preceq 0_{\mathcal{S}^n} \\
x \in \mathbb{R}^m
\end{gathered}
$$

(with the identity matrix $I \in \mathcal{S}^n$ and the Löwner partial ordering $\preceq$). Indeed, $A(x) - \lambda I$ is negative semidefinite if and only if for all eigenvalues $\lambda_1, \ldots, \lambda_n$ of $A(x)$ the inequality $\lambda_i \le \lambda$ is satisfied. Hence, with the minimization of λ we determine the smallest among the largest eigenvalues of $A(x)$.

(b) We consider a nonlinear optimization problem with a quadratic constraint in a finite dimensional setting, i.e. we have

$$\begin{array}{c} \min f(x) \\ \text{subject to the constraints} \\ (Ax+b)^T(Ax+b) - c^T x - \alpha \leq 0 \\ x \in \mathbb{R}^m \end{array} \tag{7.5}$$

with an objective function $f : \mathbb{R}^m \to \mathbb{R}$, a given matrix $A \in \mathbb{R}^{(k,m)}$, given vectors $b \in \mathbb{R}^k$ and $c \in \mathbb{R}^m$ and a real number α. If $\preceq$ denotes again the Löwner partial ordering, we consider the inequality

$$-\begin{pmatrix} I & Ax+b \\ (Ax+b)^T & c^T x + \alpha \end{pmatrix} \preceq 0_{\mathcal{S}^{k+1}} \tag{7.6}$$

($I \in \mathcal{S}^k$ denotes the identity matrix). By Lemma 7.2 this inequality is equivalent to the quadratic constraint

$$(Ax+b)^T(Ax+b) - c^T x - \alpha \leq 0.$$

If the i-th column of the matrix A (with $i \in \{1, \ldots, k\}$) is denoted by $a^{(i)} \in \mathbb{R}^m$, then we set

$$A^{(0)} := \begin{pmatrix} I & b \\ b^T & \alpha \end{pmatrix}$$

and

$$A^{(i)} := \begin{pmatrix} 0_{\mathcal{S}^k} & a^{(i)} \\ a^{(i)T} & c_i \end{pmatrix} \text{ for all } i \in \{1, \ldots, k\},$$

and the inequality (7.6) is equivalent to

$$-A^{(0)} - A^{(1)}x_1 - \cdots - A^{(k)}x_k \preceq 0_{\mathcal{S}^{k+1}}.$$

Hence, the original problem (7.5) with a quadratic constraint can be written as a semidefinite optimization problem with a linear constraint

$$\begin{array}{c} \min f(x) \\ \text{subject to the constraints} \\ -A^{(0)} - A^{(1)}x_1 - \cdots - A^{(k)}x_k \preceq 0_{\mathcal{S}^{k+1}} \\ x \in \mathbb{R}^m. \end{array}$$

Although the partial ordering used in the constraint becomes more complicated by this transformation, the type of the constraint which is now linear and not quadratic, is much simpler to handle. A similar transformation can be carried out in the case that, in addition, the objective function f is also quadratic. Then we minimize an additional variable and use this variable as an upper bound of the objective function.

(c) We consider a system of autonomous linear differential equations

$$\dot{x}(t) = Ax(t) + Bu(t) \text{ almost everywhere on } [0, \infty) \tag{7.7}$$

with given matrices $A \in \mathbb{R}^{(k,k)}$ and $B \in \mathbb{R}^{(k,l)}$. Using a feedback control

$$u(t) = Fx(t) \text{ almost everywhere on } [0, \infty)$$

with an unknown matrix $F \in \mathbb{R}^{(l,k)}$ we try to make the system (7.7) asymptotically stable , i.e. we require for every solution x of (7.7) that

$$\lim_{t\to\infty} \|x(t)\| = 0$$

for the Euclidean norm $\|\cdot\|$ in $\mathbb{R}^k$. In control theory the autonomous linear system (7.7) is called *stabilizable*, if there exists a matrix $F \in \mathbb{R}^{(l,k)}$ so that the system (7.7) is asymptotically stable.

For the determination of an appropriate matrix F we investigate the so-called Lyapunov function $v : \mathbb{R}^k \to \mathbb{R}$ with

$$v(\tilde{x}) = \tilde{x}^T P\tilde{x} \text{ for all } \tilde{x} \in \mathbb{R}^k$$

($P \in \mathcal{S}^k$ is arbitrarily chosen and should be positive definite). Since P is positive definite we have

$$v(\tilde{x}) > 0 \text{ for all } \tilde{x} \in \mathbb{R}^k \backslash \{0_{\mathbb{R}^k}\}. \tag{7.8}$$

For a solution x of the system (7.7) we obtain

$$\begin{aligned}
&\dot{v}(x(t)) \\
&= \frac{d}{dt} x(t)^T Px(t) \\
&= \dot{x}(t)^T Px(t) + x(t)^T P\dot{x}(t) \\
&= \big(Ax(t) + BFx(t)\big)^T Px(t) + x(t)^T P\big(Ax(t) + BFx(t)\big) \\
&= x(t)^T \big((A + BF)^T P + P(A + BF)\big)x(t).
\end{aligned}$$

If the matrices P and F are chosen in such a way that $(A + BF)^T P + P(A + BF)$ is negative definite, then there is a positive number α with

$$\dot{v}(x(t)) \leq -\alpha \|x(t)\|^2 \text{ for all } t \in [0, \infty). \tag{7.9}$$

The inequalities (7.8) and (7.9) imply

$$\lim_{t \to \infty} v(x(t)) = 0. \tag{7.10}$$

Since P is assumed to be positive definite, there is a positive number $\beta > 0$ with

$$v(\tilde{x}) \geq \beta \|\tilde{x}\|^2 \text{ for all } \tilde{x} \in \mathbb{R}^k.$$

Then we conclude with (7.10)

$$\lim_{t \to \infty} \|x(t)\| = 0,$$

i.e. the autonomous linear system (7.7) is stabilizable. Hence, we obtain the stabilization of (7.7) by a feedback control, if we choose a positive definite matrix $P \in \mathcal{S}^k$ and a matrix $F \in \mathbb{R}^{(l,k)}$ so that $(A + BF)^T P + P(A + BF)$ is negative definite.
In order to fulfill this requirement we consider the semidefinite optimization problem

$$\begin{gathered} \min \lambda \\ \text{subject to the constraints} \\ -\lambda I + (A + BF)^T P + P(A + BF) \preceq 0_{\mathcal{S}^k} \\ -\lambda I - P \preceq 0_{\mathcal{S}^k} \\ \lambda \in \mathbb{R}, \quad P \in \mathcal{S}^k, \quad F \in \mathbb{R}^{(l,k)} \end{gathered} \tag{7.11}$$

($I \in \mathcal{S}^k$ denotes the identity matrix and recall that $\preceq$ denotes the Löwner partial ordering). By a suitable transformation this problem formally fits into the class (7.3) of semidefinite problems. Here G has to be defined in an appropriate way. It is important to note that it is not necessary to solve the problem (7.11). Only a feasible solution with $\lambda < 0$ is requested. Then the matrices P and F fulfill the requirements for the stabilization of the autonomous linear system (7.7).

(d) Finally we discuss an applied problem from structural optimization and consider a structure of k elastic bars connecting a set of p nodes (see Fig. 7.1). The design variables x_i $(i = 1, \ldots, k)$ are the cross-sectional areas of the bars. We assume that nodal load forces $f_1, \ldots, f_p$ are given.

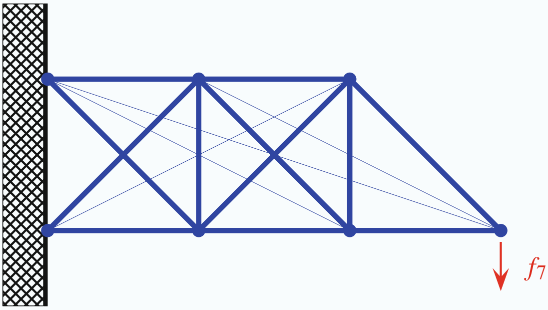

Fig. 7.1 Cantilever with seven nodes and the load force f_7

$l_1, \ldots, l_k$ denote the length of the bars, v is the maximal volume, and $\underline{x}_i > 0$ and $\bar{x}_i$ are the lower and upper bounds of the cross-sectional areas. The so-called stiffness matrix $A(x) \in \mathcal{S}^p$ is positive definite for all $x_1, \ldots, x_k > 0$. We want to find a feasible structure with minimal elastic stored energy. Then we obtain the optimization problem

$$\begin{gathered} \min f^T A(x)^{-1} f \\ \text{subject to the constraints} \\ \sum_{i=1}^{k} l_i x_i \leq v \\ \underline{x}_i \leq x_i \leq \bar{x}_i \ \text{ for all } i \in \{1, \ldots, k\} \end{gathered}$$

or

$$\begin{gathered} \min \lambda \\ \text{subject to the constraints} \\ f^T A(x)^{-1} f - \lambda \leq 0 \\ \sum_{i=1}^{k} l_i x_i \leq v \\ \underline{x}_i \leq x_i \leq \bar{x}_i \ \text{ for all } i \in \{1, \ldots, k\}. \end{gathered}$$

By Lemma 7.2 the inequality constraint

$$f^T A(x)^{-1} f - \lambda \leq 0$$

is equivalent to

$$-\begin{pmatrix} A(x) & f \\ f^T & \lambda \end{pmatrix} \preceq 0_{\mathcal{S}^{p+1}}$$

(recall that $\preceq$ denotes the Löwner partial ordering). Hence, we get a standard semidefinite optimization problem with an additional linear inequality constraint and upper and lower bounds:

$$\begin{gathered} \min \lambda \\ \text{subject to the constraints} \\ -\begin{pmatrix} A(x) & f \\ f^T & \lambda \end{pmatrix} \preceq 0_{\mathcal{S}^{p+1}} \\ \sum_{i=1}^{k} l_i x_i \leq v \\ \underline{x}_i \leq x_i \leq \bar{x}_i \;\; \text{for all } i \in \{1, \ldots, k\}. \end{gathered}$$

Although the Löwner partial ordering is mostly used for describing the inequality constraint (7.2), we mainly investigate the more general conic optimization problem (7.3) covering the standard semidefinite problem. For the application of the general theory of this book we now investigate properties of the presented ordering cones in more detail.

Lemma 7.4 (properties of the Löwner ordering cone).

For the Löwner ordering cone $\mathcal{S}^n_+$ we have:

(a) $int(\mathcal{S}^n_+) = \{X \in \mathcal{S}^n \mid X$ is positive definite$\}$
(b) $(\mathcal{S}^n_+)^* = \mathcal{S}^n_+$, i.e. $\mathcal{S}^n_+$ is self-dual.

Proof

(a) First, we show the inclusion $\text{int}(\mathcal{S}^n_+) \subset \{X \in \mathcal{S}^n \mid X \text{ is positive definite}\}$. Let $X \in \text{int}(\mathcal{S}^n_+)$ be arbitrarily chosen. Then we get for a sufficiently small $\lambda > 0$ $\;X - \lambda I \in \mathcal{S}^n_+$ ($I \in \mathcal{S}^n$ denotes the identity matrix), i.e.

$$0 \leq x^T (X - \lambda I)x = x^T X x - \lambda x^T x \;\; \text{for all } x \in \mathbb{R}^m$$

implying

$$x^T X x \geq \lambda x^T x > 0 \;\; \text{for all } x \in \mathbb{R}^m \backslash \{0_{\mathbb{R}^m}\}.$$

Consequently, the matrix X is positive definite.

Next we prove the converse inclusion. Let a positive definite matrix $X \in \mathcal{S}^n$ be arbitrarily given. Then all eigenvalues of X are positive. Since the minimal eigenvalue continuously depends on the elements of the matrix, it follows immediately that X belongs to the interior of $\mathcal{S}^n_+$.

(b) First, we show the inclusion $(\mathcal{S}^n_+)^* \subset \mathcal{S}^n_+$. Let an arbitrary matrix $X \in (\mathcal{S}^n_+)^*$ be chosen and assume that $X \notin \mathcal{S}^n_+$. Then there exists some $y \in \mathbb{R}^m$ so that $y^T X y < 0$. If we set $Y := yy^T$, we have $Y \in \mathcal{S}^n_+$ and we obtain

$$\langle X, Y \rangle = \text{trace}(Xyy^T) = y^T X y < 0,$$

a contradiction to $X \in (\mathcal{S}^n_+)^*$.

Now, we prove the converse inclusion. Let $X \in \mathcal{S}^n_+$ be arbitrarily given. Choose any $Y \in \mathcal{S}^n_+$. Since X and Y are symmetric and positive semidefinite it is known that there are matrices $\sqrt{X}, \sqrt{Y} \in \mathcal{S}^n_+$ with $(\sqrt{X})^2 = X$ and $(\sqrt{Y})^2 = Y$ and we obtain

$$\begin{aligned}\langle X, Y \rangle &= \text{trace}(\sqrt{X}\sqrt{X}\sqrt{Y}\sqrt{Y}) \\ &= \text{trace}(\sqrt{X}\sqrt{Y}\sqrt{Y}\sqrt{X}) \\ &= \langle \sqrt{X}\sqrt{Y}, \sqrt{X}\sqrt{Y} \rangle \\ &\geq 0.\end{aligned}$$

Hence, we conclude $X \in (\mathcal{S}^n_+)^*$. □

The result of Lemma 7.4,(b) is also called *Féjèr theorem* in the special literature. For the K-copositive ordering cone we obtain similar results.

Lemma 7.5 (properties of the K-copositive ordering cone).

Let $K \subset \mathbb{R}^n$ be a convex cone. For the K-copositive ordering cone C^n_K we have:

(a) $\{X \in \mathcal{S}^n \mid X$ is positive definite$\} \subset int(C^n_K)$.
(b) In addition, if K is closed, then for $H_K :=$ convex hull $\{xx^T \mid x \in K\}$
   (i) H_K is closed
   (ii) $(C^n_K)^* = H_K$.

Proof

(a) By definition we have $S^n_+ \subset C^n_K$. Consequently, the assertion follows from Lemma 7.4,(a).

(b) (i) Let an arbitrary sequence $X_k \in H_K$ be chosen with the limit $X \in \mathcal{S}^n$ (with respect to the spectral norm). Since K is a cone, for every $k \in \mathbb{N}$ there are vectors $x^{(1_k)}, \ldots, x^{(p_k)} \in K$ with the property

$$X_k = \sum_{i=1}^{p} x^{(i_k)} x^{(i_k)^T}$$

(notice that by the Carathéodory theorem the number p of vectors is bounded by $n+1$). Every $x^{(i_k)} \in K$ ($i \in \{1, \ldots, p\}$, $k \in \mathbb{N}$) can be written as

$$x^{(i_k)} = \mu_{i_k} s^{(i_k)}$$

with $\mu_{i_k} \geq 0$ and

$$s^{(i_k)} \in K \cap \{x \in \mathbb{R}^n \mid \|x\| = 1\}$$

($\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$). This set is compact because K is assumed to be closed. Consequently, we obtain for every $k \in \mathbb{N}$

$$X_k = \sum_{i=1}^{p} \mu_{i_k}^2 s^{(i_k)} s^{(i_k)^T}.$$

Since $s^{(1_k)}, \ldots, s^{(p_k)}$ belong to a compact set and $(X_k)_{k \in \mathbb{N}}$ converges to X, the numbers $\mu_{1_k}, \ldots, \mu_{p_k}$ are bounded and there are subsequences $(s^{(i_{k_l})})_{l \in \mathbb{N}}$ and $(\mu_{i_{k_l}})_{l \in \mathbb{N}}$ (with $i \in \{1, \ldots p\}$) converging to $s^{(i)} \in K$ and $\mu_i \in \mathbb{R}$, respectively, with the property

$$X = \sum_{i=1}^{p} \mu_i^2 s^{(i)} s^{(i)^T}.$$

This implies $X \in H_K$. Hence, H_K is a closed set.

(ii) First we show the inclusion $H_K \subset (C_K^n)^*$. For an arbitrary $X \in H_K$ we have the representation

$$X = \sum_{i=1}^{p} x^{(i)} x^{(i)^T} \text{ for some } x^{(1)}, \ldots, x^{(p)} \in K$$

(notice here that K is a cone). Then we obtain for every $Y \in C_K^n$

$$\begin{aligned}\langle Y, X \rangle &= \text{trace}(Y \cdot X) \\ &= \text{trace}\left(Y \sum_{i=1}^{p} x^{(i)} x^{(i)^T}\right)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{p} \mathrm{trace}(Y x^{(i)} {x^{(i)}}^T) \\
&= \sum_{i=1}^{p} {x^{(i)}}^T Y x^{(i)} \\
&\geq 0,
\end{aligned}
$$

i.e. $X \in (C_K^n)^*$.
For the proof of the converse inclusion we first show $H_K^* \subset C_K^n$. Let an arbitrary $X \notin C_K^n$ be given. Then there is some $y \in K$ with $y^T X y < 0$. If we set $Y := yy^T$, then we have $Y \in H_K$ and

$$\langle Y, X \rangle = \mathrm{trace}(Y \cdot X) = \mathrm{trace}(X y y^T) = y^T X y < 0,$$

i.e. $X \notin H_K^*$. Consequently $H_K^* \subset C_K^n$ and for the dual cones we get

$$(C_K^n)^* \subset (H_K^*)^*. \tag{7.12}$$

Next, we show that $(H_K^*)^* \subset H_K$. For this proof let $Z \in (H_K^*)^*$ be arbitrarily given and assume that $Z \notin H_K$. Since H_K is closed by part (i) and convex, by Theorem C.3 there exists some $V \in \mathcal{S}^n \backslash \{0_{\mathcal{S}^n}\}$ with

$$\langle V, Z \rangle < \inf_{U \in H_K} \langle V, U \rangle. \tag{7.13}$$

This inequality implies

$$\langle V, Z \rangle < 0, \tag{7.14}$$

if we set $U = 0_{\mathcal{S}^n}$. Now assume that $V \notin H_K^*$. Then there is some $\tilde{U} \in H_K$ with $\langle V, \tilde{U} \rangle < 0$. Since H_K is a cone, we have $\lambda \tilde{U} \in H_K$ for all $\lambda > 0$ and

$$0 > \lambda \langle V, \tilde{U} \rangle = \langle V, \lambda \tilde{U} \rangle \ \text{ for all } \lambda > 0.$$

Consequently, $\langle V, \lambda \tilde{U} \rangle$ can be made arbitrarily small contradicting to the inequality (7.13). So $V \in H_K^*$ and because of $Z \in (H_K^*)^*$ we obtain $\langle V, Z \rangle \geq 0$ contradicting (7.14). Hence we get $Z \in H_K$. With the inclusions $(H_K^*)^* \subset H_K$ and (7.12) we then conclude $(C_K^n)^* \subset H_K$ which has to be shown. □

In the special literature elements in the dual cone $(C_{\mathbb{R}_+^n}^n)^* = H_{\mathbb{R}_+^n}$ (i.e. we set $K = \mathbb{R}_+^n$) are called *completely positive matrices*.

Finally we present similar results for the nonnegative ordering cone and the doubly nonnegative ordering cone.

Lemma 7.6 (properties of the nonnegative and doubly nonnegative ordering cone).

For the nonnegative ordering cone N^n and the doubly nonnegative ordering cone D^n we have:

(a) $\mathrm{int}(N^n) = \{X \in \mathcal{S}^n \mid X_{ij} > 0 \text{ for all } i, j \in \{1, \ldots, n\}\}$
(b) $(N^n)^* = N^n$, *i.e.* N^n *is self-dual*
(c) $\mathrm{int}(D^n) = \{X \in \mathcal{S}^n \mid X \text{ is positive definite and elementwise positive}\}$
(d) $(D^n)^* = D^n$, *i.e.* D^n *is self-dual.*

Proof
(a) This part is obvious.
(b) (i) Let $X \in N^n$ be arbitrarily chosen. Then we get for all $M \in N^n$

$$\langle X, M\rangle = \mathrm{trace}(X \cdot M) = \sum_{i=1}^{n}\sum_{j=1}^{n} \underbrace{X_{ij}}_{\geq 0} \cdot \underbrace{M_{ji}}_{\geq 0} \geq 0.$$

Consequently, we have $X \in (N^n)^*$.
(ii) Now let $X \in (N^n)^*$ be arbitrarily chosen. If we consider $M \in N^n$ with

$$M_{ij} = \begin{cases} 1 \text{ for } i = k \text{ and } j = l \\ 0 \text{ otherwise} \end{cases}$$

for arbitrary $k, l \in \{1, \ldots, n\}$, then we conclude

$$0 \leq \langle X, M\rangle = X_{kl}.$$

So, we obtain $X \in N^n$.
(c) With Lemma 7.4,(a) and part (a) of this lemma we get

$$\begin{aligned} \mathrm{int}(D^n) &= \mathrm{int}(\mathcal{S}^n_+ \cap N^n) \\ &= \mathrm{int}(\mathcal{S}^n_+) \cap \mathrm{int}(N^n) \\ &= \{X \in \mathcal{S}^n_+ \mid X \text{ positive definite and elementwise positive}\}. \end{aligned}$$

(d) With Lemma 7.4,(b) and part (b) of this lemma we obtain

$$(D^n)^* = (\mathcal{S}^n_+)^* \cap (N^n)^* = \mathcal{S}^n_+ \cap N^n = D^n. \qquad \square$$

7.2 Optimality Conditions

The necessary optimality conditions presented in Sect. 5.2 are now applied to the conic optimization problem (7.3) with the partial ordering $\preccurlyeq$ inducing the ordering cone C. To be more specific, let $f : \mathbb{R}^m \to \mathbb{R}$ and $G : \mathbb{R}^m \to \mathcal{S}^n$ be given functions and consider the conic optimization problem

$$\begin{gathered} \min f(x) \\ \text{subject to the constraints} \\ G(x) \preccurlyeq 0_{\mathcal{S}^n} \\ x \in \mathbb{R}^m. \end{gathered}$$

First, we answer the question under which assumptions the matrix function G is Fréchet differentiable.

Lemma 7.7 (Fréchet derivative of G).

Let the matrix function $G : \mathbb{R}^m \to \mathcal{S}^n$ be elementwise differentiable at some $\bar{x} \in \mathbb{R}^m$. Then the Fréchet derivative of G at $\bar{x}$ is given by

$$G'(\bar{x})(h) = \sum_{i=1}^{m} G_{x_i}(\bar{x})\, h_i \ \text{ for all } h \in \mathbb{R}^m$$

with

$$G_{x_i} := \begin{pmatrix} \frac{\partial}{\partial x_i} G_{11} & \cdots & \frac{\partial}{\partial x_i} G_{1n} \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_i} G_{n1} & \cdots & \frac{\partial}{\partial x_i} G_{nn} \end{pmatrix} \ \text{ for all } \ i \in \{1, \ldots, m\}.$$

Proof Let $h \in \mathbb{R}^m$ be arbitrarily chosen. Since G is elementwise differentiable at $\bar{x} \in \mathbb{R}^m$, we obtain for the Fréchet derivative of G

$$G'(\bar{x})(h) = \begin{pmatrix} \nabla G_{11}(\bar{x})^T h & \cdots & \nabla G_{1n}(\bar{x})^T h \\ \vdots & & \vdots \\ \nabla G_{n1}(\bar{x})^T h & \cdots & \nabla G_{nn}(\bar{x})^T h \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{m} G_{11x_i}(\bar{x})\, h_i & \cdots & \sum_{i=1}^{m} G_{1nx_i}(\bar{x})\, h_i \\ \vdots & & \vdots \\ \sum_{i=1}^{m} G_{n1x_i}(\bar{x})\, h_i & \cdots & \sum_{i=1}^{m} G_{nnx_i}(\bar{x})\, h_i \end{pmatrix}$$

$$= \sum_{i=1}^{m} G_{x_i}(\bar{x})\, h_i\,. \qquad \square$$

Now we present the Lagrange multiplier rule for the conic optimization problem (7.3).

Theorem 7.8 (Lagrange multiplier rule).

Let $f : \mathbb{R}^m \to \mathbb{R}$ and $G : \mathbb{R}^m \to \mathcal{S}^n$ be given functions, and let $\bar{x} \in \mathbb{R}^m$ be a minimal solution of the conic optimization problem (7.3). Let f be differentiable at $\bar{x}$ and let G be elementwise differentiable at $\bar{x}$. Then there are a real number $\mu \geq 0$ and a matrix $L \in C^*$ with $(\mu, L) \neq (0, 0_{\mathcal{S}^n})$,

$$\mu \nabla f(\bar{x}) + \begin{pmatrix} \langle L, G_{x_1}(\bar{x})\rangle \\ \vdots \\ \langle L, G_{x_m}(\bar{x})\rangle \end{pmatrix} = 0_{\mathbb{R}^m} \tag{7.15}$$

and

$$\langle L, G(\bar{x})\rangle = 0. \tag{7.16}$$

If, in addition to the above assumptions the equality

$$G'(\bar{x})(\mathbb{R}^m) + cone\,(C + \{G(\bar{x})\}) = \mathcal{S}^n \tag{7.17}$$

is satisfied, then it follows $\mu > 0$.

Proof Because of the differentiability assumptions we have that f and G are Fréchet differentiable at $\bar{x}$. Then we apply Corollary 5.4 and obtain the existence of a real number $\mu \geq 0$ and a matrix $L \in C^*$ with $(\mu, L) \neq (0, 0_{\mathcal{S}^n})$,

$$\mu \nabla f(\bar{x}) + L \circ G'(\bar{x}) = 0_{\mathbb{R}^m} \tag{7.18}$$

and

$$\langle L, G(\bar{x})\rangle = 0.$$

For every $h \in \mathbb{R}^m$ we obtain with Lemma 7.7

$$\begin{aligned}(L \circ G'(\bar{x}))(h) &= \langle L, G'(\bar{x})(h)\rangle \\ &= \langle L, \sum_{i=1}^{m} G_{x_i}(\bar{x})h_i\rangle \\ &= \sum_{i=1}^{m}\langle L, G_{x_i}(\bar{x})\rangle h_i \\ &= \begin{pmatrix} \langle L, G_{x_1}(\bar{x})\rangle \\ \vdots \\ \langle L, G_{x_m}(\bar{x})\rangle \end{pmatrix}^T h.\end{aligned}$$

Then the equality (7.18) implies

$$\mu \nabla f(\bar{x}) + \begin{pmatrix} \langle L, G_{x_1}(\bar{x})\rangle \\ \vdots \\ \langle L, G_{x_m}(\bar{x})\rangle \end{pmatrix} = 0_{\mathbb{R}^m}.$$

Hence, one part of the assertion is shown. If we consider the Kurcyusz-Robinson-Zowe regularity assumption (5.9) for the special problem (7.3), we have $\hat{S} = \mathbb{R}^m$ and $\text{cone}(\hat{S} - \{\bar{x}\}) = \mathbb{R}^m$. So, the equality (7.17) is equivalent to the regularity assumption (5.9). This completes the proof. □

In the case of $\mu > 0$ we can set $U := \frac{1}{\mu}L \in C^*$ and the equalities (7.15) and (7.16) can be written as

$$f_{x_i}(\bar{x}) + \langle U, G_{x_i}(\bar{x})\rangle = 0 \;\text{ for all } i \in \{1, \ldots, m\}$$

and

$$\langle U, G(\bar{x})\rangle = 0.$$

This gives the extended Karush-Kuhn-Tucker conditions to matrix space problems.

In Theorem 7.8 the Lagrange multiplier L is a matrix in the dual cone C^*. According to the specific choice of the ordering cone C discussed in Lemmas 7.4, 7.5 and 7.6 we take the dual cones given in Lemmas 7.4,(b), 7.5,(b),(ii) and 7.6,(b),(d). For instance, if C denotes the Löwner ordering cone, then the multiplier L is positive semidefinite.

Instead of the regularity assumption (7.17) used in Theorem 7.8 we can also consider a simpler condition.

Lemma 7.9 (regularity condition).

Let the assumption of Theorem 7.8 be satisfied and let C denote the K-copositive ordering cone C_K^n for an arbitrary convex cone K. If there exists a vector $\hat{x} \in \mathbb{R}^m$ so that $G(\bar{x}) + \sum_{i=1}^{m} G_{x_i}(\bar{x})(\hat{x}_i - \bar{x}_i)$ is negative definite, then the regularity assumption in Theorem 7.8 is fulfilled.

Proof By Lemma 7.5,(a) we have

$$G(\bar{x}) + G'(\bar{x})(\hat{x} - \bar{x}) = G(\bar{x}) + \sum_{i=1}^{m} G_{x_i}(\bar{x})(\hat{x}_i - \bar{x}_i) \in -\mathrm{int}(C_K^n)$$

and with Theorem 5.6 the Kurcyusz-Robinson-Zowe regularity assumption is satisfied, i.e. the equality (7.17) is fulfilled. □

It is obvious that in the case of the Löwner partial ordering $S_+^n = C_{\mathbb{R}^n}^n$ Lemma 7.9 is also applicable. Notice that a similar result can be shown for the ordering cones discussed in Lemma 7.6. For the interior of these cones we can then use the results in Lemma 7.6,(a) and (c).

Next, we answer the question under which assumptions the Lagrange multiplier rule given in Theorem 7.8 as a necessary optimality condition is a sufficient optimality condition for the conic optimization problem (7.3).

Theorem 7.10 (sufficient optimality condition).

Let $f : \mathbb{R}^m \to \mathbb{R}$ and $G : \mathbb{R}^m \to \mathcal{S}^n$ be given functions. Let for some $\bar{x} \in \mathbb{R}^m$ f be differentiable and pseudoconvex at $\bar{x}$ and let G be elementwise differentiable and $(-C + cone(\{G(\bar{x})\}) - cone(\{G(\bar{x})\}))$-quasiconvex at $\bar{x}$. If there is a matrix $L \in C^*$ with

$$\nabla f(\bar{x}) + \begin{pmatrix} \langle L, G_{x_1}(\bar{x}) \rangle \\ \vdots \\ \langle L, G_{x_m}(\bar{x}) \rangle \end{pmatrix} = 0_{\mathbb{R}^m} \tag{7.19}$$

and

$$\langle L, G(\bar{x}) \rangle = 0,$$

then $\bar{x}$ is a minimal solution of the conic optimization problem (7.3).

Proof With Lemma 7.7 the equality (7.19) implies

$$\nabla f(\bar{x}) + L \circ G'(\bar{x}) = 0_{\mathbb{R}^m}.$$

By Lemma 5.16 and Corollary 5.15 the assertion follows immediately. □

The quasiconvexity assumption in Theorem 7.10 (compare Definition 5.12) means that for all feasible $x \in \mathbb{R}^m$

$$\begin{aligned} & G(x) - G(\bar{x}) \in -C + \text{cone}(\{G(\bar{x})\}) - \text{cone}(\{G(\bar{x})\}) \\ \Longrightarrow \quad & \sum_{i=1}^{m} G_{x_i}(\bar{x})(x_i - \bar{x}_i) \in -C + \text{cone}(\{G(\bar{x})\}) - \text{cone}(\{G(\bar{x})\}). \end{aligned}$$

For all feasible $x \in \mathbb{R}^m$ this implication can be rewritten as

$$\begin{aligned} & G(x) + (\alpha - 1 - \beta)G(\bar{x}) \in -C \ \text{ for some } \alpha, \beta \geq 0 \\ \Longrightarrow \quad & \sum_{i=1}^{m} G_{x_i}(\bar{x})(x_i - \bar{x}_i) + (\gamma - \delta)G(\bar{x}) \in -C \ \text{ for some } \gamma, \delta \geq 0 \end{aligned}$$

or

$$\begin{aligned} & G(x) + \bar{\alpha} G(\bar{x}) \in -C \ \text{ for some } \bar{\alpha} \in \mathbb{R} \\ \Longrightarrow \quad & \sum_{i=1}^{m} G_{x_i}(\bar{x})(x_i - \bar{x}_i) + \bar{\gamma} G(\bar{x}) \in -C \ \text{ for some } \bar{\gamma} \in \mathbb{R}. \end{aligned}$$

7.3 Duality

The duality theory developed in Chap. 6 is now applied to the conic optimization problem (7.3) with given functions $f : \mathbb{R}^m \to \mathbb{R}$ and $G : \mathbb{R}^m \to \mathcal{S}^n$ and the partial ordering $\preccurlyeq$ inducing the ordering cone C.

For convenience we recall the primal optimization problem

$$\begin{gathered} \min f(x) \\ \text{subject to the constraints} \\ G(x) \preccurlyeq 0_{\mathcal{S}^n} \\ x \in \mathbb{R}^m. \end{gathered}$$

According to Sect. 6.1 the dual problem can be written as

$$\max_{U \in C^*} \inf_{x \in \mathbb{R}^m} f(x) + \langle U, G(x) \rangle \tag{7.20}$$

or equivalently

$$\begin{gathered}\max \lambda \\ \text{subject to the constraints} \\ f(x) + \langle U, G(x) \rangle \geq \lambda \;\text{ for all } x \in \mathbb{R}^m \\ \lambda \in \mathbb{R},\;\; U \in C^*.\end{gathered}$$

We are now able to formulate a *weak duality theorem* for the conic optimization problem (7.3).

Theorem 7.11 (weak duality theorem).

For every feasible $\hat{x}$ of the primal problem (7.3) and for every feasible $\hat{U}$ of the dual problem (7.20) the following inequality is satisfied

$$\inf_{x \in \mathbb{R}^m} f(x) + \langle \hat{U}, G(x) \rangle \leq f(\hat{x}).$$

Proof This result follows immediately from Theorem 6.7. □

The following *strong duality theorem* is a direct consequence of Theorem 6.8.

Theorem 7.12 (strong duality theorem).

Let the composite mapping $(f, G) : \mathbb{R}^m \to \mathbb{R} \times \mathcal{S}^n$ be convex-like and let the ordering cone have a nonempty interior $int(C)$. If the primal problem (7.3) is solvable and the generalized Slater condition is satisfied, i.e., there is a vector $\hat{x} \in \mathbb{R}^m$ with $G(\hat{x}) \in -int(C)$, then the dual problem (7.20) is also solvable and the extremal values of the two problems are equal.

For instance, if the ordering cone C is the K-copositive ordering cone C_K^n for some convex cone $K \subset \mathbb{R}^n$, then by Lemma 7.5,(a) the generalized Slater condition in Theorem 7.12 is satisfied whenever $G(\hat{x})$ is negative definite for some $\hat{x} \in \mathbb{R}^m$. In this case a duality gap cannot appear.

With the investigations in Sect. 6.4 it is simple to state the dual problem of a linear semidefinite optimization problem. If we specialize the problem (7.3) to the linear problem

$$\begin{gathered}\min \; c^T x \\ \text{subject to the constraints} \\ B \preccurlyeq A(x) \\ x_1, \ldots, x_m \geq 0\end{gathered} \tag{7.21}$$

with $c \in \mathbb{R}^m$, a linear mapping $A : \mathbb{R}^m \to \mathcal{S}^n$ and a matrix $B \in \mathcal{S}^n$. Since A is linear, there are matrices $A^{(1)}, \ldots, A^{(m)} \in \mathcal{S}^n$ so that

$$A(x) = A^{(1)}x_1 + \ldots + A^{(m)}x_m \text{ for all } x \in \mathbb{R}^m.$$

Then the primal linear problem (7.21) can also be written as

$$\begin{gathered} \min \ c^T x \\ \text{subject to the constraints} \\ B \preccurlyeq A^{(1)}x_1 + \ldots + A^{(m)}x_m \\ x_1, \ldots, x_m \geq 0. \end{gathered} \tag{7.22}$$

For the formulation of the dual problem of (7.22) we need the adjoint mapping $A^* : \mathcal{S}^n \to \mathbb{R}^m$ defined by

$$\begin{aligned} A^*(U)(x) &= \langle U, A(x)\rangle \\ &= \langle U, A^{(1)}x_1 + \ldots + A^{(m)}x_m\rangle \\ &= \langle U, A^{(1)}\rangle x_1 + \ldots + \langle U, A^{(m)}\rangle x_m \\ &= \Big(\langle U, A^{(1)}\rangle, \ldots, \langle U, A^{(m)}\rangle\Big) \cdot x \\ &\quad \text{for all } x \in \mathbb{R}^m \text{ and all } U \in \mathcal{S}^n. \end{aligned}$$

Using the general formulation (6.19) we then obtain the dual problem

$$\begin{gathered} \max \ \langle B, U\rangle \\ \text{subject to the constraints} \\ \langle A^{(1)}, U\rangle \leq c_1 \\ \vdots \\ \langle A^{(m)}, U\rangle \leq c_m \\ U \in C^*. \end{gathered} \tag{7.23}$$

In the special literature on semidefinite optimization the dual problem (7.23) is very often the primal problem with $C^* = \mathcal{S}^n_+$. In this case our primal problem is then the dual problem in the literature.

Exercises

(7.1) Show that the Löwner ordering cone $\mathcal{S}^n_+$ is closed and pointed.
(7.2) Show for the Löwner ordering cone

$$\mathcal{S}^n_+ = \text{convex hull } \{xx^T \mid x \in \mathbb{R}^n\}.$$

(7.3) As an extension of Lemma 7.2 prove the following result: Let $X = \begin{pmatrix} A & B^T \\ B & C \end{pmatrix} \in \mathcal{S}^{k+l}$ with $A \in \mathcal{S}^k$, $C \in \mathcal{S}^l$ and $B \in \mathbb{R}^{(l,k)}$ be given, and assume that A is positive definite. Then we have for an arbitrary convex cone $K \subset \mathbb{R}^l$:

$$X \in C^{k+l}_{\mathbb{R}^k \times K} \quad \Longleftrightarrow \quad C - BA^{-1}B^T \in C^l_K.$$

(7.4) Show for arbitrary $A, B \in \mathcal{S}^n_+$

$$\langle A, B \rangle = 0 \quad \Longleftrightarrow \quad AB = 0_{\mathcal{S}^n}.$$

(7.5) Let A be a given symmetric (n, n) matrix. Show for an arbitrary $(j-i+1, j-i+1)$ block matrix A^{ij} $(1 \leq i \leq j \leq n)$ with

$$A^{ij}_{kl} = A_{i+k-1,\, i+l-1} \text{ for all } k, l \in \{1, \ldots, j-i+1\} :$$

$$A \text{ positive semidefinite} \quad \Longrightarrow \quad A^{ij} \text{ positive semidefinite.}$$

(7.6) Show that the linear semidefinite optimization problem

$$\begin{gathered} \min\ x_2 \\ \text{subject to the constraints} \\ -\begin{pmatrix} x_1 & 1 \\ 1 & x_2 \end{pmatrix} \preceq 0_{\mathcal{S}^2} \\ x_1, x_2 \in \mathbb{R} \end{gathered}$$

(where $\preceq$ denotes the Löwner partial ordering) is not solvable.

(7.7) Let the linear mapping $G : \mathbb{R}^2 \to \mathcal{S}^2$ with

$$G(x_1, x_2) = \begin{pmatrix} x_1 & x_2 \\ x_2 & 0 \end{pmatrix} \text{ for all } (x_1, x_2) \in \mathbb{R}^2$$

be given. Show that G does not fulfill the generalized Slater condition given in Theorem 7.12 for $C = \mathcal{S}^2_+$.

(7.8) Let $c \in \mathbb{R}^m$, $B \in \mathcal{S}^n$ and a linear mapping $A : \mathbb{R}^m \to \mathcal{S}^n$ with

$$A(x) = A^{(1)}x_1 + \ldots + A^{(m)}x_m \text{ for all } x \in \mathbb{R}^m$$

for $A^{(1)}, \ldots, A^{(m)} \in \mathcal{S}^n$ be given. Show that for the linear problem

$$\begin{gathered} \min\ c^T x \\ \text{subject to the constraints} \\ B \preccurlyeq A(x) \\ x \in \mathbb{R}^m \end{gathered}$$

the dual problem is given by

$$\begin{gathered} \max\ \langle B, U \rangle \\ \text{subject to the constraints} \\ \langle A^{(1)}, U \rangle = c_1 \\ \vdots \\ \langle A^{(m)}, U \rangle = c_m \\ U \in C^*. \end{gathered}$$

(7.9) Consider the linear semidefinite optimization problem

$$\begin{gathered} \min\ x_1 \\ \text{subject to the constraints} \\ \begin{pmatrix} 0 & -x_1 & 0 \\ -x_1 & -x_2 & 0 \\ 0 & 0 & -x_1 - 1 \end{pmatrix} \preceq 0_{\mathcal{S}^3} \\ x_1, x_2 \in \mathbb{R} \end{gathered}$$

(where $\preceq$ denotes the Löwner partial ordering). Give the corresponding dual problem and show that the extremal values of the primal and dual problem are not equal. Why is Theorem 7.12 not applicable?

8 Extension to Discrete-Continuous Problems

Many optimization problems in practice have continuous and discrete variables. Formally, one can plug the discrete variables in the superset (called $\hat{S}$) of the constraint set, but then this set is nonconvex and the Lagrange multiplier rule in Chap. 5 is not applicable and the duality theory in Chap. 6 is only limitedly applicable. For a Lagrange theory and a duality theory in discrete-continuous nonlinear optimization one needs a different approach,[13] which is developed in this chapter. The main key for such a theory is a special separation theorem for discrete sets. Using this theory we present optimality conditions as well as duality results.

It is known from Chap. 5 that the Lagrange multiplier rule in continuous optimization is based on the fact that to each constraint <u>one</u> Lagrange multiplier is associated, which is then used for the formulation of the Lagrange functional. In mixed discrete-continuous optimization it turns out that this point is completely different. In general, to each constraint of a discrete-continuous problem <u>more than one</u> Lagrange multiplier are associated.

8.1 Problem Formulation

In this chapter we follow the definition of a discrete set often used in discrete optimization.

[13] J. Jahn and M. Knossalla, "Lagrange theory of discrete-continuous nonlinear optimization", *Journal of Nonlinear and Variational Analysis* 2 (2018) 317–342.

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3_8

Definition 8.1 (discrete set).

A nonempty subset of a real linear space is called *discrete*, if its number of elements is either finite or countably infinite.

This definition differs from that in a topological setting where for every vector of the considered set there is a neighborhood so that no other vector of this set belongs to this neighborhood. In a real topological linear space every vector of such a set is also called *isolated*.

Example 8.2 (discrete set).

The number of elements of the set $A := \{0\} \cup \{\frac{1}{n}\}_{n \in \mathbb{N}}$ is countably infinite and, therefore, A is a discrete set. But there is no neighborhood of 0 so that no nonzero element of A belongs to such a neighborhood, i.e. 0 is not an isolated point.

Based on Definition 8.1 we generally write a discrete subset A of a real linear space as $A = \{a^i\}_{i \in N}$ with $N := \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$ or $N := \mathbb{N}$, where $a^1, a^2, \ldots$ are elements of the real linear space.

One can also extend Definition 8.1 to a more general class of sets.

Definition 8.3 (extendedly discrete set).

A nonempty subset A of a real linear space E is called *extendedly discrete*, if $A = \bigcup_{i \in N} A_i$ for $N := \{1, 2, \ldots, n\}$ (with $n \in \mathbb{N}$) or $N := \mathbb{N}$ and for nonempty sets $A_i \subset E$ $(i \in N)$.

It is obvious that every discrete set is also extendedly discrete. The concept of extendedly discrete sets is needed in the next section.

Next, we investigate discrete-continuous nonlinear optimization problems under the following assumption where discrete and continuous variables are indexed by a subscript d and c, respectively. For real normed spaces we also use these subscripts in a similar way.

The standard assumption of this chapter reads as follows:

$$\left.\begin{array}{l}\text{Let } (X_d, \|\cdot\|_{X_d}), (X_c, \|\cdot\|_{X_c}), (Y, \|\cdot\|_Y) \text{ and } (Z, \|\cdot\|_Z) \\ \text{be real normed spaces,} \\ \text{and let } C \subset Y \text{ be a convex cone in } Y; \\ \text{let } S_d \text{ be a discrete subset of } X_d, \text{ i.e. } S_d = \{x_d^i\}_{i\in N} \text{ for} \\ N := \{1, 2, \ldots, n\} (\text{with } n \in \mathbb{N}) \text{ or } N := \mathbb{N} \\ \text{with elements } x_d^1, x_d^2, \ldots \in X_d; \\ \text{let } S_c \text{ be a nonempty subset of } X_c; \\ \text{let } f : S_d \times S_c \to \mathbb{R}, g : S_d \times S_c \to Y \text{ and} \\ h : S_d \times S_c \to Z \text{ be given mappings;} \\ \text{furthermore let the constraint set} \\ S := \{(x_d, x_c) \in S_d \times S_c \mid g(x_d, x_c) \in -C,\ h(x_d, x_c) = 0_Z\} \\ \text{be nonempty.}\end{array}\right\} \tag{8.1}$$

Under this assumption we study the discrete-continuous optimization problem

$$\min_{(x_d, x_c)\in S} f(x_d, x_c). \tag{8.2}$$

Such a discrete-continuous optimization problem is much more complicated than standard problems of continuous optimization. For instance, the constraint set S of this problem can be written as

$$S = \bigcup_{i\in N} \{(x_d^i, x_c) \in X_d \times S_c \mid g(x_d^i, x_c) \in -C,\ h(x_d^i, x_c) = 0_Z\}$$

so that the constraint set is a union of layered subsets of $X_d \times X_c$. Figure 8.1 illustrates this multi-layered structure. Even if these layered subsets are convex, its union is not convex in general. Therefore, these nonconvex constraint sets cannot be treated with standard approaches of continuous optimization. It seems to be obvious that separation theorems with certain nonlinear separating functionals are needed.

8.2 Separation Theorems for Discrete Sets

It is well-known that separation results are the fundamental tool for the theoretical investigation of continuous optimization problems. In discrete-continuous optimization one needs special separation results being the topic of this section. Our first separation theorem can be formulated for extendedly discrete sets.

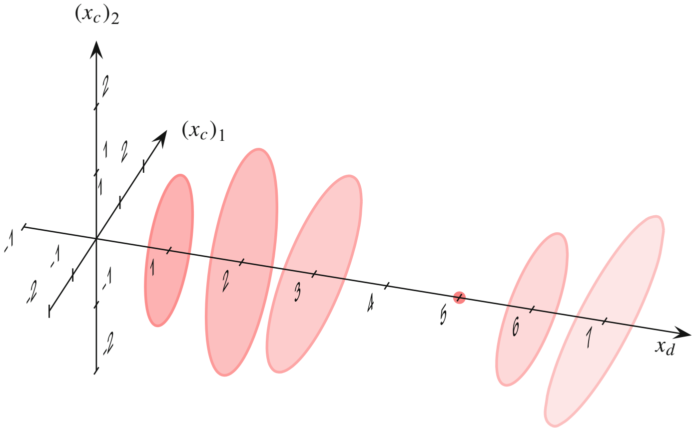

Fig. 8.1 Illustration of a possible constraint set of problem (8.2) with $S_d = \{1, 2, \ldots, 7\}$ and $S_c = \mathbb{R}^2$

Theorem 8.4 (first separation version).

Let $(E, \|\cdot\|)$ be a real normed space, let A be a nonempty subset of E, and let B be an extendedly discrete subset of E with the representation $B = \bigcup_{i\in N} int(conv(B_i))$ for $N := \{1, 2, \ldots, n\}$ (with $n \in \mathbb{N}$) or $N := \mathbb{N}$ where $B_i \subset E$ and $int(conv(B_i)) \neq \emptyset$ for all $i \in N$. Then $conv(A) \cap B = \emptyset$ if and only if for every $i \in N$ there exist a continuous linear functional $\ell^i \in E^* \backslash \{0_{E^*}\}$ and a real number α^i with

$$0 < \ell^i(y) + \alpha^i \text{ for all } y \in int(conv(B_i)), \tag{8.3}$$

and the inequalities

$$\sup_{i\in N}\{\ell^i(x) + \alpha^i\} \leq 0 \leq \inf_{i\in N}\{\ell^i(y) + \alpha^i\} \text{ for all } x \in conv(A) \text{ and all } y \in \bigcup_{i\in N} conv(B_i) \tag{8.4}$$

are fulfilled.

Proof

(a) For all $i \in N$ let there exist a continuous linear functional $\ell^i \in E^* \backslash \{0_{E^*}\}$ and a real number α^i so that the inequalities (8.3) and in (8.4) are fulfilled, i.e. we have for all $k \in N$

$$\sup_{i \in N}\{\ell^i(x) + \alpha^i\} \leq 0 < \ell^k(y) + \alpha^k \text{ for all } x \in \text{conv}(A) \text{ and all } y \in B.$$

Then it is obvious that $\text{conv}(A) \cap B = \emptyset$.

(b) Now assume that $\text{conv}(A) \cap B = \emptyset$, i.e. $\text{conv}(A) \cap \left(\bigcup_{i \in N} \text{int}(\text{conv}(B_i)) \right) = \emptyset$. This implies

$$\text{conv}(A) \cap \text{int}(\text{conv}(B_i)) = \emptyset \text{ for all } i \in N.$$

Let $i \in N$ be arbitrarily chosen. By Eidelheit's separation theorem (Theorem C.2) there exist a continuous linear functional $\ell^i \in E^* \backslash \{0_{E^*}\}$ and a real number $-\alpha^i$ with

$$\ell^i(x) \leq -\alpha^i \leq \ell^i(y) \text{ for all } x \in \text{conv}(A) \text{ and all } y \in \text{conv}(B_i)$$

and

$$-\alpha^i < \ell^i(y) \text{ for all } y \in \text{int}(\text{conv}(B_i)).$$

Then we get

$$\ell^i(x) + \alpha^i \leq 0 \leq \ell^i(y) + \alpha^i \text{ for all } x \in \text{conv}(A) \text{ and all } y \in \text{conv}(B_i) \tag{8.5}$$

and

$$0 < \ell^i(y) + \alpha^i \text{ for all } y \in \text{int}(\text{conv}(B_i)).$$

So, the inequality (8.3) is shown. Since the left inequality in (8.5) holds for all $i \in N$, we conclude

$$\sup_{i \in N}\{\ell^i(x) + \alpha^i\} \leq 0 \text{ for all } x \in \text{conv}(A)$$

and the left inequality in (8.4) is shown. The right inequality in (8.5) immediately implies

$$0 \leq \inf_{i \in N}\{\ell^i(y) + \alpha^i\} \text{ for all } y \in \bigcup_{i \in N} \text{conv}(B_i)$$

and the right inequality in (8.4) is shown. □

Figure 8.2 illustrates the separation result of Theorem 8.4. Since the set B in Theorem 8.4 may be nonconvex, a separation is carried out with a nonlinear

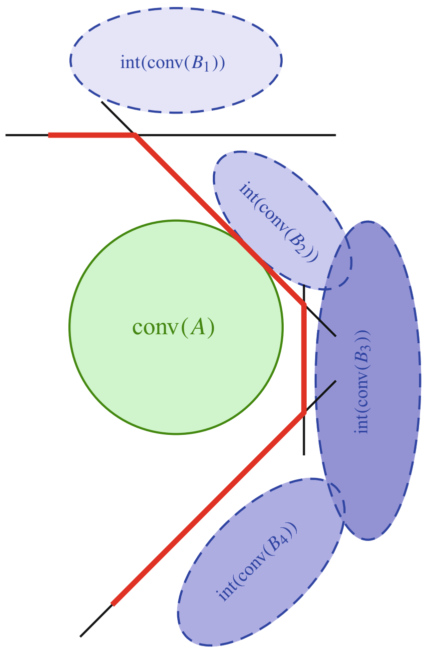

Fig. 8.2 Separation of the sets conv(A) and $B := \bigcup_{i\in\{1,2,3,4\}} \text{int}(\text{conv}(B_i))$ by Theorem 8.4

functional. From a geometrical point of view the separation of the sets conv(A) and B is done with a separating hyperface.

Remark 8.5 (first separation version).

(a) The proof of Theorem 8.4 is based on the standard case of convex separation. But here we work with the supremum and infimum of continuous affine linear functionals. If the index set N consists of finitely many indices, then the inf and sup terms reduce to a min and max term, respectively.
(b) Since the sets $A, B_1, B_2, \ldots$ in Theorem 8.4 may be discrete sets, the separation result can be applied to certain discrete problems.
(c) Theorem 8.4 can also be extended to a real linear space without normed structure.

Next we specialize Theorem 8.4 to a singleton set A.

Corollary 8.6 (separation with a singleton).

Let an element x of a real normed space $(E, \|\cdot\|)$ be given, and let B be an extendedly discrete subset of E with the representation $B = \bigcup_{i\in N} int(conv(B_i))$ for $N := \{1, 2, \ldots, n\}$ (with $n \in \mathbb{N}$) or $N := \mathbb{N}$ where $B_i \subset E$ and $int(conv(B_i)) \neq \emptyset$ for all $i \in N$. Then $x \notin B$ if and only if for every $i \in N$ there exists a continuous linear functional $\ell^i \in E^*\backslash\{0_{E^*}\}$ with

$$0 < \ell^i(y - x) \text{ for all } y \in int(conv(B_i)). \tag{8.6}$$

In this case we additionally have

$$0 \leq \inf_{i\in N}\{\ell^i(y - x)\} \text{ for all } y \in \bigcup_{i\in N} conv(B_i). \tag{8.7}$$

Proof

(a) First, for every $i \in N$ let $\ell^i \in E^*\backslash\{0_{E^*}\}$ be a continuous linear functional so that the inequality (8.6) holds. If we assume that $x \in B$, then we get a contradiction from the inequality (8.6), if we set $y = x$.

(b) Next let $x \notin B$, which is equivalent to the condition $0_E \notin B - \{x\}$. This condition is again equivalent to $\text{conv}(A) \cap (B - \{x\}) = \emptyset$ for $A := \{0_E\}$. By Theorem 8.4 for every $i \in N$ there exist a continuous linear functional $\ell^i \in E^*\backslash\{0_{E^*}\}$ and a real number α^i with

$$0 < \ell^i(y - x) + \alpha^i \text{ for all } y \in \text{int}(\text{conv}(B_i)), \tag{8.8}$$

and the inequalities

$$\sup_{i\in N}\{\alpha^i\} \leq 0 \leq \inf_{i\in N}\{\ell^i(y - x) + \alpha^i\} \text{ for all } y \in \bigcup_{i\in N} \text{conv}(B_i) \tag{8.9}$$

are fulfilled. The left inequality in (8.9) implies

$$\alpha^i \leq 0 \text{ for all } i \in N \tag{8.10}$$

and the right inequality in (8.9) gives

$$0 \leq \inf_{i\in N}\{\ell^i(y - x) + \alpha^i\} \leq \inf_{i\in N}\{\ell^i(y - x)\} \text{ for all } y \in \bigcup_{i\in N} \text{conv}(B_i).$$

So, the inequality (8.7) is shown. The proof of the inequality (8.6) follows from the inequality (8.8) together with the inequality (8.10). This completes the proof. □

Another strict separation result can be shown for a singleton set.

Theorem 8.7 (separation with a singleton).

Let an element x of a real normed space $(E, \|\cdot\|)$ be given, and let B be an extendedly discrete subset of E with the representation $B = \bigcup_{i\in N} cl(conv(B_i))$ for $N := \{1, 2, \ldots, n\}$ (with $n \in \mathbb{N}$) or $N := \mathbb{N}$ where $conv(B_i)$ are nonempty subsets of E for all $i \in N$. Then $x \notin B$ if and only if for every $i \in N$ there exists a continuous linear functional $\ell^i \in E^* \backslash \{0_{E^*}\}$ with

$$0 < \ell^i(y - x) \text{ for all } y \in cl(conv(B_i)). \tag{8.11}$$

Proof

(a) As in the part (a) of the proof of Corollary 8.6 we obtain that the condition (8.11) implies $x \notin B$.

(b) Now assume that $x \notin B$. Then for every $i \in \mathbb{N}$ the origin 0_E does not belong to the set $\text{cl}(\text{conv}(B_i)) - \{x\}$. By a strict separation theorem (Theorem C.3) for every $i \in \mathbb{N}$ there is a continuous linear functional $\ell^i \in E^* \backslash \{0_{E^*}\}$ with

$$0 = \ell^i(0_E) < \ell^i(y) \text{ for all } y \in \text{cl}(\text{conv}(B_i)) - \{x\}$$

implying

$$0 < \ell^i(y - x) \text{ for all } y \in \text{cl}(\text{conv}(B_i)),$$

which has to be shown. □

We now present a variant of the separation result of Theorem 8.4.

Theorem 8.8 (second separation version).

Let $(E, \|\cdot\|)$ be a real normed space, let A be a subset of E with $int(conv(A)) \neq \emptyset$, and let B be an extendedly discrete subset of E with the representation $B = \bigcup_{i\in N} conv(B_i)$ for $N := \{1, 2, \ldots, n\}$ (with $n \in \mathbb{N}$) or $N := \mathbb{N}$ where $\emptyset \neq B_i \subset E$ for all $i \in N$. Then $int(conv(A)) \cap B = \emptyset$ if and only if for every $i \in N$ there exist a continuous linear functional $\ell^i \in E^* \backslash \{0_{E^*}\}$ and a real number α^i with

$$0 < \ell^i(x) + \alpha^i \text{ for all } x \in int(conv(A)), \tag{8.12}$$

and the inequalities

$$\sup_{i\in N}\{\ell^i(y)+\alpha^i\} \leq 0 \leq \inf_{i\in N}\{\ell^i(x)+\alpha^i\} \text{ for all } x \in \mathrm{conv}(A) \text{ and all } y \in B \tag{8.13}$$

are fulfilled.

Proof

(a) First we assume that for every $i \in N$ there exist a continuous linear functional $\ell^i \in E^*\backslash\{0_{E^*}\}$ (with $i \in N$) and a real number α^i (with $i \in N$) so that the inequalities in (8.13) and (8.12) are fulfilled. Then we have for every $k \in N$

$$\sup_{i\in N}\{\ell^i(y)+\alpha^i\} \leq 0 < \ell^k(x)+\alpha^k \text{ for all } x \in \mathrm{int}(\mathrm{conv}(A)) \text{ and all } y \in B,$$

and this implies $\mathrm{int}(\mathrm{conv}(A)) \cap B = \emptyset$.

(b) Next, we assume that

$$\emptyset = \mathrm{int}(\mathrm{conv}(A)) \cap B = \mathrm{int}(\mathrm{conv}(A)) \cap \left(\bigcup_{i\in N} \mathrm{conv}(B_i)\right).$$

Then we have

$$\mathrm{int}(\mathrm{conv}(A)) \cap \mathrm{conv}(B_i) = \emptyset \text{ for all } i \in N.$$

For an arbitrarily chosen $i \in N$ we now apply Eidelheit's separation theorem (Theorem C.2). Then there exist a continuous linear functional $-\ell^i \in E^*\backslash\{0_{E^*}\}$ and a real number α^i with

$$-\ell^i(x) \leq \alpha^i \leq -\ell^i(y) \text{ for all } x \in \mathrm{conv}(A) \text{ and all } y \in \mathrm{conv}(B_i)$$

and

$$-\ell^i(x) < \alpha^i \text{ for all } x \in \mathrm{int}(\mathrm{conv}(A)).$$

Then it follows

$$\ell^i(y)+\alpha^i \leq 0 \leq \ell^i(x)+\alpha^i \text{ for all } x \in \mathrm{conv}(A) \text{ and all } y \in \mathrm{conv}(B_i)$$

and

$$0 < \ell^i(x) + \alpha^i \text{ for all } x \in \operatorname{int}(\operatorname{conv}(A)).$$

Hence, the inequality (8.12) is shown. Following the lines of the proof of Theorem 8.4 we conclude

$$\sup_{i \in N}\{\ell^i(y) + \alpha^i\} \leq 0 \leq \inf_{i \in N}\{\ell^i(x) + \alpha^i\} \text{ for all } x \in \operatorname{conv}(A) \text{ and all } y \in B,$$

which completes the proof. □

In contrast to the first separation result (Theorem 8.4) Theorem 8.8 has the advantage that the sets B_i (with $i \in N$) may be singletons, i.e. $B_i := \{b^i\}$ for some vectors $b^i \in E$. In this special case the extended discrete set $B = \bigcup_{i \in N} \operatorname{conv}(B_i) = \{b^1, b^2, \ldots\}$ is a discrete set. If A is a discrete set as well, i.e. $A := \{a^1, a^2, \ldots\}$ for some $a^1, a^2, \ldots \in E$, then we can separate the sets $\operatorname{int}(\operatorname{conv}\{a^1, a^2, \ldots\})$ and $\{b^1, b^2, \ldots\}$. Figure 8.3 illustrates this special type of separation. The nonlinear separating functional makes use of only three affine linear functionals where one of them could be dropped. So, in Fig. 8.3 one needs only two affine linear functionals for the nonlinear separation. This shows that the number of involved affine linear functionals may be less than expected.

It is well-known for a separation result in the finite dimensional space $\mathbb{R}^m$ that the interior of one of the two sets is not needed to be nonempty. This fact simplifies the two Theorems 8.4 and 8.8. This additional result is formulated in the following theorem.

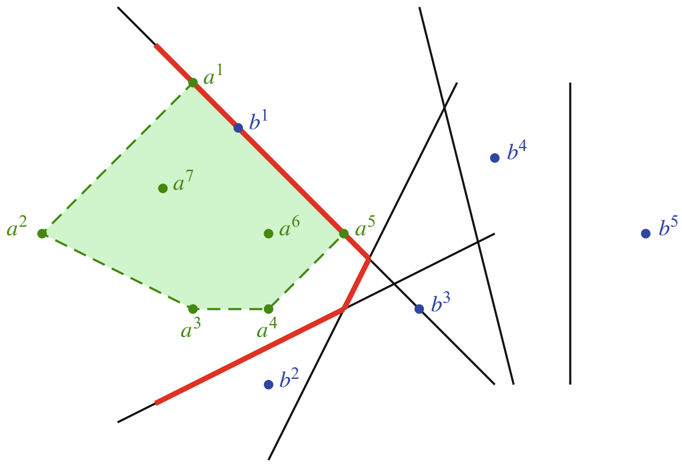

Fig. 8.3 Separation of the sets $\operatorname{int}(\operatorname{conv}\{a^1, a^2, \ldots, a^7\})$ and $\{b^1, b^2, \ldots, b^5\}$ by Theorem 8.8

Theorem 8.9 (separation in $\mathbb{R}^m$).

Let A be a nonempty subset of $\mathbb{R}^m$ (with $m \in \mathbb{N}$), let B be an extendedly discrete subset of $\mathbb{R}^m$ with the representation $B = \bigcup_{i\in N} conv(B_i)$ for $N := \{1, 2, \ldots, n\}$ (with $n \in \mathbb{N}$) or $N := \mathbb{N}$ where $\emptyset \neq B_i \subset \mathbb{R}^m$ for all $i \in N$. If $conv(A) \cap B = \emptyset$, then for every $i \in N$ there exist a vector $\ell^i \in \mathbb{R}^m \backslash \{0_{\mathbb{R}^m}\}$ and a real number α^i with

$$\sup_{i\in N}\{{\ell^i}^T y + \alpha^i\} \leq 0 \leq \inf_{i\in N}\{{\ell^i}^T x + \alpha^i\} \text{ for all } x \in conv(A) \text{ and all } y \in B.$$

Proof The proof follows the lines of the proof of Theorem 8.4 or 8.8. Assume that the sets $\text{conv}(A)$ and B are disjunct. This implies

$$\text{conv}(A) \cap \text{conv}(B_i) = \emptyset \text{ for all } i \in \mathbb{N}.$$

For every $i \in \mathbb{N}$ we then apply a finite dimensional separation theorem (e.g., see [134, Satz 2.22]). Hence, for every $i \in \mathbb{N}$ there exist a vector $\ell^i \in \mathbb{R}^m \backslash \{0_{\mathbb{R}^m}\}$ and a number $\alpha^i \in \mathbb{R}$ so that

$$-{\ell^i}^T x \leq \alpha^i \leq -{\ell^i}^T y \text{ for all } x \in \text{conv}(A) \text{ and all } y \in \text{conv}(B_i).$$

As in the proof of Theorem 8.4 we then get

$$\sup_{i\in N}\{{\ell^i}^T y + \alpha^i\} \leq 0 \leq \inf_{i\in N}\{{\ell^i}^T x + \alpha^i\} \text{ for all } x \in \text{conv}(A) \text{ and all } y \in B$$

and the proof is complete. □

8.3 Optimality Conditions

In this section optimality conditions in discrete-continuous nonlinear optimization are given, which are based on the separation result of Corollary 8.6. We begin with a first necessary optimality condition for the optimization problem (8.2).

Theorem 8.10 (necessary optimality condition).

Let the assumption (8.1) be satisfied and in addition, let $int(S_c) \neq \emptyset$ and $int(C) \neq \emptyset$. Let $\bar{x} = (x_d^j, \bar{x}_c)$ (for some $j \in N$) be a minimal solution of the discrete-continuous optimization problem (8.2). For every $i \in N$ let the set

$$B_i := \left\{ \begin{pmatrix} f(x_d^i, x_c) - f(\bar{x}) + \alpha \\ g(x_d^i, x_c) + y \\ h(x_d^i, x_c) \end{pmatrix} \in \mathbb{R} \times Y \times Z \;\middle|\; x_c \in int(S_c), \right.$$

$$\left. \alpha > 0,\ \ y \in int(C) \right\}$$

be convex and let $h(x_d^i, int(S_c))$ be an open set. Then for every $i \in N$ there exist a real number $\mu^i \geq 0$ and continuous linear functionals $\ell_g^i \in C^*$ and $\ell_h^i \in Z^*$ with $(\mu^i, \ell_g^i, \ell_h^i) \neq (0, 0_{Y^*}, 0_{Z^*})$, and the inequality

$$0 \leq \inf_{i \in N} \left\{ \mu^i \left(f(x_d^i, x_c) - f(\bar{x}) \right) + \ell_g^i \left(g(x_d^i, x_c) \right) + \ell_h^i \left(h(x_d^i, x_c) \right) \right\}$$
$$\text{for all } x_c \in S_c \tag{8.14}$$

and the equality

$$\ell_g^j \left(g(\bar{x}) \right) = 0 \tag{8.15}$$

are fulfilled.

Proof Let $\bar{x} = (x_d^j, \bar{x}_c)$ (for some $j \in N$) be a minimal solution of the optimization problem (8.2). Choose an arbitrary $i \in N$. For the nonempty set B_i we show $(0, 0_Y, 0_Z) \notin B_i$. Assume that $(0, 0_Y, 0_Z) \in B_i$. Then there are some $x_c \in \text{int}(S_c)$, some $\alpha > 0$ and some $y \in \text{int}(C)$ with

$$f(x_d^i, x_c) - f(\bar{x}) = -\alpha \tag{8.16}$$

$$g(x_d^i, x_c) = -y \tag{8.17}$$

$$h(x_d^i, x_c) = 0_Z. \tag{8.18}$$

With $x_c \in \text{int}(S_c)$ and $y \in \text{int}(C)$ the equations (8.17) and (8.18) imply $(x_d^i, x_c) \in S$. Because of $\alpha > 0$ the equation (8.16) contradicts the assumption that $\bar{x}$ is a minimal solution of problem (8.2). Hence, we conclude

$$(0, 0_Y, 0_Z) \notin B_i \text{ for all } i \in N,$$

which implies

$$(0, 0_Y, 0_Z) \notin \bigcup_{i \in N} B_i.$$

Since the sets B_i (with $i \in N$) are convex and open, the separation result in Corollary 8.6 is applicable and for every $i \in N$ there exist a real number μ^i and continuous linear functionals $\ell_g^i \in Y^*$ and $\ell_h^i \in Z^*$ with $(\mu^i, \ell_g^i, \ell_h^i) \neq (0, 0_{Y^*}, 0_{Z^*})$, and the inequality

$$\begin{aligned} 0 \le \inf_{i \in N} \Big\{ & \mu^i \left(f(x_d^i, x_c) - f(\bar{x}) + \alpha \right) + \ell_g^i \left(g(x_d^i, x_c) + y \right) \\ & + \ell_h^i \left(h(x_d^i, x_c) \right) \Big\} \text{ for all } x_c \in S_c,\ \alpha \ge 0 \text{ and } y \in C \end{aligned} \tag{8.19}$$

is fulfilled (notice that $S_c \subset \text{cl}(\text{int}(S_c))$ and $C \subset \text{cl}(\text{int}(C))$). For an arbitrary $k \in N$ we get with $x_c = \bar{x}_c$ and $y = 0_Y$ from the inequality (8.19)

$$\begin{aligned} 0 &\le \inf_{i \in N} \left\{ \mu^i \left(f(x_d^i, \bar{x}_c) - f(\bar{x}) + \alpha \right) + \ell_g^i \left(g(x_d^i, \bar{x}_c) \right) + \ell_h^i \left(h(x_d^i, \bar{x}_c) \right) \right\} \\ &\le \mu^k \left(f(x_d^k, \bar{x}_c) - f(\bar{x}) + \alpha \right) + \ell_g^k \left(g(x_d^k, \bar{x}_c) \right) + \ell_h^k \left(h(x_d^k, \bar{x}_c) \right) \\ &\quad \text{for all } \alpha \ge 0. \end{aligned}$$

This implies for some $\beta \in \mathbb{R}$

$$\mu^k \alpha \ge \beta \text{ for all } \alpha \ge 0$$

and $\mu^k \ge 0$. Since $k \in N$ is arbitrarily chosen we conclude

$$\mu^i \ge 0 \text{ for all } i \in N.$$

If we set $x_c = \bar{x}_c$ and $\alpha = 0$, we obtain from the inequality (8.19) for an arbitrary $k \in N$

$$\begin{aligned} 0 &\le \inf_{i \in N} \left\{ \mu^i \left(f(x_d^i, \bar{x}_c) - f(\bar{x}) \right) + \ell_g^i \left(g(x_d^i, \bar{x}_c) + y \right) + \ell_h^i \left(h(x_d^i, \bar{x}_c) \right) \right\} \\ &\le \mu^k \left(f(x_d^k, \bar{x}_c) - f(\bar{x}) \right) + \ell_g^k \left(g(x_d^k, \bar{x}_c) + y \right) + \ell_h^k \left(h(x_d^k, \bar{x}_c) \right) \\ &\quad \text{for all } y \in C. \end{aligned}$$

We then conclude for some $\gamma \in \mathbb{R}$

$$\ell_g^k(y) \geq \gamma \text{ for all } y \in C.$$

This inequality implies $\ell_g^k \in C^*$ and since $k \in N$ is arbitrarily chosen we have

$$\ell_g^i \in C^* \text{ for all } i \in N.$$

If we set $\alpha = 0$ and $y = 0_Y$ in the inequality (8.19), we immediately get the assertion (8.14). From the inequality (8.14) we obtain with $x_c = \bar{x}_c$

$$\begin{aligned} 0 &\leq \inf_{i \in N} \left\{ \mu^i \left(f(x_d^i, \bar{x}_c) - f(\bar{x}) \right) + \ell_g^i \left(g(x_d^i, \bar{x}_c) \right) + \ell_h^i \left(h(x_d^i, \bar{x}_c) \right) \right\} \\ &\leq \mu^j \underbrace{(f(\bar{x}) - f(\bar{x}))}_{=0} + \ell_g^j \left(g(\bar{x}) \right) + \underbrace{\ell_h^j(\underbrace{h(\bar{x})}_{=0_Z})}_{=0} \\ &= \ell_g^j \left(g(\bar{x}) \right) \end{aligned}$$

which implies $\ell_g^j(g(\bar{x})) \geq 0$. Since $g(\bar{x}) \in -C$ and $\ell_g^j \in C^*$, we also get $\ell_g^j(g(\bar{x})) \leq 0$. Hence, we conclude $\ell_g^j(g(\bar{x})) = 0$, and the proof is complete. □

Remark 8.11 (necessary optimality condition).

In Theorem 8.10 it is assumed that for every $i \in N$ the set B_i is convex. It is well-known that the set B_i (for every $i \in N$) is convex, if the set S_c is convex, the objective functional f is convex, the constraint mapping g is convex and the mapping h is affine linear. But these conditions can be weakened, if one extends the notion of convex-likeness (see Definition 6.5) to the interior of a cone.
Moreover, for every $i \in N$ it is assumed that the set $h(x_d^i, \text{int}(S_c))$ is open. For instance, this assumption is satisfied by the open mapping theorem, if X_c and Z are real Banach spaces and the mapping $h(x_d^i, \cdot)$ is continuous, linear and surjective.

Corollary 8.12 (necessary optimality condition with CQ).

Let the assumptions of Theorem 8.10 be satisfied. Again, let $\bar{x} = (x_d^j, \bar{x}_c)$ *(for some* $j \in N$ *and some* $\bar{x}_c \in S_c$*) denote a minimal solution of the discrete-continuous optimization problem (8.2). In addition, we assume*

$$\forall\, i \in N \;\; \exists\, x_c \in S_c : \;\; 0 > \ell_g^i \left(g(x_d^i, x_c) \right) + \ell_h^i \left(h(x_d^i, x_c) \right). \tag{8.20}$$

Then for every $i \in N$ *there exist continuous linear functionals* $\bar{\ell}_g^i \in C^*$ *and* $\bar{\ell}_h^i \in Z^*$*, and the inequality*

$$f(\bar{x}) \leq \inf_{i \in N} \left\{ f(x_d^i, x_c) + \bar{\ell}_g^i \left(g(x_d^i, x_c) \right) + \bar{\ell}_h^i \left(h(x_d^i, x_c) \right) \right\} \quad \text{for all } x_c \in S_c \tag{8.21}$$

and the equality

$$\bar{\ell}_g^j \left(g(\bar{x}) \right) = 0 \tag{8.22}$$

are fulfilled.

Proof By Theorem 8.10 for every $i \in N$ there exist a real number $\mu^i \geq 0$ and continuous linear functionals $\ell_g^i \in C^*$ and $\ell_h^i \in Z^*$ with $(\mu^i, \ell_g^i, \ell_h^i) \neq (0, 0_{Y^*}, 0_{Z^*})$, and the inequality (8.14) and the equality (8.15) are fulfilled. We first show that

$$\mu^i > 0 \text{ for all } i \in N. \tag{8.23}$$

Assume that there exists some $k \in N$ with $\mu^k = 0$. Then we conclude with the additional assumption (8.20) for some $x_c \in S_c$

$$\begin{aligned}
0 &> \ell_g^k \left(g(x_d^k, x_c) \right) + \ell_h^k \left(h(x_d^k, x_c) \right) \\
&= \mu^k \left(f(x_d^k, x_c) - f(\bar{x}) \right) + \ell_g^k \left(g(x_d^k, x_c) \right) + \ell_h^k \left(h(x_d^k, x_c) \right) \\
&\geq \inf_{i \in N} \left\{ \mu^i \left(f(x_d^i, x_c) - f(\bar{x}) \right) + \ell_g^i \left(g(x_d^i, x_c) \right) + \ell_h^i \left(h(x_d^i, x_c) \right) \right\}
\end{aligned}$$

which contradicts the inequality (8.14). So, the inequality (8.23) is proven.

For every $i \in N$ we now set $\bar{\ell}_g^i := \frac{1}{\mu^i}\ell_g^i \in C^*$ and $\bar{\ell}_h^i := \frac{1}{\mu^i}\ell_h^i \in Z^*$. From the inequality (8.14) it then follows

$$0 \leq \inf_{i\in N}\left\{f(x_d^i, x_c) - f(\bar{x}) + \bar{\ell}_g^i\left(g(x_d^i, x_c)\right) + \bar{\ell}_h^i\left(h(x_d^i, x_c)\right)\right\}$$
$$\text{for all } x_c \in S_c,$$

which implies

$$f(\bar{x}) \leq \inf_{i\in N}\left\{f(x_d^i, x_c) + \bar{\ell}_g^i\left(g(x_d^i, x_c)\right) + \bar{\ell}_h^i\left(h(x_d^i, x_c)\right)\right\}$$
$$\text{for all } x_c \in S_c.$$

Hence, the inequality (8.21) is shown. The equality (8.22) immediately follows from the equality (8.15). □

The condition (8.20) ensures that the multipliers μ^i (for all $i \in N$) are nonzero, i.e. it is a constraint qualification (CQ). This CQ extends the known Slater condition to discrete-continuous optimization problems.

The necessary optimality condition (8.21) can also be given in a different form.

Corollary 8.13 (necessary optimality condition).

Let the assumptions of Corollary 8.12 be satisfied, and let $\bar{x} = (x_d^j, \bar{x}_c)$ (for some $j \in N$ and some $\bar{x}_c \in S_c$) denote a minimal solution of the discrete-continuous optimization problem (8.2). Then for every $i \in N$ there exist continuous linear functionals $\bar{\ell}_g^i \in C^*$ and $\bar{\ell}_h^i \in Z^*$, and the inequality

$$\inf_{i\in N}\left\{f(x_d^i, \bar{x}_c) + \bar{\ell}_g^i\big(g(x_d^i, \bar{x}_c)\big) + \bar{\ell}_h^i\big(h(x_d^i, \bar{x}_c)\big)\right\}$$
$$\leq \inf_{i\in N}\left\{f(x_d^i, x_c) + \bar{\ell}_g^i\left(g(x_d^i, x_c)\right) + \bar{\ell}_h^i\left(h(x_d^i, x_c)\right)\right\}$$
$$\textit{for all } x_c \in S_c \qquad (8.24)$$

and the equality

$$\bar{\ell}_g^j\,(g(\bar{x})) = 0$$

are fulfilled.

Proof This corollary is a direct consequence of Corollary 8.12, if we prove the inequality (8.24). With the inequality (8.21) we obtain with $\bar{x} = (x_d^j, \bar{x}_c)$

$$
\begin{aligned}
&\inf_{i\in N} \left\{ f(x_d^i, \bar{x}_c) + \bar{\ell}_g^i\big(g(x_d^i, \bar{x}_c)\big) + \bar{\ell}_h^i\big(h(x_d^i, \bar{x}_c)\big) \right\} \\
&\le f(x_d^j, \bar{x}_c) + \bar{\ell}_g^j\big(g(x_d^j, \bar{x}_c)\big) + \bar{\ell}_h^j\big(h(x_d^j, \bar{x}_c)\big) \\
&= f(\bar{x}) + \underbrace{\bar{\ell}_g^j(g(\bar{x}))}_{=0} + \underbrace{\bar{\ell}_h^j(h(\bar{x}))}_{=0} \\
&= f(\bar{x}) \\
&\le \inf_{i\in N} \left\{ f(x_d^i, x_c) + \bar{\ell}_g^i\big(g(x_d^i, x_c)\big) + \bar{\ell}_h^i\big(h(x_d^i, x_c)\big) \right\} \text{ for all } x_c \in S_c,
\end{aligned}
$$

which has to be shown. □

Corollary 8.13 motivates the following definition of a Lagrange functional in discrete-continuous optimization.

Definition 8.14 (Lagrange functional).

Let the assumption (8.1) be satisfied. The functional $L : S_c \times \bigtimes_{i\in N} C^* \times \bigtimes_{i\in N} Z^* \to \mathbb{R}$ with

$$
\begin{aligned}
&L\big(x_c, (\ell_g^i)_{i\in N}, (\ell_h^i)_{i\in N}\big) \\
&= \inf_{i\in N} \left\{ f(x_d^i, x_c) + \ell_g^i\big(g(x_d^i, x_c)\big) + \ell_h^i\big(h(x_d^i, x_c)\big) \right\} \\
&\quad \text{for all } x_c \in S_c, \quad \ell_g^1, \ell_g^2, \ldots \in C^*, \quad \ell_h^1, \ell_h^2, \ldots \in Z^*
\end{aligned}
$$

is called *Lagrange functional* of the discrete-continuous optimization problem (8.2).

Remark 8.15 (Lagrange functional).

Using the Lagrange functional the result of Corollary 8.13 can be simplified. Under the assumptions of Corollary 8.13 for every minimal solution $\bar{x} = (x_d^j, \bar{x}_c)$ of the discrete-continuous optimization problem (8.2) there are continuous linear functionals $\bar{\ell}_g^i \in C^*$ (for all $i \in N$) and $\bar{\ell}_h^i \in Z^*$ (for all $i \in N$) with

$$
L\big(\bar{x}_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big) = \min_{x_c\in S_c} L\big(x_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big).
$$

Corollary 8.16 (saddle point property).

Let the assumptions of Corollary 8.13 be satisfied, and let $\bar{x} = (x_d^j, \bar{x}_c)$ *(for some* $j \in N$ *and some* $\bar{x}_c \in S_c$*) denote a minimal solution of the discrete-continuous optimization problem (8.2). Then for every* $i \in N$ *there exist continuous linear functionals* $\bar{\ell}_g^i \in C^*$ *and* $\bar{\ell}_h^i \in Z^*$*, and the Lagrange functional L fulfills the inequalities*

$$L\big(\bar{x}_c, (\ell_g^i)_{i\in N}, (\ell_h^i)_{i\in N}\big) \le L\big(\bar{x}_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big)$$
$$\le L\big(x_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big)$$

$$\text{for all } x_c \in S_c, \quad \ell_g^1, \ell_g^2, \ldots \in C^*, \quad \ell_h^1, \ell_h^2, \ldots \in Z^*. \tag{8.25}$$

Proof The right inequality in (8.25) is shown by Corollary 8.13. For the proof of the left inequality in (8.25) choose for every $i \in N$ arbitrary linear functionals $\ell_g^i \in C^*$ and $\ell_h^i \in Z^*$. Then we get with Corollary 8.12

$$\begin{aligned}
&L\big(\bar{x}_c, (\ell_g^i)_{i\in N}, (\ell_h^i)_{i\in N}\big)\\
&= \inf_{i\in N} \Big\{ f(x_d^i, \bar{x}_c) + \ell_g^i\big(g(x_d^i, \bar{x}_c)\big) + \ell_h^i\big(h(x_d^i, \bar{x}_c)\big)\Big\}\\
&\le f(\bar{x}) + \underbrace{\ell_g^j(g(\bar{x}))}_{\le 0} + \underbrace{\ell_h^j(h(\bar{x}))}_{=0}\\
&\le f(\bar{x})\\
&\le L\big(\bar{x}_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big),
\end{aligned}$$

which completes the proof. □

The inequalities in (8.25) mean that $\big(\bar{x}_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big)$ is a saddle point of the Lagrange functional L. Corollary 8.16 implies a necessary optimality condition using the subdifferential of the Lagrangian.

Corollary 8.17 (subdifferential version).

Let the assumptions of Corollary 8.13 be satisfied, and in addition, let $S_c = X_c$*. Let* $\bar{x} = (x_d^j, \bar{x}_c)$ *(for some* $j \in N$ *and some* $\bar{x}_c \in X_c$*) denote a minimal solution of the discrete-continuous optimization problem (8.2). Then for every* $i \in N$ *there exist continuous linear functionals* $\bar{\ell}_g^i \in C^*$ *and* $\bar{\ell}_h^i \in Z^*$ *with*

$$0_{X_c^*} \in \partial_{x_c} L\big(\bar{x}_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big) \tag{8.26}$$

and

$$\bar{\ell}_g^j\,(g(\bar{x})) = 0, \tag{8.27}$$

where

$$\begin{aligned}
&\partial_{x_c} L\big(\bar{x}_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big)\\
&:= \Big\{\ell \in X_c^* \mid L\big(x_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big) \geq L\big(\bar{x}_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big)\\
&\qquad\qquad + \ell(x_c - \bar{x}_c) \text{ for all } x_c \in X_c\Big\}
\end{aligned}$$

denotes the subdifferential (w.r.t. x_c) of the Lagrange functional $L\big(\cdot, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big)$ at $\bar{x}_c$.

Proof By Corollary 8.16 for every $i \in N$ there exist continuous linear functionals $\bar{\ell}_g^i \in C^*$ and $\bar{\ell}_h^i \in Z^*$ with

$$L\big(x_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big) \geq L\big(\bar{x}_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big) + 0_{X_c^*}(x_c - \bar{x}_c)$$
$$\text{for all } x_c \in X_c,$$

and the condition (8.26) is proven. The condition (8.27) is already shown in Corollary 8.13. □

Using the Lagrange functional L the inequality (8.21) as part of a necessary optimality condition in Corollary 8.12 can be written as

$$f(\bar{x}) \leq L\big(x_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big) \text{ for all } x_c \in S_c.$$

The next theorem says that this inequality is even a sufficient optimality condition.

Theorem 8.18 (sufficient optimality condition).

Let the assumption (8.1) be satisfied, and let $\bar{x} := (x_d^j, \bar{x}_c)$ be a feasible vector of the discrete-continuous optimization problem (8.2). If for every $i \in N$ there exist continuous linear functionals $\bar{\ell}_g^i \in C^*$ and $\bar{\ell}_h^i \in Z^*$ and if the inequality

$$f(\bar{x}) \leq L\big(x_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big) \text{ for all } x_c \in S_c \tag{8.28}$$

is satisfied, then $\bar{x}$ is a minimal solution of problem (8.2).

Proof Let $(x_d^k, x_c) \in S$ (with $k \in N$) be arbitrarily chosen. Then it follows from the inequality (8.28)

$$\begin{aligned} f(\bar{x}) &\leq L\big(x_c, (\bar{\ell}_g^i)_{i\in N}, (\bar{\ell}_h^i)_{i\in N}\big) \\ &= \inf_{i\in N} \left\{ f(x_d^i, x_c) + \bar{\ell}_g^i \left(g(x_d^i, x_c)\right) + \bar{\ell}_h^i \left(h(x_d^i, x_c)\right)\right\} \\ &\leq f(x_d^k, x_c) + \underbrace{\bar{\ell}_g^k\big(g(x_d^k, x_c)\big)}_{\leq 0} + \underbrace{\bar{\ell}_h^k\big(h(x_d^k, x_c)\big)}_{=0} \\ &\leq f(x_d^k, x_c). \end{aligned}$$

This implies that $\bar{x}$ is a minimal solution of problem (8.2). □

8.3.1 Specialization to Discrete Sets with Finite Cardinality

The optimality conditions are now specialized to discrete-continuous optimization problems with finitely many discrete points, i.e. we have $S_d = \{x_d^1, \ldots, x_d^n\}$ for some $n \in \mathbb{N}$.

We now specialize the assumption (8.1):

$$\left.\begin{array}{l}\text{Let the assumption (8.1) be satisfied with } N := \{1, \ldots, n\} \\ \text{for some } n \in N.\end{array}\right\} \tag{8.29}$$

Theorem 8.19 (multiplier rule).

Let the assumption (8.29) and the assumptions of Corollary 8.12 be satisfied, and let $\bar{x} = (x_d^j, \bar{x}_c)$ (for some $j \in \{1, \ldots, n\}$ and some $\bar{x}_c \in S_c$) denote a minimal solution of the discrete-continuous optimization problem (8.2). For every $i \in \{1, \ldots, n\}$ let the functional $f(x_d^i, \cdot)$ and the mappings $g(x_d^i, \cdot)$ and $h(x_d^i, \cdot)$ be Fréchet differentiable at $\bar{x}_c$. Then there exist continuous linear functionals $\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n \in C^*$ and $\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n \in Z^*$, and the inequality

$$\min_{i\in I(\bar{x}_c)} \left\{\left(f'(x_d^i, \bar{x}_c) + \bar{\ell}_g^i\big(g'(x_d^i, \bar{x}_c)\big) + \bar{\ell}_h^i\big(h'(x_d^i, \bar{x}_c)\big)\right)(x_c - \bar{x}_c)\right\}$$
$$\geq 0 \text{ for all } x_c \in S_c \tag{8.30}$$

and the equality

$$\bar{\ell}_g^j\,(g(\bar{x})) = 0 \tag{8.31}$$

are fulfilled. Here we use the abbreviation

$$I(\bar{x}_c) := \Big\{ i \in \{1, \ldots, n\} \;\Big|\; f(x_d^i, \bar{x}_c) + \bar{\ell}_g^i \Big(g(x_d^i, \bar{x}_c) \Big) + \bar{\ell}_h^i \Big(h(x_d^i, \bar{x}_c) \Big) \\ = \min_{1 \le k \le n} \big\{ f(x_d^k, \bar{x}_c) + \bar{\ell}_g^k \big(g(x_d^k, \bar{x}_c) \big) + \bar{\ell}_h^k \big(h(x_d^k, \bar{x}_c) \big) \big\} \Big\}. \tag{8.32}$$

Proof Let $\bar{x} = (x_d^j, \bar{x}_c)$ (for some $j \in \{1, \ldots, n\}$ and some $\bar{x}_c \in S_c$) be a minimal solution of problem (8.2). By Remark 8.15 there exist continuous linear functionals $\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n \in C^*$ and $\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n \in Z^*$ with

$$L \Big(\bar{x}_c, (\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n), (\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n) \Big) = \min_{x_c \in S_c} L \Big(x_c, (\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n), (\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n) \Big)$$

where L denotes the Lagrange functional, i.e.

$$\begin{aligned} & L \Big(x_c, (\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n), (\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n) \Big) \\ &= \inf_{1 \le i \le n} \Big\{ f(x_d^i, x_c) + \bar{\ell}_g^i \Big(g(x_d^i, x_c) \Big) + \bar{\ell}_h^i \Big(h(x_d^i, x_c) \Big) \Big\} \\ &= \min_{1 \le i \le n} \Big\{ f(x_d^i, x_c) + \bar{\ell}_g^i \Big(g(x_d^i, x_c) \Big) + \bar{\ell}_h^i \Big(h(x_d^i, x_c) \Big) \Big\} \text{ for all } x_c \in S_c. \end{aligned}$$

Since the equation (8.31) is already shown in Corollary 8.12, it remains to prove the inequality (8.30). $\bar{x}_c$ is a minimal point of the Lagrange functional $L\big(\cdot, (\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n), (\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n)\big)$ on S_c and, therefore, we obtain the known necessary optimality condition

$$L' \Big(\bar{x}_c, (\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n), (\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n) \Big) (x_c - \bar{x}_c) \ge 0 \text{ for all } x_c \in S_c \tag{8.33}$$

where L' denotes the Gâteaux derivative (compare Theorem 3.8,(a) for directional derivatives). For simplification, for every $i \in \{1, \ldots, n\}$ we use the functional $\varphi_i : S_c \to \mathbb{R}$ with

$$\varphi_i(x_c) = f(x_d^i, x_c) + \bar{\ell}_g^i \Big(g(x_d^i, x_c) \Big) + \bar{\ell}_h^i \Big(h(x_d^i, x_c) \Big) \text{ for all } x_c \in S_c.$$

Since $f(x_d^i,\cdot)$, $\bar{\ell}_g^i\left(g(x_d^i,\cdot)\right)$ and $\bar{\ell}_h^i\left(h(x_d^i,\cdot)\right)$ are Fréchet differentiable at $\bar{x}_c$, we obtain the Fréchet derivative $\varphi_i'(\bar{x}_c)(\cdot)$ with

$$\begin{aligned}
&\varphi_i'(\bar{x}_c)(d)\\
&\quad= f'(x_d^i,\cdot)(\bar{x}_c)(d) + \left(\bar{\ell}_g^i\left(g(x_d^i,\cdot)\right)\right)'(\bar{x}_c)(d) + \left(\bar{\ell}_h^i\left(h(x_d^i,\cdot)\right)\right)'(\bar{x}_c)(d)\\
&\quad= \left(f'(x_d^i,\bar{x}_c) + \bar{\ell}_g^i\left(g'(x_d^i,\bar{x}_c)\right) + \bar{\ell}_h^i\left(h'(x_d^i,\bar{x}_c)\right)\right)(d)\\
&\qquad \text{for all } d \in X_c. \qquad (8.34)
\end{aligned}$$

The Gâteaux derivative of the Lagrange functional at $\bar{x}_c$ can be written as

$$L'\left(\bar{x}_c, (\bar{\ell}_g^1,\ldots,\bar{\ell}_g^n), (\bar{\ell}_h^1,\ldots,\bar{\ell}_h^n)\right)(d) = \left(\min_{1\le i\le n}\{\varphi_i(\cdot)\}\right)'(\bar{x}_c)(d) \quad \text{for all } d \in X_c. \qquad (8.35)$$

If we notice that with

$$I(\bar{x}_c) = \left\{i \in \{1,\ldots,n\} \,\middle|\, \varphi_i(\bar{x}_c) = \min_{1\le k\le n}\{\varphi_k(\bar{x}_c)\}\right\}$$

we have

$$\min_{1\le k\le n}\{\varphi_k(\bar{x}_c)\} = \varphi_i(\bar{x}_c) \text{ for all } i \in I(\bar{x}_c)$$

and because of the continuity of the functionals $\varphi_1,\ldots,\varphi_n$ at $\bar{x}_c$

$$\min_{1\le i\le n}\{\varphi_i(\bar{x}_c+\lambda d)\} = \min_{i\in I(\bar{x}_c)}\{\varphi_i(\bar{x}_c+\lambda d)\} \text{ for all } d \in X_c \text{ and all } \lambda \text{ sufficiently close to } 0,$$

we then get the Gâteaux derivative

$$\begin{aligned}
&\left(\min_{1\le i\le n}\{\varphi_i(\cdot)\}\right)'(\bar{x}_c)(d)\\
&= \lim_{\lambda\to 0}\frac{1}{\lambda}\left(\min_{1\le i\le n}\{\varphi_i(\bar{x}_c+\lambda d)\} - \min_{1\le i\le n}\{\varphi_i(\bar{x}_c)\}\right)\\
&= \lim_{\lambda\to 0}\frac{1}{\lambda}\left(\min_{i\in I(\bar{x}_c)}\{\varphi_i(\bar{x}_c+\lambda d)\} - \min_{1\le i\le n}\{\varphi_i(\bar{x}_c)\}\right)\\
&= \lim_{\lambda\to 0}\ \min_{i\in I(\bar{x}_c)}\left\{\frac{1}{\lambda}\Big(\varphi_i(\bar{x}_c+\lambda d)\} - \varphi_i(\bar{x}_c)\Big)\right\}\\
&= \min_{i\in I(\bar{x}_c)}\left\{\varphi_i'(\bar{x}_c)(d)\right\} \text{ for all } d \in X_c
\end{aligned}$$

(compare Exercise (8.3) for directional derivatives) where φ_i' (with $i \in I(\bar{x}_c)$) denotes the Fréchet derivative, which equals the Gâteaux derivative. With the equation (8.35) we then conclude

$$L'\left(\bar{x}_c, (\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n), (\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n)\right)(d) = \min_{i \in I(\bar{x}_c)} \left\{\varphi_i'(\bar{x}_c)(d)\right\} \text{ for all } d \in X_c,$$

and with the inequality (8.33) and the equality (8.34) we obtain

$$\min_{i \in I(\bar{x}_c)} \left\{\left(f'(x_d^i, \bar{x}_c) + \bar{\ell}_g^i\big(g'(x_d^i, \bar{x}_c)\big) + \bar{\ell}_h^i\big(h'(x_d^i, \bar{x}_c)\big)\right)(x_c - \bar{x}_c)\right\} \geq 0$$
$$\text{for all } x_c \in S_c,$$

which has to be shown. □

If the set S_c equals the whole linear space X_c, we can simplify the result of Theorem 8.19.

Corollary 8.20 (specialized multiplier rule).

Let the assumptions of Theorem 8.19 be satisfied and, in addition, let $S_c = X_c$. Again, let $\bar{x} = (x_d^j, \bar{x}_c)$ (for some $j \in \{1, \ldots, n\}$ and some $\bar{x}_c \in S_c$) denote a minimal solution of the discrete-continuous optimization problem (8.2). Then there exist continuous linear functionals $\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n \in C^*$ and $\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n \in Z^*$, and the inequality

$$f'(x_d^i, \bar{x}_c) + \bar{\ell}_g^i\big(g'(x_d^i, \bar{x}_c)\big) + \bar{\ell}_h^i\big(h'(x_d^i, \bar{x}_c)\big) = 0_{X_c^*} \text{ for all } i \in I(\bar{x}_c) \tag{8.36}$$

(where $I(\bar{x}_c)$ is defined in (8.32)) and the equality

$$\bar{\ell}_g^j\,(g(\bar{x})) = 0 \tag{8.37}$$

are fulfilled.

Proof We write the inequality (8.30) as

$$\left(f'(x_d^i, \bar{x}_c) + \bar{\ell}_g^i\big(g'(x_d^i, \bar{x}_c)\big) + \bar{\ell}_h^i\big(h'(x_d^i, \bar{x}_c)\big)\right)(x_c - \bar{x}_c) \geq 0$$
$$\text{for all } i \in I(\bar{x}_c) \text{ and } x_c \in S_c.$$

With $S_c = X_c$ and the linearity of the Fréchet derivative we then get

$$f'(x_d^i, \bar{x}_c) + \bar{\ell}_g^i\big(g'(x_d^i, \bar{x}_c)\big) + \bar{\ell}_h^i\big(h'(x_d^i, \bar{x}_c)\big) = 0_{X_c^*} \text{ for all } i \in I(\bar{x}_c)$$

and the condition (8.36) is shown. □

The optimality conditions (8.36), (8.37) which are valid in infinite dimensional spaces (also compare Corollary 5.4) are already very close to the Karush-Kuhn-Tucker (KKT) conditions in finite dimensional spaces. KKT conditions are formulated in the next corollary, which is a direct consequence of Corollary 8.20.

Corollary 8.21 (KKT conditions).

Let the assumptions of Theorem 8.19 be satisfied and, in addition, let $X_d = \mathbb{R}^{n_d}$, $X_c = S_c = \mathbb{R}^{n_c}$, $Y = \mathbb{R}^m$, $Z = \mathbb{R}^p$ and $C = \mathbb{R}^{n_c}_+$. Again, let $\bar{x} = (x_d^j, \bar{x}_c)$ (for some $j \in \{1, \ldots, n\}$ and some $\bar{x}_c \in S_c$) denote a minimal solution of the discrete-continuous optimization problem (8.2). For every $i \in \{1, \ldots, n\}$ let $f(x_d^i, \cdot)$, $g_1(x_d^i, \cdot), \ldots, g_m(x_d^i, \cdot)$ and $h_1(x_d^i, \cdot), \ldots, h_p(x_d^i, \cdot)$ be differentiable at $\bar{x}_c$. Then there exist vectors $u^1, \ldots, u^n \in \mathbb{R}^{n_c}_+$ and $v^1, \ldots, v^n \in \mathbb{R}^{n_c}$ with

$$\nabla_{x_c} f(x_d^i, \bar{x}_c) + \sum_{k=1}^{m} u_k^i \nabla_{x_c} g_k(x_d^i, \bar{x}_c) + \sum_{k=1}^{p} v_k^i \nabla_{x_c} h_k(x_d^i, \bar{x}_c) = 0_{\mathbb{R}^{n_c}}$$

$$\text{for all } i \in I(\bar{x}_c)$$

(where $I(\bar{x}_c)$ is defined in (8.32)) and

$$u_k^j g_k(x_d^j, \bar{x}_c) = 0 \text{ for all } k \in \{1, \ldots, m\}.$$

Example 8.22 (KKT conditions).

In the assumption (8.29) we set $X_d = X_c = Y = \mathbb{R}$, $S_c = \mathbb{R}$ and $S_d = \{-10, -8, -6, \ldots, 10\}$ (i.e. we have $n = 11$). Let $\psi : S_d \to \mathbb{R}$ be an arbitrary function with

$$\psi(x_d^i) > 0 \text{ for all } i \in \{1, \ldots, n\},$$

and let x_d^j for some $j \in \{1, \ldots, 11\}$ be a unique minimal point of ψ, i.e.

$$\psi(x_d^j) < \psi(x_d^i) \text{ for all } i \in \{1, \ldots, 11\} \backslash \{j\}.$$

Then we consider the discrete-continuous optimization problem

$$\min\ \psi(x_d)\cdot {x_c}^2$$
subject to the constraints
$$x_c \geq 1$$
$$x_d \in \{-10,-8,\ldots,10\},\quad x_c \in \mathbb{R}.$$

It is evident that $\bar{x} := (x_d^j, 1)$ is the unique minimal solution of this optimization problem. Next, we define the objective function $f : S_d \times \mathbb{R} \to \mathbb{R}$ with

$$f(x_d, x_c) = \psi(x_d)\cdot {x_c}^2 \text{ for all } x_d \in S_d \text{ and all } x_c \in \mathbb{R}$$

and the constraint function $g : S_d \times \mathbb{R} \to \mathbb{R}$ with

$$g(x_d, x_c) = 1 - x_c \text{ for all } x_d \in S_d \text{ and all } x_c \in \mathbb{R}.$$

Notice that there is no equality constraint. Then the Lagrangian L can be written as

$$\begin{aligned} L\big(x_c, (u^1,\ldots,u^{11})\big) &= \inf_{1\leq i\leq 11}\Big\{f(x_d^i, x_c) + u^i g(x_d^i, x_c)\Big\} \\ &= \min_{1\leq i\leq 11}\Big\{\psi(x_d^i)\cdot {x_c}^2 + u^i(1-x_c)\Big\} \\ &\qquad \text{for all } x_c \in \mathbb{R} \text{ and all } u^1,\ldots,u^{11} \in \mathbb{R}_+. \end{aligned}$$

For the determination of the index set $I(1)$ we notice that

$$\begin{aligned} f(x_d^j, 1) + u^j g(x_d^j, 1) &= \psi(x_d^j) \\ &= \min_{1\leq i\leq 11} \psi(x_d^i) \\ &= \min_{1\leq i\leq 11}\Big\{f(x_d^i, 1) + u^i g(x_d^i, 1)\Big\}, \end{aligned}$$

i.e. $j \in I(1)$. Since x_d^j is a unique minimal point of ψ, we obtain for every $i \in \{1,\ldots,11\}\backslash\{j\}$

$$\begin{aligned} f(x_d^i, 1) + u^i g(x_d^i, 1) &= \psi(x_d^i) \\ &> \min_{1\leq k\leq 11} \psi(x_d^k) \\ &= \min_{1\leq k\leq 11}\Big\{f(x_d^k, 1) + u^k g(x_d^k, 1)\Big\}, \end{aligned}$$

i.e. $i \notin I(1)$. Hence, we conclude $I(1) = \{j\}$. Then we get the KKT condition

$$\begin{aligned} 0 &= \nabla_{x_c} f(x_d^j, 1) + u^j \nabla_{x_c} g(x_d^j, 1) \\ &= 2\psi(x_d^j) - u^j, \end{aligned}$$

which gives $u^j = 2\psi(x_d^j) > 0$. So, the minimal solution $\bar{x}$ fulfills the necessary optimality conditions in Corollary 8.21.

Finally we show under which assumptions the optimality conditions in Theorem 8.19 are sufficient optimality conditions, at least in a local sense.

Theorem 8.23 (sufficient optimality condition).

Let Assumption 8.29 be satisfied, and let $\bar{x} = (x_d^j, \bar{x}_c) \in S$ (for some $j \in \{1, \ldots, n\}$ and some $\bar{x}_c \in S_c$) be a feasible vector of the discrete-continuous optimization problem (8.2). For every $i \in \{1, \ldots, n\}$ let the functional $f(x_d^i, \cdot)$ and the mappings $g(x_d^i, \cdot)$ and $h(x_d^i, \cdot)$ be Fréchet differentiable at $\bar{x}_c$. Let there exist continuous linear functionals $\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n \in C^*$ and $\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n \in Z^*$, and let the inequality (8.30) and the equality (8.31) be satisfied. Moreover, let there exist some $\varepsilon > 0$ so that

$$j \in I(x_c) \text{ for all } x_c \in S_c \cap B(\bar{x}_c, \varepsilon)$$

(where $I(x_c)$ is defined in (8.32) and $B(\bar{x}_c, \varepsilon)$ denotes the ball with center $\bar{x}_c$ and radius ε) and the functional $f(x_d^j, \cdot) + \bar{\ell}_g^j\big(g(x_d^j, \cdot)\big) + \bar{\ell}_h^j\big(h(x_d^j, \cdot)\big)$ is convex on $S_c \cap B(\bar{x}_c, \varepsilon)$. Then $\bar{x}$ is a local minimal solution of the discrete-continuous optimization problem (8.2).

Proof From the proof of Theorem 8.19 it is evident that the inequality (8.30) is equivalent to the inequality (8.33). Since by assumption the Lagrange functional $L\left(\cdot, (\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n), (\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n)\right)$ with

$$\begin{aligned} &L\left(x_c, (\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n), (\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n)\right) \\ &= \min_{1 \le 1 \le n} \left\{ f(x_d^i, x_c) + \bar{\ell}_g^i \left(g(x_d^i, x_c) \right) + \bar{\ell}_h^i \left(h(x_d^i, x_c) \right) \right\} \\ &= f(x_d^j, x_c) + \bar{\ell}_g^j \left(g(x_d^j, x_c) \right) + \bar{\ell}_h^j \left(h(x_d^j, x_c) \right) \text{ for all } x_c \in S_c \cap B(\bar{x}_c, \varepsilon) \end{aligned}$$

is convex on $S_c \cap B(\bar{x}, \varepsilon)$, we conclude with Theorem 3.16 and the equality (8.33)

$$\begin{aligned}
& L\left(x_c, (\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n), (\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n)\right) \\
& \geq L\left(\bar{x}_c, (\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n), (\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n)\right) \\
& \quad \underbrace{+ L'\left(\bar{x}_c, (\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n), (\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n)\right)(x_c - \bar{x}_c)}_{\geq 0} \\
& \geq L\left(\bar{x}_c, (\bar{\ell}_g^1, \ldots, \bar{\ell}_g^n), (\bar{\ell}_h^1, \ldots, \bar{\ell}_h^n)\right) \\
& = f(x_d^j, \bar{x}_c) + \bar{\ell}_g^j\left(g(x_d^j, \bar{x}_c)\right) + \bar{\ell}_h^j\left(h(x_d^j, \bar{x}_c)\right) \\
& = f(\bar{x}) + \underbrace{\bar{\ell}_g^j(g(\bar{x}))}_{=0} + \underbrace{\bar{\ell}_h^j(h(\bar{x}))}_{=0} \\
& = f(\bar{x}) \text{ for all } x_c \in S_c \cap B(\bar{x}_c, \varepsilon).
\end{aligned}$$

If we then apply Theorem 8.18 with the set $S_c \cap B(\bar{x}_c, \varepsilon)$ instead of the set S_c, $\bar{x}$ is a locally minimal solution of the discrete-continuous optimization problem (8.2). □

Remark 8.24 (sufficient optimality condition).

Since in general, the constraint set of a discrete-continuous optimization problem is nonconvex (compare Fig. 8.1), a local minimal solution does not need to be a (global) minimal solution, if the objective functional is convex (compare Theorem 2.16).

Example 8.25 (sufficient optimality condition).

We investigate the discrete-continuous optimization problem in Example 8.22. Since x_d^j (for some $j \in \{1, \ldots, 11\}$) is a unique minimal point of ψ, we have

$$\psi(x_d^j) < \min_{\substack{1 \leq i \leq 11 \\ i \neq j}} \left\{\psi(x_d^i)\right\}.$$

It is already discussed in Example 8.22 that $\bar{x} = (x_d^j, 1)$ fulfills the KKT conditions. Notice that the Lagrange multipliers $u^1, \ldots, u^{11} \geq 0$ are given as $u^i = 2\psi(x_d^i)$ for all $i \in \{1, \ldots, 11\}$. For a sufficiently small $\varepsilon > 0$ we then conclude

$$\begin{aligned} L\big(x_c, (u^1, \ldots, u^{11})\big) &= \min_{1 \le i \le 11} \Big\{ \psi(x_d^i) \cdot {x_c}^2 + u^i \cdot (1 - x_c) \Big\} \\ &= \min_{1 \le i \le 11} \Big\{ \psi(x_d^i) \cdot {x_c}^2 + 2\psi(x_d^i) \cdot (1 - x_c) \Big\} \\ &= \min_{1 \le i \le 11} \Big\{ \psi(x_d^i) + \psi(x_d^i) \cdot (x_c - 1)^2 \Big\} \\ &= \psi(x_d^j) + \psi(x_d^j) \cdot (x_c - 1)^2 \\ &\quad \text{for all } x_c \in [1 - \varepsilon, 1 + \varepsilon]. \end{aligned}$$

This means that

$$j \in I(x_c) \text{ for all } x_c \in [1 - \varepsilon, 1 + \varepsilon].$$

The functional $f(x_d^j, \cdot) + u^j \cdot g(x_d^j, \cdot)$ with

$$f(x_d^j, x_c) + u^j \cdot g(x_d^j, x_c) = \psi(x_d^j) \cdot x_c^2 + u^j \cdot (1 - x_c) \text{ for all } x_c \in \mathbb{R}$$

is convex so that all assumptions of Theorem 8.23 are fulfilled. Hence, $\bar{x} = (x_d^j, 1)$ is a local minimal solution of the considered discrete-continuous optimization problem.

8.4 Duality

Under the assumption (8.1) we continue to investigate the discrete-continuous optimization problem (8.2)

$$\begin{aligned} &\min f(x_d, x_c) \\ &\text{subject to the constraints} \\ &g(x_d, x_c) \in -C \\ &h(x_d, x_c) = 0_Z \\ &(x_d, x_c) \in S_d \times S_c, \end{aligned} \tag{8.38}$$

which is now called *primal problem* (8.38). To this problem we associate a so-called *dual problem* (8.39) given by

$$\begin{aligned} \max \inf_{\substack{i \in N \\ x_c \in S_c}} & \left\{ f(x_d^i, x_c) + \ell_g^i\big(g(x_d^i, x_c)\big) + \ell_h^i\big(h(x_d^i, x_c)\big) \right\} \\ & \text{subject to the constraints} \\ & \ell_g^i \in C^* \text{ for all } i \in N \\ & \ell_h^i \in Z^* \text{ for all } i \in N. \end{aligned} \tag{8.39}$$

Notice that the variables of the dual problem (8.39) are elements of the dual spaces Y^* and Z^*, respectively.

A first relationship between the primal and dual problem is given by the following *weak duality theorem*, which is simple to prove.

Theorem 8.26 (weak duality theorem).

Let the assumption (8.1) be satisfied. For every $\hat{x} \in S$ (feasible element of (8.38)) and for all $\hat{\ell}_g^1, \hat{\ell}_g^2, \ldots \in C^*$ and $\hat{\ell}_h^1, \hat{\ell}_h^2, \ldots \in Z^*$ (feasible elements of (8.39)) the following inequality holds

$$\inf_{\substack{i \in N \\ x_c \in S_c}} \left\{ f(x_d^i, x_c) + \ell_g^i\big(g(x_d^i, x_c)\big) + \ell_h^i\big(h(x_d^i, x_c)\big) \right\} \leq f(\hat{x}).$$

Proof For an arbitrary feasible vector $\hat{x} = (x_d^j, \hat{x}_c)$ (for some $j \in N$ and some $\hat{x}_c \in S_c$) and arbitrary linear functionals $\hat{\ell}_g^1, \hat{\ell}_g^2, \ldots \in C^*$ and $\hat{\ell}_h^1, \hat{\ell}_h^2, \ldots \in Z^*$ we conclude

$$\begin{aligned} \inf_{\substack{i \in N \\ x_c \in S_c}} & \left\{ f(x_d^i, x_c) + \hat{\ell}_g^i\big(g(x_d^i, x_c)\big) + \hat{\ell}_h^i\big(h(x_d^i, x_c)\big) \right\} \\ & \leq f(\hat{x}) + \underbrace{\hat{\ell}_g^j\big(g(\hat{x})\big)}_{\leq 0} + \underbrace{\hat{\ell}_h^j\big(h(\hat{x})\big)}_{=0} \\ & \leq f(\hat{x}), \end{aligned}$$

which has to be shown. □

The next *strong duality theorem* answers the question under which assumptions the dual problem is solvable.

Theorem 8.27 (strong duality theorem).

Let the assumptions of Corollary 8.12 be satisfied and let $\bar{x} = (x_d^j, \bar{x}_c) \in S$ (for some $j \in N$ and some $\bar{x}_c \in S_c$) denote a minimal solution of the primal problem (8.38). Then the dual problem (8.39) is solvable and the extremal values of both problems are equal.

Proof By Corollary 8.12 we get for some continuous linear functionals $\bar{\ell}_g^i \in C^*$ (with $i \in N$) and $\bar{\ell}_h^i \in Z^*$ (with $i \in N$)

$$f(\bar{x}) \leq \inf_{i \in N} \left\{ f(x_d^i, x_c) + \bar{\ell}_g^i \left(g(x_d^i, x_c) \right) + \bar{\ell}_h^i \left(h(x_d^i, x_c) \right) \right\} \text{ for all } x_c \in S_c,$$

which can also be written as

$$f(\bar{x}) \leq \inf_{\substack{i \in N \\ x_c \in S_c}} \left\{ f(x_d^i, x_c) + \bar{\ell}_g^i \big(g(x_d^i, x_c) \big) + \bar{\ell}_h^i \big(h(x_d^i, x_c) \big) \right\}.$$

By the weak duality theorem (Theorem 8.26) we then obtain

$$f(\bar{x}) = \inf_{\substack{i \in N \\ x_c \in S_c}} \left\{ f(x_d^i, x_c) + \bar{\ell}_g^i \big(g(x_d^i, x_c) \big) + \bar{\ell}_h^i \big(h(x_d^i, x_c) \big) \right\}.$$

This means that the dual problem (8.39) is solvable and the extremal values of both problems are equal. □

8.4.1 Specialization to Extendedly Linear Problems

Since the dual problem (8.39) seems to be quite general, we now investigate linear discrete-continuous optimization problems in an extended form. In this case the dual problem (8.39) has a simpler structure.

Now we assume:

$$\left.\begin{array}{l}
\text{Let } (X_d, \|\cdot\|_{X_d}), (X_c, \|\cdot\|_{X_c}), (Y, \|\cdot\|_Y) \text{ and} \\
(Z, \|\cdot\|_Z) \text{ be real normed spaces;} \\
\text{let } C_Y \subset Y \text{ and } C_{X_c} \subset X_c \text{ be convex cones} \\
\text{in } Y \text{ and } X_c, \text{ respectively;} \\
\text{let } S_d \text{ be a discrete subset of } X_d, \\
\text{i.e. } S_d = \{x_d^i\}_{i\in N} \text{ for } N := \{1, 2, \ldots, n\} \\
\text{(with } n \in \mathbb{N}) \text{ or } N := \mathbb{N} \text{ with } x_d^1, x_d^2, \ldots \in X_d; \\
\text{let } c \in X_c^* \text{ be a continuous linear functional;} \\
\text{let } A : X_c \to Y \text{ be a linear mapping;} \\
\text{let } b \in Y \text{ be a given element;} \\
\text{let } f_d : S_d \to \mathbb{R} \text{ be a given functional;} \\
\text{let } g_d : S_d \to Y \text{ be a given mapping;} \\
\text{and let the constraint set} \\
S := \{(x_d, x_c) \in S_d \times C_{X_c} \mid g_d(x_d) + A(x_c) - b \in C_Y\} \\
\text{be nonempty.}
\end{array}\right\} \quad (8.40)$$

Notice that f_d and g_d do not need to be linear or convex. Under the assumption (8.40) we now consider the discrete-continuous optimization problem

$$\begin{array}{c}
\min \ f_d(x_d) + c(x_c) \\
\text{subject to the constraints} \\
g_d(x_d) + A(x_c) - b \in C_Y \\
x_d \in S_d, \quad x_c \in C_{X_c}.
\end{array} \quad (8.41)$$

This is an *extendedly linear primal problem*, which is more general than standard linear optimization problems. According to our general formulation of the dual problem (8.39) we associate the *extendedly linear dual problem*

$$\begin{array}{c}
\max \inf_{\substack{i\in N \\ x_c \in C_{X_c}}} \left\{ f_d(x_d^i) + c(x_c) + \ell^i\big(-g_d(x_d^i) - A(x_c) + b\big)\right\} \\
\text{subject to the constraints} \\
\ell^i \in C_Y^* \text{ for all } i \in N
\end{array} \quad (8.42)$$

to the original problem (8.41). Then the following lemma simplifies this problem in a way that it is more suitable in applications.

Lemma 8.28 (simplification of the extendedly linear dual problem).

Under the assumption (8.40) the extendedly linear dual problem (8.42) is equivalent to the problem

$$\begin{array}{c}\max \inf_{i\in N} \Big\{ f_d(x_d^i) - \big(\ell^i \circ g_d\big)(x_d^i) + \ell^i(b) \Big\} \\ \text{subject to the constraints} \\ c - \ell^i \circ A \in C_{X_c}^* \text{ for all } i \in N \\ \ell^i \in C_Y^* \text{ for all } i \in N. \end{array} \tag{8.43}$$

Proof

(a) Problem (8.42) is equivalent to the problem

$$\begin{array}{c}\max \lambda \\ \text{subject to the constraints} \\ \lambda \le \inf\limits_{\substack{i\in N \\ x_c \in C_{X_c}}} \Big\{ f_d(x_d^i) + c(x_c) + \ell^i\big(- g_d(x_d^i) - A(x_c) + b\big)\Big\} \\ \lambda \in \mathbb{R}, \quad \ell^i \in C_Y^* \text{ for all } i \in N. \end{array} \tag{8.44}$$

The inequality constraint in problem (8.44) can also be written as

$$\begin{array}{c}\lambda \le f_d(x_d^i) + c(x_c) + \ell^i\big(- g_d(x_d^i) - A(x_c) + b\big) \\ \text{for all } i \in N \text{ and all } x_c \in C_{X_c}\end{array}$$

or

$$c(x_c) - \big(\ell^i \circ A\big)(x_c) \ge \lambda - f_d(x_d^i) + \big(\ell^i \circ g_d\big)(x_d^i) - \ell^i(b) \quad \text{for all } i \in N \text{ and all } x_c \in C_{X_c}. \tag{8.45}$$

Since the right hand side of the inequality (8.45) does not depend on x_c, we conclude

$$c(x_c) - \big(\ell^i \circ A\big)(x_c) \ge 0 \text{ for all } x_c \in C_{X_c},$$

i.e.

$$c - \ell^i \circ A \in C_{X_c}^*,$$

and the constraint in problem (8.43) is shown. If we set $x_c = 0_{X_c}$ in the inequality (8.45), we obtain

$$\lambda \leq f_d(x_d^i) - \big(\ell^i \circ g_d\big)(x_d^i) + \ell^i(b) \text{ for all } i \in N$$

or

$$\lambda \leq \inf_{i \in N} \left\{ f_d(x_d^i) - \big(\ell^i \circ g_d\big)(x_d^i) + \ell^i(b) \right\}.$$

By maximization of λ we then get the formulation in (8.43).

(b) In a similar way we rewrite problem (8.43) in an equivalent form and we get the constraints

$$\begin{aligned} c(x_c) - \big(\ell^i \circ A\big)(x_c) &\geq 0 \\ &\geq \lambda - f_d(x_d^i) + \big(\ell^i \circ g_d\big)(x_d^i) - \ell^i(b) \\ &\quad \text{for all } i \in N \text{ and all } x_c \in C_{X_c}. \end{aligned}$$

This leads to problem (8.42). □

Remark 8.29 (simplification of the extendedly linear dual problem).

In the finite dimensional case, i.e. $X_d = \mathbb{R}^{n_d}$, $X_c = \mathbb{R}^{n_c}$, $Y = \mathbb{R}^m$, $A \in \mathbb{R}^{(m,n_c)}$, $c \in \mathbb{R}^{n_c}$ and $b \in \mathbb{R}^m$, the equivalent dual problem (8.43) can be written as

$$\begin{gathered} \max \inf_{i \in N} \left\{ f_d(x_d^i) + {\ell^i}^T \big(b - g_d(x_d^i)\big) \right\} \\ \text{subject to the constraints} \\ c - A^T \ell^i \in C^*_{\mathbb{R}^{n_c}} \text{ for all } i \in N \\ \ell^i \in C^*_{\mathbb{R}^m} \text{ for all } i \in N. \end{gathered}$$

In Example 6.2 we have already investigated a discrete problem with the standard duality theory based on only one Lagrange multiplier per constraint. With the theory of this chapter we investigate this problem again and we find out that the duality gap disappears.

Example 8.30 (extendedly linear dual problem).

As a very simple example we investigate the linear problem in Example 6.2 with only discrete variables:

$$\begin{gathered}\min\ -2(x_d)_1 + (x_d)_2\\ \text{subject to the constraints}\\ (x_d)_1 + (x_d)_2 - 3 \leq 0\\ x_d \in S_d := \left\{ \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}0\\4\end{pmatrix}, \begin{pmatrix}4\\4\end{pmatrix}, \begin{pmatrix}4\\0\end{pmatrix}, \begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix} \right\}.\end{gathered} \tag{8.46}$$

Here we have $X_d = \mathbb{R}^2$, $Y = \mathbb{R}$, $C_Y = \mathbb{R}_+$, $f_d(\cdot) = (-2, 1)\big(\cdot\big)$, $g_d(\cdot) = (-1, -1)\big(\cdot\big)$ and $b = -3$. By Lemma 8.28 the equivalent dual problem reads

$$\max_{\ell^1,\ldots,\ell^6 \geq 0}\ \min_{1 \leq i \leq 6} \left\{ (-2, 1)x_d^i + \ell^i(1, 1)x_d^i - 3\ell^i \right\}. \tag{8.47}$$

The objective function of this problem can be written as

$$\begin{aligned}&\min_{1 \leq i \leq 6} \left\{ (-2, 1)x_d^i + \ell^i(1, 1)x_d^i - 3\ell^i \right\}\\ &= \min \left\{ -3\ell^1,\ 4 + \ell^2,\ -4 + 5\ell^3,\ -8 + \ell^4,\ 0,\ -3 \right\},\end{aligned}$$

and problem (8.47) is equivalent to

$$\begin{gathered}\max\ \lambda\\ \text{subject to the constraints}\\ \lambda \leq -3\ell^1\\ \lambda \leq 4 + \ell^2\\ \lambda \leq -4 + 5\ell^3\\ \lambda \leq -8 + \ell^4\\ \lambda \leq 0\\ \lambda \leq -3\\ \lambda \in \mathbb{R},\quad \ell^1, \ldots, \ell^4 \geq 0.\end{gathered}$$

Notice that the second and fifth inequality constraint are redundant so that the variable ℓ^2 can also be dropped. Then we get the equivalent dual problem

$$
\begin{aligned}
&\max \lambda \\
&\text{subject to the constraints} \\
&\lambda \leq -3\ell^1 \\
&\lambda \leq -4 + 5\ell^3 \\
&\lambda \leq -8 + \ell^4 \\
&\lambda \leq -3 \\
&\lambda \in \mathbb{R}, \quad \ell^1, \ell^3, \ell^4 \geq 0.
\end{aligned}
$$

It is evident that the maximal value of this problem equals -3. Notice that the primal problem (8.46) has only the three feasible points $(0, 0)$, $(1, 2)$ and $(2, 1)$ with the minimal value -3. Hence, the extremal values of the primal and dual problem coincide. In contrast to the classical Lagrange theory in continuous optimization (compare Example 6.2) there is no duality gap because we work with more than one Lagrange multiplier.

8.5 Application to Discrete-Continuous Semidefinite and Copositive Optimization

It is known that semidefinite and copositive optimization have important applications in practice (compare Chap. 7). Now we extend these problems in such a way that the variables may be discrete and continuous. Since these problems are optimization problems in a finite dimensional Hilbert space, the discrete-continuous theory can be used for this problem class.

As in Chap. 7 let $\mathcal{S}^k$ (for some $k \in \mathbb{N}$) denote the real linear space of symmetric (k, k)-matrices. In the following let $C \subset \mathcal{S}^k$ denote either the Löwner ordering cone, i.e.

$$C := \left\{ M \in \mathcal{S}^k \mid M \text{ is positive semidefinite} \right\} =: \mathcal{S}^k_+, \tag{8.48}$$

or the copositive ordering cone, i.e.

$$C := \left\{ M \in \mathcal{S}^k \mid y^T M y \geq 0 \text{ for all } y \in \mathbb{R}^k_+ \right\} \tag{8.49}$$

(compare Remark 7.1). Furthermore, we assume:

$$\left.\begin{array}{l}\text{Let } m_d, m_c \in \mathbb{N} \text{ be given integers;}\\ \text{let } S_d := \left\{x_d^1, x_d^2, \ldots, x_d^n\right\} \subset \mathbb{R}^{m_d} \text{ be a discrete set with } n \in \mathbb{N};\\ \text{let } f : S_d \times \mathbb{R}^{m_c} \to \mathbb{R} \text{ be a given function;}\\ \text{let } G : S_d \times \mathbb{R}^{m_c} \to \mathcal{S}^k \text{ be a given mapping;}\\ \text{and let the constraint set}\\ S := \{(x_d, x_c) \in S_d \times \mathbb{R}^{m_c} \mid G(x_d, x_c) \in -C\}\\ \text{be nonempty.}\end{array}\right\} \tag{8.50}$$

As mentioned on page 189 the space $\mathcal{S}^k$ is a finite dimensional Hilbert space with the scalar product $\langle\cdot,\cdot\rangle$ defined by

$$\langle A, B\rangle = \text{trace}\,(A \cdot B) \text{ for all } A, B \in \mathcal{S}^k.$$

Under the assumption (8.50) we investigate the semidefinite/copositive optimization problem

$$\begin{array}{c}\min f(x_d, x_c)\\ \text{subject to the constraints}\\ G(x_d, x_c) \in -C\\ (x_d, x_c) \in S_d \times \mathbb{R}^{m_c}.\end{array} \tag{8.51}$$

If the cone C is given by (8.48), then problem (8.51) is a discrete-continuous semidefinite optimization problem. If C is given by (8.49), then problem (8.51) is a discrete-continuous copositive optimization problem.

We restrict our investigation to the formulation of extended KKT conditions for problem (8.51).

Theorem 8.31 (extended KKT conditions).

Let the assumption (8.50) be satisfied and let $\bar{x} = (x_d^j, \bar{x}_c)$ (for some $j \in \{1, \ldots, n\}$) be a minimal solution of the discrete-continuous semidefinite/copositive optimization problem (8.51). For every $i \in \{1, \ldots, n\}$ let the set

$$B_i := \left\{ \begin{pmatrix} f(x_d^i, x_c) - f(\bar{x}) + \alpha \\ G(x_d^i, x_c) + Y \end{pmatrix} \in \mathbb{R} \times \mathcal{S}^k \;\middle|\; x_c \in \mathbb{R}^{m_c}, \right.$$
$$\left. \alpha > 0,\ Y \in int(C) \right\}$$

be convex. For every $i \in \{1, \ldots, n\}$ let the function $f(x_d^i, \cdot)$ be differentiable at $\bar{x}_c$ and let the mapping $G(x_d^i, \cdot)$ be Fréchet differentiable at $\bar{x}_c$. Then there exist real numbers $\mu^1, \ldots, \mu^n \geq 0$ and matrices $L^1, \ldots, L^n \in C^*$ with $(\mu^i, L^i) \neq (0, 0_{\mathcal{S}^k})$ for all $i \in \{1, \ldots, n\}$, and the equality

$$\mu^i \nabla_{x_c} f(x_d^i, \bar{x}_c) + \begin{pmatrix} \langle L^i, G_{(x_c)_1}(\bar{x}) \rangle \\ \vdots \\ \langle L^i, G_{(x_c)_{m_c}}(\bar{x}) \rangle \end{pmatrix} = 0_{\mathbb{R}^{m_c}} \text{ for all } i \in I(\bar{x}_c)$$

with

$$G_{(x_c)_i} := \begin{pmatrix} \frac{\partial}{\partial (x_c)_i} G_{11} & \cdots & \frac{\partial}{\partial (x_c)_i} G_{1k} \\ \vdots & & \vdots \\ \frac{\partial}{\partial (x_c)_i} G_{k1} & \cdots & \frac{\partial}{\partial (x_c)_i} G_{kk} \end{pmatrix} \text{ for all } i \in \{1, \ldots, m_c\}.$$

and

$$\begin{aligned} I(\bar{x}_c) := \Big\{ i' \in \{1, \ldots, n\} \;\Big|\; & f(x_d^{i'}, \bar{x}_c) + \langle L^{i'}, G(x_d^{i'}, \bar{x}_c) \rangle \\ & = \min_{1 \leq i \leq n} \left\{ f(x_d^i, \bar{x}_c) + \langle L^i, G(x_d^i, \bar{x}_c) \rangle \right\} \Big\} \end{aligned}$$

is fulfilled together with the equality

$$\langle L^j, G(\bar{x}) \rangle = 0.$$

If in addition to the above assumptions the condition

$$\forall\, i \in \{1, \ldots, n\} \;\; \exists\, x_c \in \mathbb{R}^{m_c} : \;\; 0 > \langle L^i, G(x_d^i, x_c) \rangle$$

is satisfied, then it follows

$$\mu^i > 0 \text{ for all } i \in \{1, \ldots, n\}.$$

Proof To the minimal solution $\bar{x} = (x_d^j, \bar{x}_c)$ we apply Theorem 8.10 and follow the proof of Theorem 8.19 and Corollary 8.20. Using the known formula for the Fréchet derivative of $G(x_d^i, \cdot)$ for every $i \in \{1, \ldots, n\}$ (see Lemma 7.7) we then get the first part of the assertion. The second part is a consequence of the CQ given in Corollary 8.12. □

Theorem 8.31 holds for the semidefinite variant of problem (8.51) and for the copositive case. The only difference is the dual cone C^*. Since the Löwner ordering cone is self-dual, the condition $L^1, \ldots, L^n \in C^*$ means that the matrices $L^1, \ldots, L^n$ are positive semidefinite. If C equals the copositive ordering cone, then its dual cone is given by

$$C^* = \text{convex hull}\left\{yy^T \mid y \in \mathbb{R}^k_+\right\}$$

and elements of C^* are called *completely positive matrices* (compare Lemma 7.5 and the subsequent remark).

Finally, we turn our attention to the dual problem of the semidefinite/copositive problem (8.51). This dual problem is given by

$$\begin{aligned} &\max \inf_{\substack{1 \le i \le n \\ x_c \in \mathbb{R}^{m_c}}} \left\{ f(x_d^i, x_c) + \langle L^i, G(x_d^i, x_c)\rangle \right\} \\ &\text{subject to the constraints} \\ &L^1, \ldots, L^n \in C^*. \end{aligned} \tag{8.52}$$

So, the n variables of the dual semidefinite problem are positive semidefinite symmetric (k, k)-matrices and in the copositive case the dual variables are completely positive matrices. The weak duality theorem (Theorem 8.26) and the strong duality theorem (Theorem 8.27) are directly applicable to the primal problem (8.51) and its dual formulation (8.52).

Exercises

(8.1) Are the two discrete-continuous optimization problems

$$\begin{gathered} \min \frac{1}{x_1^4} + \frac{1}{x_2^2} \\ \text{subject to} \\ x_1 \in \mathbb{N},\ x_2 \in \mathbb{R} \end{gathered}$$

and

$$\begin{gathered} \min \frac{x_2^2}{x_1^4} \\ \text{subject to} \\ x_1 \in \mathbb{N},\ x_2 \in \mathbb{R} \end{gathered}$$

solvable?

(8.2) Let $(X_d, \|\cdot\|_{X_d})$ be a real normed space, and let $(X_c, \|\cdot\|_{X_c})$ be a reflexive real Banach space. Let the set $S_d \subset X_d$ consist of finitely many elements, and let $S_c \subset X_c$ be a nonempty, convex, closed and bounded set. Moreover, let the functional $f : S_d \times S_c \to \mathbb{R}$ have the property that for every $x_d \in S_d$

the functional $f(x_d, \cdot)$ is continuous and quasiconvex. Prove that the discrete-continuous optimization problem

$$\min_{(x_d,x_c)\in S_d\times S_c} f(x_d, x_c)$$

is solvable.

(8.3) Let S be a nonempty subset of a real normed space $(X, \|\cdot\|)$ and for some $n \in \mathbb{N}$ let $\varphi_1, \ldots, \varphi_n : S \to \mathbb{R}$ be given functionals being continuous and directionally differentiable at an arbitrary $\bar{x} \in S$. Show that the directional derivative of the functional $f : S \to \mathbb{R}$ with

$$f(x) = \min_{1\le i\le n} \{\varphi_i(x)\} \text{ for all } x \in S$$

at $\bar{x} \in S$ is given by

$$\left(\min_{1\le i\le n} \{\varphi_i(\cdot)\}\right)'(\bar{x})(h) = \min_{i\in I(\bar{x})} \left\{\varphi_i'(\bar{x})(h)\right\} \text{ for all } h \in X$$

where

$$I(\bar{x}) := \left\{ i \in \{1, \ldots, n\} \;\middle|\; \varphi_i(\bar{x}) = \min_{1\le k\le n} \{\varphi_k(\bar{x})\} \right\}.$$

(8.4) Consider the discrete-continuous optimization problem

$$\begin{gathered} \min \ \big(1 + (x_d)_1\big)^2 \big(2 + (x_d)_2\big)^4 \big(1 + (x_c)_1 + (x_c)_2^3\big) \\ \text{subject to the constraints} \\ (x_d)_1, (x_d)_2 \in \{1, 2, \ldots, 20\} \\ (x_c)_1 \ge 0,\ (x_c)_2 \ge 0. \end{gathered}$$

Determine the unique minimal solution of this problem and the Lagrange multipliers associated to this minimal solution.

(8.5) Consider the simple primal discrete-continuous optimization problem

$$\begin{gathered} \min \ (x_d)_1^2 + \ln (x_d)_2 + x_c \\ \text{subject to the constraints} \\ e^{(x_d)_1^2 + (x_d)_2^2} + 3x_c \ge 1 \\ x_d \in S_d := \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \end{pmatrix} \right\}, \quad x_c \ge 0 \end{gathered}$$

and formulate its dual problem. Determine a maximal solution and the maximal value of the dual problem.

9 Direct Treatment of Special Optimization Problems

Many of the results derived in this book are concerned with a generally formulated optimization problem. But if a concrete problem is given which has a rich mathematical structure, then solutions or characterizations of solutions can be derived sometimes in a direct way. In this case one takes advantage of the special structure of the optimization problem and can achieve the desired results very quickly.

In this final chapter we present two special optimal control problems and show how they can be treated without the use of general theoretical optimization results. The first problem is a so-called linear quadratic optimal control problem. For the given quadratic objective functional one gets a minimal solution with the aid of a simple quadratic completion without using necessary optimality conditions. The second problem is a time-minimal optimal control problem which can be solved directly by the application of a separation theorem.

9.1 Linear Quadratic Optimal Control Problems

In this section we consider a system of autonomous linear differential equations

$$\dot{x}(t) = Ax(t) + Bu(t) \text{ almost everywhere on } [0, \hat{T}] \tag{9.1}$$

and an initial condition

$$x(0) = x^0 \tag{9.2}$$

(where $\hat{T} > 0$ and $x^0 \in \mathbb{R}^n$ are arbitrarily given). Let A and B be (n, n) and (n, m) matrices with real coefficients, respectively. Let every control $u \in L_\infty^m([0, \hat{T}])$ be feasible (i.e. the controls are unconstrained). It is our aim to steer the system (9.1), (9.2) as close to a state of rest as possible at the terminal time $\hat{T}$. In other

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3_9

words: For a given positive definite symmetric (n, n) matrix G with real coefficients the quadratic form $x(\hat{T})^T G x(\hat{T})$ should be minimal. Since we want to reach our goal with a minimal steering effort, for a given positive definite symmetric (m, m) matrix R with real coefficients the expression $\int_0^{\hat{T}} u(t)^T R u(t)\, dt$ should be minimized as well. These two goals are used for the definition of the objective functional $J : L_\infty^m([0, \hat{T}]) \to \mathbb{R}$ with

$$J(u) = x(\hat{T})^T G x(\hat{T}) + \int_0^{\hat{T}} u(t)^T R u(t)\, dt \text{ for all } u \in L_\infty^m([0, \hat{T}]).$$

Under these assumptions the considered linear quadratic optimal control problem then reads as follows:

> Minimize the objective functional J with respect to all controls $u \in L_\infty^m([0, \hat{T}])$ for which the resulting trajectory is given by the system (9.1) of differential equations and the initial condition (9.2). (9.3)

In order to be able to present an optimal control for the problem (9.3) we need two technical lemmas.

Lemma 9.1 (relationship between control and trajectory).

Let $P(\cdot)$ be a real (n, n) matrix function which is symmetric and differentiable on $[0, \hat{T}]$. Then it follows for an arbitrary control $u \in L_\infty^m([0, \hat{T}])$ and a trajectory x of the initial value problem (9.1), (9.2):

$$\begin{aligned} 0 = {x^0}^T P(0) x^0 - x(\hat{T})^T P(\hat{T}) x(\hat{T}) + \int_0^{\hat{T}} \Big[& 2u(t)^T B^T P(t) x(t) \\ & + x(t)^T \left(\dot{P}(t) + A^T P(t) + P(t) A \right) x(t) \Big] dt. \end{aligned}$$

Proof For an arbitrary control $u \in L_\infty^m([0, \hat{T}])$ and a corresponding trajectory x of the initial value problem (9.1), (9.2) and an arbitrary real matrix function $P(\cdot)$ defined on $[0, \hat{T}]$ and being symmetric and differentiable it follows:

$$\begin{aligned} \frac{d}{dt}\left[x(t)^T P(t) x(t) \right] &= \dot{x}(t)^T P(t) x(t) + x(t)^T \left(\dot{P}(t) x(t) + P(t) \dot{x}(t) \right) \\ &= \left(A x(t) + B u(t) \right)^T P(t) x(t) \end{aligned}$$

$$+x(t)^T \left(\dot{P}(t)x(t) + P(t)\left(Ax(t) + Bu(t)\right)\right)$$
$$= x(t)^T \left(\dot{P}(t) + A^T P(t) + P(t)A\right) x(t)$$
$$+2u(t)^T B^T P(t)x(t) \text{ almost everywhere on } [0, \hat{T}].$$

With the initial condition (9.2) we get immediately by integration

$$x(\hat{T})^T P(\hat{T})x(\hat{T}) - x^{0^T} P(0)x^0$$
$$= \int_0^{\hat{T}} \left[2u(t)^T B^T P(t)x(t) + x(t)^T \left(\dot{P}(t) + A^T P(t) + P(t)A\right) x(t)\right] dt$$

which implies the assertion. □

Lemma 9.2 (Bernoulli matrix differential equation).

The (n, n) matrix function $P(\cdot)$ with

$$P(t) = \left[e^{A(t-\hat{T})} G^{-1} e^{A^T(t-\hat{T})} + \int_t^{\hat{T}} e^{A(t-s)} B R^{-1} B^T e^{A^T(t-s)}\, ds\right]^{-1}$$
$$\text{for all } t \in [0, \hat{T}] \qquad (9.4)$$

is a solution of the Bernoulli matrix differential equation

$$\dot{P}(t) + A^T P(t) + P(t)A - P(t)BR^{-1}B^T P(t) = 0_{(n,n)} \text{ for all } t \in [0, \hat{T}] \qquad (9.5)$$

with the terminal condition

$$P(\hat{T}) = G. \qquad (9.6)$$

The matrix function $P(\cdot)$ defined by (9.4) is symmetric.

Proof First we define the (n, n) matrix function $Q(\cdot)$ by

$$Q(t) = e^{A(t-\hat{T})} G^{-1} e^{A^T(t-\hat{T})} + \int_t^{\hat{T}} e^{A(t-s)} B R^{-1} B^T e^{A^T(t-s)}\, ds \text{ for all } t \in [0, \hat{T}]$$

(notice that the matrix exponential function is defined as a matrix series). It is evident that $Q(\cdot)$ is a symmetric matrix function. For an arbitrary $z \in \mathbb{R}^n$, $z \neq 0_{\mathbb{R}^n}$, we obtain

$$z^T Q(t)z = \underbrace{z^T e^{A(t-\hat{T})} G^{-1} e^{A^T(t-\hat{T})} z}_{>0} + \int_t^{\hat{T}} \underbrace{z^T e^{A(t-s)} B R^{-1} B^T e^{A^T(t-s)} z}_{\geq 0} \, ds$$

$$> 0 \text{ for all } t \in [0, \hat{T}].$$

Consequently, for every $t \in [0, \hat{T}]$ the matrix $Q(t)$ is positive definite and therefore invertible, i.e. the matrix function $P(\cdot)$ with

$$P(t) = Q(t)^{-1} \text{ for all } t \in [0, \hat{T}]$$

is well-defined. Since $Q(\cdot)$ is symmetric, $P(\cdot)$ is also symmetric.

It is obvious that $P(\cdot)$ satisfies the terminal condition (9.6). Hence, it remains to be shown that $P(\cdot)$ is a solution of the Bernoulli matrix differential equation (9.5). For this proof we calculate the derivative (notice the implications for arbitrary $t \in [0, \hat{T}]$: $Q(t) \cdot Q(t)^{-1} = I \implies \dot{Q}(t)Q(t)^{-1} + Q(t)\frac{d}{dt}\left(Q(t)^{-1}\right) = 0_{(n,n)} \implies \frac{d}{dt}\left(Q(t)^{-1}\right) = -Q(t)^{-1}\dot{Q}(t)\, Q(t)^{-1}$)

$$\begin{aligned}
\dot{P}(t) &= \frac{d}{dt}\left(Q(t)^{-1}\right) \\
&= -Q(t)^{-1}\dot{Q}(t)Q(t)^{-1} \\
&= -Q(t)^{-1}\Bigg[A e^{A(t-\hat{T})} G^{-1} e^{A^T(t-\hat{T})} + e^{A(t-\hat{T})} G^{-1} e^{A^T(t-\hat{T})} A^T \\
&\quad + \int_t^{\hat{T}} \Big(A e^{A(t-s)} B R^{-1} B^T e^{A^T(t-s)} \\
&\quad + e^{A(t-s)} B R^{-1} B^T e^{A^T(t-s)} A^T\Big) ds - B R^{-1} B^T\Bigg] Q(t)^{-1} \\
&= -Q(t)^{-1}\left[A Q(t) + Q(t) A^T - B R^{-1} B^T\right] Q(t)^{-1} \\
&= -Q(t)^{-1} A - A^T Q(t)^{-1} + Q(t)^{-1} B R^{-1} B^T Q(t)^{-1} \\
&= -P(t)A - A^T P(t) + P(t) B R^{-1} B^T P(t) \text{ for all } t \in [0, \hat{T}].
\end{aligned}$$

Consequently, $P(\cdot)$ satisfies the Bernoulli matrix differential equation (9.5). □

With the aid of the two preceding lemmas it is now possible to present the optimal control of the linear quadratic problem (9.3).

Theorem 9.3 (feedback control).

The so-called feedback control $\bar{u}$ given by

$$\bar{u}(t) = -R^{-1}B^T P(t)x(t) \text{ almost everywhere on } [0, \hat{T}]$$

is the only optimal control of the linear quadratic control problem (9.3) where the matrix function $P(\cdot)$ is given by (9.4).

Proof In the following let $P(\cdot)$ be the matrix function defined by (9.4). Then we have with Lemmas 9.1 and 9.2 for every control $u \in L_\infty^m([0, \hat{T}])$ with $u \neq \bar{u}$:

$$\begin{aligned}
J(u) &= x(\hat{T})^T G x(\hat{T}) + \int_0^{\hat{T}} u(t)^T R u(t)\, dt \\
&= {x^0}^T P(0) x^0 + x(\hat{T})^T [G - P(\hat{T})] x(\hat{T}) \\
&\quad + \int_0^{\hat{T}} \Big[u(t)^T R u(t) + 2u(t)^T B^T P(t) x(t) \\
&\qquad + x(t)^T \Big(\dot{P}(t) + A^T P(t) + P(t) A \Big) x(t) \Big] dt \\
&\qquad\qquad \text{(from Lemma 9.1)} \\
&= {x^0}^T P(0) x^0 + \int_0^{\hat{T}} \Big[u(t)^T R u(t) + 2u(t)^T B^T P(t) x(t) \\
&\qquad + x(t)^T P(t) B R^{-1} B^T P(t) x(t) \Big] dt \\
&\qquad\qquad \text{(from Lemma 9.2)} \\
&= {x^0}^T P(0) x^0 + \int_0^{\hat{T}} \Big(u(t) + R^{-1} B^T P(t) x(t) \Big)^T R \\
&\qquad \Big(u(t) + R^{-1} B^T P(t) x(t) \Big) dt \\
&> {x^0}^T P(0) x^0 \\
&= J(\bar{u}).
\end{aligned}$$

Hence $\bar{u}$ is the only minimal point of the functional J. □

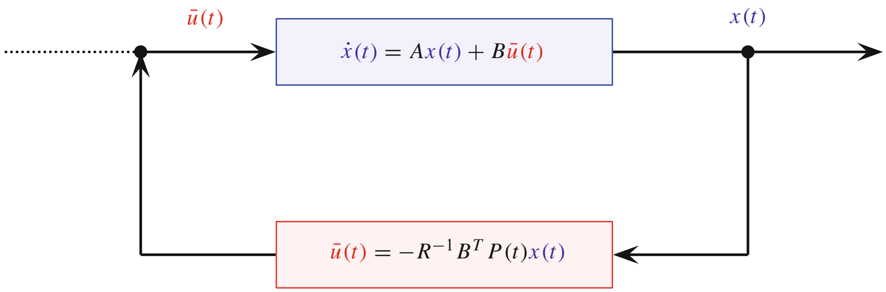

Fig. 9.1 Feedback control of Theorem 9.3

The optimal control presented in Theorem 9.3 depends on the time variable t and the current state $x(t)$. Such a control is called a *feedback* or a *closed loop control* (see Fig. 9.1).

If the control function depends only on t and not on the state $x(t)$, then it is called an *open loop control*. Feedback controls are of special importance for applications. Although feedback controls are also derived from the mathematical model, they make use of the real state of the system which is described mathematically only in an approximate way. Hence, in the case of perturbations which are not included in the mathematical model, feedback controls are often more realistic for the regulation of the system.

Since the matrix function P is analytic and the trajectory x is absolutely continuous, the optimal control $\bar{u}$ in Theorem 9.3 is an absolutely continuous vector function. In fact, a solution of the linear quadratic optimal control problem lies in a smaller subspace of $L_\infty^m([0, \hat{T}])$.

Notice that the proof of Theorem 9.3 could be done with the aid of an optimality condition. Instead of this we use a quadratic completion with Lemmas 9.1 and 9.2 which is simpler from a mathematical point of view.

The linear quadratic control problem (9.3) can be formulated more generally. If one defines the objective functional J by

$$J(u) = x(\hat{T})^T Gx(\hat{T}) + \int_0^{\hat{T}} \Big(x(t)^T Qx(t) + u(t)^T Ru(t)\Big)\, dt \quad \text{for all } u \in L_\infty^m([0, \hat{T}])$$

where Q is a positive definite symmetric (n, n) matrix with real coefficients, then the result of Theorem 9.3 remains almost true for the modified control problem.

The only difference is that then the matrix function $P(\cdot)$ is a solution of the Riccati matrix differential equation

$$\dot{P}(t) + A^T P(t) + P(t)A + Q - P(t)BR^{-1}B^T P(t) = 0_{(n,n)} \text{ for all } t \in [0, \hat{T}]$$

with the terminal condition $P(\hat{T}) = G$.

Example 9.4 (feedback control).

As a simple model we consider the differential equation

$$\dot{x}(t) = 3x(t) + u(t) \text{ almost everywhere on } [0, 1]$$

with the initial condition

$$x(0) = x^0$$

where $x^0 \in \mathbb{R}$ is arbitrarily chosen. The objective functional J reads as follows:

$$J(u) = x(1)^2 + \frac{1}{5}\int_0^1 u(t)^2\,dt \text{ for all } u \in L_\infty([0, 1]).$$

Then we obtain the function P as

$$\begin{aligned} P(t) &= \left[e^{3(t-1)}e^{3(t-1)} + 5\int_t^1 e^{3(t-s)}e^{3(t-s)}\,ds \right]^{-1} \\ &= \left[e^{6(t-1)} + 5\int_t^1 e^{6(t-s)}\,ds \right]^{-1} \\ &= \left[e^{6(t-1)} - \frac{5}{6}e^{6(t-1)} + \frac{5}{6} \right]^{-1} \\ &= \frac{6}{5 + e^{6(t-1)}} \text{ for all } t \in [0, 1]. \end{aligned}$$

Hence, the optimal control $\bar{u}$ is given by

$$\begin{aligned}\bar{u}(t) &= -5\,\frac{6}{5+e^{6(t-1)}}\,x(t)\\ &= -\frac{30}{5+e^{6(t-1)}}\,x(t) \text{ almost everywhere on } [0,1]. \qquad (9.7)\end{aligned}$$

If we plug the feedback control $\bar{u}$ in the differential equation, we can determine the trajectory x:

$$\begin{aligned}\dot{x}(t) &= 3x(t) + \bar{u}(t)\\ &= 3x(t) - \frac{30}{5+e^{6(t-1)}}x(t)\\ &= \left(3 - \frac{30}{5+e^{6(t-1)}}\right)x(t).\end{aligned}$$

Then we obtain the trajectory x as

$$\begin{aligned}x(t) &= x^0\, e^{\int_0^t \left(3-\frac{30}{5+e^{6(s-1)}}\right)ds}\\ &= x^0\, e^{\left(3s-6(s-1)+\ln(e^{6(s-1)}+5)\right)\Big|_0^t}\\ &= x^0\, e^{-3t+\ln(e^{6(t-1)}+5)-\ln(e^{-6}+5)}\\ &= \frac{x^0}{e^{-6}+5}\, e^{-3t}\, \left(e^{6(t-1)}+5\right) \text{ for all } t\in[0,1]. \qquad (9.8)\end{aligned}$$

If we plug the equation (9.8) in the equation (9.7), we get the optimal control $\bar{u}$ in the open loop form

$$\bar{u}(t) = -\frac{30x^0}{e^{-6}+5}\, e^{-3t} \text{ almost everywhere on } [0,1]$$

(compare Fig. 9.2). This optimal control is even a smooth function.

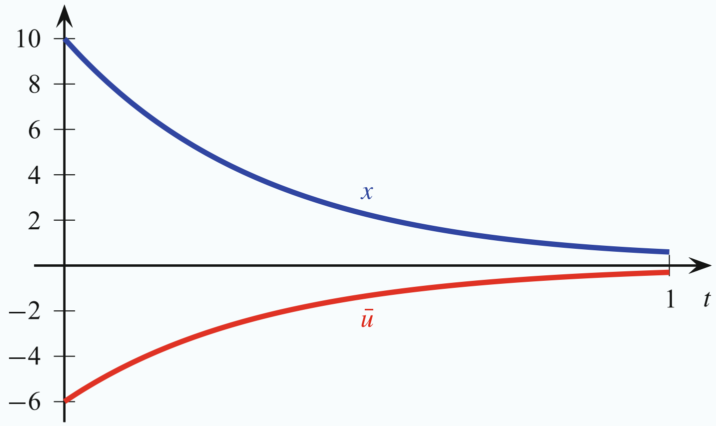

Fig. 9.2 Illustration of the optimal control $\bar{u}$ and the trajectory x

9.2 Time Minimal Control Problems

An important problem in control theory is the problem of steering a linear system with the aid of a bounded control from its initial state to a desired terminal point in minimal time. In this section we answer the questions concerning the existence and the characterization of such a time minimal control. As a necessary condition for such an optimal control we derive a so-called weak bang-bang principle. Moreover, we investigate a condition under which a time minimal control is unique.

In this section we consider the system of linear differential equations

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \text{ almost everywhere on } [0, \hat{T}] \tag{9.9}$$

with the initial condition

$$x(0) = x^0 \tag{9.10}$$

and the terminal condition

$$x(\hat{T}) = x^1 \tag{9.11}$$

where $\hat{T} > 0$, $x^0, x^1 \in \mathbb{R}^n$, A and B are (n, n) and (n, m) matrix functions with real coefficients, respectively, which are assumed to be continuous on $[0, \hat{T}]$, and controls u are chosen from $L_\infty^m([0, \hat{T}])$ with $\|u_i\|_{L_\infty([0,\hat{T}])} \leq 1$ for all $i \in \{1, \ldots, m\}$. Then we ask for a minimal time $\bar{T} \in [0, \hat{T}]$ so that the linear system (9.9) can be steered from x^0 to x^1 on the time interval $[0, \bar{T}]$.

If we consider the linear system (9.9) on a time interval $[0, T]$ with $T \in [0, \hat{T}]$ we use the abbreviation

$$U(T) := \{u \in L_\infty^m([0,T]) \mid \text{for every } k \in \{1,\ldots,m\} \text{ we have } |u_k(t)| \leq 1 \text{ almost everywhere on } [0,T]\} \quad \text{for all } T \in [0,\hat{T}] \tag{9.12}$$

for the set of all feasible controls with terminal time T.

Definition 9.5 (set of attainability).

For any $T \in [0, \hat{T}]$ consider the linear system (9.9) on $[0, T]$ with the initial condition (9.10). The set

$$K(T) := \{x(T) \in \mathbb{R}^n \mid u \in U(T) \text{ and } x \text{ satisfies the linear system (9.9) on } [0,T] \text{ and the initial condition (9.10)}\}$$

(with $U(T)$ given in (9.12)) is called the *set of attainability*.

The set of attainability consists of all terminal points to which the system can be steered from x^0 at the time T. Since we assume by (9.11) that the system can be steered to x^1 we have $x^1 \in K(\hat{T})$. Hence, the problem of finding a time minimal control for the linear system (9.9) satisfying the conditions (9.10), (9.11) can be transformed to a problem of the following type: Determine a minimal time $\bar{T} \in [0, \hat{T}]$ for which $x^1 \in K(\bar{T})$ (see Fig. 9.3).

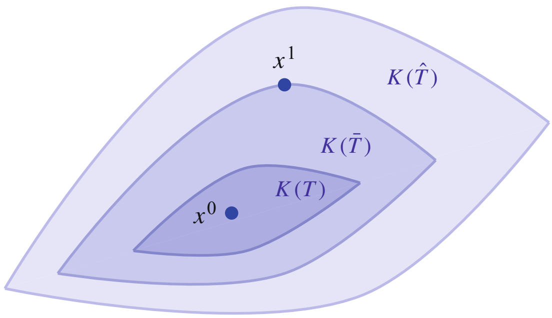

Fig. 9.3 Illustration of the set of attainability with $T \in (0, \bar{T})$

Before going further we recall that for an arbitrary $u \in L_\infty^m([0, T])$ the solution of the initial value problem (9.9), (9.10) with respect to the time interval $[0, T]$, $T \in [0, \hat{T}]$, can be written as

$$x(t) = \Phi(t)x^0 + \Phi(t)\int_0^t \Phi(s)^{-1}B(s)u(s)\,ds \text{ for all } t \in [0, \bar{T}]$$

where Φ is the fundamental matrix with

$$\dot{\Phi}(t) = A(t)\Phi(t) \text{ for all } t \in [0, T],$$

$$\Phi(0) = I \text{ (identity matrix)}^{14}.$$

Notice that in the case of a time independent matrix A, the fundamental matrix Φ is given as

$$\Phi(t) = e^{At} = \sum_{i=0}^{\infty} A^i \frac{t^i}{i!} \text{ for all } t \in [0, T].$$

In the following, for reasons of simplicity, we use the abbreviations

$$Y(t) := \Phi^{-1}(t)B(t) \text{ for all } t \in [0, T]$$

and

$$R(T) := \left\{ \int_0^T Y(t)u(t)dt \;\middle|\; u \in U(T) \right\} \text{ for all } T \in [0, \hat{T}].$$

The set $R(T)$ is sometimes called the *reachable set*. A connection between K and R is given by

$$\begin{aligned} K(T) &= \Phi(T)\left(x^0 + R(T)\right) \\ &= \{\Phi(T)x^0 + \Phi(T)y \mid y \in R(T)\} \text{ for all } T \in [0, \hat{T}]. \end{aligned} \tag{9.13}$$

First we investigate properties of the set of attainability.

[14] A proof of this existence result can be found e.g. in [212, p. 121–122].

Lemma 9.6 (properties of the set of attainability).

For every $T \in [0, \hat{T}]$ *the set* $K(T)$ *of attainability for the initial value problem (9.9), (9.10) with respect to the time interval* $[0, T]$ *is nonempty, convex and compact.*

Proof We present a proof of this lemma only in a short form. Let some $T \in [0, \hat{T}]$ be arbitrarily given. Because of the initial condition (9.10) it is obvious that $R(T) \neq \emptyset$. Next we show that the reachable set

$$R(T) = \left\{ \int_0^T Y(t)u(t)\,dt \;\middle|\; u \in U(T) \right\}$$

is convex and compact. $U(T)$ is the closed unit ball in $L_\infty^m([0, T])$ and therefore weak*-compact. Next we define the linear mapping $L : L_\infty^m([0, T]) \to \mathbb{R}^n$ with

$$L(u) = \int_0^T Y(t)u(t)\,dt \text{ for all } u \in L_\infty^m([0, T]).$$

L is continuous with respect to the norm topology in $L_\infty^m([0, T])$, and therefore it is also continuous with respect to the weak*-topology in $L_\infty^m([0, T])$. Since L is continuous and linear and the set $U(T)$ is weak*-compact and convex, the image $R(T) = L(U(T))$ is compact and convex. Because of the equation (9.13) the set $K(T)$ is also compact and convex. □

As a first important result we present an existence theorem for time minimal controls.

Theorem 9.7 (existence of a time minimal control).

If there is a control which steers the linear system (9.9) with the initial condition (9.10) to a terminal state x^1 *within a time* $\tilde{T} \in [0, \hat{T}]$*, then there is also a time minimal control with this property.*

Proof We assume that $x^1 \in K(\tilde{T})$. Next we set

$$\bar{T} := \inf\{T \in [0, \hat{T}] \mid x^1 \in K(T)\}.$$

Then we have $\bar{T} \leq \tilde{T}$, and there is a monotonically decreasing sequence $(T_i)_{i\in\mathbb{N}}$ with the limit $\bar{T}$ and a sequence $(u^i)_{i\in\mathbb{N}}$ of feasible controls with

$$x^1 =: x(T_i, u^i) \in K(T_i)$$

(let $x(T_i, u^i)$ denote the terminal state at the time T_i with the control u^i). Then it follows

$$\begin{aligned}
&\|x^1 - x(\bar{T}, u^i)\| \\
&= \|x(T_i, u^i) - x(\bar{T}, u^i)\| \\
&= \left\| \Phi(T_i)x^0 + \Phi(T_i)\int_0^{T_i} Y(t)u^i(t)\,dt - \Phi(\bar{T})\int_0^{T_i} Y(t)u^i(t)\,dt \right. \\
&\quad \left. -\Phi(\bar{T})x^0 - \Phi(\bar{T})\int_0^{\bar{T}} Y(t)u^i(t)\,dt + \Phi(\bar{T})\int_0^{T_i} Y(t)u^i(t)\,dt \right\| \\
&\leq \left\|(\Phi(T_i) - \Phi(\bar{T}))x^0\right\| + \left\|(\Phi(T_i) - \Phi(\bar{T}))\int_0^{T_i} Y(t)u^i(t)\,dt\right\| \\
&\quad + \left\|\Phi(\bar{T})\int_{\bar{T}}^{T_i} Y(t)u^i(t)\,dt\right\|
\end{aligned}$$

which implies because of the continuity of Φ

$$x_1 = \lim_{i\to\infty} x(\bar{T}, u^i).$$

Since $x(\bar{T}, u^i) \in K(\bar{T})$ for all $i \in \mathbb{N}$ and the set $K(\bar{T})$ is closed, we get $x^1 \in K(\bar{T})$ which completes the proof. □

In our problem formulation we assume that the terminal condition (9.11) is satisfied. Therefore Theorem 9.7 ensures that a time minimal control exists without additional assumptions. For the presentation of a necessary condition for such a time minimal control we need some lemmas given in the following.

Lemma 9.8 (property of the set of attainability).

Let the linear system (9.9) with the initial condition (9.10) be given. Then the set-valued mapping $K : [0, \hat{T}] \to 2^{\mathbb{R}^n}$ (where $K(\cdot)$ denotes the set of attainability) is continuous (with respect to the Hausdorff distance).

Proof First we prove the continuity of the mapping R. For that proof let $\bar{T}, T \in [0, \hat{T}]$, with $\bar{T} \neq T$, be arbitrarily chosen. Without loss of generality we assume $\bar{T} < T$. Then for an arbitrary $\bar{y} \in R(\bar{T})$ there is a feasible control $\bar{u}$ with

$$\bar{y} = \int_0^{\bar{T}} Y(t)\bar{u}(t)\,dt.$$

For the feasible control u given by

$$u(t) = \left\{ \begin{array}{l} \bar{u}(t) \text{ almost everywhere on } [0, \bar{T}] \\ (1, \ldots, 1)^T \text{ for all } t \in (\bar{T}, T] \end{array} \right\}$$

we have

$$\int_0^{T} Y(t)u(t)\,dt \in R(T).$$

Consequently we get

$$\begin{aligned} d(\bar{y}, R(T)) &:= \min_{y \in R(T)} \|\bar{y} - y\| \\ &\leq \left\| \bar{y} - \int_0^T Y(t)u(t)\,dt \right\| \\ &= \left\| \int_{\bar{T}}^T Y(t)(1, \ldots, 1)^T dt \right\| \\ &\leq \sqrt{m} \int_{\bar{T}}^T |||Y(t)|||\,dt \end{aligned}$$

and

$$\max_{\bar{y} \in R(\bar{T})} d(\bar{y}, R(T)) \leq \sqrt{m} \int_{\bar{T}}^T |||Y(t)|||\,dt$$

(here $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$ and $|||\cdot|||$ denotes the spectral norm). Similarly one can show

$$\max_{y\in R(T)} d(R(\bar{T}), y) \leq \sqrt{m} \int_{\bar{T}}^{T} |||Y(t)||| \, dt.$$

Hence, we obtain for the metric ϱ:

$$\begin{aligned}\varrho(R(\bar{T}), R(T)) &:= \max_{\bar{y}\in R(\bar{T})} \min_{y\in R(T)} \|\bar{y}-y\| + \max_{y\in R(T)} \min_{\bar{y}\in R(\bar{T})} \|\bar{y}-y\| \\ &\leq 2\sqrt{m} \int_{\bar{T}}^{T} |||Y(t)||| \, dt.\end{aligned}$$

Since the matrix function Y is continuous, there is a constant $\alpha > 0$ with

$$|||Y(t)||| \leq \alpha \text{ for all } t \in [0, \hat{T}].$$

Then we get

$$\varrho(R(\bar{T}), R(T)) \leq 2\alpha\sqrt{m}(T-\bar{T}).$$

Consequently, the set-valued mapping R is continuous. Since the fundamental matrix Φ is continuous and the images of the set-valued mapping R are bounded sets, we obtain with the equation (9.13) (notice for $\bar{T}, T \in [0, \hat{T}]$ and a constant $\beta > 0$ the inequality $\varrho(K(\bar{T}), K(T)) \leq \beta|||\Phi(\bar{T}) - \Phi(T)||| + |||\Phi(\bar{T})|||\, \varrho(R(\bar{T}), R(T)))$ that the mapping K is continuous. □

Lemma 9.9 (property of the set of attainability).

Let the linear system (9.9) with the initial condition (9.10) and some $\bar{T} \in [0, \hat{T}]$ be given. Let $\bar{y}$ be a point in the interior of the set $K(\bar{T})$ of attainability, then there is a time $T \in (0, \bar{T})$ so that $\bar{y}$ is also an interior point of $K(T)$.

Proof Let $\bar{y}$ be an interior point of the set $K(\bar{T})$ (this implies $\bar{T} > 0$). Then there is an $\varepsilon > 0$ so that $B(\bar{y}, \varepsilon) \subset K(\bar{T})$ for the closed ball $B(\bar{y}, \varepsilon)$ around $\bar{y}$ with radius ε. Now we assume that for all $T \in (0, \bar{T})$ $\bar{y}$ is not an interior point of the set $K(T)$. For every $T \in (0, \bar{T})$ the set $K(T) \subset \mathbb{R}^n$ is closed and convex. Then for every $T \in (0, \bar{T})$ there is a hyperplane separating the set $K(T)$ and the point $\bar{y}$ (compare Theorems C.5 and C.3). Consequently, for every $T \in (0, \bar{T})$ there is a point $y_T \in B(\bar{y}, \varepsilon)$ whose distance to the set $K(T)$ is at least ε. But this contradicts the continuity of the set-valued mapping K. □

The next lemma is the key for the proof of a necessary condition for time minimal controls. For the formulation of this result we use the function sgn : $\mathbb{R} \to \mathbb{R}$ given by

$$\mathrm{sgn}(y) = \left\{ \begin{array}{r} 1 \text{ for } y > 0 \\ 0 \text{ for } y = 0 \\ -1 \text{ for } y < 0 \end{array} \right\}.$$

Lemma 9.10 (property of the set of attainability).

Let the linear system (9.9) with the initial condition (9.10) and some $\bar{T} \in (0, \hat{T}]$ be given. If $\bar{x}(\bar{T}, \bar{u}) \in \partial K(\bar{T})$ for some $\bar{u} \in U(\bar{T})$, then there is a vector $\eta \neq 0_{\mathbb{R}^n}$ so that for all $k \in \{1, \ldots, m\}$:

$$\bar{u}_k(t) = sgn[\eta^T Y_k(t)] \text{ almost everywhere on } \{t \in [0, \bar{T}] \mid \eta^T Y_k(t) \neq 0\}$$

($\bar{x}(\bar{T}, \bar{u})$ denotes the state at the time $\bar{T}$ with respect to the control $\bar{u}$; $Y_k(t)$ denotes the k-th column of the matrix $Y(t)$).

Proof Let an arbitrary point $\bar{y} := \bar{x}(\bar{T}, \bar{u}) \in \partial K(\bar{T})$ be given. Since the set $K(\bar{T})$ is a convex and closed subset of $\mathbb{R}^n$, by a separation theorem (see Theorem C.5) there is a vector $\bar{\eta} \neq 0_{\mathbb{R}^n}$ with the property

$$\bar{\eta}^T \bar{y} \geq \bar{\eta}^T y \text{ for all } y \in K(\bar{T}).$$

Because of

$$\bar{\eta}^T \bar{y} = \bar{\eta}^T \Phi(\bar{T}) x^0 + \bar{\eta}^T \Phi(\bar{T}) \int_0^{\bar{T}} Y(t) \bar{u}(t)\, dt$$

and

$$\bar{\eta}^T y = \bar{\eta}^T \Phi(\bar{T}) x^0 + \bar{\eta}^T \Phi(\bar{T}) \int_0^{\bar{T}} Y(t) u(t)\, dt \text{ for all } y \in K(\bar{T})$$

we obtain for $\eta^T := \bar{\eta}^T \Phi(\bar{T})$

$$\eta^T \int\limits_0^{\bar{T}} Y(t)\bar{u}(t)\,dt \geq \eta^T \int\limits_0^{\bar{T}} Y(t)u(t)\,dt \tag{9.14}$$

for all feasible controls steering the linear system (9.9) with the initial condition (9.10) to a state in the set $K(\bar{T})$ of attainability. From the inequality (9.14) we conclude

$$\eta^T Y(t)\bar{u}(t) \geq \eta^T Y(t)u(t) \text{ almost everywhere on } [0, \bar{T}]. \tag{9.15}$$

For the proof of the implication "(9.14) $\Longrightarrow$ (9.15)" we assume that the inequality (9.15) is not true. Then there is a feasible control u and a set $M \subset [0, \bar{T}]$ with positive measure so that

$$\eta^T Y(t)\bar{u}(t) < \eta^T Y(t)u(t) \text{ almost everywhere on } M.$$

If one defines the feasible control u^* by

$$u^*(t) = \left\{ \begin{array}{ll} \bar{u}(t) & \text{almost everywhere on } [0, \bar{T}] \setminus M \\ u(t) & \text{almost everywhere on } M \end{array} \right\},$$

then it follows

$$\begin{aligned} \eta^T \int\limits_0^{\bar{T}} Y(t)u^*(t)\,dt &= \eta^T \int\limits_M Y(t)u(t)\,dt + \eta^T \int\limits_{[0,\bar{T}]\setminus M} Y(t)\bar{u}(t)\,dt \\ &> \eta^T \int\limits_M Y(t)\bar{u}(t)\,dt + \eta^T \int\limits_{[0,\bar{T}]\setminus M} Y(t)\bar{u}(t)\,dt \\ &= \eta^T \int\limits_0^{\bar{T}} Y(t)\bar{u}(t)\,dt \end{aligned}$$

which contradicts the inequality (9.14). Hence, the inequality (9.15) is true.

From the inequality (9.15) we get for all $k \in \{1, \ldots, m\}$

$$\bar{u}_k(t) = \text{sgn}\,[\eta^T Y_k(t)] \text{ almost everywhere on } \{t \in [0, \bar{T}] \mid \eta^T Y_k(t) \neq 0\}. \qquad \Box$$

Now we present the afore-mentioned necessary condition for time minimal controls.

Theorem 9.11 (necessary condition for time minimal controls).

Let the linear system (9.9) with the initial condition (9.10) and the terminal condition (9.11) be given. If $\bar{u}$ is a time minimal control with respect to the minimal terminal time $\bar{T} \in [0, \hat{T}]$, then there is a vector $\eta \neq 0_{\mathbb{R}^n}$ so that for all $k \in \{1, \ldots, m\}$:

$$\bar{u}_k(t) = sgn[\eta^T Y_k(t)] \text{ almost everywhere on } \{t \in [0, \bar{T}] \mid \eta^T Y_k(t) \neq 0\}. \tag{9.16}$$

Proof The assertion is obvious for $\bar{T} = 0$. Therefore we assume $\bar{T} > 0$ for the following. We want to show that

$$\bar{y} := \Phi(\bar{T})x^0 + \Phi(\bar{T}) \int_0^{\bar{T}} Y(t)\bar{u}(t)\, dt \;\in\; \partial K(\bar{T}). \tag{9.17}$$

Suppose that $\bar{y}$ were an interior point of the set $K(\bar{T})$ of attainability. Then by Lemma 9.9 there is a time $T \in (0, \bar{T})$ so that $\bar{y}$ is also an interior point of the set $K(T)$. But this contradicts the fact that $\bar{T}$ is the minimal time. Hence, the condition (9.17) is true. An application of Lemma 9.10 completes the proof. □

The statement (9.16) is also called a *weak bang-bang principle*. If the measure of the set $\{t \in [0, \bar{T}] \mid \eta^T Y_k(t) = 0\}$ equals 0 for every $k \in \{1, \ldots, m\}$, the statement (9.16) is called a *strong bang-bang principle*. Theorem 9.11 can also be formulated as follows:

For every time minimal control $\bar{u}$ there is a vector $\eta \neq 0_{\mathbb{R}^n}$
so that $\bar{u}$ satisfies the weak bang-bang principle (9.16).

The next example illustrates the applicability of Theorem 9.11.

Example 9.12 (necessary condition for time minimal controls).

We consider the harmonic oscillator mathematically formalized by

$$\ddot{y}(t) + y(t) = u(t) \text{ almost everywhere on } [0, \hat{T}],$$

$$\|u\|_{L_\infty([0,\hat{T}])} \leq 1$$

where $\hat{T} > 0$ is sufficiently large. An initial condition is not given explicitly. The corresponding linear system of first order reads

$$\dot{x}(t) = \underbrace{\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}}_{=:\,A} x(t) + \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{=:\,B} u(t).$$

We have

$$\Phi(t) = e^{At} = \sum_{i=0}^{\infty} A^i \frac{t^i}{i!} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$$

and

$$Y(t) = \Phi(t)^{-1} B = e^{-At} B = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}.$$

Then we obtain for an arbitrary vector $\eta \neq 0_{\mathbb{R}^n}$

$$\eta^T Y(t) = -\eta_1 \sin t + \eta_2 \cos t.$$

Consequently, we get for a number $\alpha \in \mathbb{R}$ and a number $\delta \in [-\pi, \pi]$

$$\eta^T Y(t) = \alpha \sin(t + \delta)$$

and therefore

$$\text{sgn}[\eta^T Y(t)] = \text{sgn}[\alpha \sin(t + \delta)]$$

(see Fig. 9.4).

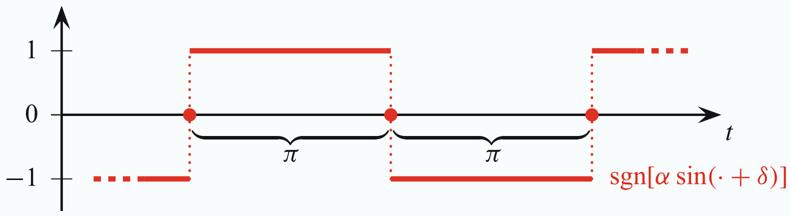

Fig. 9.4 Illustration of the time optimal control

Conclusion: If there is a time minimal control $\bar{u}$, then it fulfills the strong bang-bang principle, and therefore it is unique. After π time units one always gets a change of the sign of $\bar{u}$.
With a standard result from control theory one can see that the considered linear system is null controllable (i.e., it can be steered to the origin in a finite time). Hence, by Theorem 9.7 there is also a time minimal control $\bar{u}$ which steers this system into a state of rest, and therefore the preceding results are applicable.

Now we present an example for which the necessary condition for time minimal controls does not give any information.

Example 9.13 (necessary condition for time minimal controls).

We investigate the simple linear system

$$\left.\begin{aligned} \dot{x}_1(t) &= x_1(t) + u(t) \\ \dot{x}_2(t) &= x_2(t) + u(t) \end{aligned}\right\} \text{ almost everywhere on } [0, \hat{T}]$$

with

$$\|u\|_{L_\infty[0,\hat{T}]} \leq 1$$

and $\hat{T} > 0$. Here we set

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I \text{ and } B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Then we obtain

$$Y(t) = e^{-At} B = e^{-t} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and for any vector $\eta \neq 0_{\mathbb{R}^2}$ we get

$$\eta^T Y(t) = (\eta_1 + \eta_2) e^{-t}.$$

For example, for $\eta = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ we conclude

$$\eta^T Y(t) = 0 \text{ for all } t \in [0, \hat{T}],$$

and Theorem 9.11 does not give a suitable necessary condition for time minimal controls.

Next we investigate the question under which conditions time minimal controls are unique. For this investigation we introduce the notion of normality.

Definition 9.14 (normal linear system).

(a) The linear system (9.9) is called *normal on* $[0, T]$ (with $T \in [0, \hat{T}]$), if for every vector $\eta \neq 0_{\mathbb{R}^n}$ the sets

$$G_k(\eta) = \{t \in [0, T] \mid \eta^T Y_k(t) = 0\} \text{ with } k \in \{1, \ldots, m\}$$

have the measure 0. $Y_k(t)$ denotes again the k-th column of the matrix $Y(t)$.

(b) The linear system (9.9) is called *normal*, if for every $T \in [0, \hat{T}]$ this system is normal on $[0, T]$.

Theorem 9.15 (uniqueness of a time minimal control).

Let the linear system (9.9) with the initial condition (9.10) and the terminal condition (9.11) be given. If $\bar{u}$ is a time minimal control with respect to the minimal terminal time $\bar{T} \in [0, \hat{T}]$ and if the linear system (9.9) is normal on $[0, \bar{T}]$, then $\bar{u}$ is the unique time minimal control.

Proof By Theorem 9.11 for every time minimal control $\bar{u}$ there is a vector $\eta \neq 0_{\mathbb{R}^n}$ so that for all $k \in \{1, \ldots, m\}$:

$$\bar{u}_k(t) = \text{sgn}[\eta^T Y_k(t)] \text{ almost everywhere on } [0, \bar{T}] \setminus G_k(\eta).$$

Then the assertion follows from the normality assumption (notice that in the proof of Lemma 9.10 the vector η depends on the terminal state and not on the control). □

A control $\bar{u}$ which satisfies the assumptions of Theorem 9.15 fulfills the strong bang-bang principle

$$\bar{u}(t) = \text{sgn}[\eta^T Y_k(t)] \text{ almost everywhere on } [0, \bar{T}].$$

One obtains an interesting characterization of the concept of normality in the case of an autonomous linear system (9.9) with constant matrix functions A and B.

Theorem 9.16 (characterization of normality).

The autonomous linear system (9.9) with constant matrix functions A and B is normal if and only if for every $k \in \{1, \ldots, m\}$ either

$$\text{rank}\,(B_k, AB_k, \ldots, A^{n-1}B_k) = n \tag{9.18}$$

or

$$\text{rank}\,(A - \lambda I, B_k) = n \text{ for all eigenvalues } \lambda \text{ of } A. \tag{9.19}$$

Here B_k denotes the k-th column of the matrix B.

Proof We fix an arbitrary terminal time $T \in [0, \hat{T}]$. First notice that for every $k \in \{1, \ldots, m\}$ and every $\eta \in \mathbb{R}^n$

$$\eta^T Y_k(t) = \eta^T e^{-At} B_k \text{ for all } t \in [0, T].$$

Consequently, the real-valued analytical function $\eta^T Y_k(\cdot)$ on $[0, T]$ is either identical to 0 or it has a finite number of zeros on this interval. Therefore, the autonomous linear system (9.9) is normal on $[0, T]$ if and only if the following implication is satisfied:

$$\eta^T e^{-At} B_k = 0 \text{ for all } t \in [0, T] \text{ and some } k \in \{1, \ldots, m\} \Rightarrow \eta = 0_{\mathbb{R}^n}. \tag{9.20}$$

Next we show that the implication (9.20) is equivalent to the condition (9.18). For this proof we assume that the condition (9.18) is satisfied. Let a vector $\eta \in \mathbb{R}^n$ with

$$\eta^T e^{-At} B_k = 0 \text{ for all } t \in [0, T] \text{ and some } k \in \{1, \ldots, m\}$$

be arbitrarily given. By repeated differentiation and setting "$t = 0$" we get

$$\eta^T (B_k, AB_k, \ldots, A^{n-1}B_k) = 0_{\mathbb{R}^n}^T \text{ for some } k \in \{1, \ldots, m\}.$$

By assumption the system of row vectors of the matrix $(B_k, AB_k, \ldots, A^{n-1}B_k)$ is linear independent, and therefore we get $\eta = 0_{\mathbb{R}^n}$. Hence, the implication (9.20) is satisfied, i.e. the autonomous linear system (9.9) is normal on $[0, T]$.

Now we assume that the condition (9.18) is not satisfied. This means that for some $k \in \{1, \ldots, m\}$ the system of row vectors of the matrix $(B_k, AB_k, \ldots, A^{n-1}B_k)$ is linear dependent. Then there is a vector $\eta \neq 0_{\mathbb{R}^n}$ with

$$\eta^T (B_k, AB_k, \ldots, A^{n-1}B_k) = 0_{\mathbb{R}^n}^T$$

which implies

$$\eta^T B_k = \eta^T A B_k = \cdots = \eta^T A^{n-1} B_k = 0. \tag{9.21}$$

The Cayley-Hamilton theorem states that the matrix A satisfies its characteristic equation, i.e.

$$A^n = \alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1} A^{n-1}$$

with appropriate coefficients $\alpha_0, \alpha_1, \ldots, \alpha_{n-1} \in \mathbb{R}$. Then we obtain with (9.21)

$$\eta^T A^n B_k = \alpha_0 \eta^T B_k + \alpha_1 \eta^T A B_k + \cdots + \alpha_{n-1} \eta^T A^{n-1} B_k = 0$$

and by induction

$$\eta^T A^l B_k = 0 \text{ for all } l \geq n. \tag{9.22}$$

Equations (9.21) and (9.22) imply

$$\eta^T A^l B_k = 0 \text{ for all } l \geq 0$$

which leads to

$$\eta^T e^{-At} B_k = \eta^T \left(\sum_{i=0}^{\infty} A^i \frac{(-t)^i}{i!} \right) B_k = 0 \text{ for all } t \in [0, T].$$

Consequently, the implication (9.20) is not satisfied, i.e. the autonomous linear system (9.9) is not normal on $[0, T]$.

Finally we show the equivalence of the two rank conditions (9.18) and (9.19). Let $k \in \{1, \ldots, m\}$ be arbitrarily chosen.

Assume that the condition (9.19) is not satisfied, i.e. for some possibly complex eigenvalue λ of A we have

$$\text{rank}\,(A - \lambda I, B_k) \neq n.$$

Then there is a vector $z \in \mathbb{R}^n$ with $z \neq 0_{\mathbb{R}^n}$ and

$$z^T (A - \lambda I, B_k) = 0^T_{\mathbb{R}^{n+1}},$$

i.e.

$$z^T A = \lambda z^T \tag{9.23}$$

and

$$z^T B_k = 0. \tag{9.24}$$

With the equations (9.23) and (9.24) we conclude

$$z^T A B_k = \lambda z^T B_k = 0,$$

and by induction we get

$$z^T A^l B_k = 0 \text{ for all } l \geq 0.$$

Hence we have

$$\text{rank } (B_k, AB_k, \ldots, A^{n-1} B_k) \neq n.$$

Conversely, we assume now that the equation (9.18) is not satisfied. Then there is a $z \neq 0_{\mathbb{R}^n}$ with

$$z^T B_k = 0, \; z^T A B_k = 0, \ldots, \; z^T A^{n-1} B_k = 0.$$

Again with the Cayley-Hamilton theorem we conclude immediately

$$z^T A^l B_k = 0 \text{ for all } l \geq 0.$$

Consequently, the linear subspace

$$S := \{\tilde{z} \in \mathbb{R}^n \mid \tilde{z}^T A^l B_k = 0 \text{ for all } l \geq 0\}$$

has the dimension ≥ 1. Since the set S is invariant under A^T (i.e. $A^T S \subset S$), one eigenvector $\bar{z}$ of A^T belongs to S. Hence, there is an eigenvalue λ of A^T which is also an eigenvalue of A so that

$$A^T \bar{z} = \lambda \bar{z}$$

or alternatively

$$\bar{z}^T (A - \lambda I) = 0^T_{\mathbb{R}^n}. \tag{9.25}$$

Because of $\bar{z} \in S$ we obtain with $l = 0$

$$\bar{z}^T B_k = 0. \tag{9.26}$$

Equations (9.25) and (9.26) imply

$$\text{rank}\,(A - \lambda I, B_k) \neq n \text{ for some eigenvalue } \lambda \text{ of } A.$$

This completes the proof. □

In control theory the condition

$$\text{rank}\,(B, AB, \ldots, A^{n-1}B) = n$$

is called the *Kalman condition*. It is obvious that the condition

$$\text{rank}\,(B_k, AB_k, \ldots, A^{n-1}B_k) = n \text{ for all } k \in \{1, \ldots, m\}$$

which is given in Theorem 9.16 implies the Kalman condition. Moreover, in control theory the condition

$$\text{rank}\,(A - \lambda I, B) = n \text{ for all eigenvalues } \lambda \text{ of } A$$

is called the *Hautus condition* which is implied by the condition

$$\text{rank}\,(A - \lambda I, B_k) = n \text{ for all } k \in \{1, \ldots, m\} \text{ and all eigenvalues } \lambda \text{ of } A.$$

One can show with the same arguments as in the proof of Theorem 9.16 that the Kalman and Hautus conditions are equivalent. In control theory one proves that the Kalman condition (or the Hautus condition) characterizes the controllability of an autonomous linear system, i.e. in this case there is an unconstrained control which steers the autonomous linear system from an arbitrary initial state to an arbitrary terminal state in finite time.

The following example shows that the Kalman condition (or the Hautus condition) does not imply the condition (9.18) (and (9.19), respectively).

Example 9.17 (Kalman condition).

The following autonomous linear system satisfies the Kalman condition but it is not normal:

$$\left.\begin{aligned} \dot{x}_1(t) &= -x_1(t) + u_1(t) \\ \dot{x}_2(t) &= -2x_2(t) + u_1(t) + u_2(t) \end{aligned}\right\} \text{ almost everywhere on } [0, \hat{T}]$$

with some $\hat{T} > 0$. Here we set

$$A = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Then we have

$$B_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \; AB_1 = \begin{pmatrix} -1 \\ -2 \end{pmatrix},$$

$$B_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \; AB_2 = \begin{pmatrix} 0 \\ -2 \end{pmatrix}.$$

The matrix (B_2, AB_2) has the rank 1, and therefore the linear system is not normal. On the other hand we have

$$\text{rank } (B, AB) = 2,$$

i.e. the Kalman condition is satisfied.

Exercises

(9.1) Consider the differential equation

$$\dot{x}(t) = 2x(t) - 3u(t) \text{ almost everywhere on } [0, 2]$$

with the initial condition

$$x(0) = x^0$$

for an arbitrarily chosen $x^0 \in \mathbb{R}$. Determine an optimal control $\bar{u} \in L_\infty([0, 2])$ as a minimal point of the objective functional $J : L_\infty([0, 2]) \to \mathbb{R}$ with

$$J(u) = \frac{1}{2}x(1)^2 + 2\int_0^2 u(t)^2 \, dt \text{ for all } u \in L_\infty([0, 2]).$$

(9.2) ([51, p. 132–133]) Let the initial value problem

$$\dot{x}(t) = u(t) \text{ almost everywhere on } [0, 1],$$

$$x(0) = 1$$

be given. Determine an optimal control $u \in L_\infty([0,1])$ for which the objective functional $J : L_\infty([0,1]) \to \mathbb{R}$ with

$$J(u) = \int_0^1 \left(u(t)^2 + x(t)^2\right) dt \text{ for all } u \in L_\infty([0,1])$$

becomes minimal.

(9.3) Consider the linear differential equation of n-th order

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_0 y(t) = u(t)$$
$$\text{almost everywhere on } [0,\hat{T}]$$

where $\hat{T} > 0$ and $a_0, \ldots, a_{n-1} \in \mathbb{R}$ are given constants. The control u is assumed to be an $L_\infty([0,\hat{T}])$ function. Show that the system of linear differential equations of first order which is equivalent to this differential equation of n-th order satisfies the Kalman condition.

(9.4) ([216, p. 22–24]) Let the system of linear differential equations

$$\dot{x}(t) = Ax(t) + Bu(t) \text{ almost everywhere on } [0,\hat{T}]$$

with

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\alpha & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 \\ -\beta \\ 0 \\ \gamma \end{pmatrix}$$

be given where $\hat{T} > 0$, $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ are constants. It is assumed that $u \in L_\infty([0,\hat{T}])$. Show that this system satisfies the Hautus condition.

(9.5) For the linear system in exercise (9.4) assume in addition that the terminal time $\hat{T}$ is sufficiently large. Moreover, let the initial condition

$$x(0) = x^0$$

with $x^0 \in \mathbb{R}^4$ and the terminal condition

$$x(\hat{T}) = 0_{\mathbb{R}^4}$$

be given. For the control u we assume

$$\|u\|_{L_\infty([0,\hat{T}])} \leq 1.$$

It can be proved with a known result from control theory that this system can be steered from x^0 to $0_{\mathbb{R}^4}$ in finite time. Show then that a time minimal control exists which is unique, and give a characterization of this time minimal control.

Weak Convergence

A

Definition A.1 (weakly convergent sequence).

Let $(X, \|\cdot\|)$ be a normed space. A sequence $(x_n)_{n\in\mathbb{N}}$ of elements of X is called *weakly convergent* to some $\bar{x} \in X$ if for all continuous linear functionals l on X

$$\lim_{n\to\infty} l(x_n) = l(\bar{x}).$$

In this case $\bar{x}$ is called a *weak limit* of the sequence $(x_n)_{n\in\mathbb{N}}$.

In a finite dimensional normed space a sequence is weakly convergent if and only if it is convergent. In an arbitrary normed space every convergent sequence is also weakly convergent; the converse statement does not hold in general.

Example A.2 (weakly convergent sequence).

Consider the Hilbert space l_2 of all real sequences $x = (x^i)_{i\in\mathbb{N}}$ with $\sum_{i=1}^{\infty} |x^i|^2 < \infty$. In this linear space we investigate the special sequence

$$x_1 := (1, 0, 0, 0, \ldots),$$

$$x_2 := (0, 1, 0, 0, \ldots),$$

$$x_3 := (0, 0, 1, 0, \ldots),$$

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3

and so on. This sequence converges weakly to 0_{l_2} because for each continuous linear functional l on l_2 there is a $y \in l_2$ with

$$l(x) = \langle y, x \rangle \text{ for all } x \in l_2$$

so that

$$\lim_{n\to\infty} l(x_n) = \lim_{n\to\infty} \langle y, x_n \rangle = \lim_{n\to\infty} \sum_{i=1}^{\infty} y^i x_n^i = \lim_{n\to\infty} y^n = 0.$$

On the other hand the sequence $(x_n)_{n\in\mathbb{N}}$ does not converge to 0_{l_2} because

$$\|x_n - 0_{l_2}\| = \|x_n\| = \sqrt{\langle x_n, x_n \rangle} = \sqrt{\sum_{i=1}^{\infty} (x_n^i)^2} = 1 \text{ for all } n \in \mathbb{N}.$$

Definition A.3 (weakly sequentially closed set).

Let $(X, \|\cdot\|)$ be a normed space. A nonempty subset S of X is called *weakly sequentially closed* if for every weakly convergent sequence in S the weak limit also belongs to S.

Every weakly sequentially closed subset of a normed space is also closed (because every convergent sequence converges weakly to the same limit). The converse statement is not true in general. But every nonempty convex closed subset of a normed space is also weakly sequentially closed.

Definition A.4 (weakly sequentially compact set).

Let $(X, \|\cdot\|)$ be a normed space. A nonempty subset S of X is called *weakly sequentially compact* if every sequence in S contains a weakly convergent subsequence whose weak limit belongs to S.

A nonempty subset of a normed space is weakly sequentially compact if and only if it is weakly compact (i.e. compact with respect to the weak topology). In a finite dimensional normed space a nonempty subset is weakly sequentially compact if and only if it is closed and bounded.

Reflexivity of Banach Spaces

B

Definition B.1 (Banach space).

A complete normed space is called a *Banach space*.

Using a James theorem (e.g., compare [177, § 19]) a sufficient condition for the weak sequence compactness of a nonempty subset of a real Banach space can be given.

Theorem B.2 (weakly sequentially compact set).

Let S be a nonempty convex bounded closed subset of a real Banach space. If every continuous linear functional attains its supremum on S, then the set S is weakly sequentially compact.

Reflexive normed spaces are special Banach spaces. In specialist literature a normed linear space $(X, \|\cdot\|)$ is called reflexive if the canonical embedding of X into X^{**} is surjective — but here we use a known characterization for the definition of this notion.

Definition B.3 (reflexive Banach space).

A Banach space $(X, \|\cdot\|)$ is called *reflexive* if the closed unit ball $\{x \in X \mid \|x\| \leq 1\}$ is weakly sequentially compact.

Every finite dimensional normed space is reflexive. For instance, the linear space $L_1[0, 1]$ of Lebesgue integrable real-valued functions on [0,1] is a Banach space, but it is not reflexive.

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3

In a reflexive Banach space a simple sufficient condition for the weak sequence compactness of a nonempty subset can be given (for instance, compare [365, Cor. 6.1.9]).

Theorem B.4 (weakly sequentially compact set).

Every nonempty convex bounded closed subset of a reflexive Banach space is weakly sequentially compact.

Notice that in a finite dimensional normed space the assumption of convexity can be dropped.

Hahn-Banach Theorem

C

The following theorem is also called a *basic version of the Hahn-Banach theorem* (for a proof, for instance, compare [190, Thm. 3.8]).

Theorem C.1 (basic version of the Hahn-Banach theorem).

Let X be a real linear space. For every sublinear functional $f : X \to \mathbb{R}$ there is a linear functional l on X with

$$l(x) \leq f(x) \text{ for all } x \in X$$

(see Fig. C.1).

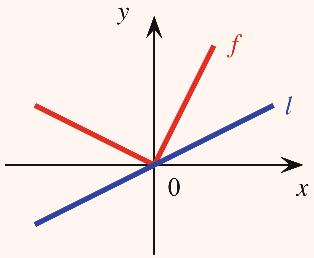

Fig. C.1 Illustration of the result of Theorem C.1

Besides this basic version there are further versions of the Hahn-Banach theorem. The following *Eidelheit separation theorem* can be deduced from Theorem C.1 (for a proof see [190, Thm. 3.16]).

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3

Theorem C.2 (Eidelheit separation theorem).

Let S and T be nonempty convex subsets of a real topological linear space X with $int(S) \neq \emptyset$. Then we have $int(S) \cap T = \emptyset$ if and only if there are a continuous linear functional $l \in X^* \setminus \{0_{X^*}\}$ and a real number γ with

$$l(s) \leq \gamma \leq l(t) \text{ for all } s \in S \text{ and all } t \in T$$

and

$$l(s) < \gamma \text{ for all } s \in int(S)$$

(see Fig. C.2).

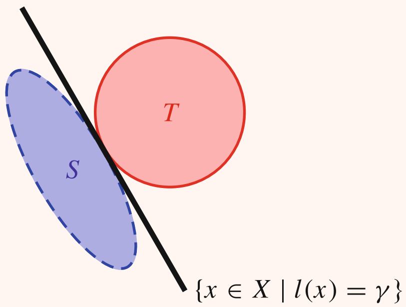

Fig. C.2 Illustration of the result of Theorem C.2

The following separation theorem can be obtained from the preceding theorem.

Theorem C.3 (strict separation theorem).

Let S be a nonempty convex and closed subset of a real locally convex space X. Then we have $x \in X \setminus S$ if and only if there is a continuous linear functional $l \in X^* \setminus \{0_{X^*}\}$ with

$$l(x) < \inf_{s \in S} l(s). \tag{C.1}$$

Proof

(a) Let any $x \in X$ be given. If there is a continuous linear functional $l \in X^* \setminus \{0_{X^*}\}$ with the property (C.1), then it follows immediately $x \notin S$.

(b) Choose an arbitrary element $x \in X \setminus S$. Since S is closed, there is a convex neighborhood N of x with $N \cap S = \emptyset$. By the Eidelheit separation theorem

(Theorem C.2) there are a continuous linear functional $l \in X^* \setminus \{0_{X^*}\}$ and a real number γ with

$$l(x) < \gamma \le l(s) \text{ for all } s \in S.$$

The inequality (C.1) follows directly from the previous inequality.

□

The next result is a special version of the Hahn-Banach theorem deduced by the Eidelheit separation theorem.

Theorem C.4 (continuous linear functional with special property).

Let $(X, \|\cdot\|_X)$ be a real normed space. For every $x \in X$ there is an $l \in X^*$ with $\|l\|_{X^*} = 1$ and $l(x) = \|x\|_X$.

Proof For $x = 0_X$ the assertion is evident. Therefore assume in the following that any $x \neq 0_X$ is arbitrarily given. Let S denote the closed ball around zero with the radius $\|x\|$, and let $T := \{x\}$. Because of $\text{int}(S) \cap T = \emptyset$ by the Eidelheit separation theorem (Theorem C.2) there are an $\bar{l} \in X^* \setminus \{0_{X^*}\}$ and a $\gamma \in \mathbb{R}$ with

$$\bar{l}(s) \le \gamma \le \bar{l}(x) \text{ for all } s \in S.$$

If we define $l := \frac{1}{\|\bar{l}\|_{X^*}} \bar{l}$, we have $\|l\|_{X^*} = 1$ and

$$l(s) \le l(x) \text{ for all } s \in S.$$

Then we get

$$\begin{aligned} \|x\|_X &= \|x\|_X \sup_{\|y\|_X \le 1} |l(y)| \\ &= \sup_{\|y\|_X \le 1} |l(\|x\|_X\, y)| \\ &= \sup_{s \in S} |l(s)| \\ &= \sup_{s \in S} l(s) \\ &\le l(x). \end{aligned} \tag{C.2}$$

Since $\|l\|_{X^*} = 1$ we have

$$\sup_{y \neq 0_X} \frac{|l(y)|}{\|y\|_X} = 1$$

resulting in

$$l(y) \le \|y\|_X \text{ for all } y \in X. \tag{C.3}$$

From the inequality (C.3) we obtain

$$l(x) \le \|x\|_X$$

and together with the inequality (C.2) we conclude

$$l(x) = \|x\|_X.$$

This completes the proof. □

Finally we present a special separation theorem in a finite dimensional space (for instance, compare [365, Thm. 3.2.6]). This result is in general not true in an infinite dimensional setting.

Theorem C.5 (separation in a finite dimensional space).

Let S be a nonempty convex and closed subset of a finite dimensional real normed space $(X, \|\cdot\|_X)$. Then for every boundary point $\bar{x} \in \partial S$ there is a continuous linear functional $l \in X^* \setminus \{0_{X^*}\}$ with

$$l(s) \le l(\bar{x}) \text{ for all } s \in S.$$

Partially Ordered Linear Spaces

D

Definition D.1 (partially ordered linear space).

Let X be a real linear space.

(a) Every nonempty subset R of the product space $X \times X$ is called a *binary relation* R on X (one writes xRy for $(x, y) \in R$).

(b) Every binary relation $\leq$ on X is called a *partial ordering* on X, if for arbitrary $w, x, y, z \in X$:

(i) $x \leq x$ (reflexivity);

(ii) $x \leq y,\ y \leq z \Rightarrow x \leq z$ (transitivity);

(iii) $x \leq y,\ w \leq z \Rightarrow x + w \leq y + z$ (compatibility with the addition);

(iv) $x \leq y,\ \alpha \in \mathbb{R}_+ \Rightarrow \alpha x \leq \alpha y$ (compatibility with the scalar multiplication).

(c) A partial ordering $\leq$ on X is called *antisymmetric*, if for arbitrary $x, y \in X$:

$$x \leq y,\ y \leq x \Rightarrow x = y.$$

(d) A real linear space equipped with a partial ordering is called a *partially ordered linear space*.

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3

Example D.2 (partially ordered linear space).

(a) If one defines the componentwise partial ordering $\leq$ on $\mathbb{R}^n$ by

$$\leq := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid x_i \leq y_i \text{ for all } i \in \{1, \ldots, n\}\},$$

then the linear space $\mathbb{R}^n$ becomes a partially ordered linear space.

(b) For $-\infty < a < b < \infty$ let $C[a, b]$ denote the linear space of all continuous real-valued functions on $[a, b]$. With the natural partial ordering $\leq$ on $C[a, b]$ given by

$$\leq := \{(x, y) \in C[a, b] \times C[a, b] \mid x(t) \leq y(t) \text{ for all } t \in [a, b]\}$$

the space $C[a, b]$ becomes a partially ordered linear space.

Notice that two arbitrary elements of a partially ordered linear space may not always be compared with each other with respect to the partial ordering. For instance, for the componentwise partial ordering $\leq$ on $\mathbb{R}^2$ we have

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} \nleq \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 \\ 1 \end{pmatrix} \nleq \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

The following theorem which is simple to prove says that partial orderings on linear spaces can be characterized by convex cones.

Theorem D.3 (characterization of a partial ordering).

Let X be a real linear space.

(a) If $\leq$ is a partial ordering on X, then the set

$$C := \{x \in X \mid 0_X \leq x\}$$

is a convex cone. If, in addition, the partial ordering is antisymmetric, then C is pointed .

(b) If C is a convex cone in X, then the binary relation

$$\leq := \{(x, y) \in X \times X \mid y - x \in C\}$$

is a partial ordering on X. If, in addition, C is pointed, then the partial ordering $\leq$ is antisymmetric .

Definition D.4 (ordering cone).

A convex cone characterizing the partial ordering on a real linear space is called an *ordering cone* (or also a *positive cone*).

Example D.5 (ordering cones).

(a) For the natural partial ordering given in Example D.2, (a) the ordering cone reads

$$C := \{x \in \mathbb{R}^n \mid x_i \geq 0 \text{ for all } i \in \{1, \ldots, n\}\} = \mathbb{R}^n_+$$

(compare Fig. D.1).

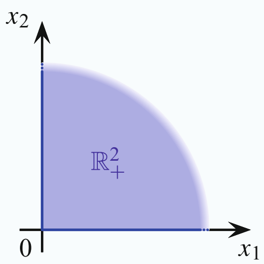

Fig. D.1 Illustration of the ordering cone $\mathbb{R}^2_+$

(b) In Example D.2, (b) the ordering cone can be written as

$$C := \{x \in C[a, b] \mid x(t) \geq 0 \text{ for all } t \in [a, b]\}$$

(compare Fig. D.2).

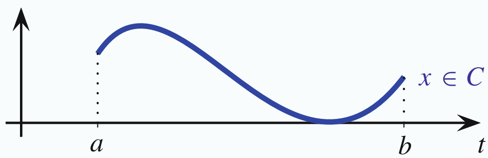

Fig. D.2 Illustration of a function $x \in C$

If a real linear space is partially ordered, then a partial ordering can also be introduced on its dual space.

Definition D.6 (dual cone).

Let X be a real linear space with an ordering cone C. The cone

$$C' := \{l \in X' \mid l(x) \geq 0 \text{ for all } x \in C\}$$

is called the *dual cone* for C (here X' denotes the algebraical dual space of X).

With the aid of the dual cone C' a partial ordering is described on the dual space X'. In the case of a real normed space $(X, \|\cdot\|)$ the dual cone in the topological dual space X^* is denoted by C^*.

Answers to the Exercises

Chapter 2

(2.1) Use Definition 2.1 and notice that in a finite dimensional normed space weak convergence is equivalent to norm convergence.

(2.2) Show for the functions $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ with

$$f_1(x) = \left\{ \begin{array}{ll} f(x) & \text{for all } x \leq -1 \\ -\frac{1}{e} & \text{for all } x > -1 \end{array} \right\}$$

and

$$f_2(x) = \left\{ \begin{array}{ll} -\frac{1}{e} & \text{for all } x < -1 \\ f(x) & \text{for all } x \geq -1 \end{array} \right\}$$

that the level sets $S_\alpha^{f_1} := \{x \in \mathbb{R} \mid f_1(x) \leq \alpha\}$ and $S_\alpha^{f_2} := \{x \in \mathbb{R} \mid f_2(x) \leq \alpha\}$ are convex for all $\alpha \in \mathbb{R}$. Then the level set $S_\alpha^f = S_\alpha^{f_1} \cap S_\alpha^{f_2}$ is convex as well.

(2.3) For the "$\Longrightarrow$" part of this proof consider the level set S_α with $\alpha := \max\{f(x), f(y)\}$. Prove the converse case by showing that S_α is convex.

(2.4) Take an arbitrary sequence $(x_n)_{n \in \mathbb{N}}$ in a proximinal set S converging to some $\bar{x}$. Then the approximation problem $\min\limits_{x \in S} \|x - \bar{x}\|$ has a solution $\tilde{x} \in S$. Since $\|\tilde{x} - \bar{x}\| \leq \|x_n - \bar{x}\| \underset{n \to \infty}{\longrightarrow} 0$, we conclude $\bar{x} = \tilde{x} \in S$.

(2.5) Apply Theorem 2.18.

(2.6) Notice the remarks at the end of Sect. 2.4.

(2.7) The constraint set S is not convex.

(2.8) In analogy to the proof of Theorem 2.23 show that f is convex and continuous. The assertion then follows from Theorem 2.12.

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3

Chapter 3

(3.1) For $h \neq 0$ we obtain

$$f'(0)(h) = \lim_{\lambda \to 0_+} \frac{1}{\lambda}\big(f(\lambda h) - f(0)\big) = \lim_{\lambda \to 0_+} \lambda h^2 \sin \frac{1}{\lambda h} = 0,$$

and for $h = 0$ we immediately get $f'(0)(h) = 0$.

(3.2) The result is trivial in the case of $\bar{x} = \hat{x}$. For $\bar{x} \neq \hat{x}$ we obtain for the directional derivative

$$\begin{aligned} & f'(\bar{x})(h) \\ &= \lim_{\lambda \to 0_+} \frac{1}{\lambda}\big(\|\bar{x} + \lambda h - \hat{x}\| - \|\bar{x} - \hat{x}\|\big) \\ &= \lim_{\lambda \to 0_+} \frac{1}{\lambda}\left(\max_{t \in M} |\bar{x}(t) + \lambda h(t) - \hat{x}(t)| - \max_{t \in M} |\bar{x}(t) - \hat{x}(t)|\right) \\ &\geq \max_{t \in M(\bar{x})} \operatorname{sgn}(\bar{x}(t) - \hat{x}(t))h(t) \text{ for all } h \in C(M). \end{aligned}$$

For every $\lambda > 0$ choose a $t_\lambda \in M$ with

$$|\bar{x}(t_\lambda) - \hat{x}(t_\lambda) + \lambda h(t_\lambda)| = \|\bar{x} - \hat{x} + \lambda h\|.$$

Then we conclude

$$\lim_{\lambda \to 0_+} |\bar{x}(t_\lambda) - \hat{x}(t_\lambda) + \lambda h(t_\lambda)| = \|\bar{x} - \hat{x}\|.$$

This implies the existence of a sequence $(\lambda_k)_{k \in \mathbb{N}}$ of positive numbers converging to 0 with $\lim_{k \to \infty} t_{\lambda_k} = t_0 \in M(\bar{x})$. Then we get for sufficiently large $k \in \mathbb{N}$

$$\begin{aligned} & \frac{1}{\lambda_k}|\bar{x}(t_{\lambda_k}) - \hat{x}(t_{\lambda_k}) + \lambda_k h(t_{\lambda_k})| - |\bar{x}(t_0) - \hat{x}(t_0)| \\ &\qquad \leq \operatorname{sgn}(\bar{x}(t_0) - \hat{x}(t_0))h(t_{\lambda_k}) \end{aligned}$$

implying

$$f'(\bar{x})(h) \leq \max_{t \in M(\bar{x})} \operatorname{sgn}(\bar{x}(t) - \hat{x}(t))h(t).$$

(3.3) From Theorem 3.16 we obtain

$$f(x) \geq f(\bar{x}) + f'(\bar{x})(x - \bar{x}) \text{ for all } x \in X.$$

If $f'(\bar{x}) = 0_{X^*}$, then $\bar{x}$ is a minimal point of f on X. The converse statement follows from Theorem 3.17.

(3.4) $\partial f(0) = \{l \in \mathbb{R} \mid |l| \leq 1\} = [-1, 1]$.

(3.5) One proves for $l_1, l_2 \in \partial f(\bar{x})$ and $\lambda \in [0, 1]$ that $\lambda l_1 + (1 - \lambda) l_2 \in \partial f(\bar{x})$.

(3.6) Since

$$f(x) - f(\bar{x}) \geq \lim_{\lambda \to 0_+} \frac{1}{\lambda}\big(f(\bar{x} + \lambda(x - \bar{x})) - f(\bar{x})\big) = \nabla f(\bar{x})^T (x - \bar{x}),$$

we conclude $\nabla f(\bar{x}) \in \partial f(\bar{x})$. For an arbitrary $v \in \partial f(\bar{x})$ one gets for all unit vectors $e_1, \ldots, e_n \in \mathbb{R}^n$

$$v_i \leq \lim_{\lambda \to 0_+} \frac{1}{\lambda}\big(f(\bar{x} + \lambda e_i) - f(\bar{x})\big) = \frac{\partial f(\bar{x})}{\partial x_i}$$

and

$$v_i \geq \lim_{\lambda \to 0_-} \frac{1}{\lambda}\big(f(\bar{x} + \lambda e_i) - f(\bar{x})\big) = \frac{\partial f(\bar{x})}{\partial x_i}$$

which results in $v = \nabla f(\bar{x})$.

(3.7) The sub- and superdifferential can be chosen as

$$\underline{\partial} f(x_1, x_2) = \begin{cases} \{(\operatorname{sgn} x_1 |x_2|,\ \operatorname{sgn} x_2 |x_1|)\} & \text{if } x_1 x_2 \neq 0 \\ \{(u, 0) \mid |u| \leq |x_2|\} & \text{if } x_1 = 0 \\ \{(0, v) \mid |v| \leq |x_1|\} & \text{if } x_2 = 0 \end{cases}$$

and $\overline{\partial} f(x_1, x_2) = \{(0, 0)\}$. Then $Df(x_1, x_2) = (\underline{\partial} f(x_1, x_2), \overline{\partial} f(x_1, x_2))$ is a quasidifferential of f at (x_1, x_2).

(3.8) Since the directional derivative $f'(\bar{x})$ is given by

$$f'(\bar{x})(h) = |h_1| - |h_2| \text{ for all } h = (h_1, h_2) \in \mathbb{R}^2,$$

the function f is quasidifferentiable at $\bar{x}$.

(3.9) Since f is a convex function, the Clarke derivative coincides with the directional derivative $f'(\bar{x})(h)$. Then we obtain for all $h \in \mathbb{R}^n$

$$\begin{aligned} f'(\bar{x})(h) &= \lim_{\lambda \to 0_+} \frac{1}{\lambda} \left(\max_{1 \leq i \leq n} \{\bar{x}_i + \lambda h_i\} - \max_{1 \leq i \leq n} \{\bar{x}_i\} \right) \\ &= \lim_{\lambda \to 0_+} \frac{1}{\lambda} \max_{i \in I(\bar{x})} \{\bar{x}_i + \lambda h_i - \bar{x}_i\} \\ &= \max_{i \in I(\bar{x})} \{h_i\}. \end{aligned}$$

Chapter 4

(4.1) The inclusion $\text{int}(C) \subset \text{int}(C) + C$ is trivial, and the converse inclusion is simple to show.

(4.2) If f is sublinear, it is simple to show that the epigraph $E(f)$ is a cone. By Theorem 2.8 this cone is convex. For the proof of the converse implication one proves that

$$f(\lambda x) = \lambda f(x) \text{ for all } \lambda > 0$$

and $f(0) = 0$. The subadditivity can be simply shown.

(4.3) Take $x_1, x_2 \in \text{cone}(S)$, i.e. $x_1 = \lambda_1 s_1$ and $x_2 = \lambda_2 s_2$ for some $\lambda_1, \lambda_2 \geq 0$ and $s_1, s_2 \in S$. Without loss of generality we assume $\lambda_1 + \lambda_2 \neq 0$. Then

$$x_1 + x_2 = (\lambda_1 + \lambda_2)\left(\frac{\lambda_1}{\lambda_1 + \lambda_2} s_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2} s_2\right) \in \text{cone}(S).$$

(4.4) $\text{cone}(S) = \mathbb{R}^2_+$.

(4.5) $T(S, (1, 2)) = \{\lambda(a, -1) \mid \lambda \geq 0,\ a \in [-1, \frac{1}{2}]\}$.

(4.6) Simply apply Definition 4.6.

(4.7) (a) Apply Definition 4.6 and notice that $S_1 \subset S_2$.
(b) By part (a) one obtains $T(S_1 \cap S_2, \bar{x}) \subset T(S_1, \bar{x})$ and $T(S_1 \cap S_2, \bar{x}) \subset T(S_2, \bar{x})$ implying $T(S_1 \cap S_2, \bar{x}) \subset T(S_1, \bar{x}) \cap T(S_2, \bar{x})$.

(4.8) See the remark under 4. on page 31 in [202].

(4.9) Apply Theorem 2.4.5 in [71]. The assertion then follows from Proposition 2.2.1 in [71]. This result is also proved on the pages 17–18 in [309, Theorem 2E].

(4.10) f is <u>not</u> pseudoconvex at $\bar{x} = 0$.

Chapter 5

(5.1) Since $x \notin S = \text{cl}(S)$ and S is convex, the separation Theorem C.3 gives the desired result.

(5.2) See Lemmas 1.1 and 2.1 in [202].

(5.3) The Slater condition given in Lemma 5.9 is satisfied (take $\bar{x} := (\frac{1}{2}, \frac{1}{2})$).

(5.4) (a) Since $x_1, x_2 \geq 0$, it follows $x_1 + x_2 \geq 0$ for all feasible $(x_1, x_2) \in \mathbb{R}^2$.
(b) No.
(c) No.

(5.5) (a) $(2, 1)$.
(b) $(1.5, 2.25)$.
(c) $(\frac{185}{768}, \frac{55}{768}, -\frac{5}{16})$.

(5.6) Yes. For all feasible (x, y) it follows

$$\frac{x+3y+3}{2x+y+6} = \frac{1}{2} + \frac{\frac{5}{2}y}{2x+y+6} \geq \frac{1}{2}.$$

Since $\frac{\frac{5}{2}y}{2x+y+6} > 0$ for $y > 0$, we conclude that there are no other solutions.

(5.7) This problem satisfies the MFCQ. Then the Karush-Kuhn-Tucker conditions give the desired assertion.

(5.8) Choose an arbitrary $x \in S$, show $\nabla f(\bar{x})^T(x - \bar{x}) \geq 0$ and conclude that $\bar{x}$ is a minimal point of f on S.

(5.9) The function p with

$$p(t) = \frac{1}{3}\left(1 - e^{t-1}\right)\begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ for all } t \in [0, 1]$$

satisfies the adjoint equation (5.36) and the transversality condition (5.37). Then an optimal control $\bar{u} = (\bar{u}_1, \bar{u}_2)$ is given as

$$\bar{u}_1(t) = \left\{ \begin{array}{cl} \frac{5}{2} - \frac{3}{4(1-e^{t-1})} & \text{almost everywhere on } [0, 1 + \ln\frac{7}{10}] \\ 0 & \text{almost everywhere on } [1 + \ln\frac{7}{10}, 1] \end{array} \right\}$$

and

$$\bar{u}_2(t) = \left\{ \begin{array}{cl} \frac{23}{8} - \frac{3}{4(1-e^{t-1})} & \text{almost everywhere on } [0, 1 + \ln\frac{17}{23}] \\ 0 & \text{almost everywhere on } [1 + \ln\frac{17}{23}, 1] \end{array} \right\}.$$

Chapter 6

(6.1) The dual problem can be written as

$$\begin{aligned} \max_{u \geq 0} \inf_{x_1, x_2 \in \mathbb{R}} & \underbrace{x_1 + 2(x_2 - 1)^2 + u(-x_1 - x_2 + 1)}_{=(1-u)x_1 + 2(x_2-1)^2 + u(-x_2+1)} \\ &= \max_{u \geq 0} \begin{cases} -\frac{1}{8}, & \text{if } u = 1 \\ -\infty, & \text{if } u \neq 1 \end{cases} \\ &= -\frac{1}{8}. \end{aligned}$$

The maximal value $-\frac{1}{8}$ of the dual problem is attained at the maximal solution $\bar{u} := 1$. $\bar{x} := \left(-\frac{1}{4}, \frac{5}{4}\right)$ is a minimal solution of the primal problem with minimal value $-\frac{1}{8}$. There is no duality gap.

(6.2) (a) The dual problem reads

$$\max \int_0^1 u(t)\,dt$$

subject to the constraints

$$\int_0^t u(s)\,ds \leq t \text{ almost everywhere on } [0,1]$$
$$u(t) \geq 0 \text{ almost everywhere on } [0,1]$$
$$\int_0^1 u(t)\,dt \leq 2$$
$$u \in L_2[0,1].$$

(b) $(\alpha, x) = (1, 0_{L_2[0,1]})$ is a solution of the primal problem with the minimal value 2, and u with

$$u(t) = 1 \text{ almost everywhere on } [0,1]$$

is a solution of the dual problem with the maximal value 1.

(6.3) The solution reads $x_1 = 2(\sqrt{2} - 1) \approx 0.8284272$ with the minimal value $1 - x_1 = 3 - 2\sqrt{2} \approx 0.1715728$.

Chapter 7

(7.1) Take any sequence $(X_i)_{i \in \mathbb{N}}$ in $\mathcal{S}^n_+$ converging to some matrix $X \in \mathcal{S}^n$. The matrix X_i is symmetric and positive semidefinite for every $i \in \mathbb{N}$ and, therefore, all eigenvalues of X_i are nonnegative. Since the eigenvalues continuously depend on the entries of a matrix, we also obtain that the eigenvalues of the matrix X are nonnegative or $X \in \mathcal{S}^n_+$. Consequently, $\mathcal{S}^n_+$ is closed.

For the proof that $\mathcal{S}^n_+$ is also pointed take an arbitrary matrix $X \in \mathcal{S}^n_+ \cap (-\mathcal{S}^n_+)$. Then all eigenvalues of X are nonnegative and nonpositive, i.e. they equal 0. So, we get $X = 0_{\mathcal{S}^n}$.

(7.2) For $K := \mathbb{R}^n$ we obtain by Lemmas 7.4,(b) and 7.5,(b),(ii)

$$\mathcal{S}_+^n = (\mathcal{S}_+^n)^* = (C_{\mathbb{R}^n}^n)^* = H_{\mathbb{R}^n} = \text{convex hull}\,\{xx^T \mid x \in \mathbb{R}^n\}.$$

(7.3) We proceed as in the proof of Lemma 7.2. We have

$$\begin{aligned} X \in C_{\mathbb{R}^k \times K}^{k+l} &\iff 0 \le (x^T, y^T) \begin{pmatrix} A & B^T \\ B & C \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &\qquad \text{for all } x \in \mathbb{R}^k \text{ and all } y \in K \\ &\iff 0 \le y^T (C - BA^{-1}B^T) y \ \text{ for all } y \in K \\ &\iff C - BA^{-1}B^T \in C_K^l. \end{aligned}$$

(7.4) Since $\langle A, B \rangle = \text{trace}\,(AB)$, the implication "$\Leftarrow$" is obvious. For the proof of the converse implication assume that $\langle A, B \rangle = 0$ is fulfilled. With Exercise (7.2) we can write for some $p \in \mathbb{N}$

$$B = \sum_{i=1}^{p} x^{(i)} x^{(i)^T} \text{ for appropriate } x^{(1)}, \ldots, x^{(p)} \in \mathbb{R}^n.$$

Since $A \in \mathcal{S}_+^n$, we have $A = \sqrt{A}\sqrt{A}$ for a matrix $\sqrt{A} \in \mathcal{S}_+^n$. Then we obtain

$$\begin{aligned} 0 &= \langle A, B \rangle \\ &= \text{trace}(AB) \\ &= \text{trace}\left(\sqrt{A}\sqrt{A} \sum_{i=1}^{p} x^{(i)} x^{(i)^T}\right) \\ &= \sum_{i=1}^{p} \text{trace}\left(x^{(i)^T} \sqrt{A}\sqrt{A}\, x^{(i)}\right) \\ &= \sum_{i=1}^{p} \underbrace{\left(\sqrt{A}\, x^{(i)}\right)^T \left(\sqrt{A}\, x^{(i)}\right)}_{\ge 0} \end{aligned}$$

implying

$$\sqrt{A}\, x^{(i)} = 0_{\mathbb{R}^n} \text{ for all } i = 1, \ldots, p.$$

With this equation we get

$$\begin{aligned} AB &= \sqrt{A}\sqrt{A}\sum_{i=1}^{p} x^{(i)}x^{(i)T} \\ &= \sqrt{A}\sum_{i=1}^{p} \underbrace{\sqrt{A}\,x^{(i)}}_{=0_{\mathbb{R}^n}} x^{(i)T} \\ &= 0_{\mathcal{S}^n}. \end{aligned}$$

(7.5) Since A is positive semidefinite, we have

$$x^T Ax \geq 0 \text{ for all } x \in \mathbb{R}^n.$$

For $x = (0, \ldots, 0, x_i, \ldots, x_j, 0, \ldots, 0) \in \mathbb{R}^n$ with arbitrary $x_i, \ldots, x_j \in \mathbb{R}$ we then obtain

$$\begin{aligned} 0 &\leq x^T Ax \\ &= (0, \ldots, 0, x_i, \ldots, x_j, 0, \ldots, 0) \begin{pmatrix} \ddots & & \\ & A^{ij} & \\ & & \ddots \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ x_i \\ \vdots \\ x_j \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ &= (x_i, \ldots, x_j) A^{ij} \begin{pmatrix} x_i \\ \vdots \\ x_j \end{pmatrix}. \end{aligned}$$

Consequently, the block matrix A^{ij} is positive semidefinite.

(7.6) For the matrix $A := \begin{pmatrix} x_1 & 1 \\ 1 & x_2 \end{pmatrix}$ we obtain the eigenvalues

$$\lambda_{1/2} = \frac{x_1 + x_2}{2} \pm \sqrt{\frac{(x_1 + x_2)^2}{4} - x_1 x_2 + 1}$$

being nonnegative if and only if

$$x_1 x_2 \geq 1, \quad x_1 > 0, \quad x_2 > 0.$$

Therefore, the feasible set of this problem can be written as

$$\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 x_2 \geq 1, \quad x_1 > 0, \quad x_2 > 0\}.$$

It is obvious that the objective function has the lower bound 0 on this set but this value is not attained at a point of the constraint set.

(7.7) By Lemma 9.4,(a) we have

$$-\text{int}(\mathcal{S}^2_+) = \{X \in \mathcal{S}^2 \mid X \text{ is negative definite}\}.$$

For arbitrary $x_1, x_2 \in \mathbb{R}$ the eigenvalues of $G(x_1, x_2)$ are

$$\lambda_1 = \frac{x_1}{2} + \sqrt{\frac{x_1^2}{4} + x_2^2}$$

and

$$\lambda_2 = \frac{x_1}{2} - \sqrt{\frac{x_1^2}{4} + x_2^2}.$$

If $x_1 \leq 0$, then we get $\lambda_1 \geq 0$ and in the case of $x_1 > 0$ we have $\lambda_1 > 0$. Hence, there is no vector $(\hat{x}_1, \hat{x}_2) \in \mathbb{R}^2$ with $G(\hat{x}_1, \hat{x}_2) \in -\text{int}(\mathcal{S}^2_+)$, i.e. the generalized Slater condition is not satisfied.

(7.8) For an arbitrary $x \in \mathbb{R}^m$ we write

$$x_i = y_i - z_i \text{ for all } i \in \{1, \ldots, m\}$$

with $y_1, \ldots, y_m, z_1, \ldots, z_m \geq 0$. Then the primal problem can be written as

$$\begin{gathered} \min\ (c, -c)^T \begin{pmatrix} y \\ z \end{pmatrix} \\ \text{subject to the constraints} \\ B \preccurlyeq (A, -A) \begin{pmatrix} y \\ z \end{pmatrix} \\ y_1, \ldots, y_m, z_1, \ldots, z_m \geq 0. \end{gathered}$$

This problem has the form of the primal problem (7.21) and its dual is given by (7.23) as

$$\begin{gathered}\max \langle B, U\rangle \\ \text{subject to the constraints} \\ \langle A^{(1)}, U\rangle \leq c_1 \\ \vdots \\ \langle A^{(m)}, U\rangle \leq c_m \\ -\langle A^{(1)}, U\rangle \leq -c_1 \\ \vdots \\ -\langle A^{(m)}, U\rangle \leq -c_m \\ U \in C^*.\end{gathered}$$

This problem can be simplified to the dual problem

$$\begin{gathered}\max \langle B, U\rangle \\ \text{subject to the constraints} \\ \langle A^{(1)}, U\rangle = c_1 \\ \vdots \\ \langle A^{(m)}, U\rangle = c_m \\ U \in C^*.\end{gathered}$$

(7.9) The primal problem equals the primal problem in Exercise (7.8), if we set $c = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$ and

$$A(x) = A^{(1)}x_1 + A^{(2)}x_2$$

with

$$A^{(1)} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad A^{(2)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The eigenvalues of the matrix $B - A(x)$ are

$$\lambda_1 = -\frac{x_2}{2} + \sqrt{\frac{x_2^2}{4} + x_1^2},$$

$$\lambda_2 = -\frac{x_2}{2} - \sqrt{\frac{x_2^2}{4} + x_1^2}$$

and

$$\lambda_3 = -x_1 - 1.$$

$B - A(x)$ is negative semidefinite if and only if $\lambda_1, \lambda_2, \lambda_3 \leq 0$. These eigenvalues are nonpositive if and only if $x_1 = 0$ and $x_2 \geq 0$. So, the constraint set of the primal problem can be written as $\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = 0,\ x_2 \geq 0\}$ and, therefore, the extremal value of the primal problem equals 0.

With Exercise (7.8) the dual problem can be written in this special case as

$$\begin{gathered} \max -U_{33} \\ \text{subject to the constraints} \\ 2U_{12} + U_{33} = 1 \\ U_{22} = 0 \\ U \in \mathcal{S}^3_+ \end{gathered}$$

or equivalently

$$\begin{gathered} \max -U_{33} \\ \text{subject to the constraint} \\ \begin{pmatrix} U_{11} & \frac{1}{2}(1 - U_{33}) & U_{31} \\ \frac{1}{2}(1 - U_{33}) & 0 & U_{32} \\ U_{31} & U_{32} & U_{33} \end{pmatrix} \in \mathcal{S}^3_+. \end{gathered}$$

Since the matrix defining the constraint is positive semidefinite, by Exercise (7.5) the leading block matrices $U^{11} := (U_{11})$ and

$$U^{12} := \begin{pmatrix} U_{11} & \frac{1}{2}(1 - U_{33}) \\ \frac{1}{2}(1 - U_{33}) & 0 \end{pmatrix}$$

have to be positive semidefinite as well. The eigenvalues of the matrix U^{12} are

$$\lambda_{1/2} = \frac{U_{11}}{2} \pm \sqrt{\left(\frac{U_{11}}{2}\right)^2 + \frac{1}{4}(1 - U_{33})^2}.$$

They are nonnegative if and only if $U_{11} \geq 0$ and $U_{33} = 1$. Then the extremal value of the dual problem equals -1. So, the extremal values of the primal and dual problem do not coincide.

We consider an arbitrary $x \in \mathbb{R}^2$ for which the matrix $B - A(x)$ is negative semidefinite. Then one eigenvalue of this matrix equals 0. Therefore, $B - A(x)$ is not negative definite. Hence, the generalized Slater condition is not satisfied and Theorem 7.12 is not applicable.

Chapter 8

(8.1) The infimal value of the first problem is not attained because

$$\frac{1}{x_1^4} + \frac{1}{x_2^2} > 0 \text{ for all } x_1 \in \mathbb{N} \text{ and } x_2 \in \mathbb{R}$$

and

$$\inf_{x_1 \in \mathbb{N},\, x_2 \in \mathbb{R}} \frac{1}{x_1^4} + \frac{1}{x_2^2} = \underbrace{\inf_{x_1 \in \mathbb{N}} \frac{1}{x_1^4}}_{=0} + \underbrace{\inf_{x_2 \in \mathbb{R}} \frac{1}{x_2^2}}_{=0} = 0.$$

Hence, the first problem is not solvable.

For the second optimization problem we have

$$\inf_{x_1 \in \mathbb{N},\, x_2 \in \mathbb{R}} \frac{x_2^2}{x_1^4} = \underbrace{\inf_{x_1 \in \mathbb{N}} \frac{1}{x_1^4}}_{=0} \cdot \underbrace{\inf_{x_2 \in \mathbb{R}} x_2^2}_{=0} = 0.$$

This infimal value is attained at every vector $(x_1, 0)$ with arbitrary $x_1 \in \mathbb{N}$ and, therefore, the second optimization problem is solvable.

(8.2) Let for some $n \in \mathbb{N}$ the discrete set S_d be written as $S_d =: \left\{x_d^1, \ldots, x_d^n\right\}$ with $x_d^1, \ldots, x_d^n \in X_d$. Then we get

$$\min_{(x_d, x_c) \in S_d \times S_c} f(x_d, x_c) = \min_{1 \le i \le n} \left\{ \min_{x_c \in S_c} f(x_d^i, x_c) \right\}.$$

By Theorem 2.12 the interior optimization problems are solvable. Thus, the exterior min term exists.

(8.3) The functionals $\varphi_1, \ldots, \varphi_n$ are assumed to be continuous at $\bar{x}$ and, therefore, we get for an arbitrary $h \in X$

$$\min_{1 \le i \le n} \{\varphi_i(\bar{x} + \lambda h)\} = \min_{i \in I(\bar{x})} \{\varphi_i(\bar{x} + \lambda h)\}$$
$$\text{for all } \lambda > 0 \text{ being sufficiently close to } 0$$

because for $i \notin I(\bar{x})$ we have

$$\varphi_i(\bar{x}) > \min_{1 \le k \le n} \{\varphi_k(\bar{x})\}$$

implying

$$\varphi_i(\bar{x} + \lambda h) > \min_{1 \le k \le n} \{\varphi_k(\bar{x} + \lambda h)\}$$
$$\text{for all } \lambda > 0 \text{ being sufficiently close to } 0.$$

Then the directional derivative of f at $\bar{x}$ is given by

$$\begin{aligned}
&\left(\min_{1\le i\le n}\{\varphi_i(\cdot)\}\right)'(\bar{x})(h)\\
&= \lim_{\lambda\to 0_+}\frac{1}{\lambda}\left(\min_{i\in I(\bar{x})}\{\varphi_i(\bar{x}+\lambda h)\} - \min_{1\le i\le n}\{\varphi_i(\bar{x})\}\right)\\
&= \min_{i\in I(\bar{x})}\left\{\varphi_i'(\bar{x})(h)\right\} \text{ for all } h\in X.
\end{aligned}$$

(8.4) It is obvious that the vector $(\bar{x}_d, \bar{x}_c) := \big((1,1),(0,0)\big)$ is the unique minimal solution of this optimization problem. If we define the sets $S_d := \{1,\ldots,20\}$ and $S_c := \mathbb{R}^2$, the objective function $f : S_d \times S_c :\to \mathbb{R}$ with

$$f(x_d,x_c) = \big(1+(x_d)_1\big)^2\big(2+(x_d)_2\big)^4\big(1+(x_c)_1+(x_c)_2^3\big)$$
$$\text{for all } x_d\in S_d \text{ and } x_c\in S_c$$

and the constraint functions $g_1,\ g_2 : S_d\times S_c :\to \mathbb{R}$ with

$$g_1(x_d,x_c) = -\,(x_c)_1 \text{ for all } x_d\in S_d \text{ and } x_c\in S_c$$

and

$$g_2(x_d,x_c) = -\,(x_c)_2 \text{ for all } x_d\in S_d \text{ and } x_c\in S_c,$$

then we obtain the KKT conditions at the minimal solution

$$\begin{aligned}
&\nabla_{x_c} f\big((1,1),(0,0)\big) + u_1\nabla_{x_c} g_1\big((1,1),(0,0)\big)\\
&+u_2\nabla_{x_c} g_2\big((1,1),(0,0)\big) = \begin{pmatrix}0\\0\end{pmatrix}.
\end{aligned}$$

with nonnegative Lagrange multipliers u_1 and u_2. By a simple calculation we then get

$$u_1\begin{pmatrix}-1\\0\end{pmatrix} + u_2\begin{pmatrix}0\\-1\end{pmatrix} = \begin{pmatrix}-324\\0\end{pmatrix}$$

with the unique Lagrange multipliers $u_1 = 324$ and $u_2 = 0$. The remaining Lagrange multipliers are obtained in analogy where the number -324 in the right hand side of the linear system is replaced by another negative integer.

(8.5) It is obvious that the inequality constraint is redundant and the minimal value of the primal problem equals 1. With $x_d^1 := \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $x_d^2 := \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $x_d^3 := \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ the dual problem is given by Lemma 8.28 as

$$\max \inf_{i\in\{1,2,3\}} \left\{ \left(x_d^i\right)_1^2 + \ln\left(x_d^i\right)_2 - \ell^i e^{(x_d)_1^2+(x_d)_2^2} + \ell^i \right\}$$
$$\text{subject to the constraints}$$
$$1 - 3\ell^i \geq 0 \text{ for all } i \in \{1,2,3\}$$
$$\ell^i \geq 0 \text{ for all } i \in \{1,2,3\}.$$

The inequality constraints mean that $\ell^1, \ell^2, \ell^3 \in \left[0, \frac{1}{3}\right]$. Then the dual problem can be written as

$$\max_{\ell^1,\ell^2,\ell^3\in\left[0,\frac{1}{3}\right]} \min\left\{1 + \ell^1(1-e^2),\ 4 + \ell^2(1-e^5),\right.$$
$$\left. 9 + \ln 2 + \ell^3(1-e^{13})\right\} = 1.$$

The maximal value of the dual problem equals the minimal value of the primal problem. A possible maximal solution of the dual problem is $\left(\ell^1, \ell^2, \ell^3\right) = (0, 0, 0)$.

Chapter 9

(9.1) An optimal (feedback) control $\bar{u}$ is given by

$$\bar{u}(t) = \frac{12}{7e^{4(t-2)} + 9}\, x(t) \text{ almost everywhere on } [0, 2].$$

(9.2) An optimal (feedback) control $\bar{u}$ is given by

$$\bar{u}(t) = -\tanh(1-t)\, x(t) \text{ almost everywhere on } [0, 1].$$

(9.3) The equivalent system of linear differential equations of first order reads

$$\dot{x}(t) = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{pmatrix} x(t) + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} u(t)$$

almost everywhere on $[0, \hat{T}]$.

This system satisfies the Kalman condition.

(9.4) The eigenvalues of A are $\lambda_1 = \sqrt{\alpha}\, i,\ \lambda_2 = -\sqrt{\alpha}\, i,\ \lambda_3 = \lambda_4 = 0$. For every eigenvalue of A we get the implication

$$z^T (A - \lambda_j I, B) = 0^T_{\mathbb{R}^5} \implies z = 0_{\mathbb{R}^4}$$

resulting in

$$\text{Rank}\,(A - \lambda_j I, B) = 4 \text{ for } j = 1, \ldots, 4.$$

Hence, the Hautus condition is fulfilled.

(9.5) By Theorem 9.7 there is a time minimal control $\bar{u}$, and by Theorem 9.11 there is a vector $\eta \neq 0_{\mathbb{R}^4}$ with

$$\bar{u}(t) = \text{sgn}\left[\frac{\eta_1 \beta}{\sqrt{\alpha}} \sin t\sqrt{\alpha} - \eta_2 \beta \cos t\sqrt{\alpha} - \eta_3 \gamma t + \eta_4 \gamma\right]$$

almost everywhere on $[0, \bar{T}]$

($\bar{T}$ denotes the minimal time). Since the term in brackets has only finitely many zeros, the time minimal control $\bar{u}$ is unique.

Bibliography

1. J. Abadie, *Nonlinear programming* (North-Holland, Amsterdam, 1967).
2. J. Abadie, *Integer and nonlinear programming* (North-Holland, Amsterdam, 1970).
3. P.R. Adby and M.A. Dempster, *Introduction to optimization methods* (Chapman and Hall, London, 1974).
4. D. Alevras and M.W. Padberg, *Linear optimization and extensions - problems and solutions* (Springer, Berlin, 2001).
5. W. Alt, *Nichtlineare Optimierung - Eine Einführung in Theorie, Verfahren und Anwendungen* (Vieweg, 2002).
6. E.J. Anderson and P. Nash, *Linear programming in infinite-dimensional spaces* (Wiley, Chichester, 1987).
7. T.S. Angell and A. Kirsch, *Optimization methods in electromagnetic radiation* (Springer, New York, 2004).
8. Q.H. Ansari, E. Köbis and J.-C. Yao, *Vector variational inequalities and vector optimization* (Springer, Berlin, 2018).
9. J.S. Arora, *Introduction to optimum design* (McGraw-Hill, New York, 1989).
10. K.J. Arrow and L. Hurwicz, *Studies in linear and non-linear programming* (Stanford University Press, Stanford, 1958).
11. J.-P. Aubin, *Analyse convexe et ses applications* (Springer, Berlin, 1974).
12. J.-P. Aubin, *Applied abstract analysis* (Wiley, New York, 1977).
13. J.-P. Aubin, *Explicit methods of optimization* (gauthier-villars, Paris, 1984).
14. J.-P. Aubin, *Optima and equilibra - an introduction to nonlinear analysis* (Springer, Berlin, 2002).
15. J.-P. Aubin, P. Nepomiastchy and A.-M. Charles, *Méthodes explicites de l'optimisation* (Dunod, Paris, 1982).
16. A.V. Balakrishnan, *Introduction to optimization theory in a Hilbert space* (Springer, Berlin, 1971).
17. V.K. Balakrishnan, *Network optimization* (Chapman & Hall, New York, 1994).
18. R. Baldick, *Applied optimization - formulation and algorithms for engineering systems* (Cambridge University Press, 2006).
19. B. Bank, J. Guddat, D. Klatte, B. Kummer and K. Tammer, *Non-linear parametric optimization* (Birkhäuser, Basel, 1983).
20. V. Barbu, *Mathematical methods in optimization of differential systems* (Kluwer, Dordrecht, 1994).
21. V. Barbu and T. Precupanu, *Convexity and optimization in Banach spaces* (Editura Acad., Bucaresti, 1986).
22. J.F. Bard, *Practical bilevel optimization - algorithms and applications* (Springer, 1999).
23. M. Bartholomew-Biggs, *Nonlinear optimization with financial applications* (Springer, 2005).
24. M.S. Bazaraa and C.M. Shetty, *Foundations of optimization* (Springer, Berlin, 1976).

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3

25. M.S. Bazaraa and C.M. Shetty, *Nonlinear programming* (Wiley, New York, 1979).
26. E.M.L. Beale, *Introduction to optimization* (Wiley, Chichester, 1988).
27. E.J. Beltrami, *An algorithmic approach to nonlinear analysis and optimization* (Academic Press, New York, 1970).
28. M.P. Bendsøe and O. Sigmund, *Topology optimization - theory, methods, and applications* (Springer, Berlin, 2004).
29. A. Ben-Israel, A. Ben-Tal and S. Zlobec, *Optimality in nonlinear programming* (Wiley, New York, 1981).
30. H. Benker, *Mathematische Optimierung mit Computeralgebrasystemen - Einführung für Ingenieure, Naturwissenschaftler und Wirtschaftswissenschaftler unter Anwendung von Mathematica, Maple, Mathcad, Matlab und Excel* (Springer, 2003).
31. A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization - analysis, algorithms, and engineering applications* (SIAM, Philadelphia, 2001).
32. A. Berman, *Cones, matrices and mathematical programming* (Springer, Berlin, 1973).
33. D.P. Bertsekas, *Constrained optimization and Lagrange multiplier methods* (Athena Scientific, Belmont, 1996).
34. D.P. Bertsekas, *Nonlinear programming* (Athena Scientific, Belmont, 2004).
35. D.P. Bertsekas with A. Nedic and A.E. Ozdaglar, *Convex analysis and optimization* (Athena Scientific, Belmont, 2003).
36. J.T. Betts, *Practical methods for optimal control using nonlinear programming* (SIAM, 2001).
37. G.S. Beveridge and R.S. Schechter, *Optimization* (McGraw-Hill, London, 1970).
38. M.A. Bhatti, *Practical optimization methods - with Mathematica applications* (Springer, New York, 2000).
39. E. Blum and W. Oettli, *Mathematische Optimierung* (Springer, Berlin, 1975).
40. A.I. Bojarinov and V.V. Kafarov, *Optimierungsmethoden in der chemischen Technologie* (Verlag Chemie, Weinheim, 1972).
41. V.G. Boltjanskij, *Mathematische Methoden der optimalen Steuerung* (Hanser, München, 1972).
42. J.F. Bonnans, J.C. Gilbert, C. Lemaréchal and C.A. Sagastizábal, *Numerical optimization - theoretical and practical aspects* (Springer, Berlin, 2003).
43. J.F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems* (Springer, New York, 2000).
44. K.H. Borgwardt, *Optimierung, Operations Research, Spieltheorie - Mathematische Grundlagen* (Birkhäuser, 2001).
45. J.M. Borwein and A.S. Lewis, *Convex analysis and nonlinear optimization - theory and examples* (Springer, New York, 2006).
46. R.I. Boţ, S.-M. Grad, G. Wanka, *Duality in vector optimization* (Springer, Berlin, 2009).
47. M.J. Box, D. Davies and W.H. Swann, *Non-linear optimization techniques* (Oliver & Boyd, Edinburgh, 1969).
48. S. Boyd and L. Vandenberghe, *Convex optimization* (Cambridge University Press, Cambridge, 2004).
49. J. Bracken and G.P. McCormick, *Selected applications of nonlinear programming* (Wiley, New York, 1968).
50. S.J. Britvec, *Stability and optimization of flexible space structures* (Birkhäuser, Boston, 1995).
51. R.W. Brockett, *Finite dimensional linear systems* (Wiley, New York, 1970).
52. B. Brosowski, *Parametric semi-infinite optimization* (Lang, Frankfurt, 1982).
53. D. Bucur and G. Buttazzo, *Variational methods in shape optimization problems* (Birkhäuser, Boston, 2005).
54. R. Bulirsch, A. Miele, J. Stoer and K.H. Well, *Optimal control* (Birkhäuser, Basel, 1992).
55. B.D. Bunday, *Basic optimisation methods* (Arnold, London, 1985).
56. R.E. Burkard and U.T. Zimmermann, *Einführung in die Mathematische Optimierung* (Springer, Berlin, 2012).
57. E.R. Caianiello, *Functional analysis and optimization* (Academic Press, New York, 1966).

58. N. Cameron, *Introduction to linear and convex programming* (Cambridge University Press, Cambridge, 1985).
59. M.D. Canon, C.D. Cullum and E. Polak, *Theory of optimal control and mathematical programming* (McGraw-Hill, New York, 1970).
60. K.W. Cattermole, *Optimierung in der Nachrichtentechnik* (Verlag Chemie, Weinheim, 1990).
61. J. Cea, *Optimisation* (Dunod, Paris, 1971).
62. J. Cea, *Lectures on optimization* (Springer, Berlin, 1978).
63. Y. Censor and S.A. Zenios, *Parallel optimization - theory, algorithms, and applications* (Oxford University Press, New York, 1997).
64. L. Cesari, *Optimization, theory and applications* (Springer, New York, 1983).
65. G.-y. Chen, X. Huang and X. Yang, *Vector optimization - set-valued and variational analysis* (Springer, 2005).
66. E.K.P. Chong and S.H. Zak, *An introduction to optimization* (Wiley, 2001).
67. V. Chvátal, *Linear programming* (Freeman, New York, 1983).
68. P.G. Ciarlet, *Introduction to numerical linear algebra and optimisation* (Cambridge University Press, Cambridge, 1989).
69. S.J. Citron, *Elements of optimal control* (Holt, New York, 1969).
70. F.H. Clarke, *Methods of dynamic and nonsmooth optimization* (SIAM, Philadelphia, 1989).
71. F.H. Clarke, *Optimization and nonsmooth analysis* (SIAM, 1990).
72. F. Clarke, *Necessary conditions in dynamic optimization* (AMS, Providence, 2005).
73. T.F. Coleman, *Large sparse numerical optimization* (Springer, Berlin, 1984).
74. L. Collatz and W. Wetterling, *Optimierungsaufgaben* (Springer, Berlin, 1971).
75. L. Collatz and W. Wetterling, *Optimization problems* (Springer, Heidelberg, 1975).
76. Y. Collette and P. Siarry, *Multiobjective optimization - principles and case studies* (Springer, Berlin, 2004).
77. A.R. Conn, N.I.M. Gould and P.L. Toint, *Trust-region methods* (SIAM, Philadelphia, 2000).
78. G. Cornuejols and R. Tutuncu, *Optimization methods in finance* (Cambridge University Press, 2007).
79. B.D. Craven, *Mathematical programming and control theory* (Chapman and Hall, London, 1978).
80. B.D. Craven, *Fractional programming* (Heldermann, Berlin, 1988).
81. B.D. Craven, *Control and optimization* (Chapman & Hall, London, 1995).
82. R.F. Curtain and A.J. Pritchard, *Functional analysis in modern applied mathematics* (Academic Press, London, 1977).
83. R.F. Curtain and H.J. Zwart, *An introduction to infinite-dimensional linear systems theory* (Springer, 1995).
84. T.R. Cuthbert, *Optimization using personal computers* (Wiley, New York, 1987).
85. J.W. Daniel, *The approximate minimization of functionals* (Prentice-Hall, Englewood Cliffs, 1971).
86. R.W. Daniels, *An introduction to numerical methods and optimization techniques* (North-Holland, New York, 1978).
87. S. Danoe, *Nonlinear and dynamic programming* (Springer, Wien, 1975).
88. R.B. Darst, *Introduction to linear programming - applications and extensions* (Dekker, New York).
89. K. Deb, *Multi-objective optimization using evolutionary algorithms* (Wiley, Chichester, 2004).
90. D.G. de Figueiredo, *The Ekeland variational principle with applications and detours* (Springer, Berlin, 1989).
91. S. Dempe, *Foundations of bilevel programming* (Kluwer, Dordrecht, 2002).
92. S. Dempe, V. Kalashnikov, G.A. Pérez-Valdés and N. Kalashnykova, *Bilevel programming problems - algorithms and applications to energy networks* (Springer, Berlin, 2015).
93. V.F. Dem'yanov and A.M. Rubinov, *Approximate methods in optimization problems* (Elsevier, New York, 1970).
94. V.F. Demyanov and A.M. Rubinov, *Constructive nonsmooth analysis* (Lang, Frankfurt, 1995).

95. V.F. Demyanov, G.E. Stavroulakis, L.N. Ployakova and P.D. Panagiotopoulos, *Quasidifferentiability and nonsmooth modelling in mechanics, engineering and economics* (Kluwer, Dordrecht, 1996).
96. V.F. Dem'yanov and L.V. Vasil'ev, *Nondifferentiable optimization* (Optimization Software, New York, 1985).
97. M.M. Denn, *Optimization by variational methods* (McGraw-Hill, New York, 1969).
98. J.B. Dennis, *Mathematical programming and electrical networks* (MIT Press, Cambridge, 1959).
99. J.E. Dennis and R.B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations* (SIAM, Philadelphia, 1996).
100. W. Dinkelbach, *Sensitivitätsanalysen und parametrische Programmierung* (Springer, Berlin, 1969).
101. W. Dinkelbach, *Entscheidungsmodelle* (de Gruyter, Berlin, 1982).
102. S. Dischinger, *Pivotauswahlverfahren in der Linearen Programmierung* (Lang, Frankfurt, 1995).
103. U. Diwekar, *Introduction to applied optimization* (Springer, 2003).
104. L.C. Dixon, *Numerical optimisation of dynamic systems* (North- Holland, Amsterdam, 1980).
105. A. Dontchev and T. Zolezzi, *Well-posed optimization problems* (Springer, Berlin, 1993).
106. D.-Z. Du, P.M. Pardalos and W. Wu, *Mathematical theory of optimization* (Springer, 2001).
107. R.J. Duffin, E.L. Peterson and C. Zener, *Geometric programming, theory and application* (Wiley, New York, 1967).
108. P. Dyer and S.R. McReynolds, *The computation and theory of optimal control* (Academic Press, New York, 1970).
109. T.F. Edgar and D.M. Himmelblau, *Optimization of chemical processes* (McGraw-Hill, New York, 1988).
110. M. Ehrgott, *Multicriteria optimization* (Springer, Berlin, 2005).
111. G. Eichfelder, *Adaptive scalarization methods in multiobjective optimization* (Springer, Berlin, 2008).
112. G. Eichfelder, *Variable ordering structures in vector optimization* (Springer, Berlin, 2014).
113. H.A. Eiselt, G. Pederzoli and C.-L. Sandblom, *Continuous optimization models* (de Gruyter, Berlin, 1987).
114. I. Ekeland and R. Témam, *Convex analysis and variational problems* (SIAM, 1999).
115. I. Ekeland and T. Turnbull, *Infinite-dimensional optimization and convexity* (University of Chicago Press, Chicago, 1983).
116. K.-H. Elster, *Nichtlineare Optimierung* (Teubner, Leipzig, 1978).
117. K.-H. Elster, R. Reinhardt, M. Schäuble and G. Donath, *Einführung in die nichtlineare Optimierung* (Teubner, Leipzig, 1977).
118. V.W. Eveleigh, *Adaptive control and optimization techniques* (McGraw-Hill, New York, 1967).
119. B.S. Everitt, *Introduction to optimization methods and their application in statistics* (Chapman and Hall, London, 1987).
120. J.G. Evtusenko, *Numerical optimization techniques* (Optimization Software, New York, 1985).
121. H.O. Fattorini, *Infinite dimensional optimization and control theory* (Cambridge University Press, Cambridge, 1999).
122. A.V. Fiacco and G.P. McCormick, *Nonlinear programming* (Wiley, New York, 1968).
123. R. Fletcher, *Practical methods of optimization* (Wiley, Chichester, 2003).
124. M. Florenzano and C. Le Van, *Finite dimensional convexity and optimization* (Springer, 2001).
125. C.A. Floudas, *Deterministic global optimization - theory, methods and applications* (Kluwer, Dordrecht, 2000).
126. C.A. Floudas and P.M. Pardalos, *A collection of test problems for constrained global optimization algorithms* (Springer, Berlin, 1990).
127. L.R. Foulds, *Optimization techniques* (Springer, New York, 1981).

128. R.L. Fox, *Optimization methods for engineering design* (Addison-Wesley, Reading, 1971).
129. W. Frank, *Mathematische Grundlagen der Optimierung* (Oldenbourg, München, 1969).
130. H. Fränkel, *Diskrete optimale Steuerungsprobleme und konvexe Optimierung* (de Gruyter, Berlin, 1971).
131. A. Fromm, *Nichtlineare Optimierungsmodelle* (Deutsch, Frankfurt, 1975).
132. R.H. Gallagher, *Optimum structural design* (Wiley, Chichester, 1977).
133. C. Geiger and C. Kanzow, *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben* (Springer, 1999).
134. C. Geiger and C. Kanzow, *Theorie und Numerik restringierter Optimierungsaufgaben* (Springer, 2002).
135. A.M. Geoffrion, *Perspectives on optimization* (Addison-Wesley, Reading, 1972).
136. M. Gerdts and F. Lempio, *Mathematische Optimierungsverfahren des Operations Research* (De Gruyter, Berlin, 2011).
137. P. Gessner and K. Spremann, *Optimierung in Funktionenräumen* (Springer, Berlin, 1972).
138. F. Giannessi, *Constrained optimization and image space analysis - volume 1: separation of sets and optimality conditions* (Springer, 2005).
139. P.E. Gill, *Numerical methods for constrained optimization* (Academic Press, London, 1974).
140. P.E. Gill, W. Murray and M.H. Wright, *Practical optimization* (Academic Press, London, 1981).
141. G. Giorgi, A. Guerraggio and J. Thierfelder, *Mathematics of optimization: smooth and nonsmooth case* (Elsevier, 2004).
142. I.V. Girsanov, *Lectures on mathematical theory of extremum problems* (Springer, Berlin, 1972).
143. K. Glashoff and S.-A. Gustafson, *Linear optimization and approximation* (Springer, New York, 1983).
144. F. Glover, N. Phillips and D. Klingman, *Network models in optimization and their applications in practice* (Wiley, New York, 1992).
145. A. Göpfert, *Mathematische Optimierung in allgemeinen Vektorräumen* (Teubner, Leipzig, 1973).
146. A. Göpfert, L. Bittner, K.-H. Elster, F. Nožička, J. Piehler and R. Tichatschke (eds.), *Lexikon der Optimierung* (Akademie-Verlag, Berlin, 1986).
147. A. Göpfert and R. Nehse, *Vektoroptimierung - Theorie, Verfahren und Anwendungen* (Teubner, Leipzig, 1990).
148. A. Göpfert, H. Riahi, C. Tammer and C. Zalinescu, *Variational methods in partially ordered spaces* (Springer, 2003).
149. B.S. Gottfried and J. Weismann, *Introduction to optimization theory* (Prentice-Hall, Englewood Cliffs, 1973).
150. S.-M. Grad, *Vector Optimization and Monotone Operators via Convex Duality* (Springer, Berlin, 2015).
151. C. Grossmann and A.A. Kaplan, *Strafmethoden und modifizierte Lagrangefunktionen in der nichtlinearen Optimierung* (Teubner, Leipzig, 1979).
152. C. Grossmann and H. Kleinmichel, *Verfahren der nichtlinearen Optimierung* (Teubner, Leipzig, 1976).
153. C. Großmann and J. Terno, *Numerik der Optimierung* (Teubner, Stuttgart, 1993).
154. W.A. Gruver and E. Sachs, *Algorithmic methods in optimal control* (Pitman, Boston, 1980).
155. J. Guddat, F.G. Vasquez and H.Th. Jongen, *Parametric optimization: singularities, pathfollowing and jumps* (Wiley, Chichester, 1990).
156. J. Guddat, F.G. Vasquez, K. Tammer and K. Wendler, *Multiobjective and stochastic optimization based on parametric optimization* (Akademie-Verlag, Berlin, 1985).
157. I. Gumowski and C. Mira, *Optimization in control theory and practice* (Cambridge University Press, Cambridge, 1968).
158. G. Hadley, *Nichtlineare und dynamische Programmierung* (Physica-Verlag, Würzburg, 1969).
159. H.W. Hamacher, *Mathematische Lösungsverfahren für planare Standortprobleme* (Vieweg, Wiesbaden, 1995).

160. H.W. Hamacher and K. Klamroth, *Lineare und Netzwerk-Optimierung - ein bilinguales Lehrbuch* (Vieweg, Braunschweig, 2000).
161. R. Hartley, *Linear and nonlinear programming* (Harwood, Chichester, 1985).
162. L. Hasdorff, *Gradient optimization and nonlinear control* (Wiley, New York, 1976).
163. J. Haslinger and R.A.E. Mäkinen, *Introduction to shape optimization - theory, approximation, and computation* (SIAM, Philadelphia, 2003).
164. S. Helbig, *Parametrische semi-infinite Optimierung in totalgeordneten Gruppen* (Fischer, Frankfurt, 1987).
165. U. Helmke and J.B. Moore, *Optimization and dynamical systems* (Springer, Berlin, 1994).
166. W.S. Hemp, *Optimum structures* (Clarendon Press, Oxford, 1973).
167. M.R. Hestenes, *Optimization theory* (Wiley, New York, 1975).
168. M.R. Hestenes, *Conjugate direction methods in optimization* (Springer, Berlin, 1980).
169. R. Hettich and P. Zencke, *Numerische Methoden der Approximation und semi-infiniten Optimierung* (Teubner, Stuttgart, 1982).
170. C. Hillermeier, *Nonlinear multiobjective optimization - a generalized homotopy approach* (Birkhäuser, 2001).
171. D.M. Himmelblau, *Applied nonlinear programming* (McGraw-Hill, New York, 1972).
172. J.-B. Hiriart-Urruty and C. Lemarechal, *Convex analysis and minimization algorithms, Part 1 and 2* (Springer, New York, 1993).
173. W. Hock and K. Schittkowski, *Test examples for nonlinear programming codes* (Springer, Berlin, 1981).
174. E. Hofer and R. Lunderstädt, *Numerische Methoden der Optimierung* (Oldenbourg, München, 1975).
175. U. Hoffmann and H. Hofmann, *Einführung in die Optimierung* (Verlag Chemie, Weinheim, 1971).
176. R.B. Holmes, *A course on optimization and best approximation* (Springer, Berlin, 1972).
177. R.B. Holmes, *Geometric functional analysis and its applications* (Springer, New York, 1975).
178. C.S. Hong and Z. Quan, *Integral global optimization* (Springer, Berlin, 1988).
179. R. Horst, *Nichtlineare Optimierung* (Hanser, München, 1979).
180. R. Horst, P.M. Pardalos and N. V. Thoai, *Introduction to global optimization* (Springer, 2001).
181. R. Horst and H. Tuy, *Global optimization - deterministic approaches* (Springer, 1996).
182. L. Huang, *Second-order directional derivatives in nonsmooth optimization* (Chinese University Press, Hong Kong, 1995).
183. P. Hupfer, *Optimierung von Baukonstruktionen* (Teubner, Stuttgart, 1970).
184. M.D. Intriligator, *Mathematical optimization and economic theory* (SIAM, 2002).
185. A.D. Ioffe and V.M. Tichomirov, *Theorie der Extremalaufgaben* (Deutscher Verlag der Wissenschaften, Berlin, 1979).
186. G. Isac, *Complementarity problems* (Springer, 1992).
187. G. Isac and V. Postolică, *The best approximation and optimization in locally convex spaces* (Lang, Frankfurt, 1993).
188. S.L. Jacoby, J.S. Kowalik and J.T. Pizzo, *Iterative methods for nonlinear optimization problems* (Prentice-Hall, Englewood Cliffs, 1972).
189. J. Jahn, *Mathematical vector optimization in partially ordered linear spaces* (Lang, Frankfurt, 1986).
190. J. Jahn, *Vector optimization - theory, applications, and extensions* (Springer, Berlin, 2011).
191. B. Jansen, *Interior point techniques in optimization - complementarity, sensitivity and algorithms* (Kluwer, Dordrecht, 1997).
192. F. Jarre and J. Stoer, *Optimierung* (Springer, 2004).
193. M.W. Jeter, *Mathematical programming* (Dekker, New York, 1986).
194. H.T. Jongen, P. Jonker and F. Twilt, *Nonlinear optimization in finite dimensions - Morse theory, Chebyshev approximation, transversality, flows, parametric aspects* (Springer, 2001).
195. H.T. Jongen, K. Meer and E. Triesch, *Optimization theory* (Springer, 2004).
196. I. Kaliszewski, *Quantitative Pareto analysis by cone separation technique* (Kluwer, Boston, 1994).

197. E.L. Kaplan, *Mathematical programming and games* (Wiley, New York, 1982).
198. M.H. Karwan, *Redundancy in mathematical programming* (Springer, Berlin, 1983).
199. C.T. Kelley, *Iterative methods for linear and nonlinear equations* (SIAM, Philadelphia, 1995).
200. C.T. Kelley, *Iterative methods for optimization* (SIAM, Philadelphia, 1999).
201. A.A. Khan, C. Tammer, C. Zălinescu, *Set-valued optimization* (Springer, Berlin, 2015).
202. A. Kirsch, W. Warth and J. Werner, *Notwendige Optimalitätsbedingungen und ihre Anwendung* (Springer, Berlin, 1978).
203. K.-P. Kistner, *Optimierungsmethoden* (Physica, Heidelberg, 1988).
204. K.C. Kiwiel, *Methods of descent for nondifferentiable optimization* (Springer, Berlin, 1985).
205. D. Klatte and B. Kummer, *Nonsmooth equations in optimization - regularity, calculus, methods and applications* (Springer, 2002).
206. B. Kolman and R.E. Beck, *Elementary linear programming with applications* (Academic Press, 1995).
207. D. Koo, *Elements of optimization* (Springer, Berlin, 1977).
208. P. Kosmol, *Methoden zur numerischen Behandlung nichtlinearer Gleichungen und Optimierungsaufgaben* (Teubner, Stuttgart, 1989).
209. P. Kosmol, *Optimierung und Approximation* (de Gruyter, Berlin, 1991).
210. J.S. Kowalik and M.R. Osborne, *Methods for unconstrained optimization problems* (Elsevier, New York, 1968).
211. W. Krabs, *Optimierung und Approximation* (Teubner, Stuttgart, 1975).
212. W. Krabs, *Einführung in die Kontrolltheorie* (Wissenschaftliche Buchgesellschaft, Darmstadt, 1978).
213. W. Krabs, *Optimization and approximation* (Wiley, Chichester, 1979).
214. W. Krabs, *Einführung in die lineare und nichtlineare Optimierung für Ingenieure* (Teubner, Stuttgart, 1983).
215. W. Krabs, *On moment theory and controllability of one-dimensional vibrating systems and heating processes* (Springer, Berlin, 1992).
216. W. Krabs, *Optimal control of undamped linear vibrations* (Heldermann, Lemgo, 1995).
217. W. Krabs and S.T. Pickl, *Analysis, controllability and optimization of time-discrete systems and dynamical games* (Springer, Berlin, 2003).
218. B. Kreko, *Optimierung* (Deutscher Verlag der Wissenschaften, Berlin, 1974).
219. H.P. Kuenzi and W. Krelle, *Nichtlineare Programmierung* (Springer, Berlin, 1962).
220. H.P. Kuenzi, W. Krelle and R. von Randow, *Nichtlineare Programmierung* (Springer, Berlin, 1979).
221. H.P. Kuenzi and W. Oettli, *Nichtlineare Optimierung* (Springer, Berlin, 1969).
222. H.P. Kuenzi, H.G. Tzschach and C.A. Zehnder, *Numerische Methoden der mathematischen Optimierung mit ALGOL- und FORTRAN-Programmen* (Teubner, Stuttgart, 1967).
223. H.P. Kuenzi, H.G. Tzschach and C.A. Zehnder, *Numerical methods of mathematical optimization* (Academic Press, New York, 1971).
224. J.L. Kuester and J.H. Mize, *Optimization techniques with Fortran* (McGraw-Hill, New York, 1973).
225. A.H. Land and S. Powell, *Fortran codes for mathematical programming* (Wiley, London, 1973).
226. K. Lange, *Optimization* (Springer, 2004).
227. E. Laporte and P. Le Tallec, *Numerical methods in sensitivity analysis and shape optimization* (Birkhäuser, 2002).
228. L.S. Lasdon, *Optimization theory for large systems* (Macmillan, London, 1970).
229. P.-J. Laurent, *Approximation et optimisation* (Herman, Paris, 1972).
230. E.B. Lee and L. Markus, *Foundations of optimal control theory* (Wiley, New York, 1967).
231. G. Leitmann, *Topics in optimization* (Academic Press, New York, 1967).
232. G. Leitmann, *Einführung in die Theorie optimaler Steuerung und der Differentialspiele* (Oldenbourg, München, 1974).

233. E.S. Levitin, *Perturbation theory in mathematical programming and its applications* (Wiley, Chichester, 1994).
234. K. Littger, *Optimierung* (Springer, Berlin, 1992).
235. A. Löhne, *Vector optimization with infimum and supremum* (Springer, Berlin, 2011).
236. G. Lorenzen, *Parametrische Optimierungen und einige Anwendungen* (Oldenbourg, München, 1974).
237. D.T. Luc, *Theory of vector optimization* (Springer, Berlin, 1989).
238. D.T. Luc, *Multiobjective linear programming - an introduction* (Springer, Berlin, 2016).
239. D.G. Luenberger, *Optimization by vector space methods* (Wiley, New York, 1998).
240. D.G. Luenberger, *Introduction to linear and nonlinear programming* (Addison-Wesley, Reading, 1973).
241. D.G. Luenberger, *Linear and nonlinear programming* (Addison-Wesley, Reading, 1984).
242. K. Malanowski, *Stability of solutions to convex problems of optimization* (Springer, Berlin, 1987).
243. O.L. Mangasarian, *Nonlinear programming* (SIAM, 1994).
244. K. Marti, *Descent directions and efficient solutions in discretely distributed stochastic programs* (Springer, Berlin, 1988).
245. K. Marti, *Stochastic optimization methods* (Springer, Berlin, 2005).
246. K. Marti and D. Gröger, *Einführung in die lineare und nichtlineare Optimierung* (Physica, 2000).
247. B. Martos, *Nonlinear programming theory and methods* (North-Holland, Amsterdam, 1975).
248. A. Marzollo, *Periodic optimization* (Springer, Wien, 1972).
249. A. Marzollo, *Controllability and optimization* (Springer, Wien, 1972).
250. G.P. McCormick, *Nonlinear programming* (Wiley, New York, 1983).
251. C. McMillan, *Mathematical programming* (Wiley, New York, 1970).
252. C.W. Merriam, *Optimization theory and the design of feedback control systems* (McGraw-Hill, New York, 1964).
253. K. Miettinen, *Nonlinear multiobjective optimization* (Kluwer, Boston, 2004).
254. R.E. Miller, *Optimization - foundations and applications* (Wiley, New York, 2000).
255. M. Minoux, *Programmation mathématique* (Dunod, Paris, 1983).
256. M. Minoux, *Mathematical programming* (Wiley, Chichester, 1986).
257. K.V. Mital, *Optimization methods in operations research and systems analysis* (Wiley, New Delhi, 1976).
258. K.V. Mital, *Optimization methods* (Wiley, New Delhi, 1977).
259. G. Mitra, *Theory and application of mathematical programming* (Academic Press, London, 1976).
260. J. Mockus, *Bayesian approach to global optimization - theory and applications* (Kluwer, Dordrecht, 1989).
261. B.S. Mordukhovich, *Variational analysis and generalized differentiation I - basic theory* (Springer, Berlin, 2006).
262. B.S. Mordukhovich, *Variational analysis and generalized differentiation II - applications* (Springer, Berlin, 2006).
263. J.J. Moré and S.J. Wright, *Optimization software guide* (SIAM, Philadelphia, 1993).
264. I.H. Mufti, *Computational methods in optimal control problems* (Springer, Berlin, 1970).
265. W.A. Murray, *Numerical methods for unconstrained optimization* (Academic Press, London, 1972).
266. K.G. Murty, *Linear Programming* (Wiley, New York, 1983).
267. H. Nakayama and T. Tanino, *Theory and applications of multiobjective programming (in Japanese)* (Society of Instrument and Control Engineers, Tokyo, 1994).
268. H. Nakayama, Y. Yun, M. Yoon, *Sequential approximate multiobjective optimization using computational intelligence* (Springer, Berlin, 2009).
269. J.L. Nazareth, *Differentiable optimization and equation solving - a treatise on algorithmic science and the Karmarkar revolution* (Springer, New York, 2003).
270. J.L. Nazareth, *An optimization primer - on models, algorithms, and duality* (Springer, 2004).

271. Y. Nesterov, *Introductory lectures on convex optimization - a basic course* (Springer, 2004).
272. Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming* (SIAM, Philadelphia, 1994).
273. L.W. Neustadt, *Optimization* (Princeton University Press, Princeton, 1976).
274. J. Nocedal and S.J. Wright, *Numerical optimization* (Springer, New York, 1999).
275. T. Okabe, *Evolutionary multi-objective optimization - on the distribution of offspring in parameter and fitness space* (Shaker, Aachen, 2004).
276. G.M. Ostrovskij and J.M. Volin, *Methoden zur Optimierung chemischer Reaktoren* (Akademie Verlag, Berlin, 1973).
277. A. Osyczka, *Evolutionary algorithms for single and multicriteria design optimization* (Physica, Heidelberg, 2002).
278. J. Outrata, M. Kocvara and J. Zowe, *Nonsmooth approach to optimization problems with equilibrium constraints - theory, applications and numerical results* (Kluwer, Dordrecht, 1998).
279. M. Padberg, *Linear optimization and extensions* (Springer, Berlin, 1999).
280. D. Pallaschke and S. Rolewicz, *Foundations of mathematical optimization - convex analysis without linearity* (Kluwer, Dordrecht, 1997).
281. M.J. Panik, *Classical optimization* (North-Holland, Amsterdam, 1976).
282. M. Papageorgiou, *Optimierung* (Oldenbourg, München, 1991).
283. P.M. Pardalos and J.B. Rosen, *Constrained global optimization* (Springer, Berlin, 1987).
284. P. Pedregal, *Introduction to optimization* (Springer, New York, 2004).
285. A.L. Peressini, F.E. Sullivan and J.J. Uhl, *The mathematics of nonlinear programming* (Springer, New York, 1988).
286. G.C. Pflug, *Optimization of stochastic models - the interface between simulation and optimization* (Kluwer, Boston, 1996).
287. R.R. Phelps, *Convex functions, monotone operators and differentiability* (Springer, Berlin, 1993).
288. D.A. Pierre, *Optimization theory with applications* (Wiley, New York, 1969).
289. D.A. Pierre and M.J. Lowe, *Mathematical programming via augmented Lagrangians* (Addison-Wesley, Reading, 1975).
290. E. Polak, *Computational methods in optimization* (Academic Press, New York, 1971).
291. E. Polak, *Optimization - algorithms and consistent approximations* (Springer, New York, 1997).
292. J.P. Ponstein, *Approaches to the theory of optimization* (Cambridge University Press, 2004).
293. W. Prager, *Introduction to structural optimization* (Springer, Wien, 1972).
294. B.N. Pšeničnyj, *Notwendige Optimalitätsbedingungen* (Oldenbourg, München, 1972).
295. B.N. Pšeničnyj and J.M. Danilin, *Numerische Methoden für Extremalaufgaben* (Deutscher Verlag der Wissenschaften, Berlin, 1982).
296. B.N. Pshenichnyj, *The linearization method for constraint optimization* (Springer, Berlin, 1994).
297. L. Pun, *Introduction to optimization practice* (Wiley, New York, 1969).
298. J. Ramík and M. Vlach, *Generalized concavity in fuzzy optimization and decision analysis* (Springer, 2001).
299. S.S. Rao, *Optimization* (Wiley, New Delhi, 1978).
300. T. Rapcsák, *Smooth nonlinear optimization in* $\mathbb{R}^n$ (Kluwer, Dordrecht, 1997).
301. H. Ratschek and J. Rokne, *New computer methods for global optimization* (Ellis Horwood Limited, Chichester, 1988).
302. B.S. Razumikhin, *Physical models and equilibrium methods in programming and economics* (Reidel, Dordrecht, 1984).
303. G.V. Reklaitis, K.M. Ravindran and A. Raysdell, *Engineering optimization* (Wiley, New York, 1983).
304. C. Richter, *Optimierungsverfahren und BASIC-Programme* (Akademie-Verlag, Berlin, 1988).
305. C. Richter, K. Schittkowski and R. Hettich, *Sequential quadratic programming - theory and applications* (Akademie-Verlag, Berlin, 1993).

306. K.-J. Richter, *Methoden der Optimierung* (Fachbuchverlag, Leipzig, 1970).
307. R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, 1970).
308. R.T. Rockafellar, *Conjugate duality and optimization* (SIAM, Philadelphia, 1974).
309. R.T. Rockafellar, *The theory of subgradients and its applications to problems of optimization: convex and nonconvex functions* (Heldermann, Berlin, 1981).
310. R.T. Rockafellar, *Network flows and monotropic optimization* (Wiley, New York, 1984).
311. R.T. Rockafellar and R.J.-B. Wets, *Variational analysis* (Springer, 2004).
312. C. Roos, T. Terlaky and J.-Ph. Vial, *Theory and algorithms for linear optimization - an interior point approach* (Wiley, Chichester, 1997).
313. C. Roos, T. Terlaky and J.-P. Vial, *Interior point methods for linear optimization* (Springer, 2006).
314. A. Rubinov, *Abstract convexity and global optimization* (Springer, 2000).
315. A. Rubinov and X.-q. Yang, *Lagrange-type functions in constrained non-convex optimization* (Springer, 2003).
316. D.L. Russell, *Optimization theory* (Benjamin, New York, 1970).
317. B. Rustem, *Algorithms for nonlinear programming and multiple-objective decisions* (Wiley, Chichester, 1998).
318. M. Sakawa, *Fuzzy sets and interactive multiobjective optimization* (Plenum, New York, 1993).
319. H.-J. Sander, *Dualität bei Optimierungsaufgaben* (Oldenbourg, München, 1973).
320. Y. Sawaragi, H. Nakayama and T. Tanino, *Theory of multiobjective optimization* (Academic Press, Orlando, 1985).
321. S. Schäffler, *Global optimization - a stochastic approach* (Springer, New York, 2012).
322. W. Schirotzek, *Differenzierbare Extremalprobleme* (Teubner, Leipzig, 1989).
323. K. Schittkowski, *More test examples for nonlinear programming codes* (Springer, Berlin, 1987).
324. R. Schmidt, *Advances in nonlinear parameter optimization* (Springer, Berlin, 1982).
325. E. Seiffart and K. Manteuffel, *Lineare Optimierung* (Teubner, 1991).
326. J.F. Shapiro, *Mathematical programming* (Wiley, New York, 1979).
327. H.D. Sherali, *Optimization with disjunctive constraints* (Springer, Berlin, 1980).
328. N.Z. Shor, *Minimization methods for non-differentiable functions* (Springer, Berlin, 1985).
329. N.Z. Shor, *Nondifferentiable optimization and polynomial problems* (Springer, 1998).
330. G. Sierksma, *Linear and integer programming - theory and practice* (Dekker, New York, 1996).
331. I. Singer, *Duality for nonconvex approximation and optimization* (Springer, New York, 2006).
332. S.M. Sinha, *Mathematical programming - theory and methods* (Elsevier, 2006).
333. M. Sniedovich, *Dynamic programming - foundations and principles* (CRC Press, Boca Raton, 2010).
334. J.A. Snyman, *Practical mathematical optimization - an introduction to basic optimization theory and classical and new gradient-based algorithms* (Springer, 2005).
335. P. Spellucci, *Numerische Verfahren der nichtlinearen Optimierung* (Birkhäuser, Basel, 1993).
336. K. Stahl and N. Schulz, *Mathematische Optimierung und mikro-ökonomische Theorie* (Springer, Berlin, 1981).
337. O. Stein, *Grundzüge der Nichtlinearen Optimierung* (Springer, Berlin, 2018).
338. O. Stein, *Grundzüge der Globalen Optimierung* (Springer, Berlin, 2018).
339. J. Stoer and C. Witzgall, *Convexity and optimization in finite dimensions* (Springer, Berlin, 1970).
340. R.G. Strongin and Y.D. Sergeyev, *Global optimization with non-convex constraints - sequential and parallel algorithms* (Springer, 2000).
341. W. Sun and Y.-x. Yuan, *Optimization theory and methods - nonlinear programming* (Springer, 2006).
342. R.K. Sundaram *A first course in optimization theory* (Cambridge University Press, 1996).
343. D. Sworder, *Optimal adaptive control systems* (Academic Press, New York, 1966).

344. M. Tawarmalani and N.V. Sahinidis, *Convexification and global optimization in continuous and mixed-integer nonlinear programming - theory, algorithms, software, and applications* (Springer, 2003).
345. V.M. Tichomirov, *Fundamental principles of the theory of extremal problems* (Wiley, Chichester, 1986).
346. H. Tolle, *Optimierungsverfahren für Variationsaufgaben mit gewöhnlichen Differentialgleichungen als Nebenbedingungen* (Springer, Berlin, 1971).
347. H. Tolle, *Optimization methods* (Springer, Berlin, 1975).
348. A. Törn and A. Žilinskas, *Global optimization* (Springer, Berlin, 1989).
349. F. Tröltzsch, *Optimality conditions for parabolic control problems and applications* (Teubner, Leipzig, 1984).
350. F. Tröltzsch, *Optimal control of partial differential equations - theory, methods and applications* (AMS, Providence, 2010).
351. J.L. Troutman, *Variational calculus and optimal control - optimization with elementary convexity* (Springer, Berlin, 1995).
352. V. Tsurkov, *Large-scale optimization - problems and methods* (Springer, 2001).
353. H. Tuy, *Convex analysis and global optimization* (Kluwer, Dordrecht, 1998).
354. M. Ulbrich and S. Ulbrich, *Nichtlineare Optimierung* (Springer, Basel, 2012).
355. R.J. Vanderbei, *Linear programming - foundations and extensions* (Kluwer, Boston, 1996).
356. G.N. Vanderplaats, *Numerical optimization techniques for engineering design* (McGraw-Hill, New York, 1984).
357. J. Varga, *Praktische Optimierung* (Oldenbourg, München, 1974).
358. R.S. Varga, *Functional analysis and approximation theory in numerical analysis* (SIAM, Philadelphia, 1991).
359. S.A. Vavasis, *Nonlinear optimization: complexity issues* (Oxford University Press, 1992).
360. T.L. Vincent and W.J. Grantham, *Optimality in parametric systems* (Wiley, New York, 1981).
361. W. Vogel, *Vektoroptimierung in Produkträumen* (Hain, Meisenheim am Glan, 1977).
362. J. von Neumann and O. Morgenstern, *Theory of games and economic behavior* (Princeton University Press, Princeton, 1944).
363. M. Walk, *Theory of duality in mathematical programming* (Springer, Wien, 1989).
364. G.R. Walsh, *Methods of optimization* (Wiley, London, 1975).
365. J. Werner, *Optimization - theory and applications* (Vieweg, Braunschweig, 1984).
366. R.J. Wets, *Grundlagen konvexer Optimierung* (Springer, Berlin, 1976).
367. D.J. White, *Optimality and efficiency* (Wiley, Chichester, 1982).
368. D.J. Wilde, *Optimum seeking methods* (Prentice-Hall, Englewood Cliffs, 1964).
369. D.J. Wilde, *Globally optimal design* (Wiley, New York, 1978).
370. D.J. Wilde and C.S. Beightler, *Foundations of optimization* (Prentice-Hall, Englewood Cliffs, 1967).
371. H.P. Williams, *Model building in mathematical programming* (Wiley, Chichester, 1985).
372. D.A. Wismer, *Optimization methods for large-scale systems* (McGraw-Hill, New York, 1971).
373. D.A. Wismer and R. Chattergy, *Introduction to nonlinear optimization* (North-Holland, New York, 1978).
374. S.J. Wright, *Primal-dual interior-point methods* (SIAM,1997).
375. Z.B. Zabinsky, *Stochastic adaptive search for global optimization* (Springer, 2003).
376. F. Zach, *Technisches Optimieren* (Springer, Wien, 1974).
377. R. Zielinski and P. Neumann, *Stochastische Verfahren zur Suche nach dem Minimum einer Funktion* (Akademie-Verlag, Berlin, 1983).
378. G. Zoutendijk, *Mathematical programming methods* (North-Holland, Amsterdam, 1976).
379. S.I. Zuchovickij and L.I. Avdeeva, *Linear and convex programming* (Saunders, Philadelphia, 1966).

Index

J. Jahn, *Introduction to the Theory of Nonlinear Optimization*,
https://doi.org/10.1007/978-3-030-42760-3