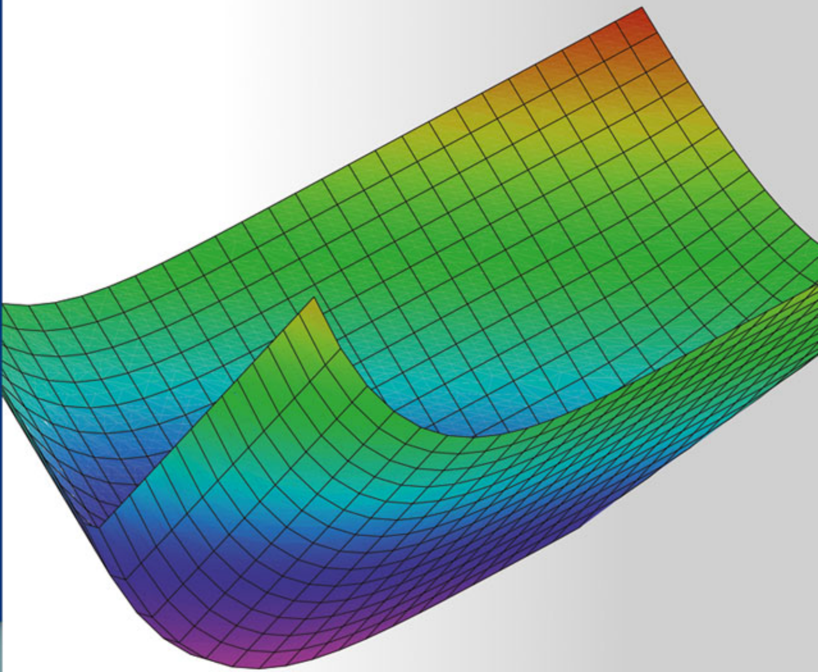
DE GRUYTER

TEXTBOOK

Peter Zörnig

NONLINEAR PROGRAMMING

AN INTRODUCTION



Peter Zörnig
Nonlinear Programming
De Gruyter Textbook

Peter Zörnig

Nonlinear Programming

An Introduction

DE GRUYTER

Mathematics Subject Classification 2010
90C30, 49Kxx, 46N10, 47N10, 90C26, 90C46

Author
Prof. Dr. Peter Zörnig
University of Brasilia
Institute of Exact Sciences
Statistical Department
Asa Norte
70910-900 Brasilia-DF
BRAZIL
peter@unb.br

ISBN 978-3-11-031527-1
e-ISBN 978-3-11-031528-8

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de.

Typesetting: le-tex publishing services GmbH, Leipzig
Printing and binding: CPI buch bücher.de GmbH, Birkach
♾Printed on acid-free paper
Printed in Germany

www.degruyter.com

Preface

The present book is based on my nonlinear programming text in Portuguese “Introdução à Programação Não Linear”, edited by the publisher of the University of Brasilia (Editora UnB) in 2011. The text originated from my lecture notes for various courses for undergraduate and graduate students, which I have taught at the Statistical Department of the University of Brasilia since 1998. The book is written primarily for undergraduate students from the fields of applied mathematics, engineering, economy, computation, etc., that want to become familiar with nonlinear optimization. I have highlighted important readings to create a comprehensive bibliography. In addition to the references, diverse web sites with information about nonlinear programming (in particular software) are cited.

The philosophy of the present book is to present an illustrative didactical text that makes access to the discipline as easy as possible. In order to motivate the student, many examples and indications of practical applications are presented. Moreover, the book contains 123 illustrative figures and 123 exercises with detailed solutions at the end of the text. Without sacrificing the necessary mathematical rigor, excessive formalisms are avoided, making the text adequate for individual studies.

The text is organized into two parts. Part I consists of Chapter 2 to Chapter 4 and provides the basic theory, introducing diverse optimality conditions and duality concepts, where special attention is given to the convex problem. Part II consists of Chapter 5 to Chapter 11 and studies solution methods for diverse specific forms of the nonlinear programming problem. After introducing evaluation criteria for iterative procedures, classical and some more advanced solution methods are presented: the minimization of a one-dimensional function is discussed in Chapter 6, the unrestricted minimization in Chapter 7, and Chapters 8 and 9 deal with linearly constrained and quadratic optimization, respectively. In Chapter 10, “classical” approaches to solve the general (continuous) optimization problem are studied: the penalty and the barrier method and sequential quadratic programming. Finally, Chapter 11 is devoted to two branches of nonlinear programming which have evolved immensely in the last few decades: nondifferential and global optimization. For these two areas, examples of applications and basic solution methods are indicated, including “modern” heuristic methods like, for example, genetic algorithms.

The only prerequisites necessary to understand the text are some basic concepts of linear algebra, calculus and linear programming. Occasionally some of these prerequisites are recalled during the development of the text.

I thank my colleague Prof. Hilton Vieira Machado and Prof. Pablo Guerrero García of the University of Malaga, Spain, for their valuable suggestions on earlier versions of the Portuguese text. My friend Hans Martin Duseberg realized the enormous work of text processing in LaTeX, including the construction of all figures. I am very grateful for his excellent work over several years.

Contents

Notations

Vectors

$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \ x^k = \begin{pmatrix} x_1^k \\ \vdots \\ x_n^k \end{pmatrix}$	(vectors are column vectors unless stated otherwise)
$\mathbb{R}^n$	set of column vectors with n components
$x^T = (x_1, \dots, x_n)$, $x^{kT} = (x_1^k, \dots, x_n^k)$	transpose of the vector x or x^k, respectively
$x^T y = x_1 y_1 + \dots + x_n y_n$	scalar product
$\lvert x \rvert = \sqrt{x^T x}$	Euclidean norm of x
$x \le y$	means that $x_i \le y_i$ for $i = 1, \dots, n$ if $x, y \in \mathbb{R}^n$
$\operatorname{grad} f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}$	gradient of function f at point x
$\operatorname{grad}_x \Phi(x,u) = \begin{pmatrix} \frac{\partial \Phi(x,u)}{\partial x_1} \\ \vdots \\ \frac{\partial \Phi(x,u)}{\partial x_n} \end{pmatrix}$	gradient of Φ with respect to x

Matrices

$A = (a_{ij}) = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$	$m \times n$ matrix
$(y^1, \dots, y^n)$	matrix with columns $y^1, \dots, y^n$
$\mathbb{R}^{m \times n}$	set of $m \times n$ matrices
A^T	transpose of matrix A
A^{-1}	inverse of matrix A
$\lvert A \rvert$	determinant of matrix A
$\operatorname{rank}(A)$	rank of matrix A (maximum number of linearly independent columns or rows)
I, I_n	$n \times n$ identity matrix
$Hf(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}$	Hessian matrix of the function f at point x
$H_x \Phi(x,u)$	Hessian matrix of Φ with respect to x
$J_f(x)$	Jacobian matrix of the function f at point x

Sets

$A \subset B$	means that any element of A is contained in B
$\mathbb{R}_+$	set of nonnegative real numbers
$\mathbb{R}_+^m = \{x \in \mathbb{R}^m \mid x \geq 0\}$	set of vectors with m nonnegative components
$Z(x^*)$	cone of feasible directions in x^*
$\overline{M}$	closure of the set M
$A(x^*)$	set of active constraints in x^*
$L(x^*)$	cone of linearized constraints in x^*
$L_0(x^*)$	interior of $L(x^*)$
$C(x^*)$	cone of tangent directions
$U_\varepsilon(x^*) = \{x \in \mathbb{R}^n \| \|x - x^*\| \leq \varepsilon\}$	ε-neighborhood of the point x
$\left[a, b\right]$	closed interval for $a < b$
(a, b)	open interval for $a < b$
$\left[x^1, x^2\right]$	line segment joining the points $x^1, x^2 \in \mathbb{R}^n$
(x^1, x^2)	relative interior of $[x^1, x^2]$
$I(a, b)$	n-dimensional interval
$\mathrm{conv}\ (K)$	convex hull of $K \subset \mathbb{R}^n$
$\mathrm{Nu}(A)$	nucleus of the matrix A
$\mathrm{Im}(A^T)$	image of the matrix A^T

Functions

$f : M \to \mathbb{R}$	real-valued function defined on the set M
$f : \mathbb{R}^n \to \mathbb{R}$	real-valued function of n variables (the domain is the entire space $\mathbb{R}^n$ or an unspecified subset)
$z := f(x),\ x := \text{“ expression ”}$	means that z is defined by $f(x)$ or by the expression on the right-hand side
$\Phi(x, u, v)$ $= f(x) + u^T g(x) + v^T h(x)$	(generalized) Lagrange function
$g^+(x)$	positive part of the function g
$\partial f(x)$	subdifferential of f at x (set of subgradients of f at x)
$epi\ (f)$	epigraph of the function f
$Df(x^*, y)$	directional derivative of the function f at point x in the direction y
$D^+f(x^*, y),\ D^-f(x^*, y)$	unilateral directional derivatives
$f_{(y)}(t) = f(x^* + ty)$	unidimensional function with variable t for $x^*, y \in \mathbb{R}^n$
$\min_{x \in M} f(x)$	problem of minimizing f over M (or minimum of f over M)
$\max_{x \in M} f(x)$	problem of maximizing f over M (or maximum of f over M)
$\sup_{x \in M} f(x)$	supremum of f over M
$\inf_{x \in M} f(x)$	infimum of f over M

1 Introduction

1.1 The model

Many problems of operations research, engineering and diverse other quantitative areas may be formulated as an optimization problem. Basically, such a problem consists of maximizing or minimizing a function of one or several variables such that the variables satisfy certain constraints in the form of equations or inequalities.

Let us consider the following *standard form* of an optimization problem:

$$\min f(x)$$

subject to

$$g_i(x) \leq 0\,, \quad i = 1, \dots, m\,, \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$ ($i = 1, \dots, m$, see the Notations at the beginning of the book) are real-valued functions of n variables. In this book we will restrict ourselves to the case in which all these functions are continuous. We are searching for a point $\overline{x} \in \mathbb{R}^n$ such that $g_i(\overline{x}) \leq 0$ for $i = 1, \dots, m$ and $f(\overline{x}) \leq f(x)$ for all x satisfying the inequalities in (1.1).

An example for (1.1) with $n = 3$ and $m = 2$ is the problem

$$\min 2x_1 + x_2^3 - x_3^5/\sqrt{x_1}$$

subject to

$$\begin{aligned} x_1^2 + x_2 \ln x_1 - 5 &\leq 0 \\ x_1 + x_3^2 - 20 &\leq 0\,. \end{aligned}$$

In this case we have $x = (x_1, x_2, x_3)^{\mathrm{T}}$, $f(x_1, x_2, x_3) = 2x_1 + x_2^3 - x_3^5/\sqrt{x_1}$, $g_1(x_1, x_2, x_3) = x_1^2 + x_2 \ln x_1 - 5$, $g_2(x_1, x_2, x_3) = x_1 + x_3^2 - 20$. (Note that $f(x)$ consists of the total expression following the symbol "min".)

Instead of (1.1) we may use one of the compact forms:

$$\min f(x)$$

subject to

$$x \in M \tag{1.2}$$

with $M := \{x \in \mathbb{R}^n | g_i(x) \leq 0; i = 1, \dots, m\}$ or

$$\min f(x)$$

subject to

$$g(x) \leq 0\,, \tag{1.3}$$

where $g_1(x) := (g_1(x), \dots, g_m(x))^T$ is a vector-valued function and 0 is the null vector of $\mathbb{R}^m$.

The function f will be called the *objective function*, and the inequalities in (1.1) are called *constraints* or *restrictions*. The elements of the set M in (1.2) are called *feasible solutions* or *feasible points* and M is referred to as a *feasible set* or *feasible region*. A point $\overline{x} \in M$ is called an *optimal solution* or *optimal point* of (1.1) if

$$f(\overline{x}) \leq f(x)$$

holds for all $x \in M$. The value $f(\overline{x})$ of an optimal solution $\overline{x}$ is called the *optimal value*. Whenever we refer to "the solution" of an optimization problem, an optimal solution is meant.

Various optimization problems, which are not written in the form (1.1) can be reformulated as a problem in standard form. For example, in the case of maximization we can transform

$$\max_{x \in M} f(x) \tag{1.4}$$

into the equivalent problem

$$\min_{x \in M} -f(x)\,. \tag{1.5}$$

Of course, the optimal values of (1.4) and (1.5) differ in sign. An inequality $g_i(x) \geq 0$ can be written as $-g_i(x) \leq 0$, and the equation $g_i(x) = 0$ can be substituted by the two inequalities

$$g_i(x) \leq 0 \quad \text{and} \quad -g_i(x) \leq 0\,.$$

We emphasize that in practice equations are maintained and not substituted by inequalities, however the transformation of several cases into a common form allows to develop a uniform theory.

Example 1.1. The following optimization problem is not in standard form:

$$\begin{aligned} \max 5x_1 - 2x_2^2x_3 \\ 4x_1^3 - x_2 \geq 12 \\ x_1x_3 + x_2^2 = 5 \\ x_2 \leq 0\,. \end{aligned}$$

Exercise 1.2. *Rewrite the above problem in the form (1.1).*

For the sake of completeness we mention that problem (1.1) can be generalized, allowing that one or both of the parameters n, m are infinite. In these cases a problem of *semiinfinite programming* or *infinite programming*, respectively, is the result. However, such generalizations cannot be treated in the present introductory book.

An important special case of (1.1) is the *linear problem*. In this case all the functions f and g_i are of the form $a^T x + b$ with $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ (linear function including additive constant).

Definition 1.3. Problem (1.1) is called a *(continuous) nonlinear problem*, if at least one of the functions $f, g_1, \ldots, g_m$ is *not* of the form $a^T x + b$ with $a \in \mathbb{R}^n$, $b \in \mathbb{R}$.

In order to illustrate the practical relevance of the nonlinear model we will now consider a simple economic problem with a single variable.

Example 1.4 (Determination of the optimal price). Let x be the price of one unit of a certain product and $y(x)$ the estimated sale in dependence of x. Suppose that the simple relation

$$y(x) = \alpha x + \beta\,,$$

holds which is frequently assumed in economics. The gross income $r(x)$ is given by the nonlinear function

$$r(x) = xy(x) = \alpha x^2 + \beta x\,.$$

In a realistic model the unit price x cannot have arbitrary values: for example, we can assume that the restrictions $c_1 \le x \le c_2$ are valid. In order to determine the unit price x that maximizes the gross income we have to solve the nonlinear problem

$$\begin{aligned} \max\ & \alpha x^2 + \beta x \\ & c_1 \le x \le c_2 \end{aligned}$$

which can be written in the standard form

$$\begin{aligned} \min\ & -\alpha x^2 - \beta x \\ & x - c_2 \le 0 \\ & c_1 - x \le 0\,. \end{aligned}$$

This problem which involves only one variable can be easily solved by means of elementary methods of calculus. A case with several variables will be studied in Example 1.9.

In the following section we will present some illustrative examples of application of the nonlinear model in various areas. In Section 1.3 we point out some complications encountered in the solution of nonlinear problems, using graphical illustrations. References, including internet sites for Chapter 1 are indicated in Section 1.4.

1.2 Special cases and applications

Models of mathematical programming have been used in various scientific areas, technology and business. In particular, a number of diverse actual Brazilian projects are based on nonlinear models. For example, such a model is used for simulation, optimization and control of the interconnected Coremas and Mãe D'Água reservoir system in the State of Paraíba. Another (mixed integer) nonlinear programming model serves as a support tool to obtain optimal operational policies in water distribution systems. Nonlinear models are also used in production and distribution of electrical energy to optimize the flow of power, considering various objectives such as, for example, minimization of the power generation cost. A problem of special interest is the *economic dispatch problem*, i.e. minimizing the cost of fuel used in a thermal power plant which has to produce a given amount of electric energy. Nonlinear models are also employed in regulating traffic on metro lines. Finally, the project of implementation of an automated mail sorting system of the correspondence of the Brazilian Company of Post and Telegraphs led to a nonlinear model to determine the optimal locations of several operational centers.

Obviously, a practical model must be sufficiently complex and sophisticated in order to represent the most relevant aspects of reality. The modeling process of complex systems includes features of an "art" and will not be examined in detail in this introductory text. In this section we will restrict ourselves to simple and illustrative examples which make clear that the nonlinear model has practical importance. In the following we present applications of the separable and the quadratic model (Sections 1.2.1 and 1.2.2, respectively).

1.2.1 Separable problem

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *separable*, if it is the sum of expressions which individually depend on a unique variable, i.e. $f(x_1, \ldots, x_n) = \sum_{i=1}^n f_i(x_i)$, where $f_i(x_i)$ is a function of x_i for $i = 1, \ldots, n$.

For example, $f(x_1, x_2, x_3) = \sqrt{x_1} + x_2 + x_2^2 + \ln x_3$ is a separable function with $f_1(x_1) = \sqrt{x_1}, f_2(x_2) = x_2 + x_2^2, f_3(x_3) = \ln x_3$. In particular, every linear function is separable. An optimization problem (1.1) is called a *separable problem*, if all the functions $f, g_1, \ldots, g_m$ are separable. We will consider two applications of the (nonlinear) separable problem in economy.

Example 1.5 (Investment planning). A total value of at most a can be applied in n possible investments. The expected profit of investment j is a function of the applied value x_j, represented by $f_j(x_j)$. In order to be economically reasonable, $f_j(x_j)$ must be an increasing function with $f_j(0) = 0$ and decreasing slope, i.e. concave (see Section 3.2). In economy it is frequently assumed that $f_j(x_j)$ has one of the following forms satisfying these conditions:

$$f_j(x_j) = \alpha_j(1 - e^{-\beta_j x_j}), f_j(x_j) = \alpha_j \ln(1 + \beta_j x_j) \text{ or } f_j(x_j) = \frac{\alpha_j x_j}{x_j + \beta_j}$$

with $\alpha_j, \beta_j > 0$.

In order to determine the investment deposits x_j which maximize the total expected profit, we have to solve the nonlinear problem:

$$\max \sum_{j=1}^{n} f_j(x_j)$$
$$\sum_{j=1}^{n} x_j \leq a$$
$$x_j \geq 0\,, \quad j = 1, \ldots, n.$$

For illustration purpose we consider a simple numerical example: an investor has the value $a = \$\,200\,000$ available which can be applied in two types of investments with expected monthly profit

$$f_1(x_1) = \frac{2000x_1}{x_1 + 6000} \quad \text{and} \quad f_2(x_2) = \frac{4000x_2}{x_2 + 30\,000}\,,$$

respectively.

Thus the nonlinear model is:

$$\max \frac{2000x_1}{x_1 + 6000} + \frac{4000x_2}{x_2 + 30000}$$
$$x_1 + x_2 \leq 200\,000$$
$$x_1, x_2 \geq 0\,.$$

This simple example can be easily solved with elementary methods (see Exercise 4.36). We obtain the optimal investment deposits $x_1^* = \$\,50\,699.72$ and $x_2^* = \$\,149\,300.28$ with corresponding expected monthly profit of $\$\,5119.09.

Example 1.6 (Nonlinear transportation problem). A product P is manufactured by the factories (origins) $A_1, \ldots, A_m$, and $B_1, \ldots, B_n$ represent centers of consumption (destinations) of P. Let a_i and b_j denote the quantity of P available in factory A_i and the quantity required at destination B_j, respectively $(i = 1, \ldots, m; j = 1, \ldots, n)$. The cost to transport the quantity x_{ij} from A_i to B_j will be denoted by $f_{ij}(x_{ij})$ (see Figure 1.1).

The problem consists of determining the quantities x_{ij} to be transported from A_i to B_j such that

- the quantities available in the origins are not exceeded
- the demands in the various destinations are at least satisfied and
- the total transportation cost is minimized.

We assume that the total amount available of the product is not less than the total amount required, i.e.

$$\sum_{i=1}^{m} a_i \geq \sum_{j=1}^{n} b_j\,, \tag{1.6}$$

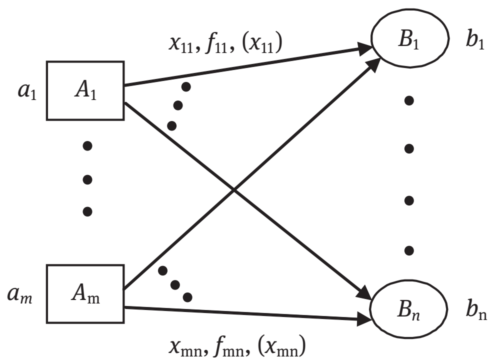


Fig. 1.1. Graphical illustration of the transportation problem.

which is clearly a necessary condition for the existence of a feasible solution. We obtain the model:

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}\ (x_{ij}) \tag{1.7a}$$

$$\sum_{j=1}^{n} x_{ij} \le a_i \qquad (i = 1, \dots, m) \tag{1.7b}$$

$$\sum_{i=1}^{m} x_{ij} \ge b_j \qquad (j = 1, \dots, n) \tag{1.7c}$$

$$x_{ij} \ge 0 \qquad (i = 1, \dots, m; j = 1, \dots, n)\,. \tag{1.7d}$$

For economic reasons a feasible solution cannot be optimal if at least one of the destinations receives a larger amount than required. Thus, the constraints (1.7c) can be substituted by the equations

$$\sum_{i=1}^{m} x_{ij} = b_j \quad (j = 1, \dots, n)\,. \tag{1.8}$$

In the unrealistic case in which (1.6) is satisfied as equation, i.e. $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$, any feasible solution of (1.7) satisfies $\sum_{j=1}^{n} x_{ij} = a_i$ for $i = 1, \dots, m$ and these equations can be used to substitute the conditions (1.7b). If $\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$, we proceed as follows: introducing slack variables $x_{io} \ge 0$ we can convert the inequalities (1.7b) into equations

$$\sum_{j=1}^{n} x_{ij} + x_{io} = a_i \quad (i = 1, \dots, m)\,, \tag{1.9}$$

where x_{i0} presents the “superfluous production” of factory A_i. Obviously the total superfluous production of all factories is the difference between the total quantity available and the total quantity required of P:

$$\sum_{i=1}^{m} x_{io} = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j\,. \tag{1.10}$$

We introduce an artificial consumption center B_0 which requires exactly the total superfluous production $b_0 := \sum_{i=1}^{m} x_{io}$ and to which the factories can transport at zero cost. By substituting the conditions (1.7b) and (1.7c) of the previous model by (1.8) and (1.9), respectively and including (1.10) we obtain

$$\begin{aligned}
&\min \sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}\,(x_{ij}) \\
&\sum_{j=0}^{n} x_{ij} = a_i \qquad (i = 1, \dots, m) \\
&\sum_{i=1}^{m} x_{ij} = b_j \qquad (j = 0, 1, \dots, n) \\
&\quad x_{ij} \ge 0 \qquad (i = 1, \dots, m\,; \quad j = 0, 1, \dots, n)\,.
\end{aligned} \tag{1.11}$$

The partial cost function for a given route, $f_{ij}(x_{ij})$ can assume diverse forms. In the simplest case, this cost is proportional to the transported quantity, i.e. $f_{ij}(x_{ij}) = c_{ij}x_{ij}$. In this case (1.11) becomes a special linear model which will not be treated here. Instead we consider the following situation, yielding a nonlinear problem.

Suppose that a product P is transported from A_i to B_j (i, j fixed) by means of trucks of the same type, each of which can transport the maximal quantity of $S = 5$ tons of the product. Moreover, suppose that the cost of using any of these trucks is $C = \$2000$, independent of the carried quantity of P. Thus the cost $f_{ij}(x_{ij})$ for transporting the quantity x_{ij} of the product from A_i to B_j is a step function (piecewise constant function)

$$f_{ij}(x_{ij}) = \begin{cases} 0 & \text{for } x_{ij} = 0 \\ C & \text{for } 0 < x_{ij} \le S \\ 2C & \text{for } S < x_{ij} \le 2S \\ 3C & \text{for } 2S < x_{ij} \le 3S \\ \quad\vdots & \qquad\vdots \end{cases}$$

The two previous examples have linear restrictions and a separable objective function, i.e. they represent specific separable problems which can be written in the form

$$\begin{aligned} &\min f(x) \\ &Ax \le b \end{aligned} \tag{1.12}$$

with $f(x) = \sum_{i=1}^{n} f_i(x_i), x := (x_1, \dots, x_n)^T$. In certain cases, a problem with linear constraints and an objective function which is not in separable form can be rewritten in the form (1.12). Consider, for example, a problem with objective function

$$f(x_1, x_2) = x_1 x_2 \,. \tag{1.13}$$

By means of the variable transformation

$$y_1 = \frac{x_1 + x_2}{2}\,, \quad y_2 = \frac{x_1 - x_2}{2}$$

which is equivalent to

$$x_1 = y_1 + y_2\,, \quad x_2 = y_1 - y_2$$

we obtain

$$f(x_1, x_2) = x_1 x_2 = (y_1 + y_2)(y_1 - y_2) = y_1^2 - y_2^2\,,$$

i.e. a separable function of the variables y_1 and y_2. Note that the considered transformation preserves the linearity of any linear constraint, since x_1 and x_2 are substituted by the linear terms $y_1 + y_2$ and $y_1 - y_2$, respectively.

Example 1.7. Given the problem

$$\begin{aligned} \min x_1 x_2 - {} & x_3^2 \\ x_1 + x_2 + 2x_3 & \le 20 \\ x_2 \quad & \ge 15\,. \end{aligned}$$

The considered variable transformation transforms the objective function in a separable function and preserves the linearity of the constraints, i.e. a problem of the form (1.12) with the variables y_1, y_2, x_3 results in:

$$\begin{aligned} \min y_1^2 - y_2^2 - x_3^2 & \\ 2y_1 + \quad 2x_3 & \le 20 \\ y_1 - \quad y_2 \quad & \ge 15\,. \end{aligned}$$

For separable problems, there exist efficient specialized resolution techniques. Some of them approximate nonlinear functions of the model by piecewise linear functions and employ an adaptation of the simplex method.

Exercise 1.8. *Rewrite the following problem in the form (1.12):*

$$\begin{aligned} \min 2x_1 x_2 - x_3 x_4 & \\ x_1 + x_2 \quad + 2x_4 & \le 30 \\ x_1 \quad + x_3 - x_4 & \ge 20\,. \end{aligned}$$

1.2.2 Problem of quadratic optimization

We will present another economic application:

Example 1.9 (Determination of optimal prices). A monopolist produces the products $P_1, \dots, P_n$ with unit prices $x_1, \dots, x_n$ to be determined. The raw materials

$M_1, \dots, M_m$ are used in the production which are available in the quantities $r_1, \dots, r_m$, respectively. Let

r_{ij}: the amount of M_i required to produce one unit of P_j $(i = 1, \dots, m; j = 1, \dots, n)$ and

y_j: the amount of P_j to be produced (and sold) in the planning period $(j = 1, \dots, n)$.

Considering the vectors $x = (x_1, \dots, x_n)^T$, $y = (y_1, \dots, y_n)^T$, $r = (r_1, \dots, r_m)^T$ and the matrix

$$R = \begin{pmatrix} r_{11} & \dots & r_{1n} \\ \vdots & & \vdots \\ r_{m1} & \dots & r_{mn} \end{pmatrix},$$

we have the following nonnegativity restrictions and limitations of raw materials:

$$\begin{aligned} y &\geq 0 \\ Ry &\leq r\,. \end{aligned} \tag{1.14}$$

Generalizing the one-dimensional relation $x(c) = \alpha c + \beta$ of Example 1.4 we assume here the linear relation

$$y = Qx + q \tag{1.15}$$

between unit prices and sales, where $Q \in \mathbb{R}^{n \times n}$ and $q \in \mathbb{R}^n$. The elements of the matrix Q and the vector q can be estimated empirically by the study of consumer behavior. Substituting y in (1.14) by the expression in (1.15) we obtain

$$\begin{aligned} Qx + q &\geq 0 \\ R(Qx + q) &\leq r\,, \end{aligned} \tag{1.16}$$

which is equivalent to

$$\begin{aligned} -Qx &\leq q \\ RQx &\leq r - Rq\,. \end{aligned} \tag{1.17}$$

The objective is to maximize the gross income given by

$$z = \sum_{i=1}^{n} x_i y_i = x^T y = x^T(Qx + q) = x^T Qx + q^T x\,.$$

We obtain the following model to determine the optimal prices:

$$\begin{aligned} &\max x^T Qx + q^T x \\ &-Qx \leq q \\ &RQx \leq r - Rq \\ &x \geq 0\,. \end{aligned} \tag{1.18}$$

Let us illustrate the model by means of the following hypothetical example: a bakery produces three types of cake: P_1: wheat cake, P_2: apple cake and P_3: lemon cake. Among others, the following ingredients are needed, which are available in limited quantities: M_1: wheat, M_2: eggs, M_3: butter, M_4: sugar. In previous years the bakery has sold the three types of cake for different prices and studied the dependence of sales from the prices which led to the following relation (see (1.15)):

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} -3 & 0 & 1 \\ 0 & -4 & 0 \\ 2 & 0 & -5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} 100 \\ 120 \\ 160 \end{pmatrix} . \tag{1.19}$$

Note that y_i and q_i have the unit kg and x_i has the unit \$/kg. Consequently, the elements of Q have the unit kg^2/\$. Considering, for example, the first line of (1.19), we have

$$y_1 = -3x_1 + x_3 + 100 ,$$

i.e. the sale of cake P_1 does not only depend on the unit price x_1 of this cake, but also on the unit price x_3 of cake P_3. If, for example, x_1 increases by one unit, while x_3 remains unchanged, the sale y_1 of P_1 reduces by 3 kg. If x_3 increases by one unit, while x_1 remains unchanged, the sale y_1 increases by 1 kg. This can be explained by the fact that some clients who used to by lemon cake decide to by wheat cake instead, if the former becomes more expensive. Similarly, one can explain the third line of (1.19): $y_3 = 2x_1 - 5x_3 + 160$.

The second line shows that the sale y_2 depends exclusively on the price $x_2 : y_2 = -4x_2 + 120$. The clients who like apple cake would never exchange it for another cake, no matter how cheap the others might be.

The quantities of monthly available ingredients are $(r_1, \dots, r_4) = (20, 15, 13, 18)$ (in kg), and the matrix of consumption is

$$R = \begin{pmatrix} 0.4 & 0.3 & 0.35 \\ 0.15 & 0.1 & 0.15 \\ 0.25 & 0.3 & 0.2 \\ 0.1 & 0.2 & 0.1 \end{pmatrix} .$$

Since the elements of the vector Rq have the unit kg (see the second restriction of (1.18)), the elements of the matrix R do not have units. For example, the first column of R represents the consumption of wheat cake: 1 kg of this cake needs 0.4 kg of wheat, 0.15 kg of eggs, 0.25 kg of butter and 0.1 kg of sugar. We obtain

$$RQ = \begin{pmatrix} -0.5 & -1.2 & -1.35 \\ -0.15 & -0.4 & -0.6 \\ -0.35 & -1.2 & -0.75 \\ -0.1 & -0.8 & -0.4 \end{pmatrix} \quad \text{and} \quad r - Rq = \begin{pmatrix} -112 \\ -36 \\ -80 \\ -32 \end{pmatrix} ,$$

resulting in the following model (see (1.18)):

$$
\begin{aligned}
\max -3x_1^2 - 4x_2^2 - 5x_3^2 + 3x_1x_3 + 100x_1 + 120x_2 + 160x_3 \\
\end{aligned}
$$

$$
\begin{array}{rrrl}
33x_1 & & -x_3 & \le 100 \\
& 4x_2 & & \le 120 \\
-2x_1 & & +5x_3 & \le 160 \\
-0.5x_1 & -1.2x_2 & -1.35x_3 & \le -112 \\
-0.15x_1 & -0.4x_2 & -0.6x_3 & \le -36 \\
-0.35x_1 & -1.2x_2 & -0.75x_3 & \le -80 \\
-0.1x_1 & -0.8x_2 & -0.4x_3 & \le -32 \\
& & x_1, x_2, x_3 & \ge 0\,.
\end{array}
$$

Using adequate software, for example LINGO, we obtain the optimal unit prices $x_1^* = 44.68\,\$/\text{kg}$, $x_2^* = 28.24\,\$/\text{kg}$, $x_3^* = 41.32\,\$/\text{kg}$ resulting in the optimal gross income $f(x_1^*, x_2^*, x_3^*) = \$\,2291.81$ per month.

Problem (1.18) has a quadratic objective function and linear constraints. Such a problem occurs frequently in mathematical programming and is called the *quadratic optimization problem*. We consider the standard form

$$
\begin{aligned}
\min\ & \frac{1}{2}x^TQx + q^Tx \\
& Ax \le b \\
& x \ge 0\,,
\end{aligned}
\tag{1.20}
$$

where $Q \in \mathbb{R}^{n\times n}$ is a nonzero matrix, $q \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. The factor $\frac{1}{2}$ is introduced in problem (1.20) so that the matrix Q is equal to the Hessian matrix of the objective function: observe that the *Hessian matrix* of $(\frac{1}{2}x^TQx + q^Tx)$ is Q, while the Hessian matrix of $(x^TQx + q^Tx)$ would be $2Q$.

Comment 1.10. Without loss of generality we can assume that the matrix in (1.20) is symmetric. Otherwise, i.e. if $f(x) = x^TPx$ with a nonsymmetric matrix P, we can write, by using the relation $x^TPx = x^TP^Tx$:

$$
f(x) = x^TPx = \frac{1}{2}(x^TPx + x^TP^Tx) = x^T\frac{P + P^T}{2}x\,,
$$

where the matrix $Q := \frac{P+P^T}{2}$ is symmetric.

Problems of quadratic optimization will be discussed in Chapter 9.

Exercise 1.11. *Determine the optimal quantities y_1^*, y_2^* and y_3^* of the cakes which must be produced in the previous example and verify that the limitation constraints of raw materials are satisfied.*

Exercise 1.12.

(a) *Rewrite the term* $(x_1, x_2)\begin{pmatrix} 1 & 4 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ *in the form* $(x_1, x_2)Q\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ *with a symmetric matrix* Q.

(b) *Rewrite the function*

$$f(x_1, x_2) = 3x_1 + 4x_2 - 2x_1^2 + 5x_1x_2 + 8x_2^2$$

in the form $f(x) = \frac{1}{2}x^TQx + q^Tx$ *with a symmetric matrix* Q.

Exercise 1.13. *Which of the following problems are of type (a) linear, (b) separable, (c) quadratic, (d) nonlinear (a problem can be of various types):*

(i)
$$\begin{aligned} &\min x_1 + 4x_2 + 2x_1^2 - x_2^2 \\ &2x_1 + x_2 \le 8 \\ &x_1, x_2 \ge 0 \,. \end{aligned}$$

(ii)
$$\begin{aligned} &\min 3x_1x_2 - x_1 \\ &x_1 + x_2 \le 5 \\ &x_1, x_2 \ge 0 \,. \end{aligned}$$

(iii)
$$\begin{aligned} &\min 8x_1 - 2x_2 \\ &2x_1 + x_2 \ge 10 \\ &x_1, x_2 \ge 0 \,. \end{aligned}$$

(iv)
$$\begin{aligned} &\min x_1 e^{x_2} + x_1^2 \\ &2x_1 + x_2^2 \le 5 \,. \end{aligned}$$

(v)
$$\begin{aligned} &\min \, e^{x_1} + \ln x_2 \\ &x_1^5 + 2e^{x_2} = 3 \,. \end{aligned}$$

1.2.3 Further examples of practical applications

We still indicate some possible applications of the nonlinear model which do not have any of the preceding forms. For example, in various statistical applications there occurs the problem of estimating parameters: we assume that the variable z depends on n variables $x_1, \dots, x_n$ such that

$$z = h(x_1, \dots, x_n; \beta_1, \dots, \beta_p) + \varepsilon \,,$$

where h is a function of the n variables $x_1, \dots, x_n$ which depends also on p parameters, and ε represents a random deviation. The values $h(x_1, \dots, x_n; \beta_1, \dots, \beta_p)$ and

$h(x_1, \dots, x_n; \beta_1, \dots, \beta_p) + \varepsilon$ can be interpreted as the "theoretical" and observed value of z, respectively. In m points $x^i = (x_1^i, \dots, x_n^i)^T$ the values of $z_i = h(x^i; \beta_1, \dots, \beta_p) + \varepsilon_i$ are determined by means of empirical studies $(i = 1, \dots, m)$.

The problem is to determine the parameters $\beta_1, \dots, \beta_p$, such that the absolute deviations $|z_i - h(x^i, \beta_1, \dots, \beta_p)|$ between observed and theoretical values are minimized. The *least square method* minimizes these deviations, solving the problem

$$\min \sum_{i=1}^{m} y_i^2 \tag{1.21}$$

$$y_i = |z_i - h(x^i, \beta_1, \dots, \beta_p)| \quad \text{for } i = 1, \dots, m.$$

Of course, (1.21) can be formulated as an unrestricted problem by substituting y_i in the objective function by $|z_i - h(x^i, \beta_1, \dots, \beta_p)|$.

Example 1.14 (A meteorological model). A meteorological observatory has determined the temperatures in various weather stations in different altitudes. All measurements have been obtained on the same day at the same time. The (hypothetical) results are presented in the following table and diagram (Figure 1.2).

Table 1.1. Data of the meteorological observatory.

i	x_i **(altitude, in m)**	z_i **(temperature, in °C)**
1	300	14
2	400	10
3	700	10
4	1000	6
5	1100	8
6	1500	4

The observed values suggest that the temperature decreases approximately linearly with the altitude x. It is reasonable to adapt the model

$$z = \beta_1 x + \beta_2 + \varepsilon,$$

where ε is a random variation of the temperature. Using the least square method, we obtain estimates for the parameters β_1 and β_2, by solving the problem

$$\min \sum_{i=1}^{6} y_i^2$$

$$y_i = |z_i - (\beta_1 x_i + \beta_2)|.$$

This is the special case of (1.21) with $n = 1, p = 2, m = 6$ and $h(x; \beta_1, \beta_2) = \beta_1 x + \beta_2$. This problem will be solved in Section 3.5 (Example 3.52). In the general case, the

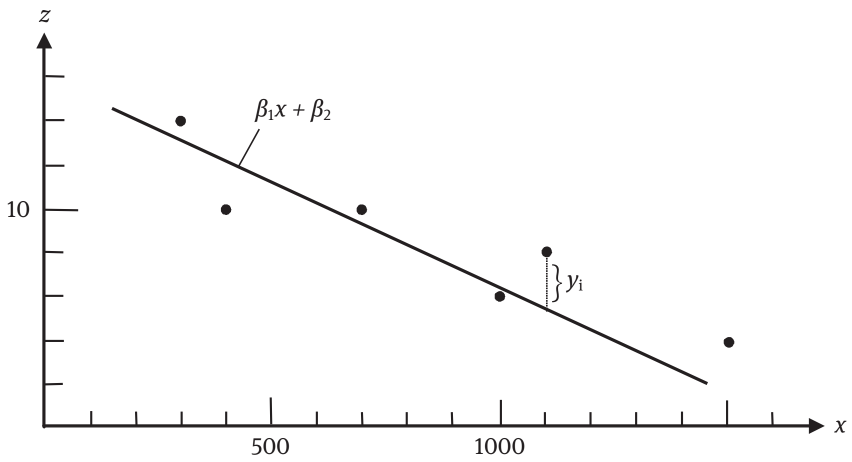


Fig. 1.2. Dispersion diagram.

exponent 2 in the objective function of (1.21) can be substituted by any number $q \geq 1$. Moreover we can associate weights $w_i > 0$ to the deviations, resulting in the nonlinear problem:

$$\min \sum_{i=1}^{m} w_i y_i^q \tag{1.22}$$

$$y_i = |z_i - h(x^i; \beta_1, \ldots, \beta_p)| \quad \text{for } i = 1, \ldots, m\,.$$

One can easily verify that (1.22) is equivalent to

$$\min \sum_{i=1}^{m} w_i y_i^q \tag{1.23}$$

$$\begin{aligned} y_i &\geq z_i - h(x^i; \beta_1, \ldots, \beta_p) && \text{for } i = 1, \ldots, m \\ y_i &\geq -z_i + h(x^i; \beta_1, \ldots, \beta_p) && \text{for } i = 1, \ldots, m\,. \end{aligned}$$

Since many solution procedures require that the functions of the optimization model are differentiable (see Chapter 11), model (1.23) has computational advantages in comparison with (1.22). Note that the right-hand sides of the constraints in (1.23) are differentiable if h is differentiable, while the absolute value in (1.22) is not everywhere differentiable.

Example 1.15 (Optimal measures of a box). A metallurgical company manufactures boxes for a cookie factory. The boxes are produced without metal cover and must have a square base and the volume $V = 2.000\,\text{cm}^3$ (Figure 1.3). The problem is to determine the measurements of the box for which the required metal foil is minimal.

Since the boxes are open at the top, the area $x_1^2 + 4x_1x_2$ of foil is needed for any of them. The volume of a box is given by $V = x_1^2 x_2$. Hence, the optimal measures are obtained

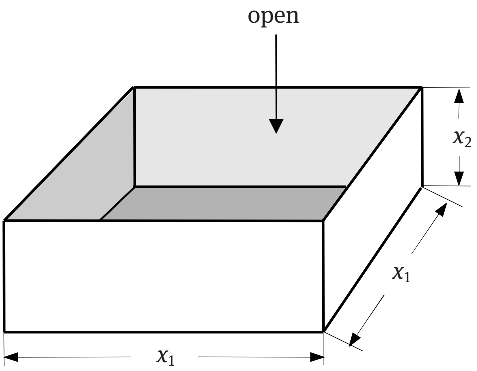


Fig. 1.3. Measurements of a box.

by solving the nonlinear problem

$$\begin{aligned} \min\, & x_1^2 + 4x_1x_2 \\ & x_1^2x_2 = 2000 \\ & x_1, x_2 \geq 0\,. \end{aligned}$$

The variable x_1 can be eliminated by substituting x_2 in the objective function by $2000/x_1^2$. Thus, we obtain the problem

$$\begin{aligned} \min\, & x_1^2 + \frac{8000}{x_1} \\ & x_1 > 0\,. \end{aligned}$$

Exercise 1.16. *Determine the optimal measures x_1^* and x_2^* of the box in the previous example. What is the minimal amount of metal foil necessary to produce one box?*

Example 1.17. We consider the following modification of Example 1.15. The box is also open at the top and the base consists of a rectangle and two semicircles attached on opposite sides, where $x_1 \leq x_2$ (Figure 1.4).

For a given volume V of a box we want to determine the measures x_1, x_2, x_3, which minimize the foil required for production. We now obtain the nonlinear optimization problem

$$\begin{aligned} \min\, & x_1x_2 + \frac{\pi}{4}x_1^2 + (2x_2 + \pi x_1)x_3 \\ & (x_1x_2 + \frac{\pi}{4}x_1^2)x_3 = V \\ & x_1 \leq x_2 \\ & x_1 \geq 0\,. \end{aligned}$$

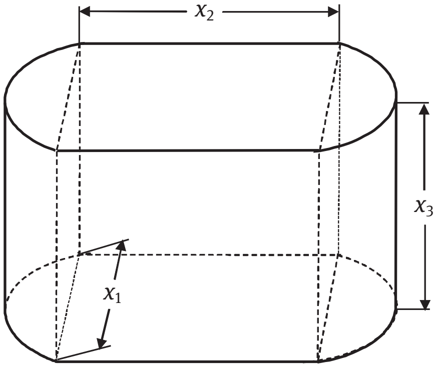


Fig. 1.4. Measures of the modified box.

Similar to the previous example we can eliminate one variable, expressing x_3 in terms of x_1 and x_2, yielding the problem:

$$\begin{aligned}\min x_1 x_2 + \frac{\pi}{4}x_1^2 + \frac{2x_2 + \pi x_1}{x_1 x_2 + \frac{\pi}{4}x_1^2}V& \\ x_1 \leq x_2& \\ x_1 \geq 0\,.&\end{aligned}$$

This problem will be solved in Exercise 4.18.

Example 1.18 (Determination of optimal location). To satisfy the demand of n gas stations $C_1, \dots, C_n$ the construction of an oil refinery D is planned. The objective is to find the location of D which minimizes the total cost of transportations between D and the C_i (Figure 1.5).

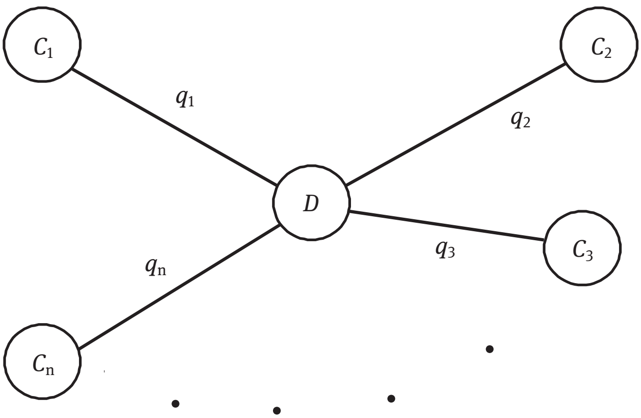


Fig. 1.5. Locations of gas stations and demands.

Let

q_i: monthly gasoline demand of C_i,

(a_i, b_i): given geographical coordinates of C_i,

(x_1, x_2): coordinates of D, to be determined.

Assume that the distance d_i between D and C_i is

$$d_i = \sqrt{(x_1 - a_i)^2 + (x_2 - b_i)^2}\,,$$

(Euclidean distance). Moreover, suppose that the transport of gasoline costs s \$ per km and cubic meter. Hence, the monthly cost of transportation between D and C_i is

$$f_i(x_1, x_2) = s q_i d_i\,,$$

resulting in the total monthly transportation cost

$$f(x_1, x_2) = s \sum_{i=1}^{n} q_i \sqrt{(x_1 - a_i)^2 + (x_2 - b_i)^2}\,.$$

By eliminating the constant s we obtain the unrestricted problem

$$\min \sum_{i=1}^{n} q_i \sqrt{(x_1 - a_i)^2 + (x_2 - b_i)^2}\,,$$

which is known in the literature as the *Weber problem*. Due to the convexity (see Section 3.2) of the objective function, the problem can be easily solved, using classical methods of nonlinear programming such as, for example, the gradient method (Section 7.2).

For further illustration we consider 5 gas stations with coordinates (in km):

$$(a_1, \ldots, a_5) = (0, 300, 500, 400, 100), (b_1, \ldots, b_5) = (500, 600, 200, -100, 100)$$

and demands (in cubic meter per month): $(q_1, \ldots, q_5) = (60, 140, 100, 40, 80)$. The software LINGO determines the optimal coordinates of the refinery: $x_1^* = 282.771$ km and $x_2^* = 365.729$ km.

If the unit cost of transportation is, for example, $s = \$\,500/(\text{km}\,\text{m}^3)$ the minimal monthly cost for all necessary transportations is \$ 62 001 470. A simpler variant of the location problem will be solved in Section 3.5 (Example 3.54).

We finally present an application in optimal control.

Example 1.19 (Construction of a road). A straight road with given width must be constructed over an uneven terrain. Suppose that the construction cost is proportional to the necessary movement of soil (volume of soil to be removed or topped up). Let T be the total length of the road and $c(t)$ the height of the terrain at the point $t \in [0, T]$. We want to determine the optimal height $x(t)$ of the road for all $t \in [0, T]$. In order to avoid excessive slopes in the road we require $|\frac{\partial x(t)}{\partial t}| \leq b_1$, and to avoid abrupt changes in the slope we also require $|\frac{\partial^2 x(t)}{\partial t^2}| \leq b_2$. Since the road must join the two extreme points $t = 0$ and $t = T$, the heights of which may not be changed, the constraints $x(0) = c(0)$ and $x(T) = c(T)$ must be included. The problem of optimizing the longitudinal profile of the road can be modeled as follows:

$$\begin{aligned} &\min \int_0^T |x(t) - c(t)|dt \\ &\left|\frac{\partial x(t)}{\partial t}\right| \le b_1 \quad \text{for } t \in [0, T] \\ &\left|\frac{\partial^2 x(t)}{\partial t^2}\right| \le b_2 \quad \text{for } t \in [0, T] \\ &x(0) = c(0) \\ &x(T) = c(T). \end{aligned} \tag{1.24}$$

In order to discretize the problem, we subdivide the interval $[0, T]$ into k subintervals of length $\Delta t := T/k$. By defining $c_i := c(i\Delta t)$ and $x_i := x(i\Delta t)$ for $i = 0, \dots, k$ and substituting the derivatives by differential quotients we obtain the following approximation for model (1.24):

$$\begin{aligned} &\min \sum_{i=1}^{k} |x_i - c_i|\Delta t \\ &|x_i - x_{i-1}| \le \Delta t b_1 && \text{for } i = 1, \dots, k \\ &|x_i - 2x_{i-1} + x_{i-2}| \le \Delta t^2 b_2 && \text{for } i = 2, \dots, k \\ &x_0 = c_0 \\ &x_k = c_k\,. \end{aligned} \tag{1.25}$$

Table 1.2. Heights of terrain and road.

Point i	**Position**	**Terrain height c_i**	**Road height x_i**
0	0	552.05	552.05
1	100	583.24	582.05
2	200	645.99	612.05
3	300	679.15	630.56
4	400	637.09	629.08
5	500	545.60	607.59
6	600	490.00	606.11
7	700	535.61	624.62
8	800	654.62	654.62
9	900	738.69	684.62
10	1000	697.54	697.54

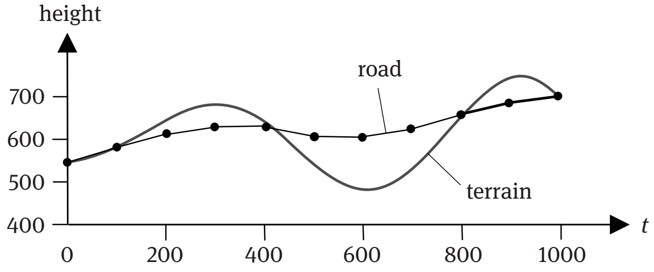


Fig. 1.6. Terrain and optimal road height.

We consider a numerical example with a total length $T = 1000$ m and subdivide the interval $[0, T]$ into $k = 10$ subintervals. The hypothetical values of the terrain altitude (in meters) are shown in the third column of Table 1.2. For the control constants $b_1 = 0.3$ and $b_2 = 0.002$, the solution of problem (1.25), determined by LINGO, is shown in the last column.

Figure 1.6 shows that a great part of the construction cost in this example is caused by topping up soil instead of removing it.

1.3 Complications caused by nonlinearity

Usually a student of nonlinear programming is already familiar with the basic concepts of linear programming. We will now illustrate why solving a nonlinear problem is much more difficult than solving a linear one. From the computational point of view, the linear problem is characterized by the fact that some methods (the simplex method, for example) can solve any special case, at least in principle. Such a universal method exists, because the linear model has several particularities:

1. The feasible region is convex and has a finite number of vertices (also called extreme points, see Section 3.1)
2. The contour lines of the objective function are parallel hyperplanes.
3. At least one vertex of the feasible region is an optimal point (assuming that a finite optimal solution exists and that the feasible region has at least one vertex).
4. Any local maximum (minimum) point is also a global maximum (minimum) point (see the following Definition 1.22).

We illustrate in this section that all four conditions may be violated in the case of nonlinear programming: the following example shows that the feasible region need not be convex and not even connected. Under such circumstances it is difficult to apply a line-search algorithm (see Section 8.1, for example) that tries to approximate the optimal solution iteratively, moving along straight lines. Such a procedure can easily get out of the feasible region, and hence the solution process may be unable to continue.

Example 1.20.

(a) Let

$$x_2 \le \frac{1}{2}(x_1 - 2)^2 + 1$$
$$x_1, x_2 \ge 0 .$$

be the constraints of a nonlinear problem. The feasible region M is not convex, i.e. condition (1) is violated (Figure 1.7).

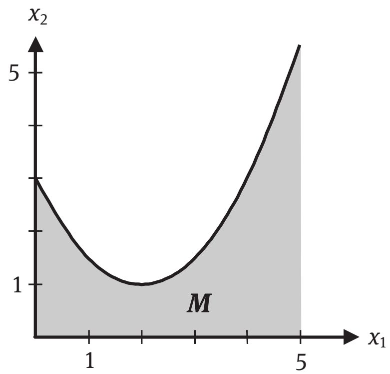


Fig. 1.7. Nonconvex feasible region.

(b) If the modified constraints

$$x_2 \le \frac{1}{2}(x_1 - 2)^2 - \frac{1}{2}$$
$$x_1, x_2 \ge 0 ,$$

are considered, the feasible region is not even connected (Figure 1.8).

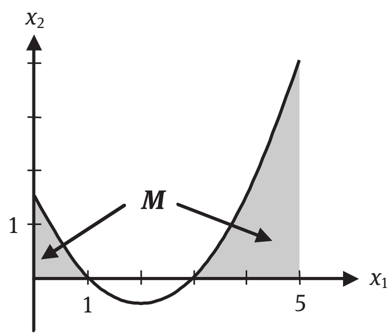


Fig. 1.8. Nonconnected feasible region.

Example 1.21. Consider the problem

$$\begin{aligned} \min(x_1-1)^2+(x_2-1)^2 \\ x_1+x_2 \le 4 \\ -x_1+x_2 \le 2 \\ x_1, x_2 \ge 0\,. \end{aligned}$$

The optimal point $\bar{x} = (1,1)^T$ lies in the interior of the feasible region M (Figure 1.9), and the contour lines of the objective function are circles with center $\overline{x}$, i.e. the conditions (2) and (3) are violated.

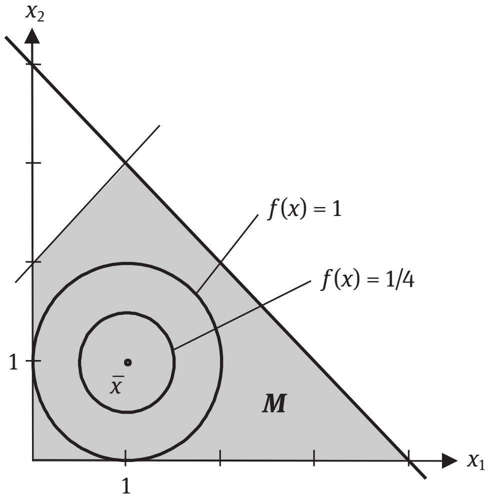


Fig. 1.9. Optimal point in the interior of M.

This example shows in particular that, unlike the simplex method, a solution method for nonlinear problems may not restrict the search of optimal points to vertices of the feasible region. In Chapter 2 we will study criteria which characterize local optimal points. Prior to showing an example in which condition (4) is violated, we recall the following concepts.

Definition 1.22. Let $f : M \to \mathbb{R}$, $M \subset \mathbb{R}^n$ be a function.

(i) A point $x^* \in M$ is called a *local minimum point* (*of f over M*), if there is an $\varepsilon > 0$ such that

$$f(x^*) \le f(x) \tag{1.26}$$

for all $x \in M \cap \{x \mid |x - x^*| < \varepsilon\}$.

(ii) $x^* \in M$ is called a *strict* (or *isolated*) *local minimum point*, if there is an $\varepsilon > 0$ such that

$$f(x^*) < f(x) \tag{1.27}$$

for all $x \in M \cap \{x| |x - x^*| < \varepsilon\}$ with $x \neq x^*$.

(iii) A point $x^* \in M$ is called a *global minimum point*, if there is an $\varepsilon > 0$ such that

$$f(x^*) \le f(x) \tag{1.28}$$

for all $x \in M$.

The corresponding concepts for a maximum point are defined analogously.

Example 1.23. Consider the problem (Figure 1.10):

$$
\begin{aligned}
&\max (x_1 - 2)^2 + (x_2 - 1)^2 \\
&-x_1 + 2x_2 \le 2 && (1)\\
&x_1 + 2x_2 \le 6 && (2)\\
&x_1 + 2x_2 \ge 2 && (3)\\
&-x_1 + 2x_2 \ge -2. && (4)
\end{aligned}
$$

The last example shows that a local maximum need not be global: the points $x^3 = (0,1)^T$ and $x^4 = (4,1)^T$ are global maxima, while $x^1 = (2,0)^T$ and $x^2 = (2,2)^T$ are only local maxima. Thus, condition (4) is violated. The fact that a local optimum need not be global causes considerable difficulties. Generally, classical solution methods of nonlinear programming can only determine a local optimum. There is no guarantee that the result of an optimization process is a global optimum. In Section 11.2 we will present "modern" methods of global optimization which are able to find a global optimum in the presence of other local optima.

Because of the possible occurrence of various complications illustrated above, no general solution method exists for nonlinear programming. Thus several specialized methods have been developed to solve specific problems (see Part II of the book).

Exercise 1.24. *Determine the local minima of the following functions. Which of them are global and/or strict minima, respectively.*

(i) $f(x) = senx + x/\sqrt{2}, x \in [-3, 6]$
(ii) $f(x) = |x| + |x - 1|, x \in \mathbb{R}$
(iii) $f(x_1, x_2) = x_1^2 + x_2^4, x \in \mathbb{R}^2$
(iv) $f(x_1, x_2) = x_1(x_1 - 1)(x_1 - 2), x \in \mathbb{R}^2$.

Illustrate the functions geometrically!

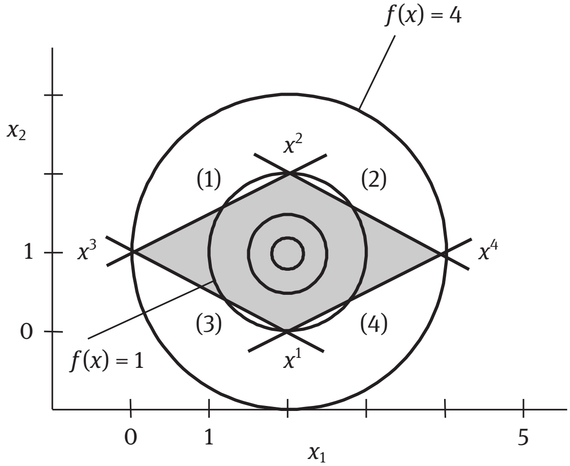


Fig. 1.10. Local and global minimum points.

1.4 References for Chapter 1

Special forms of nonlinear problems and applications are to be found in Bracken and McCormick (1968) and more sophisticated models used in practice are described in Goffin and Rousseau (1982).

Optimization methods in engineering and mechanical design are studied in Fox (1971) and Johnson (1971).

Applications in optimal control theory can be found, for example, in Canon and Eaton (1966) and Canon, Cullum and Polak (1970). The model of Example 1.19 (construction of a road) is studied in Bazaraa, Sherali and Shetty (2006, Section 1.2.A) and in the article of Sacoman:
http://www.dco.fc.unesp.br/~sacoman/artigos/oti/oti07.html.

With respect to nonlinear models in electric power production, in particular the problem of economic dispatch see, for example, Abou-Taleb et al. (1974), Sasson and Merrill (1974), the recent article by Da Silva et al. (2002) and the site of the National Seminar on Production and Transmission of Electricity (SNPTEE):
http://www.itaipu.gov.br/cd_snptee/xvsnptee/grupoiv/gat02.pdf.

Transportation problems with various types of nonlinear objective functions are considered, for example, in Michalewicz (1991).

Applications of the location model in the modernization of telegraph facilities and in determining the optimum location of an oil refinery in northeastern Brazil are described in the articles of Rosa and Thomaz et al.:
http://www.eps.ufsc.br/disserta96/altamir/docs/capit1.doc
http://www.nupeltd.ufc.br/nupeltd_artigo_refinaria.pdf

An article by Corrêa and Milani on the traffic regulation in subway lines can be found in the site:
http://www.fee.unicamp.br/revista_sba/vol12/v12a305.pdf.

Models of nonlinear optimization in the management of water resources are studied in Haimes (1977) and Yu and Haimes (1974).

Planning and optimization of water resources systems are actual problems, as is shown by the collection of Porto (2002) and the sites:
http://www.ct.ufpb.br/serea/download/resumenes_serea_espanha.pdf, http://www.ctufpb.br/serea/trabalhos/A14_06.pdf, http://www.adjacir.net/pesquisas.htm.

Generalized separable problems are considered in Luenberger (2003, Section 13.3) and Bazaraa, Sherali and Shetty (2006, Section 11.3).

Finally, an introduction into semiinfinite programming with several applications in mathematical physics, economics and engineering can be found in Still (2004). Applications of this model are also described in a thesis, available at:
https://repositorium.sdum.uminho.pt/handle/1822/203.

With respect to the definition of the semiinfinite model see the "Mathematical Programming Glossary" at:
http://carbon.cudenver.edu/hgreenbe~/glossary/index.php?page=V.html.

Part I: **Theoretical foundations**

2 Optimality conditions

In this chapter we will introduce some optimality conditions which are used to identify a local minimum. In formulating such conditions the concept of feasible direction is essential.

2.1 Feasible directions

A direction is called feasible at a given point of a set M, if a sufficiently small shift in this direction does not leave M. We obtain the following mathematically rigorous definition:

Definition 2.1. Given a set $M \subset \mathbb{R}^n$ and a point $x^* \in M$. A vector $d \in \mathbb{R}^n \backslash \{0\}$ is called a *feasible direction* at x^* if a $\lambda_0 > 0$ exists such that

$$x^* + \lambda d \in M$$

for $0 < \lambda \leq \lambda_0$.

The set of feasible directions at a point x^* will be denoted by $Z(x^*)$. Figure 2.1 illustrates feasible and nonfeasible directions at a given point.

Obviously, at an interior point of M each direction is feasible, hence any $d \in \mathbb{R}^2 \backslash \{0\}$ is feasible in x^2 (see Figure 2.1 (a)). At point x^1 the directions d^1, d^2 and d^3 are feasible, and d^4, d^5 and d^6 are not. In particular, d^4 coincides with the tangent T of the boundary curve of M at point x^1. In Figure 2.1 (b) the directions $d^1, \ldots, d^4$ are feasible at the point x^* and d^5 is not.

For $d \in Z(x^*)$ and $\alpha > 0$ the direction αd is also feasible, i.e. $Z(x^*)$ is a cone, called the *cone of feasible directions*. As Figure 2.1 (b) illustrates, $Z(x^*)$ need not be convex.

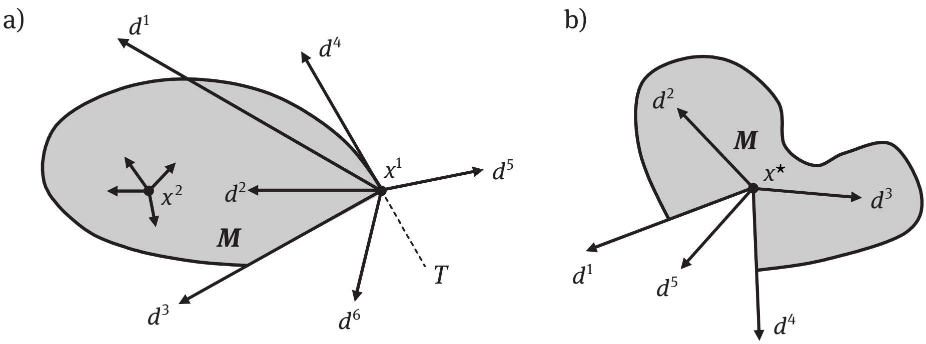


Fig. 2.1. Feasible and nonfeasible directions.

In the following we will characterize the cone of feasible directions in the case when M is of the form $M = \{x \in \mathbb{R}^n | g_i(x) \le 0, i = 1, \dots, m\}$ (see (1.2)).

Example 2.2.

(a) Let

$$M = \{x \in \mathbb{R}^2 | g_1(x_1, x_2) := x_2 - 2 \le 0\,,$$
$$g_2(x_1, x_2) := 1 + (x_1 - 1)^2 - x_2 \le 0\}\,.$$

For $x^* = (1, 1)^T \in M$, the cone of feasible directions is

$$Z(x^*) = \{d \in \mathbb{R}^2 | d_2 > 0\}\,,$$

i.e. a nonzero vector $d \in \mathbb{R}^2$ is a feasible direction, if and only if it belongs to the *open* half-plane S shown in Figure 2.2.

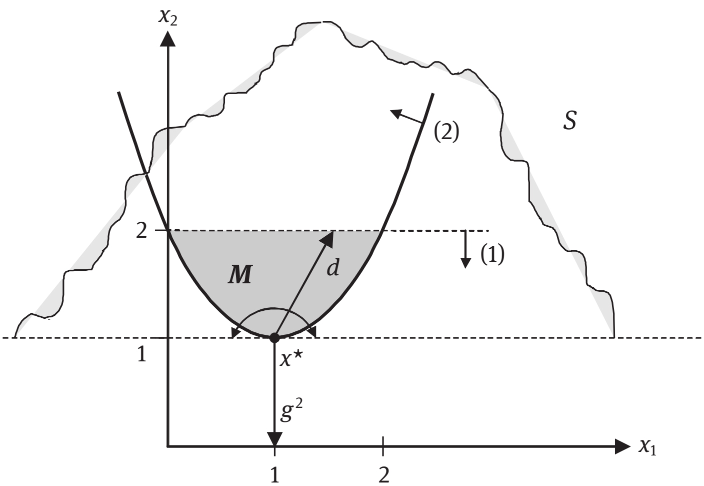


Fig. 2.2. Cone $Z(x^*)$ of Example 2.2 (a).

(b) Let

$$M = \{x \in \mathbb{R}^2 | g_1(x_1, x_2) := \frac{-x_1}{2} + x_2 \le 0,$$
$$g_2(x_1, x_2) := -x_1 - x_2 + 3 \le 0,$$
$$g_3(x_1, x_2) := x_1 - 3 \le 0\}\,.$$

For $x^* = (2, 1)^T \in M$ the cone of feasible directions is

$$Z(x^*) = \{d \in \mathbb{R}^2 | d_1 > 0, -1 \le d_2/d_1 \le 1/2\}\,.$$

A vector $d \in \mathbb{R}^2 \backslash \{0\}$ is a feasible direction at x^*, if and only if it belongs to the *closed* cone C shown in Figure 2.3.

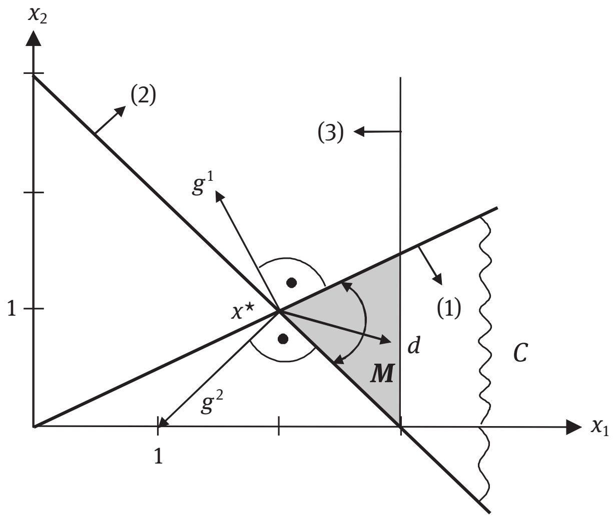


Fig. 2.3. Cone $Z(x^*)$ of Example 2.2 (b).

Definition 2.3. Let $M = \{x \in \mathbb{R}^n | g_i(x) \leq 0, i = 1, \dots, m\}$ and $x^* \in M$. A constraint $g_i(x) \leq 0$ with $g_i(x^*) = 0$ is called *active at* x^*. The set

$$A(x^*) := \{i \in \{1, \dots, m\} | g_i(x^*) = 0\}$$

is called the *set of active constraints at* x^*.

In Example 2.2 (a) only the constraint $g_2(x_1, x_2) \leq 0$ is active, while the constraints $g_1(x_1, x_2) \leq 0$ and $g_2(x_1, x_2) \leq 0$ are active in Example 2.2 (b). Figures 2.2 and 2.3 illustrate that only constraints which are active at x^* determine the cone $Z(x^*)$.

In order to characterize the cone of feasible directions, the vectors $\operatorname{grad} g_i(x^*)$ of the active constraints can be employed. (Note that the vector $\operatorname{grad} g_i(x^*)$ is perpendicular to the tangent to the curve $g_i(x) = 0$ at the point x^*). In the Example 2.2 (a) only the constraint (2) is active and we have $\operatorname{grad} g_2(x_1, x_2) = (2(x_1 - 1), -1)^T$, so the gradient of the function g_2 at point x^* is $g^2 := \operatorname{grad} g_2(x^*) = (0, -1)^T$. We observe that d is a feasible direction at x^*, if and only if the angle between g^2 and d is *greater than* 90 degrees (Figure 2.2). In Example 2.2 (b) the two constraints are active and we obtain $g^1 := \operatorname{grad} g_1(x^*) = (-1/2, 1)^T$ and $g^2 := \operatorname{grad} g_2(x^*) = (-1, -1)^T$. A direction d is feasible, if and only if d forms an angle *greater than or equal* to 90 degrees with both gradients g^1 and g^2 (Figure 2.3).

Formally, for constraints active at x^*, a feasible direction d must satisfy

$$d^T \operatorname{grad} g_i(x^*) < 0 \quad \text{or} \quad d^T \operatorname{grad} g_i(x^*) \leq 0\,,$$

respectively. These considerations result in the following concepts.

Definition 2.4. The set

$$L(x^*) = \{d \in \mathbb{R}^n \backslash \{0\} | d^T \operatorname{grad} g_i(x^*) \le 0, \quad i \in A(x^*)\}$$

is called the *cone of linearized constraints* at x^*. Furthermore we define:

$$L_0(x^*) = \{d \in \mathbb{R}^n \backslash \{0\} | d^T \operatorname{grad} g_i(x^*) < 0, \quad i \in A(x^*)\},$$

which is the interior of the cone $L(x^*)$.

One can easily verify that it holds $Z(x^*) = L_0(x^*)$ in Example 2.2 (a) and $Z(x^*) = L(x^*)$ in Example 2.2 (b). The following two theorems characterize $Z(x^*)$ in the general case.

Theorem 2.5. *Let $M = \{x \in \mathbb{R}^n | g_i(x) \le 0, i = 1, \dots, m\}$ with differentiable functions g_i and $x^* \in M$. We have*

$$L_0(x^*) \subset Z(x^*) \subset L(x^*).$$

Proof. (a) Let $d \in L_0(x^*)$, i.e. $d^T \operatorname{grad} g_i(x^*) < 0$ for all $i \in A(x^*)$. For any $i \in A(x^*)$ we consider the Taylor approximation

$$g_i(x^* + \lambda d) = g_i(x^*) + \lambda d^T \operatorname{grad} g_i(x^*) + R_1(\lambda). \tag{2.1}$$

Since $g_i(x^*) = 0$, we have

$$\frac{1}{\lambda} g_i(x^* + \lambda d) = d^T \operatorname{grad} g_i(x^*) + \frac{1}{\lambda} R_1(\lambda)$$

for $\lambda > 0$.
Considering the above assumption $d^T \operatorname{grad} g_i(x^*) < 0$ and using the fact that

$$\lim_{\lambda \to 0} \frac{1}{\lambda} R_1(\lambda) = 0,$$

we obtain

$$g_i(x^* + \lambda d) < 0 \tag{2.2}$$

for a sufficiently small $\lambda > 0$ and $i \in A(x^*)$. Since $g_i(x^*) < 0$ for $i \notin A(x^*)$, (2.2) is also satisfied for $i \notin A(x^*)$ and a sufficiently small $\lambda > 0$, so the vector d is a feasible direction.

(b) From the same Taylor approximation we obtain the following: if $d \notin L(x^*)$, i.e. $d^T \operatorname{grad} g_k(x^*) > 0$ for an index $k \in A(x^*)$, it holds

$$g_k(x^* + \lambda d) > 0$$

for a sufficiently small $\lambda > 0$, hence d is not feasible. □

We now consider the case where the set M is defined by linear functions:

Theorem 2.6. *Let* $M = \{x \in \mathbb{R}^n | g_i(x) := a^{iT}x + b_i \le 0, i = 1, \dots, m\}$ *with* $a^i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$ *and* $x^* \in M$. *It holds*

$$Z(x^*) = L(x^*)\,.$$

Proof. By Theorem 2.5, it suffices to prove that $L(x^*) \subset Z(x^*)$. Let $d \in L(x^*)$, i.e.

$$d^T a^i \le 0 \tag{2.3}$$

for all $i \in A(x^*)$. For each $i \in A(x^*)$, we have

$$\begin{aligned} g_i(x^* + \lambda d) &= a^{iT}(x^* + \lambda d) + b_i \\ &= a^{iT}x^* + b_i + \lambda a^{iT} d \\ &= \lambda d^T a^i\,, \end{aligned} \tag{2.4}$$

since $a^{iT}x^* + b_i = g_i(x^*) = 0$. The relations (2.3) and (2.4) imply

$$g_i(x^* + \lambda d) \le 0 \tag{2.5}$$

for any $\lambda > 0$. As (2.5) is also satisfied for the nonactive constraints and a sufficiently small $\lambda > 0$ (compare with the end of the proof of Theorem 2.5, part a), we obtain $d \in Z(x^*)$. □

Example 2.7. We consider

$$\begin{aligned} M = \{x \in \mathbb{R}^2 | g_1(x_1, x_2) &:= -2x_1 + x_2 - 2 \le 0\,, \quad g_2(x_1, x_2) := x_1 + x_2 - 5 \le 0\,, \\ g_3(x_1, x_2) &:= x_1 - 4x_2 \le 0\,, \quad g_4(x_1, x_2) := -x_1 \le 0\} \end{aligned}$$

with $x^* = (1, 4)^T$ (Figure 2.4).

Only the constraints (1) and (2) are active in x^*, i.e. $A(x^*) = \{1, 2\}$. We have $\operatorname{grad} g_1(x^*) = (-2, 1)^T$ and $\operatorname{grad} g_2(x^*) = (1, 1)^T$. Theorem 2.6 implies that

$$Z(x^*) = L(x^*) = \{d \in \mathbb{R}^2 \setminus \{0\} | -2d_1 + d_2 \le 0, d_1 + d_2 \le 0\}\,.$$

The set $Z(x^*)$ is represented by the cone C shown in Figure 2.4.

Theorem 2.5 states that $L_0(x^*) \subset Z(x^*) \subset L(x^*)$ is satisfied in general. In Example 2.2 we already encountered the extreme cases $Z(x^*) = L_0(x^*)$ and $Z(x^*) = L(x^*)$. The following example shows that $L_0(x^*) \ne Z(x^*) \ne L(x^*)$ is also possible.

Example 2.8. Consider the set

$$M = \left\{x \in \mathbb{R}^2 | g_1(x_1, x_2) := x_1 - x_1^2/4 - x_2 \le 0, g_2(x_1, x_2) := -4x_1 + x_1^2 + x_2 \le 0\right\}$$

with $x^* = (0, 0)^T \in M$ (Figure 2.5).

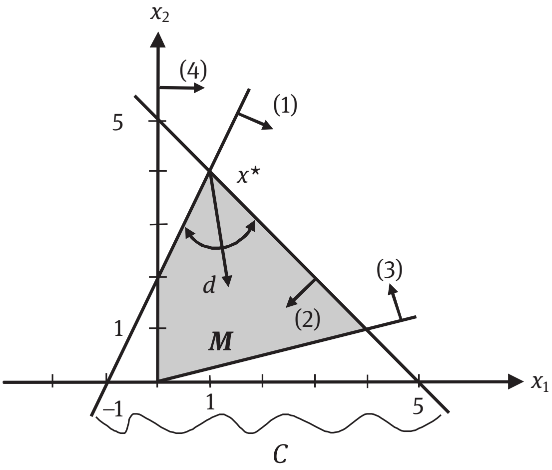


Fig. 2.4. Cone of feasible directions for linear constraints.

We have $\operatorname{grad} g_1(x^*) = (1, -1)^T$ and $\operatorname{grad} g_2(x^*) = (-4, 1)^T$, hence

$$\begin{aligned} L(x^*) &= \left\{ d \in \mathbb{R}^2 \backslash \{0\} \middle| (d_1, d_2) \begin{pmatrix} 1 \\ -1 \end{pmatrix} \le 0 \,, \quad (d_1, d_2) \begin{pmatrix} -4 \\ 1 \end{pmatrix} \le 0 \right\} \\ &= \left\{ d \in \mathbb{R}^2 \backslash \{0\} | d_1 \le d_2 \le 4d_1 \right\} \,. \end{aligned}$$

The cone $L(x^*)$ is bounded by the lines $x_2 = x_1$ and $x_2 = 4x_1$ (Figure 2.5).

Furthermore we have:

$$L_0(x^*) = \{d \in \mathbb{R}^2 | d_1 < d_2 < 4d_1\} \,.$$

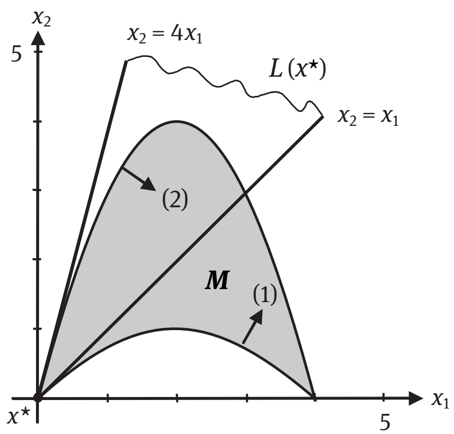


Fig. 2.5. Cone of feasible directions with $L_0(x^*) \ne Z(x^*) \ne L(x^*)$.

It is easily verified that

$$Z(x^*) = \{d \in \mathbb{R}^2 | d_1 \le d_2 < 4d_1\}$$

which is the cone $L(x^*)$ except the points on the ray $x_2 = 4x_1$ with $x_1 > 0$.

Hence, $L_0(x^*) \neq Z(x^*) \neq L(x^*)$.

The concept of feasible direction plays an important role in the construction of solution procedures: a basic idea of nonlinear programming is to determine an optimal point iteratively (see the introduction of Chapter 7 and Section 8.1). Thus, in a minimization problem, many procedures start with a feasible solution and improve the current point in any iteration by moving along a feasible descent direction, i.e. a feasible direction, in which the values of the objective function initially decrease.

Exercise 2.9. *Determine the cone of feasible directions for the following examples! Apply Theorems 2.5 and 2.6! In cases (a) and (b) illustrate geometrically the set M, the point x^* and the vectors $g^i := \operatorname{grad} g_i(x^*)$ for $i \in A(x^*)$!*

(a) $M = \{x \in \mathbb{R}^2 | x_1 + 4x_2 \le 20, x_1 + x_2 \le 8, 2x_1 + x_2 \le 14\}$ *with* $x^* = (4,4)^T$.

(b) $M = \{x \in \mathbb{R}^2 | (x_2 - 1) + (x_1 - 3)^3 \le 0, -(x_2 - 1) + (x_1 - 3)^3 \le 0\}$ *with* $x^* = (3,1)^T$.

(c) $M = \{x \in \mathbb{R}^2 | x_1^2 + x_2^2 \le 0\}$ *with* $x^* = (0,0)^T$.

2.2 First and second-order optimality conditions

We return to study the problem

$$\min_{x \in M} f(x)$$

with $M \subset \mathbb{R}^n$, where f is twice continuously differentiable. We will develop conditions which are necessary or sufficient, respectively, for a *local* minimum of the function f. The so-called first-order conditions employ only first partial derivatives while the second-order conditions require the first and second derivatives. In the latter case we need the concept of the Hessian matrix

$$Hf(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}.$$

Since f is twice continuously differentiable, $Hf(x)$ is symmetric.

Theorem 2.10 (First-order necessary optimality condition). *If x^* is a local minimum point of f over M,*

$$d^T \operatorname{grad} f(x^*) \ge 0$$

is satisfied for any feasible direction d at x^*.

Proof. Let x^* be a local minimum point of f. For any feasible direction d and any sufficiently small $\lambda > 0$ we have

$$f(x^* + \lambda d) \geq f(x^*)\,.$$

Using the Taylor expansion, we obtain

$$f(x^*) + \lambda d^T \operatorname{grad} f(x^*) + R_1(\lambda) \geq f(x^*)\,,$$

thus

$$d^T \operatorname{grad} f(x^*) + \frac{R_1(\lambda)}{\lambda} \geq 0\,. \tag{2.6}$$

Since $\lim_{\lambda\to 0} \frac{R_1(\lambda)}{\lambda} = 0$, equation (2.6) implies that $d^T \operatorname{grad} f(x^*) \geq 0$. □

Example 2.11. We consider the problem

$$\min (x_1+1)^2 + (x_2-1)^2 \quad x_1, x_2 \geq 0\,.$$

As $f(x_1, x_2)$ represents the square of the distance between $(-1, 1)^T$ and $(x_1, x_2)^T$, the point $x^* := (0, 1)^T$ is a local minimum point (and also the unique global minimum point) of f over M (Figure 2.6).

We have

$$\operatorname{grad} f(x^*) = \begin{pmatrix} 2(x_1+1) \\ 2(x_2-1) \end{pmatrix}\bigg|_{x=x^*} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}.$$

The cone of feasible directions at x^* is

$$Z(x^*) = \{d \in \mathbb{R}^2 \backslash \{0\} | d_1 \geq 0\}\,,$$

and any feasible direction d satisfies

$$d^T \operatorname{grad} f(x^*) = (d_1, d_2)\left(\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right) = 2d_1 \geq 0\,. \tag{2.7}$$

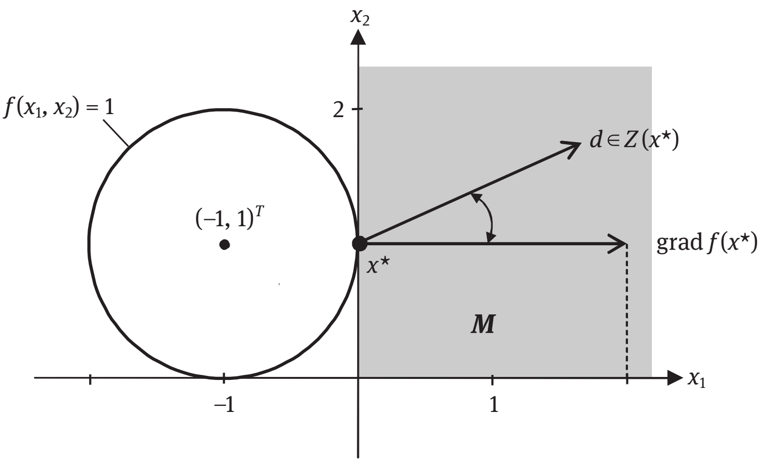


Fig. 2.6. Geometric illustration of Example 2.11.

The necessary condition (2.7) means geometrically that any feasible direction forms an angle $\leq$ 90 degrees with the vector $\operatorname{grad} f(x^*)$ (see Figure 2.6).

Theorem 2.12 (First-order necessary optimality condition for interior points).
If the inner point x^* of M is a local minimum point of f over M, then

$$\operatorname{grad} f(x^*) = 0 .$$

Proof. At interior points all directions are feasible. Theorem 2.10 implies that

$$d^T \operatorname{grad} f(x^*) \geq 0$$

for all vectors $d \in \mathbb{R}^n$. Thus, $\operatorname{grad} f(x^*) = 0$. □

Example 2.13. Obviously $x^* = (2, 3)^T$ is a local (and global) minimum point of

$$f(x_1, x_2) = 5(x_1 - 2)^2 + 2(x_2 - 3)^4$$

over $\mathbb{R}^2$. The necessary condition for an interior point is satisfied:

$$\operatorname{grad} f(x^*) = \begin{pmatrix} 10(x_1 - 2) \\ 8(x_2 - 3)^3 \end{pmatrix} \Bigg|_{x=x^*} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

Exercise 2.14. *Given the following functions over $M = \mathbb{R}^2$:*
(a) $f(x_1, x_2) = x_1^2 + 2x_2^3$
(b) $f(x_1, x_2) = x_2 e^{x_1}$
(c) $f(x_1, x_2) = \cos x_1 + x_2^2$

Determine in each case the points that satisfy the necessary condition of Theorem 2.12.

Exercise 2.15. *Given the problem*

$$\min -(x_1 - 1)^2 - (x_2 - 1)^2$$
$$x_1, x_2 \geq 0 .$$

(a) *Determine all points x^* of the feasible region*

$$M := \{x \in \mathbb{R}^2 | x_1, x_2 \geq 0\}$$

that satisfy the necessary condition of Theorem 2.10. Consider the four cases:

$$x_1^* = 0, \quad x_2^* = 0; \qquad x_1^* = 0, \quad x_2^* > 0;$$
$$x_1^* > 0, \quad x_2^* = 0; \qquad x_1^* > 0, \quad x_2^* > 0.$$

(b) *Which of the points determined in (a) are in fact local minima? Interpret the problem geometrically (see Example 2.11). Is there a global minimum point?*

Theorem 2.16 (Second-order necessary optimality conditions). *If x^* is a local minimum point of f over M, the following conditions are satisfied for any feasible direction d at x^*:*
(i) $d^T \operatorname{grad} f(x^*) \geq 0$
(ii) $d^T \operatorname{grad} f(x^*) = 0$ *implies* $d^T Hf(x^*) d \geq 0$.

Proof. Condition (i) is the assertion of Theorem 2.10. We now prove condition (ii): Let x^* be a local minimum point of f over M. Using the Taylor approximation of second order, we obtain

$$f(x^* + \lambda d) = f(x^*) + \lambda d^T \operatorname{grad} f(x^*) + \frac{\lambda^2}{2} d^T Hf(x^*) d + R_2(\lambda) \geq f(x^*)$$

for any feasible direction d and a sufficiently small $\lambda > 0$. For $d^T \operatorname{grad} f(x^*) = 0$ the inequality implies

$$\frac{\lambda^2}{2} d^T Hf(x^*) d + R_2(\lambda) \geq 0\,,$$

thus

$$d^T Hf(x^*) d + 2\frac{R_2(\lambda)}{\lambda^2} \geq 0$$

for d and λ as above. Finally, the relation

$$\lim_{\lambda \to 0} \frac{R_2(\lambda)}{\lambda^2} = 0$$

implies that $d^T Hf(x^*) d \geq 0$ for any feasible direction d. □

Example 2.17. We verify that the local minimum point $x^* = (0, 1)^T$ of Example 2.11 also satisfies the second condition of Theorem 2.16.

In fact we have

$$Hf(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

for all x. Thus

$$d^T Hf(x^*) d = 2d_1^2 + 2d_2^2 \geq 0$$

is satisfied for any feasible direction d.

Example 2.18. Given the problem

$$\min_{x_1, x_2 \geq 0} x_1 x_2\,.$$

A point $(x_1, x_2)^T$ of the feasible region $M := \{x \in \mathbb{R}^2 | x_1, x_2 \geq 0\}$ is a local minimum point if and only if $x_1 = 0$ or $x_2 = 0$. (These points are also global minima.) We verify that $x^* = (c, 0)^T$ with $c > 0$ satisfies the necessary conditions of Theorem 2.16:

(i) We have $\operatorname{grad} f(x_1, x_2) = (x_2, x_1)^T$, thus

$$\operatorname{grad} f(x^*) = (0, c)^T \text{ and } Z(x^*) = \{d \in \mathbb{R}^2 | d_2 \geq 0\}.$$

Hence,

$$d^T \operatorname{grad} f(x^*) = (d_1, d_2) \begin{pmatrix} 0 \\ c \end{pmatrix} = cd_2 \geq 0$$

for any feasible direction d.

(ii) We have $Hf(x) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ for all x. Let $(d_1, d_2)^T$ be a feasible direction such that

$$d^T \operatorname{grad} f(x^*) = (d_1, d_2) \begin{pmatrix} 0 \\ c \end{pmatrix} = cd_2 = 0\,.$$

Hence $d_2 = 0$ and this direction satisfies

$$d^T Hf(x^*) d = (d_1, d_2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = 2d_1 d_2 = 0\,.$$

Note that unlike Example 2.17, not every feasible direction d satisfies $d^T Hf(x^*) d \geq 0$.

Theorem 2.19 (Second-order necessary optimality conditions for interior points). *If the inner point x^* of M is a local minimum point of f over M, the following conditions are satisfied:*

(i) $\operatorname{grad} f(x^*) = 0$

(ii) $d^T Hf(x^*) d \geq 0$ *for all* $d \in \mathbb{R}^n$.

Proof. Theorem 2.12 implies condition (i) above. As all directions are feasible in an interior point, condition (ii) above follows from condition (ii) of Theorem 2.16. □

We observe that a point that satisfies the conditions of Theorem 2.19 need not be a local minimum point, i.e. these conditions are not sufficient. For example, consider the function $f(x) = x^3$ over $\mathbb{R}$. The point $x^* = 0$ satisfies these conditions, since $f'(x^*) = 3x^{*2} = 0$ and $f''(x^*) = 6x^* = 0$, while x^* is not a local minimum point of f over $\mathbb{R}$.

Exercise 2.20. *Show that the conditions of Theorem 2.19 are satisfied at the point $x^* = (0, 0)^T$ of the function*

$$f(x_1, x_2) = x_1^3 x_2^3$$

over $\mathbb{R}^2$, and that x^* is not a local minimum point.

Without proof we still mention the following criterion:

Theorem 2.21 (Second-order sufficient optimality conditions for interior points). *An interior point x^* of M, satisfying the conditions*
(i) $\operatorname{grad} f(x^*) = 0$
(ii) $d^T Hf(x^*) d > 0$ *for all* $d \in \mathbb{R}^n \setminus \{0\}$

is a local minimum point of the function f *over* M.

Exercise 2.22. *Given the following functions over* $\mathbb{R}^2$*:*
(a) $f(x_1, x_2) = 3(x_1 - 2)^2 + 5(x_2 - 2)^2$
(b) $f(x_1, x_2) = \cos x_1 + x_2^2$ *(see Exercise 2.14 (c)).*

Determine in each case the points that satisfy the necessary conditions of Theorem 2.19. Which of these points represent local minima?

Exercise 2.23. *Determine all feasible points of the problem in Exercise 2.15, which satisfy the necessary conditions of Theorem 2.16.*

The condition (ii) of Theorem 2.21 means that $Hf(x^*)$ is positive definite and condition (ii) of Theorem 2.19 means that $Hf(x^*)$ is positive semidefinite. We recall the necessary definitions of linear algebra:

Definition 2.24.
(i) A symmetric matrix $Q \in \mathbb{R}^{n\times n}$ is called *positive semidefinite*, if $x^T Q x \geq 0$ for all $x \in \mathbb{R}^n$.
(ii) Q is called *positive definite*, if $x^T Q x > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$.

Obviously a positive definite matrix is also positive semidefinite. Moreover we define:

Definition 2.25. A symmetric matrix $Q \in \mathbb{R}^{n\times n}$ is called *negative definite* (*negative semidefinite*) if $-Q$ is positive definite (positive semidefinite). A matrix that is neither positive semidefinite nor negative semidefinite is called *indefinite*.

Definition 2.26. Let $Q = (q_{i,j})_{i,j=1,\ldots,n}$ be a quadratic matrix. The determinants of the form

$$\begin{vmatrix} q_{i_1,i_1} & q_{i_1,i_2} & \cdots & q_{i_1,i_k} \\ q_{i_2,i_1} & q_{i_2,i_2} & \cdots & q_{i_2,i_k} \\ \vdots & \vdots & & \vdots \\ q_{i_k,i_1} & q_{i_k,i_2} & \cdots & q_{i_k,i_k} \end{vmatrix} \tag{2.8}$$

with $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ and $k \in \{1, \dots, n\}$ are called *principal minors* of the matrix Q. In particular, the determinants of the form

$$\begin{vmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,k} \\ q_{2,1} & q_{2,2} & \cdots & q_{2,k} \\ \vdots & \vdots & & \vdots \\ q_{k,1} & q_{k,2} & \cdots & q_{k,k} \end{vmatrix} \tag{2.9}$$

with $k \in \{1, \dots, n\}$ are called *successive principal minors* of Q.

Example 2.27. For the matrix

$$Q = \begin{pmatrix} q_{1,1} & q_{1,2} & q_{1,3} \\ q_{2,1} & q_{2,2} & q_{2,3} \\ q_{3,1} & q_{3,2} & q_{3,3} \end{pmatrix} = \begin{pmatrix} 1 & 3 & 5 \\ 2 & -2 & 4 \\ 3 & 0 & 6 \end{pmatrix},$$

we obtain the following principal minors:

$$\begin{array}{llll} k = 1, & i_1 = 1: & & q_{1,1} = 1 \\ k = 1, & i_1 = 2: & & q_{2,2} = -2 \\ k = 1, & i_1 = 3: & & q_{3,3} = 6 \\ k = 2, & i_1 = 1, & i_2 = 2: & \begin{vmatrix} q_{1,1} & q_{1,2} \\ q_{2,1} & q_{2,2} \end{vmatrix} = 1 \cdot (-2) - 2 \cdot 3 = -8 \\ k = 2, & i_1 = 1, & i_2 = 3: & \begin{vmatrix} q_{1,1} & q_{1,3} \\ q_{3,1} & q_{3,3} \end{vmatrix} = 1 \cdot 6 - 3 \cdot 5 = -9 \\ k = 2, & i_1 = 2, & i_2 = 3: & \begin{vmatrix} q_{2,2} & q_{2,3} \\ q_{3,2} & q_{3,3} \end{vmatrix} = -2 \cdot 6 - 0 \cdot 4 = -12 \\ k = 3, & i_1 = 1, & i_2 = 2, \quad i_3 = 3: & \det(Q) = 18 \end{array}$$

The successive principal minors are 1, -8 and 18.

Definition 2.28. Let $Q \in \mathbb{R}^{n \times n}$. A number $\lambda \in \mathbb{R}$ is called an *eigenvalue* of Q, if a vector $y \in \mathbb{R}^n$ with $y \neq 0$ exists such that

$$Qy = \lambda y\,. \tag{2.10}$$

The vector y in (2.10) is called the *eigenvector* associated with the eigenvalue λ. Due to the equivalences

$$Qy = \lambda y \Leftrightarrow (Q - \lambda I)y = 0 \Leftrightarrow (Q - \lambda I) \text{ is singular } \Leftrightarrow |Q - \lambda I| = 0\,, \tag{2.11}$$

the eigenvalues of Q are the roots of the polynomial $p(\lambda) = |Q - \lambda I|$ in λ, called the *characteristic polynomial* of Q.

It can be shown that all eigenvalues of a symmetric matrix $Q \in \mathbb{R}^{n\times n}$ are real values. The following theorem is used to check if a given matrix has one of the above features.

Theorem 2.29. *Let $Q = (q_{i,j})_{i,j=1,\ldots,n}$ be a symmetric matrix.*
(i) *If Q is positive definite, then $q_{i,i} > 0$ for $i = 1, \ldots, n$.*
If Q is positive semidefinite, then $q_{i,i} \geq 0$ for $i = 1, \ldots, n$.
(ii) *Q is positive definite, if and only if all successive principal minors are positive.*
Q is positive semidefinite, if and only if all principal minors are nonnegative.
(iii) *Q is positive definite, if and only if all eigenvalues are positive.*
Q is positive semidefinite, if and only if all eigenvalues are nonnegative.

Exercise 2.30. *Which of the following matrices is positive (semi)definite?*

$$\text{(a)} \begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix} \quad \text{(b)} \begin{pmatrix} 9 & 6 \\ 6 & 4 \end{pmatrix} \quad \text{(c)} \begin{pmatrix} 2 & -4 \\ -4 & 3 \end{pmatrix} \quad \text{(d)} \begin{pmatrix} 15 & 3 & 2 \\ 3 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \quad \text{(e)} \begin{pmatrix} 13 & 2 & 1 \\ 2 & -3 & 4 \\ 1 & 0 & 5 \end{pmatrix}.$$

Exercise 2.31. *Determine all principal minors of the matrix*

$$\begin{pmatrix} 2 & 3 & -1 & 5 \\ 3 & 8 & 2 & 1 \\ -1 & 2 & 4 & 6 \\ 5 & 1 & 6 & 15 \end{pmatrix}.$$

Which of them are successive principal minors? The matrix is positive (semi)definite?

Exercise 2.32. *Show that the function*

$$f(x_1, x_2) = \frac{3}{2}x_1^2 + 5x_2^2 + 4x_1x_2$$

over $\mathbb{R}^2$ has the unique local minimum point $x = 0$.

We emphasize the following fact: the partial derivatives of a function f at a point x^* only allow statements about the "behavior" of f in a neighborhood of x^*. Therefore, in the general case, the above optimality conditions can only identify a *local* minimum and are not useful to decide whether a given point is a global minimum point. The situation is different in the case of minimizing a convex function, studied in the next chapter.

3 The convex optimization problem

In this chapter we study the minimization of convex and concave functions. Such functions arise in many practical applications, in particular in optimal location (Examples 1.18 and 3.54), optimal control (Example 1.19), parameter estimation (Example 3.52) and investment planning (Examples 1.5 and 4.35). Convex functions have an important feature: every local minimum point is also a global minimum point. When the function f is convex, we can then determine a global minimum point with the aid of the optimality criteria of Chapter 2.

After the study of convex sets (Section 3.1) we define convex and concave functions (Section 3.2). Differentiable convex functions are characterized in Section 3.3. Then we introduce the concepts of subgradient and directional derivative (Section 3.4). Finally, Section 3.5 is devoted to the determination of minima of convex and concave functions.

3.1 Convex sets

In this section we introduce the concepts of "convex set", "extreme point" and "convex hull".

We denote the *line segment joining* the points x^1 and x^2 by

$$\begin{aligned}[x^1, x^2] &:= \{x^1 + \lambda(x^2 - x^1) \,|\, 0 \le \lambda \le 1\} \\ &= \{(1-\lambda)x^1 + \lambda x^2 \,|\, 0 \le \lambda \le 1\}\end{aligned}$$

(see Figure 3.1) and

$$(x^1, x^2) := [x^1, x^2] \setminus \{x^1, x^2\}$$

denotes the relative interior of $[x^1, x^2]$.

Definition 3.1. A set $K \subset \mathbb{R}^n$ is called *convex*, if K contains the line segment $[x^1, x^2]$ for any two points $x^1, x^2 \in K$.

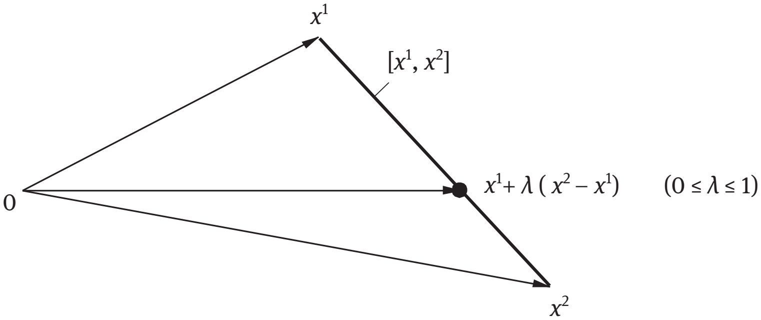


Fig. 3.1. Line segment $[x^1, x^2]$.

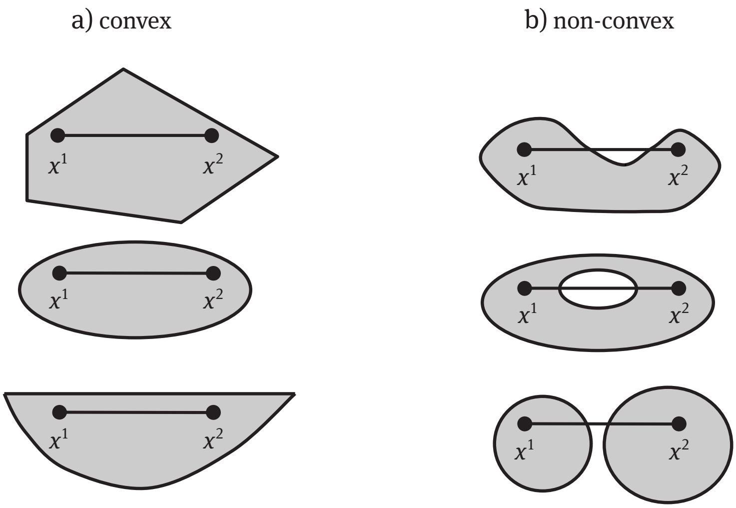


Fig. 3.2. Convex and nonconvex sets.

Some convex and nonconvex sets are illustrated in Figure 3.2. Note that for every set K in Figure 3.2 (b) exist points $x^1, x^2 \in K$ for which $[x^1, x^2]$ is not entirely contained in K. In particular, the union of two convex sets need not be convex and not even connected.

Comment 3.2. The intersection of two convex sets is convex.

Exercise 3.3. *Prove the previous statement.*

We now study some specific convex sets:

(a) A *half-space* of $\mathbb{R}^n$, i.e. a set of the form

$$S := \{x \in \mathbb{R}^n | a^T x \leq b\}$$

with $a \in \mathbb{R}^n$, $b \in \mathbb{R}$.
For example, in the case $n = 2$, $a = (-1, 1)^T$ and $b = 2$ we obtain the half-plane

$$S := \{(x_1, x_2)^T \in \mathbb{R}^2 | -x_1 + x_2 \leq 2\}$$

of $\mathbb{R}^2$.

(b) A *hyperplane* of $\mathbb{R}^n$, i.e. a set of the form

$$H := \{x \in \mathbb{R}^n | a^T x = b\}$$

with $a \in \mathbb{R}^n$, $b \in \mathbb{R}$.
For n, a and b as above we obtain the straight line

$$H := \{(x_1, x_2)^T \in \mathbb{R}^2 | -x_1 + x_2 = 2\} .$$

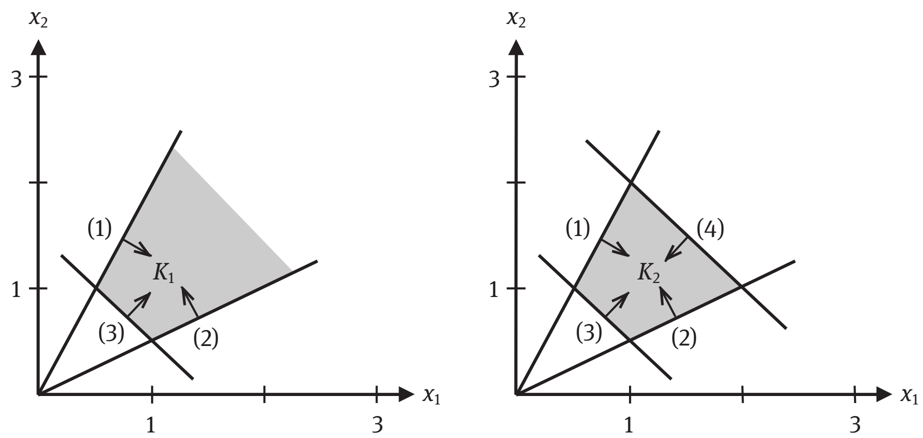


Fig. 3.3. Convex polyhedral sets.

(c) A set of the form

$$K := \{x \in \mathbb{R}^n | A^T x \leq b\}$$

with $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^m$. The set K is the intersection of the half-spaces

$$S_i := \{x \in \mathbb{R}^n | {a_i}^T x \leq b_i\}$$

for $i = 1, \ldots, m$, where a_i denotes the ith column of A and $b = (b_1, \ldots, b_m)^T$. Therefore K is convex (see Comment 3.2).

Definition 3.4. A set K of the form (c) above is called a *convex polyhedral set*. A bounded convex polyhedral set is called a *convex polytope*.

Figure 3.3 illustrates the convex polyhedral set

$$K_1 := \{(x_1, x_2)^T \in \mathbb{R}^2 | -2x_1 + x_2 \leq 0, x_1/2 - x_2 \leq 0, -x_1 - x_2 \leq -3/2\}$$

and the convex polytope

$$K_2 := \{(x_1, x_2)^T \in \mathbb{R}^2 | -2x_1 + x_2 \leq 0, x_1/2 - x_2 \leq 0, -x_1 - x_2 \leq -3/2, x_1 + x_2 \leq 3\}\,.$$

Definition 3.5. Let $K \in \mathbb{R}^n$ be a convex set. A point $x \in K$ is called an *extreme point* of $x \in K$ no two points $x^1, x^2 \in K$ exist with $x \in (x^1, x^2)$.

Clearly, an extreme point is always a boundary point. Therefore, open convex sets do not have extreme points. In the following we determine the extreme points of some

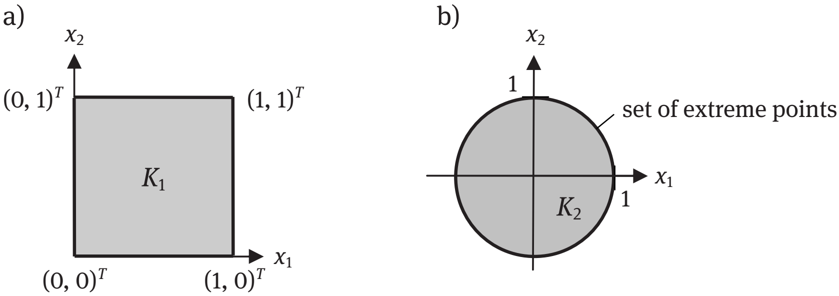


Fig. 3.4. Extreme points of convex sets.

convex sets (Figure 3.4):

(a) The extreme points of the convex polyhedron

$$K_1 := \{(x_1, x_2)^T \in \mathbb{R}^2 | 0 \le x_1, x_2 \le 1\}$$

are the four vertices $(0, 0)^T$, $(0, 1)^T$, $(1, 0)^T$ and $(1, 1)^T$ (Figure 3.4 (a)).

(b) The closed circle

$$K_2 := \{(x_1, x_2)^T \in \mathbb{R}^2 | x_1^2 + x_2^2 \le 1\}$$

has an infinite number of extreme points (Figure 3.4 (b)): any point of the boundary $\{(x_1, x_2)^T \in \mathbb{R}^2 | x_1^2 + x_2^2 = 1\}$ is an extreme point, as none of these points is contained in a line segment (x^1, x^2) with $x^1, x^2 \in K_2$.

(c) The open circle

$$\{(x_1, x_2)^T \in \mathbb{R}^2 | x_1^2 + x_2^2 < 1\}$$

has no extreme point.

Definition 3.6.

(i) Given the points $x^1, \dots, x^m \in \mathbb{R}^n$, a point

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

with $\lambda_i \ge 0$ for $i = 1, \dots, m$ and $\sum_{i=1}^m \lambda_i = 1$ is called a *convex combination* of the points $x^1, \dots, x^m$.

(ii) Given a set $K \subset \mathbb{R}^n$, the *convex hull* of K, denoted by $\operatorname{conv}(K)$ is the set of all convex combinations of points of K, i.e. $\operatorname{conv}(K)$ consists of all points of the form

$$x = \sum_{i=1}^{m} \lambda_i x^i ,$$

where m is a positive integer, $\lambda_i \ge 0$ for $i = 1, \dots, m$, $\sum_{i=1}^m \lambda_i = 1$ and $x^1, \dots, x^m \in K$.

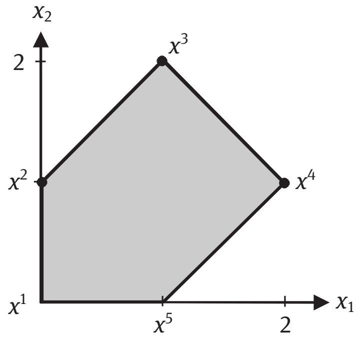


Fig. 3.5. Convex hull of a finite set.

Exercise 3.7.
(i) *Given a convex set $M \subset \mathbb{R}^n$ and points $x^1, \dots, x^m \in M$, show that M contains the convex hull of these points.*
(ii) *Show that* $\operatorname{conv}(K)$ *is a convex set for any set $K \subset \mathbb{R}^n$.*

Figure 3.5 shows the convex hull of the set $K = \{x^1, x^2, x^3, x^4, x^5\}$, where $x^1 = (0,0)^T$, $x^2 = (0,1)^T$, $x^3 = (1,2)^T$, $x^4 = (2,1)^T$ and $x^5 = (1,0)^T$. In Figure 3.6 the convex hulls of some nonfinite sets in the plane are illustrated.

In particular, Figure 3.6 illustrates the following statement:

Theorem 3.8. *Given the set $K \subset \mathbb{R}^n$. The convex hull of K is the smallest convex set that contains K, i.e. for each convex set M with $K \subset M$ holds* $\operatorname{conv}(K) \subset M$.

We note that the convex hull of a finite set K is a convex polytope (see Figure 3.5). Vice versa, every convex polytope is the convex hull of the finite set of its extreme points (vertices).

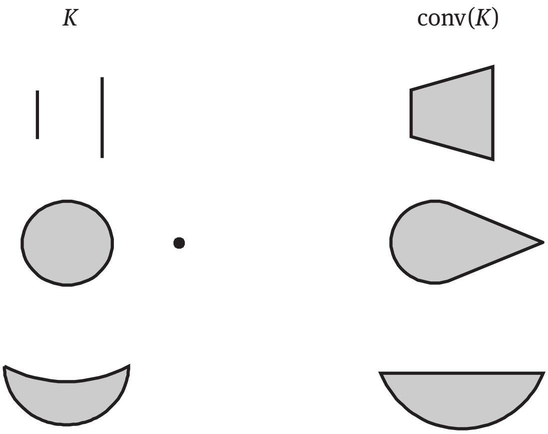


Fig. 3.6. Convex hulls of nonfinite sets.

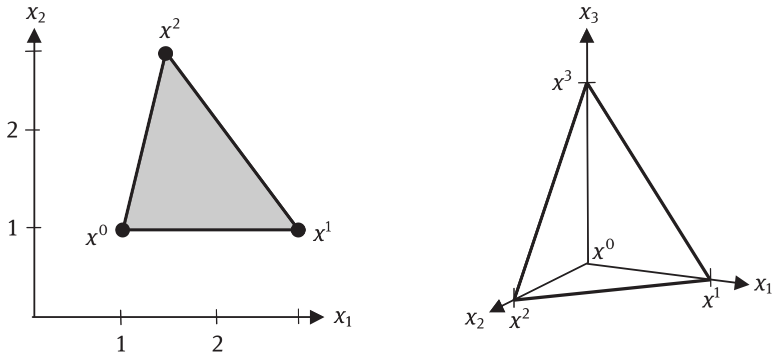


Fig. 3.7. Simplex of dimension 2 and 3.

We will now consider a specific convex polytope:

Definition 3.9. The convex hull of points $x^0, x^1, \dots, x^k \in \mathbb{R}^n$ is called a *simplex of dimension k*, if

$$x^1 - x^0, x^2 - x^0, \dots, x^k - x^0$$

are linearly independent vectors.

Figure 3.7 illustrates the simplex in the $\mathbb{R}^2$ generated by the points $x^0 = (1,1)^T$, $x^1 = (3,1)^T$, $x^2 = (3/2,3)^T$ and the simplex in the $\mathbb{R}^3$ generated by the points $x^0 = (0,0,0)^T$, $x^1 = (1,0,0)^T$, $x^2 = (0,1,0)^T$, $x^3 = (0,0,1)^T$.

We conclude this section with two theorems that reveal fundamental characteristics of the convex hull.

Theorem 3.10 (Carathéodory's Theorem). *Given the set $K \subset \mathbb{R}^n$. Each point $x \in \operatorname{conv}(K)$ can be represented as a convex combination of (at most) $n+1$ points of K, i. e. for each $x \in \operatorname{conv}(K)$ there exist points $x^1, \dots, x^{n+1} \in K$ such that*

$$x = \sum_{i=1}^{n+1} \lambda_i x^i$$

with $\lambda_i \geq 0$ for $i = 1, \dots, n+1$ and $\sum_{i=1}^{n+1} \lambda_i = 1$.

Proof. For $x \in \operatorname{conv}(K)$ there exist k points $x^1, \dots, x^k \in K$ such that

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

with $\lambda_i \geq 0$ for $i = 1, \dots, k$ and $\sum_{i=1}^{k} \lambda_i = 1$. Suppose that $k > n+1$ (otherwise nothing has to be proved). Thus the following system with $n+1$ linear equations in the variables $\lambda_1, \dots, \lambda_k$ has a feasible solution

$$\begin{pmatrix} x^1 & x^2 & \dots & x^k \\ 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{pmatrix} = \begin{pmatrix} x \\ 1 \end{pmatrix} \tag{3.1}$$

$$\lambda_i \geq 0 \quad \text{for } i = 1, \dots, k\,.$$

Suppose that the matrix in (3.1) has maximal rank, i.e. rank $n+1$. From linear programming it is known that (3.1) has a basic feasible solution, because there is a feasible solution.

To simplify the notation suppose that $(\lambda_1, \dots, \lambda_{n+1}, 0, \dots, 0)^T \in \mathbb{R}^k$ is a basic feasible solution, i.e.

$$\lambda_1 \begin{pmatrix} x^1 \\ 1 \end{pmatrix} + \dots + \lambda_{n+1} \begin{pmatrix} x^{n+1} \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ 1 \end{pmatrix}$$

$$\lambda_i \geq 0 \quad \text{for } i = 1, \dots, n+1\,.$$

Thus, x can be represented as

$$x = \sum_{i=1}^{n+1} \lambda_i x^i$$

with $\lambda_i \geq 0$ for $i = 1, \dots, n+1$, $\sum_{i=1}^{n+1} \lambda_i = 1$ and points $x^1, \dots, x^{n+1} \in K$.

When the rank of the matrix in (3.1) is less than $n+1$, we can slightly modify the above idea, determining a basic solution of the reduced system resulting from (3.1), by eliminating equations which are linearly dependent from others. □

Without a proof we present the following statement:

Theorem 3.11. *Let $K \subset \mathbb{R}^n$ be a convex, closed and bounded set and E the set of extreme points of K. It holds*

$$K = \operatorname{conv}(E)\,.$$

This result generalizes the above observation that a convex polytope is the convex hull of its extreme points.

Combining the two last theorems we obtain the following result, important for determining a global minimum of a concave function (see Theorem 3.58).

Theorem 3.12. *Each point x of a convex, closed and bounded set $K \subset \mathbb{R}^n$ can be represented as*

$$x = \sum_{i=1}^{n+1} \lambda_i x^i$$

with $\lambda_i \geq 0$ for $i = 1, \dots, n+1$, $\sum_{i=1}^{n+1} \lambda_i = 1$ and extreme points $x^1, \dots, x^{n+1} \in K$.

Theorem 3.12 is illustrated by the examples of Figure 3.8 in the case $n = 2$.

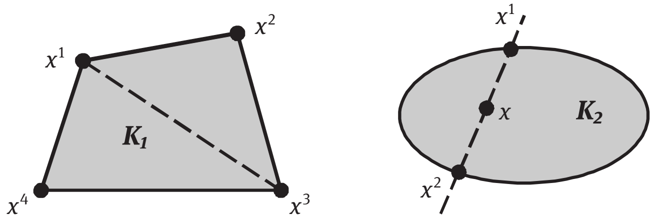


Fig. 3.8. Illustration of Theorem 3.12.

For example, a point $x \in K_1$ is a convex combination of the three extreme points x^1, x^2 and x^3, when x is above or on the line segment $[x^1, x^3]$; otherwise x is a convex combination of the three extreme points x^1, x^3 and x^4. Obviously, each point $x \in K_2$ is a convex combination of two extreme points x^1 and x^2.

Exercise 3.13.

(i) *Show that the algebraic sum of two sets $K_1, K_2 \subset \mathbb{R}^n$, defined by*

$$K_1 + K_2 := \{x + y | x \in K_1, y \in K_2\}\,,$$

is a convex set, when K_1 and K_2 are convex.

(ii) *Represent K_1, K_2 and $K_1 + K_2$ geometrically for*

$$K_1 = \{(x_1, x_2)^T \in \mathbb{R}^2 | x_1^2 + x_2^2 \le 1\}$$

and

$$K_2 = \{(y_1, y_2)^T \in \mathbb{R}^2 \,| 2 \le y_1 \le 3, 0 \le y_2 \le 1\}\,.$$

3.2 Convex and concave functions

This section deals with convex and concave functions. Furthermore, the convex optimization problem is introduced.

Definition 3.14. Let f be a function defined on a convex set $M \subset \mathbb{R}^n$.

(i) f is called *convex* (over M), if

$$f((1-\lambda)x^1 + \lambda x^2) \le (1-\lambda)f(x^1) + \lambda f(x^2) \tag{3.2}$$

is satisfied for any two different points $x^1, x^2 \in M$ and every λ with $0 < \lambda < 1$.

(ii) f is called *strictly convex* (on M), if (3.2) is true as a strict inequality, i.e. valid after replacing the sign "$\le$" by "$<$".

(iii) f is called *(strictly) concave*, if $(-f)$ is (strictly) convex.

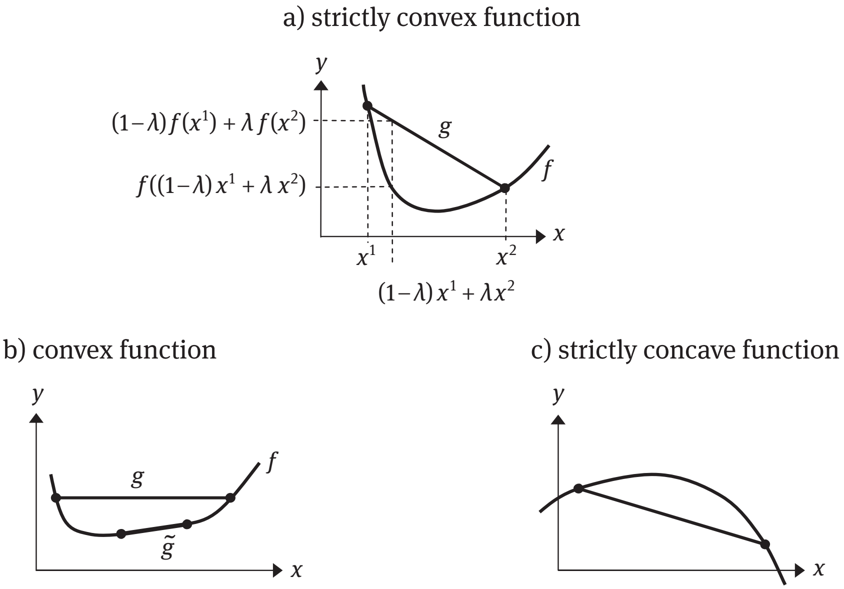


Fig. 3.9. Convex and concave functions.

Figure 3.9 illustrates the above concepts. In geometric terms a function f is convex if any line segment g, joining two points on the graph of f, lies above this graph; f is strictly convex, if this line segment may not coincide with the graph of f. Hence, the functions in Figure 3.9 (a), (b) are both convex, but only the former is strictly convex.

We study some numerical examples:

(a) The functions $f(x) = e^x$ and $f(x) = x^2$ are strictly convex over $\mathbb{R}$, $f(x_1, x_2) = x_1^2 + x_2^2$ is strictly convex over $\mathbb{R}^2$.
(b) $f(x) = \ln x$ is strictly concave over $(0, \infty)$.
(c) $f(x) = a^T x + b$ with $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ is convex and concave over $\mathbb{R}^n$.
(d) $f(x) = x^3$ is neither convex nor concave over $\mathbb{R}$, but strictly convex over $[0, \infty)$ and strictly concave over $(-\infty, 0]$.
(e) $f(x) = 1/x$ is strictly convex over $(0, \infty)$ and strictly concave over $(-\infty, 0)$.
(f) For $a, b \in \mathbb{R}$ the function $f(x_1, x_2) = \sqrt{(x_1 - a)^2 + (x_2 - b)^2}$ is convex but not strictly convex over $\mathbb{R}^2$.

Exercise 3.15.
(i) *Prove the strict convexity of the function* $f(x) = x^2$.
(ii) *Let* $f : M \to \mathbb{R}$ *be a nonnegative convex function* ($M \subset \mathbb{R}^n$ *convex*). *Show that the function* $g(x) := (f(x))^2$ *is convex.*
(iii) *Show that the function in the above example* (f) *is convex but not strictly convex.*

The next two results are used to decide whether a composite function is convex.

Theorem 3.16. *Let $f_1, \dots, f_m$ be convex functions over the convex set $M \subset \mathbb{R}^n$. Then the sum*

$$f(x) = \sum_{i=1}^{m} f_i(x)$$

is convex over M. Furthermore, f is strictly convex, if at least one of the functions f_i is strictly convex.

Theorem 3.17. *Let f be a (strictly) convex function over the convex set $M \subset \mathbb{R}^n$. For $\alpha > 0$ the function αf is (strictly) convex.*

Exercise 3.18. *Prove the two preceding theorems and formulate analogous statements for concave functions.*

Theorems 3.16 and 3.17 imply in particular that the function $f(x_1, x_2)$ in Example 1.18, representing the total cost of transportations, is convex (see Exercise 3.15 (iii)).

Theorem 3.19. *Given a concave function $f : M \to \mathbb{R}$ and a decreasing convex function $g : \overline{M} \to \mathbb{R}$ such that $f(M) \subset \overline{M}$ ($M, \overline{M} \subset \mathbb{R}^n$ convex). Then the composite function $g \circ f$ is convex over M.*

Proof. Since f is concave, we have

$$f((1-\lambda)x^1 + \lambda x^2) \geq (1-\lambda)f(x^1) + \lambda f(x^2)$$

for $x^1, x^2 \in M$, which implies

$$g(f((1-\lambda)x^1 + \lambda x^2)) \leq g((1-\lambda)f(x^1) + \lambda f(x^2)), \tag{3.3}$$

because g is decreasing. Since g is convex, we have

$$g((1-\lambda)f(x^1) + \lambda f(x^2)) \leq (1-\lambda)g(f(x^1)) + \lambda g(f(x^2)). \tag{3.4}$$

Combining (3.3) and (3.4), we obtain

$$g \circ f((1-\lambda)x^1 + \lambda x^2) \leq (1-\lambda)g \circ f(x^1) + \lambda g \circ f(x^2),$$

i.e. $g \circ f$ is convex. □

If the function g in Theorem 3.19 is of the form $g(x) = 1/x, x > 0$ (see the above example (e)), we obtain the following corollary.

Corollary 3.20. *Let $f : M \to \mathbb{R}$ be a concave function over the convex set $M \subset \mathbb{R}^n$ with $f(x) > 0$ for all $x \in M$. Then the function $g(x) = \frac{1}{f(x)}$ is convex over M.*

For example, as $f(x) = 5 - x^2$ is concave over $M = [-2, 2]$ with positive values, the function

$$g(x) = \frac{1}{f(x)} = \frac{1}{5 - x^2}$$

is convex over M.

Definition 3.21. The problem

$$\min_{x \in M} f(x)$$

with $M := \{x \in \mathbb{R}^n | g_i(x) \leq 0, i = 1, \ldots, m\}$ is called a *convex optimization problem*, if all functions $f, g_1, \ldots, g_m$ are convex.

Comment 3.22. The feasible region of a convex optimization problem is convex.

Proof. We consider the sets

$$M_i := \{x \in \mathbb{R}^n | g_i(x) \leq 0\}$$

with a convex function g_i. For two points $x^1, x^2 \in M_i$ and λ with $0 < \lambda < 1$ we have

$$g_i((1-\lambda)x^1 + \lambda x^2) \leq (1-\lambda)g_i(x^1) + \lambda g_i(x^2) \leq 0\,.$$

Therefore, the point $(1-\lambda)x^1 + \lambda x^2$ is also a point of M_i, thus M_i is convex for all $i = 1, \ldots, m$. Hence, $M = M_1 \cap \cdots \cap M_m$ is convex. □

Exercise 3.23. *Let $M \subset \mathbb{R}^n$ be a convex set and $x^* \in M$. Show that $d = x - x^*$ is a feasible direction at x^* for all $x \in M$.*

Exercise 3.24. *Investigate the following functions with respect to convexity and concavity:*

(i) $f(x) = \dfrac{1}{\ln x} + 3e^x + \dfrac{1}{x} \quad M = \{x \in \mathbb{R} | x > 1\}$,

(ii) $f(x) = \begin{cases} 0 & \textit{for } x = 0 \\ 1 + 2\sqrt{x} & \textit{for } x > 0 \end{cases} \quad M = \mathbb{R}_+$,

(iii) $f(x) = \begin{cases} x^2 & \textit{for } x \leq 0 \\ 0 & \textit{for } 0 < x < 1 \\ x - 1 & \textit{for } x \geq 1 \end{cases} \quad M = \mathbb{R}$.

3.3 Differentiable convex functions

It can be proved that a convex function over a convex set $M \subset \mathbb{R}^n$ is continuous on the interior of M. Discontinuities can only occur at boundary points (see Exercise 3.24 (ii)). However, a convex function need not be differentiable at inner points. For example, the function of Exercise 3.24 (iii) is not differentiable at point 1, and the convex function $f(x_1, x_2)$ of Example 1.18 is not differentiable at the points $(a_i, b_i)^T$, $i = 1, \ldots n$.

In this section we study some characteristics of differentiable convex functions. We present the following two results without proof:

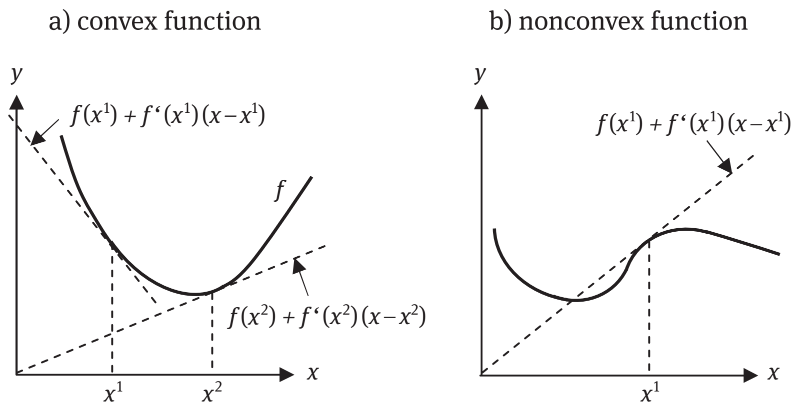


Fig. 3.10. Illustration of Theorem 3.25.

Theorem 3.25. *A continuously differentiable function f over the convex set $M \subset \mathbb{R}^n$ is convex if and only if*

$$f(x) \geq f(x^*) + (x - x^*)^T \operatorname{grad} f(x^*) \tag{3.5}$$

for all $x, x^* \in M$.

The graph of the right-hand side of (3.5) represents the tangent hyperplane of the graph of f at the point corresponding to x^*. It is intuitive that f is convex if and only if all these hyperplanes lie below the graph of f.

The fact is illustrated in Figure 3.10 for the case $n = 1$, in which the tangent hyperplanes are the straight lines $f(x^i) + f'(x^i)(x - x^i)$.

Theorem 3.26.
(i) *A twice continuously differentiable function f over the convex set $M \subset \mathbb{R}^n$ is convex if and only if, the Hessian matrix $Hf(x)$ is positive semidefinite for all $x \in M$.*
(ii) *If $Hf(x)$ is positive definite for all $x \in M$, then f is strictly convex over M.*

We illustrate the main idea of the proof of part (i):

Proof. Replacing $f(x)$ in condition (3.5) by the second-order Taylor approximation we get the following equivalent condition for convexity:

$$\begin{aligned} f(x^*) + (x - x^*)^T \operatorname{grad} f(x^*) + \frac{1}{2}(x - x^*)^T Hf(x^*)(x - x^*) \\ \geq f(x^*) + (x - x^*)^T \operatorname{grad} f(x^*) \quad \text{for all } x, x^* \in M\,. \end{aligned}$$

Assuming that this approximation is exact, we obtain the condition

$$(x - x^*)^T Hf(x^*)(x - x^*) \geq 0 \quad \text{for all } x, x^* \in M\,,$$

i.e. $Hf(x^*)$ is positive semidefinite for all $x^* \in M$.

For $x \to x^*$ the error caused by using the quadratic approximation converges rapidly to zero, thus the last condition is in fact equivalent to convexity, if f is twice continuously differentiable. $\square$

The converse of part (ii) in the above theorem is not true. For example, the function $f(x) = x^4$ over $\mathbb{R}$ is strictly convex, but $Hf(x) = f''(x) = 12x^2$ is not positive definite for $x = 0$.

Example 3.27. We consider the following functions over $M = \mathbb{R}^2$:
(a) $f(x_1, x_2) = x_1^2 + x_1x_2 + x_2^2$
(b) $f(x_1, x_2) = x_1^2 + x_1x_2^3$.

In case (a) the Hessian matrix is

$$Hf(x_1, x_2) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} .$$

Since $2 > 0$ and $\left| \begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix} \right| = 3 > 0$, $Hf(x_1, x_2)$ is positive definite at all points (see Theorem 2.29 (ii)). Thus, f is strictly convex according to Theorem 3.26 (ii).

In case (b) we obtain

$$Hf(x_1, x_2) = \begin{pmatrix} 2 & 3x_2^2 \\ 3x_2^2 & 6x_1x_2 \end{pmatrix} .$$

Theorem 2.29 (i) implies that $Hf(1, -1)$ is not positive semidefinite. Therefore, f is not convex (Theorem 3.26 (i)).

Exercise 3.28. *Which of the following functions is convex over $\mathbb{R}^2$?*
(a) $f(x_1, x_2) = 2x_1^2 - x_1x_2 + 5x_2^2$
(b) $f(x_1, x_2) = x_1^2x_2^2$.

Exercise 3.29. *Formulate analogous conditions to those of Theorems 3.25 and 3.26 that characterize the concavity of a differentiable function.*

3.4 Subgradient and directional derivative

Firstly we introduce the concept of the subgradient, which generalizes the gradient of a convex function f to points at which f is not differentiable. For this purpose we define the epigraph and the supporting hyperplane of a convex set:

Definition 3.30. Let $f : M \to \mathbb{R}$ be a function over $M \subset \mathbb{R}^n$. The *epigraf* of f is defined by

$$\text{epi}\,(f) = \{(x, r) \in \mathbb{R}^{n+1} | x \in M, r \geq f(x)\} .$$

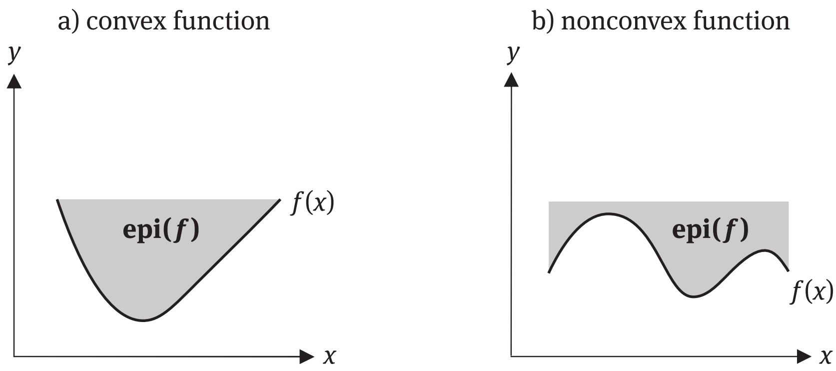


Fig. 3.11. The epigraph of a function.

For $n = 1$ and $n = 2$ the epigraph is the set of all points above the graph of f. Obviously a function is convex if and only if its epigraph is convex (Figure 3.11).

Definition 3.31. Given a convex set $M \subset \mathbb{R}^n$ and a hyperplane

$$H := \{x \in \mathbb{R}^n | a^T x = b\}$$

with $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. The half-spaces generated by H are

$$S^+ = \{x \in \mathbb{R}^n | a^T x \geq b\} \quad \text{and} \quad S^- = \{x \in \mathbb{R}^n | a^T x \leq b\}\,.$$

The hyperplane H is called a *supporting hyperplane* of M at the point x^0if $x^0 \in H \cap M$ and M is entirely contained in one of the half-spaces S^+ or S^-.

Example 3.32. Figure 3.12 illustrates supporting hyperplanes for three convex sets in the plane.

(i) $H = \{(x_1, x_2)^T \in \mathbb{R}^2 | (1,1)\binom{x_1}{x_2} = \sqrt{2}\}$ is a supporting hyperplane of the convex set $M = \{(x_1, x_2)^T \in \mathbb{R}^2 | x_1^2 + x_2^2 \leq 1\}$ at point $x^0 = (1/\sqrt{2}, 1/\sqrt{2})^T$ (Figure 3.12 (a)).

(ii) $H = \{(x_1, x_2)^T \in \mathbb{R}^2 | x_2 = 0\}$ is a supporting hyperplane of

$$M = \left\{(x_1, x_2)^T \in \mathbb{R}^2 | x_1 + x_2 \leq 1, x_1 \geq 0, x_2 \geq 0\right\}$$

at point $(\alpha, 0)^T$ for all α with $0 \leq \alpha \leq 1$ (Figure 3.12 (b)).

(iii) For each $\alpha \geq 0$, the hyperplane $H(\alpha) = \{(x_1, x_2)^T \in \mathbb{R}^2 | x_2 = 1 + \alpha - \alpha x_1\}$ is a supporting hyperplane of

$$M = \left\{(x_1, x_2)^T \in \mathbb{R}^2 | 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1\right\}$$

at point $x^0 = (1,1)^T$ (Figure 3.12 (c)). In addition there is the hyperplane $\{(x_1, x_2) \in \mathbb{R}^2 | x_1 = 1\}$.

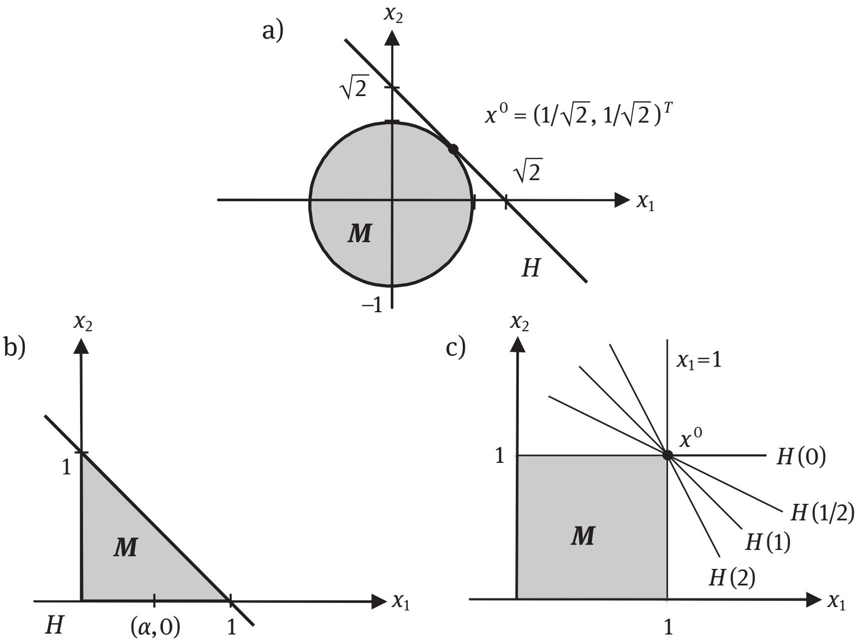


Fig. 3.12. Supporting hyperplanes of convex sets.

We now study supporting hyperplanes of the epigraph of a convex function. Obviously the graph of an affine-linear function $g(x) = a^T x + b$ is a supporting hyperplane of a set epi(f) for appropriately chosen $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. For ease of language we introduce the following concept:

Definition 3.33. Let $f : M \to \mathbb{R}$ be a convex function ($M \subset \mathbb{R}^n$ convex). The affine-linear function $g(x) = a^T x + b$ is called a *support function* of f at point $x^* \in M$, if the graph of $g(x)$ is a supporting hyperplane of epi(f) at point $(x^*, f(x^*))^T$, i.e. if

$$f(x^*) = g(x^*) \tag{3.6}$$

and

$$f(x) \geq g(x) \quad \text{for all } x \in M \tag{3.7}$$

(see Figure 3.13).

Because of (3.6), the parameter b of a support function satisfies $b = f(x^*) - a^T x^*$. Thus, the support functions of a convex function f at point x^* are the functions

$$g(x) = a^T x + f(x^*) - a^T x^* = f(x^*) + a^T (x - x^*)\,, \tag{3.8}$$

where the vector a satisfies (see (3.7)):

$$f(x) \geq f(x^*) + a^T (x - x^*) \quad \text{for all } x \in M\,. \tag{3.9}$$

Note that the vector a is the gradient of the function $g(x)$.

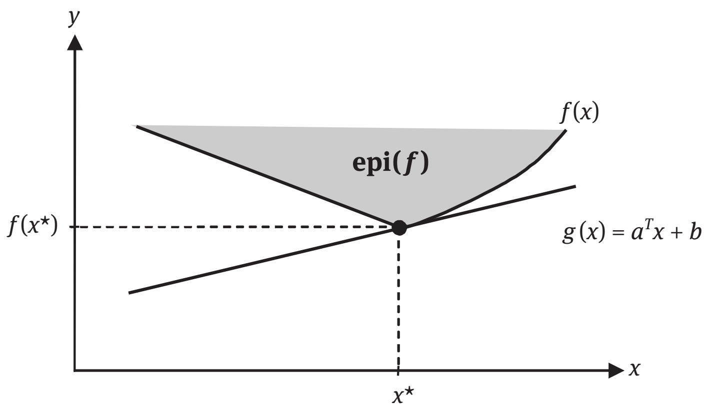


Fig. 3.13. Support function of f at x^*.

The above considerations suggest the following concept:

Definition 3.34. Let $f : M \to \mathbb{R}$ be a convex function ($M \subset \mathbb{R}^n$ convex). The vector $a \in \mathbb{R}^n$ is called a *subgradient* of f at the point x^*, if it is the gradient of a support function of f at x^*, i.e. if a satisfies (3.9). The set of subgradients of f at x^*, denoted by $\partial f(x^*)$, is called the *subdifferential* of f at x^*.

Example 3.35. We determine the subdifferential of the convex function

$$f(x) = \begin{cases} -x & \text{for } x < 0 \\ x^2 & \text{for } x \geq 0 \end{cases}$$

at the three points -1, 0 and 1 (see Figure 3.14).

Clearly, there is a unique support function at a differentiable point x^*. For example, for $x^* = -1$ and $x^* = 1$ we obtain the support functions $r = -x$ and $r = 2x - 1$ with the gradients -1 and 2, respectively (Figure 3.14 (a), (b)).

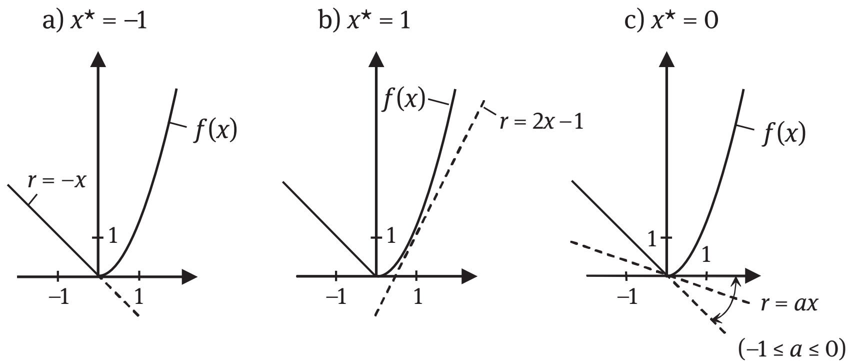


Fig. 3.14. Support functions of the function f of Example 3.35.

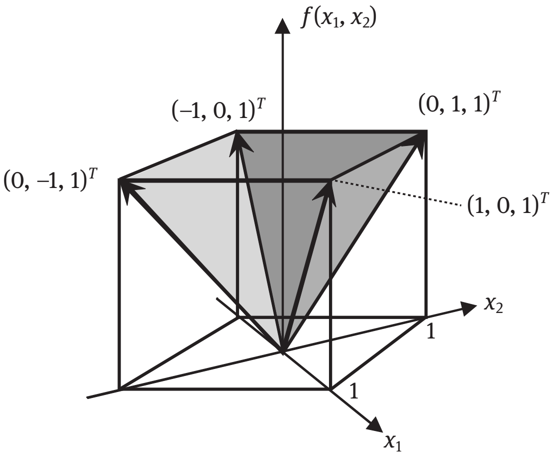


Fig. 3.15. The function $f(x_1, x_2) = |x_1| + |x_2|$.

Thus the function f has the unique subgradient -1 at point -1 and the unique subgradient 2 at point 1.

At the point $x^* = 0$ each function $r = ax$ with $-1 \le a \le 0$ is a support function (Figure 3.14 (c)), i.e. $[-1, 0]$ is the subdifferential of f at point 0.

Generalizing these observations we can state that the $\partial f(x^*) = \{gradf(x^*)\}$ when f is differentiable at x^*, otherwise $\partial f(x^*)$ is an infinite set.

Exercise 3.36. *Determine the subgradients in Example 3.35 formally, i.e. determine all vectors a that satisfy (3.9) for the cases* $x^* = -1$, $x^* = 0$ *and* $x^* = 1$.

Example 3.37. We determine the subdifferential of the convex function

$$f(x_1, x_2) = |x_1| + |x_2|$$

at point $x^* = 0$ (Figure 3.15). The graph of f is the surface of the cone generated by the vectors $(1, 0, 1)^T$, $(0, 1, 1)^T$, $(-1, 0, 1)^T$ and $(0, -1, 1)^T$.

For the above function f and $x^* = 0$, relation (3.9) becomes

$$|x_1| + |x_2| \ge a_1x_1 + a_2x_2 \quad \text{for all } x_1, x_2 \in \mathbb{R}\,, \tag{3.10}$$

which is equivalent to

$$|x_1| \ge a_1x_1 \text{and } |x_2| \ge a_2x_2 \quad \text{for all } x_1, x_2 \in \mathbb{R}\,,$$

in turn equivalent to

$$|x_1| \ge |a_1x_1| \text{and } |x_2| \ge |a_2x_2| \quad \text{for all } x_1, x_2 \in \mathbb{R}\,,$$

i.e.

$$|a_1| \le 1 \quad \text{and} \quad |a_2| \le 1\,.$$

Hence a vector $a = (a_1, a_2)^T$ is a subgradient of f at point 0 if and only if

$$|a_1| \leq 1 \quad \text{and} \quad |a_2| \leq 1 .$$

Exercise 3.38. *Determine geometrically the subdifferential of the convex function*

$$f(x) = \begin{cases} 1 - 2x & \text{for } x < 1 , \\ -3/2 + x/2 & \text{for } x \geq 1 \end{cases}$$

at point $x^* = 1$.

Exercise 3.39. *Determine the subdifferential of the function*

$$f(x_1, x_2) = 2|x_1| + 3|x_2|$$

at point $x^* = 0$.

Subgradients play an important role in the theory of convex optimization and development of algorithms. The so-called subgradient method for minimizing nondifferentiable convex functions, developed in the 1970s is an iterative procedure searching along subgradient directions (to a certain resemblance to the gradient method, see Section 7.2). Subgradients are also used in many practical applications, see the literature cited in Section 4.5.

In the remainder of this section we study the concept of a directional derivative that allows a characterization of the subdifferential, among others (see Theorem 3.43 (iii)).

Given a function $f : M \to \mathbb{R}, M \subset \mathbb{R}^n$ and an inner point $x^* \in M$. For any vector $y \in \mathbb{R} \backslash \{0\}$ we consider the one-dimensional function

$$f_{(y)}(t) := f(x^* + ty) .$$

The directional derivative of f can be interpreted as the derivative of this one-dimensional function at $t = 0$:

Definition 3.40.

(i) The *directional derivative* $Df(x^*, y)$ of the function f at point x^* in the direction y is defined by

$$Df(x^*, y) = f'_{(y)}(0) = \lim_{t \to 0} \frac{f_{(y)}(t) - f_{(y)}(0)}{t} = \lim_{t \to 0} \frac{f(x^* + ty) - f(x^*)}{t} .$$

(ii) The *unilateral derivatives* $D^+f(x^*, y)$ and $D^-f(x^*, y)$ of f at x^* in the direction y are similarly defined by

$$D^+f(x^*, y) = \lim_{t \to 0+} \frac{f_{(y)}(t) - f_{(y)}(0)}{t}$$

and

$$D^-f(x^*, y) = \lim_{t \to 0-} \frac{f_{(y)}(t) - f_{(y)}(0)}{t} .$$

For a convex function $f : M \to \mathbb{R}(M \subset \mathbb{R}^n$ convex) and an interior point $x^* \in M$ it can be proved that the unilateral derivatives exist and that

$$D^-f(x^*, y) \le D^+f(x^*, y)$$

holds for all $y \in \mathbb{R}\backslash\{0\}$. The directional derivative $Df(x^*, y)$ exists if and only if $D^-f(x^*, y) = D^+f(x^*, y)$. The term $\frac{1}{|y|}Df(x^*, y)$ can be geometrically interpreted as a measure for the rise/decline of the function f at point x^* in the direction y. Note that

$$\frac{1}{|y|}Df(x^*, y) = \lim_{t\to 0} \frac{f(x^* + ty) - f(x^*)}{|ty|},$$

where the last quotient represents the variation of f between x^* and $x^* + ty$ relative to the distance $|ty|$ between these points.

Example 3.41. Given the function

$$f(x_1, x_2) = x_1^2 + 2x_1x_2, \quad M = \mathbb{R}^2 .$$

We determine $Df(x^*, y)$ for $x^* = (1, 1)^T$ and $y = (2, 3)^T$:

$$\begin{aligned} f_{(y)}(t) &= f(x^* + ty) = f(1 + 2t, 1 + 3t) = (1 + 2t)^2 + 2(1 + 2t)(1 + 3t) \Rightarrow \\ f'_{(y)}(t) &= 2(1 + 2t)2 + 2[2(1 + 3t) + (1 + 2t)3] \Rightarrow \\ Df(x^*, y) &= f'_{(y)}(0) = 14 . \end{aligned}$$

Exercise 3.42.

(i) *Determine $Df(x^*, y)$ for*

$$f(x_1, x_2) = 2x_1 - 3x_1x_2 \quad (M = \mathbb{R}^2), \quad x^* = (2, 3)^T \text{ and } y = (1, 2)^T .$$

(ii) *Determine $D^+f(x^*, y)$ and $D^-f(x^*, y)$ for*

$$f(x_1, x_2) = \sqrt{x_1^2 + x_2^2} \quad (M = \mathbb{R}^2), \quad x^* = (0, 0)^T \text{ and } y = (1, 1)^T .$$

Does $Df(x^*, y)$ exist?

(iii) *Determine $D^+f(x^*, y)$ for*

$$f(x_1, x_2) = |x_1| + |x_2| \quad (M = \mathbb{R}^2), x^* = (0, 0)^T \text{ and } y = (y_1, y_2)^T \in \mathbb{R}^2 .$$

The following theorem shows some properties of unilateral derivatives:

Theorem 3.43. *Given a convex function $f : M \to \mathbb{R}(M \subset \mathbb{R}$ convex) and an interior point $x^* \in M$. It holds*

(i) $-D^+f(x^*, -y) = D^-f(x^*, y)$ *for all* $y \in \mathbb{R}^n$.

(ii) $D^+f(x^*, \lambda y) = \lambda D^+f(x^*, y)$ *and* $D^-f(x^*, \lambda y) = \lambda D^-f(x^*, y)$ *for* $\lambda \ge 0$ *and all* $y \in \mathbb{R}^n$.

(iii) *The vector $a \in \mathbb{R}^n$ is a subgradient of f at point x^* if and only if*

$$D^+f(x^*, y) \geq a^T y \quad \textit{for all} y \in \mathbb{R}^n\,.$$

Proof. (i) It holds

$$\begin{aligned}-D^+f(x^*, -y) &= -\lim_{t\to 0+} \frac{f(x^* - ty) - f(x^*)}{t} = \lim_{t\to 0+} \frac{f(x^* - ty) - f(x^*)}{-t} \\ &= \lim_{t\to 0-} \frac{f(x^* + ty) - f(x^*)}{t} = D^-f(x^*, y)\,.\end{aligned}$$

(ii) We get

$$\begin{aligned}D^+f(x^*, \lambda y) &= \lim_{t\to 0+} \frac{f(x^* + t\lambda y) - f(x^*)}{t} = \lambda \lim_{t\to 0+} \frac{f(x^* + t\lambda y) - f(x^*)}{\lambda t} \\ &= \lambda \lim_{t\to 0+} \frac{f(x^* + ty) - f(x^*)}{t} = \lambda D^+f(x^*, y)\,.\end{aligned}$$

Similarly one can prove the second assertion in (ii).

(iii) Let $a \in \mathbb{R}^n$ be a subgradient of f at x^*. It holds

$$f(x) \geq f(x^*) + a^T(x - x^*) \quad \text{for all } x \in \mathbb{R}\,.$$

Replacing x by $x^* + ty$, we obtain

$$f(x^* + ty) - f(x^*) \geq a^T ty$$

for a sufficiently small $t > 0$. The last relation implies

$$D^+f(x^*, y) = \lim_{t\to 0+} \frac{f(x^* + ty) - f(x^*)}{t} \geq a^T y\,.$$

Similarly one can prove that $D^+f(x^*, y) < a^T y$ for some $y \in \mathbb{R}^n$, when a is not a subgradient of f at x^*. □

Example 3.44. Given the convex function

$$f(x_1, x_2) = \begin{cases} -x_1 + x_2 & \text{for } x_1 \leq 0, \\ x_1^2 + x_2 & \text{for } x_1 > 0 \end{cases}$$

in $\mathbb{R}^2$. We determine the subdifferential of f at $x^* = 0$ by means of Theorem 3.43 (iii). It holds

$$D^+f(x^*, y) = \lim_{t\to 0+} \frac{f(x^* + ty) - f(x^*)}{t} = \lim_{t\to 0+} \frac{f(ty)}{t}.$$

Thus

$$D^+f(x^*, y) = \lim_{t\to 0+} \frac{-ty_1 + ty_2}{t} = -y_1 + y_2 \quad \text{for } y_1 \leq 0$$

and

$$D^+f(x^*, y) = \lim_{t \to 0+} \frac{t^2 y_1^2 + t y_2}{t} = y_2 \quad \text{for } y_1 > 0\,.$$

Hence, a vector $a = (a_1, a_2)^T \in \mathbb{R}^n$ is a subgradient of f at $x^* = 0$ if and only if

$$\begin{aligned} -y_1 + y_2 &\ge a_1 y_1 + a_2 y_2 \quad \text{for all } y_1 \le 0 \text{ and all } y_2 \in \mathbb{R} \\ y_2 &\ge a_1 y_1 + a_2 y_2 \quad \text{for all } y_1 > 0 \text{ and all } y_2 \in \mathbb{R}\,. \end{aligned}$$

The system is equivalent to

$$0 \ge (a_1 + 1) y_1 + (a_2 - 1) y_2 \quad \text{for all } y_1 \le 0 \text{ and all } y_2 \in \mathbb{R} \tag{3.11}$$

$$0 \ge a_1 y_1 + (a_2 - 1) y_2 \quad \text{for all } y_1 > 0 \text{ and all } y_2 \in \mathbb{R}\,. \tag{3.12}$$

As (3.11) must be satisfied for $y_1 = 0$ and all $y_2 \in \mathbb{R}$, we obtain $a_2 = 1$, and (3.11), (3.12) reduce to

$$\begin{aligned} 0 &\ge (a_1 + 1) y_1 \quad \text{for all } y_1 \le 0 \\ 0 &\ge a_1 y_1 \quad \text{for all } y_1 > 0, \end{aligned}$$

which is equivalent to $a_1 + 1 \ge 0$ and $a_1 \le 0$. Hence, the vector $a = (a_1, a_2)^T$ is a subgradient of f at $x^* = 0$ if and only if $a_2 = 1$ and $-1 \le a_1 \le 0$.

Exercise 3.45. *Determine the subdifferential of the convex function*

$$f(x_1, x_2) = |x_1| + x_2^2$$

at $x^* = 0$, using Theorem 3.43 (iii).

Corollary 3.46. *Given a convex function $f : M \to \mathbb{R}$ ($M \subset \mathbb{R}$ convex) and an interior point $x^* \in M$. The real number $a \in \mathbb{R}$ is a subgradient of f at x^* if and only if*

$$D^-f(x^*, 1) \le a \le D^+f(x^*, 1)\,,$$

i.e.

$$\partial f(x^*) = \left[D^-f(x^*, 1), D^+f(x^*, 1)\right]\,.$$

Note that the limits of the interval $\partial f(x^*)$ are the common one-sided derivatives of the one-dimensional function f at the point x^*. The corollary generalizes the observations made in Example 3.35 and Exercise 3.38

Proof. Part (ii) of the above theorem implies for $y > 0$, $\lambda := 1/y$:

$$\begin{aligned} &D^+f(x^*, 1) = \frac{1}{y} D^+f(x^*, y) \Rightarrow \\ &D^+f(x^*, y) = y D^+f(x^*, 1)\,. \end{aligned}$$

Similarly we obtain for $y < 0$, $\lambda := -1/y$:

$$D^+f(x^*, y) = -yD^+f(x^*, -1) = yD^-f(x^*, 1) \quad \text{(Theorem 3.43 (i))}.$$

Thus, the number a is a subgradient, if and only if (Theorem 3.43 (iii))

$$\begin{aligned} yD^+f(x^*, 1) &\geq ay \quad \text{for all } y > 0 \\ yD^-f(x^*, 1) &\geq ay \quad \text{for all } y < 0 . \end{aligned}$$

Dividing the inequalities by y results in the assertion of the theorem. □

Exercise 3.47. *Determine the subdifferential of the convex function*

$$f(x) = \begin{cases} -x^3 & \text{for } x < 0 , \\ 2x & \text{for } x \geq 0 . \end{cases}$$

at the point $x^* = 0$.

Finally, we note that there exist concepts for concave functions which are analogous to epigraph and subgradient for convex functions.

3.5 Minima of convex and concave functions

When a function is convex, there is no distinction between local and global minima:

Theorem 3.48. *Given a convex function $f : M \to \mathbb{R}(M \subset \mathbb{R}^n$ convex). Any local minimum of f is a global minimum.*

Proof. Suppose that $x^1 \in M$ is a local but not global minimum point of a function f. Then there is a point $x^2 \in M$ with $f(x^2) < f(x^1)$ (see Figure 3.16) and for a sufficiently small $\lambda > 0$ we have

$$f((1-\lambda)x^1 + \lambda x^2) > f(x^1) = (1-\lambda)f(x^1) + \lambda f(x^1) > (1-\lambda)f(x^1) + \lambda f(x^2) ,$$

i.e. f is not convex (see Definition 3.14).

□

We now observe that for a differentiable convex function the first-order necessary optimality condition (Theorem 2.10) is also sufficient for a global minimum.

Theorem 3.49. *Let $f : M \to \mathbb{R}$ be a differentiable convex function ($M \subset \mathbb{R}^n$ convex). The point $x^* \in M$ is a global minimum point if and only if*

$$d^T \operatorname{grad} f(x^*) \geq 0 \tag{3.13}$$

is satisfied for all feasible directions d at x^*.

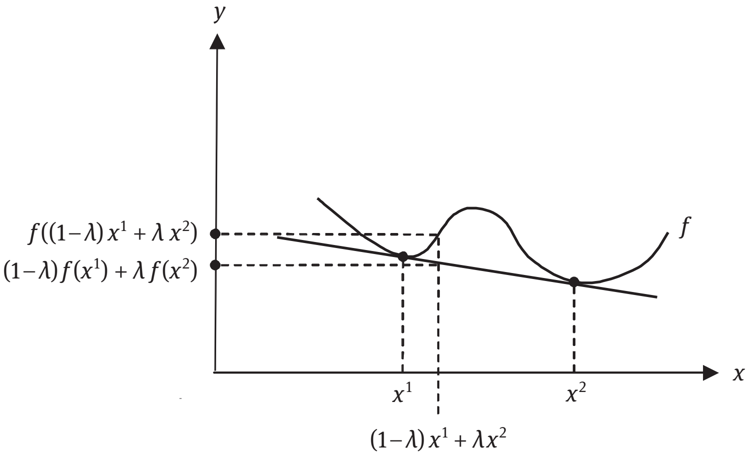


Fig. 3.16. Illustration of Theorem 3.48.

Proof. If x^* is a global minimum point, (3.13) must be satisfied (see Theorem 2.10). Let (3.13) be valid for all feasible directions. Since $(x - x^*)$ is a feasible direction for all $x \in M$ (see Exercise 3.23) we obtain

$$(x - x^*)^T \operatorname{grad} f(x^*) \geq 0 \quad \text{for all } x \in M .$$

Now Theorem 3.25 implies

$$f(x) \geq f(x^*) + (x - x^*)^T \operatorname{grad} f(x^*) \geq f(x^*) \quad \text{for all } x \in M ,$$

i.e. x^* is a global minimum point. □

Corollary 3.50. *Let f be a function as in Theorem 3.49. The interior point $x^* \in M$ is a global minimum point if and only if*

$$\operatorname{grad} f(x^*) = 0 .$$

Proof. The assertion follows from Theorem 3.49, since any direction is feasible at an interior point. □

Example 3.51. We determine a global minimum point x^* of the function

$$f(x_1, x_2, x_3) = 3x_1^2 + 2x_1x_2 + x_2^2 + x_2x_3 + 2x_3^2 - 8x_1 - 6x_2 - x_3 + 12$$

on $\mathbb{R}^3$. We obtain

$$\operatorname{grad} f(x) = \begin{pmatrix} 6x_1 + 2x_2 - 8 \\ 2x_1 + 2x_2 + x_3 - 6 \\ x_2 + 4x_3 - 1 \end{pmatrix} \quad \text{and} \quad Hf(x) = \begin{pmatrix} 6 & 2 & 0 \\ 2 & 2 & 1 \\ 0 & 1 & 4 \end{pmatrix} .$$

Theorem 2.29 (ii) implies that $Hf(x)$ is positive definite (the determinant of $Hf(x)$ is 26), thus f is strictly convex. Setting $\operatorname{grad} f(x)$ equal to zero, we obtain the unique global minimum point $x^* = (5/13, 37/13, -6/13)^T$.

Example 3.52. We consider the problem of parameter estimation

$$\min \sum_{i=1}^{m} y_i^2$$
$$y_i = |\beta_1 x_i + \beta_2 - z_i|$$

(see Example 1.14) which can be written as an unconstrained problem

$$\min f(\beta_1, \beta_2) := \sum_{i=1}^{m} (\beta_1 x_i + \beta_2 - z_i)^2 . \tag{3.14}$$

Defining the functions $f_i(\beta_1, \beta_2) := (\beta_1 x_i + \beta_2 - z_i)^2$ for $i = 1, \dots, m$, we obtain

$$\operatorname{grad} f_i(\beta_1, \beta_2) = \begin{pmatrix} 2x_i(\beta_1 x_i + \beta_2 - z_i) \\ 2(\beta_1 x_i + \beta_2 - z_i) \end{pmatrix},$$
$$Hf_i(\beta_1, \beta_2) = \begin{pmatrix} 2x_i^2 & 2x_i \\ 2x_i & 2 \end{pmatrix}.$$

Theorem 2.29 (ii) implies that the Hessian matrix $Hf_i(\beta_1, \beta_2)$ is positive semidefinite for all i. Therefore the functions f_i and hence the sum $f = \sum_{i=1}^m f_i$ is convex.

Applying the Corollary 3.50, we obtain that a minimum point of f satisfies

$$\operatorname{grad} f(\beta_1, \beta_2) = \begin{pmatrix} 2\sum_{i=1}^m x_i(\beta_1 x_i + \beta_2 - z_i) \\ 2\sum_{i=1}^m (\beta_1 x_i + \beta_2 - z_i) \end{pmatrix} = 0,$$

i.e. β_1, β_2 satisfy the system of linear equations

$$\beta_1 \sum_{i=1}^m x_i^2 + \beta_2 \sum_{i=1}^m x_i = \sum_{i=1}^m x_i z_i , \tag{3.15}$$
$$\beta_1 \sum_{i=1}^m x_i + \beta_2 m = \sum_{i=1}^m z_i . \tag{3.16}$$

With $\overline{x} = \frac{1}{m}\sum_{i=1}^m x_i$ and $\overline{z} = \frac{1}{m}\sum_{i=1}^m z_i$ Equation (3.16) implies

$$\beta_2 = \overline{z} - \beta_1 \overline{x} . \tag{3.17}$$

Replacing β_2 in (3.15) for the right-hand side of (3.17) we obtain

$$\beta_1 = \frac{\sum_{i=1}^m x_i z_i - m\overline{x}\,\overline{z}}{\sum_{i=1}^m x_i^2 - m\overline{x}^2} . \tag{3.18}$$

Thus the unique solution to the problem (3.14) is given by (3.17) and (3.18).

Exercise 3.53. (a) *Show that the objective function in (3.14) is strictly convex (excluding the case in which all values x_i are equal).*

(b) *Solve the problem (3.14) for the values*

$$\begin{array}{llll} x_1 = 1 & x_2 = 2.5 & x_3 = 3.5 & x_4 = 5 \\ z_1 = 2.1 & z_2 = 5.2 & z_3 = 6.9 & z_4 = 11 \quad (m = 4)\,. \end{array}$$

Illustrate the problem geometrically.

Example 3.54. We consider the following variant of the location problem of Example 1.18: the cost of transportation $f_i(x_1, x_2)$ between the oil refinery D and the gas station C_i is proportional to the square of the Euclidean distance between D and C_i, i.e.

$$f_i(x_1, x_2) = c[(x_1 - a_i)^2 + (x_2 - b_i)^2]\,.$$

The problem of minimizing the total transportation cost becomes

$$\min f(x_1, x_2) := c \sum_{i=1}^{n} [(x_1 - a_i)^2 + (x_2 - b_i)^2]\,.$$

Since $f(x_1, x_2)$ is a strictly convex function, the optimal coordinates of D satisfy

$$\operatorname{grad} f(x_1, x_2) = \begin{pmatrix} 2c \sum_{i=1}^{n} (x_1 - a_i) \\ 2c \sum_{i=1}^{n} (x_2 - b_i) \end{pmatrix} = 0\,,$$

which is equivalent to

$$x_1 = \overline{a} = \frac{1}{n} \sum_{i=1}^{n} a_i \quad \text{and} \quad x_2 = \overline{b} = \frac{1}{n} \sum_{i=1}^{n} b_i\,.$$

We also present a geometric application:

Example 3.55. Given the four points $A = (a_1, a_2, a_3)^T, B = (b_1, b_2, b_3)^T, C = (c_1, c_2, c_3)^T$ and $D = (d_1, d_2, d_3)^T$ in the three-dimensional space ($A \neq B$, $C \neq D$), we determine the distance between the straight line G which passes through the points A and B and the line H passing through the points C and D.

The points G and H can be expressed by

$$\begin{aligned} G(x_1) &= x_1 A + (1 - x_1) B\,, \quad x_1 \in \mathbb{R} \\ \text{and} \quad H(x_2) &= x_2 C + (1 - x_2) D\,, \quad x_2 \in \mathbb{R}\,, \end{aligned}$$

respectively. Denoting the norm of a vector v by $|v|$, the distance between the points $G(x_1)$ and $H(x_2)$ is given by

$$\begin{aligned} f(x_1, x_2) = |G(x_1) - H(x_2)| &= \left\{ \sum_{i=1}^{3} [x_1 a_i + (1 - x_1) b_i - x_2 c_i - (1 - x_2) d_i]^2 \right\}^{1/2} \\ &= \left\{ \sum_{i=1}^{3} [b_i - d_i + (a_i - b_i) x_1 + (d_i - c_i) x_2]^2 \right\}^{1/2}, \end{aligned} \tag{3.19}$$

and the distance between the lines G and H is the minimum value of the function f.

Since the square root is an increasing function, minimizing $f(x_1, x_2)$ is equivalent to minimizing

$$g(x_1, x_2) = \sum_{i=1}^{3} [b_i - d_i + (a_i - b_i)x_1 + (d_i - c_i)x_2]^2 .$$

With the methods of this chapter it is easily verified that $g(x_1, x_2)$ is convex. A minimum point is given by

$$\begin{aligned} \operatorname{grad} g(x_1, x_2) &= 0 \quad \Leftrightarrow \\ 2\sum_{i=1}^{3}[b_i - d_i + (a_i - b_i)x_1 + (d_i - c_i)x_2](a_i - b_i) &= 0 \\ 2\sum_{i=1}^{3}[b_i - d_i + (a_i - b_i)x_1 + (d_i - c_i)x_2](d_i - c_i) &= 0 . \end{aligned}$$

The last condition is equivalent to the system of linear equations in the variables x_1, x_2:

$$\begin{aligned} x_1 \sum_{i=1}^{3}(a_i - b_i)^2 + x_2 \sum_{i=1}^{3}(a_i - b_i)(d_i - c_i) &= \sum_{i=1}^{3}(d_i - b_i)(a_i - b_i) \\ x_1 \sum_{i=1}^{3}(a_i - b_i)(d_i - c_i) + x_2 \sum_{i=1}^{3}(d_i - c_i)^2 &= \sum_{i=1}^{3}(d_i - b_i)(d_i - c_i) , \end{aligned}$$

which can be written as

$$\begin{aligned} x_1|A - B|^2 + x_2(A - B)^T(D - C) &= (D - B)^T(A - B) \\ x_1(A - B)^T(D - C) + x_2|D - C|^2 &= (D - B)^T(D - C) . \end{aligned} \tag{3.20}$$

The distance between the lines G and H is now given by $|G(x_1^*) - H(x_2^*)|$ where x^* is a solution of (3.20).

Exercise 3.56. *Determine the distance between the lines G and H for* $A = (1, 0, 2)^T$, $B = (2, 0, 2)^T$, $C = (0, 1, 3)^T$ *and* $D = (1, 1, 1)^T$.

Exercise 3.57. *Show that the following function is convex and determine the global minima:*

$$f(x_1, x_2) = 3x_1^2 + 5x_2^2 + 2x_1x_2 - 10x_1 - 22x_2 \quad (x \in \mathbb{R}^2) .$$

Finally we indicate an important result about concave functions in the following.

Theorem 3.58. *Given a convex, closed and bounded set* $M \subset \mathbb{R}^n$ *and a concave function* $f : M \to \mathbb{R}$, *then there exists an extreme point* $x^* \in M$ *which is a global minimum point of* f *on* M.

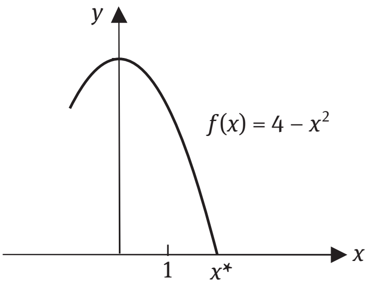


Fig. 3.17. Minimum point of a concave function.

Proof. Given $x \in M$, Theorem 3.12 implies that x can be represented as

$$x = \sum_{i=1}^{n+1} \lambda_i x^i$$

with $\lambda_i \geq 0$ for $i = 1, \ldots, n+1$, $\sum_{i=1}^{n+1} \lambda_i = 1$ and extreme points $x^1, \ldots, x^{n+1}$ of M.

Without loss of generality assume that

$$f(x^1) \leq f(x^i) \quad \text{for } i = 2, \ldots, n+1 .$$

The definition of concavity implies by induction (see Exercise 3.64):

$$f(x) = f\left(\sum_{i=1}^{n+1} \lambda_i x^i\right) \geq \sum_{i=1}^{n+1} \lambda_i f(x^i) \geq \sum_{i=1}^{n+1} \lambda_i f(x^1) = f(x^1) . \qquad \square$$

Example 3.59. The extreme point $x^* = 2$ is a global minimum point of the concave function $f(x) = 4 - x^2$ on the interval $M = [-1, 2]$ (Figure 3.17).

Exercise 3.60. *Find a global minimum point of the function $f : M \to \mathbb{R}$:*
(a) $f(x_1, x_2) = 1 - x_1^2 - 3x_2^2$, $M = \text{conv}\{x^1, \ldots, x^4\}$, *with* $x^1 = (-1, 0)^T$, $x^2 = (1, 3)^T$, $x^3 = (4, 2)^T$ *and* $x^4 = (4, -3)^T$.
(b) $f(x_1, x_2) = 1 - 2x_1^2 - 5x_2^2$, $M = \{(x_1, x_2)^T | x_1^2 + x_2^2 \leq 1\}$.

Exercise 3.61. *Does every convex function $f : M \to \mathbb{R}$ $(M \subset \mathbb{R}^n \text{convex})$ have a minimum point?*

Exercise 3.62. *Show that the following statements are equivalent for a convex function $f : M \to \mathbb{R}^n$.*
(i) *x^* is a global minimum point of f,*
(ii) *the vector $0 \in \mathbb{R}^n$ is a subgradient of f at x^*,*
(iii) *$D^+ f(x^*, y) \geq 0$ for all $y \in \mathbb{R}^n$.*

Exercise 3.63. *Show that a strictly convex function cannot have more than one global minimum point.*

Exercise 3.64. *Let* $f : M \to \mathbb{R}$ *be a concave function* $(M \subset \mathbb{R}^n \text{convex})$. *Show that*

$$f(\sum_{i=1}^{m} \lambda_i x^i) \geq \sum_{i=1}^{m} \lambda_i f(x^i)$$

for any points $x^1, \dots, x^m \in M$ *and* $\sum_{i=1}^{m} \lambda_i = 1, \lambda_i \geq 0$ *for* $i = 1, \dots, m$.

4 Karush–Kuhn–Tucker conditions and duality

In the following we express the first-order necessary optimality condition (Theorem 2.10) in a form which is more suitable for resolution procedures (see Theorem 4.6).

4.1 Karush–Kuhn–Tucker conditions

We consider the general problem of nonlinear programming (see (1.1))

$$\begin{aligned} &\min f(x) \qquad\qquad\qquad\qquad (4.1a)\\ &g_i(x) \le 0 \quad \text{for } i = 1, \dots, m \end{aligned}$$

with functions $f, g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \dots, m$. The feasible region is denoted by

$$M = \{x \in \mathbb{R}^n | g_i(x) \le 0 \quad \text{for } i = 1, \dots, m\}. \tag{4.1b}$$

We proved in Theorem 2.10 that a local minimum point x^* of a differentiable function satisfies the necessary condition

$$d^T \operatorname{grad} f(x^*) \ge 0 \quad \text{for all } d \in Z(x^*), \tag{4.2}$$

where $Z(x^*)$ is the cone of feasible directions at x^*. We make use of the following topological concept: the *closure* $\overline{M}$ of a set $M \subset \mathbb{R}^n$ is the union of M and its boundary. As an illustration we consider the cone $Z(x^*) = \{d \in \mathbb{R}^2 | d_1 \le d_2 < 4d_1\}$ of Example 2.8 which is "open at the left-hand side" (see Figure 2.5). The *closure* of this set is the closed cone $\overline{Z(x^*)} = \{d \in \mathbb{R}^2 | d_1 \le d_2 \le 4d_1\}$. Whenever the cone $L(x^*)$ is nonempty (see Definition 2.4) it holds $\overline{L(x^*)} = L(x^*) \cup \{0\}$, but in general $\overline{Z(x^*)} \ne Z(x^*) \cup \{0\}$. Since $d^T \operatorname{grad} f(x^*)$ is a continuous function at d we can write condition (4.2) as

$$d^T \operatorname{grad} f(x^*) \ge 0 \quad \text{for all } d \in \overline{Z(x^*)}. \tag{4.3}$$

From topology it is known that $M \subset N \subset \mathbb{R}^n$ implies $\overline{M} \subset \overline{N}$, hence it follows from Theorem 2.5 that

$$\overline{Z(x^*)} \subset \overline{L(x^*)}.$$

In applications these sets are usually equal, however as illustrated in the following, $\overline{Z(x^*)}$ can be a proper subset of $\overline{L(x^*)}$.

Definition 4.1. Given the set M in (4.1 (b)), the point $x^* \in M$ is called *regular* if

$$\overline{Z(x^*)} = \overline{L(x^*)}.$$

The subsequent transformations of the optimality conditions assume that x^* is regular, however, this condition is not very restrictive. There exist many sufficient conditions for regularity in the literature. We formulate some of the most important conditions without a proof.

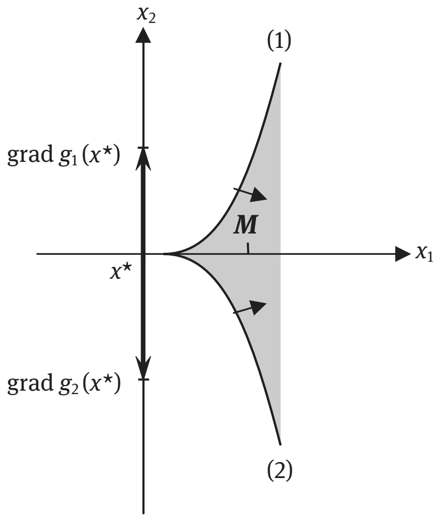


Fig. 4.1. Nonregular point $x^* = 0$.

Theorem 4.2. *Assume that the functions g_i in (4.1) are continuously differentiable.*
(a) *If one of the following conditions is satisfied, every point of M is regular:*
  (i) *The functions g_i are convex and there is a point $\tilde{x} \in M$ with $g_i(\tilde{x}) < 0$ for $i = 1, \ldots, m$ (Slater condition).*
  (ii) *The functions g_i are linear (see Theorem 2.6).*
(b) *The point $x^* \in M$ is regular, if the gradients* grad $g_i(x^*)$ *with $i \in A(x^*)$ are linearly independent.*

Example 4.3. Given the set (see Figure 4.1)

$$M = \{(x_1, x_2)^T \in \mathbb{R}^2 | g_1(x_1, x_2) := x_2 - x_1^3 \le 0, g_2(x_1, x_2) := -x_2 - x_1^3 \le 0\} .$$

It holds that grad $g_1(x_1, x_2) = (-3x_1^2, 1)^T$ and grad $g_2(x_1, x_2) = (-3x_1^2, -1)^T$. These vectors are linearly independent for $x_1 > 0$ and dependent for $x_1 = 0$. Theorem 4.2 (b) implies that every point of M is regular except for the point $x^* = 0$. As $Z(x^*) = \overline{Z(x^*)} = \{(d_1, 0)^T | d_1 \ge 0\}$ and $L(x^*) = \overline{L(x^*)} = \{(d_1, 0)^T | d_1 \in \mathbb{R}\} \ne \overline{Z(x^*)}$, x^* is not regular.

Exercise 4.4. *Show by modifying the last example adequately that the condition of Theorem 4.2 (b) is not necessary for regularity. Determine the irregular points for every set M in Exercise 2.9.*

We call attention to the fact that the concept of regular point can be defined in different ways. For example, some authors call a point x^* regular when the sufficient condition of Theorem 4.2 (b) is satisfied. Instead of saying that a point x^* is regular (with respect to a given criterion) it is frequently said that x^* satisfies a certain *constraint qualification*: for example, the sufficient condition of Theorem 4.2 (b) is called *linear independence constraint qualification (LICQ)*.

If x^* is a regular point, condition (4.3) can now be reformulated as

$$(\operatorname{grad} f(x^*))^T d \geq 0 \quad \text{for all } d \in \overline{L(x^*)}\,. \tag{4.4}$$

Denoting the indices of active constraints at x^* by $i_1, \ldots, i_p$, (4.4) can be written as

$$(\operatorname{grad} f(x^*))^T d \geq 0 \quad \text{for all } d \in \mathbb{R}^n \tag{4.5}$$

with

$$A^T d \leq 0\,,$$

where $A \in \mathbb{R}^{n \times p}$ is the matrix whose jth column is $a^j := \operatorname{grad} g_{i_j}(x^*)$ for $j = 1, \ldots, p$.

We make use of the following classical result on linear inequalities:

Theorem 4.5 (Farkas lemma). *For $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times p}$ the following statements are equivalent:*
(i) $b^T x \geq 0$ *for all* $x \in \mathbb{R}^n$ *with* $A^T x \leq 0$,
(ii) *there exists a vector* $u^* \in \mathbb{R}^p_+$ *with* $Au^* = -b$.

Farkas lemma can be interpreted geometrically: we consider the convex cone $X = \{x \in \mathbb{R}^n | A^T x \leq 0\}$, i.e. the set of vectors $x \in \mathbb{R}^n$ forming an angle ≥ 90 degrees with every vector a^j (see Figure 4.2) and the convex cone $C = \{Au | u \in \mathbb{R}^p_+\}$, i.e. the cone generated by the vectors a^j. The lemma states that a vector b forms an angle $\leq$ 90 degrees with each vector of X, if and only if $-b$ is in the cone C.

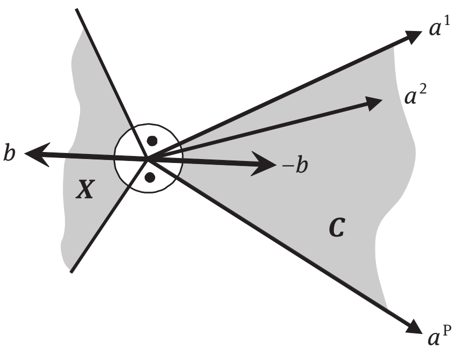


Fig. 4.2. Illustration of the Farkas lemma.

Applying the Farkas lemma to condition (4.5), we obtain the following equivalent condition. There are numbers $u_{i_j} \geq 0$ such that

$$-\operatorname{grad} f(x^*) = \sum_{j=1}^{p} u_{i_j} \operatorname{grad} g_{i_j}(x^*)\,. \tag{4.6}$$

Finally, the sum in (4.6) may be replaced by

$$\sum_{i=1}^{m} u_i^* \operatorname{grad} g_i(x^*)$$

if we set

$$u_i^* = 0 \quad \text{for } i \notin A(x^*)\,. \tag{4.7}$$

Since $g_i(x^*) = 0$ for $i \in A(x^*)$ and $g_i(x^*) < 0$ for $i \notin A(x^*)$, (4.7) is equivalent to

$$u_i^* g_i(x^*) = 0 \quad \text{for } i = 1, \ldots, m\,. \tag{4.8}$$

We obtain the following *necessary* condition for a local minimum point.

Theorem 4.6. *Given the problem (4.1) with continuously differentiable functions f, g_i and a local minimum point x^* of f over M, if x^* is regular, then there exists a vector $u^* = (u_1^*, \ldots, u_m^*)^T \in \mathbb{R}^m$ such that*

$$\operatorname{grad} f(x^*) + \sum_{i=1}^{m} u_i^* \operatorname{grad} g_i(x^*) = 0 \tag{4.9}$$

$$u_i^* g_i(x^*) = 0 \quad \textit{for } i = 1, \ldots, m \tag{4.10}$$

$$g_i(x^*) \leq 0 \quad \textit{for } i = 1, \ldots, m \tag{4.11}$$

$$u_i^* \geq 0 \quad \textit{for } i = 1, \ldots, m\,. \tag{4.12}$$

The relations (4.9)–(4.12) are called Karush–Kuhn–Tucker conditions (KKT conditions, also known as Kuhn–Tucker conditions; see the references in Section 4.5). In particular, (4.11) is called primal feasibility, (4.9) and (4.12) are called dual feasibility and (4.10) is called complementary slackness. The components $u_1^*, \ldots, u_m^*$ of the vector u^* are called Lagrange multipliers.

In the following we interpret the KKT conditions geometrically. For this purpose we introduce the concept of descent direction (compare with Definition 2.1):

Definition 4.7. Given a function $f : \mathbb{R}^n \to \mathbb{R}$ and a point $x^* \in \mathbb{R}^n$, a vector $d \in \mathbb{R}^n$ is called a *descent direction* of f (at x^*), if there is a $\lambda_0 > 0$ such that

$$f(x^* + \lambda d) < f(x^*) \tag{4.13}$$

is satisfied for $0 < \lambda \leq \lambda_0$.

If f is differentiable, $f(x^* + \lambda d)$ can be expressed by using the Taylor expansion (see the proof of the Theorem 2.10):

$$f(x^* + \lambda d) = f(x^*) + \lambda d^T \operatorname{grad} f(x^*) + R_1(\lambda)\,.$$

In this case (4.13) is equivalent to

$$d^T \operatorname{grad} f(x^*) + \frac{R_1(\lambda)}{\lambda} < 0\,.$$

Since $\lim_{\lambda\to 0} \frac{R_1(\lambda)}{\lambda} = 0$, the condition

$$d^T \operatorname{grad} f(x^*) < 0 \tag{4.14}$$

is sufficient for d to be a descent direction of f at x^*.

In analogy to Definition 4.7, an *ascent direction* can be defined by substituting (4.13) for $f(x^* + \lambda d) > f(x^*)$. In this case, the condition $d^T \operatorname{grad} f(x^*) > 0$ is sufficient for d to be an ascent direction of f at x^*.

If $d^T \operatorname{grad} f(x^*) = 0$, the vector d can be descent direction or ascent direction. For example, if $f(x_1, x_2) = -x_1^2 - x_2^2 + x_1 + x_2$, then $d = (-1, 1)^T$ is a descent direction of f at x^* with $d^T \operatorname{grad} f(x^*) = 0$. On the other hand, if $f(x_1, x_2) = x_1^2 + x_2^2 - x_1 - x_2$, then $d = (-1, 1)^T$ is an ascent direction of f at $x^* = 0$ with $d^T \operatorname{grad} f(x^*) = 0$.

In consequence of these considerations a vector d will be called a *direction of strict descent (strict ascent)* if $d^T \operatorname{grad} f(x^*) < 0 (d^T \operatorname{grad} f(x^*) > 0)$.[1] Geometrically, a direction of strict descent (strict ascent) of f at x^* forms an acute (obtuse) angle with the vector $-\operatorname{grad} f(x^*)$.

Now the KKT conditions (in the form (4.6)) may be interpreted as follows: In a local minimum point x^*, the vector $-\operatorname{grad} f(x^*)$ is a nonnegative linear combination of the vectors $\operatorname{grad} g_i(x^*)$ corresponding to active constraints at x^*, i.e. $-\operatorname{grad} f(x^*)$ lies in the cone spanned by these vectors (see Figure 4.3). Hence, any feasible direction forms an angle ≥ 90 degrees with the vector $-\operatorname{grad} f(x^*)$. (Note that in Figure 4.3 the vector $\operatorname{grad} g_i(x^*)$ is perpendicular to the tangent T_i to the curve $g_i = 0$ at x^*.) In other words, no feasible direction is a direction of strict descent. This is obviously necessary for x^* to be a local minimum point of f.

Example 4.8. We determine all solutions of the KKT conditions, given the problem

$$\begin{aligned} \min -(x_1+2)^2 - (x_2-1)^2 \\ -x_1 + x_2 - 2 \leq 0 \\ x_1^2 - x_2 \leq 0. \end{aligned}$$

1 Some authors define a descent (ascent) direction by the condition $d^T \operatorname{grad} f(x^*) < 0$ $(d^T \operatorname{grad} f(x^*) > 0)$.

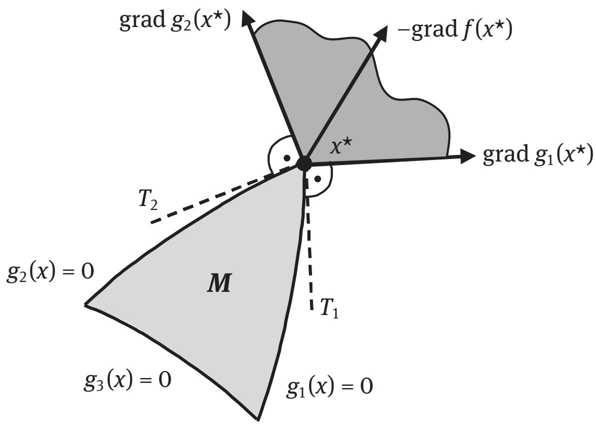


Fig. 4.3. Geometrical illustration of KKT conditions.

We obtain the system

$$\begin{aligned} -2x_1 - 4 - u_1 + 2x_1u_2 &= 0 \\ -2x_2 + 2 + u_1 - u_2 &= 0 \\ u_1(-x_1 + x_2 - 2) &= 0 \\ u_2(x_1^2 - x_2) &= 0 \end{aligned} \tag{4.15}$$

$$\begin{aligned} -x_1 + x_2 - 2 &\le 0 \\ x_1^2 - x_2 &\le 0 \\ u_1, u_2 &\ge 0\,. \end{aligned} \tag{4.16}$$

The basic idea of resolution, which can be easily generalized, is the following: we consider the four cases $u_1 = u_2 = 0$; $u_1 = 0, u_2 > 0$; $u_1 > 0, u_2 = 0$; $u_1, u_2 > 0$. For each case we solve the subsystem of equations and check whether these solutions also satisfy the inequalities.

(i) Case $u_1 = u_2 = 0$:
The system (4.15)–(4.16) reduces to

$$\begin{aligned} -2x_1 - 4 &= 0 \\ -2x_2 + 2 &= 0 \end{aligned} \tag{4.17}$$

$$\begin{aligned} -x_1 + x_2 - 2 &\le 0 \\ x_1^2 - x_2 &\le 0 \end{aligned} \tag{4.18}$$

which has no solution since the only solution $(x_1, x_2) = (-2, 1)$ of (4.17) does not satisfy (4.18).

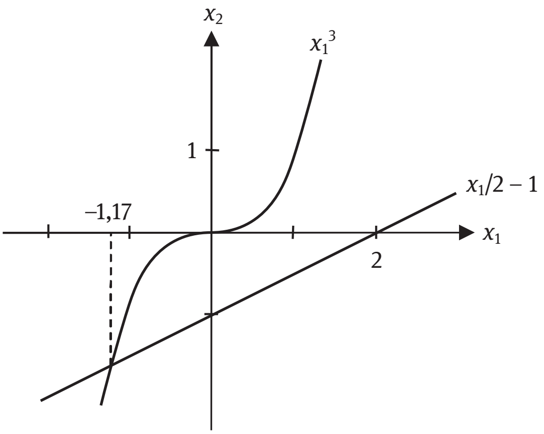


Fig. 4.4. Illustration of the relation (4.22).

(ii) Case $u_1 = 0, u_2 > 0$:
The system (4.15)–(4.16) results in

$$\begin{aligned} -2x_1 - 4 + 2x_1u_2 &= 0 \\ -2x_2 + 2 - u_2 &= 0 \\ x_1^2 &= x_2 \end{aligned} \tag{4.19}$$

$$-x_1 + x_2 - 2 \leq 0 \tag{4.20}$$
$$u_2 > 0\,.$$

The system (4.19) implies

$$\begin{aligned} x_2 &= x_1^2 \\ u_2 = 2 - 2x_2 &= 2(1 - x_1^2)\,. \end{aligned} \tag{4.21}$$

By substituting u_2 for $2(1 - x_1^2)$ in the first equation of (4.19) we obtain

$$-2x_1 - 4 + 2x_1 2(1 - x_1^2) = 0\,,$$

which is equivalent to

$$x_1^3 = \frac{x_1}{2} - 1\,. \tag{4.22}$$

Since the only real solution of (4.22) satisfies $x_1 < -1$ (see Figure 4.4), (4.21) implies $u_2 < 0$, contradicting (4.20).

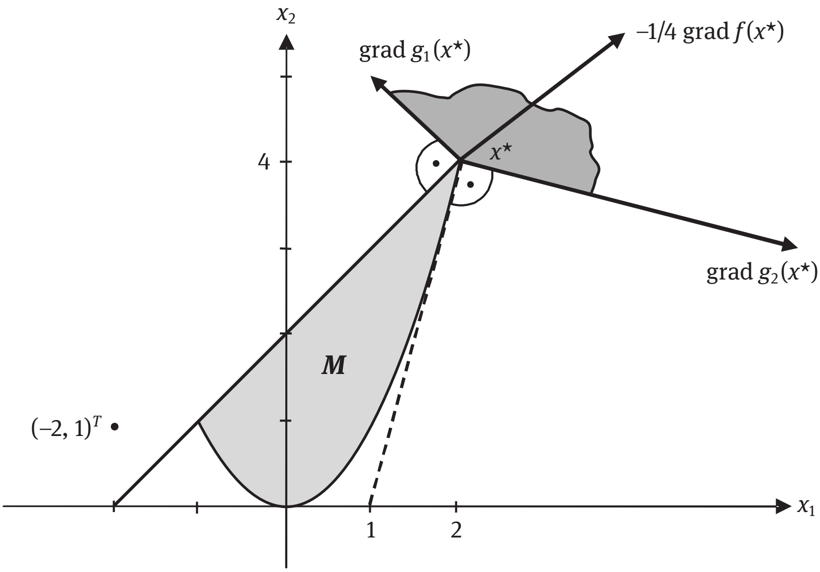


Fig. 4.5. Illustration of Example 4.8.

(iii) Case $u_1 > 0, u_2 = 0$:
The system (4.15)–(4.16) results in

$$\begin{aligned} -2x_1 - 4 - u_1 &= 0 \\ -2x_2 + 2 + u_1 &= 0 \\ -x_1 + x_2 - 2 &= 0 \end{aligned} \tag{4.23}$$

$$\begin{aligned} x_1^2 - x_2 &\le 0 \\ u_1 &> 0\,. \end{aligned} \tag{4.24}$$

The unique solution $(x_1, x_2, u_1) = (-3/2, 1/2, -1)$ of (4.23) does not satisfy (4.24).

(iv) Case $u_1, u_2 > 0$:
The system (4.15)–(4.16) results in

$$\begin{aligned} -2x_1 - 4 - u_1 + 2x_1u_2 &= 0 \\ -2x_2 + 2 + u_1 - u_2 &= 0 \\ -x_1 + x_2 &= 2 \\ x_1^2 &= x_2 \end{aligned} \tag{4.25}$$

$$u_1, u_2 > 0\,. \tag{4.26}$$

The last two equations of (4.25) imply

$$\begin{aligned} & x_1 = 2 \quad \text{and} \quad x_2 = 4 \\ \text{or} \quad & x_1 = -1 \quad \text{and} \quad x_2 = 1\,. \end{aligned}$$

For $x_1 = -1$ and $x_2 = 1$, the system (4.25) results in $u_1 = u_2 = -2/3$, i.e. (4.26) is not satisfied. For $x_1 = 2$ and $x_2 = 4$ we obtain $u_1 = 32/3$ and $u_2 = 14/3$. This solution satisfies (4.26).

Hence, the only solution of the KKT conditions (4.15)–(4.16) is given by

$$(x_1, x_2, u_1, u_2) = (2, 4, 32/3, 14/3)\,. \tag{4.27}$$

It is easily verified in Figure 4.5 that $x^* = (2,4)^T$ is in fact a local (and global) minimum point. The value $-f(x)$ is the square of the distance between the points x and $(-2,1)^T$. This value is minimized by the point x^*.

Finally, Figure 4.5 shows that $-\operatorname{grad} f(x^*) = (8,6)^T$ is contained in the cone generated by $\operatorname{grad} g_1(x^*) = (-1,1)^T$ and $\operatorname{grad} g_2(x^*) = (4,-1)^T$, i.e. the KKT conditions are satisfied, interpreting them geometrically.

Exercise 4.9. *Solve the KKT conditions for the problem*

$$\begin{aligned} &\min -x_1 - x_2 \\ &-x_1 + x_2 - 1 \le 0 \\ &x_1 + x_2^2 - 1 \le 0\,. \end{aligned}$$

Verify geometrically that the solution obtained is indeed a global minimum point.

Exercise 4.10. *Obviously the point $x^* = 0 \in \mathbb{R}^2$ is a global minimum point for the problem (see Example 4.3):*

$$\begin{aligned} &\min x_1 \\ &x_2 - x_1^3 \le 0 \\ &-x_2 - x_1^3 \le 0\,. \end{aligned}$$

Show that the KKT conditions have no solution. Explain!

We also consider some consequences and expansions of Theorem 4.6. For example, in the case of a convex optimization problem, the necessary condition for a local minimum is also sufficient for a global minimum (see Section 3.5). Therefore, Theorem 4.6 implies the following result:

Theorem 4.11. *Assume that the problem (4.1) is convex (Definition 3.21) with continuously differentiable functions f, g_i, and that all points of M are regular. Then the point x^* is a global minimum point of f on M, if and only if there is a vector u^* such that (x^*, u^*) satisfies the KKT conditions.*

Moreover it can be shown that the criterion is sufficient even when x^* is not regular, i.e. when the functions f and g_i are differentiable and convex and (x^*, u^*) satisfies the KKT conditions for $u^* \in \mathbb{R}^m$, then x^* is a global minimum point.

Without a proof we present a generalization of Theorem 4.6 for problems including equality constraints.

Theorem 4.12. *Given continuously differentiable functions* $f, g_1, \dots, g_m, h_1, \dots, h_k : \mathbb{R}^n \to \mathbb{R}$ *and the feasible region*

$$M = \{x \in \mathbb{R}^n | g_i(x) \le 0 \quad \text{for } i = 1, \dots, m, \quad h_j(x) = 0 \quad \text{for } j = 1, \dots, k\}.$$

Let $x^* \in M$ *be a local minimum point of* f *over* M *and*

$$A(x^*) = \{i \le m | g_i(x) = 0\}$$

the set of active inequalities. Suppose that the vectors $\operatorname{grad} h_j(x^*)$ *are linearly independent for* $j = 1, \dots, k$ *and that there is a vector* $z \in \mathbb{R}^n$ *such that*

$$\begin{aligned} z^T \operatorname{grad} g_i(x^*) &< 0 \quad \text{for all } i \in A(x^*) \quad \text{and} \\ z^T \operatorname{grad} h_j(x^*) &= 0 \quad \text{for all } j \le k \end{aligned}$$

(i.e. the so-called Mangasarian–Fromovitz constraint qualifications are satisfied). There exist vectors $u^* = (u_1^*, \dots, u_m^*)^T \in \mathbb{R}^m$ *and* $v^* = (v_1^*, \dots, v_k^*)^T \in \mathbb{R}^k$ *such that*

$$\operatorname{grad} f(x^*) + \sum_{i=1}^m u_i^* \operatorname{grad} g_i(x^*) + \sum_{j=1}^k v_j^* \operatorname{grad} h_j(x^*) = 0 \tag{4.28}$$

$$u_i^* g_i(x^*) = 0 \quad \text{for } i \le m \tag{4.29}$$

$$g_i(x^*) \le 0 \quad \text{for } i \le m \tag{4.30}$$

$$h_j(x^*) = 0 \quad \text{for } j \le k \tag{4.31}$$

$$u_i^* \ge 0 \quad \text{for } i \le m. \tag{4.32}$$

It can be shown that the above conditions of Mangasarian–Fromovitz are satisfied, if the vectors $\operatorname{grad} g_i(x^*)$ for $i \in A(x^*)$ and $\operatorname{grad} h_j(x^*)$ for $j = 1, \dots, k$ are linearly independent.

Exercise 4.13. *Prove the last statement.*

Example 4.14. Consider the problem

$$\begin{aligned} &\min(x_1 - 1)^2 + (x_2 - 2)^2 \\ g(x_1, x_2) &= -x_1^2 + x_2 \le 0 \\ h(x_1, x_2) &= x_1 + x_2 - 2 = 0. \end{aligned}$$

It holds that $\operatorname{grad} g(x) = (-2x_1, 1)^T$ and $\operatorname{grad} h(x) = (1, 1)^T$, i.e. for every feasible point x, these vectors are linearly independent (see Figure 4.6) and the conditions of Mangasarian–Fromovitz are satisfied. Thus, a local minimum point satisfies the following KKT conditions:

$$\begin{aligned}
2x_1 - 2 - 2ux_1 + v &= 0\\
2x_2 - 4 + u + v &= 0\\
u(-x_1^2 + x_2) &= 0\\
-x_1^2 + x_2 &\le 0\\
x_1 + x_2 &= 2\\
u &\ge 0\,.
\end{aligned} \tag{4.33}$$

Case $u = 0$:
We obtain the system:

$$\begin{aligned}
2x_1 - 2 + v &= 0\\
2x_2 - 4 + v &= 0\\
-x_1^2 + x_2 &\le 0\\
x_1 + x_2 &= 2\,.
\end{aligned}$$

It can be easily verified that there is no solution, since the only solution $(x_1, x_2, v) = (1/2, 3/2, 1)$ of the subsystem of equations does not satisfy the inequality.

Case $u > 0$:
We obtain the system:

$$\begin{aligned}
2x_1 - 2 - 2ux_1 + v &= 0\\
2x_2 - 4 + u + v &= 0\\
-x_1^2 + x_2 &= 0\\
x_1 + x_2 &= 2\\
u &> 0.
\end{aligned}$$

Solving first the last two equations, we obtain the following solutions of the subsystem of equations which also satisfy the inequality $u > 0$:

$$(x_1, x_2, u, v) = (1, 1, 2/3, 4/3)\,, \quad (x_1, x_2, u, v) = (-2, 4, 10/3, -22/3)\,.$$

Hence, the KKT conditions (4.33) provide the two “candidates” $x^1 = (1, 1)^T$ and $x^2 = (-2, 4)^T$ for local minimum points. The objective value $f(x)$ corresponds to the square of the distance between the points x and $(1, 2)^T$ (Figure 4.6). Thus, x^1 and x^2 are local minimum points of f. In fact, x^1 is a global minimum point.

Exercise 4.15. *Determine the solutions to the KKT conditions for problem*

$$\begin{aligned}
\min -2x_1 - x_2&\\
g(x_1, x_2) = x_1 + 2x_2 - 6 &\le 0\\
h(x_1, x_2) = x_1^2 - 2x_2 &= 0\,.
\end{aligned}$$

Which of these solutions represent minimum points? Illustrate the problem graphically.

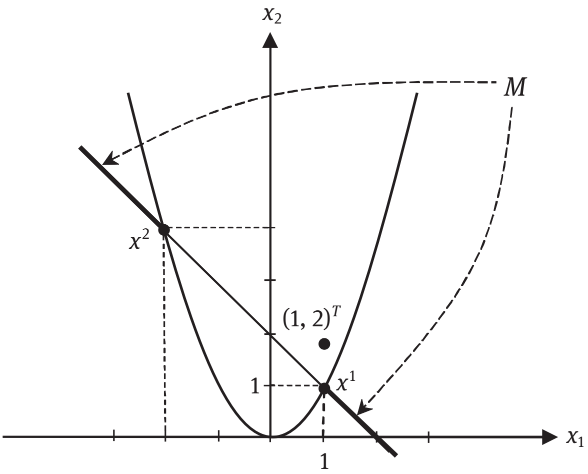


Fig. 4.6. Illustration of Example 4.14.

Finally we apply the theory of this section to solve the quadratic problem

$$\begin{aligned} &\min \tfrac{1}{2}x^T Qx + c^T x \\ &A^T x = b\,, \end{aligned} \tag{4.34}$$

where $Q \in \mathbb{R}^{n\times m}$ is symmetric and positive definite, $A \in \mathbb{R}^{n\times m}$ is a matrix with $\text{rank}(A) = m \le n$, and $b \in \mathbb{R}^m$. Theorem 3.26 implies that f is strictly convex, thus (4.33) is a convex problem. Moreover, every feasible point of (4.34) is regular, since the constraints are linear. Hence, the global minima correspond biuniquely to solutions of the KKT conditions and can be written as (see Theorems 4.11 and 4.12):

$$\begin{aligned} Qx^* + c + Av^* &= 0 \ (v^* \in \mathbb{R}^m) \\ A^T x^* &= b\,. \end{aligned} \tag{4.35}$$

Exercise 4.16. *Derive the conditions (4.35) step by step!*

From the first equation of (4.35) we get

$$x^* = -Q^{-1}(c + Av^*) \tag{4.36}$$

and the second implies:

$$\begin{aligned} -A^T Q^{-1}(c + Av^*) &= b \Rightarrow \\ A^T Q^{-1} A v^* &= -(b + A^T Q^{-1} c) \Rightarrow \\ v^* &= -(A^T Q^{-1} A)^{-1}(b + A^T Q^{-1} c)\,. \end{aligned} \tag{4.37}$$

Combining (4.36) and (4.37) we can express the unique global minimum point of (4.34) by

$$x^* = Q^{-1}[A(A^T Q^{-1} A)^{-1}(b + A^T Q^{-1} c) - c]\,. \tag{4.38}$$

Exercise 4.17. *Determine the global minimum point of the quadratic problem*

$$\min x_1^2 + x_1x_2 + \tfrac{3}{2}x_2^2 - 3x_1 + 5x_2$$
$$2x_1 + x_2 = 5\,.$$

(a) *Write the objective function in the form (4.34) and apply (4.38).*
(b) *Solve the problem alternatively by substitution of variables.*

Exercise 4.18.
(a) *Solve the last optimization problem of Example 1.17 with the aid of the KKT conditions.*
(b) *What are the optimal measures of a box? How much metal foil is needed for the construction of such a box?*
(c) *Let us modify the problem of part (a) by replacing the restriction "$x_2 \geq x_1$" for "$x_2 = x_1$". Solve this one-dimensional problem and compare the result with the solution of part (a). Explain!*

The examples and exercises above show that not every optimization problem can be "easily" solved with the aid of the KKT conditions. Even simple problems can result in a complicated KKT system, the resolution of which requires a lot of creativity. But usually not all cases of the resolution need to be exhaustively studied. By means of evaluation of the objective function the search region can be restricted and one can identify constraints which can not be active at the optimum. In the case of bidimensional functions, contour lines (see Section 7.1) are useful to identify optimal solutions. The theory outlined above provides the basis for most of the solution procedures of nonlinear programming.

The optimality conditions of this section are based on classical results of Karush, Kuhn and Tucker developed in the 1930s and 1950s. The investigation of optimality conditions for specific nonlinear programming problems is the subject of several recent and current research projects (see Section 4.5).

4.2 Lagrange function and duality

In this section we will formulate optimality conditions alternatively by introducing the concept of the *saddle point of the Lagrange function* (Theorems 4.21 and 4.22). The following results play an important role in game theory and economics. Moreover, we can define a dual problem for any nonlinear problem via the Lagrange function.

We consider again the nonlinear problem

$$\begin{aligned} \min f(x) \\ g(x) \leq 0 \end{aligned} \tag{4.39}$$

with $g(x) = (g_1(x), \dots g_m(x))^T$ (see (4.1)).

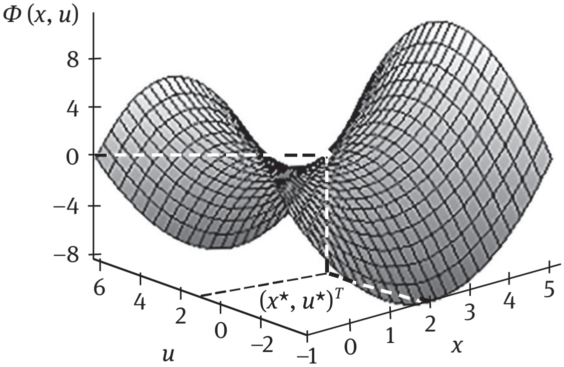


Fig. 4.7. Saddle point.

Definition 4.19. The function $\Phi : \mathbb{R}^{n+m} \to \mathbb{R}$ with

$$\Phi(x,u) = f(x) + u^T g(x)\,,$$

$x \in \mathbb{R}^n$, $u \in \mathbb{R}^m_+$ is called the *Lagrange function* of problem (4.39).

A point $(x^*, u^*) \in \mathbb{R}^{n+m}$ with $u^* \geq 0$ is called a *saddle point of the Lagrange function* $\Phi(x,u)$, if

$$\Phi(x^*, u) \leq \Phi(x^*, u^*) \leq \Phi(x, u^*) \tag{4.40}$$

is satisfied for all $x \in \mathbb{R}^n$ and all $u \in \mathbb{R}^m_+$.

Condition (4.40) is equivalent to

$$\Phi(x^*, u^*) = \max_{u \in \mathbb{R}^m_+} \Phi(x^*, u) = \min_{x \in \mathbb{R}^n} \Phi(x, u^*)\,. \tag{4.41}$$

Figure 4.7 illustrates schematically a Lagrange function with $n = m = 1$ and saddle point $(x^*, u^*)^T = (2,1)^T$.

The following statement establishes the relation between the existence of a saddle point of Φ and the KKT conditions:

Theorem 4.20. *The point $(x^*, u^*) \in \mathbb{R}^{n+m}$ is a saddle point of the Lagrange function $\Phi(x,u)$, if and only if the following conditions are satisfied:*

$$\Phi(x^*, u^*) = \min_{x \in \mathbb{R}^n} \Phi(x, u^*) \tag{i}$$

$$g(x^*) \leq 0 \tag{ii}$$

$$u^{*T} g(x^*) = 0 \tag{iii}$$

$$u^* \geq 0. \tag{iv}$$

Proof. (a) Let (i)–(iv) be satisfied. Condition (i) implies the second inequality of (4.40), and from (ii)–(iv) results for all $u \geq 0$:

$$\Phi(x^*, u) = f(x^*) + u^T g(x^*) \leq f(x^*) + u^{*T} g(x^*) = \Phi(x^*, u^*)\,, \tag{4.42}$$

i.e. the first inequality of (4.40) is also satisfied.

(b) Let (x^*, u^*) be a saddle point. Then (i) and (iv) are obviously satisfied and

$$\Phi(x^*, u) \leq \Phi(x^*, u^*)$$

holds for all $u \geq 0$. We obtain (see (4.42))

$$u^T g(x^*) \leq u^{*T} g(x^*) \tag{4.43}$$

for all $u \geq 0$.
For $u = u^* + e^i \quad (i = 1, \dots, m; e^i$ denotes the ith unit vector of $\mathbb{R}^m$)
we obtain from (4.43):

$$\begin{gathered}(u^{*T} + e^{iT}) g(x^*) \leq u^{*T} g(x^*) \Rightarrow \\ e^{iT} g(x^*) \leq 0 \Rightarrow \\ g_i(x^*) \leq 0.\end{gathered}$$

Hence, (ii) is satisfied. Equation (4.43) implies for $u = 0$:

$$0 \leq u^{*T} g(x^*)\,.$$

Moreover, $u^* \geq 0$ and $g(x^*) \leq 0$ imply

$$u^{*T} g(x^*) \leq 0\,.$$

We obtain $u^{*T} g(x^*) = 0$, i.e. (iii) is satisfied. □

We observe that the conditions (ii)–(iv) of Theorem 4.20 are equivalent to the KKT conditions (4.10)–(4.12) of Theorem 4.6.

Theorem 4.21. *If (x^*, u^*) is a saddle point of the Lagrange function $\Phi(x, u)$, then x^* is an optimal solution of problem (4.39).*

Proof. If (x^*, u^*) is a saddle point, the conditions (i)–(iv) of Theorem 4.20 are valid. Condition (i) implies

$$\Phi(x^*, u^*) = f(x^*) + u^{*T} g(x^*) \leq \Phi(x, u^*) = f(x) + u^{*T} g(x) \quad \text{for all } x\,.$$

From (iii) we obtain

$$f(x^*) \leq f(x) + u^{*T} g(x) \quad \text{for all } x\,.$$

Because of (ii) and (iv), the term $u^{*T} g(x)$ is negative, thus

$$f(x^*) \leq f(x) \quad \text{for all } x\,.$$ □

Suppose now that the functions $f, g_1, \dots, g_m$ in (4.39) are continuously differentiable and *convex*. For a fixed $u^* \in \mathbb{R}_+^m$ the Lagrange function

$$\Phi(x, u^*) = f(x) + \sum_{i=1}^{m} u_i^* g_i(x)$$

is continuously differentiable and convex and achieves its minimum at a point x^*, satisfying $\operatorname{grad}_x \Phi(x^*, u^*) = 0$. Thus condition (i) of Theorem 4.20 is equivalent to

$$\operatorname{grad}_x \Phi(x^*, u^*) = \operatorname{grad} f(x^*) + \sum_{i=1}^{m} u_i^* \operatorname{grad} g_i(x^*) = 0\,,$$

which is the KKT condition (4.9) of Theorem 4.6.

Using Theorem 4.11 we obtain the following result:

Theorem 4.22. *Let (4.39) be a convex problem with continuously differentiable functions f, g_i, such that every feasible point is regular. Then x^* is a global minimum point of (4.39), if and only if there exists a vector $u^* \in \mathbb{R}_+^m$ such that (x^*, u^*) is a saddle point of the Lagrange function.*

Before defining the dual problem of a nonlinear problem we recall here the concepts of supremum and infimum of a function.

Definition 4.23. Given a function $h : M \to \mathbb{R}, M \subset \mathbb{R}^n$, the *supremum* of h (over M), denoted by $\sup_{x \in M} h(x)$, is defined as the smallest number $m \in \mathbb{R} \cup \{\infty\}$ such that $m \geq h(x)$ for all $x \in M$. By analogy, the *infimum* of h (over M), denoted by $\inf_{x \in M} h(x)$, is defined as the largest number $m \in \mathbb{R} \cup \{-\infty\}$ such that $m \leq h(x)$ for all $x \in M$.

Note that supremum and infimum always exist and that these values are equal to the maximum or minimum, respectively, when the latter exist.

Example 4.24.
(i) Consider the function $h(x) = 1/x$ e $M = (0, \infty)$.
We have $\sup_{x \in M} h(x) = \infty$, $\inf_{x \in M} h(x) = 0$. Neither the maximum nor the minimum exists.
(ii) Given $h(x) = x^2$ over $M = [-1, 1]$.
It holds $\sup_{x \in M} h(x) = \max_{x \in M} h(x) = 1$, $\inf_{x \in M} h(x) = \min_{x \in M} h(x) = 0$.

Now we reformulate the problem (4.39):

For the Lagrange function $\Phi(x, u) = f(x) + u^T g(x)$ it holds that

$$\sup_{u \in \mathbb{R}_+^m} \Phi(x, u) = \begin{cases} f(x) & \text{for } g(x) \leq 0 \\ \infty & \text{otherwise,} \end{cases}$$

where $g(x) \le 0$ means that every component of the vector $g(x)$ is nonpositive. Thus, (4.39) is equivalent to

$$\min_{x \in \mathbb{R}^n} \sup_{u \in \mathbb{R}^m_+} \Phi(x, u)\,. \tag{P}$$

This problem is called the *primal problem*.

Definition 4.25. The problem

$$\max_{u \in \mathbb{R}^m_+} \inf_{x \in \mathbb{R}^n} \Phi(x, u) \tag{D}$$

is called the *(Lagrange) dual problem* of (P).

The following example illustrates the last definition and the results that will be developed below.

Example 4.26. Consider the primal problem (see Figure 4.8)

$$\begin{aligned} &\min (x_1 + 3)^2 + x_2^2 \\ &x_1^2 - x_2 \le 0\,. \end{aligned} \tag{P}$$

For each point $(x_1, x_2)^T$ of the feasible set $M = \{(x_1, x_2)^T | x_1^2 \le x_2\}$, the value $f(x_1, x_2)$ is the square of the distance to the point $(-3, 0)^T$. The point $x^* = (-1, 1)^T$ minimizes the function $f(x_1, x_2)$ over M. In fact, x^* is the unique optimal point of (P) and $f(x^*) = 5$.

The Lagrange function is

$$\Phi(x_1, x_2, u) = (x_1 + 3)^2 + x_2^2 + u(x_1^2 - x_2)\,.$$

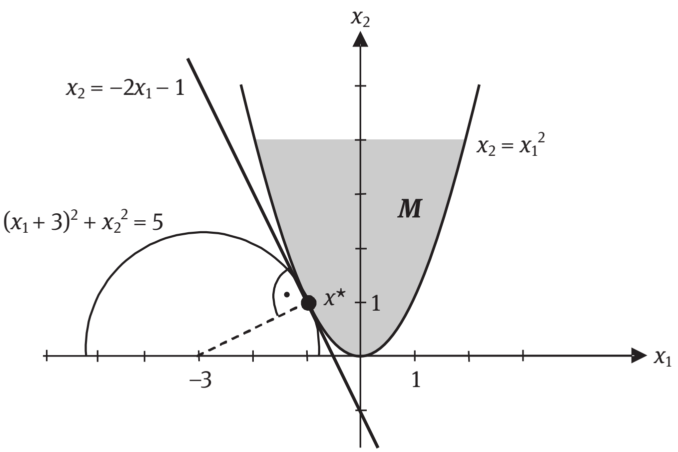


Fig. 4.8. Geometric illustration of (P).

For a fixed value $u \geq 0$, $\Phi(x_1, x_2, u)$ is convex and the unique minimum point $(x_1^*, x_2^*)^T$ is given by

$$\operatorname{grad}_x \Phi(x_1^*, x_2^*, u) = \begin{pmatrix} 2(x_1^* + 3) + 2x_1^* u \\ 2x_2 - u \end{pmatrix} = 0\,.$$

This relationship implies $x_1^* = -3/(u+1)$ and $x_2^* = u/2$, thus

$$\begin{aligned}\inf_{x\in\mathbb{R}^2} \Phi(x_1, x_2, u) &= \Phi(x_1^*, x_2^*, u)\\ &= \left(\frac{-3}{u+1} + 3\right)^2 + \frac{u^2}{4} + u\left(\frac{9}{(u+1)^2} - \frac{u}{2}\right) = \frac{9u}{u+1} - \frac{u^2}{4}\,.\end{aligned}$$

Hence, the dual problem is (see Figure 4.9)

$$\max_{u\geq 0} \frac{9u}{u+1} - \frac{u^2}{4}\,. \tag{D}$$

The first and second derivatives of the objective function $d(u) = \frac{9u}{u+1} - \frac{u^2}{4}$ are

$$\begin{aligned} d'(u) &= \frac{9}{(u+1)^2} - \frac{u}{2}\,,\\ d''(u) &= -\frac{18}{(u+1)^3} - \frac{1}{2}\,. \end{aligned}$$

It holds $d''(u) < 0$ for $u \geq 0$, i.e. the function $d(u)$ is convex for nonnegative values of u. Since $d'(2) = 0$, the unique optimal point of (D) is $u^* = 2$ with objective value $d(u^*) = 5$.

We will now see that Definition 4.25 generalizes the definition of duality in linear programming.

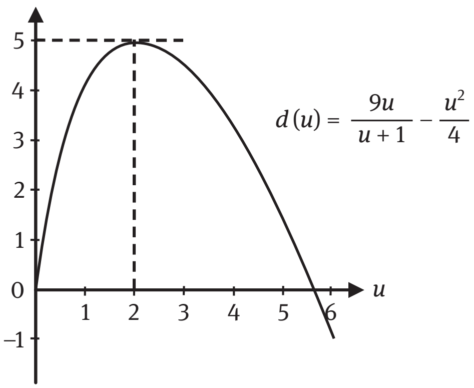


Fig. 4.9. Geometric illustration of (D).

Example 4.27. We consider the linear problem in canonical form

$$\begin{aligned} \max c^T x & \\ Ax &\le b \\ x &\ge 0 \end{aligned} \tag{P}$$

which can be written as (see (4.39)):

$$\begin{aligned} \min -c^T x & \\ Ax - b &\le 0 \\ -x &\le 0 . \end{aligned}$$

The Lagrange function is

$$\Phi(x, u, v) = -c^T x + u^T (Ax - b) - v^T x = -u^T b + (u^T A - c^T - v^T) x ,$$

where u and v are the Lagrange multipliers corresponding to the constraints $Ax - b \le 0$ and $-x \le 0$, respectively. Therefore,

$$\inf_{x \in \mathbb{R}^n} \Phi(x, u, v) = \begin{cases} -u^T b & \text{for } u^T A - c^T - v^T = 0 , \\ -\infty & \text{otherwise.} \end{cases}$$

Thus the dual problem is (see Definition 4.25)

$$\max_{u \in \mathbb{R}^n_+} \begin{cases} -u^T b & \text{for } u^T A - c^T - v^T = 0 , \\ -\infty & \text{otherwise,} \end{cases}$$

which can be written as

$$\begin{aligned} \max -\, u^T b & \\ u^T A - c^T - v^T &= 0 \\ u, v &\ge 0 . \end{aligned}$$

Finally, we obtain the canonical form of the dual problem in linear programming:

$$\begin{aligned} \min b^T u & \\ A^T u &\ge c \\ u &\ge 0 . \end{aligned} \tag{D}$$

As an immediate consequence of Definition 4.25 we obtain:

Theorem 4.28 (Weak Duality Theorem). *Let $\overline{x}, \overline{u}$ be feasible points of (P) and (C), respectively. The objective function values $f(\overline{x})$ and $d(\overline{u}) = \inf_{x \in \mathbb{R}^n} \Phi(x, \overline{u})$ satisfy*

$$d(\overline{u}) \leq f(\overline{x}) .$$

Proof. The feasible points $\overline{x}, \overline{u}$ satisfy $g(\overline{x}) \leq 0, \overline{u} \geq 0$. Therefore

$$d(\overline{u}) = \inf_{x \in \mathbb{R}^n} \Phi(x, \overline{u}) = \inf_{x \in \mathbb{R}^n} [f(x) + \overline{u}^T g(x)] \leq f(\overline{x}) + \overline{u}^T g(\overline{x}) \leq f(\overline{x}) . \qquad \square$$

Let us consider again Example 4.26. Here it holds for any feasible points $\overline{x}$, $\overline{u}$ of (P) and (D), respectively:

$$f(\overline{x}) \geq 5 \quad \text{and} \quad d(\overline{u}) \leq 5 , \quad \text{i.e. } d(\overline{u}) \leq f(\overline{x}) .$$

The theorem has important consequences for solution procedures: Let $\overline{x}, \overline{u}$ be defined as above and let z_{ot} and w_{ot} denote the optimal value of (P) and (D), respectively. Then the theorem implies that

$$d(\overline{u}) \leq f(x^*) = z_{\text{ot}} ,$$

for an optimal solution x^* de (P). Hence

$$d(\overline{u}) \leq z_{\text{ot}} \leq f(\overline{x}) . \tag{4.44}$$

In the same way we get

$$d(\overline{u}) \leq w_{\text{ot}} \leq f(\overline{x}) . \tag{4.45}$$

An algorithm that produces iteratively feasible solutions of (P) and (D), can terminate when $f(\overline{x}) - d(\overline{u})$ is sufficiently small. This value is an upper limit for the absolute error in the objective function value. In particular, if $f(\overline{x}) = d(\overline{u})$, then $\overline{x}$ and $\overline{u}$ are optimal points for (P) and (D), respectively.

We present without proof a result which has many applications in game theory:

Theorem 4.29. *The following statements are equivalent:*
(i) *The point $(x^*, u^*)^T$ is a saddle point of the Lagrange function $\Phi(x, u)$*
(ii) *x^*, u^* are optimal solutions of (P) and (D), respectively, and*

$$\Phi(x^*, u^*) = z_{\text{ot}} = w_{\text{ot}} .$$

In Example 4.26 it holds $\Phi(x_1^*, x_2^*, u^*) = z_{\text{ot}} = w_{\text{ot}} = 5$ for the optimal solutions $(x_1^*, x_2^*) = (-1, 1)$ and $u^* = 2$. Since $\Phi(x_1^*, x_2^*, u) = \Phi(x_1^*, x_2^*, u^*)$ for all $u \geq 0$, relation (4.40) is equivalent to

$$5 \leq (x_1 + 3)^2 + x_2^2 + 2(x_1^2 - x_2)$$

for all $(x_1, x_2)^T \in \mathbb{R}^2$. From the calculations of Example 4.26 it follows that this relation is satisfied, i.e. $(x_1^*, x_2^*, u^*)^T$ is a saddle point of the Lagrange function.

Combining Theorems 4.20, 4.22 and 4.29, we obtain:

Theorem 4.30 (Duality Theorem). *Let the functions f and g_i of the primal problem (P) be continuously differentiable and convex and let x^* be a regular optimal point of (P). Then it holds*

(i) *The dual problem (D) has an optimal solution.*

(ii) *The optimal values of (P) and (D) are identical.*

(iii) *Let u^* be an optimal solution of (D). Any point $\overline{x}$, satisfying*

$$\overline{x} = \inf_{x \in \mathbb{R}^n} \Phi(x, u^*)\,, \quad g(\overline{x}) \le 0 \quad \text{and} \quad u^{*T} g(\overline{x}) = 0$$

is an optimal solution of (P).

Using part (iii) of the theorem we can determine an optimal solution of (P) from an optimal solution of (D). The procedure is straightforward when $\Phi(x, u^*)$ is strictly convex in x. Then the problem

$$\min_{x \in \mathbb{R}^n} \Phi(x, u^*)$$

has a unique solution that must be equal to x^*. Since x^* is optimal for (P), the conditions $g(x^*) \le 0$ and $u^{*T} g(x^*) = 0$ are satisfied. We can then determine x^* by setting $\text{grad}_x\, \Phi(x, u^*)$ equal to zero.

We illustrate the procedure in the following example:

Example 4.31. We consider the quadratic problem

$$\begin{aligned} \min f(x) &= \frac{1}{2} x^T Q x + c^T x \\ A^T x &\le b\,, \end{aligned} \tag{P}$$

where $Q \in \mathbb{R}^{n \times n}$ is symmetric and positive definite, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^m$. One can prove that (P) has an optimal solution if a feasible solution exists.

The dual problem is

$$\max_{u \in \mathbb{R}^m_+} \inf_{x \in \mathbb{R}^n} \Phi(x, u) \tag{D}$$

with $\Phi(x, u) = \frac{1}{2} x^T Q x + c^T x + u^T (A^T x - b)$.

For any fixed $u \in \mathbb{R}^m_+$, the Lagrange function $\Phi(x, u)$ is strictly convex, since the Hessian matrix Q is positive definite. The unique global minimum point satisfies

$$\begin{aligned} \text{grad}_x\, \Phi(x, u) = Qx + c + Au = 0 \Rightarrow \\ x = -Q^{-1}(c + Au)\,. \end{aligned} \tag{4.46}$$

Inserting the term (4.46) in the Lagrange function we get

$$\begin{aligned} \inf_{x \in \mathbb{R}^n} \Phi(x, u) = {} & \frac{1}{2} [Q^{-1}(c + Au)]^T Q Q^{-1}(c + Au) - c^T Q^{-1}(c + Au) \\ & - u^T [A^T Q^{-1}(c + Au) + b]\,, \end{aligned} \tag{4.47}$$

and after some algebraic transformations:

$$\inf_{x\in\mathbb{R}^n} \Phi(x,u) = -\frac{1}{2}c^T Q^{-1} c - \frac{1}{2}u^T P u - h^T u\,, \tag{4.48}$$

where $P = A^T Q^{-1} A$, $h = A^T Q^{-1} c + b$.

Eliminating the additive constant $-\frac{1}{2}c^T Q^{-1} c$, which does not influence the optimal solution, we can write the dual problem as

$$\max_{u\in\mathbb{R}^m_+} -\frac{1}{2}u^T P u - h^T u\,.$$

Even for a large number of constraints, this problem can be solved more easily than the primal problem (P), because now the constraints are much simpler.

If an optimal solution u of (D) is determined, we obtain an optimal solution of (P), using (4.46).

Exercise 4.32. *Perform the transformation of (4.47) in (4.48) step by step.*

Exercise 4.33. *Given the quadratic problem*

$$\begin{aligned} &\min (x_1-2)^2 + x_2^2 \\ &2x_1 + x_2 - 2 \le 0\,. \end{aligned} \tag{P}$$

(a) *Determine the optimal solution of (P) using the dual (D).*
(b) *Solve (P) geometrically.*

Exercise 4.34. *Given the problem*

$$\begin{aligned} &\min -3x_1 - x_2 \\ &x_1^2 + 2x_2^2 - 2 \le 0\,. \end{aligned} \tag{P}$$

(a) *Determine the dual problem.*
(b) *Determine an optimal solution of (D).*
(c) *Solve (P), using Theorem 4.30 (iii).*
(d) *Solve (P) geometrically.*

Finally we apply the theory of this section to solve a problem of investment planning (see Example 1.5). We make use of the fact that the duality theorem remains valid when the primal problem is

$$\min_{g(x)\le 0,\, x\in K} f(x) \Leftrightarrow \min_{x\in K} \sup_{u\in\mathbb{R}^m_+} \Phi(x,u) \tag{P}$$

with $K \subset \mathbb{R}^n$. The dual problem for this generalized form of (4.39) is

$$\max_{u\in\mathbb{R}^m_+} \inf_{x\in K} \Phi(x,u)\,. \tag{D}$$

Example 4.35. We consider the problem

$$
\begin{aligned}
&\max \sum_{j=1}^{n} f_j(x_j) \\
&\sum_{j=1}^{n} x_j \le a \\
&x_j \ge 0\,, \quad j = 1, \dots, n
\end{aligned}
\tag{4.49}
$$

of Example 1.5 with concave functions $f_j(x_j)$ which can be written as

$$
\begin{aligned}
&\min \sum_{j=1}^{n} -f_j(x_j) \\
&\sum_{j=1}^{n} x_j \le a \\
&x \in K\,,
\end{aligned}
\tag{4.50}
$$

where $K = \{x = (x_1, \dots, x_n)^T \in \mathbb{R}^n | 0 \le x_i \le a \text{ for } i = 1, \dots, n\}$. The dual problem is then

$$\max_{u \in \mathbb{R}_+} \inf_{x \in K} \Phi(x, u)$$

with

$$
\begin{aligned}
\Phi(x, u) &= \sum_{j=1}^{n} -f_j(x_j) + u \left(\sum_{j=1}^{n} x_j - a \right) \\
&= \sum_{j=1}^{n} \left[-f_j(x_j) + u x_j \right] - ua\,.
\end{aligned}
$$

We obtain the dual objective function

$$d(u) = \inf_{x \in K} \Phi(x, u) = \sum_{j=1}^{n} \min_{0 \le x_i \le a} [-f_j(x_j) + u x_j] - ua$$

(note that the minima exist).

Since the f_j are concave, the functions

$$h_j(x_j) = -f_j(x_j) + u x_j$$

of the last sum are convex and the minimum point x_j^* of $h_j(x_j)$ is an extreme point of the interval $[0, a]$ or satisfies $h_j'(x_j^*) = 0$.

To solve the dual problem, the one-dimensional function $d(u)$ must be maximized. As illustrated in the following, $d(u)$ is always differentiable and concave. Hence we can solve (4.49) with the aid of the duality theorem, using a method for optimizing a one-dimensional function. (Such methods will be studied in Chapter 6). Consider the following numerical example:

$$\max 3(1 - e^{-2x_1}) + \frac{5x_2}{x_2 + 3}$$
$$x_1 + x_2 \le 2$$
$$x_1, x_2 \ge 0\,,$$

i.e. $f_1(x_1) = 3(1 - e^{-2x_1}), f_2(x_2) = \frac{5x_2}{x_2+3}, a = 2$. We have to determine the minimum of the functions $h_1(x_1) = 3(e^{-2x_1} - 1) + ux_1$ and $h_2(x_2) = -\frac{5x_2}{x_2+3} + ux_2$ over the interval $[0, 2]$.

We get $h_1'(x_1) = u - 6\,e^{-2x_1}$. Thus, $h_1'(x_1) = 0 \Leftrightarrow x_1 = -\frac{1}{2}\ln\frac{u}{6}$. However, the value $-\frac{1}{2}\ln\frac{u}{6}$ lies in the interval $[0, 2]$, if and only if $u \in [6\,e^{-4}, 6]$. For $u < 6\,e^{-4} \approx 0.11$, the function $h_1'(x_1)$ is negative, i.e. $h_1(x_1)$ is decreasing over $[0, 2]$. Similarly, $h_1(x_1)$ is increasing over $[0, 2]$ for $u > 6$ (see Figure 4.10). Summarizing these considerations, we obtain

$$\min_{0\le x_1\le 2} h_1(x_1) = \begin{cases} h_1(2) = 3\,e^{-4} - 3 + 2u & \text{for } u < 6\,e^{-4} \\ h_1(-\frac{1}{2}\ln\frac{u}{6}) = \frac{u}{2}(1 - \ln\frac{u}{6}) - 3 & \text{for } 6\,e^{-4} \le u \le 6 \\ h_1(0) = 0 & \text{for } u > 6\,. \end{cases}$$

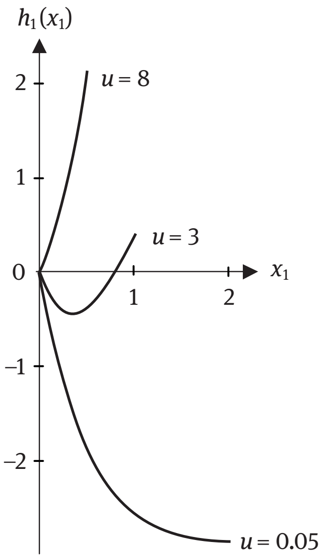


Fig. 4.10. Function $h_1(x_1)$ for various values of u.

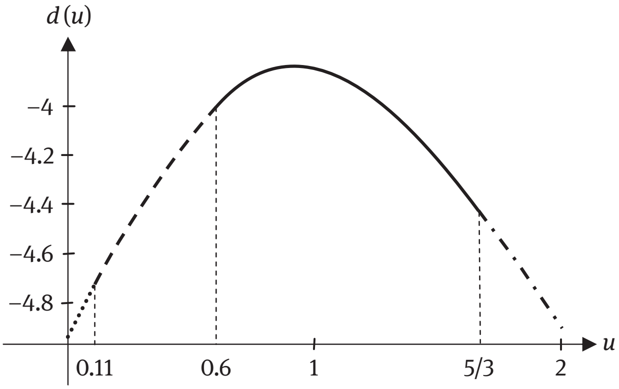


Fig. 4.11. Dual objective function.

Similarly we get

$$\min_{0\le x_2\le 2} h_2(x_2) = \begin{cases} h_2(2) = 2(u-1) & \text{for } u < 3/5 \\ h_2(\sqrt{15/u}-3) = \sqrt{60u} - 5 - 3u & \text{for } 3/5 \le u \le 5/3 \\ h_2(0) = 0 & \text{for } u > 5/3\,. \end{cases}$$

Thus the dual objective function is (Figure 4.11)

$$\begin{aligned} d(u) &= \min_{0\le x_1\le 2} h_1(x_1) + \min_{0\le x_2\le 2} h_2(x_2) - 2u \\ &= \begin{cases} 3\,\mathrm{e}^{-4} - 5 + 2u & \text{for } u < 6\,\mathrm{e}^{-4} \approx 0.11 \\ \frac{u}{2}(1 - \ln\frac{u}{6}) - 5 & \text{for } 6\,\mathrm{e}^{-4} \le u \le 3/5 \\ \frac{u}{2}(1 - \ln\frac{u}{6}) + \sqrt{60u} - 8 - 5u & \text{for } 3/5 \le u \le 5/3 \\ \frac{u}{2}(1 - \ln\frac{u}{6}) - 3 - 2u & \text{for } 5/3 \le u \le 6 \\ -2u & \text{for } u > 6\,. \end{cases} \end{aligned}$$

We obtain

$$d'(u) = \begin{cases} 2 & \text{for } u < 6\,\mathrm{e}^{-4} \approx 0.11 \\ -\frac{1}{2}\ln\frac{u}{6} & \text{for } 6\,\mathrm{e}^{-4} \le u \le 3/5 \\ -\frac{1}{2}\ln\frac{u}{6} + \sqrt{15/u} - 5 & \text{for } 3/5 \le u \le 5/3 \\ -\frac{1}{2}\ln\frac{u}{6} - 2 & \text{for } 5/3 \le u \le 6 \\ -2 & \text{for } u > 6. \end{cases}$$

The condition $d'(u) = 0$ can be satisfied only in the interval $[3/5, 5/3]$.

With Newton's method to be studied in Section 6.2, we obtain the solution $u^* \approx 0.9111$.

Since $\Phi(x,u) = 3(\mathrm{e}^{-2x_1} - 1) - \frac{5x_2}{x_2+3} + u(x_1 + x_2 - 2)$, the solution x^* of the primal problem is given by

$$\operatorname{grad}_x \Phi(x,u^*) = \begin{pmatrix} -6\,\mathrm{e}^{-2x_1} + u^* \\ -15/(x_2+3)^2 + u^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

implying $x_1^* \approx 0.9424$, $x_2^* \approx 1.058$.

Exercise 4.36. *A problem of type (4.49) with two variables can be solved directly, by using the fact that an optimal solution $(x_1^*, x_2^*)^T$ satisfies $x_1^* + x_2^* = a$ (which is intuitively clear in Example 4.35). Thus the variable x_2 may be replaced by $a - x_1$. Check the solution obtained above with the aid of this one-dimensional problem!*

4.3 The Wolfe dual problem

Beyond the (Lagrange) dual problem (Definition 4.25) there exist various other possibilities to define a dual problem. We will discuss briefly a concept credited to Wolfe (see Section 4.5). Recall that the primal problem (4.39) can be written as

$$\min_{x\in\mathbb{R}^n} \sup_{u\in\mathbb{R}^m_+} \Phi(x,u) \tag{P}$$

with $\Phi(x,u) = f(x) + u^T g(x)$ and that the (Lagrange) dual problem is

$$\max_{u\in\mathbb{R}^m_+} \inf_{x\in\mathbb{R}^n} \Phi(x,u). \tag{D}$$

Assume that all functions $f, g_1, \dots, g_m$ are continuously differentiable and convex, i.e. the "interior optimization problem" of (D) is *convex* for all $u \in \mathbb{R}^m_+$. We can make use of the fact that the infimum of $\Phi(x,u)$ is attained if and only if the gradient with respect to x is zero. This idea leads to the so-called *Wolfe dual problem*:

$$\begin{aligned} &\max \Phi(x,u) \\ &\operatorname{grad}_x \Phi(x,u) = 0 \\ &u \geq 0\,. \end{aligned} \tag{DW}$$

Generally the equality constraints are not linear. In this case (DW) is not convex and the dual of the dual does not exist.

Example 4.37. For the primal problem of Example 4.26 we get

$$\begin{aligned} \Phi(x_1,x_2,u) &= (x_1+3)^2 + x_2^2 + u(x_1^2 - x_2)\,, \\ \operatorname{grad}_x \Phi(x_1,x_2,u) &= \begin{pmatrix} 2(x_1+3) + 2x_1 u \\ 2x_2 - u \end{pmatrix}. \end{aligned}$$

The corresponding Wolfe dual is not convex:

$$\max(x_1+3)^2 + x_2^2 + u(x_1^2 - x_2)$$

$$\begin{aligned} x_1 + 3 + x_1 u &= 0 \\ 2x_2 - u &= 0 \\ u &\geq 0. \end{aligned} \tag{4.51}$$

By means of variable elimination (see Section 8.2) we can transform (4.51) in a single variable problem. The second constraint yields

$$u = 2x_2\,. \tag{4.52}$$

By substituting u in the first constraint we get

$$x_1 + 3 + 2x_1x_2 = 0\,,$$

hence

$$x_2 = -\frac{x_1+3}{2x_1}\,. \tag{4.53}$$

By eliminating u and x_2 in the objective function of (4.51) we obtain

$$d(x_1) = (x_1+3)^2 + \frac{(x_1+3)^2}{4x_1^2} - \frac{x_1+3}{x_1}\left(x_1^2 + \frac{x_1+3}{2x_1}\right) = (x_1+3)\frac{12x_1^2 - x_1 - 3}{4x_1^2}\,.$$

Since $u \geq 0$, the conditions (4.52) and (4.53) imply $-3 \leq x_1 < 0$. and (4.51) can be expressed in the equivalent form

$$\max_{-3\leq x_1<0} d(x_1)\,. \tag{4.54}$$

Since $d'(x_1) = [3(x_1+1)(x_1^2 - x_1 + 3/2)]/x_1^3$, the unique solution of (4.54) is $x_1^* = -1$. Hence, (4.51) has the solution $(x_1^*, x_2^*, u^*) = (-1, 1, 2)$ with optimal value $w_{\text{ot}} = z_{\text{ot}} = 5$.

For the pair of problems (P) and (DW) the weak duality theorem is valid:

Theorem 4.38. *Assume that the functions $f, g_1, \ldots, g_m$ of (P) are continuously differentiable and convex, and let $\hat{x}$ and $(\overline{x}, \overline{u})$ be feasible solutions of (P) and (DW), respectively. Then*

$$\Phi(\overline{x}, \overline{u}) \leq f(\hat{x})\,.$$

Proof. The feasible points satisfy the conditions $\operatorname{grad}_x \Phi(\overline{x}, \overline{u}) = 0, \overline{u} \geq 0, g(\hat{x}) \leq 0$. Due to the convexity this implies that $\overline{x}$ minimizes the function $\Phi(x, \overline{u})$ for all $\overline{u} \geq 0$. We also get $\overline{u}^T g(\hat{x}) \leq 0$, thus

$$\Phi(\overline{x}, \overline{u}) \leq \Phi(\hat{x}, \overline{u}) = f(\hat{x}) + \overline{u}^T g(\hat{x}) \leq f(\hat{x})\,. \qquad \square$$

We easily verify the weak duality for the previous example, where $\Phi(\overline{x}, \overline{u}) \leq 5 \leq f(\hat{x})$ holds.

We now determine the Wolfe dual for a linear and a quadratic problem.

Example 4.39. For the linear problem

$$\begin{aligned} \max c^T x& \\ Ax \leq& 0 \\ x \geq& 0\,, \end{aligned} \tag{P}$$

we get (see Example 4.27)

$$\Phi(x, u, v) = -u^T b + (u^T A - c^T - v^T)x\,,$$

thus the Wolfe dual is

$$\begin{aligned} \max -u^T b + (u^T A - c^T - v^T)x& \\ u^T A - c^T - v^T = 0& \\ u, v \geq 0&\,, \end{aligned}$$

which is equivalent to

$$\begin{aligned} \min b^T u& \\ A^T u \geq c& \\ u \geq 0&\,. \end{aligned} \tag{DW}$$

Thus for the linear problem (P) above, the Wolfe dual is identical with the Lagrange dual.

Example 4.40. Consider the quadratic problem with nonnegativity constraints

$$\begin{aligned} \min f(x) = \frac{1}{2} x^T Q x + c^T x& \\ A^T x \leq b& \\ x \geq 0&\,, \end{aligned} \tag{P}$$

where $Q \in \mathbb{R}^{n \times n}$ is symmetric and positive semidefinite, $c \in \mathbb{R}^n, A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^m$. Since

$$\Phi(x, u, v) = \frac{1}{2} x^T Q x + c^T x + u^T (A^T x - b) - v^T x\,,$$

the Wolfe dual is

$$\begin{aligned} \max \frac{1}{2} x^T Q x + c^T + u^T (A^T x - b) - v^T x& \\ Qx + c + Au - v = 0& \\ u, v \geq 0&\,. \end{aligned}$$

By substituting the vector c in the objective function with the aid of the first constraint we get:

$$\begin{aligned}\Phi(x,u,v) &= \frac{1}{2}x^TQx + (v - Qx - Au)^Tx + u^T(A^Tx - b) - v^Tx \\ &= \frac{1}{2}x^TQx + v^Tx - x^TQx - u^TA^Tx + u^TA^Tx - u^Tb - v^Tx \\ &= -\frac{1}{2}x^TQx - u^Tb\,.\end{aligned}$$

Hence, the Wolfe dual can be written as

$$\begin{aligned}\max -\frac{1}{2}x^TQx - u^Tb& \\ Qx + c + Au \geq 0& \\ u \geq 0&\end{aligned} \tag{DW}$$

or

$$\begin{aligned}\min \frac{1}{2}x^TQx + u^Tb& \\ Qx + c + Au \geq 0& \\ u \geq 0\,.&\end{aligned}$$

We call attention to the fact that for the existence of the Wolfe dual in Example 4.40 it is only required that Q is positive semidefinite (see the definition at the beginning of this section and Example 4.31).

Exercise 4.41. *Determine the Wolfe dual* (DW) *for*

$$\begin{aligned}\min x_1^2 + x_2^2 - 6x_2& \\ x_1 \leq 2& \\ x_2 \leq 1& \\ x_1, x_2 \geq 0\,.&\end{aligned} \tag{P}$$

Solve the problems (P) *and* (DW)*. Determine the optimal values z_{ot} and w_{ot} of (P) and* (DW)*, respectively, and verify the weak duality theorem (Theorem 4.38). Show that there exists a so-called duality gap, i.e. $w_{ot} < z_{ot}$.*

Exercise 4.42. *Is it possible to formulate a "Duality Theorem" for the Wolfe dual analogous to Theorem 4.30?*

Exercise 4.43. *Determine the Wolfe dual for the quadratic problem* (P) *of Example 4.31. Show that the Wolfe dual is the standard form of a linear problem for* $Q = 0$.

4.4 Second-order optimality criteria

In Theorem 4.12 we already considered first-order necessary optimality conditions for the problem

$$\begin{aligned} &\min f(x) \\ g_i(x) &\le 0 \quad \text{for } i = 1, \dots, m \\ h_j(x) &= 0 \quad \text{for } j = 1, \dots, k\,. \end{aligned} \tag{4.55}$$

Without proof we now present some conditions of second order. In order to express them in a condensed form we employ the *generalized Lagrange function* of (4.55)

$$\Phi(x, u, v) = f(x) + u^T g(x) + v^T h(x)\,,$$

where $x \in \mathbb{R}^n, u \in \mathbb{R}^m_+, v \in \mathbb{R}^k, g(x) = (g_1(x), \dots, g_m(x))^T, h(x) = (h_1(x), \dots, h_k(x))^T$ (see Definition 4.19). Deriving this function two times with respect to x we get

$$\operatorname{grad}_x \Phi(x, u, v) = \operatorname{grad} f(x) + \sum_{i=1}^m u_i \operatorname{grad} g_i(x) + \sum_{j=1}^k v_j \operatorname{grad} h_j(x), \tag{4.56}$$

$$H_x \Phi(x, u, v) = Hf(x) + \sum_{i=1}^m u_i H g_i(x) + \sum_{j=1}^k v_j H h_j(x)\,. \tag{4.57}$$

By using (4.56) we can rewrite the first-order necessary condition (4.28) as

$$\operatorname{grad}_x \Phi(x^*, u^*, v^*) = 0\,. \tag{4.58}$$

Recall that in Section 4.1 we defined the linear independence constraint qualification for problems with inequality constraints. We now generalize this concept to the problem with mixed constraints (4.55):

Definition 4.44. Let x^* be a feasible point of (4.55). We say that the restrictions satisfy the *linear independence constraint qualification* (LICQ) at point x^*, if the vectors $\operatorname{grad} g_i(x^*)$ and $\operatorname{grad} h_j(x^*)$ are linearly independent for all $i \in A(x^*)$ (see Definition 3.2) and all $j = 1, \dots, k$.

Recall that condition LICQ is sufficient for the regularity of point x^* (Theorem 4.2).

Theorem 4.45 (Necessary optimality conditions). *Let the functions f, g_i and h_j of (4.55) be twice continuously differentiable and let x^* be a local minimum point at which the conditions* LICQ *are satisfied. Then there exist vectors $u^* \in \mathbb{R}^m$ and $v^* \in \mathbb{R}^k$*

such that

$$\operatorname{grad}_x \Phi(x^*, u^*, v^*) = 0$$

$$\begin{aligned} u_i^* g_i(x^*) &= 0 \quad \textit{for } i = 1, \ldots, m \\ g_i(x^*) &\le 0 \quad \textit{for } i = 1, \ldots, m \\ h_j(x^*) &= 0 \quad \textit{for } j = 1, \ldots, k \\ u_i^* &\ge 0 \quad \textit{for } i = 1, \ldots, m\,, \end{aligned} \tag{4.59}$$

and

$$d^T H_x \Phi(x^*, u^*, v^*) d \ge 0 \quad \textit{for all } d \in C(x^*)\,, \tag{4.60}$$

where

$$\begin{aligned} C(x^*) = \{d \in \mathbb{R}^n | d^T \operatorname{grad} g_i(x^*) &\le 0 \quad \textit{for } i \in A(x^*) \textit{ and } u_i^* = 0, \\ d^T \operatorname{grad} g_i(x^*) &= 0 \quad \textit{for } i \in A(x^*) \textit{ and } u_i^* > 0, \\ d^T \operatorname{grad} h_j(x^*) &= 0 \quad \textit{for } j = 1, \ldots, k\}\,. \end{aligned}$$

Theorem 4.46 (Sufficient optimality conditions). *If the functions f, g_i and h_j of (4.55) are twice continuously differentiable and if there exist vectors $u^* \in \mathbb{R}^m$ and $v^* \in \mathbb{R}^k$ such that (4.59) and*

$$d^T H_x \Phi(x^*, u^*, v^*) d > 0 \quad \textit{for all } d \in C(x^*) \setminus \{0\} \tag{4.61}$$

are satisfied, then x^* is a strict local minimum point of (4.55).

Note that conditions (4.59) correspond to the first-order necessary conditions of Section 4.1; (4.60) and (4.61) are called *necessary and sufficient second-order optimality conditions* and state that the Hessian matrix $H_x\Phi(x^*, u^*, v^*)$ must be positive semidefinite or positive definite, respectively, “over the set C(x*)”. This set is called the *cone of tangent directions*.

In the following we illustrate the above theorems by means of diverse examples. At first we consider the specific cases of (4.55) with $k = 0$ and $m = 0$ (Examples 4.47 and 4.50). A case with mixed constraints is studied in Example 4.51.

Example 4.47. Given the problem

$$\min -c_1 x_1 - x_2$$

$$x_1^2 + x_2 - 2 \le 0 \tag{1}$$

$$x_1 - x_2 \le 0 \tag{2}$$

we obtain grad $f(x) = (-c_1, -1)^T$, grad $g_1(x) = (2x_1, 1)^T$, grad $g_2(x) = (1, -1)^T$, and the first-order necessary conditions are:

$$\begin{aligned} -c_1 + 2u_1x_1 + u_2 &= 0 \\ -1 + u_1 - u_2 &= 0 \\ u_1(x_1^2 + x_2 - 2) &= 0 \\ u_2(x_1 - x_2) &= 0 \\ x_1^2 + x_2 - 2 &\leq 0 \\ x_1 - x_2 &\leq 0 \\ u_1, u_2 &\geq 0\,. \end{aligned} \tag{4.62}$$

Case (a)
Let $c_1 = 3$:
Since the objective function is linear, we easily verify by means of Figure 4.12 that $(x_1^*, x_2^*) = (1, 1)$ is the unique optimal solution, and because $-$ grad $f(x^*)$ is in the interior of the cone generated by $g^1 :=$ grad $g_1(x^*)$ and $g^2 :=$ grad $g_2(x^*)$, both La-

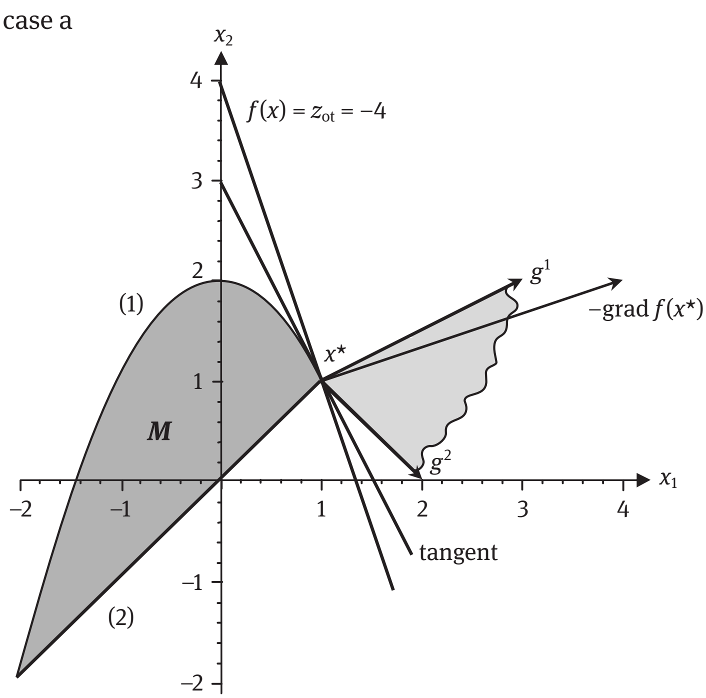


Fig. 4.12. Geometrical solution of Example 4.47 (case a).

grange multipliers must be positive. Since $x_1^* = 1$, the first two equations of (4.62) yield $u_1^* = 4/3$ and $u_2^* = 1/3$. In order to verify the second-order conditions we observe that

$$\Phi(x, u) = -c_1 x_1 - x_2 + u_1(x_1^2 + x_2 - 2) + u_2(x_1 - x_2)\,,$$

thus

$$H_x\Phi(x^*, u^*) = \begin{pmatrix} 2u_1^* & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 8/3 & 0 \\ 0 & 0 \end{pmatrix}.$$

The cone of tangent directions consists of a unique point:

$$C(x^*) = \left\{ d \in \mathbb{R}^2 \,\middle|\, (d_1, d_2)\begin{pmatrix} 2 \\ 1 \end{pmatrix} = 0,\ (d_1, d_2)\begin{pmatrix} 1 \\ -1 \end{pmatrix} = 0 \right\} = \{0\}\,,$$

and (4.61) is trivially satisfied.

Case (b)
Let $c_1 = 2$:
As in the previous case, the unique solution is $(x_1^*, x_2^*) = (1, 1)$ (see Figure 4.13), but now it holds $-\operatorname{grad} f(x^*) = g^1 = (2, 1)^T$, i.e. $-\operatorname{grad} f(x^*)$ and g^1 are collinear and therefore $u_2^* = 0$. From the second equation of (4.62) we obtain $u_1^* = 1$. The cone of tangent directions is now

$$C(x^*) = \left\{ d \in \mathbb{R}^2 \,\middle|\, (d_1, d_2)\begin{pmatrix} 2 \\ 1 \end{pmatrix} = 0,\ (d_1, d_2)\begin{pmatrix} 1 \\ -1 \end{pmatrix} \le 0 \right\} = \left\{ \alpha \begin{pmatrix} -1 \\ 2 \end{pmatrix} \,\middle|\, \alpha \ge 0 \right\}. \tag{4.63}$$

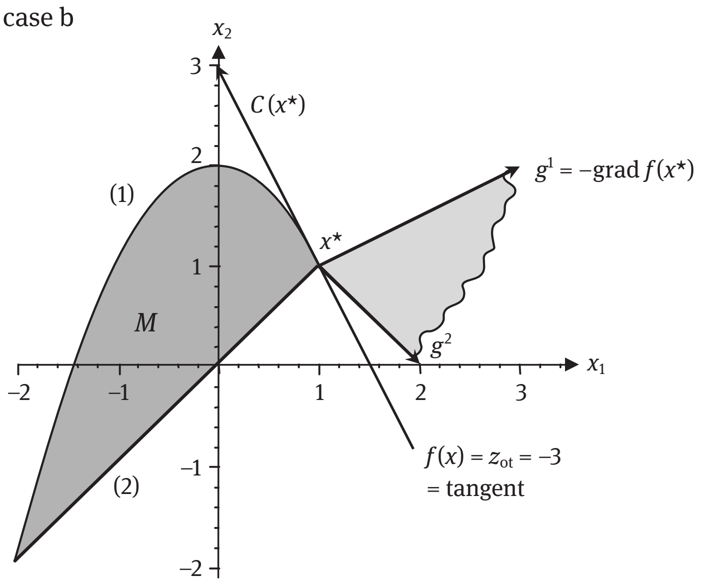


Fig. 4.13. Geometrical solution of Example 4.47 (case b).

This ray (displaced to the point x^*) coincides with the tangent to the curve $g_1(x) = 0$ at point x^*.

The second-order sufficient conditions are satisfied, since

$$d^T H_x \Phi(x^*, u^*) d = d^T \begin{pmatrix} 2u_1^* & 0 \\ 0 & 0 \end{pmatrix} d = 2u_1^* d_1^2 = 2d_1^2 > 0$$

holds for all $d \in C(x^*) \setminus \{0\}$. Note that the Hessian matrix

$$H_x \Phi(x^*, u^*) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

is not positive definite, it is only "positive definite over the cone (4.63)".

We suggest that the reader solves the system (4.62) for the two preceding cases also by means of the method of Example 4.8, i.e. without using geometrical arguments.

Exercise 4.48. *Solve the problem of Example 4.47 for the case $c_1 = -1$. Verify the second-order sufficient conditions.*

Exercise 4.49. *Solve the problem*

$$\begin{aligned} \min c_1 x_1 + c_2 x_2 + c_3 x_3 \\ \frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \frac{x_3^2}{a_3^2} \le 1 \end{aligned}$$

($c = (c_1, c_2, c_3)^T \neq 0$ and $a_1, a_2, a_3 > 0$) with the help of the preceding theorems. Interpret the solution x^* and the cone $C(x^*)$ (shifted to the point x^*) geometrically. Make use of the concept of a supporting hyperplane (Definition 3.31).

Example 4.50. We determine the distance between the parables A and B of Figure 4.14. We have to solve the problem

$$\begin{aligned} \min d(P_1, P_2) \\ P_1 \in A\,, \quad P_2 \in B\,, \end{aligned}$$

equivalent to

$$\begin{aligned} \min d^2(P_1, P_2) \\ P_1 \in A\,, \quad P_2 \in B\,, \end{aligned}$$

where $d(P_1, P_2)$ represents the Euclidean distance between the points P_1 and P_2. Setting $P_1 = (x_1, x_2)^T$ and $P_2 = (x_3, x_4)^T$, we obtain the equivalent problem in four variables

$$\begin{aligned} \min (x_1 - x_3)^2 + (x_2 - x_4)^2 \\ (x_1 + 2)^2 + 1 + x_2 = 0 \\ 2(x_3 - 1)^2 - x_4 = 0\,. \end{aligned}$$

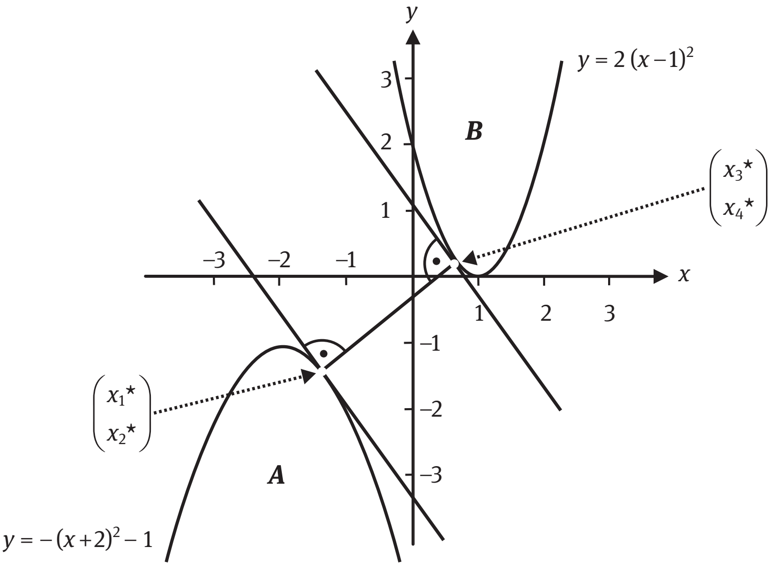


Fig. 4.14. Illustration of Example 4.50.

It holds that

$$\begin{aligned} \operatorname{grad} f(x) &= 2(x_1 - x_3, x_2 - x_4, x_3 - x_1, x_4 - x_2)^T, \\ \operatorname{grad} h_1(x) &= (2(x_1 + 2), 1, 0, 0)^T, \\ \operatorname{grad} h_2(x) &= (0, 0, 4(x_3 - 1), -1)^T \end{aligned}$$

and the first-order necessary conditions are

$$x_1 - x_3 + v_1(x_1 + 2) = 0 \tag{1}$$
$$2(x_2 - x_4) + v_1 = 0 \tag{2}$$
$$x_3 - x_1 + 2v_2(x_3 - 1) = 0 \tag{3}$$
$$2(x_4 - x_2) - v_2 = 0 \tag{4}$$
$$(x_1 + 2)^2 + 1 + x_2 = 0 \tag{5}$$
$$2(x_3 - 1)^2 - x_4 = 0\,. \tag{6}$$

In order to solve this system we first observe that the variables x_2, x_4, v_1, v_2 can be expressed in terms of x_1 and x_3 (see (1), (3), (5), (6)). In particular,

$$v_1 = \frac{x_3 - x_1}{x_1 + 2}, \tag{7}$$
$$v_2 = \frac{x_1 - x_3}{2(x_3 - 1)}\,. \tag{8}$$

Comparing (2) and (4) we get

$$v_1 = v_2\,, \tag{9}$$

and (7)–(9) result in

$$\frac{x_3 - x_1}{x_1 + 2} = \frac{x_1 - x_3}{2(x_3 - 1)} \Rightarrow$$
$$x_3 = -\frac{x_1}{2}\,. \tag{10}$$

Combining (7) and (10) we obtain

$$v_1 = \frac{-3x_1}{2(x_1 + 2)}\,. \tag{11}$$

By substituting the variables in (2) with the help of (5), (6), (11), (10) we get

$$\begin{aligned} 0 &= 2(x_2 - x_4) + v_1 = 2[-(x_1+2)^2 - 1 - 2(x_3 - 1)^2] - \frac{3x_1}{2(x_1+2)} \Rightarrow \\ 0 &= -2(x_1+2)^2 - 2 - 4\left(\frac{x_1+2}{2}\right)^2 - \frac{3x_1}{2(x_1+2)}\,. \end{aligned}$$

By means of elementary transformations we obtain the equivalent cubic equation

$$0 = x_1^3 + 6x_1^2 + \frac{79}{6}x_1 + \frac{28}{3}$$

with the unique real solution

$$x_1^* \approx -1.364\,.$$

Now it is easy to obtain the values of the other variables:

$$\begin{aligned} x_2^* &\approx -1.405 \\ x_3^* &\approx 0.6818 \\ x_4^* &\approx 0.2024 \\ v_1^* = v_2^* &\approx 3.215\,. \end{aligned}$$

The cone of tangent directions is

$$\begin{aligned} C(x^*) &= \left\{ d \in \mathbb{R}^4 \;\middle|\; \begin{pmatrix} 2(x_1^*+2) & 1 & 0 & 0 \\ 0 & 0 & 4(x_3^*-1) & -1 \end{pmatrix} \begin{pmatrix} d_1 \\ \vdots \\ d_4 \end{pmatrix} = 0 \right\} \\ &= \left\{ \begin{pmatrix} d_1 \\ -2d_1(x_1^*+2) \\ d_3 \\ -2d_3(x_1^*+2) \end{pmatrix} \;\middle|\; d_1, d_3 \in \mathbb{R} \right\} \subset \mathbb{R}^4\,. \end{aligned} \tag{4.64}$$

We get

$$\begin{aligned} \Phi(x, v) &= f(x) + v^T h(x) \\ &= (x_1 - x_3)^2 + (x_2 - x_4)^2 + v_1[(x_1+2)^2 + 1 + x_2] + v_2[2(x_3-1)^2 - x_4]\,, \end{aligned}$$

thus

$$H_x\Phi(x,v) = \begin{pmatrix} 2(1+v_1) & 0 & -2 & 0 \\ 0 & 2 & 0 & -2 \\ -2 & 0 & 2(1+2v_2) & 0 \\ 0 & -2 & 0 & 2 \end{pmatrix},$$

and for $d \in C(x^*) \setminus \{0\}$ we get (see (4.64)):

$$\begin{aligned} d^T H_x\Phi(x^*, v^*)d &= 2((1+v_1^*)d_1^2 + d_2^2 + (1+2v_2^*)d_3^2 + d_4^2) - 4(d_1d_3 + d_2d_4) \\ &\approx 11.67d_1^2 - 10.48d_1d_3 + 18.10d_3^2 \\ &= (d_1, d_3)\begin{pmatrix} 11.67 & -5.24 \\ -5.24 & 18.10 \end{pmatrix}\begin{pmatrix} d_1 \\ d_3 \end{pmatrix}. \end{aligned}$$

This expression is positive, since the last matrix is positive definite (observe that $H_x\Phi(x^*, v^*)$ is not positive definite). Hence the second-order sufficient conditions are satisfied for the above solution x^*. The distance between the parables is $\sqrt{f(x^*)} \approx 2.602$ (see Figure 4.14).

We suggest that the reader solves the last example alternatively by means of substitution of variables.

Example 4.51. Consider again the problem of Example 4.14:

$$\begin{gathered} \min (x_1-1)^2 + (x_2-2)^2 \\ -x_1^2 + x_2 \leq 0 \\ x_1 + x_2 - 2 = 0\,. \end{gathered}$$

We have shown that there exist the two solution candidates $x^1 = (1,1)^T$ and $x^2 = (-2,4)^T$. Since the Lagrange multiplier u of the inequality is positive for x^1 and x^2, we obtain

$$\begin{aligned} C(x^1) &= \left\{ d \in \mathbb{R}^2 \,\middle|\, (d_1,d_2)\begin{pmatrix} -2 \\ 1 \end{pmatrix} = 0, (d_1,d_2)\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0 \right\} = \{0\}, \\ C(x^2) &= \left\{ d \in \mathbb{R}^2 \,\middle|\, (d_1,d_2)\begin{pmatrix} 4 \\ 1 \end{pmatrix} = 0, (d_1,d_2)\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0 \right\} = \{0\}\,, \end{aligned}$$

and the second-order sufficient conditions are trivially satisfied for these points.

We still consider a modification of the above problem, substituting the objective function by $f(x) := (x_1-2)^2 + (x_2-2)^2$. The points x^1 and x^2 remain optimal points and $C(x^2) = \{0\}$ remains unchanged. But for x^1 the multiplier associated to the inequality is now $u = 0$, thus

$$C(x^1) = \left\{ d \in \mathbb{R}^2 \,\middle|\, (d_1,d_2)\begin{pmatrix} -2 \\ 1 \end{pmatrix} \leq 0, (d_1,d_2)\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0 \right\} = \left\{ \alpha\begin{pmatrix} 1 \\ -1 \end{pmatrix} \,\middle|\, \alpha \geq 0 \right\}\,.$$

The Hessian matrix

$$H_x\Phi(x^*, u^*, v^*) = \begin{pmatrix} 2(1-u^*) & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

is positive definite, hence the second-order sufficient conditions are satisfied.

We suggest that the reader studies also the case $f(x) = (x_1-2)^2 + (x_2-1)^2$.

Exercise 4.52. *Which solution candidates of Exercise 4.15 satisfy the second-order sufficient conditions?*

We finally study some modifications of optimality conditions. When at least one of the Lagrange multipliers u_i is zero, the definition of the cone $C(x^*)$ in (4.60) contains inequalities and $C(x^*)$ is not a vector space (see Examples 4.47 (b) and 4.51). For large dimensions, the evaluation of the matrix $H_x\Phi(x^*, u^*, v^*)$ over the cone $C(x^*)$ can be computationally demanding.

In order to obtain "simpler" optimality conditions, we define the vector subspaces

$$\begin{aligned} C^-(x^*) = \{d \in \mathbb{R}^n | d^T \operatorname{grad} g_i(x^*) &= 0 \text{ for } i \in A(x^*)\,, \\ d^T \operatorname{grad} h_j(x^*) &= 0 \text{ for } j = 1, \dots, k\}\,, \\ C^+(x^*) = \{d \in \mathbb{R}^n | d^T \operatorname{grad} g_i(x^*) &= 0 \text{ for } i \in A(x^*) \text{ and } u_i^* > 0\,, \\ d^T \operatorname{grad} h_j(x^*) &= 0 \text{ for } j = 1, \dots, k\} \end{aligned}$$

of $\mathbb{R}^n$ which satisfy the relation

$$C^-(x^*) \subset C(x^*) \subset C^+(x^*)\,.$$

Therefore, if $H_x\Phi(x^*, u^*, v^*)$ is positive semidefinite over $C(x^*)$, it is also positive semidefinite over $C^-(x^*)$. Thus we obtain a necessary optimality condition by substituting $C(x^*)$ by $C^-(x^*)$ in (4.60). Similarly, we obtain a sufficient condition by substituting $C(x^*)$ for $C^+(x^*)$ in (4.61).

Exercise 4.53. *Given the problem*

$$\begin{aligned} \min (x_1-2)^2 + x_2^2 + x_3^2 \\ -x_1^2 + x_2^2 + x_3^2 \le 1 \\ x_3 \ge 0 \\ x_1 + x_2 + x_3 = 1\,. \end{aligned}$$

Determine $C(x^*)$, $C^-(x^*)$ and $C^+(x^*)$ for the point $x^* = (1, 0, 0)^T$.

4.5 References for Part I

A mathematically demanding representation of the theory introduced in Chapter I and diverse extensions can be found, for example, in Avriel (2003), Bazaraa and Shetty (1979), Bazaraa, Sherali and Shetty (2006), Luenberger (2003) and Bertsekas (2004).

The theory of nonlinear programming is also studied in the books of Fritzsche (1978), Mahey (1987) and Mateus and Luna (1986).

A first systematic study of convex sets was performed by Minkowski (1911). Generalizations of the convexity are considered, for example, in Bazaraa, Sherali and Shetty (2006, Section 3.5), Fritzsche (1978) and Mahey (1987). The theory of convexity and various applications in optimization are also intensively studied in Rockafellar (1996).

The subgradient method is described, for example, in a paper by Boyd et. al.: http://www.stanford.edu/class/ee3920/subgrad_method.pdf.

Applications of this method to solve the traveling salesman problem are described in Held and Karp (1970) and in the site
http://www.lac.inpe.br/~lorena/marcelo/Gera-colunas-TSP.pdf.

Applications in capacitated location problems can be found in:
http://www.po.ufrj.br/projeto/papers/loccap1.html.

Subgradients are also applied to optimization problems in elastic-plastic analysis of structural and mechanical components, see:
http://www.mec.pucrio.br/~teses/teses/MildredHecke.htm.

Other applications of subgradients can be found in Held, Wolfe and Crowder (1974), Wolfe (1976) and in the more recent article Larsson, Patriksson and Strömberg (1996).

The KKT conditions were independently developed by Karush (see the unpublished thesis of Karush (1939)) and Kuhn and Tucker (1951). A summary of optimality conditions is given in McCormick (1983), and the historical development is documented in Kuhn (1976). Since 1996 several studies have been devoted to the formulation of optimality conditions for specific nonlinear problems. For example, for optimal control problems (see the websites: http://www.epubs.siam.org/sam-bin/dbq/article/26198 and http://www.math.ohiou.edu/~mvoisei/fnnc.pdf) for optimization with multiple criteria (see: http://www.bellaassali.chez.tiscali.fr/sicon.pdf, http://www.cambric.ec.unipi.it/reports/Rep.103.pdf) for optimization with complementarity constraints (see: http://www.epubs.siam.org/sam-bin/dbq/article/32188), and semidefinite optimization (see: http://www.citeseer.ist.psu.Edu/120763.html).

Second-order optimality conditions are intensively studied in Avriel (2003), Bertsekas (2004), Conn, Gould and Toint (2000) and in the site:
http://www.numerical.rl.ac.uk/nimg/course/lectures/paper/paper.pdf.

An article about higher order optimality conditions can be found in:
http://www.ima.umn.edu/~olga/ICCSA03.pdf.

A detailed discussion of duality with possible applications is given in Geoffrion (1971). Diverse duality concepts are listed in the “Mathematical Programming Glossary” of the INFORMS Computing Society:
http://www.glossary.computing.society.informs.org.

In particular, the Wolfe dual is studied in the course “Nonlinear Programming” of Kees Roos:
http://www.isa.ewi.tudelft.nl/~roos/courses.

Recent developments regarding duality can be found in Bertsekas (2004).

Part II: **Solution methods**

5 Iterative procedures and evaluation criteria

The basic idea to solve an optimization problem iteratively is the following: choose an initial point x^0 that can be considered as a first approximation for the optimal point x^*. Apply repeatedly a procedure that (usually) improves the actual approximation. In this way, a sequence $x^0, x^1, x^2, \ldots$ of points is generated which in the ideal case converges to x^*. In many practical applications the convergence to an optimal point cannot be guaranteed, but the limit point $\lim_{k\to\infty} x^k$ satisfies at least the first-order necessary optimality conditions. Basically, the theory of iterative procedures can be subdivided into three topics which are naturally not disjointed.

The first topic treats the construction of algorithms. Making use of the respective particularities of the considered problem, the algorithm must be designed such that it works as efficient as possible. In the following chapters we will present various procedures which are convergent under determined conditions. We prefer to omit the proofs of convergence which are available in the literature, cited in Section 11.3. Most of the procedures studied in this book are "descent procedures", i.e. the condition

$$f(x^{k+1}) < f(x^k)$$

is satisfied for $k = 0, 1, 2, \ldots$. In many applications this condition causes convergence to the point x^*. In practice, the procedure terminates when a certain stopping criterion is satisfied. Among others, the following criteria are frequently used:

$$\begin{aligned} &\Delta_k \le \varepsilon\,, &&\frac{\Delta_k}{|x^{k+1}|} \le \varepsilon\,, &&\frac{\Delta_k}{\Delta_{k-1}} \le \varepsilon\,,\\ &\Delta f_k \le \varepsilon\,, &&\frac{\Delta f_k}{|f(x^{k+1})|} \le \varepsilon\,, &&\frac{\Delta f_k}{\Delta f_{k-1}} \le \varepsilon\,, \end{aligned}$$

where $\Delta_k := |x^{k+1} - x^k|$ represents the distance between two consecutive points of the sequence and $\Delta f_k := |f(x^{k+1}) - f(x^k)|$ denotes the absolute difference between the corresponding objective function values.

The second topic of the theory of iterative procedures analyzes the question for which starting point x^0 the sequence $x^0, x^1, x^2 \ldots$ converges to the optimal solution x^*. If $\lim_{k\to\infty} x^k = x^*$ for any x^0, the procedure is called globally convergent. If this relation holds only for any x^0, chosen from a neighborhood of x^*, we are talking about local convergence. Questions concerning convergence will be treated for diverse procedures introduced in part II of this book (Theorems 6.11, 7.13 and 7.21, among others). Finally, the third topic of the theory treats the evaluation of the "quality" of a computational procedure. The remainder of the chapter is dedicated to this subject.

We introduce some criteria to evaluate the performance of a procedure. We assume that the sequence $\{x^i\}$ converges to x^* and that f is a continuous function, and

therefore $\{f(x^i)\}$ converges to $f(x^*)$. We consider the three criteria:
(C1) convergence speed
(C2) robustness
(C3) stability

(C1) In the study of convergence speed we consider the error sequence

$$r_k = |x^k - x^*| \quad \text{or} \quad r_k = |f(x^k) - f(x^*)|,$$

which, by hypothesis, converges to 0. The convergence speed (at iteration k) can be measured, for example, by the smallest number of iterations q, necessary to reduce the error r_k at least by the factor 1/10. So we look for the smallest integer q such that

$$\frac{r_{k+q}}{r_k} \leq 1/10\,.$$

Example 5.1. Consider three error sequences that satisfy, respectively, the following conditions for $k = 1, 2, \ldots$
(a) $r_k \leq 0.99 r_{k-1}$
(b) $r_k \leq (1/3) r_{k-1}$
(c) $r_k \leq (1/2)^k r_{k-1}$.

(a) It holds that $r_{k+q} \leq 0.99^q r_k$. Therefore we look for the smallest integer q such that

$$0.99^q \leq 1/10 \Leftrightarrow$$
$$q \geq \ln 0.1/\ln 0.99 = 229.1\,.$$

Independent from the iteration k we obtain $q = 230$, i.e. in the worst case we need 230 iterations to reduce the error by the factor 1/10. If for example $r_k = 0.02$, then up to 230 iterations may be necessary to make the error smaller then 0.002. The number of iterations might be less if $r_k < 0.99 r_{k-1}$ holds for some iterations.

(b) Similar to the previous case we obtain $r_{k+q} \leq (1/3)^q r_k$ and

$$(1/3)^q \leq 1/10 \Leftrightarrow$$
$$q \geq \ln 0.1/\ln(1/3) = 2.096\,.$$

Hence, $q = 3$. In this case only three iterations are sufficient to reduce the error by the factor 1/10.

(c) For a given k it holds

$$r_{k+q} \leq (1/2)^{(k+q)+\cdots+(k+1)} r_k = (1/2)^{(2k+q+1)q/2} r_k\,.$$

The integer q satisfies

$$(1/2)^{(2k+q+1)q/2} \leq 1/10 \Leftrightarrow$$
$$q^2/2 + (k + 1/2)q \geq \ln 0.1/\ln 0.5 = 3.322 \Leftrightarrow$$
$$q \geq -(k + 0.5) + \sqrt{(k + 0.5)^2 + 6.644}\,. \tag{5.1}$$

Now the convergence speed increases rapidly with the number k of realized iterations. For example, for $k = 1$ relation (5.1) yields

$$q \geq -1.5 + \sqrt{1.5^2 + 6.644} = 1.48 \Rightarrow q = 2\,.$$

For $k = 3$ we get

$$q \geq -3.5 + \sqrt{3.5^2 + 6.644} = 0.85 \Rightarrow q = 1\,.$$

Obviously the procedure (b) is much better than (a), and (c) is better than (b). Procedure (a) is not applicable in practice because of its slow convergence.

Exercise 5.2. *An error sequence satisfies $r_k = 0.9$ and*

$$r_k = 0.8(r_{k-1})^2 \quad (\textit{for } k = 1, 2, \dots)\,.$$

Determine the convergence speed for $k = 2$.

We still introduce some criteria, frequently used to describe the convergence speed of an algorithm.

Definition 5.3. Let $\{r_k\}$ be the error sequence of a procedure. We say that the procedure is linearly convergent (with convergence rate not exceeding β) if

$$\frac{r_{k+1}}{r_k} \leq \beta$$

holds for a sufficiently large k $(0 < \beta < 1)$.

The procedure is called *superlinearly convergent* if

$$\lim_{k \to \infty} \frac{r_{k+1}}{r_k} = 0\,.$$

If

$$\frac{r_{k+1}}{(r_k)^p} \leq y$$

for $y > 0, p > 1$ and sufficiently large k, we say that the procedure converges *with order not inferior to p*. In particular, for $p = 2$ and $p = 3$ we say that the convergence is *quadratic* or *cubic*, respectively.

Note that the constant y need not be smaller than 1.

For example, the procedures (a) and (b) of Example 5.1 converge linearly, (c) converges superlinearly, and the procedure of Exercise 5.2 converges quadratically.

Exercise 5.4. *In order to compare linear and quadratic convergence, calculate the first five values for the following error sequences:*

(a) $r_k = 0.9 r_{k-1}$, $r_0 = 1$,

(b) $r_k = 0.9(r_{k-1})^2$, $r_0 = 1$.

Exercise 5.5. *Show that*
(a) *quadratic convergence implies superlinear convergence*
(b) *superlinear convergence implies linear convergence.*

We now describe the concepts robustness and stability in general terms.
(C2) A procedure is called *robust*, if it works for a large class of relevant problems, i.e. the functions of the optimization problem need to satisfy only a few conditions which are generally satisfied in applications.
(C3) A procedure is called *stable*, if rounding errors do not seriously influence the convergence.

In addition to the criteria (C1)–(C3) above, the computational work per iteration (say, time spent) and memory demand (say, physicals space for storage of intermediate data) play an important role in the evaluation of procedures. Finally, a meaningful evaluation requires several practical tests.

Modern implementations of optimization algorithms are usually very sophisticated. Among others, the application of error theory and numerical analysis in general can improve their stability. In this introductory text we can only illustrate the basic principles of the methods.

6 Unidimensional minimization

In this chapter we minimize a function of a single variable. This topic might appear to be of minor practical interest, since realistic models usually have multiple variables. However, a one-dimensional optimization appears as a subproblem in most methods to solve a multidimensional problem. We will now introduce some of the most common procedures to minimize a one-dimensional function. The first two methods make only use of the function values and can therefore be applied to nondifferentiable functions. After that we present two methods using the first derivative. In Section 6.2 we introduce Newtons's method that needs the first and second derivatives. Later we minimize a function by approximating it by a polynomial (Section 6.3). The chapter ends with some considerations about the practical use of the methods (Section 6.4).

Let $f : \mathbb{R} \to \mathbb{R}$ be function. We consider the problems

$$\min_{x \in \mathbb{R}_+} f(x) \tag{6.1}$$

with $\mathbb{R}_+ = \{x \in \mathbb{R} \,|\, x \geq 0\}$ and

$$\min_{x \in [a,b]} f(x) \tag{6.2}$$

with $a, b \in \mathbb{R}, a < b$.

To ensure that an iterative procedure converges to a global minimum point, it is often assumed that the function f is unimodal, a condition quite restrictive, but often satisfied in applications:

Definition 6.1. A real function f is called *unimodal* over the interval $I = [a, b]$, $a < b$, if there is a point $\overline{x} \in I$ such that
(i) f is strictly decreasing over $[a, \overline{x}]$, and
(ii) f is strictly increasing over $[\overline{x}, b]$.

A unimodal function is illustrated in Figure 6.1. In general, $\overline{x}$ can be any point of the interval $[a, b]$. The extreme cases, i.e. the cases $\overline{x} = a$ and $\overline{x} = b$ correspond to a strictly increasing or strictly decreasing function, respectively.

Obviously, $\overline{x}$ is the unique global minimum point of a unimodal function. In particular, a strictly convex function over I is unimodal, but the converse is not true, since inflection points may occur (see Figure 6.1). If f is unimodal and differentiable over $[a, b]$ with $a < \overline{x} < b$, it holds $f'(x) \leq 0$ over $[a, \overline{x})$, $f'(x) \geq 0$ over $(\overline{x}, b]$ and $f'(\overline{x}) = 0$. Note that the derivative can be zero in a point different from $\overline{x}$ (see Exercise 6.8).

Definition 6.1 can be extended to an unbounded interval I, permitting the values $a = -\infty$ and $b = \infty$.

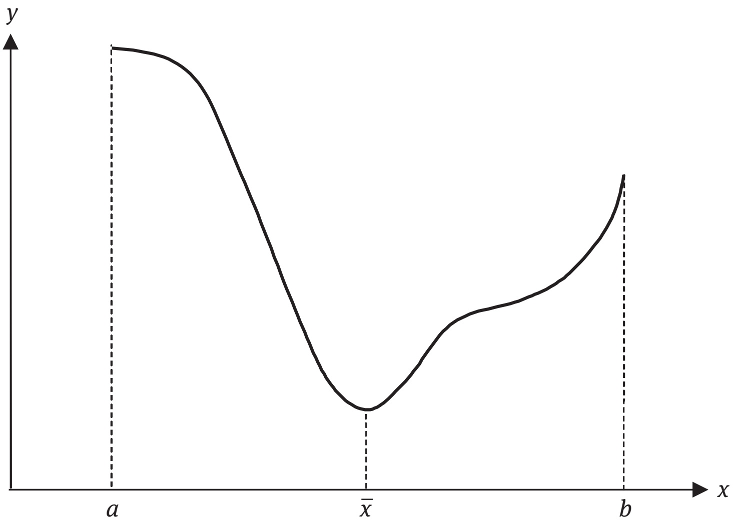


Fig. 6.1. Unimodal function.

6.1 Delimitation of the search region

If f is unimodal over $\mathbb{R}_+$ the following procedure reduces the problem (6.1) to the form (6.2).

Delimitation algorithm I
Choose a real number $\rho > 0$ and calculate $f(0)$ and $f(\rho)$. If $f(\rho) \geq f(0)$, the global minimum point $\overline{x}$ is an element of the interval $[a, b] := [0, \rho]$. Otherwise, compute $f(2\rho), f(3\rho)$, etc., until $f(k\rho)$ is larger or equal to $f((k-1)\rho)$. In this case, $\overline{x}$ belongs to the interval $[a, b] := [(k-2)\rho, k\rho]$.

We emphasize that in general the minimum $\overline{x}$ can be located anywhere in the interval $[(k-2)\rho, k\rho]$. In part (a) of the next example, $\overline{x}$ is an element of $[(k-1)\rho, k\rho]$ and in part (b) it holds $\overline{x} \in [(k-2)\rho, (k-1)\rho]$. An interval containing the optimal solution is frequently called a *bracket*.

Example 6.2. Given is the unimodal function $f(x) = (x-2)^2$ over $\mathbb{R}_+$ with $\overline{x} = 2$.

(a) For $\rho = 0.8$ we obtain (see Figure 6.2 (a))
$$f(0) = 4 > f(\rho) = 1.44 > f(2\rho) = 0.16 \leq f(3\rho) = 0.16\,,$$
i.e. $k = 3$ and $\overline{x} \in [(k-1)\rho, k\rho] = [1.6, 2.4]$.

(b) For $\rho = 0.7$ we obtain (see Figure 6.2 (b))
$$f(0) = 4 > f(\rho) = 1.69 > f(2\rho) = 0.36 > f(3\rho) = 0.01 \leq f(4\rho) = 0.64\,,$$
i.e. $k = 4$ and $\overline{x} \in [(k-2)\rho, (k-1)\rho] = [1.4, 2.1]$.

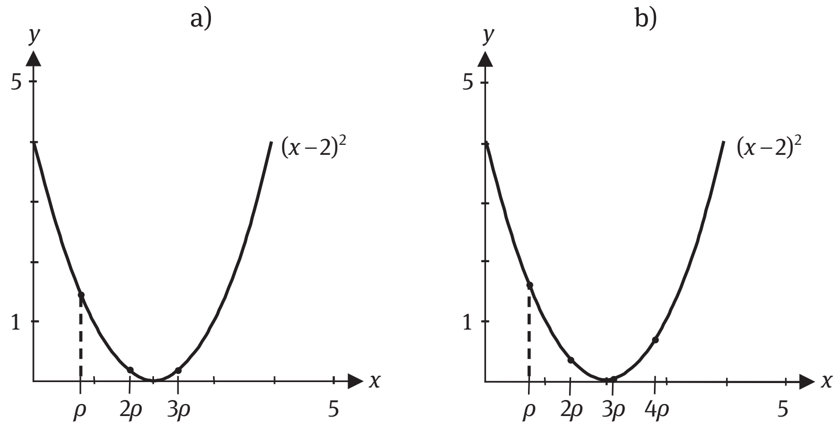


Fig. 6.2. Locating of $\overline{x}$.

The above algorithm continues until $\overline{x}$ is delimited. Clearly, the number ρ should not be "too large", and in order to terminate the process quickly, the value of ρ should not be "too small". However, we will not pursue the question what is the most appropriate value of ρ, since the next procedure operates efficiently to shorten the interval containing $\overline{x}$.

Golden Section algorithm

We now solve the problem (6.2), where f is unimodal. Starting with the interval $[a_0, b_0] := [a, b]$, a sequence of intervals $[a_k, b_k]$ is constructed such that

(i) any interval contains $\overline{x}$,

(ii) $[a_{k+1}, b_{k+1}] \subset [a_k, b_k]$ for $k = 0, 1, 2, \ldots$

(iii) the length of the interval $L_k := b_k - a_k$ converges to 0, i.e. $\lim_{k\to\infty} L_k = 0$.

Note that, in this way, $\overline{x}$ can be determined with arbitrary accuracy. In constructing $[a_{k+1}, b_{k+1}]$ from $[a_k, b_k]$, two intermediate points v_k and w_k are calculated, satisfying

$$a_k < v_k < w_k < b_k\,,$$

i.e. $v_k = a_k + \alpha L_k$ and $w_k = a_k + \beta L_k$ (where α, β are arbitrarily chosen with $0 < \alpha < \beta < 1$ and independent of k). Since f is unimodal, we get (see Figure 6.3):

$$f(v_k) < f(w_k) \Rightarrow \overline{x} \in [a_k, w_k]\,, \qquad \text{(case A)}$$

$$f(v_k) \geq f(w_k) \Rightarrow \overline{x} \in [v_k, b_k]\,. \qquad \text{(case B)}$$

So we set $[a_{k+1}, b_{k+1}] := [a_k, w_k]$ in case A and $[a_{k+1}, b_{k+1}] := [v_k, b_k]$ in case B.

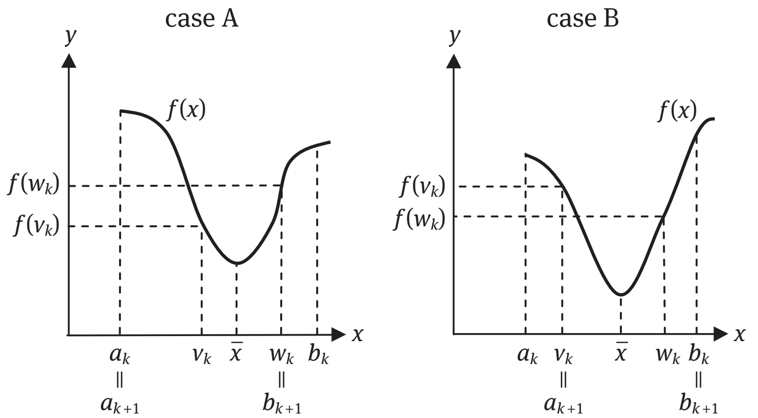


Fig. 6.3. Construction of the new interval $[a_{k+1}, b_{k+1}]$.

To reduce the computational work on building successive intervals, it is reasonable to use one of the intermediate points of $[a_k, b_k]$ as an intermediate point of $[a_{k+1}, b_{k+1}]$ instead of constructing two new points. In case A we use v_k as an intermediate point of the interval $[a_{k+1}, b_{k+1}]$ and in case B we use w_k.

For this purpose we choose α and β such that $w_{k+1} = v_k$ in case A and $v_{k+1} = w_k$ in case B. These conditions determine α and β uniquely:

In case A we have

$$\begin{aligned} w_{k+1} &= a_{k+1} + \beta L_{k+1} = a_k + \beta(w_k - a_k) = a_k + \beta^2 L_k\,, \\ v_k &= a_k + \alpha L_k\,. \end{aligned}$$

Thus, if $\alpha = \beta^2$ we have $w_{k+1} = v_k$ in the case A (and we only need to compute v_{k+1}).

In case B we have

$$\begin{aligned} v_{k+1} &= a_{k+1} + \alpha L_{k+1} = v_k + \alpha(b_k - v_k) = a_k + \alpha L_k + \alpha(b_k - a_k - \alpha L_k) \\ &= a_k + \alpha L_k + \alpha L_k - \alpha^2 L_k = a_k + (2\alpha - \alpha^2) L_k\,, \\ w_k &= a_k + \beta L_k\,. \end{aligned}$$

And therefore, if $2\alpha - \alpha^2 = \beta$, we have $v_{k+1} = w_k$ in case B (and need only calculate w_{k+1}).

Summarizing the above considerations, α and β must satisfy the system

$$\begin{aligned} \alpha &= \beta^2 \\ 2\alpha - \alpha^2 &= \beta \\ 0 < \alpha &< \beta < 1\,. \end{aligned}$$

The two equations imply

$$\beta^4 - 2\beta^2 + \beta = \beta(\beta - 1)(\beta^2 + \beta - 1) = 0\,,$$

and because of the inequalities the unique solution is

$$\beta = \frac{\sqrt{5}-1}{2} \approx 0.618\,, \quad \alpha = \beta^2 = \frac{3-\sqrt{5}}{2} \approx 0.382\,. \tag{6.3}$$

The construction of intervals with the above definition of α and β is called the *Golden Section algorithm*.

With this division, in either case, the new points v_k and w_k divide the interval $[a_k, b_k]$ symmetrically, i.e. $v_k - a_k = b_k - w_k$. The name of the algorithm is explained by the fact that

$$\frac{w_k - a_k}{b_k - a_k} = \frac{b_k - w_k}{w_k - a_k}\,.$$

This ratio is known since antiquity and is considered aesthetically perfect which explains the name "Golden Section".

The algorithm terminates when $L_k < 2\varepsilon$, where $\varepsilon > 0$ is a given error threshold. By setting $\tilde{x} := (a_k + b_k)/2$, we obtain an approximation for the optimal point $\bar{x}$, satisfying $|\tilde{x} - \bar{x}| < \varepsilon$.

Example 6.3. To illustrate the procedure we minimize the function $f(x) = (x-2)^2$ over the interval $[1, 4]$. We indicate the first three iterations in four-digit arithmetic:

Table 6.1. Golden Section.

k	a_k	b_k	v_k	w_k	Case
0	1	4	2.146	2.854	A
1	1	2.854	1.708	2.146	B
2	1.708	2.854	2.146	2.416	A
3	1.708	2.416	1.978	2.146	A

Note that from the first iteration only the values v_1, w_2 and v_3 need to be calculated. The other values w_1, v_2 and w_3 can be derived from the previous line. The actual approximation for the optimal solution $\bar{x} = 2$ is $\tilde{x} = (a_3 + b_3)/2 = 2.062$.

Exercise 6.4. *Minimize the function* $f(x) = |x/2 - 1|$ *over the interval* $[0, 3]$ *with the Golden Section algorithm! Perform three iterations!*

Exercise 6.5. *Show that any iteration of the Golden Section algorithm reduces the length of the interval by the factor β in (6.3), i.e.*

$$\frac{L_{k+1}}{L_k} = \beta \quad \textit{for } k = 0, 1, 2, \ldots$$

This means that with respect to the length of the delimiting interval, the convergence is linear (with convergence rate not exceeding β). However, the error sequence need not be decreasing (see next exercise).

Exercise 6.6. *How many iterations are required in Example 6.3 to make sure that the absolute error* $r_k := |\overline{x} - \widetilde{x}_k|$ *is at most* 10^{-4}*? Verify that the sequence* $\{r_k\}$ *is not decreasing!*

If the function f in (6.1) or (6.2) is unimodal and differentiable, the following simple alternatives to the above two methods can be used. In the following we assume that

$$f'(x) \neq 0 \text{ for } x \neq \overline{x} \quad \text{and } a < \overline{x} < b\,.$$

Delimitation algorithm II

Problem (6.1) is reduced to (6.2). For a given $\rho > 0$, it is calculated $f'(\rho), f'(2\rho)$, $f'(3\rho)$, etc., until $f'(k\rho) \geq 0$. If $f'(k\rho) = 0$ we have $\overline{x} = k\rho$. If $f'(k\rho) > 0$, then $\overline{x}$ is an element of the interval $[a, b] := [(k-1)\rho, k\rho]$. (Compare with the delimitation algorithm I!)

Problem (6.2) can be solved by the following procedure:

Bisection method

We start with $[a_0, b_0] := [a, b]$. In each iteration we calculate $m_k := (a_k + b_k)/2$. If $f'(m_k) = 0$, it holds $\overline{x} = m_k$. We set $[a_{k+1}, b_{k+1}] := [a_k, m_k]$ if $f'(m_k) > 0$ and $[a_{k+1}, b_{k+1}] := [m_k, b_k]$ if $f'(m_k) < 0$.

The sequence of intervals $[a_k, b_k]$ satisfies the conditions (i)–(iii) of the Golden Section algorithm and the procedure terminates when $L_k < 2\varepsilon$. As before, the approximation for the optimal point $\overline{x}$ is given by $\widetilde{x} := (a_k + b_k)/2$ with absolute error $|\widetilde{x} - \overline{x}| < \varepsilon$.

Exercise 6.7. *Apply the bisection algorithm to Example 6.3. Realize three iterations. How many iterations are required to make sure that* $|\widetilde{x} - \overline{x}| \leq 10^{-4}$ *(see Exercise 6.6).*

Exercise 6.8. *Find a function* f *that is unimodal and differentiable on an interval* $[a, b] (a < b)$ *and does not satisfy the condition* $f'(x) \neq 0$ *for* $x \neq \overline{x}$ *assumed above. Show that the bisection algorithm can fail in this case.*

6.2 Newton's method

A minimum point $\overline{x}$ of a differentiable function g satisfies the necessary condition $g'(\overline{x}) = 0$ (Section 2.2). Therefore, to minimize a function g one must search for a root of the derivative $f = g'$. We now present a very intuitive and efficient method to determine a root x^* of a continuously differentiable function $f : [a, b] \to \mathbb{R}$. Later we will apply this procedure to minimize a function g.

We assume that $f(a) < 0 < f(b)$ and $f'(x) > 0$ for all $x \in [a, b]$. Starting with a point x_0 we construct a sequence $x_0, x_1, x_2, \ldots$ of points which usually converges to

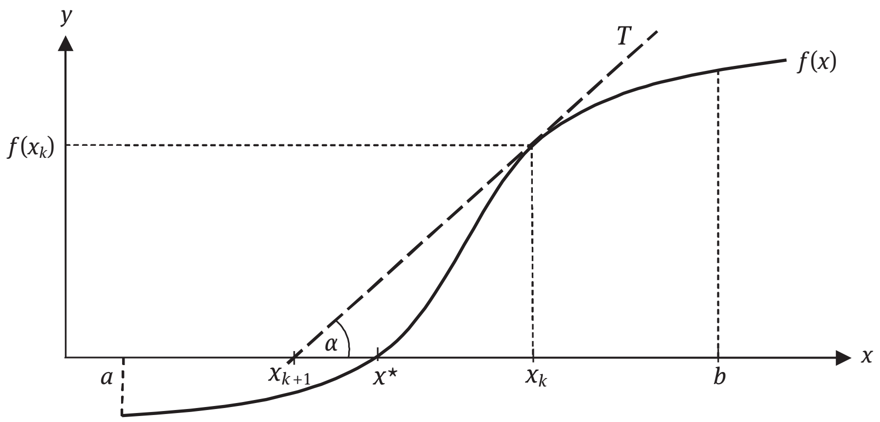


Fig. 6.4. Iteration of Newton's method.

x^* if x_0 lies "sufficiently close to x^*". For a given point x_k, constructed is the tangent line T of the graph of f at the point $(x_k, f(x_k))^T$, and x_{k+1} is chosen as the intersection between T and the x-axis (Figure 6.4).

Since $f'(x_k)$ is the tangent of the angle α (the slope of T), we have

$$\begin{aligned} f'(x_k) &= \frac{f(x_k)}{x_k - x_{k+1}} \Leftrightarrow \\ x_{k+1} &= x_k - \frac{f(x_k)}{f'(x_k)}\,. \end{aligned} \tag{6.4}$$

The iterative construction of the points x_k according to (6.4) is called *Newton's method*. In practice the procedure ends when $|f(x_k)|$ is sufficiently small.

Example 6.9. We determine a root x^* of the function

$$f(x) = x^3 + x - 1\,.$$

Since $f'(x) = 3x^2 + 1$, we have

$$x_{k+1} = x_k - \frac{x_k^3 + x_k - 1}{3x_k^2 + 1} = \frac{2x_k^3 + 1}{3x_k^2 + 1}\,.$$

With the initial value $x_0 = 1$, we obtain in three-digit arithmetic $x_1 = 0.75$, $x_2 \approx 0.686$, $x_3 \approx 0.682$, $x_4 \approx 0.682, \ldots$ We conclude that the root is $x^* \approx 0.682$.

Now we determine a minimum point of a function g, applying the above procedure to $f := g'$: Assume that f satisfies the conditions established above, i.e. g is twice continuously differentiable, and

$$g'(a) < 0 < g'(b)\,, \quad g''(x) > 0 \quad \text{for all } x \in [a, b]\,. \tag{6.5}$$

Such a function is strictly convex and has a unique global minimum point $\overline{x}$, satisfying $g'(\overline{x}) = 0$ (see Section 3.5). Analogously to (6.4) we now calculate

$$x_{k+1} = x_k - \frac{g'(x_k)}{g''(x_k)} . \tag{6.6}$$

One can easily verify that x_{k+1} is the minimum point of the Taylor approximation

$$g(x_k) + g'(x_k)(x - x_k) + \tfrac{1}{2}g''(x_k)(x - x_k)^2 \tag{6.7}$$

of the function g around the point x_k.

The procedure typically converges to the minimum point $\overline{x}$ of g (Corollary 6.12). The stopping criterion is usually $|g'(x_k)| \le \varepsilon$ for a predetermined $\varepsilon > 0$.

Example 6.10. We consider the function

$$g(x) = x^2 + 3e^{-x} \quad \text{over } [0, 3] .$$

We have $g'(x) = 2x - 3e^{-x}, g''(x) = 2 + 3e^{-x}$. Thus, $g'(0) = -3 < 0$, $g'(3) = 6 - 3e^{-3} > 0$ and $g''(x) > 0$ for all $x \in [0, 3]$, i.e. (6.5) is satisfied, g is strictly convex and has a unique global minimum point.

The relation (6.6) yields

$$x_{k+1} = x_k - \frac{2x_k - 3e^{-x_k}}{2 + 3e^{-x_k}} . \tag{6.8}$$

Starting, for example, with $x_0 = 1$, the recurrence formula (6.8) generates the sequence:

$$x_0 = 1 , \quad x_1 \approx 0.711 , \quad x_2 \approx 0.726 , \quad x_3 \approx 0.726 , \ldots$$

One can observe a very rapid convergence to the minimum point $\overline{x} \approx 0.726$.

In the remainder of the section we will discuss some convergence properties. Newton's method is not globally convergent, i.e. does not determine an optimal solution for all starting values x_0. Figure 6.5 illustrates the generation of a useless point for the procedure (6.4). If x_0 is not sufficiently close to the root x^*, the slope of the tangent line T may be very small, resulting in a point x_1 which is far away from x^*.

Another problem which can occur theoretically, is a certain type of cycling. If the function f is, for example, given by

$$f(x) = \begin{cases} c\sqrt{x} & \text{for } x \ge 0 , \\ -c\sqrt{-x} & \text{for } x < 0 (c \ne 0) , \end{cases}$$

we have

$$f'(x) = \frac{c}{2\sqrt{|x|}} \quad \text{for } x \ne 0 ,$$

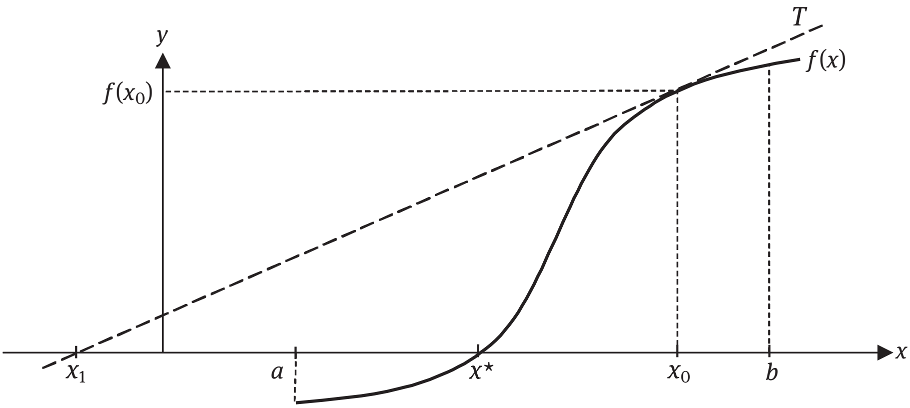


Fig. 6.5. Construction of a useless value.

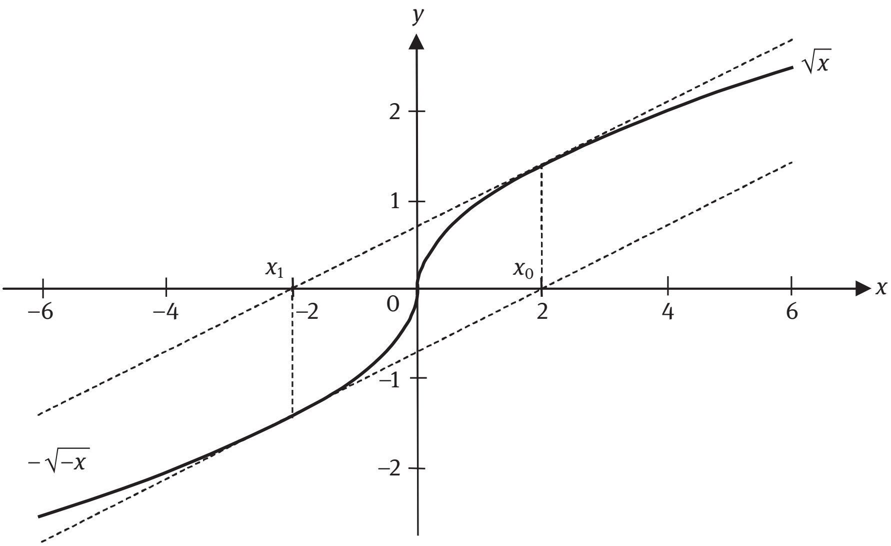


Fig. 6.6. Cycling of Newton's method.

resulting in

$$x - \frac{f(x)}{f'(x)} = -x \quad \text{for all } x \neq 0\,.$$

In this case, the procedure (6.4) generates an alternating sequence of points $x_0, x_1 = -x_0, x_2 = x_0, x_3 = -x_0, \ldots$ Figure 6.6 illustrates such a sequence geometrically for $c = 1$ and $x_0 = 2$.

We now present a theorem with respect to the convergence of procedure (6.4).

Theorem 6.11. *Let $f : [a,b] \to \mathbb{R}$ be a twice continuously differentiable function, satisfying the following conditions:*
(i) $f(a) < 0 < f(b), f'(x) > 0$ *for all* $x \in [a,b]$,
(ii) $f''(x) \le 0$ *for all* $x \in [a,b]$, *or* $f''(x) \ge 0$ *for all* $x \in [a,b]$,
(iii) $|f(a)/f'(a)| < b - a$ *and* $|f(b)/f'(b)| < b - a$.

Then, for every $x_0 \in [a,b]$, Newton's method (6.4) converges quadratically to the unique root of f.

The condition (ii) above implies in particular that the function f is strictly concave or strictly convex over $[a,b]$. Note that the function in Figure 6.6 does not satisfy this assumption.

For the minimization of a function g we get immediately:

Corollary 6.12. *Let $g : [a,b] \to \mathbb{R}$ be a three times continuously differentiable function, satisfying the following conditions:*
(i) $g'(a) < 0 < g'(b), g''(x) > 0$ *for all* $x \in [a,b]$,
(ii) $g'''(x) \le 0$ *for all* $x \in [a,b]$, *or* $g'''(x) \ge 0$ *for all* $x \in [a,b]$,
(iii) $|g'(a)/g''(a)| < b - a$ *and* $|g'(b)/g''(b)| < b - a$.

Then, for every $x_0 \in [a,b]$, Newton's method (6.6) converges quadratically to the unique global minimum point of g.

Example 6.13. Using the corollary, we now prove that the procedure in Example 6.10 converges regardless of the initial point $x_0 \in [0,3]$. The condition (i) of the corollary has already been proved. Since $g'''(x) = -3e^{-x}$, (ii) is satisfied. Furthermore we have

$$\left|\frac{g'(0)}{g''(0)}\right| = \left|\frac{-3}{5}\right| < 3 \quad \text{and} \quad \left|\frac{g'(3)}{g''(3)}\right| = \frac{6-3e^{-3}}{2+3e^{-3}} < \frac{6}{2} = 3\,,$$

that is, (iii) is also satisfied.

A generalization of Newton's method for functions of several variables will be studied in Section 7.2.

Exercise 6.14. *Determine the root of the function*

$$f(x) = x^2 + \sqrt{x} - 15\,, \quad x > 0$$

by method (6.4). Choose $x_0 = 4$ and perform three iterations.

Exercise 6.15.

(a) *Solve the problem*

$$\min 1/x^2 + x\,, \quad x > 0$$

by method (6.6). Choose $x_0 = 1$ *and perform four iterations.*

(b) *Prove that the procedure converges to the minimum point if* $0.1 \le x_0 \le 1.9$. *Show that it does not converge to a finite value for* $x_0 = 3$.

Exercise 6.16 (Hypothesis Testing). *The inhabitants of an island may be descendants of the civilization A (null hypothesis,* H_0*) or of the civilization B (alternative hypothesis,* H_1*). To decide which hypothesis has to be adopted, the average height* $\overline{x}$ *of a random sample of 100 adult residents of the island has been determined. If* $\overline{x}$ *exceeds a certain height* t *(to be determined),* H_1 *is accepted, otherwise* H_0 *is accepted. Mean and standard deviations of the heights of the two civilizations are:*

$$A\colon \mu = 176\,\text{cm}\,, \quad \sigma = 15\,\text{cm}\,,$$
$$B\colon \mu = 180\,\text{cm}\,, \quad \sigma = 12\,\text{cm}\,.$$

To determine the critical value t*, the following types of errors are considered:*

error I : *accept* H_1 *when, in reality,* H_0 *is true,*

error II : *accept* H_0 *when, in reality,* H_1 *is true.*

It is known from the statistical literature that:

- $P(\text{error } I) = P(\overline{X} > t \,|\, \mu = 176) = 1 - \Phi(\frac{t-176}{15/\sqrt{100}})$,
- $P(\text{error } II) = P(\overline{X} \le t \,|\, \mu = 180) = \Phi(\frac{t-180}{12/\sqrt{100}})$.

According to the statistical conventions we denote by

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

the density of the reduced normal distribution, and by

$$\Phi(x) = \int_{-\infty}^{x} \varphi(s)\,ds$$

the corresponding cumulative probability function.

Determine the critical value t*, such that* $P(\text{error I}) = P(\text{error II})$! *Apply method (6.4) with* $t_o = 178$. *(Make use of an appropriate software or normal distribution table to calculate the function* $\Phi(x)$*.)*

6.3 Interpolation methods

Several iterative procedures to minimize a one-dimensional function f, which are quite efficient in practice, are based on the following idea: the function f is successively approximated by polynomials or other “simple” functions, the minimum points of which converge to the minimum point $\overline{x}$ of f. The procedure may work in the following way or similarly: in the first iteration, the function f is approximated by a polynomial p of degree r, such that f and p coincide at different points $x_0, x_1, \ldots, x_r$, chosen from a neighborhood of $\overline{x}$. Then the minimum point of p, denoted by x_{r+1} is determined which represents a first approximation to $\overline{x}$. In the next iteration a new polynomial p is determined such that f and p coincide at the points $x_1, \ldots, x_{r+1}$. The minimum point of this polynomial is denoted by x_{r+2} and represents a better approximation for $\overline{x}$, etc.

In practice, p is generally of second or third degree. In these cases we talk about *quadratic* or *cubic interpolation*, respectively. We first illustrate the quadratic interpolation, where $p(x)$ can be written as

$$p(x) = \frac{a}{2}x^2 + bx + c\,.$$

If $a > 0$, the minimum point of p is $-b/a$. At iteration k, the polynomial p is uniquely determined by

$$\begin{aligned} p(x_{k-2}) &= f(x_{k-2})\,, \\ p(x_{k-1}) &= f(x_{k-1})\,, \\ p(x_k) &= f(x_k)\,. \end{aligned} \tag{6.9}$$

Example 6.17. We illustrate the method by minimizing the function

$$f(x) = x^2 + 3e^{-x}$$

(Example 6.10) with the initial values $x_0 = 0.5, x_1 = 0.7, x_2 = 1$.

For $k = 2$, condition (6.9) results in the system of linear equations

$$\begin{aligned} p(0.5) &= f(0.5)\,, \\ p(0.7) &= f(0.7)\,, \\ p(1) &= f(1)\,, \end{aligned}$$

equivalent to

$$\begin{aligned} 0.125a + 0.5b + c &\approx 2.070\,, \\ 0.245a + 0.7b + c &\approx 1.980\,, \\ 0.5a + b + c &\approx 2.104\,. \end{aligned}$$

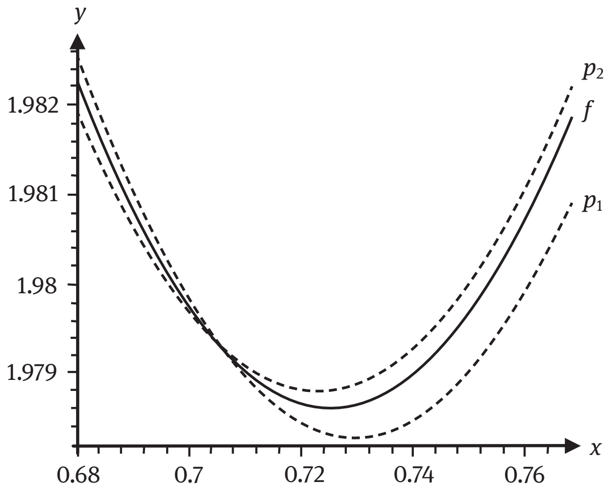


Fig. 6.7. Interpolation polynomials of Example 6.17.

The solution is given by $a \approx 3.453$, $b \approx -2.522$, $c \approx 2.899$, i.e. the first interpolation polynomial is $p_1(x) \approx 1.726x^2 - 2.522x + 2.899$ with minimum point $x_3 = -b/a \approx 0.7304$.

In the next iteration we solve (6.9) for $k = 3$, resulting in the system

$$\begin{aligned} 0.245a + \quad 0.7b + c &\approx 1.980\,, \\ 0.5a + \quad\quad b + c &\approx 2.104\,, \\ 0.2667a + 0.7304b + c &\approx 1.979 \end{aligned}$$

with the solution $a \approx 3.277$, $b \approx -2.372$, $c \approx 2.837$. The second interpolation polynomial is $p_2(x) \approx 1.638x^2 - 2.372x + 2.837$ with minimum point $x_4 = -b/a \approx 0.7238$ (Figure 6.7). The point x_4 is already "close" to the minimum point 0.726 of f (see Example 6.10).

There exist multiple possible variations to construct the polynomial p. One can require three other independent conditions. Thus for example, instead of (6.9) one can choose

$$\begin{aligned} p(x_{k-1}) &= f(x_{k-1}) \\ p'(x_{k-1}) &= f'(x_{k-1}) \\ p(x_k) &= f(x_k) \end{aligned} \tag{6.10}$$

or

$$\begin{aligned} p(x_k) &= f(x_k) \\ p'(x_k) &= f'(x_k) \\ p''(x_k) &= f''(x_k)\,. \end{aligned} \tag{6.11}$$

In these cases one needs only two starting points x_0 and x_1 or one starting point x_0, respectively.

Exercise 6.18. *Solve the problem of Example 6.17 with (6.10) instead of (6.9). Perform two iterations.*

Exercise 6.19. *Show that the quadratic interpolation is identical to Newton's method, if the polynomial p is constructed according to (6.11).*

The cubic interpolation works similarly. Then the polynomial $p(x)$ is of the form

$$p(x) = \frac{a}{3}x^3 + \frac{b}{2}x^2 + cx + d$$

and for determining it, one needs four independent conditions, for example,

$$\begin{aligned} p(x_k) &= f(x_k)\,, \\ p'(x_k) &= f'(x_k)\,, \\ p''(x_k) &= f''(x_k)\,, \\ p'''(x_k) &= f'''(x_k)\,. \end{aligned} \tag{6.12}$$

or

$$\begin{aligned} p(x_{k-1}) &= f(x_{k-1})\,, \\ p'(x_{k-1}) &= f'(x_{k-1})\,, \\ p(x_k) &= f(x_k)\,, \\ p'(x_k) &= f'(x_k)\,. \end{aligned} \tag{6.13}$$

The polynomial p, defined by (6.12) corresponds to the third-order Taylor approximation of the function f at point x_k, i.e. $p(x) = \sum_{i=0}^{3} \frac{f^{(i)}(x_k)}{i!}(x-x_k)^i$. The conditions (6.13) define the so-called *Hermite interpolation polynomial* (of degree ≤ 3). While the Taylor polynomial approximates f well in a neighborhood of x_k, the idea underlying conditions (6.13) is to construct a polynomial that approximates f well on the whole interval between the points x_{k-1} and x_k.

Example 6.20. Consider the convex function (Figure 6.8)

$$f(x) = x^2 + \frac{1}{x}, \quad x > 0$$

with minimum point $\bar{x} = 1/2^{1/3} \approx 0.7937$.

(a) Let $x_0 = 0.75$. For $k = 0$ the system (6.12) becomes

$$\begin{aligned} \frac{0.75^3}{3}a + \frac{0.75^2}{2}b + 0.75c + d &= f(0.75) &&\approx 1.896\,, \\ 0.75^2 a + 0.75b + c &= f'(0.75) &&\approx -0.2778\,, \\ 1.5a + b &= f''(0.75) &&\approx 6.741\,, \\ 2a &= f'''(0.75) &&\approx -18.96\,, \end{aligned}$$

and has the solution $a \approx -9.481$, $b \approx 20.96$, $c \approx -10.67$, $d \approx 5.333$, resulting in the Taylor approximation

$$p(x) \approx -3.160x^3 + 10.48x^2 - 10.67x + 5.333. \tag{6.14}$$

The minimum point of p over the interval [0, 1] is $x_1 \approx 0.7947$, yielding a good approximation for $\overline{x}$. For $k = 1$, the minimum point of the polynomial defined by (6.12) approximates $\overline{x}$ even better (the verification is left to the reader).

(b) Let $x_0 = 0.6$ and $x_1 = 1$. By using (6.13), we obtain for $k = 1$:

$$\begin{aligned}
\frac{0.6^3}{3}a + \frac{0.6^2}{2}b + 0.6c + d &= f(0.6) \approx 2.027\,,\\
0.6^2 a + 0.6b + c &= f'(0,6) \approx -1.578\,,\\
\frac{1}{3}a + \frac{1}{2}b + c + d &= f(1) = 2\,,\\
a + b + c &= f'(1) = 1\,.
\end{aligned}$$

The solution is $a \approx -8.333$, $b \approx 19.78$, $c \approx -10.44$, $d \approx 5.333$, yielding the first Hermite interpolation polynomial

$$p(x) \approx -2.778x^3 + 9.889x^2 - 10.44x + 5.333 \tag{6.15}$$

with minimum point $x_2 \approx 0.7925$. The polynomials (6.14) and (6.15) (interrupted lines in Figure 6.8) are almost coincident over the interval $[0, 1]$ and approximate f well over $[0.6, 1]$.

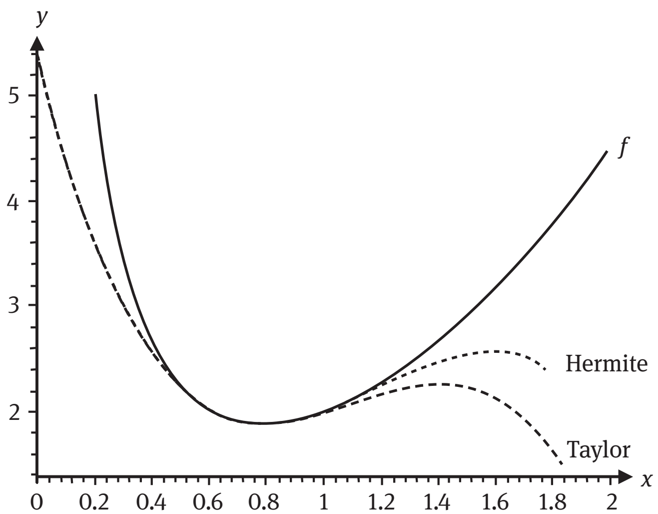


Fig. 6.8. Interpolation polynomials of Example 6.20.

Exercise 6.21. *Determine the first Hermite interpolation polynomial for the function*

$$f(x) = \frac{1}{x^2} + e^x, \quad x > 0$$

for $x_0 = 0.5$ *and* $x_1 = 1$. *Illustrate graphically!*

6.4 On the use of the methods in practice

Using the delimitation algorithm I and the Golden Section algorithm is recommended when a function is not differentiable or when the derivatives are complicated. Otherwise, the delimitation algorithm II and the bisection method should be preferred. An advantage of the Golden Section and bisection algorithms is that the number of required iterations may be estimated in advance (see Exercises 6.6 and 6.7).

All methods are robust and stable, except the Golden Section algorithm. In the latter method, rounding errors may cause that the relations $w_{k+1} = v_k$ or $v_{k+1} = w_k$ are not valid with sufficient accuracy, i.e. the minimum point may "disappear" from the calculated intervals. Disadvantages of Newton's method are that the first and second derivatives must be calculated for each iteration and that it does not converge for all starting values. However, Newton's method has the highest convergence speed among the introduced procedures. In practice it is recommended to start with the methods of Section 6.1 and to continue with Newton's method, when the convergence of the latter is guaranteed.

The procedures discussed in this chapter are also used to minimize functions that are not unimodal. As we will see in the next chapter, the one-dimensional minimization is an important part of multidimensional optimization. In many practical cases an approximate solution of the one-dimensional problem is sufficient (see Section 7.6).

7 Unrestricted minimization

In this chapter we minimize an unrestricted function with n variables, i.e. we solve the problem

$$\min_{x\in\mathbb{R}^n} f(x)\,. \tag{7.1}$$

After the study of quadratic functions (Section 7.1), which play an important role in the development of algorithms, we present diverse iterative methods to resolve (7.1). Except trust region methods (Section 7.7), all methods of this chapter have in common that an iteration consists of three steps of the following type:

Step 1:
For a given point x^k, a strict descent direction $d^k \in \mathbb{R}^n$ of f at point x^k is determined (see Definition 4.7 and the subsequent remarks).

Step 2:
Determine a $\lambda_k \in \mathbb{R}$ such that

$$f(x^k + \lambda_k d^k) < f(x^k)$$

(usually $\lambda_k > 0$).

Step 3:
Set $x^{k+1} = x^k + \lambda_k d^k$.

The value λ_k is called the *step size*, and its determination is called *line search* (*in the direction d^k*). When λ_k is determined by resolving the unidimensional problem

$$\min_{\lambda\in\mathbb{R}_+} f(x^k + \lambda d^k)\,, \tag{7.2}$$

we talk about *exact line search*.

7.1 Analysis of quadratic functions

Various iterative optimization methods make use of the fact that a function $f : \mathbb{R}^n \to \mathbb{R}$ can often be well approximated by a quadratic function

$$q(s) = \frac{1}{2}s^TQs + c^Ts + d\,. \tag{7.3}$$

For example, if f is twice continuously differentiable in a neighborhood of a point $\overline{x} \in \mathbb{R}^n$, then the second-order Taylor approximation

$$f(\overline{x} + s) \approx \frac{1}{2}s^THf(\overline{x})s + (\operatorname{grad} f(\overline{x}))^Ts + f(\overline{x}) \text{ for } s \in U_\varepsilon(0) \tag{7.4}$$

is of the form (7.3) with $Q := Hf(\overline{x})$, $c := \operatorname{grad} f(\overline{x})$ and $d := f(\overline{x})$. In a neighborhood of $\overline{x}$, the function (7.3) can be considered as a “model” of the “complicated” function f and serves, e.g. to approximate the minimum point of f (see the beginning of Section 7.3) or to evaluate the function f approximately in a certain region (see Section 7.7).

We first recall some results of linear algebra.

Definition 7.1. A matrix $M \in \mathbb{R}^{n\times n}$ is called *orthogonal*, if

$$M^T M = I\,, \tag{7.5}$$

i.e. the columns of M are normalized and mutually orthogonal.

Obviously (7.5) is equivalent to $M^T = M^{-1}$, and when M is orthogonal, the same holds for M^T. From (7.5) it follows that $\det(M^T M) = (\det(M))^2 = \det(I) = 1$, i.e. the determinant of an orthogonal matrix is $+1$ or -1. In these two cases the matrix is called *proper* or *improper orthogonal*, respectively.

For an orthogonal matrix M, the coordinate transformation

$$x \mapsto \tilde{x} := Mx \tag{7.6}$$

represents a rotation of the axis for $\det(M) = 1$ and a rotation followed by a reflection for $\det(M) = -1$. In both cases the new coordinate axes are aligned with the column vectors of M.

Example 7.2. It can be easily verified that the matrices

$$M_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad M_2 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 & 3 \\ 3 & -1 \end{pmatrix}$$

are proper and improper orthogonal, respectively. The coordinate transformation is illustrated in Figure 7.1. In case (a), the coordinates x_1 and x_2 are transformed into $\tilde{x}_1$ and $\tilde{x}_2$ by means of a rotation, and in case (b) we have a rotation followed by a reflection.

Theorem 7.3 (Spectral Theorem). *Let $Q \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then Q has n real eigenvalues $\lambda_1, \dots, \lambda_n$ (not necessarily different, see the end of Chapter 2) and there exist associated eigenvectors $y^1, \dots, y^n$, such that the matrix $Y = (y^1, \dots, y^n)$ is orthogonal. The matrix Q is diagonalizable, i.e. satisfies*

$$Q = YDY^T\,, \tag{7.7}$$

where D is the diagonal matrix defined by

$$D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

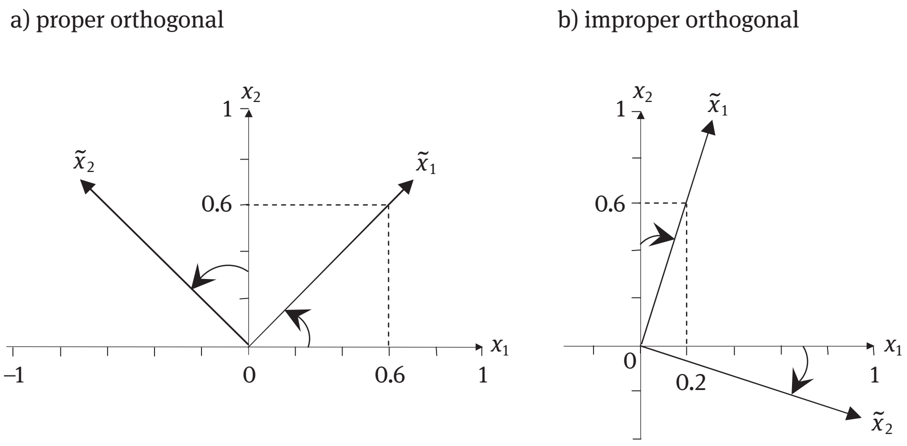


Fig. 7.1. Orthogonal coordinate transformations.

Example 7.4. For

$$Q = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}$$

the characteristic polynomial is

$$p(\lambda) = \begin{vmatrix} 2-\lambda & 2 \\ 2 & 5-\lambda \end{vmatrix} = (2-\lambda)(5-\lambda) - 4 = \lambda^2 - 7\lambda + 6 = (\lambda - 1)(\lambda - 6)\,,$$

thus $\lambda_1 = 1$ and $\lambda_2 = 6$. An eigenvector $y = (\alpha, \beta)^T$, associated with λ_1, satisfies

$$(Q - \lambda_1 I)y = 0 \Leftrightarrow \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = 0\,,$$

hence $y = (\alpha, -\alpha/2)^T$. Setting $\alpha = 2/\sqrt{5}$ yields the normalized eigenvector

$$y^1 = (2/\sqrt{5}, -1/\sqrt{5})^T\,.$$

In the same way we obtain the normalized eigenvector $y^2 = (1/\sqrt{5}, 2/\sqrt{5})^T$, associated with λ_2. It is easily verified that $y^{1T}y^2 = 0$ and

$$Y = (y^1, y^2) = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}\,.$$

The diagonalization of the matrix Q is

$$Q = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix} = YDY^T = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}\frac{1}{\sqrt{5}}\begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}\,.$$

Example 7.5. For

$$Q = \frac{1}{6}\begin{pmatrix} 7 & -1 & 2 \\ -1 & 7 & -2 \\ 2 & -2 & 10 \end{pmatrix}$$

we get $p(\lambda) = -\lambda^3 + 4\lambda^2 - 5\lambda + 2 = -(\lambda - 1)^2(\lambda - 2)$ and the eigenvalues are $\lambda_1 = \lambda_2 = 1$ and $\lambda_3 = 2$. In order to obtain an eigenvector associated to the eigenvalue 1 (with multiplicity 2) the system

$$(Q - 1I)y = 0 \Leftrightarrow (6Q - 6I)y = 0 \Leftrightarrow \begin{pmatrix} 1 & -1 & 2 \\ -1 & 1 & -2 \\ 2 & -2 & 4 \end{pmatrix} y = 0$$

must be solved. Since the last matrix has rank 1, the solution set is $S = \{y \in \mathbb{R}^3 | y_3 = \frac{y_2 - y_1}{2}, y_1 = y_2 \in \mathbb{R}\}$. We can choose $y^1 \in S$ arbitrarily such that $|y^1| = 1$, for example $y^1 = (1/\sqrt{2})(1, 1, 0)^T$. A vector chosen from S that is orthogonal to y^1, has the form $(-\alpha, \alpha, \alpha)^T$, yielding the normalized eigenvector $y^2 = (1/\sqrt{3})(-1, 1, 1)^T$ for $\alpha = 1/\sqrt{3}$.

An eigenvector associated to $\lambda_3 = 2$ satisfies the system

$$(Q - 2I)y = 0 \Leftrightarrow (6Q - 12I)y = 0 \Leftrightarrow \begin{pmatrix} -5 & -1 & 2 \\ -1 & -5 & -2 \\ 2 & -2 & -2 \end{pmatrix} y = 0 .$$

Any solution has the form $(\alpha, -\alpha, 2\alpha)^T$, yielding the normalized eigenvector

$$y^3 = 1/\sqrt{6}(1, -1, 2)^T \quad \text{for } \alpha = 1/\sqrt{6} .$$

Hence, the matrix Y is given by

$$Y = (y^1, y^2, y^3) = \frac{1}{\sqrt{6}}\begin{pmatrix} \sqrt{3} & -\sqrt{2} & 1 \\ \sqrt{3} & \sqrt{2} & -1 \\ 0 & \sqrt{2} & 2 \end{pmatrix},$$

resulting in the diagonalization

$$\begin{aligned} Q &= \frac{1}{6}\begin{pmatrix} 7 & -1 & 2 \\ -1 & 7 & -2 \\ 2 & -2 & 10 \end{pmatrix} = YDY^T \\ &= \frac{1}{\sqrt{6}}\begin{pmatrix} \sqrt{3} & -\sqrt{2} & 1 \\ \sqrt{3} & \sqrt{2} & -1 \\ 0 & \sqrt{2} & 2 \end{pmatrix}\begin{pmatrix} 1 & & \\ & 1 & \\ & & 2 \end{pmatrix}\frac{1}{\sqrt{6}}\begin{pmatrix} \sqrt{3} & \sqrt{3} & 0 \\ -\sqrt{2} & \sqrt{2} & \sqrt{2} \\ 1 & -1 & 2 \end{pmatrix} . \end{aligned}$$

The following concept is useful for the geometric illustration of functions.

Definition 7.6. For a function $f : \mathbb{R}^n \mapsto \mathbb{R}$, the set $\{x \in \mathbb{R}^n | f(x) = z\}$ is called the *level set* z. In particular, for $n = 2$ and $n = 3$ we call this set a *level curve* or a *level surface*, respectively.

We are now able to analyze a pure quadratic function $q(x) = x^T Qx$. We consider the coordinate transformation

$$\widetilde{x} := Y^T x\,, \tag{7.8}$$

where Y is the orthogonal matrix of (7.7). Without loss of generality we may assume that $\det(Y) = 1$, i.e. (7.8) represents a rotation. (Otherwise, i.e. in the case $\det(Y) = -1$, we can simply substitute one of the eigenvectors y^i for $-y^i$, yielding $\det(Y) = 1$). Now (7.7) implies

$$x^T Qx = x^T YDY^T x = \widetilde{x}^T D\widetilde{x} = \lambda_1 \widetilde{x}_1^2 + \cdots + \lambda_n \widetilde{x}_n^2\,. \tag{7.9}$$

If Q is positive definite, all eigenvalues λ_i are positive (Theorem 2.29 (iii)), and for all $z > 0$ the solution set of

$$x^T Qx = z \Leftrightarrow \lambda_1 \widetilde{x}_1^2 + \cdots + \lambda_n \widetilde{x}_n^2 = z \tag{7.10}$$

is an *n-dimensional ellipsoid*, the principal axes of which are aligned with the eigenvectors y^i. In particular, for $n = 2$ these sets are ellipses. By setting $r_i := \sqrt{z/\lambda_i}$, (7.10) becomes the standard form of the ellipsoid

$$\frac{\widetilde{x}_1^2}{r_1^2} + \cdots + \frac{\widetilde{x}_n^2}{r_n^2} = 1\,, \tag{7.11}$$

where r_i is the *ith radius*, i.e. half length of the ith principal axis.

Figures 7.2 and 7.3 illustrate level sets for the function $q(x) = x^T Qx$, when Q is chosen as in Examples 7.4 or 7.5, respectively. In the second case, Q has an eigenvalue with multiplicity 2, therefore the ellipsoid in Figure 7.3 has equal radii r_1 and r_2 for a given level z and is called a *spheroid*.

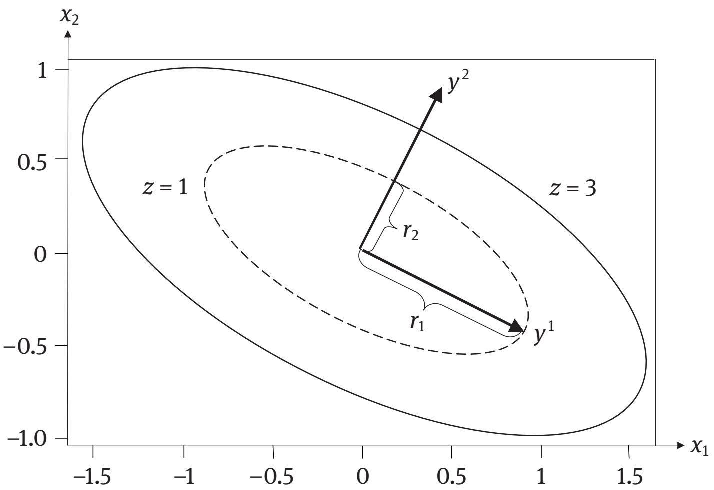


Fig. 7.2. Level curves of the function $2x_1^2 + 4x_1x_2 + 5x_2^2$.

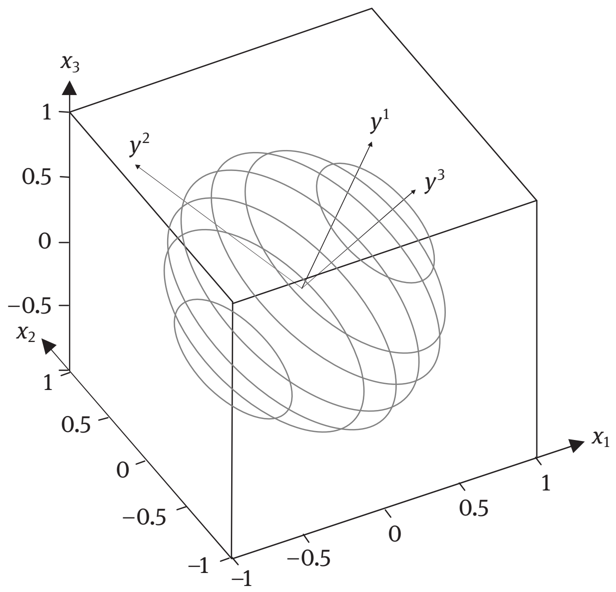


Fig. 7.3. Level surface $z = 6$ for the function $7x_1^2 - 2x_1x_2 + 4x_1x_3 + 7x_2^2 - 4x_2x_3 + 10x_3^2$.

Exercise 7.7. *Illustrate the level curves of the function x^TQx with*

$$Q = \begin{pmatrix} 3 & \sqrt{2} \\ \sqrt{2} & 4 \end{pmatrix}.$$

Determine the principal axis of the ellipses and the radii for $z = 1$.

We finally consider the general quadratic function

$$q(x) = \frac{1}{2}x^TQx + c^Tx\,. \tag{7.12}$$

If Q is symmetric and positive definite, this function is strictly convex and the minimum point x^* satisfies grad $q(x^*) = Qx^* + c = 0$, i.e.

$$x^* = -Q^{-1}c\,. \tag{7.13}$$

By shifting the coordinate system to the minimum point, i.e. by setting $\overline{x} = x - x^* = x + Q^{-1}c$, the linear term c^Tx in (7.12) is eliminated. We obtain the relation

$$\frac{1}{2}x^TQx + c^Tx = \frac{1}{2}\overline{x}^TQ\overline{x} - \frac{1}{2}c^TQ^{-1}c\,, \tag{7.14}$$

which can easily be verified by the reader.

Exercise 7.8. *Add the linear term $4x_1 + 7x_2$ to the function of the Exercise 7.7. Illustrate the displaced ellipses and determine their center.*

Exercise 7.9. *Illustrate the level curves of a quadratic function with positive semidefinite and indefinite Hessian matrix, respectively. The curves are ellipses?*

7.2 The gradient method

Obviously we get a direction of strict descent $d^k \in \mathbb{R}^n$ of f at point x^k, choosing

$$d^k = -\operatorname{grad} f(x^k) \neq 0$$

(see (4.14)). This vector can be interpreted as the *direction of maximum descent*. In order to justify this, we consider the expression $g(y) = \frac{1}{|y|} Df(x^k, y)$ (see Definition 3.40 and the subsequent remarks) which can be considered as a measure of ascent/descent of the function f at point x^k in the direction $y (y \neq 0)$.

By using the first-order Taylor approximation

$$f(x^k + ty) = f(x^k) + ty^T \operatorname{grad} f(x^k) + R_1(t)\,,$$

where the error term $R_1(t)$ satisfies $\lim_{t\to 0} \frac{R_1(t)}{t} = 0$, we obtain

$$\begin{aligned} g(y) &= \frac{1}{|y|} \lim_{t\to 0} \frac{f(x^k + ty) - f(x^k)}{t} = \frac{1}{|y|} \lim_{t\to 0} \frac{ty^T \operatorname{grad} f(x^k) + R_1(t)}{t} \\ &= \frac{1}{|y|} y^T \operatorname{grad} f(x^k) = \frac{1}{|y|} |y|\, |\operatorname{grad} f(x^k)| \cos\alpha = |\operatorname{grad} f(x^k)| \cos\alpha\,, \end{aligned}$$

where α is the angle between the vectors y and $\operatorname{grad} f(x_k)$. The expression $g(y)$ is minimal for $\alpha = \pi$, i.e. for $y = -\operatorname{grad} f(x_k)$.

If λ_k is determined by exact line search, we obtain the classical *gradient method*:

(1) Choose $x^0 \in \mathbb{R}^n$ and set $k = 0$.
(2) Compute $d^k = -\operatorname{grad} f(x^k)$.
(3) If $d^k = 0$ (or $|d^k| \le \varepsilon$), stop. Otherwise, go to step (4).
(4) Determine a solution λ_k of the problem $\min_{\lambda\in\mathbb{R}_+} f(x^k + \lambda d^k)$.
(5) Set $x^{k+1} = x^k + \lambda_k d^k$, $k = k+1$ and go to step (2).

We illustrate the procedure by means of the following two examples. The problems can be easily solved in a straight forward way with the help of Section 3.5, but here they serve to illustrate the operation of the described method.

Example 7.10. We perform the first iteration to solve the problem

$$\min_{x\in\mathbb{R}^2} x_1^2 + e^{x_2} - x_2$$

starting with $x^0 = (2, 1)^T$.

Since $\operatorname{grad} f(x_1, x_2) = \binom{2x_1}{e^{x_2}-1}$, we obtain $\operatorname{grad} f(x^0) = \binom{4}{e-1}$ and $d^0 = \binom{-4}{1-e}$. We now have to minimize the one-dimensional function in λ:

$$f(x^0 + \lambda d^0) = (2 - 4\lambda)^2 + e^{1+\lambda(1-e)} - 1 - \lambda(1 - e)\,.$$

Using, for example, Newton's method of Section 6.2 we obtain in three-digit arithmetic the minimum point $\lambda_0 \approx 0.507$. Thus,

$$x^1 = x^0 + \lambda_0 d^0 \approx \begin{pmatrix} 2 \\ 1 \end{pmatrix} + 0.507 \begin{pmatrix} -4 \\ -1.72 \end{pmatrix} \approx \begin{pmatrix} -0.0280 \\ 0.128 \end{pmatrix}.$$

This point is already "close" to the minimum point $x^* = (0,0)^T$.

An important special case is that of minimizing a quadratic function.

$$f(x) = \frac{1}{2} x^T Q x + c^T x \tag{7.15}$$

with a positive definite matrix Q (see Section 7.1). Then the result of the line search can be expressed explicitly as:

$$\lambda_k = \frac{g^{kT} g^k}{g^{kT} Q g^k}, \tag{7.16}$$

where $g^k := \operatorname{grad} f(x^k)$ is assumed to be nonzero. Since $\operatorname{grad} f(x) = Qx + c$, we obtain:

$$\begin{aligned} \frac{d}{d\lambda} f(x^k - \lambda g^k) &= (\operatorname{grad} f(x^k - \lambda g^k))^T (-g^k) = -(Q(x^k - \lambda g^k) + c)^T g^k \\ &= -(g^k - \lambda Q g^k)^T g^k = -g^{kT} g^k + \lambda g^{kT} Q g^k . \end{aligned}$$

The last expression is zero when λ is given by (7.16).

Example 7.11. We apply the method to the problem

$$\min_{x \in \mathbb{R}^2} (x_1 - 1)^2 + 2(x_2 - 2)^2 ,$$

starting with $x^0 = (0,3)^T$. Except for an additive constant which does not influence the minimization, this problem is of the form (7.15) with

$$Q = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \quad \text{and} \quad c = \begin{pmatrix} -2 \\ -8 \end{pmatrix} .$$

Since

$$\operatorname{grad} f(x_1, x_2) = \begin{pmatrix} 2(x_1 - 1) \\ 4(x_2 - 2) \end{pmatrix},$$

we obtain

$$g^0 = \operatorname{grad} f(x^0) = \begin{pmatrix} -2 \\ 4 \end{pmatrix}$$

and

$$d^0 = \begin{pmatrix} 2 \\ -4 \end{pmatrix}.$$

Formula (7.16) results in

$$\lambda_0 = \frac{(-2,4) \begin{pmatrix} -2 \\ 4 \end{pmatrix}}{(-2,4) \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} -2 \\ 4 \end{pmatrix}} = 5/18 .$$

Therefore

$$x^1 = x^0 + \lambda_0 d^0 = \begin{pmatrix} 0 \\ 3 \end{pmatrix} + \frac{5}{18} \begin{pmatrix} 2 \\ -4 \end{pmatrix} = \begin{pmatrix} 5/9 \\ 17/9 \end{pmatrix} .$$

The calculations of the first three iterations are summarized as follows:

Table 7.1. Gradient method.

k	x^k	$f(x^k)$	d^k	λ_k
0	$\binom{0}{3}$	3	$\binom{2}{-4}$	5/18
1	$\frac{1}{9}\binom{5}{17}$	2/9	$\frac{1}{9}\binom{8}{4}$	5/12
2	$\frac{1}{27}\binom{25}{56}$	$4/3^5$	$\frac{1}{27}\binom{4}{-8}$	5/18
3	$\frac{1}{243}\binom{235}{484}$	$8/3^8$	$\frac{1}{243}\binom{16}{8}$	5/12

The sequence of iteration points x^k is illustrated in Figure 7.4, showing the level curves of function f. Observe that two consecutive search directions d^k and d^{k+1} are orthogonal.

Exercise 7.12. *Show that the directions d^k and d^{k+1} are orthogonal in the general case.*

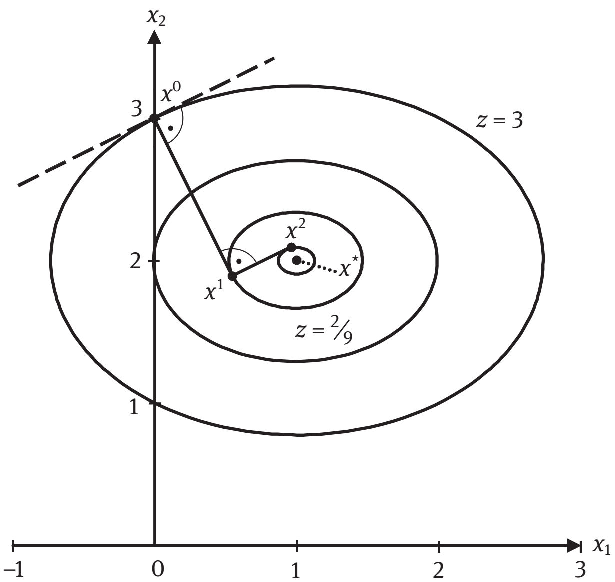


Fig. 7.4. Minimization process of the gradient method.

If the initial point x^0 is adequately chosen (see the following theorem), the gradient method converges to a point $x^* \in \mathbb{R}^n$ with $\operatorname{grad} f(x^*) = 0$. If the Hessian matrix $Hf(x^*)$ is positive definite, f is convex in a neighborhood of x^* and x^* is a local minimum point. Otherwise, x^* can be a saddle point, and in practice, another procedure must be applied for searching a point "close" to x^* with $f(\tilde{x}) < f(x^*)$. After this "perturbation" the gradient method can be applied again with $x^0 := \tilde{x}$.

The convergence of the gradient method is established in the following theorem.

Theorem 7.13. *Given the problem (7.1) with a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$. If the set*

$$C(x^0) := \{x \in \mathbb{R}^n | f(x) \le f(x^0)\} \tag{7.17}$$

is bounded, the gradient method converges to a point x^* with $\operatorname{grad} f(x^*) = 0$.

If $Hf(x^*)$ is positive definite with smallest and largest eigenvalues μ and ν, respectively, then the error sequence $r_k := f(x^k) - f(x^*)$ converges linearly, with rate not exceeding

$$\left(\frac{\nu - \mu}{\nu + \mu}\right)^2 < 1 . \tag{7.18}$$

Example 7.14. We apply the theorem to Examples 7.10 and 7.11.

(a) Since the objective function of Example 7.10 is convex, the set $C(x^0)$ is bounded. The convergence to the point $x^* = (0, 0)^T$ is assured since it is the only point in which the gradient is zero. The Hessian matrix

$$Hf(x^*) = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

has the eigenvalues $\mu = 1$ and $\nu = 2$. The theorem states that the error sequence converges linearly with rate not exceeding

$$\left(\frac{2-1}{2+1}\right)^2 = \frac{1}{9} .$$

We verify this for the first iteration:

$$r_0 = f(x^0) - f(x^*) = (4 + \mathrm{e} - 1) - 1 = 2 + \mathrm{e} \approx 4.72 ,$$
$$r_1 = f(x^1) - f(x^*) \approx (0.0280^2 + \mathrm{e}^{0.128} - 0.128) - 1 \approx 9.34 \cdot 10^{-3} ,$$

hence

$$\frac{r_1}{r_0} \approx 1.98 \cdot 10^{-3} \le \frac{1}{9} .$$

(b) Similar to part (a), the set $C(x^0)$ of Example 7.11 is bounded, therefore the convergence to the unique minimum point $x^* = (1, 2)^T$ is assured. The Hessian matrix at this point is

$$Hf(x^*) = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$$

with the eigenvalues $\mu = 2$ and $\nu = 4$. The theorem assures a linear convergence with rate not exceeding

$$\left(\frac{4-2}{4+2}\right)^2 = \frac{1}{9}.$$

In fact we have $r_k = f(x^k) - f(x^*) = f(x^k) = 3\left(\frac{2}{27}\right)^k$ (see Table 7.1). Therefore,

$$\frac{r_{k+1}}{r_k} = \frac{2}{27} \le \frac{1}{9}.$$

The gradient method is relatively stable. The one-dimensional problem in step (4) does not need a very accurate solution. If the value (7.18) is close to 1, the convergence speed may by insufficient for practical applications (see Example 5.1). However, it is possible to decrease this value with the aid of suitable coordinate transformations, which aim to reduce the difference between the eigenvalues μ and ν (see Exercise 7.17 and, e.g. Luenberger (2003, Section 7.7)).

There also exist several variants of the gradient method with constant or variable step size, which dispense the solution of a one-dimensional minimization problem (see Sections 7.6 and 11.3).

Exercise 7.15. *Given the problem*

$$\min_{x \in \mathbb{R}^2} x_1^2 + x_2^4 .$$

Perform the first iteration of the gradient method with $x^0 = (2, 1)^T$.

Exercise 7.16. *Apply the gradient method to the problem*

$$\min_{x \in \mathbb{R}^2} 2x_1^2 + 3x_2^2 .$$

Perform three iterations, starting with $x^0 = (3, 2)^T$ and using formula (7.16). Verify the statement about the convergence speed in Theorem 7.13.

Exercise 7.17. *Determine coordinate transformations such that the function $f(x_1, x_2)$ of Example 7.11 takes the form $f(y_1, y_2) = \frac{1}{2}(y_1^2 + y_2^2)$. What is the value of (7.18) for the transformed problem? Show that the gradient method solves the transformed problem for any starting point x^0 in a unique step.*

7.3 Multidimensional Newton's method

Newton's method to minimize one-dimensional functions (Section 6.2) can be generalized to solve the unconstrained problem (7.1). Essentially one needs only to replace the first and second derivatives $f'(x_k)$ and $f''(x_k)$ in (6.6) for the gradient $\operatorname{grad} f(x^k)$ and the Hessian matrix $Hf(x^k)$, respectively. This idea results in the recurrence formula

$$x^{k+1} = x^k - (Hf(x^k))^{-1} \operatorname{grad} f(x^k) \tag{7.19}$$

for minimizing a twice continuously differentiable function f. If $Hf(x^k)$ is positive definite, then x^{k+1} is the minimum point of the quadratic approximation

$$f(x^k) + \operatorname{grad} f(x^k)^T (x - x^k) + \frac{1}{2}(x - x^k)^T Hf(x^k)(x - x^k)$$

of the function f (see (6.7) and Section 7.1).

In this way we obtain the following algorithm to minimize multidimensional functions:

(1) Choose $x^0 \in \mathbb{R}^n$ and set $k = 0$.
(2) Calculate $g^k = \operatorname{grad} f(x^k)$.
(3) If $g^k = 0$ (or $|g^k| \le \varepsilon$), stop. Otherwise, go to step (4).
(4) Set $x^{k+1} = x^k - (Hf(x^k))^{-1} g^k$, $k = k + 1$ and go to step (2).

Example 7.18. We apply the method to the problem of Example 7.10. We have

$$\begin{aligned}
f(x_1, x_2) &= x_1^2 + \mathrm{e}^{x_2} - x_2, \\
\operatorname{grad} f(x_1, x_2) &= \begin{pmatrix} 2x_1 \\ \mathrm{e}^{x_2} - 1 \end{pmatrix}, \\
Hf(x_1, x_2) &= \begin{pmatrix} 2 & 0 \\ 0 & \mathrm{e}^{x_2} \end{pmatrix}, \\
(Hf(x_1, x_2))^{-1} &= \begin{pmatrix} 1/2 & 0 \\ 0 & \mathrm{e}^{-x_2} \end{pmatrix}.
\end{aligned}$$

Then the recurrence formula is

$$\begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \end{pmatrix} = \begin{pmatrix} x_1^k \\ x_2^k \end{pmatrix} - \begin{pmatrix} 1/2 & 0 \\ 0 & \mathrm{e}^{-x_2^k} \end{pmatrix} \begin{pmatrix} 2x_1^k \\ \mathrm{e}^{x_2^k} - 1 \end{pmatrix} = \begin{pmatrix} 0 \\ x_2^k - 1 + \mathrm{e}^{-x_2^k} \end{pmatrix},$$

where

$$x^k = \begin{pmatrix} x_1^k \\ x_2^k \end{pmatrix}.$$

For

$$x^0 = \begin{pmatrix} 2 \\ 1 \end{pmatrix},$$

the first two iterations in three-digit arithmetic result in:

$$x^1 = \begin{pmatrix} 0 \\ 1 - 1 + e^{-1} \end{pmatrix} = \begin{pmatrix} 0 \\ e^{-1} \end{pmatrix} \approx \begin{pmatrix} 0 \\ 0.368 \end{pmatrix},$$

$$x^2 = \begin{pmatrix} 0 \\ e^{-1} - 1 + e^{-e^{-1}} \end{pmatrix} \approx \begin{pmatrix} 0 \\ 0.0601 \end{pmatrix}.$$

Exercise 7.19. *Apply Newton's method to the problem of Exercise 7.15. Perform three iterations, starting with $x^0 = (2, 1)^T$.*

Exercise 7.20. *Verify that for each starting point x^0, Newton's method solves the problem of Example 7.11 in a single step.*

In practice it is not calculated the inverse of the Hessian matrix $(Hf(x^k))^{-1}$ in step (4). Instead of this, the point x^{k+1} is obtained as $x^{k+1} = x^k + d^k$, where d^k is the solution of the system of linear equations

$$Hf(x^k)d = -g^k\,. \tag{7.20}$$

As in the one-dimensional case, the procedure converges quadratically to the solution x^* of the system $\operatorname{grad} f(x) = 0$, if the starting point x^0 is "close enough" to x^*. However, the computational work per iteration, i.e. calculation of $Hf(x^k)$ and resolution of (7.20), may be considerable.

If the matrix $Hf(x^k)$ is positive definite for all k, Newton's method is a descent method with descent direction $d^k = -(Hf(x^k))^{-1} \operatorname{grad} f(x^k)$ and constant step size $\lambda_k = 1$. If λ_k is determined as in the previous section, we obtain *Newton's method with exact line search*:

(1) Choose $x^0 \in \mathbb{R}^n$ and set $k = 0$.
(2) Calculate $g^k = \operatorname{grad} f(x^k)$.
(3) If $g^k = 0$ (or $|g^k| \le \varepsilon$), stop. Otherwise, go to step (4).
(4) Calculate a solution d^k of the system $Hf(x^k)d = -g^k$.
(5) Determine a solution λ_k of the problem $\min_{\lambda \in \mathbb{R}_+} f(x^k + \lambda d^k)$.
(6) Set $x^{k+1} = x^k + \lambda_k d^k$, $k = k + 1$ and go to step (2).

The convergence of this procedure is established as follows (compare with Theorem 7.13):

Theorem 7.21. *Consider the problem (7.1) with a three times continuously differentiable function $f: \mathbb{R}^n \to \mathbb{R}$. Assume that the Hessian matrix $Hf(x)$ is positive definite in each point x, i.e. f is strictly convex. If*

$$C(x^0) := \{x \in \mathbb{R}^n \,|\, f(x) \le f(x^0)\}$$

is bounded, then Newton's method with exact line search converges quadratically to the unique global minimum point x^*.

Note that in a neighborhood of x^*, this method corresponds approximately to the common Newton's method.

Exercise 7.22. *Given the problem*

$$\min_{x \in \mathbb{R}^2} (x_1 - 2)^4 + (x_2 - 5)^4 .$$

(a) *Show that the recurrence formula of Newton's method can be written as*

$$x^{k+1} = \frac{1}{3}\left(2x^k + \begin{pmatrix} 2 \\ 5 \end{pmatrix}\right) \quad \textit{for } k = 0, 1, 2, \dots$$

(b) *Show that*

$$x^k = \left(1 - \left(\frac{2}{3}\right)^k\right)\begin{pmatrix} 2 \\ 5 \end{pmatrix}$$

if $x^0 = (0, 0)^T$.

(c) *Illustrate the points* $x^0, \dots, x^5$ *and the minimum point* $x^* = (2, 5)^T$ *graphically in the plane (see Figure 7.4).*

Exercise 7.23. *Show that Newton's method with exact line search solves the problem of Exercise 7.22 in a single step, if* $x^0 = (0, 0)^T$. *Is a calculation necessary to prove this fact? (Use your solution of the previous exercise).*

Newton's method can be further generalized, by substituting the Hessian matrix in step (4) by a symmetric positive definite matrix $B_k = Hf(x^k) + \sigma I$ with $\sigma > 0$. This results in the modified Newton method which cannot be fully addressed at this place.

7.4 Conjugate directions and quasi-Newton methods

As we have seen in the two previous sections, Newton's method converges usually much faster than the gradient method, but requires more computational work per iteration. While the former needs to calculate the Hessian matrix and to solve the system of linear equations (7.20), the latter has the advantage that it requires only the first partial derivatives of the objective function.

In the following we introduce two procedures which aim to combine the respective advantages of both methods. Since in a neighborhood of a minimum point, general functions can often be well approximated by quadratic functions, we restrict ourselves in the theoretical analysis to the problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T Q x + c^T x \tag{7.21}$$

(see Section 7.1), where $Q \in \mathbb{R}^{n \times n}$ is symmetrical and positive definite, and $c \in \mathbb{R}^n$.

The gradient method cannot solve (7.21) in a finite number of iterations, except in the case when all eigenvalues of Q are equal (see the transformed problem in Exercise 7.17). The convergence speed may be insufficient, when the value $(\frac{\nu-\mu}{\nu+\mu})^2$ is close to 1 (see Theorem 7.13). Newton's method solves (7.21) in a single step:

$$x^1 = x^0 - (Hf(x^0))^{-1} \operatorname{grad} f(x^0) = x^0 - Q^{-1}(Qx^0 + c) = -Q^{-1}c = x^*$$

(see Exercise 7.20), but therefore the system of linear equations (7.20) must be solved.

The procedures of this section aim to solve (7.21) in a *finite number of iterations*, using *only first partial derivatives* of the objective function.

Conjugate direction methods

The following definition generalizes the concept of orthogonality of vectors.

Definition 7.24. Let $Q \in \mathbb{R}^{n\times n}$ be a symmetrical matrix. The vectors $d^0, d^1, \ldots, d^m \in \mathbb{R}^n$ are called *Q-conjugate* (or *Q-orthogonal*), if

$$d^{iT} Qd^j = 0 \quad \text{for } i \neq j\,; \quad i, j = 0, 1, \ldots, m\,.$$

Obviously, Q-conjugate vectors are orthogonal for $Q = I$.

Proposition 7.25. *If the vectors $d^0, \ldots, d^{n-1} \in \mathbb{R}^n \setminus \{0\}$ are Q-conjugate for a positive definite matrix $Q \in \mathbb{R}^{n\times n}$, then these vectors are linearly independent.*

Proof. Let $\lambda_0, \ldots, \lambda_{n-1}$ be real numbers such that

$$\lambda_0 d^0 + \cdots + \lambda_{n-1} d^{n-1} = 0\,.$$

Then it holds for all $k = 0, \ldots, n-1$

$$d^{kT} Q(\lambda_0 d^0 + \cdots + \lambda_{n-1} d^{n-1}) = \lambda_k d^{kT} Qd^k = 0\,.$$

Since Q is positive definite, it holds $d^{kT} Qd^k > 0$ and therefore $\lambda_k = 0$. □

We now present the so-called conjugate direction theorem.

Theorem 7.26. *Let $f(x)$ be the objective function of problem (7.21), and let the vectors $d^0, \ldots, d^{n-1} \in \mathbb{R}^n \setminus \{0\}$ be Q-conjugate. Then the procedure*

$$x^0 \in \mathbb{R}^n\,, \quad x^{k+1} = x^k + \lambda_k d^k \quad \text{for } k = 0, 1, \ldots, n-1\,, \tag{7.22}$$

where $\lambda_k = -(d^{kT} g^k)/(d^{kT} Qd^k)$, $g^k = Qx^k + c$ (gradient of f at x^k) converges to the unique global minimum point x^* of $f(x)$ in n steps, i.e. $x^n = x^*$.

Proof. Since the vectors $d^0, \ldots, d^{n-1}$ are linearly independent, there exist uniquely determined real numbers $\lambda_0, \ldots, \lambda_{n-1}$ such that

$$x^* - x^0 = \lambda_0 d^0 + \cdots + \lambda_{n-1} d^{n-1}\,.$$

We prove the theorem by showing that the λ_i are of the form in (7.22). We first observe that for $k = 0, \dots, n-1$ holds

$$d^{kT}Q(x^* - x^0) = d^{kT}Q(\lambda_0 d^0 + \cdots + \lambda_{n-1}d^{n-1}) = \lambda_k d^{kT}Qd^k\,,$$

implying

$$\lambda_k = \frac{d^{kT}Q(x^* - x^0)}{d^{kT}Qd^k}\,. \tag{7.23}$$

From the recurrence relation it follows that

$$x^k - x^0 = \sum_{i=0}^{k-1} x^{i+1} - x^i = \lambda_0 d^0 + \cdots + \lambda_{k-1}d^{k-1}\,.$$

Multiplying by $d^{kT}Q$ yields

$$d^{kT}Q(x^k - x^0) = d^{kT}Q(\lambda_0 d^0 + \cdots + \lambda_{k-1}d^{k-1}) = 0\,. \tag{7.24}$$

By using (7.24) and the fact that $g^* = \operatorname{grad} f(x^*) = Qx^* + c = 0$, we can rewrite the numerator of (7.23) as follows:

$$d^{kT}Q(x^* - x^0) = d^{kT}Q(x^* - x^k + x^k - x^0) = d^{kT}Q(x^* - x^k) = -d^{kT}g^k\,. \qquad \square$$

A procedure of the type (7.22) is called a *conjugate direction method*.

Exercise 7.27. *Show that the value λ_k in (7.22) is a solution of the problem*

$$\min_{\lambda\in\mathbb{R}} f(x^k + \lambda d^k)\,.$$

Example 7.28. For

$$Q = \begin{pmatrix} 3 & 1 \\ 1 & 5 \end{pmatrix}$$

and

$$c = -\begin{pmatrix} 11 \\ 13 \end{pmatrix}\,,$$

the vectors $d^0 = (2,1)^T$ and $d^1 = (3,-3)^T$ are Q-conjugate. Applying the procedure (7.22) for $x^0 = (2,4)^T$, we obtain:

Iteration 1:

$$g^0 = \begin{pmatrix} 3 & 1 \\ 1 & 5 \end{pmatrix}\begin{pmatrix} 2 \\ 4 \end{pmatrix} - \begin{pmatrix} 11 \\ 13 \end{pmatrix} = \begin{pmatrix} -1 \\ 9 \end{pmatrix}.$$

$$\lambda_0 = -\frac{d^{0T}g^0}{d^{0T}Qd^0} = -\frac{(2,1)\begin{pmatrix} -1 \\ 9 \end{pmatrix}}{(2,1)\begin{pmatrix} 3 & 1 \\ 1 & 5 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix}} = -\frac{1}{3}.$$

Thus,

$$x^1 = x^0 + \lambda_0 d^0 = \begin{pmatrix} 4/3 \\ 11/3 \end{pmatrix}.$$

Iteration 2:

$$g^1 = \begin{pmatrix} 3 & 1 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 4/3 \\ 11/3 \end{pmatrix} - \begin{pmatrix} 11 \\ 13 \end{pmatrix} = \begin{pmatrix} -10/3 \\ 20/3 \end{pmatrix},$$

$$\lambda_1 = -\frac{(3,-3)\begin{pmatrix} -10/3 \\ 20/3 \end{pmatrix}}{(3,-3)\begin{pmatrix} 3 & 1 \\ 1 & 5 \end{pmatrix}\begin{pmatrix} 3 \\ -3 \end{pmatrix}} = \frac{5}{9},$$

$$x^2 = x^1 + \lambda_1 d^1 = \begin{pmatrix} 3 \\ 2 \end{pmatrix} = x^* .$$

The following *conjugate gradient method* of Fletcher and Reeves applies a procedure of the form (7.22) repeatedly to solve the general unrestricted problem (7.1). An initial search direction is defined by $d^0 = -\operatorname{grad} f(x^0)$, i.e. the first iteration corresponds to a step of the gradient method. The following search directions are recursively determined by

$$d^{k+1} = -\operatorname{grad} f(x^{k+1}) + \beta_k d^k ,$$

i.e. d^{k+1} is the sum of the negative gradient of f at x^{k+1} and a multiple of the previous direction d^k (see the steps (8) and (9) of the following algorithm). After a cycle of n iterations, a new cycle starts by setting $x^0 := x^n$.

It can be shown that the directions constructed in that way are Q-conjugate, if the algorithm is applied to problem (7.21), therefore the method solves a quadratic problem in n iterations (see Theorem 7.26).

Conjugate gradient method

(1) Choose $x^0 \in \mathbb{R}^n$, calculate $g^0 = \operatorname{grad} f(x^0)$.
(2) If $g^0 = 0$ (or $|g^0| \leq \varepsilon$), stop. Otherwise set $d^0 = -g^0$, $k = 0$ and go to step (3).
(3) Determine λ_k by means of exact line search.
(4) Set $x^{k+1} = x^k + \lambda_k d^k$.
(5) Calculate $g^{k+1} = \operatorname{grad} f(x^{k+1})$.
(6) If $g^{k+1} = 0$ (or $|g^{k+1}| \leq \varepsilon$), stop. Otherwise go to step (7).
(7) If $k < n-1$, go to step (8); if $k = n-1$, go to step (10).
(8) Calculate $\beta_k = \frac{g^{(k+1)T} g^{k+1}}{g^{kT} g^k}$.
(9) Set $d^{k+1} = -g^{k+1} + \beta_k d^k$, $k = k+1$ and go to step (3).
(10) Set $x^0 = x^n$, $d^0 = -g^n$, $k = 0$ and go to step (3).

Example 7.29. We apply the algorithm to the problem of Example 7.11, i.e. we solve (7.21) for

$$Q = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$$

and $c = -(2,8)^T$, choosing $x^0 = (0,3)^T$.

The calculations are summarized as follows:

Table 7.2. Conjugate gradient method (quadratic problem).

k	x^k	g^k	d^k	λ_k	β_k
0	$\binom{0}{3}$	$\binom{-2}{4}$	$\binom{2}{-4}$	5/18	4/81
1	$\binom{5/9}{17/9}$	$\binom{-8/9}{-4/9}$	$\binom{80/81}{20/81}$	9/20	
2	$\binom{1}{2}$				

The optimal solution is achieved after a cycle of $n = 2$ iterations and the constructed directions are Q-conjugate, i.e.

$$d^{0T} Q d^1 = (2, -4) \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 80/81 \\ 20/81 \end{pmatrix} = 0\,.$$

There exist various alternative formulas for the parameter β_k in step (8), e.g. the formulas

$$\beta_k^{PR} = \frac{g^{(k+1)T}(g^{k+1} - g^k)}{g^{kT} g^k} \tag{7.25}$$

of Polak-Ribière and

$$\beta_k^{HS} = \frac{g^{(k+1)T}(g^{k+1} - g^k)}{g^{kT}(g^{k+1} - g^k)} \tag{7.26}$$

of Hestenes and Stiefel.

Example 7.30. We now apply the method to the nonquadratic function

$$f(x) = (x_1 - 1)^4 + (2x_1 + x_2)^2\,.$$

Table 7.3 shows the calculations for the starting point $x^0 = (-2, 1)^T$. After every cycle of two iterations we set $x^0 := x^2$. The convergence to the optimal point $x^* = (1, -2)^T$ is illustrated in Figure 7.5.

Exercise 7.31. *Apply the above method to the quadratic problem (7.21) with*

$$Q = \begin{pmatrix} 3 & 4 \\ 4 & 6 \end{pmatrix}$$

and $c = -(36, 38)^T$, *starting with* $x^0 = (6, 3)^T$. *Verify that the constructed directions are Q-conjugate!*

Table 7.3. Conjugate gradient method (nonquadratic problem).

Cycle	k	x^k	g^k	d^k	λ_k	β_k
1	0	$\begin{pmatrix}-2\\1\end{pmatrix}$	$\begin{pmatrix}-120\\-6\end{pmatrix}$	$\begin{pmatrix}120\\6\end{pmatrix}$	0.01646	0.0003060
	1	$\begin{pmatrix}-0.02463\\1.099\end{pmatrix}$	$\begin{pmatrix}-0.1050\\2.099\end{pmatrix}$	$\begin{pmatrix}0.1417\\-2.097\end{pmatrix}$	0.6485	
2	0	$\begin{pmatrix}0.06724\\-0.2612\end{pmatrix}$	$\begin{pmatrix}-3.753\\-0.2535\end{pmatrix}$	$\begin{pmatrix}3.753\\0.2535\end{pmatrix}$	0.06029	0.03304
	1	$\begin{pmatrix}0.2935\\-0.2460\end{pmatrix}$	$\begin{pmatrix}-0.04608\\0.6822\end{pmatrix}$	$\begin{pmatrix}0.1701\\-0.6738\end{pmatrix}$	1.353	
3	0	$\begin{pmatrix}0.5236\\-1.157\end{pmatrix}$	$\begin{pmatrix}-0.8735\\-0.2205\end{pmatrix}$	$\begin{pmatrix}0.8735\\0.2205\end{pmatrix}$	0.08533	0.01742
	1	$\begin{pmatrix}0.5981\\-1.139\end{pmatrix}$	$\begin{pmatrix}-0.02910\\0.1153\end{pmatrix}$	$\begin{pmatrix}0.04431\\-0.1114\end{pmatrix}$	4.218	
4	0	$\begin{pmatrix}0.7850\\-1.609\end{pmatrix}$	$\begin{pmatrix}-0.1942\\-0.07720\end{pmatrix}$	$\begin{pmatrix}0.1942\\0.07720\end{pmatrix}$	0.09647	0.004221
	1	$\begin{pmatrix}0.8037\\-1.601\end{pmatrix}$	$\begin{pmatrix}-0.005016\\0.01262\end{pmatrix}$	$\begin{pmatrix}0.005836\\-0.01229\end{pmatrix}$	21.38	
5	0	$\begin{pmatrix}0.9285\\-1.864\end{pmatrix}$				

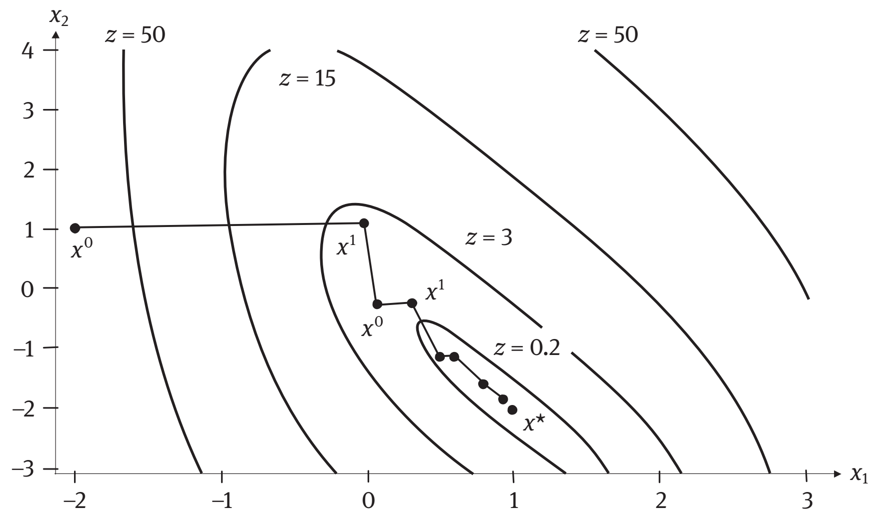


Fig. 7.5. Solution of Example 7.30 by the conjugate gradient method.

Quasi-Newton methods

A quasi-Newton method is a procedure of the form

$$x^0 \in \mathbb{R}^n, x^{k+1} = x^k + \lambda_k d^k \,, \tag{7.27}$$

where λ_k is determined by exact line search; the search direction is calculated as $d^k = -M_k g^k$, where $M_k \in \mathbb{R}^{n\times n}$ is a symmetric and positive definite matrix which must be updated in each step. The matrices M_k are constructed such that

$$M_n = Q^{-1} \tag{7.28}$$

is satisfied when solving a quadratic problem (7.21). It is desirable that the construction of these matrices requires the lowest possible computational work and uses only the first derivatives of the objective function. The fact that M_k is positive definite, ensures that d^k is a descent direction. Due to (7.28), n steps of a quasi-Newton method correspond approximately to one step of the Newton method with exact line search.

The following is a procedure of Davidon, Fletcher and Powell (called *DFP method* or *variable metric method*) which is both a quasi-Newton method as well as a method of conjugate directions. The basic structure is identical to that of the conjugate gradient method. The algorithm searches successively in directions $d^0, d^1, \ldots, d^{n-1}$, where $d^0 = -\operatorname{grad} f(x^0)$ and $d^k = -M_k g^k$ for $k \geq 1$. When a cycle of n iterations is completed, a new cycle begins, setting $x^0 := x^n$. The method generalizes the secant method for multidimensional functions, but this fact will not be further explored at this point.

DFP Method

(1) Choose $x^0 \in \mathbb{R}^n$, calculate $g^0 = \operatorname{grad} f(x^0)$.
(2) If $g^0 = 0$ (or $|g^0| \leq \varepsilon$), stop. Otherwise go to step (3).
(3) Set $M_0 = I$ ($n \times n$ unit matrix), $k = 0$.
(4) Set $d^k = -M_k g^k$.
(5) Determine λ_k by exact line search.
(6) Set $x^{k+1} = x^k + \lambda_k d^k$.
(7) Calculate $g^{k+1} = \operatorname{grad} f(x^{k+1})$.
(8) If $g^{k+1} = 0$ (or $|g^{k+1}| \leq \varepsilon$), stop. Otherwise go to step (9).
(9) If $k < n - 1$, go to step (10); if $k = n - 1$, go to step (12).
(10) Set $y^k = g^{k+1} - g^k, p^k = x^{k+1} - x^k$,

$$M_{k+1} = M_k + \frac{p^k p^{kT}}{p^{kT} y^k} - \frac{(M_k y^k)(M_k y^k)^T}{y^{kT} M_k y^k} .$$

(11) Set $k = k + 1$ and go to step (4).
(12) Set $x^0 = x^n, g^0 = g^n, M_0 = I, k = 0$ and go to step (4).

We call attention to the fact that the second line in step (10) involves a sum of matrices. While the denominators are scalars, the numerators are matrices. In particular, $p^k p^{kT}$ is the matrix with elements $a_{ij} = p_i^k p_j^k$, where $p^k = (p_1^k, \ldots, p_n^k)^T$.

The two methods of this section generate the same points x^k in the application to a quadratic problem. If $C(x^0) := \{x \in \mathbb{R}^n | f(x) \leq f(x^0)\}$ is bounded, both procedures converge superlinearly to a point x^* with $\operatorname{grad} f(x^*) = 0$.

Table 7.4. DFP method.

Cycle	k	x^k	M_k	d^k	λ_k
1	0	$\begin{pmatrix}-2\\1\end{pmatrix}$	I	$\begin{pmatrix}120\\6\end{pmatrix}$	0.01646
	1	$\begin{pmatrix}-0.02463\\1.099\end{pmatrix}$	$\begin{pmatrix}0.02096 & -0.06642\\-0.06642 & 0.9955\end{pmatrix}$	$\begin{pmatrix}0.1416\\-2.097\end{pmatrix}$	0.6487
2	0	$\begin{pmatrix}0.06724\\-0.2612\end{pmatrix}$	I	$\begin{pmatrix}3.753\\0.2535\end{pmatrix}$	0.06029
	1	$\begin{pmatrix}0.2935\\-0.2460\end{pmatrix}$	$\begin{pmatrix}0.1199 & -0.2332\\-0.2332 & 0.9404\end{pmatrix}$	$\begin{pmatrix}0.1646\\-0.6523\end{pmatrix}$	1.397
3	0	$\begin{pmatrix}0.5236\\-1.157\end{pmatrix}$	I	$\begin{pmatrix}0.8735\\0.2205\end{pmatrix}$	0.08533
	1	$\begin{pmatrix}0.5981\\-1.139\end{pmatrix}$	$\begin{pmatrix}0.2167 & -0.3231\\-0.3231 & 0.8686\end{pmatrix}$	$\begin{pmatrix}0.04355\\-0.1095\end{pmatrix}$	4.291
4	0	$\begin{pmatrix}0.7850\\-1.609\end{pmatrix}$	I	$\begin{pmatrix}0.1942\\0.7720\end{pmatrix}$	0.09647
	1	$\begin{pmatrix}0.8037\\-1.601\end{pmatrix}$	$\begin{pmatrix}0.2673 & -0.3544\\-0.3544 & 0.8292\end{pmatrix}$	$\begin{pmatrix}0.005811\\-0.01224\end{pmatrix}$	21.47
5	0	$\begin{pmatrix}0.9285\\-1.864\end{pmatrix}$			

Example 7.32. For the function

$$f(x) = (x_1 - 1)^4 + (2x_1 + x_2)^2$$

of the previous example, the method DFP performs the calculations in Table 7.4, when $x^0 = (-2, 1)^T$. In four-digit arithmetic, the iteration points and gradients are identical to those of Table 7.3.

Exercise 7.33. *Apply the DFP method to the quadratic problem of Exercise 7.31, starting with $x^0 = (6, 3)^T$. Compare with the calculations of that exercise. After the regular termination of the algorithm, realize once again the step (10), calculating the matrix M_2. Is the condition (7.28) satisfied?*

Exercise 7.34. *Minimize the function $f(x) = x_1^4 + x_2^4$ for $x^0 = (3, 5)^T$*
(a) *using the conjugate gradient method,*
(b) *using the DFP method.*

We close the section with some remarks on numerical aspects and the BFGS method.

While the gradient method is slow but stable, Newton's method is fast but causes extensive computational work per iteration. The application of the latter to functions with many variables is usually not practicable. Due to the fast convergence in the neighborhood of a solution, Newton's method may be used to minimize functions with a few variables and a "simple" Hessian matrix. Usually, it is recommended to apply a conjugate gradient or quasi-Newton method. But these methods are not stable with respect to errors in the calculation of the step size.

In extensive numerical tests with various quasi-Newton methods, good results have been observed when using a method suggested by Broyden, Fletcher, Goldfarb and Shannon (*BFGS method*, see Luenberger (2003, Section 9.4)). In order to make the procedure more stable, they substituted the formula in step (10) of the DFP method by

$$M_{k+1} = M_k + C_k - D_k$$

with

$$C_k = \left(1 + \frac{y^{kT} M_k y^k}{p^{kT} y^k}\right) \frac{p^k p^{kT}}{p^{kT} y^k}\,,$$

and

$$D_k = \frac{p^k (M_k y^k)^T + (M_k y^k) p^{kT}}{p^{kT} y^k}\,.$$

Exercise 7.35. *Solve the Exercise 7.33 once again, calculating the matrices M_k by the above formula.*

The computations may be more stable, calculating iteratively the inverses of the matrices M_k. It can be proved that these inverses, denoted by B_k, satisfy the recurrence relation

$$B_{k+1} = B_k + \frac{y^k y^{kT}}{y^{kT} p^k} - \frac{(B_k p^k)(B_k p^k)^T}{p^{kT} B_k p^k}\,. \tag{7.29}$$

In this case, the formula of step (10) of the DFP method is substituted by (7.29), and condition $d^k = -M_k g^k$ in step (4) is substituted by the equivalent system of linear equations $B_k d^k = -g^k$.

7.5 Cyclic coordinate search techniques

All procedures of this chapter, studied so far, need to calculate the gradient of the objective function at every iteration point. However, practical applications involve frequently the minimization of a nondifferentiable function f (see also Section 11.1). Moreover, even when f is differentiable, the calculation of the gradient may be cumbersome. Therefore, various methods have been proposed to solve problem (7.1), which do not require the derivatives of f. The simplest one is the following cyclic minimization method in directions of coordinate axes. This procedure is more robust than all of the previous methods, but it converges slowly.

Starting with an initial point x^0 in the first iteration, we minimize a function over the coordinate axis x_1, i.e. the one-dimensional problem

$$\min_{\lambda \in \mathbb{R}} f(x^0 + \lambda e^1)\,, \tag{7.30}$$

is solved, where e^i denotes the ith unit vector of $\mathbb{R}^n$ for $i = 1, \dots, n$.

Let λ_0 be a minimum point of (7.30), then x^1 is defined by $x^1 = x^0 + \lambda_0 e^1$. In the second iteration we minimize over the coordinate axis x_2, i.e. we solve the problem

$$\min_{\lambda \in \mathbb{R}} f(x^1 + \lambda e^2) \tag{7.31}$$

and the point x^2 is defined by $x^2 = x^1 + \lambda_1 e^2$, where λ_1 is a minimum point of (7.31), etc.

After n iterations a cycle is completed and a new cycle begins by minimizing successively over the coordinate axes $x_1, \dots, x_n$. The procedure ends when a whole cycle has not diminished the value of the objective function f sufficiently.

Example 7.36. Given is the problem

$$\min_{x \in \mathbb{R}^2} (1 + x_1^2)(2 + 3x_2^2)$$

with $x^0 = (2, 3)^T$.

Iteration 1:
We minimize the one-dimensional function in λ:

$$f(x^0 + \lambda e^1) = f(2 + \lambda, 3) = 29(1 + (2 + \lambda)^2)\,.$$

We get $\lambda_0 = -2$, thus

$$x^1 = x^0 + \lambda_0 e^1 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}.$$

Iteration 2:
The function to be minimized is:

$$f(x^1 + \lambda e^2) = f(0, 3 + \lambda) = 2 + 3(3 + \lambda)^2$$

yielding $\lambda_1 = -3$, therefore

$$x^2 = x^1 + \lambda_1 e^2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Since $f(x^2 + \lambda e^1) = 2(1 + \lambda^2)$ is minimized by $\lambda_2 = 0$, the following iteration does not diminish the value of f. The same holds for the fourth iteration. Therefore the method ends with the point $(0, 0)^T$, which is in fact the unique global minimum point.

Exercise 7.37. *Apply the method to the problem*

$$\min_{x \in \mathbb{R}^2} |x_1| + 2x_1^2 + |x_2 - 1|$$

starting with $x^0 = (4, 2)^T$.

The following problem cannot be solved in a single cycle of iterations.

Example 7.38. We consider the problem

$$\min_{x \in \mathbb{R}^2} (x_1 - 1)^2 + (x_1 - x_2)^2$$

with $x^0 = (3, 2)^T$.

Iteration 1:
The function

$$f(x^0 + \lambda e^1) = f(3 + \lambda, 2) = (2 + \lambda)^2 + (1 + \lambda)^2$$

has the minimum point $\lambda_0 = -3/2$. Therefore

$$x^1 = x^0 + \lambda_0 e^1 = \begin{pmatrix} 3/2 \\ 2 \end{pmatrix} .$$

Iteration 2:
The minimum point of

$$f(x^1 + \lambda e^2) = f(3/2, 2 + \lambda) = \left(\frac{1}{2}\right)^2 + (-\lambda - \frac{1}{2})^2$$

is $\lambda_1 = -1/2$, thus

$$x^2 = x^1 + \lambda_1 e^2 = \begin{pmatrix} 3/2 \\ 3/2 \end{pmatrix} .$$

The first six iterations are summarized as follows:

Table 7.5. Cyclic minimization.

k	x^k	$f(x^k)$	λ_k
0	$\binom{3}{2}$	5	$-3/2$
1	$\binom{3/2}{2}$	1/2	$-1/2$
2	$\binom{3/2}{3/2}$	1/4	$-1/4$
3	$\binom{5/4}{3/2}$	1/8	$-1/4$
4	$\binom{5/4}{5/4}$	1/16	$-1/8$
5	$\binom{9/8}{5/4}$	1/32	$-1/8$
6	$\binom{9/8}{9/8}$	1/64	

The convergence of the points x^i to the minimum point $x^* = (1, 1)^T$ is illustrated graphically in Figure 7.6.

Exercise 7.39. *Apply the method to the problem*

$$\min_{x \in \mathbb{R}^2} (x_1 - 2)^2 + (3x_1 - x_2)^2$$

with $x^0 = (3, 5)^T$. *Realize six iterations!*

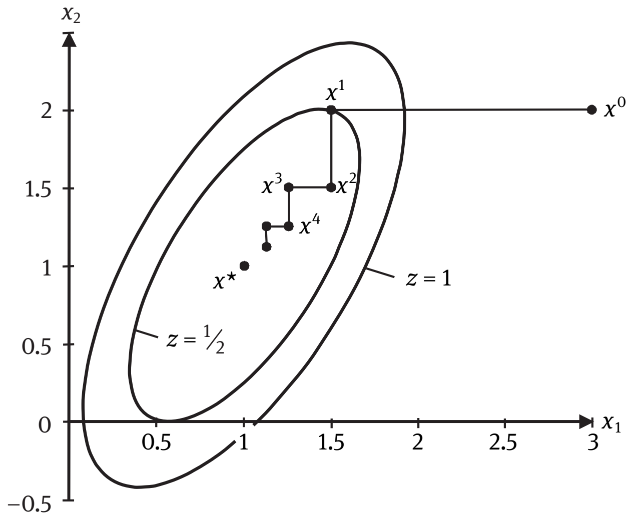


Fig. 7.6. Solution of Example 7.38.

7.6 Inexact line search

Except Newton's method (see the beginning of Section 7.3), all procedures of this chapter determine the step size λ_k with exact line search. This is reasonable, if one wants to perform as few iterations as possible. But we have seen already, that the computational work per iteration may be considerable in this case, since the exact solution of a one-dimensional optimization problem is required. Therefore, modern optimization techniques frequently avoid the one-dimensional optimization and try to determine with "few effort" a λ_k such that

$$f(x^k + \lambda_k d^k) < f(x^k) \tag{7.32}$$

(see the three steps described at the beginning of Chapter 7). Such an *inexact line search* aims to reduce the total time of resolution. However, the following example demonstrates that not every step size satisfying (7.32) is useful.

Example 7.40. To solve the problem

$$\min_{x \in \mathbb{R}^2} x_1^2 + 2x_2^2 \tag{7.33}$$

with optimal point $x^* = (0, 0)^T$, we apply a modification of the gradient method (Section 7.2) in which the search direction is $d^k = -\operatorname{grad} f(x^k)$ and the step size λ_k is predetermined in a specific manner. Denoting the iteration point by $x^k = (x_1^k, x_2^k)^T$, the recurrence formula

$$x^{k+1} = x^k - \lambda_k \operatorname{grad} f(x^k)$$

becomes

$$x_1^{k+1} = x_1^k - 2\lambda_k x_1^k ,$$
$$x_2^{k+1} = x_2^k - 4\lambda_k x_2^k ,$$

i.e.

$$\begin{aligned} x_1^k &= (1-2\lambda_0)(1-2\lambda_1)\dots(1-2\lambda_{k-1})x_1^0 ,\\ x_2^k &= (1-4\lambda_0)(1-4\lambda_1)\dots(1-4\lambda_{k-1})x_2^0 , \end{aligned} \tag{7.34}$$

where $x^0 = (x_1^0, x_2^0)^T$ is the starting point.

Table 7.6. Inexact line search (case a).

k	x^k	$f(x^k)$
0	$\binom{2}{3}$	22
1	$\binom{1.5}{1.5}$	6.75
2	$\binom{1.313}{1.125}$	4.255
3	$\binom{1.230}{0.9844}$	3.451
4	$\binom{1.192}{0.9229}$	3.124

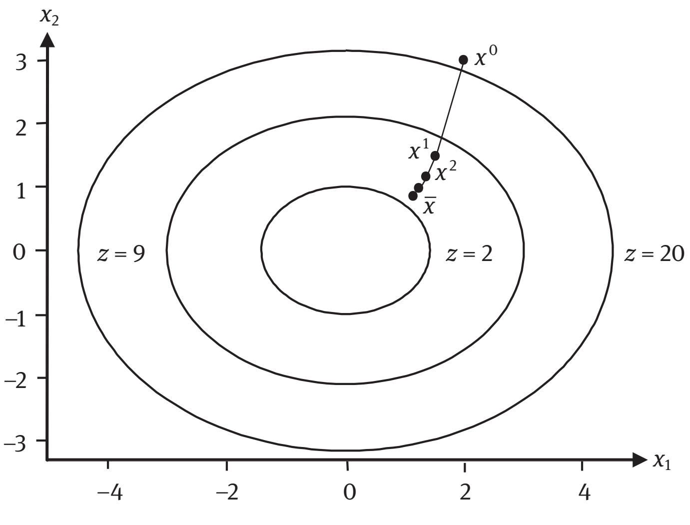


Fig. 7.7. Convergence to a nonoptimal point.

(a) Choosing $x^0 = (2,3)^T$ and $\lambda_k = \frac{1}{2^{k+3}}$, formula (7.34) yields the following iteration points in four-digit arithmetic:
For $c := \prod_{k=1}^{\infty} \frac{2^k-1}{2^k} \approx 0.2888$ the iteration points converge to the point $\overline{x} = (4c, 3c)^T \approx (1.155, 0.8664)^T$ which is not optimal (Figure 7.7).

(b) For the same starting point x^0 and $\lambda_k = \frac{1}{2} - \frac{1}{2^{k+3}}$ we obtain the following iterations.

Table 7.7. Inexact line search (case b).

k	x^k	$f(x^k)$
0	$\binom{2}{3}$	22
1	$\binom{0.5}{-1.5}$	4.75
2	$\binom{0.0625}{1.125}$	2.535
3	$\binom{0.003906}{-0.9844}$	1.938
4	$\binom{0.0001221}{0.9229}$	1.703

We now get $\lim_{k\to\infty} x_1^k = 0$, $\lim_{k\to\infty} x_2^{2k} = 3c \approx 0.8664$, $\lim_{k\to\infty} x_2^{2k+1} = -3c$, i.e. for a large k the algorithm "oscillates" between points close to $\tilde{x} := (0, 3c)^T$ and $-\tilde{x}$ (Figure 7.8).

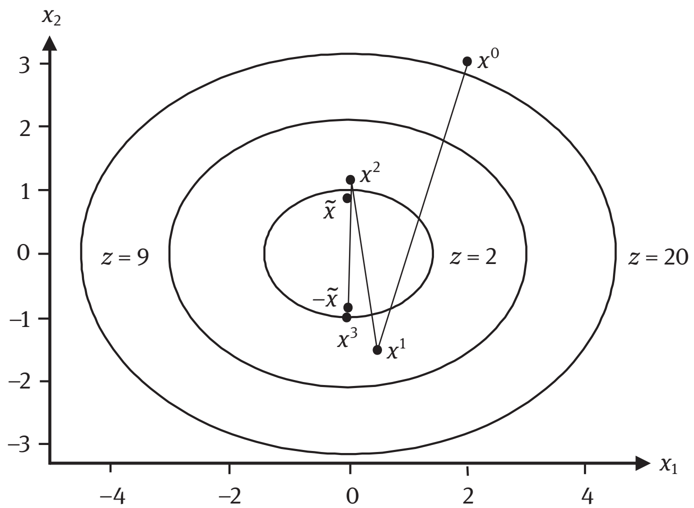


Fig. 7.8. Oscillation of the algorithm.

In both of the preceding cases the algorithm does not achieve the optimum, though it searches in "correct directions". In case (a) the steps are too small compared to the decrease of the objective value. In case (b) the steps are too large, and in each iteration the algorithm passes near the optimum x^* but terminates far away from it.

We now develop strategies which result in an adequate step size. At first we substitute the condition (7.32) by the so-called *Armijo condition*

$$f(x^k + \lambda_k d^k) < f(x^k) + \alpha\,\lambda_k d^{kT}\,\mathrm{grad} f(x^k)\,, \tag{7.35}$$

where d^k denotes any descent direction which may be chosen, e.g. as in the gradient or Newton's method, and α is any value between 0 and 1. The condition (7.35) is equivalent to

$$f(x^k) - f(x^k + \lambda_k d^k) > -\alpha\,\lambda_k d^{kT}\,\mathrm{grad} f(x^k)\,. \tag{7.36}$$

We assume that d^k is a strict descent direction (see (4.14) and the subsequent considerations). In this case, (7.36) means that the decrease (left side) must be larger than the positive value to the right. This condition is stricter than (7.32) which only requires a positive decrease. Therefore the Armijo condition is also called *condition of sufficient decrease*. The larger the step size λ_k, the greater is the required decrease. Hence, λ_k may not be too large relative to the decrease.

In order to avoid too small steps, we proceed as follows. The algorithm starts with a sufficiently large initial step size $\lambda^{(0)}$, and choosing a value $\tau \in (0, 1)$, say $\tau = 1/2$, the search direction is reversed by checking whether the values $\tau\lambda^{(0)}$, $\tau^2\lambda^{(0)}$, $\tau^3\lambda^{(0)}, \ldots$ also satisfy the Armijo condition (Figure 7.9).

Fig. 7.9. Reversal strategy.

The first of these values satisfying (7.35) is chosen as the step size. We obtain the following *backtracking procedure* to determine the step size:

(1) Choose $\lambda^{(0)} > 0$; $\alpha, \tau \in (0, 1)$ and set $i = 0$.
(2) Until $f(x^k + \lambda^{(i)} d^k) < f(x^k) + \alpha\lambda^{(i)} d^{kT}\,\mathrm{grad} f(x^k)$ do $\lambda^{(i+1)} = \tau\lambda^{(i)}$, $i = i + 1$.
(3) Set $\lambda_k = \lambda^{(i)}$.

So we get the *generic line search method* for unrestricted minimization:

(1) Choose a starting point x^0 and set $k = 0$.
(2) Until convergence
    Determine a descent direction d^k at x^k.
    Determine the step size λ_k with backtracking.
    Set $x^{k+1} = x^k + \lambda_k d^k$, $k = k + 1$.

Example 7.41. We apply the generic line search method to problem (7.33) with parameters $\lambda^{(0)} = 1$, $\alpha = 1/5$, $\tau = 1/2$ (defining x^0 and d^k as above). Thus the Armijo condition is

$$f(x^k - \lambda^{(i)} \operatorname{grad} f(x^k)) < f(x^k) - \frac{1}{5}\lambda^{(i)} |\operatorname{grad} f(x^k)|^2 . \tag{7.37}$$

$k = 0$: Condition (7.37) becomes

$$f\left(\begin{pmatrix}2\\3\end{pmatrix} - \lambda^{(i)}\begin{pmatrix}4\\12\end{pmatrix}\right) < f\left(\begin{pmatrix}2\\3\end{pmatrix}\right) - \frac{1}{5}\lambda^{(i)} 160$$

$$\Leftrightarrow \left(2 - 4\lambda^{(i)}\right)^2 + 2\left(3 - 12\lambda^{(i)}\right)^2 < 22 - 32\lambda^{(i)} . \tag{7.38}$$

The smallest i for which (7.38) holds, is $i = 2$, thus $\lambda_0 = 1/4$ (in Example 7.40 we obtained $\lambda_0 = 1/8$ and $\lambda_0 = 3/8$, respectively), and

$$x^1 = x^0 + \lambda_0 d^0 = \begin{pmatrix}2\\3\end{pmatrix} - \frac{1}{4}\begin{pmatrix}4\\12\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix} .$$

$k = 1$: Condition (7.37) has the form

$$f\left(\begin{pmatrix}1\\0\end{pmatrix} - \lambda^{(i)}\begin{pmatrix}2\\0\end{pmatrix}\right) < f\left(\begin{pmatrix}1\\0\end{pmatrix}\right) - \frac{4}{5}\lambda^{(i)} \Leftrightarrow \left(1 - 2\lambda^{(i)}\right)^2 < 1 - \frac{4}{5}\lambda^{(i)} . \tag{7.39}$$

This condition is not satisfied for $\lambda^{(0)} = 1$, but for $\lambda^{(1)} = 1/2$, thus $\lambda_1 = 1/2$ and

$$x^2 = x^1 + \lambda_1 d^1 = \begin{pmatrix}1\\0\end{pmatrix} - \frac{1}{2}\begin{pmatrix}2\\0\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} = x^* .$$

This point is already optimal.

We will make use of the following concept.

Definition 7.42. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called *Lipschitz continuous* (over the set $M \subset \mathbb{R}^n$) with *Lipschitz constant* $L > 0$, if

$$|f(x) - f(y)| \le L|x - y| \quad \text{for } x, y \in M . \tag{7.40}$$

The number L is an upper limit for the "variation speed" of the function f. For $m = 1$, a continuously differentiable function f is Lipschitz continuous over M, if and only if

$$\sup_{x \in M} \{|\operatorname{grad} f(x)|\} \tag{7.41}$$

is finite. In this case, (7.41) is the smallest Lipschitz constant.

For practical purposes of optimization it is usually sufficient to show that a Lipschitz constant exists. It may be difficult to determine the smallest of them for $m > 1$.

Exercise 7.43. *Decide if the function is Lipschitz continuous over M. If yes, determine the smallest Lipschitz constant.*
(i) $f(x) = x^2$, $M = \mathbb{R}$ *and* $M = [-1, 3]$, *respectively,*
(ii) $f(x) = \frac{1}{x}$, $M = (0, \infty)$,
(iii) $f(x_1, x_2) = 2x_1^2 + 3x_2$, $M = \{x \in \mathbb{R}^2 | -3 \le x_1 \le 5, -2 \le x_2 \le 6\}$ *and* $M = \mathbb{R}^2$, *respectively.*

Without proof we now present the following theorems of global convergence.

Theorem 7.44. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable such that* $\operatorname{grad} f(x)$ *is Lipschitz continuous. For the generic line search method using the maximum descent direction $d^k = -\operatorname{grad} f(x^k)$ it holds*

$$\lim_{k\to\infty} f(x^k) = -\infty \tag{7.42}$$

or

$$\lim_{k\to\infty} \operatorname{grad} f(x^k) = 0\,. \tag{7.43}$$

If f is bounded below, then the procedure converges to a point in which the first-order optimality conditions are satisfied.

Theorem 7.45. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable such that* $\operatorname{grad} f(x)$ *is Lipschitz continuous. For the generic line search method using a modified Newton direction, i.e. $d^k = -B_k^{-1} \operatorname{grad} f(x^k)$, where B_k is symmetric and positive definite, one of the conditions (7.42) or (7.43) is satisfied when the eigenvalues of B_k are uniformly limited, i.e. there exist constants c_1 and c_2 with $0 < c_1 < c_2$ such that for an arbitrary k, all eigenvalues of B_k are contained in the interval $[c_1, c_2]$.*

We call attention to the following fact. We have seen that the generic line search is a strict descent method, i.e. $f(x^{k+1}) < f(x^k)$ for $k = 0, 1, 2, \dots$ Therefore all iteration points x^k belong to the set $C(x^0) = \{x \in \mathbb{R}^n | f(x) < f(x^0)\}$ and the properties of differentiability and continuity of the function f in the last two theorems need to be satisfied only over the set $C(x^0)$. These conditions are not very restrictive, if $C(x^0)$ is bounded (see Exercise 7.43).

Exercise 7.46. *Apply the generic line search method to the problem*

$$\min_{x\in\mathbb{R}^2} \frac{1}{2}x_1^2 + \frac{9}{2}x_2^2\,.$$

Choose $x^0 = (9, 1)^T$, $\lambda^{(0)} = 1$, $\alpha = 1/4$, $\tau = 1/2$ and perform three iterations. Determine the descent direction by
(a) $d^k = -\operatorname{grad} f(x^k)$,
(b) $d^k = -(Hf(x^k))^{-1} \operatorname{grad} f(x^k)$ *(Newton direction).*

Illustrate the iterations graphically.

7.7 Trust region methods

For a given iteration point x^k, all previous methods of this chapter only search along a given line to find a "better" point, i.e. a point x^{k+1}, satisfying $f(x^{k+1}) < f(x^k)$. An alternative approach is to expand the search to arbitrary directions, i.e. to select a point x^{k+1} "close" to x^k, which results in a sufficient decrease of the objective value. Clearly, a complete evaluation of a complex function f in a whole neighborhood of x^k is not realizable, but f can be approximately evaluated in a certain region by means of a linear or quadratic model. The search region may not be too large so that we can "trust" that the quadratic approximation is good enough in this region. We will now present a basic version of a *trust region method*. We restrict ourselves to a quadratic model, i.e. we approximate $f(x^k + s)$ in a neighborhood of x^k by

$$q_k(s) = \frac{1}{2}s^T Q_k s + c^{kT} s + f(x^k)\,, \tag{7.44}$$

where $Q_k{:=}Hf(x^k)$ and $c^k{:=}\operatorname{grad} f(x^k)$ (see (7.3)). As *trust region* we choose

$$\{s \in \mathbb{R}^n | \quad |s| \le \Delta_k\}\,. \tag{7.45}$$

In order to determine x^{k+1}, we calculate a solution s^k of the subproblem

$$\min_{|s| \le \Delta_k} q_k(s)\,. \tag{7.46}$$

We accept $x^{k+1} = x^k + s^k$ as the new iteration point, if the quotient

$$\rho_k = \frac{f(x^k) - f(x^k + s^k)}{f(x^k) - q_k(s^k)} \tag{7.47}$$

of the real decrease $f(x^k) - f(x^k + s^k)$ and the predicted decrease $f(x^k) - q_k(s^k)$ is "sufficiently large", i.e. $\rho_k \ge \alpha$ for $0 < \alpha < 1$. If ρ_k is "very large", i.e. $\rho_k \ge \beta$ for $\alpha \le \beta < 1$, we assume that increasing the trust region is beneficial for the following iterations. Finally, if ρ_k is "very small", i.e. $\rho_k < \alpha$, we reject the provisional new point, i.e. we set $x^{k+1} = x^k$ and resolve (7.46) for a smaller radius Δ_k. We obtain the following algorithm:

(1) Choose $x^0, \Delta_0, \alpha, \beta$ and the parameters p, q to control the radius ($\Delta_0 > 0, 0 < \alpha \le \beta < 1, p < 1 < q$) and set $k = 0$.

(2) Until convergence
Determine a solution s^k of (7.46).
Calculate ρ_k in (7.47).

$$\text{Set} \begin{cases} x^{k+1} = x^k + s^k\,, \ \Delta_{k+1} = q\Delta_k & \text{for } \rho_k \ge \beta\,, \\ x^{k+1} = x^k + s^k\,, \ \Delta_{k+1} = \Delta_k & \text{for } \alpha \le \rho_k < \beta\,, \\ x^{k+1} = x^k \qquad\quad\ , \ \Delta_{k+1} = p\Delta_k & \text{for } \rho_k < \alpha, \end{cases}$$

and set $k = k + 1$.

The three cases for ρ_k correspond to a great success, a success and a failure, respectively, in the approximation of the function f by the model.

In order to solve the subproblem (7.46), we can make use of the following statement which will be generalized below.

Theorem 7.47. *Let $Q \in \mathbb{R}^{n\times n}$ be symmetric and positive definite and $c \in \mathbb{R}^n$. The unique minimum point s^* of the convex problem*

$$\min_{|s|\leq\Delta} \frac{1}{2}s^T Qs + c^T s \tag{7.48}$$

satisfies

$$s^* = -Q^{-1}c \tag{7.49}$$

if $|Q^{-1}c| \leq \Delta$, and

$$s^* = -(Q+\lambda^* I)^{-1}c, |s^*| = \Delta, \lambda^* > 0 \tag{7.50}$$

otherwise.

Proof. The problem (7.48) is equivalent to

$$\begin{aligned} \min f(s) &:= \frac{1}{2}s^T Qs + c^T s \\ g(s) &:= s_1^2 + \cdots + s_n^2 - \Delta^2 \leq 0\,. \end{aligned}$$

We get $\operatorname{grad} f(s) = Qs + c$, $\operatorname{grad} g(s) = 2s$, thus the KKT conditions (Theorem 4.6) are

$$\begin{aligned} Qs + c + 2us &= 0\,, \\ u(s_1^2 + \cdots + s_n^2 - \Delta^2) &= 0\,, \\ s_1^2 + \cdots + s_n^2 - \Delta^2 &\leq 0\,, \\ u &\geq 0\,. \end{aligned}$$

For $\lambda :=$ 2u we get the equivalent system

$$\begin{aligned} (Q+\lambda I)s &= -c\,, \\ \lambda(|s| - \Delta) &= 0\,, \\ |s| &\leq \Delta\,, \\ \lambda &\geq 0\,. \end{aligned}$$

For $\lambda = 0$ we obtain $s = -Q^{-1}c$, $|s| \leq \Delta$, and for $\lambda > 0$ we get

$$s = -(Q+\lambda I)^{-1}c\,, \quad |s| = \Delta\,.$$

□

Theorem 4.11 assures that the solution is a global minimum point.

To illustrate the theorem, one can distinguish two cases: If the solution $-Q^{-1}c$ of the unrestricted problem $\min \frac{1}{2}s^T Qs + c^T s$ belongs to the region $\{s \in \mathbb{R}^n \,||s| \leq \Delta\}$, then $-Q^{-1}c$ obviously solves also (7.48). Otherwise the solution s^* of (7.48) lies on the boundary of that region, i.e. $|s^*| = \Delta$ (see Figure 7.10).

In order to calculate s^* in (7.50), we must first determine the value λ^*. This can be done with the aid of the Spectral Theorem. Let $Q = YDY^T$ be the diagonalization of Q (see Theorem 7.3). Since Y is orthogonal, we obtain:

$$\begin{aligned} Q + \lambda^* I = YDY^T + Y\lambda^* IY^T = Y(D + \lambda^* I)Y^T \Rightarrow \\ (Q + \lambda^* I)^{-1} = Y(D + \lambda^* I)^{-1} Y^T. \end{aligned} \tag{7.51}$$

Combining this with (7.50), we obtain

$$\Delta^2 = |s^*|^2 = |(Q + \lambda^* I)^{-1}c|^2 = |Y(D + \lambda^* I)^{-1} Y^T c|^2 = |(D + \lambda^* I)^{-1} Y^T c|^2,$$

and since $|Yx| = |x|$ for all x, we get

$$\Delta^2 = \left| \begin{pmatrix} \frac{1}{\lambda_1 + \lambda^*} & & \\ & \ddots & \\ & & \frac{1}{\lambda_n + \lambda^*} \end{pmatrix} Y^T c \right|^2 = \left| \left(\frac{\mathrm{e}^{1T} Y^T c}{\lambda_1 + \lambda^*}, \dots, \frac{\mathrm{e}^{nT} Y^T c}{\lambda_n + \lambda^*} \right) \right|^2 = \sum_{i=1}^{n} \frac{\alpha_i^2}{(\lambda_i + \lambda^*)^2}, \tag{7.52}$$

where the e^i denote the unit vectors and $\alpha_i = \mathrm{e}^{iT} Y^T c = c^T Y \mathrm{e}^i = c^T y^i$. Now the value λ^* is uniquely determined by (7.52), since Δ^2 is a decreasing function of λ^*, and s^* can be calculated by using (7.50).

Example 7.48. We solve (7.48) for

$$Q = \begin{pmatrix} 4 & 4 \\ 4 & 10 \end{pmatrix}$$

(see Example 7.4), $c = -(16, 34)^T$ and $\Delta = 2$. Since $Q^{-1}c = -(1, 3)^T$, it holds $|Q^{-1}c| = \sqrt{10} > \Delta$, and the solution can be calculated by (7.50). We get the diagonalization $Q = YDY^T$ with

$$Y = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}, \quad D = \begin{pmatrix} 2 & 0 \\ 0 & 12 \end{pmatrix},$$

thus

$$\begin{aligned} &\lambda_1 = 2, \quad \lambda_2 = 12, \quad y^1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \\ &y^2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \alpha_1 = c^T y^1 = \frac{2}{\sqrt{5}}, \quad \alpha_2 = c^T y^2 = -\frac{84}{\sqrt{5}}. \end{aligned}$$

Relation (7.52) implies

$$\Delta^2 = 4 = \sum_{i=1}^{2} \frac{\alpha_i^2}{(\lambda_i + \lambda^*)^2} = \frac{0.8}{(2 + \lambda^*)^2} + \frac{1411.2}{(12 + \lambda^*)^2} \Rightarrow \lambda^* \approx 6.807.$$

From (7.50) we now obtain

$$s^* = -(Q+\lambda^* I)^{-1}c = \begin{pmatrix} 4+\lambda^* & 4 \\ 4 & 10+\lambda^* \end{pmatrix}^{-1} \begin{pmatrix} 16 \\ 34 \end{pmatrix}$$
$$= \frac{1}{(4+\lambda^*)(10+\lambda^*)-16} \begin{pmatrix} 10+\lambda^* & -4 \\ -4 & 4+\lambda^* \end{pmatrix} \begin{pmatrix} 16 \\ 34 \end{pmatrix} \approx \begin{pmatrix} 0.8024 \\ 1.832 \end{pmatrix}$$

with objective value $\frac{1}{2} s^{*T} Q s^* + c^T s^* \approx -51.18$ (see Figure 7.10).

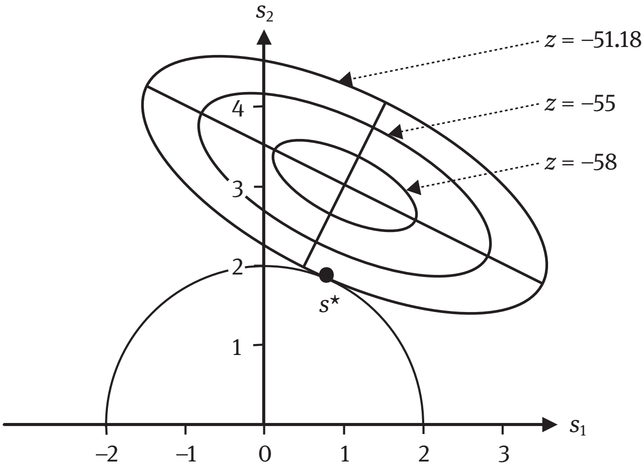


Fig. 7.10. Solution of the quadratic subproblem.

Theorem 7.47 can be generalized for a positive semidefinite or indefinite matrix Q. In these cases the solution need not be unique. The following result is due to Moré and Sorensen.

Theorem 7.49. *Let $Q \in \mathbb{R}^{n\times n}$ be symmetric and $c \in \mathbb{R}^n$. Any global minimum point s^* of (7.48) satisfies $(Q+\lambda^* I)s^* = -c$, where λ^* is such that $(Q+\lambda^* I)$ is positive semidefinite, $\lambda^* \geq 0$ and $\lambda^*(|s^*| - \Delta) = 0$.*

Example 7.50. We apply the trust region method to the function

$$f(x) = x_1^4 + x_1^2 + x_1 x_2 + (1+x_2)^2 \tag{7.53}$$

(see Figure 7.11). We get

$$\operatorname{grad} f(x) = \begin{pmatrix} 4x_1^3 + 2x_1 + x_2 \\ x_1 + 2(1+x_2) \end{pmatrix},$$
$$Hf(x) = \begin{pmatrix} 12x_1^2 + 2 & 1 \\ 1 & 2 \end{pmatrix}$$

where the Hessian matrix $Hf(x)$ is positive definite for all x. We choose $x^0 = (1,1)^T$, $\Delta_0 = 1$, $\alpha = 0.2$, $\beta = 0.9$, $p = 0.5$, $q = 2$.
$k = 0$: We get

$$Q_0 = Hf(x^0) = \begin{pmatrix} 14 & 1 \\ 1 & 2 \end{pmatrix}, \quad c^0 = \operatorname{grad} f(x^0) = \begin{pmatrix} 7 \\ 5 \end{pmatrix}.$$

Since $|Q_0^{-1}c^0| = \sqrt{50}/3 > \Delta_0$, we calculate the solution by means of (7.50). From the diagonalization $Q_0 = YDY^T$ we get $\lambda_1 \approx 14.08$, $\lambda_2 \approx 1.917$, $y^1 \approx \binom{0.9966}{0.08248}$, $y^2 \approx \binom{0.08248}{-0.9966}$, thus $\alpha_1 = c^{0T}y^1 \approx 7.389$, $\alpha_2 = c^{0T}y^2 \approx -4.406$.

Formulas (7.52) and (7.50) yield

$$\Delta_0^2 = 1 \approx \frac{7.389^2}{(14.08 + \lambda^*)^2} + \frac{4.406^2}{(1.917 + \lambda^*)^2} \Rightarrow \lambda^* \approx 2.972\,,$$
$$s^0 = -(Q_0 + \lambda^* I)^{-1}c^0 \approx -\begin{pmatrix} 16.97 & 1 \\ 1 & 4.972 \end{pmatrix}^{-1} \begin{pmatrix} 7 \\ 5 \end{pmatrix} \approx \begin{pmatrix} -0.3575 \\ -0.9337 \end{pmatrix}.$$

In order to evaluate the decrease, we calculate (see (7.47))

$$\rho_0 = \frac{f(x^0) - f(x^0 + s^0)}{f(x^0) - q_0(s^0)} \approx \frac{7 - 1.763}{7 - 1.929} \approx 1.033 \geq \beta\,.$$

The decrease is even better than predicted by the model, i.e. we set $x^1 = x^0 + s^0 \approx \binom{0.6425}{0.0663}$ and increase the trust region by setting $\Delta_1 = q\Delta_0 = 2$.
$k = 1$: We get

$$Q_1 = Hf(x^1) \approx \begin{pmatrix} 6.954 & 1 \\ 1 & 2 \end{pmatrix}, \quad c^1 = \operatorname{grad} f(x^1) \approx \begin{pmatrix} 2.412 \\ 2.775 \end{pmatrix}.$$

Now the point $-Q_1^{-1}c^1 \approx -\binom{0.1587}{1.308}$ is in the interior of the trust region, i.e. we get $s^1 \approx -\binom{0.1587}{1.308}$.

Since

$$\rho_1 = \frac{f(x^1) - f(x^1 + s^1)}{f(x^1) - q_1(s^1)} \approx \frac{1.763 - (-0.2535)}{1.763 - (-0.2435)} \approx 1.005 \geq \beta\,,$$

we set

$$x^2 = x^1 + s^1 \approx \begin{pmatrix} 0.4838 \\ -1.242 \end{pmatrix}, \Delta_2 = q\Delta_1 = 4.$$

$k = 2$: We get

$$Q_2 = Hf(x^2) \approx \begin{pmatrix} 4.809 & 1 \\ 1 & 2 \end{pmatrix},$$
$$c^2 = \operatorname{grad} f(x^2) \approx \begin{pmatrix} 0.1786 \\ -0.0002 \end{pmatrix}.$$

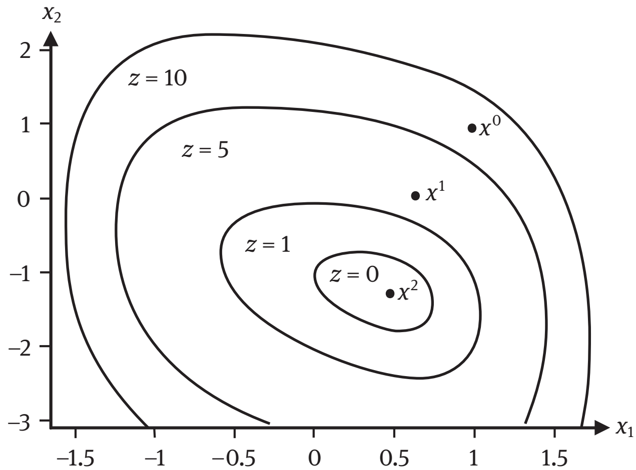


Fig. 7.11. Trust region method.

Since $-Q_2^{-1}c^2 \approx \binom{-0.04147}{0.02084}$ is in the interior of the trust region, we get

$$s^2 \approx \begin{pmatrix} -0.04147 \\ 0.02084 \end{pmatrix}.$$

It holds

$$\rho_2 \approx \frac{-0.2535 - (-0.2573)}{-0.2535 - (-0.2572)} > 1 \geq \beta\,,$$

i.e. $x^3 = x^2 + s^2 \approx \binom{0.4423}{-1.221}$. This point is already "close" to the optimal solution $x^* \approx \binom{0.4398}{-1.220}$ of (7.53).

The example shows that the trust region method can cause considerable computational work per iteration, if an exact solution of the quadratic subproblem is determined. The resolution of large-scale problems in this way is not practicable. However, there exist some approaches, not detailed here, which resolve the subproblem only approximately. One of them consists in determining a point which is not worse than the so-called *Cauchy point* (see Exercise 7.53) and another applies the conjugate gradient method. We finally present a global convergence theorem.

Theorem 7.51. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. If the solution of the subproblem is not worse than the Cauchy point and every Hessian matrix $Hf(x^k)$ is bounded, then the trust region method satisfies*

$$\lim_{k\to\infty} f(x^k) = -\infty$$

or

$$\lim_{k\to\infty} \operatorname{grad} f(x^k) = 0\,.$$

Exercise 7.52. *Apply the trust region method to the function*

$$f(x) = \frac{1}{2}x_1^4 + 3x_1^2 + 4x_2^2 .$$

Choose $x^0 = (2,1)^T$, $\Delta_0 = 1$, $\alpha = 0.2$, $\beta = 0.9, p = 0.5, q = 2$ *and perform three iterations!*

Exercise 7.53. *The Cauchy point* s_C *of problem (7.48) is obtained by one iteration of the gradient method, starting with* $s = 0$*, i.e.* $s_C = -\alpha^* \operatorname{grad} q(0) = -\alpha^* c$*, where* q *denotes the objective function of (7.48) and* α^* *is the solution of*

$$\min_{|\alpha c| \leq \Delta} q(-\alpha c) .$$

Express s_C *in terms of* Q*,* c *and* Δ*.*

Determine the Cauchy point for the first subproblem of Example 7.50 and compare the objective value with that of the exact solution s^0*. Illustrate geometrically the level curves of the function* q *and the points* s^0 *and* s_C*.*

Exercise 7.54. *Solve the problems*

(a) $\min_{|s| \leq 1} s_1^2 - s_2^2$,

(b) $\min_{|s| \leq 1} 2s_1^2 - 6s_1s_2 + s_2^2 + 2s_1 - s_2$

with the aid of Theorem 7.49! Illustrate the second problem geometrically!

8 Linearly constrained problems

As we could observe in Section 1.2, problems with linear constraints play an important role in practical applications. In the present chapter we present solution methods for this type of model. We start with the general linearly constrained problem (Section 8.1). In Section 8.2 we consider the specific case in which all constraints are equations.

8.1 Feasible direction methods

We consider the problem

$$\begin{aligned} &\min f(x) \\ a^{iT}x &\leq b_i \quad (i = 1, \dots, s) \\ a^{iT}x &= b_i \quad (i = s+1, \dots, m) \end{aligned} \tag{8.1}$$

where f is continuously differentiable, $a^i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$ for $i = 1, \dots, m$. Suppose that a feasible point x^0 is known (which can be determined with the first phase of the simplex method of linear programming). In the following we present two classical feasible direction methods for (8.1), credited to Rosen and Zoutendijk. The basic idea of both is to modify the three steps of unconstrained minimization, described at the beginning of Chapter 7, such that all iteration points x^k are feasible:

Step 1:
For a given point x^k determine a feasible descent direction $d^k \in \mathbb{R}^n$.

Step 2:
Determine a $\lambda_k > 0$ such that $x^k + \lambda_k d^k$ is a feasible point, satisfying

$$f(x^k + \lambda_k d^k) < f(x^k)\,.$$

Step 3:
Set $x^{k+1} = x^k + \lambda_k d^k$.

Note that, e.g. the simplex method of linear programming is a feasible direction method.

8.1.1 Rosen's gradient projection method

For a given feasible point x^* of (8.1) we consider the vector space

$$V^* = \left\{ d \in \mathbb{R}^n \,\middle|\, a^{iT}d = 0 \text{ for } i \in \overline{A}(x^*) \right\}\,, \tag{8.2}$$

where $A(x^*)$ is the set of constraints active at x^* (see Definition 2.3) and

$$\overline{A}(x^*) = A(x^*) \cup \{s+1, \ldots, m\}.$$

This space is parallel to the intersection of the hyperplanes $a^{iT}x = b_i$ with $i \in \overline{A}(x^*)$. By moving from x^* in a direction $d^* \in V^*$, all constraints active at x^* remain active, since $a^{iT}x^* = b_i$ implies $a^{iT}(x^* + \lambda d^*) = b_i$ for all i and all $\lambda \in \mathbb{R}$. If at least one constraint is active at x^*, Rosen's method determines the search direction by projecting the vector $-\operatorname{grad} f(x^*)$ in the space V^*. Assume that the vectors a^i in (8.1) are linearly independent. Let $q > 0$ be the number of constraints active at x^* and $A^* \in \mathbb{R}^{q \times n}$ the matrix composed of (the coefficients of) these constraints, i.e. $V^* = \operatorname{Nu}(A^*) = \{d \in \mathbb{R}^n | A^* d = 0\}$.

Furthermore we define $W^* = \operatorname{Im}(A^{*T}) = \{A^{*T}\alpha | \alpha \in \mathbb{R}^q\}$. The spaces V^* and W^* are orthogonal, called *nucleus* of A^* and *image* of A^{*T} and have dimensions $n-q$ and q, respectively. We now determine the search direction d^*. Since V^* and W^* are orthogonal, we can write the negative gradient as

$$-\operatorname{grad} f(x^*) = d^* + A^{*T}\alpha^*, \tag{8.3}$$

where $d^* \in V^*$ and $\alpha^* \in \mathbb{R}^q$ are uniquely determined. Since $d^* \in V^*$, we obtain

$$\begin{aligned} &0 = A^* d^* = A^*(-\operatorname{grad} f(x^*) - A^{*T}\alpha^*) = -A^* \operatorname{grad} f(x^*) - A^* A^{*T}\alpha^* \Rightarrow \\ &\alpha^* = -\left(A^* A^{*T}\right)^{-1} A^* \operatorname{grad} f(x^*). \end{aligned} \tag{8.4}$$

Substituting (8.4) for α^* in (8.3), we get

$$\begin{aligned} d^* &= -\operatorname{grad} f(x^*) + A^{*T}(A^* A^{*T})^{-1} A^* \operatorname{grad} f(x^*) \\ &= -\left[I - A^{*T}(A^* A^{*T})^{-1} A^*\right] \operatorname{grad} f(x^*). \end{aligned} \tag{8.5}$$

Defining the *projection matrix* P^* as

$$P^* = I - A^{*T}(A^* A^{*T})^{-1} A^*, \tag{8.6}$$

the direction d^* is given by

$$d^* = -P^* \operatorname{grad} f(x^*) \tag{8.7}$$

and this is a feasible descent direction if $d^* \neq 0$. The determination of d^* is illustrated as follows.

Example 8.1. We consider the problem (see Figure 8.1)

$$\begin{aligned} \min x_1^2 + x_2^2 + (x_3-1)^2 & \\ x_1 + x_2 + x_3 &\le 1 && (1) \\ -x_1 &\le 0 && (2) \\ -x_2 &\le 0 && (3) \\ -x_3 &\le 0\,. && (4) \end{aligned}$$

For example, at point $x^* = (1/2, 1/4, 0)^T$ only the constraint (4) is active, and V^* corresponds to the plane $x_3 = 0$. At point $\overline{x} = (0, 1/2, 0)^T$, constraints (2) and (4) are active, and $\overline{V}$ is the intersection of the planes $x_1 = 0$ and $x_3 = 0$, i.e. the axis x_2. It holds $g^* := \operatorname{grad} f(x^*) = (1, 1/2, -2)^T$ (see Figure 8.1) and the matrix composed of the active constraints is $A^* = (0, 0, -1)$. We obtain

$$A^* A^{*T} = 1, \quad \text{thus} \quad (A^* A^{*T})^{-1} = 1,$$

$$P^* = I - A^{*T} A^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} (0, 0, -1)$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and the search direction is (see (8.7))

$$d^* = -\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1/2 \\ -2 \end{pmatrix} = \begin{pmatrix} -1 \\ -1/2 \\ 0 \end{pmatrix}.$$

Figure 8.1 shows that a move from x^* in the direction d^* does not leave the plane V^*. In particular, d^* is feasible at x^* and, by definition, the angle between $\operatorname{grad} f(x^*)$ and d^* is less than 90 degrees, i.e. d^* is also a descent direction.

Similarly we get

$$\overline{g} := \operatorname{grad} f(\overline{x}) = (0, 1, -2)^T,$$

$$\overline{A} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \overline{A}\overline{A}^T = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$(\overline{A}\overline{A}^T)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\overline{P} = I - \overline{A}^T \overline{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} -1 & 0 \\ 0 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and finally

$$\overline{d} = -\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}.$$

This is a feasible descent direction, and a movement in direction $\overline{d}$ does not leave the line $\overline{V}$.

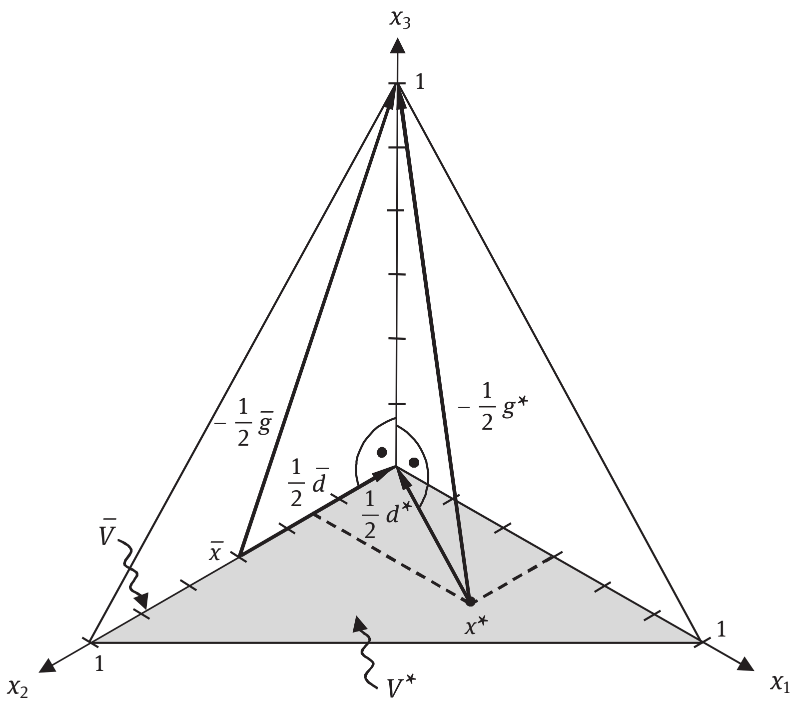


Fig. 8.1. Construction of the search direction.

We now consider the case $d^* = 0$, i.e. the projection of $-\operatorname{grad} f(x^*)$ in the space V^* is zero. Then equation (8.3) implies $-\operatorname{grad} f(x^*) = A^{*T}\alpha^*$. For $\alpha^* \geq 0$ the KKT conditions (Theorem 4.12) are satisfied and the algorithm terminates. If at least one component of $\alpha^* = (\alpha_1^*, \ldots, \alpha_q^*)^T$ is negative, x^* is not optimal. For $\alpha_j^* < 0$, a feasible descent direction at x^* can be determined by removing the constraint j and projecting $-\operatorname{grad} f(x^*)$ in the accordingly modified space V^*. We show that this projection is not null: Let $\widetilde{A}^*$ and $\widetilde{\alpha}^*$ be the matrix and the vector resulting from A^* and α^*, respectively, by removing the constraint j. Then it holds (see (8.3))

$$-\operatorname{grad} f(x^*) = \widetilde{d}^* + \widetilde{A}^{*T}\widetilde{\alpha}^*\,, \tag{8.8}$$

where $\widetilde{d}^*$ is the projection of $-\operatorname{grad} f(x^*)$, using $\widetilde{A}^*$.

Comparing (8.3) and (8.8), we observe that $\widetilde{d}^*$ cannot be null, otherwise these relations would imply that $A^{*T}\alpha^* = \widetilde{A}^{*T}\widetilde{\alpha}^*$. However, this is impossible, since the vectors a^j are linearly independent and the component eliminated from α^* is not zero. Hence, $\widetilde{d}^*$ is a feasible descent direction.

We now discuss the step size, considering the set of values $\lambda \geq 0$, such that

$$x^* + \lambda d^* \tag{8.9}$$

is feasible. By definition, this point is feasible if and only if

$$a^{iT}(x^* + \lambda d^*) \le b_i \quad (i = 1, \dots, s)$$
$$a^{iT}(x^* + \lambda d^*) = b_i \quad (i = s+1, \dots, m)\,,$$

which is equivalent to

$$\begin{aligned} \lambda a^{iT} d^* &\le b_i - a^{iT}x^* \quad (i = 1, \dots, s) \\ \lambda a^{iT} d^* &= b_i - a^{iT}x^* \quad (i = s+1, \dots, m)\,. \end{aligned} \tag{8.10}$$

Since x^* is a feasible point, the right sides of (8.10) are zero for $i \in \overline{A}(x^*)$ (see (8.1)). Since d^* is a feasible direction at x^*, the following conditions equivalent to feasibility hold:

$$\begin{aligned} a^{iT} d &\le 0 \quad \text{for } i \in A(x^*) \\ a^{iT} d &= 0 \quad \text{for } i \in \{s+1, \dots, m\} \end{aligned} \tag{8.11}$$

(see Definition 2.4 and Theorem 2.6). Thus, (8.10) is always satisfied for equations and inequalities active at x^*. Moreover, (8.10) holds for nonactive inequalities with $a^{iT}d^* \le 0$.

Finally, $x^* + \lambda d^*$ is a feasible point if and only if

$$\lambda \le \overline{\lambda} := \min \left\{ \frac{b_i - a^{iT}x^*}{a^{iT}d^*} \,\middle|\, i \in \{1, \dots, s\} \setminus A(x^*) \text{ with } a^{iT}d^* > 0 \right\}. \tag{8.12}$$

If the feasible region of problem (8.1) is unbounded, $a^{iT}d^* \le 0$ may hold for all inactive inequalities. In this case we set $\overline{\lambda} := \infty$, and in practice $\overline{\lambda}$ is a sufficiently large value. The above considerations result in the following *gradient projection algorithm*:

(0) Determine a feasible point x^0 and set $k = 0$.
(1) Determine the matrix of active constraints A_k.
(2) Calculate $P_k = I - A_k^T(A_kA_k^T)^{-1}A_k$ (if no constraint is active, set $P_k = I$) and $d^k = -P_k \operatorname{grad} f(x^k)$.
(3) If $d^k \ne 0$, calculate $\overline{\lambda}_k$ as in (8.12), substituting x^* and d^* by x^k and d^k, respectively. Determine a solution λ_k of the problem

$$\min_{0 \le \lambda \le \overline{\lambda}_k} f(x^k + \lambda d^k)\,.$$

Set $x^{k+1} = x^k + \lambda_k d^k$, $k = k+1$ and go to step (1).
(4) If $d^k = 0$, determine $\alpha^k = -(A_kA_k^T)^{-1}A_k \operatorname{grad} f(x^k)$. If $\alpha^k \ge 0$, stop. Otherwise, delete the line of A_k, corresponding to the smallest component of α^k and go to step (2).

Example 8.2. We solve the problem

$$\begin{aligned} \min 2(x_1-2\,)^2 + (x_2-3.9)^2 & \\ x_1 + 3\,x_2 &\le 9 \\ x_1 + \quad x_2 &\le 5 \\ -x_1 \qquad &\le 0 \\ -x_2 &\le 0\,, \end{aligned}$$

starting with $x^0 = (0,0)^T$. The minimization process is illustrated in Figure 8.2. The arrow starting at x^i shows the direction of $-\operatorname{grad} f(x^i)$.

Iteration 1:
We get $k = 0, x^0 = (0,0)^T$.

$$A_0 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\,, \quad A_0 A_0^T = I\,,$$

$$P_0 = I - A_0^T A_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\,,$$

$$d^0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\,,$$

$$\alpha^0 = -(A_0 A_0^T)^{-1} A_0 \operatorname{grad} f(x^0) = \operatorname{grad} f(x^0) = \begin{pmatrix} -8 \\ -7.8 \end{pmatrix}\,.$$

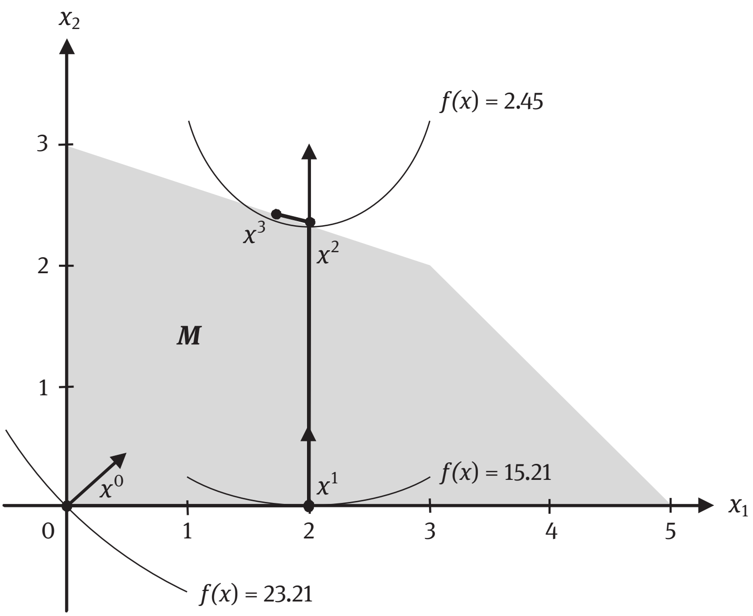


Fig. 8.2. Resolution of Example 8.2 by Rosen's method.

Since -8 is the smallest component of α^0, we delete the first line of A_0 and get

$$A_0 = (0, -1)\,, \quad A_0 A_0^T = 1\,, \quad P_0 = I - A_0^T A_0 = I - \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$d^0 = -\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} -8 \\ -7.8 \end{pmatrix} = \begin{pmatrix} 8 \\ 0 \end{pmatrix},$$

$$\overline{\lambda}_0 = \min \left\{ \frac{9 - (1,3)\begin{pmatrix} 0 \\ 0 \end{pmatrix}}{(1,3)\begin{pmatrix} 8 \\ 0 \end{pmatrix}}, \frac{5 - (1,1)\begin{pmatrix} 0 \\ 0 \end{pmatrix}}{(1,1)\begin{pmatrix} 8 \\ 0 \end{pmatrix}} \right\} = \frac{5}{8}.$$

We must solve the one-dimensional problem

$$\min_{0 \le \lambda \le 5/8} f(x^0 + \lambda d^0) := 2(8\lambda - 2)^2 + 15.21\,,$$

which has the solution $\lambda_0 = 1/4$.

Hence $x^1 = x^0 + \lambda_0 d^0 = \binom{2}{0}$.

We draw attention to the fact that the elimination of the first line of A_0 results in a search along the axis x_1. Likewise, eliminating the second line would result in a search along the axis x_2. Since the angle between $-\operatorname{grad} f(x^0)$ and the axis x_1 is smaller than that between $-\operatorname{grad} f(x^0)$ and the axis x_2, the first alternative is favorable.

Iteration 2:
$k = 1, x^1 = (2, 0)^T$.

$$A_1 = (0, -1)\,, \quad P_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$d^1 = -\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \operatorname{grad} f(x^1) = -\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ -7.8 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$\alpha^1 = (0, 1)\begin{pmatrix} 0 \\ -7.8 \end{pmatrix} = -7.8\,.$$

After the deletion of the unique line of A^1 there are no more active constraints and we set $P_1 = I$, $d^1 = -\operatorname{grad} f(x^1) = \binom{0}{7.8}$.

We get

$$\overline{\lambda}_1 = \min \left\{ \frac{9 - (1,3)\binom{2}{0}}{(1,3)\binom{0}{7.8}}, \frac{5 - (1,1)\binom{2}{0}}{(1,1)\binom{0}{7.8}} \right\} = \frac{7}{23.4}.$$

The problem

$$\min_{0 \le \lambda \le 7/23.4} f(x^1 + \lambda d^1) := (7.8\lambda - 3.9)^2$$

has the solution $\lambda_1 = 7/23.4$, since the objective function is decreasing over the considered interval.

Hence, $x^2 = x^1 + \lambda_1 d^1 = \binom{2}{7/3}$.

Iteration 3:

$$k = 2\,, \quad x^2 = (2, 7/3)^T\,.$$

$$A_2 = (1,3)\,, \quad P_2 = I - \frac{1}{10}\begin{pmatrix}1 & 3\\ 3 & 9\end{pmatrix} = \frac{1}{10}\begin{pmatrix}9 & -3\\ -3 & 1\end{pmatrix}\,,$$

$$d^2 = -\frac{1}{10}\begin{pmatrix}9 & -3\\ -3 & 1\end{pmatrix}\begin{pmatrix}0\\ -9.4/3\end{pmatrix} = \begin{pmatrix}-0.94\\ 0.94/3\end{pmatrix},$$

$$\overline{\lambda}_2 = \min\left\{\frac{0-(-1,0)\binom{2}{7/3}}{(-1,0)\binom{-0.94}{0.94/3}}\right\} = \frac{2}{0.94}\,.$$

The problem

$$\min_{0 \le \lambda \le 2/0.94} f\,(x^2 + \lambda d^2) := 2(0.94\lambda)^2 + \left(\frac{7}{3} + \frac{0.94}{3}\lambda - 3.9\right)^2$$

has the solution $\lambda_2 = \frac{5}{19}$, yielding $x^3 = x^2 + \lambda_2 d^2 = \frac{1}{19}\binom{33.3}{45.9}$.

Iteration 4:

$$k = 3, x^3 = \frac{1}{19}\begin{pmatrix}33.3\\ 45.9\end{pmatrix},$$

$$A_3 = (1,3)\,, \quad P_3 = -\frac{1}{10}\begin{pmatrix}9 & -3\\ -3 & 1\end{pmatrix}, \quad d^3 = -\frac{1}{10}\begin{pmatrix}9 & -3\\ -3 & 1\end{pmatrix}\begin{pmatrix}-18.8/19\\ -56.4/19\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix},$$

$$\alpha^3 = -\left[(1,3)\begin{pmatrix}1\\ 3\end{pmatrix}\right]^{-1}(1,3)\,\text{grad} f(x^3) = 94/95 \ge 0\,.$$

The algorithm terminates with the optimal solution x^3.

Exercise 8.3. *Solve the problem in Example 8.1 completely by the gradient projection method. Start with the points x^* and $\overline{x}$, respectively.*

Exercise 8.4. *Show that the matrix P^* in (8.6) can be written as*

$$P^* = Z^*(Z^{*T}Z^*)^{-1}Z^{*T}\,,$$

where $Z^* \in \mathbb{R}^{n\times(n-q)}$ is any matrix, the columns of which form a basis of $\text{Nu}(A)$.

Exercise 8.5. *Solve the problem*

$$\begin{aligned}
\min (x_1-1)^2 + (x_2-1)^2 + x_3^2 & \\
2x_1 + x_2 + x_3 &\le 5\\
-x_1 &\le 0\\
-x_2 &\le 0\\
-x_3 &\le 0\,,
\end{aligned}$$

starting with $x^0 = (1,2,1)^T$.

The philosophy of the above method is to maintain constraints active as long as possible. Methods based on this idea are called *active constraint methods* or *active set methods* (see, e.g. Luenberger (2003, Section 11.3)). We will return to this issue in Section 8.1.3.

In practice, the projection matrix P_k is not calculated entirely new in each iteration. There are procedures to determine P_{k+1} by updating the matrix P_k of the previous iteration which reduce the computational work considerably. Efficiency and convergence matters are detailed, e.g. in Luenberger (2003, Sections 11.4 and 11.5).

8.1.2 Zoutendijk's method

The preceding method constructs an iteration point x^k such that usually at least one constraint is active at this point, i.e. the procedure progresses preferentially along the surface of the feasible region M. In order to achieve the optimum faster, it seems reasonable to allow a move through the interior of M. The method discussed below is based on this idea.

For (8.1) the vector $d \in \mathbb{R}^n$ is a (strict) descent direction at $x^* \in \mathbb{R}^n$, if

$$(\operatorname{grad} f(x^*))^T d < 0 \tag{8.13}$$

(see Definition 4.7) and d is feasible at x^*, if (8.11) is satisfied. A feasible descent direction at x^* is obtained in an obvious manner by solving the linear problem

$$\begin{aligned} &\min (\operatorname{grad} f(x^*))^T d \\ a^{iT} d &\le 0 \quad (i \in A(x^*)) \\ a^{iT} d &= 0 \quad (i \in \{s+1, \ldots, m\}) \\ -1 \le d_i &\le 1 \quad (i = 1, \ldots, n) \,. \end{aligned} \tag{8.14}$$

The solution of (8.14) is a feasible direction (see (8.11)) with maximum descent, called *optimal direction* at point x^*. The constraints

$$-1 \le d_i \le 1 \qquad (i = 1, \ldots, n) \tag{8.15}$$

delimit the length of the vector d and thus assure that a bounded solution exists. The conditions (8.15) can be substituted by other bounds for the norm $|d|$, for example

$$d^T d \le 1 \tag{8.16}$$

or

$$d_i \le 1 \text{ for } g_i \le 0 \,, \quad -1 \le d_i \text{ for } g_i > 0 \,, \tag{8.17}$$

where g_i is the ith component of the vector $\operatorname{grad} f(x^*)$.

Example 8.6. Consider the problem

$$\begin{aligned} \min x_1^2 + x_2^2 & \\ -2x_1 + x_2 &\leq -1 && (1)\\ x_1 + x_2 &\leq 5 && (2)\\ -x_2 &\leq 1 && (3) \end{aligned}$$

with feasible points $x^* = (1,1)^T$ and $\overline{x} = (3,2)^T$ (see Figure 8.3).

It holds $f(x^*) = (2,2)^T$ and only the constraint (1) is active at x^*. So the optimal direction d^* is the solution of

$$\begin{aligned} \min 2d_1 + 2d_2 & \\ -2d_1 + d_2 &\leq 0 \\ -1 \leq d_1 &\leq 1 \\ -1 \leq d_2 &\leq 1 \end{aligned}$$

(see (8.14)), which can be easily solved by the simplex method or graphically. The unique solution is $d^* = (-1/2, -1)^T$. In practical applications (8.14) is solved by the simplex method for bounded variables. In the same way we find the optimal direction

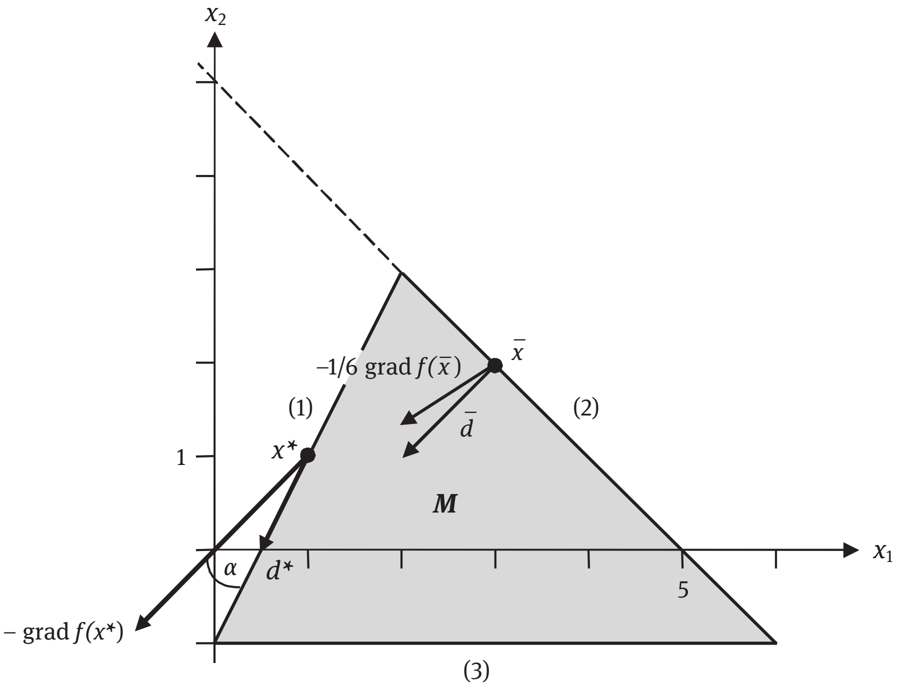


Fig. 8.3. Optimal directions.

at $\overline{x}$ by solving

$$\begin{aligned}&\min 6d_1 + 4d_2\\&d_1 + d_2 \le 0\\&-1 \le d_1 \le 1\\&-1 \le d_2 \le 1\,.\end{aligned}$$

We get $\overline{d} = (-1, -1)^T$. A move in this direction passes through the interior of the set M.

It is clear that the optimal direction depends on the constraints chosen to limit $|d|$.

Exercise 8.7. *Determine the direction in the above example, using the constraint (8.16) instead of (8.15).*

Exercise 8.8. *Assuming that $-g^*$ is a feasible direction at $x^* \in M$, where $g^* :=$ $\operatorname{grad} f(x^*)$. Show that $d^* := -\frac{1}{|g^*|}g^*$ is a solution of (8.14), if the condition (8.15) is replaced by (8.16). Compare with the result of the previous exercise.*

Exercise 8.9. *Given the problem*

$$\begin{aligned}&\min (x_1 - 1)^2 + e^{x_2} - x_2\\&-x_1 + x_2 \le 2 \qquad (1)\\&-x_1 - x_2 \le -2 \qquad (2)\end{aligned}$$

and the feasible point $x^* = (0, 2)^T$. Determine the optimal direction at point x^*, using the bounds (8.17) for $|d|$! Illustrate the problem graphically!

If the optimization problem (8.14) has a solution $d^* = 0$, then it holds

$$(\operatorname{grad} f(x^*))^T d \ge 0$$

for all feasible directions d at x^*, i.e. the first-order necessary optimality condition is satisfied (see Theorem 2.10) and the method terminates.

If the step size is determined as in Section 8.1.1, we obtain *Zoutendijk's algorithm*:

(0) Choose a feasible point x^0, set $k = 0$.
(1) Determine the solution d^k of the problem (8.14) (with $x^* = x^k$).
(2) If $d^k = 0$ (or $|d^k| \le \varepsilon$), stop. Otherwise go to step (3).
(3) Calculate $\overline{\lambda}_k$ by (8.12) (with $x^* = x^k$ and $d^* = d^k$) and determine a solution λ_k of the problem
$$\min_{0 \le \lambda \le \overline{\lambda}_k} f(x^k + \lambda d^k).$$
(4) Set $x^{k+1} = x^k + \lambda_k d^k$, $k = k + 1$ and go to step (1).

Exercise 8.10. *Apply Zoutendijk's method to the problem*

$$\min (x_1 - 5)^2 + x_2^2$$
$$-x_1 - x_2 \le -1 \quad (1)$$
$$x_1 - 2x_2 \le -1 \quad (2)$$
$$2x_1 - 2x_2 \le 1 \quad (3)$$
$$x_1 \ge 0 \quad (4)$$

starting with $x^0 = (0, 1)^T$. *Illustrate the problem graphically!*

An advantage of the above method is that it searches always along an optimal direction. But to determine such a direction, a problem of type (8.14) must be solved. Therefore, Zoutendijk's method generally requires greater computational work per iteration than Rosen's. Furthermore it is possible that the above algorithm tends to a nonoptimal point. An example was constructed by Wolfe (1972). However, Topkis and Veinott (1967) developed a modification of Zoutendijk's method that guarantees convergence (see Bazaraa and Shetty (1979, Section 10.2)).

Even when the Zoutendijk's method converges to an optimum, the convergence speed may be very low due to a zigzag movement:

Example 8.11. We apply Zoutendijk's method to the quadratic problem

$$\min x_1^2 + 3x_2^2$$
$$-x_1 + 5x_2 \le 0 \quad (1)$$
$$-x_1 - 5x_2 \le 0\,, \quad (2)$$

starting with $x^0 = (10, 2)^T$ (see Figure 8.4). To limit $|d|$ we use condition (8.16), i.e. the algorithm searches in the direction of $-\operatorname{grad} f(x^k)$, if this direction is feasible (see Exercise 8.8).

Since $-\operatorname{grad} f(x^0) = (-20, -12)^T$ is a feasible direction at x^0, the algorithm starts moving in this direction. Figure 8.4 shows that $x^0 + \lambda d^0$ is a feasible point, if and only if $\lambda \le 1/4$, and the problem

$$\min_{0 \le \lambda \le 1/4} f(x^0 + \lambda d^0)$$

has the solution $\lambda_0 = 1/4$. Thus,

$$x^1 = x^0 + \lambda_0 d^0 = \begin{pmatrix} 5 \\ -1 \end{pmatrix}.$$

Continuing the procedure in this way, the points

$$x^2 = \begin{pmatrix} 5/2 \\ 1/2 \end{pmatrix}, \quad x^3 = \begin{pmatrix} 5/4 \\ -1/4 \end{pmatrix}, \quad x^4 = \begin{pmatrix} 5/8 \\ 1/8 \end{pmatrix}, \quad \dots$$

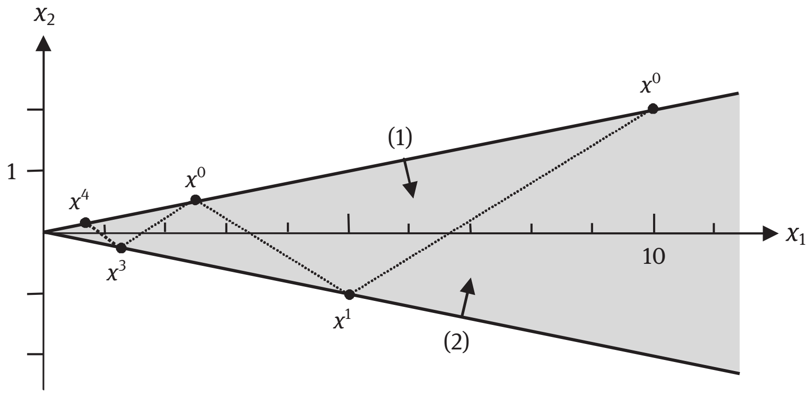


Fig. 8.4. Zigzag movement.

are generated, i.e.

$$x^k = \begin{pmatrix} 10/2^k \\ (-1)^k/2^{k-1} \end{pmatrix} .$$

This sequence converges to the optimal point $(0, 0)^T$ in a zigzag movement.

8.1.3 Advanced techniques: an outline

The above example suggests the following heuristic to avoid a zigzag movement: If possible, the feasible descent direction d^k should be chosen such that the constraints active at x^k are also active at x^{k+1}, which is the basic idea of active set methods (see Section 8.1.1).

Choosing, e.g. the direction d^0 in Example 8.11 in this way, the algorithm searches along the line (1) of Figure 8.4 and determines the solution $(0, 0)^T$ in the first iteration (see Exercise 8.12).

In the general case we obtain the following conditions for d^k:

$$a^{iT}(x^k + \lambda d^k) = b_i \tag{8.18}$$

for $i \in \overline{A}(x^k)$ and all $\lambda \geq 0$. Since $a^{iT}x^k = b_i$ for $i \in \overline{A}(x^k)$, (8.18) is equivalent to

$$a^{iT}d^k = 0 \tag{8.19}$$

for $i \in \overline{A}(x^k)$.

Refining these ideas yields the following algorithm that combines the strategy of active constraints with that of optimal directions. In each iteration one tries to determine a descent direction, satisfying (8.19). If the linear problem in the following step (2) has a solution $d^k \neq 0$, then d^k is a direction with the desired properties. If $d^k = 0$, there is no descent direction satisfying (8.19) and the search direction is de-

termined as in Section 8.1.2:
(0) Determine a feasible point x^0 and set $k = 0$.
(1) Calculate $g^k = \operatorname{grad} f(x^k)$.
(2) Determine a solution d^k of the problem

$$\begin{aligned} &\min g^{kT} d \\ &a^{iT} d = 0 \quad \text{for } i \in \overline{A}(x^k) \\ &-1 \le d_i \le 1 \quad \text{for } i = 1, \dots, n\,. \end{aligned}$$

(3) If $d^k = 0$, go to step (4), otherwise go to step (6).
(4) Determine a solution of the problem

$$\begin{aligned} &\min g^{kT} d \\ &a^{iT} d \le 0 \quad \text{for } i \in A(x^k) \\ &a^{iT} d = 0 \quad \text{for } i = s+1, \dots, m \\ &-1 \le d_i \le 1 \quad \text{for } i = 1, \dots, n\,. \end{aligned}$$

(5) If $d^k = 0$, stop (the KKT conditions are satisfied). Otherwise go to step (6).
(6) Calculate a solution λ_k of the problem

$$\min_{0 \le \lambda \le \overline{\lambda}_k} f(x^k + \lambda d^k)\,,$$

where

$$\overline{\lambda}_k = \min \left\{ \frac{b_i - a^{iT} x^k}{a^{iT} d^k} \,\middle|\, i \in \{1, \dots, s\} \setminus A(x^k) \quad \text{with } a^{iT} d^k > 0 \right\} \quad \text{(see (8.12))}\,.$$

(7) Set $x^{k+1} = x^k + \lambda_k d^k$, $k = k+1$ and go step (1).

In practice the stopping criterion $d^k = 0$ in steps (3) and (5) is often substituted by

$$g^{kT} \frac{d^k}{|d^k|} \le \varepsilon\,.$$

Exercise 8.12. *Apply the above algorithm to the problem of Example 8.11 with* $x^0 = (10, 2)^T$.

Exercise 8.13. *Solve the problem of Exercise 8.10 with the above algorithm.*

Exercise 8.14. *Devise a simple example of problem (8.1) for which the above algorithm needs more iterations than Zoutendijk's method.*

In analogy to the unconstrained optimization, the above algorithm can be improved by several modifications to accelerate the convergence. This results in quasi-Newton methods or conjugate direction methods for linearly constrained problems. We illustrate briefly how adjustments can be incorporated that result in conjugate directions: If $d^k \neq 0$ in step (3) and $\lambda_k < \overline{\lambda}_k$ in step (6), the constraints in (2) are unchanged in the next iteration. In this case, the equation

$$(g^{k+1} - g^k)^T d^{k+1} = 0 \tag{8.20}$$

is added to the linear constraints of the problem in (2). For a quadratic function $f(x) = \frac{1}{2}x^T Q x + c^T x$ we get

$$g^{k+1} - g^k = Q(x^{k+1} - x^k) = \lambda_k Q d^k\,,$$

hence in this case the condition (8.20) is equivalent to $d^{kT} Q d^{k+1} = 0$, i.e. d^k and d^{k+1} are Q-conjugate (see Section 7.4).

If in the next iteration holds $d^{k+1} \neq 0$ in step (3) and $\lambda_{k+1} < \overline{\lambda}_{k+1}$ in step (6), the two equations $(g^{k+1} - g^k)^T d^{k+2} = 0$ and $(g^{k+2} - g^{k+1})^T d^{k+2} = 0$ are added. Continuing in this way, an equation of type (8.20) is added to the constraints of problem (2) until the problem has the solution $0 \in \mathbb{R}^n$ (case 1) or $\lambda_{k+i} = \overline{\lambda}_{k+i}$ occurs (case 2). In case 1, some or all equations of type (8.20) are deleted until a nonzero solution exists, or the algorithm must continue with step (4) (this happens if the solution is still zero after the removal of all constraints of type (8.20)). In case (2), other constraints become active and all additional equations are eliminated. The procedure is illustrated in the following example.

Other arrangements to accelerate the convergence can be found in the literature (see Section 11.3).

Example 8.15. We apply the procedure to the problem

$$\min 2x_1^2 + \frac{1}{2}x_2^2 + \frac{1}{2}x_3^2 + x_1x_2 + x_1x_3 - 2x_1 - x_2 - x_3$$
$$x_1 + 2x_2 + 2x_3 \le 2$$
$$x_1 \ge 0\,, \quad x_2 \ge 0\,, \quad x_3 \ge 0\,.$$

The function f can be written in the form $f(x) = \frac{1}{2}x^T Q x + c^T x$ with

$$Q = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad c = -\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}.$$

We get

$$\operatorname{grad} f(x) = Qx + c\,.$$

Since Q is positive definite, f is convex and thus the one-dimensional function $f(x^k + \lambda d^k)$ in λ is convex. If we minimize over $[0, \infty)$ the minimum point is given by

$$\widetilde{\lambda}_k := -\frac{g^{kT} d^k}{d^{kT} Q d^k}$$

(see Exercise 7.27). Thus the step size λ_k in step (6) of the algorithm is

$$\lambda_k = \min\{\overline{\lambda}_k, \widetilde{\lambda_k}\}\,.$$

To limit the length of d we choose the conditions

$$-1 \le d_i \le 1 \quad \text{for } i = 1, \dots, n$$

(see (8.14)), and we start the method with $x^0 = (0, 0, 0)^T$.

Iteration 1:
$g^0 = Qx^0 + c = (-2, -1, -1)^T$. The linear problem of step (2)

$$\begin{aligned} &\min -2d_1 - d_2 - d_3 \\ &d_1 = d_2 = d_3 = 0 \\ &-1 \le d_i \le 1 \quad (i = 1, 2, 3) \end{aligned}$$

has the solution $d^0 = (0, 0, 0)^T$ and the problem of step (4)

$$\begin{aligned} &\min -2d_1 - d_2 - d_3 \\ &d_1, d_2, d_3 \ge 0 \\ &-1 \le d_i \le 1 \quad (i = 1, 2, 3) \end{aligned}$$

has the solution $d^0 = (1, 1, 1)^T$. We obtain

$$\widetilde{\lambda}_0 = \frac{-g^{0T} d^0}{d^{0T} Q d^0} = \frac{(2, 1, 1)\begin{pmatrix}1\\1\\1\end{pmatrix}}{(1, 1, 1)\begin{pmatrix}4 & 1 & 1\\1 & 1 & 0\\1 & 0 & 1\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix}} = \frac{2}{5}\,, \quad \overline{\lambda}_0 = \frac{2}{5}\,, \quad \lambda_0 = \frac{2}{5}\,,$$

$$x^1 = x^0 + \lambda_0 d^0 = \left(\frac{2}{5}, \frac{2}{5}, \frac{2}{5}\right)^T .$$

Iteration 2:
$g^1 = Qx^1 + c = (\frac{2}{5}, -\frac{1}{5}, -\frac{1}{5})^T$.
Step (2):

$$\begin{aligned} &\min \frac{2}{5}d_1 - \frac{1}{5}d_2 - \frac{1}{5}d_3 \\ &d_1 + 2d_2 + 2d_3 = 0 \\ &-1 \le d_i \le 1 \quad (i = 1, 2, 3) \end{aligned}$$

has the solution $d^1 = (-1, 0, \frac{1}{2})^T$,

$$\widetilde{\lambda}_1 = \frac{2}{13}\,, \quad \overline{\lambda}_1 = \frac{2}{5}\,, \quad \lambda_1 = \frac{2}{13}\,,$$

$$x^2 = x^1 + \lambda_1 d^1 = \left(\frac{6}{65}\frac{2}{5}, \frac{31}{65}\right)^T .$$

Iteration 3:
$g^2 = Qx^2 + c = (-\frac{9}{65}, -\frac{34}{65}, -\frac{18}{65})^T$.
Step (2): We add a constraint of type (8.20) to the linear problem in iteration 1. Since

$$g^2 - g^1 = \left(-\frac{7}{13}, -\frac{2}{13}, -\frac{1}{13}\right)^T ,$$

we get the problem

$$\begin{aligned} \min -\frac{9}{65}d_1 - \frac{23}{65}d_2 - \frac{18}{65}d_3 \\ d_1 + 2d_2 + 2d_3 = 0 \\ -\frac{7}{13}d_1 - \frac{2}{13}d_2 - \frac{1}{13}d_3 = 0 \\ -1 \le d_i \le 1 \quad (i = 1, 2, 3) , \end{aligned}$$

which has the solution $d^2 = (-\frac{2}{13}, 1, -\frac{12}{13})^T$,

$$\tilde{\lambda}_2 = \frac{1}{25} , \quad \overline{\lambda}_2 = \frac{31}{60} , \quad \lambda_2 = \frac{1}{25} ,$$
$$x^3 = x^2 + \lambda_2 d^2 = \left(\frac{6}{25}, \frac{11}{25}, \frac{11}{25}\right)^T .$$

Iteration 4:
$g^3 = Qx^3 + c = (-\frac{4}{25}, -\frac{8}{25}, -\frac{8}{25})^T$.
Step (2): We add a second constraint of type (8.20). Since

$$g^3 - g^2 = \left(-\frac{7}{325}, \frac{11}{325}, -\frac{14}{325}\right)^T ,$$

the new linear problem is

$$\begin{aligned} \min -\frac{4}{25}d_1 - \frac{8}{25}d_2 - \frac{8}{25}d_3 \\ d_1 + 2d_2 + 2d_3 = 0 \\ -\frac{7}{13}d_1 - \frac{2}{13}d_2 - \frac{1}{13}d_3 = 0 \\ -\frac{7}{325}d_1 + \frac{11}{325}d_2 - \frac{14}{325}d_3 = 0 \\ -1 \le d_i \le 1 \quad (i = 1, 2, 3) . \end{aligned}$$

Comparing the objective function with the first restriction, it is easily seen that the solution is $d^3 = (0, 0, 0)^T$ (even after removal of the additional equations).
Step (4): Obviously the problem

$$\begin{aligned} \min -\frac{4}{25}d_1 - \frac{8}{25}d_2 - \frac{8}{25}d_3 \\ d_1 + 2d_2 + 2d_3 \le 0 \\ -1 \le d_i \le 1 \quad (i = 1, 2, 3) \end{aligned}$$

also has the solution $d^3 = (0, 0, 0)^T$. The algorithm ends here. The point x^3 is the wanted solution. Finally we verify that the directions d^1 and d^2 are Q-conjugate:

$$d^{1T}Qd^2 = \left(-1, 0, \tfrac{1}{2}\right)\begin{pmatrix} 4 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} -2/13 \\ 1 \\ -12/13 \end{pmatrix} = 0 .$$

Exercise 8.16. *Apply the above algorithm to the problem*

$$\begin{gathered} \min (x_1 - 1)^2 + (x_2 - 1)^2 + (x_3 + 1)^2 \\ x_1 + x_2 + x_3 \leq 3 \\ x_1 \geq 0 , \quad x_2 \geq 0 , \quad x_3 \geq 0 \end{gathered}$$

with $x^0 = (0, 0, 3)^T$.

In the appropriate literature (see, e.g. Fletcher (1987), Gill, Murray and Wright (1981)) are studied several enhancements of the active set method which will not be detailed here. For example, if the problem in step (2) has the solution $d = 0$, one can eliminate an equality constraint and solve the resulting problem instead of the problem in step (4). The restriction to be removed is adequately determined by using estimates of Lagrange multipliers.

There also exist refined techniques for solving an optimization problem with linear *equality* constraints (see next section) and procedures that determine a descent direction without minimizing the function $g^{kT}d$. Other notable feasible directions methods are, e.g. the reduced gradient method of Wolfe, the convex simplex method of Zangwill and the linearization method of Frank and Wolfe.

8.2 Linear equality constraints

We now turn to the special case of (8.1) in which all constraints are equations:

$$\begin{aligned} &\min f(x) \\ &Ax = b , \end{aligned} \tag{8.21}$$

where f is twice continuously differentiable, $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = m < n$, and $b \in \mathbb{R}^m$.

Problem (8.21) can be easily transformed in an unrestricted minimization problem with $n - m$ variables: since $\operatorname{rank}(A)$ is maximal, A can by partitioned into a nonsingular matrix $B \in \mathbb{R}^{m\times m}$ and a matrix $N \in \mathbb{R}^{m\times(n-m)}$, i.e. $A = (B|N)$. Accordingly we write the vector x in the form $x = \binom{x_B}{x_N}$, i.e. x_B and x_N consist of the components of x, corresponding to B and N, respectively. The relation $Ax = b$ in (8.21) is equivalent to

$$\begin{aligned}(B|N)\begin{pmatrix} x_B \\ x_N \end{pmatrix} &= b \Leftrightarrow \\ Bx_B + Nx_N &= b \Leftrightarrow \\ x_B &= B^{-1}(b - Nx_N)\,.\end{aligned}$$

By substituting x in the objective function in terms of x_B and x_N, we obtain the following unrestricted problem, equivalent to (8.21)

$$\min_{x_N \in \mathbb{R}^{n-m}} h(x_N) := f(B^{-1}(b - Nx_N), x_N)\,. \tag{8.22}$$

Example 8.17. Given the problem

$$\begin{aligned}\min 2x_1^2 + x_2^2 + x_3^2 &+ 3x_4^2 \\ 2x_1 - 3x_2 - 4x_3 - x_4 &= 1 \\ x_1 + 5x_2 - 2x_3 + 6x_4 &= 7\,.\end{aligned}$$

By pivoting, the system $Ax = b$ can be transformed as follows:

Table 8.1. Pivoting of Example 8.17.

x_1	x_2	x_3	x_4	b
2	–3	–4	–1	1
1	5	–2	6	7
1	$-3/2$	–2	$-1/2$	$1/2$
0	$13/2$	0	$13/2$	$13/2$
1	0	–2	1	2
0	1	0	1	1

We get

$$\begin{aligned}x_1 &= 2 + 2x_3 - x_4\,, \\ x_2 &= 1 - x_4\,,\end{aligned} \tag{8.23}$$

and the function h in (8.22) is

$$h(x_3,x_4) = 2(2+2x_3-x_4)^2 + (1-x_4)^2 + x_3^2 + 3x_4^2\,,$$

yielding

$$\begin{aligned}\operatorname{grad} h(x_3,x_4) &= \begin{pmatrix} 4(2+2x_3-x_4)\cdot 2 + 2x_3 \\ -4(2+2x_3-x_4) + 2(x_4-1) + 6x_4 \end{pmatrix},\\ Hh(x_3,x_4) &= \begin{pmatrix} 18 & -8 \\ -8 & 12 \end{pmatrix}.\end{aligned}$$

Since $Hh(x_3,x_4)$ is positive definite, h is strictly convex and the unique minimum point is given by

$$\operatorname{grad} h(x_3,x_4) = 0 \Leftrightarrow \begin{pmatrix} x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -14/19 \\ 13/38 \end{pmatrix}.$$

From (8.23) we obtain

$$x_1 = 7/38\,,\quad x_2 = 25/38\,.$$

Hence the solution of the example is

$$(x_1,\dots,x_4) = \frac{1}{38}(7, 25, -28, 13)\,.$$

Exercise 8.18. *Solve the following problems by the above method:*
(a)

$$\begin{aligned} &\min x_1^2 + 2x_2^2 + 3x_2x_3 \\ &x_1 - 2x_2 \qquad = -1 \\ &3x_1 - x_2 + 5x_3 = 12\,. \end{aligned}$$

(b)

$$\begin{aligned} &\min x_1^2 + \frac{1}{4}x_2^4 \\ &x_1 + x_2 = 6\,. \end{aligned}$$

Exercise 8.19. *Determine the distance between the origin of the* $\mathbb{R}^3$ *and the plane containing the points* $(1,0,0)^T$, $(0,2,0)^T$ *and* $(0,0,3)^T$. *Formulate a problem of type (8.21) and solve it by the method of Example 8.17.*

If the objective function of (8.21) is “simple”, the solution can by explicitly expressed by using the KKT conditions. Consider the following specific case of (8.21):

$$\begin{aligned} &\min \frac{1}{2}x^TQx + c^Tx \\ &Ax = b\,. \end{aligned} \tag{8.24}$$

Assuming that $Q \in \mathbb{R}^{n\times n}$ is positive definite and $q \in \mathbb{R}^n$, the solution is uniquely defined (see, e.g. Luenberger (2003)). The KKT conditions are

$$\begin{aligned} Qx + c + A^T v &= 0 \\ Ax &= b\,. \end{aligned} \tag{8.25}$$

The first equation of (8.25) implies

$$\begin{aligned} Qx &= -c - A^T v \Rightarrow \\ x &= -Q^{-1}c - Q^{-1}A^T v\,. \end{aligned} \tag{8.26}$$

By substituting x in the second equation of (8.25) by the last expression, we get

$$\begin{aligned} -AQ^{-1}c - AQ^{-1}A^T v &= b \Rightarrow \\ AQ^{-1}A^T v &= -(AQ^{-1}c + b) \Rightarrow \\ v &= -(AQ^{-1}A^T)^{-1}(AQ^{-1}c + b)\,, \end{aligned} \tag{8.27}$$

and replacing v in (8.26) by the right side of (8.27) finally yields

$$x = -Q^{-1}c + Q^{-1}A^T(AQ^{-1}A^T)^{-1}(AQ^{-1}c + b)\,. \tag{8.28}$$

Exercise 8.20.
(a) *Formulate the KKT conditions for problem (8.21).*
(b) *Solve the problem of Exercise 8.18 (a) with the aid of these conditions.*

The direct elimination of variables is computationally demanding and may cause considerable rounding errors. We now study an alternative to transform (8.21) in an unrestricted problem which results in a more stable resolution procedure.

We use again the orthogonal spaces $\mathrm{Nu}(A) = \{x \in \mathbb{R}^n | Ax = 0\}$ and $\mathrm{Im}(A^T) = \{A^T y | y \in \mathbb{R}^m\}$ which have dimensions $n - m$ and m and satisfy

$$\begin{aligned} \mathrm{Nu}(A) + \mathrm{Im}(A^T) &= \mathbb{R}^n\,, \\ \mathrm{Nu}(A) \cap \mathrm{Im}(A^T) &= \{0\}\,. \end{aligned} \tag{8.29}$$

From linear algebra it is known that the feasible set $M = \{x \in \mathbb{R}^n | Ax = b\}$ of (8.21) can be written as

$$M = \tilde{x} + \mathrm{Nu}(A) = \{\tilde{x} + x | x \in \mathrm{Nu}(A)\}\,, \tag{8.30}$$

where $\tilde{x}$ is an arbitrary element of M. If $Z \in \mathbb{R}^{n\times(n-m)}$ is a matrix the columns of which form a basis of $Nu(A)$, we get

$$M = \{\tilde{x} + Zy | y \in \mathbb{R}^{n-m}\}\,. \tag{8.31}$$

By setting $\Psi(y) := f(\tilde{x} + Zy)$ we obtain the unrestricted problem

$$\min_{y\in\mathbb{R}^{n-m}} \Psi(y)\,, \tag{8.32}$$

equivalent to (8.21). Clearly, y^* is a local (global) minimum point of ψ, if and only if $x^* := \tilde{x} + Zy^*$ is a local (global) minimum point of (8.21), and principally (8.32) can by solved by the methods of Chapter 7.

We now use the equivalence between (8.21) and (8.32) to develop optimality conditions and a solution method for (8.21). Therefore we determine the gradient and the Hessian matrix of Ψ. It holds $\Psi(y) = f(g(y))$, where $g(y) = \tilde{x} + Zy$, and by applying the chain rule we get

$$J_\psi(y) = J_f(g(y))J_g(y) = (\operatorname{grad} f(g(y)))^T Z\,, \tag{8.33}$$

where $J_f(x)$ is the *Jacobian matrix* of f at x. Thus,

$$\operatorname{grad}\ \Psi(y) = Z^T \operatorname{grad} f(\tilde{x} + Zy)\,. \tag{8.34}$$

Applying the chain rule again, we obtain

$$H\Psi(y) = Z^T Hf(\tilde{x} + Zy)Z\,. \tag{8.35}$$

The expressions (8.34) and (8.35) are called *reduced gradient* and *reduced Hessian matrix*, respectively. In practice, the matrix Z is determined with the aid of the QR decomposition of the matrix A (see Section 11.3).

Using (8.34), (8.35) and Theorem 2.19, we obtain the following second-order necessary conditions for an optimal point x^* of (8.21):

$$Ax^* = b \tag{8.36 a}$$

$$Z^T \operatorname{grad} f(x^*) = 0 \tag{8.36 b}$$

$$Z^T Hf(x^*)Z \quad \text{is positive semidefinite.} \tag{8.36 c}$$

By substituting "positive semidefinite" in (8.36 c) for "positive definite" we obtain sufficient optimality conditions (see Theorem 2.21). Condition (8.36 b) means that the vector $\operatorname{grad} f(x^*)$ is orthogonal to the space $Nu(A)$, i.e. it is an element of $Im(A^T)$ (see (8.29)). Thus, (8.36 b) is equivalent to

$$\operatorname{grad} f(x^*) = A^T \alpha^* \tag{8.37}$$

for an $\alpha^* \in \mathbb{R}^n$, while conditions (8.36 a) and (8.37) correspond to the KKT conditions of (8.20) (see Exercise 8.20).

Example 8.21. We solve the convex problem of Example 8.17 by using the above optimality conditions. We get

$$\operatorname{Nu}(A) = \left\{ y \in \mathbb{R}^4 \,\middle|\, \begin{pmatrix} 2 & -3 & -4 & -1 \\ 1 & 5 & -2 & 6 \end{pmatrix} y = 0 \right\}.$$

It can be easily verified that the vectors $(2, 0, 1, 0)^T$ and $(-1, -1, 0, 1)^T$ form a basis of $Nu(A)$. So we get

$$Z = \begin{pmatrix} 2 & -1 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and the first-order necessary conditions (8.36 a, b) are

$$Z^T \operatorname{grad} f(x) = 0$$
$$Ax = b\,.$$

This system is equivalent to

$$\begin{aligned} 8x_1 \quad\quad +2x_3 \quad\quad &= 0 \\ -4x_1 - 2x_2 \quad\quad + 6x_4 &= 0 \\ 2x_1 - 3x_2 - 4x_3 - x_4 &= 1 \\ x_1 + 5x_2 - 2x_3 + 6x_4 &= 7 \end{aligned}$$

with the unique solution

$$(x_1, \dots, x_4) = \frac{1}{38}(7, 25, -28, 13)\,. \tag{8.38}$$

The matrix

$$Z^T Hf(x^*)Z = \begin{pmatrix} 18 & -8 \\ -8 & 12 \end{pmatrix}$$

in (8.36 c) is positive definite, i.e. (8.38) is a global minimum point.

Exercise 8.22. *Solve the problems of Exercise 8.18 in the above manner.*

We now present a feasible direction method for problem (8.21) (see, e.g. Friedlander (1994, Chapter 8)):

(0) Determine a feasible point x^0 of (8.21) and a matrix $Z \in \mathbb{R}^{n\times(n-m)}$, the columns of which form a basis of $Nu(A)$. Choose $\alpha \in (0, 1)$ and set $k = 0$.
(1) If $Z^T \operatorname{grad} f(x^k) = 0$, stop (the first-order necessary conditions are satisfied for x^k, see (8.36)).
(2) Calculate $d^k = -ZZ^T \operatorname{grad} f(x^k)$.
(3) Set $\lambda = 1$.
(4) If $f(x^k + \lambda d^k) < f(x^k) + \alpha\lambda(\operatorname{grad} f(x^k))^T d^k$, go to step (6).
(5) Choose $\bar\lambda \in [0.1\,\lambda, 0.9\,\lambda]$, set $\lambda = \bar\lambda$ and go to step (4).
(6) Set $\lambda_k = \lambda$, $x^{k+1} = x^k + \lambda_k d^k$, $k = k + 1$ and go to step (1).

The descent direction in step (2) is determined in a natural way: consider the function $\Psi(y) = f(\tilde{x} + Zy)$ in (8.32). Let ω be a descent direction of $\Psi(y)$ at $y = 0$, i.e. $(\operatorname{grad} \Psi(0))^T\omega < 0$ holds (see (8.13)).

It follows that $(Z^T \operatorname{grad} f(\tilde{x}))^T\omega = (\operatorname{grad} f(\tilde{x}))^T Z\omega < 0$, i.e. $Z\omega$ is a descent direction of f at $\tilde{x}$. Since $Z\omega \in Nu(A)$, we get $\tilde{x} + \lambda Z\omega \in M$ for all λ (see (8.30)), i.e. $Z\omega$ is also a feasible direction at $\tilde{x}$. In particular, for $\omega = -\operatorname{grad} \Psi(0) = -Z^T \operatorname{grad} f(\tilde{x})$ (see (8.34)), $Z\omega$ corresponds to the direction determined in step (2). The algorithm searches in this direction and determines the step size according to the Armijo condition (see Section 7.6).

Example 8.23. We solve the problem of Example 8.17 for $x^0 = (2, 1, 0, 0)^T$,

$$Z = \begin{pmatrix} 2 & -1 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

(see Example 8.21) and $\alpha = 0.5$. In step (5) we always choose $\bar{\lambda} = 0.2\lambda$.

Iteration 1:
We have $k = 0$,

$$Z^T \operatorname{grad} f(x^0) = \begin{pmatrix} 2 & 0 & 1 & 0 \\ -1 & -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 2 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 16 \\ -10 \end{pmatrix} \neq 0\,,$$

$$d^0 = -ZZ^T \operatorname{grad} f(x^0) = -\begin{pmatrix} 2 & -1 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 16 \\ -10 \end{pmatrix} = \begin{pmatrix} -42 \\ -10 \\ -16 \\ 10 \end{pmatrix}.$$

Since

$$\begin{aligned} f(x^0 + \lambda d^0) &= 2(2 - 42\lambda)^2 + (1 - 10\lambda)^2 + (16\lambda)^2 + 3(10\lambda)^2 \\ &= 4184\lambda^2 - 356\lambda + 9\,, \end{aligned}$$

$$f(x^0) + \alpha\lambda(\operatorname{grad} f(x^0))^T d^0 = 9 + 0.5\lambda(8, 2, 0, 0) \begin{pmatrix} -42 \\ -10 \\ -16 \\ 10 \end{pmatrix} = 9 - 178\lambda\,,$$

the Armijo condition is not satisfied for $\lambda = 1$ and $\lambda = 0.2$, but for $\lambda = 0.04$. So we obtain $\lambda_0 = 0.04$, $x^1 = x^0 + \lambda_0 d^0 = (0.32, 0.6, -0.64, 0.4)^T$.

Iteration 2:
We have $k = 1$,

$$Z^T \operatorname{grad} f(x^1) = \begin{pmatrix} 2 & 0 & 1 & 0 \\ -1 & -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1.28 \\ 1.2 \\ -1.28 \\ 2.4 \end{pmatrix} = \begin{pmatrix} 1.28 \\ -0.08 \end{pmatrix} \neq 0\,,$$

$$d^1 = -\begin{pmatrix} 2 & -1 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1.28 \\ -0.08 \end{pmatrix} = \begin{pmatrix} -2.64 \\ -0.08 \\ -1.28 \\ 0.08 \end{pmatrix}.$$

We now have

$$\begin{aligned} f(x^1 + \lambda d^1) &= 2(0.32 - 2.64\lambda)^2 + (0.6 - 0.08\lambda)^2 \\ &\quad + (0.64 + 1.28\lambda)^2 + 3(0.4 + 0.08\lambda)^2\,, \\ f(x^1) + \alpha\lambda(\operatorname{grad} f(x^1))^T d^1 &= 1.4544 - 0.8224\lambda \end{aligned}$$

and as before the Armijo condition is not satisfied for $\lambda = 1$ and $\lambda = 0.2$, but for $\lambda = 0.04$. We obtain $\lambda_1 = 0.04$, $x^2 = x^1 + \lambda_1 d^1 = (0.2144, 0.5968, -0.6912, 0.4032)^T$.

Iteration 3:
We have

$$k = 2\,, \quad Z^T \operatorname{grad} f(x^2) = \begin{pmatrix} 0.3328 \\ 0.368 \end{pmatrix} \neq 0\,,$$
$$d^2 = \begin{pmatrix} -0.2976 \\ 0.368 \\ -0.3328 \\ -0.368 \end{pmatrix}, \quad x^3 = \begin{pmatrix} 0.202496 \\ 0.61152 \\ -0.704512 \\ -0.38848 \end{pmatrix}.$$

The sequence of points x^k converges to the solution $x^* = \frac{1}{38}(7.25, -28.13)^T$.

An advantage of the method is that it avoids the one-dimensional minimization to determine the step size. The feasible descent direction of Rosen's method (Section 8.1.1) is a special case of the construction in step (2) of the above algorithm. In the former method the following holds

$$d^k = -Z(Z^T Z)^{-1} Z^T \operatorname{grad} f(x^k) \tag{8.39}$$

(see Exercise 8.4), where the right side is independent of Z (which is any $n\times(n-m)$-matrix the columns of which form a basis of $\mathrm{Nu}(A)$). If Z is orthogonal, i.e. $Z^T Z = I$, the expression (8.39) reduces to the form in step (2).

Exercise 8.24. *Perform all calculations of the last example, substituting Z for*

$$Z = \frac{1}{\sqrt{33}} \begin{pmatrix} 4 & -\sqrt{11} \\ -2 & -\sqrt{11} \\ 3 & 0 \\ 2 & \sqrt{11} \end{pmatrix}.$$

Check if this matrix is orthogonal and the search direction is that of Rosen's method.

9 Quadratic problems

In this chapter we present two methods to solve a quadratic problem. At first we study briefly a specialization of an active set method (see Section 8.1.1). Later we formulate the KKT conditions of a quadratic problem in a condensed form (Section 9.2) and present Lemke's method for solving these conditions (Section 9.3).

9.1 An active-set method

We consider the quadratic optimization problem of the form

$$\begin{aligned} \min \ & \frac{1}{2} x^T Q x + c^T x \\ & a^{iT} x \leq b_i \quad (i = 1, \dots, s) \\ & a^{iT} x = b_i \quad (i = s+1, \dots, m) \end{aligned} \tag{9.1}$$

where $Q \in \mathbb{R}^{n \times n}$ is symmetric and positive definite and the vectors a^i are linearly independent (see (8.1)).

Suppose that a feasible point x^k is known. We consider the "displacement" d^k, defined in such a way that $x^k + d^k$ minimizes the objective value over the set of points $x^k + d$, preserving the activity of the constraints active at x^k.

We get

$$\begin{aligned} f(x^k + d) &= \frac{1}{2}(x^k + d)^T Q (x^k + d) + c^T (x^k + d) \\ &= \frac{1}{2} x^{kT} Q x^k + x^{kT} Q d + \frac{1}{2} d^T Q d + c^T x^k + c^T d\,. \end{aligned}$$

By defining the expressions

$$\begin{aligned} C_k &:= \frac{1}{2} x^{kT} Q x^k + c^T x^k\,, \\ g^k &:= \operatorname{grad} f(x^k) = Q x^k + c\,, \end{aligned}$$

which do not depend on d, we obtain:

$$f(x^k + d) = C_k + \frac{1}{2} d^T Q d + g^{kT} d\,.$$

Now the displacement d^k is the solution of the problem

$$\begin{aligned} \min \ & \frac{1}{2} d^T Q d + g^{kT} d \\ & a^{iT} d = 0 \quad \text{for } i \in \overline{A}(x^k) \quad \text{(see (8.19))}. \end{aligned}$$

Denoting by A_k the matrix of constraints active at x^k, we can rewrite the last problem as

$$\min \frac{1}{2} d^T Q d + g^{kT} d$$
$$A_k d = 0 . \tag{9.2}$$

and from (8.28) we obtain

$$d^k = -Q^{-1} g^k + Q^{-1} A_k^T (A_k Q^{-1} A_k^T)^{-1} A_k Q^{-1} g^k . \tag{9.3}$$

We use d^k as the search direction in the following algorithm to minimize (9.1). The step size is determined similar to Example 8.15. By definition of d^k, we get in particular

$$f(x^k + d^k) \le f(x^k + \lambda d^k) \quad \text{for all } \lambda \ge 0 ,$$

i.e. the value $\tilde{\lambda}_k := 1$ minimizes the one-dimensional convex function $f(x^k + \lambda d^k)$ in λ, defined over $[0, \infty)$. For $d^k \ne 0$ we now obtain the step size by $\lambda_k := \min(\tilde{\lambda}_k, \overline{\lambda}_k)$, implying

$$\lambda_k = \min \left\{ 1, \frac{b_i - a^{iT} x^k}{a^{iT} d^k} \,\middle|\, i \in \{1, \dots, s\} \setminus A(x^k) \,, \text{ with } a^{iT} d^k > 0 \right\} \tag{9.4}$$

and set $x^{k+1} = x^k + \lambda_k d^k$.

If the solution of (9.3) is null, we calculate the vector of Lagrange multipliers for problem (9.2), i.e.

$$v^k = -(A_k Q^{-1} A_k^T)^{-1} A_k Q^{-1} g^k \quad \text{(see (8.27))}. \tag{9.5}$$

If $v^k \ge 0$, the actual iteration point x^k is optimal. Otherwise we eliminate the line of A_k, corresponding to the smallest component of v^k and calculate d^k by means of the new matrix A_k.

We obtain the following *active set algorithm for the quadratic problem* (9.1):

(0) Determine a feasible point x^0, set $k = 0$.
(1) Determine d^k by (9.3). (If no active constraint exists, set $d^k = -g^k$).
(2) If $d^k \ne 0$, calculate λ_k by (9.4). Set $x^{k+1} = x^k + \lambda_k d^k$, $k = k + 1$ and go to step (1).
(3) If $d^k = 0$, calculate v^k by (9.5). If $v^k \ge 0$, stop. Otherwise eliminate the line of A_k, corresponding to the smallest component of v^k and go to step (1).

Example 9.1. We apply the algorithm to the problem

$$\min x_1^2 + (x_2 - 1)^2$$
$$x_2 \le 4 \tag{1}$$
$$-x_1 - x_2 \le -2 \tag{2}$$
$$x_1 \le 2 \tag{3}$$

starting with $x^0 = (2, 4)^T$ (see Figure 9.1).

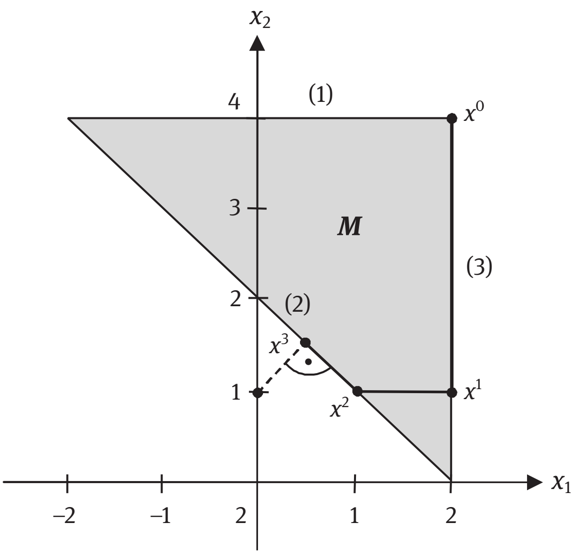


Fig. 9.1. Active set method.

Ignoring the additive constant, the objective function is of the form (9.1) with

$$Q = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, Q^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}, \quad c = \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \quad g^k = Qx^k + c = 2x^k - \begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

Iteration 1:

$$k = 0, \quad x^0 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \quad g^0 = \begin{pmatrix} 4 \\ 6 \end{pmatrix}, \quad A_0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$A_0 Q^{-1} A_0^T = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}, \quad (A_0 Q^{-1} A_0^T)^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},$$

$$\begin{aligned} d^0 &= -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 4 \\ 6 \end{pmatrix} \\ &\quad + \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 4 \\ 6 \end{pmatrix} \\ &= -\begin{pmatrix} 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

The result $d^0 = 0$ expresses the trivial fact that no direction can preserve the activity of both constraints (1) and (3).

$$v^0 = -\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 4 \\ 6 \end{pmatrix} = \begin{pmatrix} -6 \\ -4 \end{pmatrix}.$$

Since -6 is the smallest component of v^0, the first line of A_0 is eliminated. Thus

$$A_0 = (1, 0)\,, \quad A_0 Q^{-1} A_0^T = 1/2\,, \quad (A_0 Q^{-1} A_0^T)^{-1} = 2\,,$$

and the search direction becomes

$$\begin{aligned} d^0 &= -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 4 \\ 6 \end{pmatrix} + \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} 2(1, 0)\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 4 \\ 6 \end{pmatrix} \\ &= -\begin{pmatrix} 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -3 \end{pmatrix}. \end{aligned}$$

We obtain $\lambda_0 = 1$, $x^1 = x^0 + \lambda_0 d^0 = \binom{2}{1}$.

The point x^1 minimizes the objective function over the boundary line of constraint (3) (see Figure 9.1).

Iteration 2:

$$k = 1\,, \quad x^1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad g^1 = \begin{pmatrix} 4 \\ 0 \end{pmatrix}, \quad A_1 = (1, 0)$$

(compare with the previous iteration)

$$\begin{aligned} &(A_1 Q^{-1} A_1^T)^{-1} = 2\,, \\ &\qquad d^1 = -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \end{pmatrix} + \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} 2(1, 0)\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \end{pmatrix} \\ &\qquad\quad = -\begin{pmatrix} 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

We have to eliminate the unique active constraint, thus $d^1 = -g^1 = \binom{-4}{0}$.

We get $\lambda_1 = 1/4$ (see Figure 9.1), $x^2 = x^1 + \lambda_1 d^1 = \binom{1}{1}$.

Iteration 3:

$$\begin{aligned} &k = 2\,, \quad x^2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad g^2 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad A_2 = (-1, -1)\,, \quad (A_2 Q^{-1} A_2^T)^{-1} = 1\,, \\ d^2 &= -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} -1 \\ -1 \end{pmatrix} 1(-1, -1)\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \end{pmatrix} \\ &= -\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} = \begin{pmatrix} -1/2 \\ 1/2 \end{pmatrix}. \end{aligned}$$

We obtain $\lambda_2 = 1\,, \quad x^3 = x^2 + \lambda_2 d^2 = \binom{1/2}{3/2}$.

Iteration 4:

$$k = 3\,, \quad x^3 = \begin{pmatrix} 1/2 \\ 3/2 \end{pmatrix}, \quad g^3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad A_3 = (-1, -1), (A_3 Q^{-1} A_3^T)^{-1} = 1\,,$$

$$\begin{aligned} d^3 &= -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} -1 \\ -1 \end{pmatrix} 1(-1, -1)\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= -\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} + \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

We get

$$v^3 = -1(-1, -1)\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1 \geq 0\,,$$

and the algorithm terminates with the optimal solution x^3.

In practice, the search direction is not calculated by formula (9.3). Instead of this, problem (9.2) is resolved by techniques outlined in Section 8.2.

Exercise 9.2. *Apply the above algorithm to the problem*

$$\begin{aligned} \min (x_1 + 1)^2 + 2x_2^2 & \\ -x_1 + x_2 &\leq 2 \qquad (1) \\ -x_1 - x_2 &\leq -2 \qquad (2) \\ x_1 \quad &\leq 2 \qquad (3) \end{aligned}$$

starting with $x^0 = (2, 4)^T$.

9.2 Karush–Kuhn–Tucker conditions

In the remainder of this chapter we study the quadratic problem of the form

$$\begin{aligned} \min \frac{1}{2} x^T Q x + c^T x & \\ Ax \leq b & \qquad (9.6) \\ x \geq 0\,, & \end{aligned}$$

where $Q \in \mathbb{R}^{n \times n}$ is symmetric, $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$ (see (1.20) and Remark 1.10). This form is more adequate for the following transformations. It is known from linear programming that the constraints of (9.6) can be converted into that of (9.1) and vice versa.

Suppose that the matrix Q is positive semidefinite, i.e. (9.6) is a convex optimization problem. Since the constraints are linear, all feasible points are regular (see Definition 4.1 and Theorem 4.2). Thus the KKT conditions are necessary and sufficient for a global minimum point (see Theorem 4.11).

We write the constraints of (9.6) in the form

$$Ax - b \le 0 \tag{9.7}$$

$$-x \le 0 \tag{9.8}$$

and denote the vectors of Lagrange multipliers associated to (9.7) and (9.8) by $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, respectively. This yields the KKT conditions

$$\begin{aligned} Qx + c + A^T u - v &= 0 \\ Ax - b &\le 0 \\ u^T(Ax - b) &= 0 \\ v^T x &= 0 \\ u, v, x &\ge 0 \quad \text{(see Exercise 9.3).} \end{aligned} \tag{9.9}$$

By introducing the vector of slack variables $y = b - Ax \in \mathbb{R}^m$ we get

$$\begin{aligned} v - Qx - A^T u &= c \\ y + Ax &= b \\ u^T y &= 0 \\ v^T x &= 0 \\ u, v, x, y &\ge 0\,. \end{aligned} \tag{9.10}$$

Since the vectors u, v, x and y are nonnegative, the third and fourth equation in (9.10) can be substituted by the single condition $u^T y + v^T x = 0$.

Finally, by defining

$$w = \begin{pmatrix} v \\ y \end{pmatrix}, \quad z = \begin{pmatrix} x \\ u \end{pmatrix}, \quad r = \begin{pmatrix} c \\ b \end{pmatrix} \in \mathbb{R}^{n+m}$$

and

$$R = \begin{pmatrix} Q & A^T \\ -A & 0 \end{pmatrix} \in \mathbb{R}^{(n+m)\times(n+m)},$$

the KKT conditions can be written in the condensed form

$$\begin{aligned} w - Rz &= r \\ w^T z &= 0 \\ w, z &\ge 0\,. \end{aligned} \tag{9.11}$$

The nonlinear condition $w^T z = 0$ is called *complementarity condition*. Since $w, z \ge 0$, it means that at least one of the variables w_i and z_i is zero for all $i = 1, \ldots, n + m$. The problem of determining a solution of (9.11) is called the *linear complementarity problem*.

Exercise 9.3. *Show step by step that the KKT conditions of problem (9.6) have the form (9.9). (Compare with Exercise 8.20!)*

9.3 Lemke's method

This method for solving the linear complementarity problem (9.11) is based on the simplex method for linear programming. The KKT conditions are solved by means of an adequate pivoting strategy. At first an auxiliary variable $z_0 \in \mathbb{R}$ and a vector $e = (1, \dots, 1)^T \in \mathbb{R}^{n+m}$ are introduced in system (9.11):

$$\begin{aligned} w - Rz - z_0 e &= r \\ w^T z &= 0 \\ w, z, z_0 &\geq 0 \,. \end{aligned} \tag{9.12}$$

Writing (9.12) in tableau form, we obtain

Table 9.1. Initial tableau of Lemke's method.

BV	z_0	z_1	$\dots$	z_{n+m}	w_1	$\dots$	w_{n+m}	r
w_1	$-r_{1,0}$	$-r_{1,1}$	$\dots$	$-r_{1,n+m}$	1			r_1
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\ddots$		$\vdots$
w_{n+m}	$-r_{n+m,0}$	$-r_{n+m,1}$	$\dots$	$-r_{n+m,n+m}$			1	r_{n+m}

where the elements $r_{i,0}$ denote the components of the vector e, the $r_{i,j}$ denote the elements of the matrix R, and the r_i are the components of the vector $r (i = 1, \dots, n + m)$.

The variables are $z_0, z_1, \dots z_{n+m}, w_1, \dots w_{n+m}$. The first column contains the basic variables (BV), which initially are $w_1, \dots w_{n+m}$. The basic solution corresponding to this tableau is

$$(z_0, z_1, \dots z_{n+m}, w_1, \dots w_{n+m}) = (0, \dots, 0, r_1, \dots, r_{n+m}) \,,$$

which obviously satisfies the complementarity condition.

If $r_i \geq 0$ holds for $i = 1, \dots, n+m$, then the basic solution is feasible and provides a solution of the KKT conditions (9.11).

Otherwise a solution of (9.11) is obtained by performing a sequence of pivot steps in the following way:

By selecting z_0 as the entering variable, a first feasible tableau is constructed such that the basic solution satisfies the complementarity condition. In the following, pivot steps are performed that preserve feasibility as well as complementarity of the respective basic solutions. The process ends, when z_0 leaves the basis. In this case, $z_0 = 0$ holds and the basic solution satisfies (9.11). We obtain the following algorithm.

Lemke's algorithm
I First pivot step:
Choose z_0 as the pivot column. Determine the pivot row k by

$$r_k = \min \{ r_i \,|\, i = 1, \dots, n + m \} \,.$$

Perform a pivot step.

II Following iterations:
If z_0 has left the basis in the previous step, then the actual basic solution is optimal and the algorithm ends. Otherwise denote by $j \in \{1, \dots, n+m\}$ the index of the variable that left the basis in the preceding iteration, i.e. z_j or w_j has left the basis. In the first case, choose w_j as the pivot column, otherwise z_j must be the pivot column. If all elements of the pivot column are nonpositive, the KKT conditions have no solution, i.e. problem (9.6) has no bounded optimal solution (see Bazaraa and Shetty (1979, p. 452, Theorem 11.2.4)) and the algorithm ends (see Example 9.5).

Otherwise, choose the pivot row k according to the leaving variable criterion of the simplex method, i.e. such that

$$\frac{r_k}{r_{k\bullet}} = \min \left\{ \frac{r_i}{r_{i\bullet}} \, | r_{i\bullet} > 0 \right\} , \tag{9.13}$$

where $r_{i\bullet}$ and r_i denote the actual values of right side and pivot column, respectively. Perform a pivot step.

The well-known criterion (9.13) ensures that each tableau is feasible. The pivot column is selected such that in each iteration at least one of the variables z_i and w_i is nonbasic for all $i \in \{1, \dots, n+m\}$. Since nonbasic variables are always zero, any basic solution constructed by the algorithm satisfies the complementarity condition.

If each basic solution is nondegenerate (i.e. all basic variables are positive) and if the matrix Q satisfies certain conditions (e.g. Q positive definite), Lemke's algorithm ends after a finite number of iterations (see, e.g. Bazaraa and Shetty (1979, p. 452)). The procedure terminates in two cases, illustrated by the following examples: When a solution of the KKT conditions is encountered, or when a nonpositive column is found, i.e. no bounded optimal solution exists.

Example 9.4. Consider the problem

$$\begin{gathered} \min x_1^2 + 2x_2^2 + \frac{1}{3}x_3^2 + 2x_1x_2 - \frac{1}{2}x_1x_3 - 4x_1 + x_2 - x_3 \\ x_1 + 2x_2 - x_3 \le 4 \\ x_1, x_2, x_3 \ge 0 \, . \end{gathered}$$

It holds

$$Q = \begin{pmatrix} 2 & 2 & -1/2 \\ 2 & 4 & 0 \\ -1/2 & 0 & 2/3 \end{pmatrix} \quad (Q \text{ is positive definite}) \, , \quad c = \begin{pmatrix} -4 \\ 1 \\ -1 \end{pmatrix} ,$$

$$A = (1, 2, -1) \, , \quad b = 4 \, ,$$

$$R = \begin{pmatrix} Q & A^T \\ -A & 0 \end{pmatrix} = \begin{pmatrix} 2 & 2 & -1/2 & 1 \\ 2 & 4 & 0 & 2 \\ -1/2 & 0 & 2/3 & -1 \\ -1 & -2 & 1 & 0 \end{pmatrix} , \quad r = \begin{pmatrix} c \\ b \end{pmatrix} = \begin{pmatrix} -4 \\ 1 \\ -1 \\ 4 \end{pmatrix} .$$

Table 9.2. Pivot steps of Example 9.4.

BV	z_0	z_1	z_2	z_3	z_4	w_1	w_2	w_3	w_4	r
w_1	-1	-2	-2	$1/2$	-1	1	0	0	0	-4
w_2	-1	-2	-4	0	-2	0	1	0	0	1
w_3	-1	$-1/2$	0	$-2/3$	1	0	0	1	0	-1
w_4	-1	1	2	-1	0	0	0	0	1	4
z_0	1	2	2	$-1/2$	1	-1	0	0	0	4
w_2	0	0	-2	$-1/2$	-1	-1	1	0	0	5
w_3	0	$5/2$	2	$-7/6$	-2	-1	0	1	0	3
w_4	0	3	4	$-3/2$	1	-1	0	0	1	8
z_0	1	0	$2/5$	$13/30$	$-3/5$	$-1/5$	0	$-4/5$	0	$8/5$
w_2	0	0	-2	$-1/2$	-1	-1	1	0	0	5
z_1	0	1	$4/5$	$-7/15$	$4/5$	$-2/5$	0	$2/5$	0	$6/5$
w_4	0	0	$8/5$	$-1/10$	$-7/5$	$1/5$	0	$-6/5$	1	$22/5$
z_3	$30/13$	0	$12/13$	1	$-18/13$	$-6/13$	0	$-24/13$	0	$48/13$
w_2	$15/13$	0	$-20/13$	0	$-22/13$	$-16/13$	1	$-12/13$	0	$89/13$
z_1	$14/13$	1	$16/13$	0	$2/13$	$-8/13$	0	$-6/13$	0	$38/13$
w_4	$3/13$	0	$22/13$	0	$-20/13$	$2/13$	0	$-18/13$	1	$62/13$

The solution process is shown in Table 9.2. In the first iteration z_0 enters the basis, yielding a feasible tableau, i.e. the right side is nonnegative. Since w_1 has left the basis in the first iteration, z_1 must enter in the next iteration, i.e. z_1 is the pivot column. Since

$$\frac{3}{5/2} = \min\left\{\frac{4}{2}, \frac{3}{5/2}, \frac{8}{3}\right\},$$

the pivot row is w_3. In the second iteration, w_3 leaves the basis, thus z_3 is the pivot column in the third iteration. The pivot row is then z_0, since $13/30$ is the only positive element in the column. Since z_0 leaves the basis in the third iteration, the fourth tableau is optimal. The corresponding basic solution is

$$(z_0, z_1, z_2, z_3, z_4, w_1, w_2, w_3, w_4) = (0, 38/13, 0, 48/13, 0, 0, 89/13, 0, 62/13)\,.$$

Since $z = \binom{x}{u}$ and $w = \binom{v}{y}$, the solution of the KKT conditions is

$$\begin{aligned} x &= (z_1, z_2, z_3)^T = (38/13, 0, 48/13)^T\,, & u &= z_4 = 0\,, \\ v &= (w_1, w_2, w_3)^T = (0, 89/13, 0)^T\,, & y &= w_4 = 62/13\,. \end{aligned}$$

The solution of the quadratic problem is $x = (38/13, 0, 48/13)^T$.

We now consider a problem with an unbounded optimal solution.

Example 9.5. Applying Lemke's method to the problem

$$\begin{aligned} \min 4x_1^2 - 4x_1x_2 + x_2^2 - 3x_1 - x_2, \\ x_1 - x_2 \le 1 \\ -3x_1 + x_2 \le 2 \\ x_1, x_2 \ge 0 \quad \text{(see Table 9.3)}, \end{aligned}$$

we obtain

$$Q = \begin{pmatrix} 8 & -4 \\ -4 & 2 \end{pmatrix}, \; A = \begin{pmatrix} 1 & -1 \\ -3 & 1 \end{pmatrix}, \; R = \begin{pmatrix} 8 & -4 & 1 & -3 \\ -4 & 2 & -1 & 1 \\ -1 & 1 & 0 & 0 \\ 3 & -1 & 0 & 0 \end{pmatrix}, \; r = \begin{pmatrix} c \\ b \end{pmatrix} = \begin{pmatrix} -3 \\ -1 \\ 1 \\ 2 \end{pmatrix}.$$

Since w_2 leaves the basis in the second iteration, the column z_2 should be the next pivot column. But this column has only nonpositive elements, i.e. the algorithm detects that no bounded optimal solution exists.

This fact can be interpreted geometrically: the second line of the last tableau shows that selecting z_2 as the entering variable would generate a feasible point $(x_1, x_2)^T$ with $x_2 \ge 0$ and $x_1 - (1/2)x_2 = 1/6$ (note that $z_1 = x_1$ and $z_2 = x_2$). The set of these points is the ray $R = \{\binom{\alpha}{2\alpha - 1/3} \mid \alpha \ge 1/6\}$ in Figure 9.2. By increasing α, the objective value can be made arbitrarily small.

Exercise 9.6. *Solve the quadratic problem*

$$\begin{aligned} \min x_1^2 + \frac{3}{2}x_2^2 + \frac{1}{2}x_3^2 + x_1x_2 + x_2x_3 - x_1 + 5x_2 - 2x_3 \\ 2x_1 + 3x_2 - x_3 \le 5 \\ x_1, x_2, x_3 \ge 0 \end{aligned}$$

with Lemke's method.

Table 9.3. Pivot steps of Example 9.5.

BV	z_0	z_1	z_2	z_3	z_4	w_1	w_2	w_3	w_4	r
w_1	-1	-8	4	-1	3	1	0	0	0	-3
w_2	-1	4	-2	1	-1	0	1	0	0	-1
w_3	-1	1	-1	0	0	0	0	1	0	1
w_4	-1	-3	1	0	0	0	0	0	1	2
z_0	1	8	-4	1	-3	-1	0	0	0	3
w_2	0	12	-6	2	-4	-1	1	0	0	2
w_3	0	9	-5	1	-3	-1	0	1	0	4
w_4	0	5	-3	1	-3	-1	0	0	1	5
z_0	1	0	0	$-1/3$	$-1/3$	$-1/3$	$-2/3$	0	0	$5/3$
z_1	0	1	$-1/2$	$1/6$	$-1/3$	$-1/12$	$1/12$	0	0	$1/6$
w_3	0	0	$-1/2$	$-1/2$	0	$-1/4$	$-3/4$	1	0	$5/2$
w_4	0	0	$-1/2$	$1/6$	$-4/3$	$-7/12$	$-5/12$	0	1	$25/6$

Exercise 9.7. *Consider a simple variant of the location problem of Example 1.18. Given are three gas stations C_1, C_2, and C_3 with coordinates $(a_1, b_1) = (0, 4)$, $(a_2, b_2) = (3, 4)$ and $(a_3, b_3) = (3, -1)$ and equal demands $q_1 = q_2 = q_3$ (see Figure 9.3). The transportation costs between the refinery D and the gas station C_i increases quadratically with the Euclidean distance, i.e.*

$$f_i(x_1, x_2) = \alpha[(x_1 - a_i)^2 + (x_2 - b_i)^2] \quad \textit{for } i = 1, 2, 3\,,$$

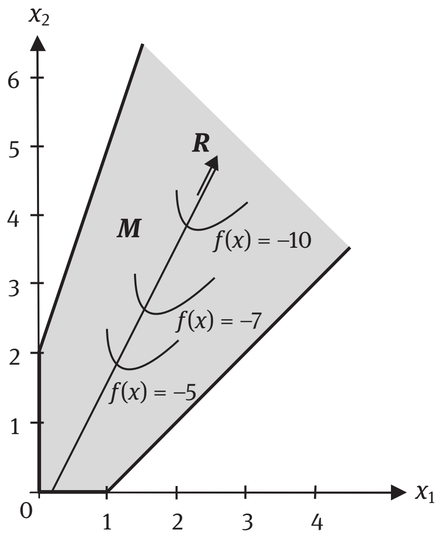


Fig. 9.2. Unbounded optimal solution.

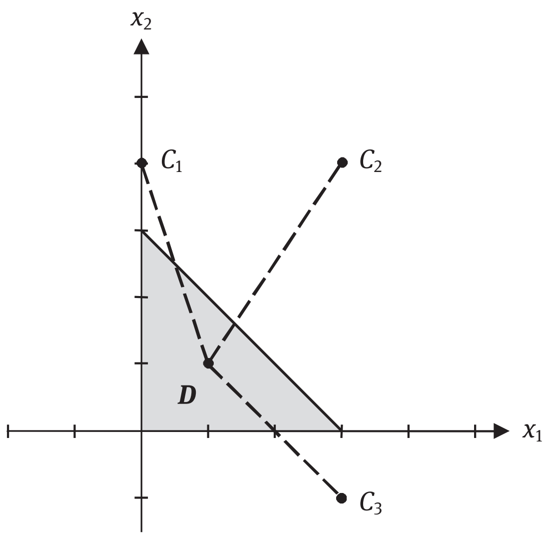


Fig. 9.3. Locations of gas stations.

where x_1, x_2 denote the coordinates of the refinery D, which must be located in the triangle shown in Figure 9.3. Determine x_1, x_2 such that the total costs $\sum_{i=1}^{3} f_i(x_1, x_2)$ are minimized. Formulate a problem of type (9.6) and apply Lemke's method!

10 The general problem

In this chapter we study three approaches to solve the general problem

$$\min_{x \in M} f(x) \tag{10.1}$$

with $M \subset \mathbb{R}^n$. The penalty method (Section 10.1) and the barrier method (Section 10.2) aim to construct a sequence of problems, "simpler" than (10.1), such that the minimum points of these problems converge to the minimum point of (10.1). The sequential quadratic programming (Section 10.3) is an iterative procedure that determines the search direction with the aid of quadratic problems.

10.1 The penalty method

Using the indicator function

$$\delta_M(x) := \begin{cases} 0 & \text{for } x \in M \\ \infty & \text{for } x \neq M\,, \end{cases}$$

we can rewrite (10.1) as an unrestricted problem

$$\min_{x \in \mathbb{R}^n} f(x) + \delta_M(x)\,. \tag{10.2}$$

Interpreting $f(x)$ economically as a cost to be minimized, the term $\delta_M(x)$ can be understood as an "infinite penalty" associated with the exit from the feasible region. However, model (10.2) cannot be used practically, not even by substituting "∞" by a "large" positive number. Therefore, $\delta_M(x)$ is approximated by a so-called *penalty term* $p(x)$. For a feasible set in the form of

$$M = \left\{x \in \mathbb{R}^n \,|g_i(x) \leq 0 \text{ for } i = 1, \ldots, m; h_j(x) = 0 \text{ for } j = 1, \ldots, l\right\} \tag{10.3}$$

it is common to define

$$p(x) = \sum_{i=1}^{m} [g_i^+(x)]^\beta + \sum_{j=1}^{l} [h_j(x)]^\beta\,, \tag{10.4}$$

where g_i^+ is the *positive part* of the function g_i, defined by

$$g_i^+(x) = \begin{cases} g_i(x) & \text{for } g_i(x) \geq 0\,, \\ 0 & \text{for } g_i(x) < 0\,. \end{cases}$$

In practice, the parameter β has usually the values 2 or 4. If the functions g_i and h_j are twice continuously differentiable, then p is continuously differentiable for $\beta = 2$

and twice continuously differentiable for $\beta = 4$ (see Exercise 10.10). In the following examples we restrict ourselves to the case $\beta = 2$. From Definition (10.4) it follows that $p(x) = 0$ for $x \in M$ and $p(x) > 0$ for $x \notin M$. For $r > 0$ it holds that $rp(x) = \delta_M(x)$ for $x \in M$, and for $x \notin M$ the term $rp(x)$ approximates $\delta_M(x)$ better, the larger r is.

Example 10.1. For the feasible region $M = [a, b]$ with $a < b$ we obtain the constraints

$$\begin{aligned} g_1(x) &:= a - x \leq 0\,, \\ g_2(x) &:= x - b \leq 0\,. \end{aligned}$$

Therefore

$$g_1^+(x) = \begin{cases} a - x & \text{for } x \leq a\,, \\ 0 & \text{for } x > a\,, \end{cases}$$

$$g_2^+(x) = \begin{cases} x - b & \text{for } x \geq b\,, \\ 0 & \text{for } x < b\,, \end{cases}$$

$$p(x) = [g_1^+(x)]^2 + [g_2^+(x)]^2 = \begin{cases} (a - x)^2 & \text{for } x \leq a\,, \\ 0 & \text{for } a < x < b\,, \\ (b - x)^2 & \text{for } x \geq b\,. \end{cases}$$

The penalty term $rp(x)$ is illustrated in Figure 10.1 for $a = 4$, $b = 8$ and $r = 1, 10, 100$.

The solution of problem (10.1) is obtained, by solving the unrestricted problems

$$\min_{x \in \mathbb{R}^n} q(x, r) \tag{10.5}$$

$$\text{with} \quad q(x, r) := f(x) + rp(x)$$

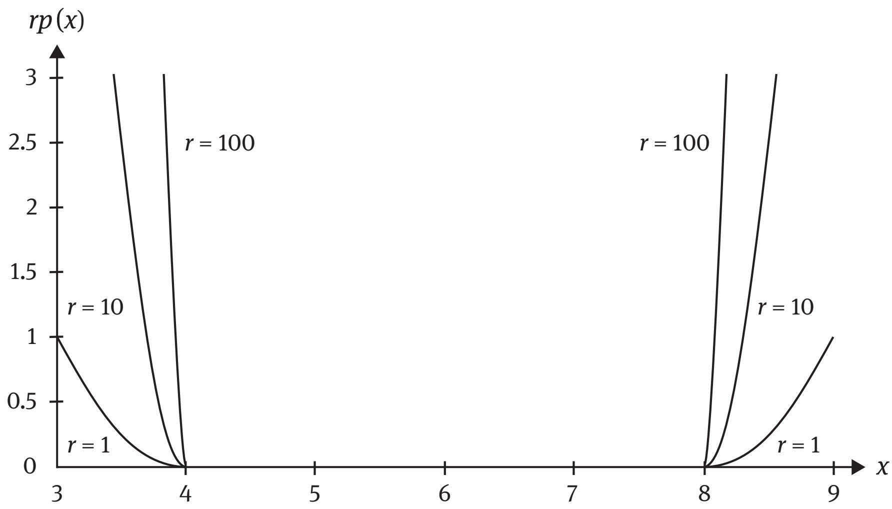


Fig. 10.1. Penalty terms.

for $r = r_0, r_1, r_2, \ldots$, where $(r_i)_{i=0,1,2,\ldots}$ is an increasing sequence of positive values with $lim_{i\to\infty} r_i = \infty$. One can choose, e.g. $r_0 > 0$ and $r_i := 10^i r_0$ for $i = 1, 2, \ldots$

Let x^0 be a global minimum point of (10.5) for $r = r_0$. If $p(x^0) = 0$, it holds $x^0 \in M$ and

$$f(x^0) = q(x^0, r_0) \le q(x, r_0) = f(x) \quad \text{for all } x \in M\,,$$

i.e. x^0 is optimal for (10.1). If $p(x^0) > 0$, we solve (10.5) for $r = r_1$. If the global minimum point x^1 satisfies $p(x^1) = 0$, then x^1 is optimal for (10.1). Otherwise we solve (10.5) for $r = r_2$, etc. We obtain the following basic procedure.

Penalty method

(0) Choose a scalar $\varepsilon > 0$ and a number $r_0 > 0$. Set $k = 0$.

(1) Determine a solution x^k of problem

$$\min_{x\in\mathbb{R}^n} q(x, r_k)\,.$$

(2) If $r_k p(x^k) \le \varepsilon$, stop ($x^k$ provides an approximation for a global minimum point of (10.1)). Otherwise go to step (3).

(3) Set $r_{k+1} = 10 r_k$, $k = k + 1$ and go to step (1).

The stopping criterion in step (2) is based on the fact that $r_k p(x^k)$ converges to zero, if $k \to \infty$ (see, e.g. Bazaraa and Shetty (1979, Section 9.2)).

As already said above, $q(x, r)$ is continuously differentiable, if f, g_i and h_j have this property. In this case, a great part of the methods for unrestricted minimization can be used to solve the problems of type (10.5). Similarly, $q(x, r)$ is convex, if f, g_i and h_j are convex (see Exercise 10.10). There exist many variants of penalty methods. Before discussing some of them, we apply the above method to some examples.

Example 10.2. The global minimum point of

$$\begin{aligned} &\min 2x^2 - x \\ &x \ge 1\,, \end{aligned} \tag{10.6}$$

is obviously $x^* = 1$. We obtain the constraint $g(x) = 1 - x \le 0$, therefore

$$g^+(x) = \begin{cases} 1 - x & \text{for } x \le 1\,, \\ 0 & \text{for } x > 1\,, \end{cases}$$

and

$$q(x, r) = f(x) + r(g^+(x))^2 = \begin{cases} 2x^2 - x + r(1-x)^2 & \text{for } x \le 1\,, \\ 2x^2 - x & \text{for } x > 1\,, \end{cases}$$

(see Figure 10.2).

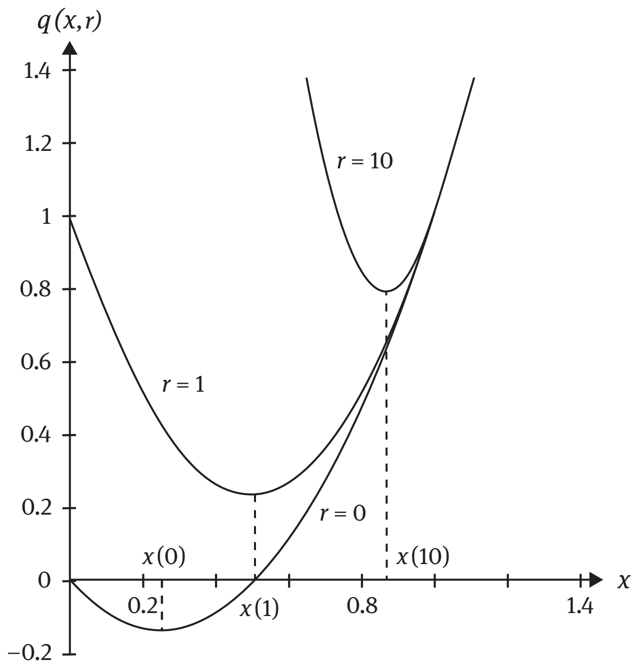


Fig. 10.2. Function $q(x,r)$ of Example 10.2.

Since $q(x,r)$ is increasing for $x \geq 1$, the solution of the problem $\min_{x\in\mathbb{R}} q(x,r)$ is given by

$$\frac{\partial}{\partial x}\left(2x^2 - x + r(1-x)^2\right) = 0 \Rightarrow$$
$$4x - 1 + 2r(x-1) = 0\,.$$

Therefore, $x(r) = \frac{2r+1}{2r+4}$. Since $x(r) < 1$ for all $r > 0$ and $\lim_{r\to\infty} x(r) = 1$, the penalty method generates a sequence of infeasible points of (10.6) that converges to the optimal point $x^* = 1$.

Example 10.3. Consider the problem

$$\min_{x\geq 0} f(x) \tag{10.7}$$

with

$$f(x) = \begin{cases} (x+1)^2 - 1 & \text{for } x < 0\,, \\ x(x-1)(x-2) & \text{for } x \geq 0\,. \end{cases}$$

It is easily verified that f is continuously differentiable over $\mathbb{R}$ and that the optimal point of (10.7) is $x^* = 1 + 1/\sqrt{3}$ (see Figure 10.3). The constraint is $g(x) := -x \leq 0$.

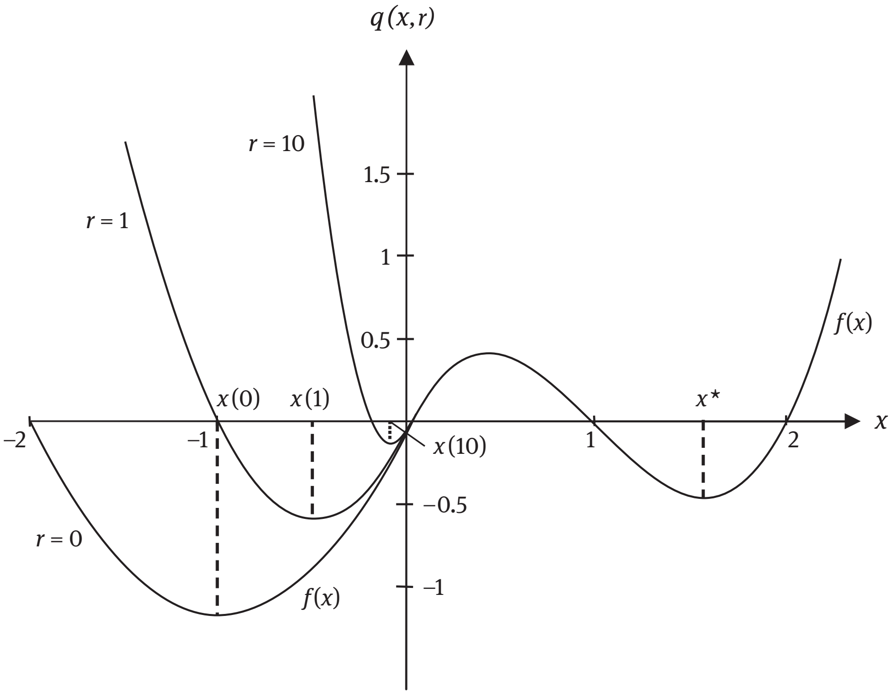


Fig. 10.3. Function $q(x, r)$ of Example 10.2.

Therefore,

$$g^+(x) = \begin{cases} -x & \text{for } x < 0\,, \\ 0 & \text{for } x \geq 0\,, \end{cases}$$

$$q(x, r) = f(x) + r(g^+(x))^2 = \begin{cases} (x+1)^2 - 1 + rx^2 & \text{for } x < 0, \\ x(x-1)(x-2) & \text{for } x \geq 0. \end{cases}$$

The function q is illustrated in Figure 10.3 for some values of r.

For a given $r > 0$, q has two local minimum points: the point $x^* = 1 + 1/\sqrt{3}$ and the point $x(r)$, given by

$$\frac{\partial}{\partial x}[(x+1)^2 - 1 + rx^2] = 2(x+1) + 2rx = 0\,,$$

i.e. $x(r) = -1/(r+1)$. The values of q at these points are

$$q(x(r), r) = \left(-\frac{1}{r+1} + 1\right)^2 - 1 + r\frac{1}{(r+1)^2} = -\frac{1}{r+1}\,,$$

$$q(x^*, r) = x^*(x^* - 1)(x^* - 2) = -\frac{2}{3\sqrt{3}} \approx -0.385\,.$$

Hence, for

$$-\frac{1}{r+1} > -\frac{2}{3\sqrt{3}} \Leftrightarrow$$
$$r > \frac{3\sqrt{3}}{2} - 1 \approx 1.6\,, \tag{10.8}$$

the point x^* is a global minimum point of the function q. Otherwise, $x(r)$ minimizes q.

We apply the above algorithm, starting with $r_0 = 1$:

Iteration 0:
The problem $\min_{x\in\mathbb{R}} q(x, r_0)$ has the global minimum $x(1) = -1/2$ (Figure 10.3). Since $p(-1/2) = [g^+(-1/2)]^2 = 1/4 > 0$, the algorithm continues with the next step.

Iteration 1:
Since $r_1 = 10$ satisfies (10.8), the problem $\min_{x\in\mathbb{R}} q(x, r_1)$ has the global minimum point x^* (see Figure 10.3). It holds $p(x^*) = 0$, and the algorithm terminates, providing the solution $x^* = 1 + 1/\sqrt{3} \approx 1.577$ of problem (10.7).

We consider another problem with two variables.

Example 10.4. The solution of

$$\begin{aligned} &\min (x_1+1)^2 + (x_2+2)^2 \\ &x_1 \geq 1 \\ &x_2 \geq 2 \end{aligned} \tag{10.9}$$

is obviously $x^* = (1, 2)^T$.

We get

$$\begin{aligned} g_1(x_1, x_2) &:= 1 - x_1\,, \\ g_2(x_1, x_2) &:= 2 - x_2\,, \\ q(x_1, x_2, r) &= f(x_1 x_2) + r[(g_1^+(x_1, x_2))^2 + (g_2^+(x_1, x_2))^2] \\ &= \begin{cases} (x_1+1)^2 + (x_2+2)^2 + r(1-x_1)^2 + r(2-x_2)^2 & \text{for } x_1 \leq 1, x_2 \leq 2\,, \\ (x_1+1)^2 + (x_2+2)^2 + r(1-x_1)^2 & \text{for } x_1 \leq 1, x_2 > 2\,, \\ (x_1+1)^2 + (x_2+2)^2 + r(2-x_2)^2 & \text{for } x_1 > 1, x_2 \leq 2\,, \\ (x_1+1)^2 + (x_2+2)^2 & \text{for } x_1 > 1, x_2 > 2\,. \end{cases} \end{aligned}$$

Since q is continuously differentiable and convex, a global minimum point is obtained by solving $\operatorname{grad}_x q(x_1, x_2, r) = 0$.

It holds

$$\frac{\partial}{\partial x_1} q(x_1, x_2, r) = \begin{cases} 2(x_1 + 1) + 2r(x_1 - 1) & \text{for } x_1 \le 1\,, \\ 2(x_1 + 1) & \text{for } x_1 > 1\,, \end{cases}$$

$$\frac{\partial}{\partial x_2} q(x_1, x_2, r) = \begin{cases} 2(x_2 + 2) + 2r(x_2 - 2) & \text{for } x_2 \le 2, \\ 2(x_2 + 2) & \text{for } x_2 > 2. \end{cases}$$

Therefore, $\operatorname{grad}_x q(x_1, x_2, r) = 0$ is equivalent to the system

$$(x_1 + 1) + r(x_1 - 1) = 0\,,$$
$$(x_2 + 2) + r(x_2 - 2) = 0\,,$$

which has the unique solution

$$x(r) = \left(\frac{r-1}{r+1}, \frac{2r-2}{r+1} \right)^T .$$

The points $x(r)$ are not feasible for (10.9) but converge to the optimal point $x^* = (1, 2)^T$ if $r \to \infty$ (see Figure 10.4).

Despite its simplicity, the penalty method is successfully applied in practice. Frequently the optimal point x^* of (10.1) belongs to the boundary of the feasible region M (see Examples 10.2 and 10.4), and the algorithm generates a sequence of infeasible

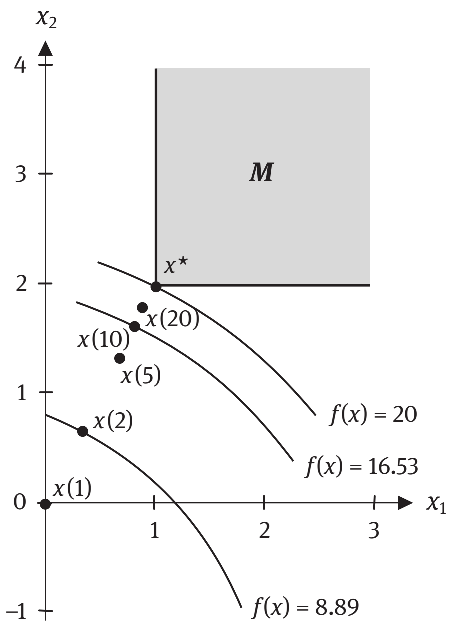


Fig. 10.4. Penalty method.

points converging to x^*. If x^* belongs to the interior of M, the algorithm may reach x^* after a finite number of iterations (see Example 10.3). In each case the iteration points $x^0, x^1, \ldots$ are outside of the feasible region. Therefore the above method is also known as the *external penalty method*. The convergence is discussed in the following theorem without a proof.

Theorem 10.5. *Given the problem (10.1), where $M \neq \emptyset$ has the form (10.3). Assume that the functions f, g_i, h_j are continuously differentiable and $f(x) \to \infty$ for $|x| \to \infty$. Then it holds*

(a) *Problem (10.5) has a global minimum point for all $r > 0$.*
(b) *The sequence of iteration points $x^0, x^1, \ldots$ converges to a global minimum point of (10.1).*
(c) *If x^* is the unique minimum point of f over M and if x^* is regular for (10.1), then the expressions*
$$u_i^{(k)} := \beta r_k [g_i^+(x^k)]^{\beta-1}$$
and
$$v_j^{(k)} := \beta r_k [h_j^+(x^k)]^{\beta-1}$$
converge to the respective Lagrange multipliers at x^*.

We illustrate part (c) of the theorem with the aid of Example 10.2.

Example 10.6. The KKT conditions of problem (10.6)

$$\begin{aligned} 4x - 1 - u &= 0 \\ u(1-x) &= 0 \\ 1 - x &\leq 0 \\ u &\geq 0 \end{aligned}$$

have the unique solution $(x, u) = (1, 3)$ (see Theorem 4.6). From Example 10.2 we know that
$$x^k = \frac{2r_k + 1}{2r_k + 4} < 1 \quad \text{for all } k\,.$$
Therefore,
$$g^+(x^k) = 1 - x^k = \frac{3}{2r_k + 4}$$
and
$$u^{(k)} = \beta r^k [g^+(x^k)]^{\beta-1} = 2r_k \frac{3}{2r_k + 4} = \frac{3r_k}{r_k + 2}\,,$$
which converges to $u = 3$ for $k \to \infty$.

As already mentioned, there exist many variants of the penalty method (see Section 11.3). Among others, alternative penalty terms have been developed to improve the efficiency of the procedure. For example, instead of a unique parameter r in the definition of the function $q(x, r)$ (see (10.4), (10.5)), one can associate a parameter to each constraint, i.e. instead of

$$q(x, r) = f(x) + r\{\sum_{i=1}^{m}[g_i^+(x)]^\beta + \sum_{j=1}^{l}[h_j(x)]^\beta\},$$

the penalty term can be defined as

$$q(x, r_1, \dots, r_{m+l}) = f(x) + \sum_{i=1}^{m} r_i[g_i^+(x)]^\beta + \sum_{j=1}^{l} r_{m+j}[h_j(x)]^\beta .$$

In the latter case a parameter r_i will be increased when the corresponding constraint g_i or h_j satisfies $g_i(x^k) > \varepsilon$ or $|h_j(x^k)| > \varepsilon$, respectively. The algorithm terminates when all values $g_1(x^k), \dots, g_m(x^k), |h_1(x^k)|, \dots, |h_l(x^k)|$ are smaller or equal than ε.

Frequently linear constraints are not included in the function $p(x)$ of (10.4). In this case, q(x, r) in (10.5) is minimized subject to these constraints.

In step (3) of the algorithm, r_{k+1} can be defined by $r_{k+1} = \alpha r_k$, where $\alpha > 1$ may be arbitrary. However, to limit the number of necessary iterations, α should be sufficiently large.

The above penalty method is not iterative in the strict sense, i.e. to calculate the point x^{k+1}, the previous point x^k need not be known. An evident idea would be to solve the unrestricted problem (10.5) only once for a large value of r. But this is not done in practice, since the minimization of $q(x, r)$ would be numerically unstable.

If in step (1) only a local minimum can be determined instead of a global, then the procedures converge to a local minimum of f over M.

Exercise 10.7.
(a) *How many iterations are necessary to solve the problem of Example 10.2 for* $r_0 = 1$ *and* $\varepsilon = 10^{-9}$? *Determine the iteration points* x^k *and the values* $p(x^k)$.
(b) *Answer the preceding question for the problem of Example 10.4.*

Exercise 10.8. *Solve the following problems by the penalty method:*
(a) $\min (x_1 - 1)^2 + x_2^2,$
$x_2 \geq x_1 + 1.$
(b) $\min x_1 + x_2,$
$x_2 = x_1^2.$

Exercise 10.9. *Verify the statement (c) of Theorem 10.5*
(a) *for problem (10.7),*
(b) *for problem (10.9).*

Exercise 10.10.
(a) *Consider the function $p(x)$ in (10.4) with $\beta = 2$. Show that $p(x)$ is continuously differentiable, if all functions g_i and h_j have this property.*
(b) *Show that the function $q(x, r)$ in (10.5) is convex, if all functions f, g_i and h_j are convex.*

We finally discuss briefly the so-called exact penalty functions. In general, the above penalty method cannot solve problem (10.1) in a finite number of steps (see Examples 10.2 and 10.4). A good approximation of the solution is only ensured if the value r in (10.5) is sufficiently large. But this may cause numerical instability.

However, the penalty term $p(x)$ in (10.4) can be chosen such that for sufficiently large r, each local minimum of (10.5) (satisfying certain conditions, detailed below) is also a local minimum of (10.1). In this case, $q(x, r)$ is called an *exact penalty function*.

The following theorem states that

$$q(x, r) = f(x) + rp(x)\,, \tag{10.10 a}$$

with

$$p(x) = \sum_{i=1}^{m} g_i^+(x) + \sum_{j=1}^{l} |h_j(x)| \tag{10.10 b}$$

is an exact penalty function.

Theorem 10.11. *Given the problem (10.1) with M defined as in (10.3) and a point $x^* \in M$ satisfying the second-order sufficient optimality conditions (see Theorem 4.46). Let u_i^* and v_j^* be the Lagrange multipliers associated with the inequalities and equations, respectively. Then x^* is a local minimum point of the function $q(x, r)$ in (10.10), if $r > C := \max\{u_1^*, \ldots, u_m^*, |v_1^*|, \ldots, |v_l^*|\}$.*

The statement is illustrated in the following example.

Example 10.12. Consider again the problem

$$\begin{aligned} &\min 2x^2 - x \\ &\quad x \geq 1 \qquad \text{(see Example 10.2)}\,. \end{aligned} \tag{10.11}$$

The solution is $x^* = 1$ and the KKT conditions

$$\begin{aligned} 4x - 1 - u &= 0 \\ u(1 - x) &= 0 \\ 1 - x &\leq 0 \\ u &\geq 0 \end{aligned}$$

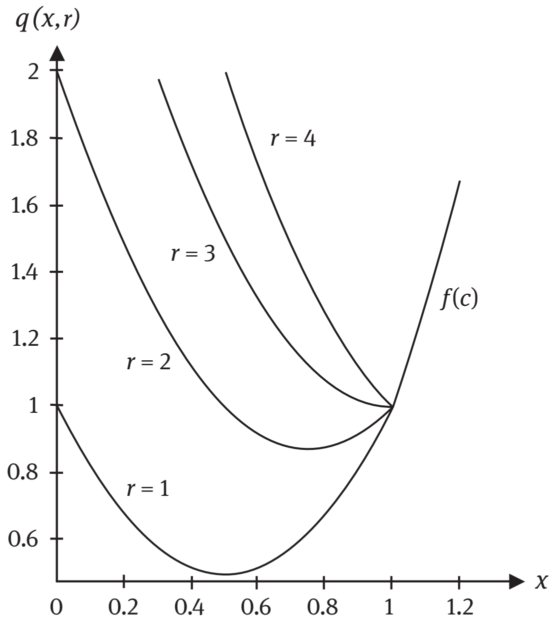


Fig. 10.5. Exact penalty function (10.12).

have the unique solution $(x^*, u^*) = (1, 3)$, therefore $C = 3$. The exact penalty function (10.10) is (see Figure 10.5):

$$q(x,r) = 2x^2 - x + rg^+(x) = \begin{cases} 2x^2 - x + r(1-x) & \text{for } x \leq 1\,, \\ 2x^2 - x & \text{for } x > 1\,. \end{cases} \tag{10.12}$$

For $x \neq 1$ this function is differentiable and

$$\frac{\partial}{\partial x} q(x,r) = \begin{cases} 4x - (r+1) & \text{for } x < 1\,, \\ 4x - 1 & \text{for } x > 1\,. \end{cases}$$

If $r > C = 3$, then $\frac{\partial}{\partial x} q(x,r) < 0$ for $x < 1$ and $\frac{\partial}{\partial x} q(x,r) > 0$ for $x > 1$, i.e. the unique global minimum point of $q(x,r)$ is $x^* = 1$.

If the above penalty algorithm is applied with $r_0 = 1$ and $p(x)$ as in (10.10 b), then the solution is found after two iterations.

In practice, the value of the constant C is unknown, but the theorem ensures that the process terminates after a *finite* number of iterations. The function $q(x,r)$ in (10.10) has the disadvantage, that it is not everywhere differentiable. Therefore most methods to solve unrestricted problems are not directly applicable. In the literature one can find also differentiable exact penalty functions which are very complex and not detailed here.

Exercise 10.13. *Consider the problem of Example 10.4:*

$$\min (x_1+1)^2+(x_2+2)^2$$
$$x_1 \geq 1$$
$$x_2 \geq 2\,.$$

(a) *Determine the Lagrange multipliers* u_1^*, u_2^* *at the minimum point* $x^* = (1,2)^T$.
(b) *Determine the exact penalty function* $q(x,r)$ *of (10.10).*
(c) *Show that* x^* *is a global minimum point* $q(x,r)$ *for* $r > \max\{u_1^*, u_2^*\}$. *Make use of Exercise 3.62.*

10.2 The barrier method

The following method is applicable to problems of type (10.1), where M has the following property:

Definition 10.14. A set $M \subset \mathbb{R}^n$ is called *robust*, if $U_\varepsilon(x) \cap int(M) \neq \emptyset$ for all $x \in M$ and all $\varepsilon > 0$, where $U_\varepsilon(x)$ and $int(M)$ denote the ε-neighborhood and the interior of M, respectively.

Intuitively the robustness of a set M means that all boundary points are "arbitrarily close" to the interior. Thus, for all $x \in M$ exists a sequence of interior points $x^1, x^2, \ldots$ that converges to x. The concept is illustrated in Figure 10.6. While in part (a) each point of M can be approximated by a sequence of interior points, this is not possible, e.g. for the point x in part (b), since the indicated ε-neighborhood does not contain interior points. In applications, the set M usually has the form

$$M = \{x \in \mathbb{R}^n \,|g_i(x) \leq 0 \quad \text{for } i = 1, \ldots, m\}\,.$$

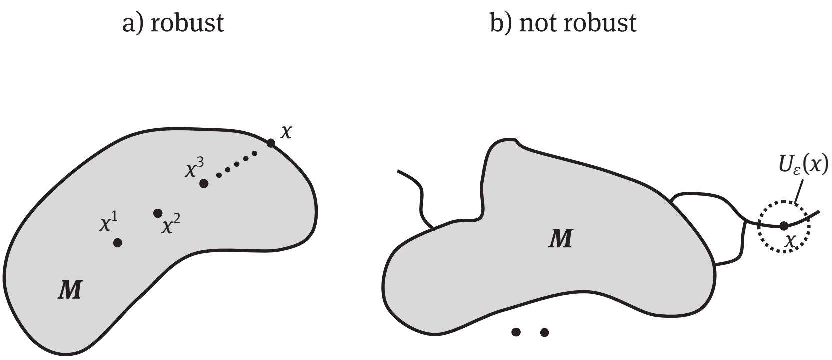


Fig. 10.6. Illustration of robustness.

Since M must have a nonempty interior, the problem may not have constraints in equality form. The basic idea of a barrier method is to add a so-called *barrier function* $b(x)$ to the objective function, preventing the search algorithm from leaving the feasible region M. The function $b(x)$ must be continuous and nonnegative over the interior of M, such that $b(x)$ converges to infinity if x approximates the boundary of M.

A barrier function frequently has the form

$$b(x) = -\sum_{i=1}^{m} \frac{1}{g_i(x)} \quad \text{for } x \in \operatorname{int}(M)\,. \tag{10.13}$$

Example 10.15. For the feasible region $M = [a, b]$ with $a < b$, we get

$$g_1(x) = a - x\,, \quad g_2(x) = x - b$$

(see Example 10.1), thus

$$b(x) = \frac{1}{x-a} + \frac{1}{b-x} \quad \text{for } a < x < b\,.$$

The term $\frac{1}{c}b(x)$ is depicted in Figure 10.7 for $a = 4$, $b = 8$ and $c = 1, 2$. Similar to the reasoning at the start of Section 10.1, we consider the *barrier term* $\frac{1}{c}b(x)$ for increasing values of c, which approximates $\delta_M(x)$ better, the larger c is. Instead of (10.5) we now solve the problem

$$\min_{x \in int(M)} s(x, c)$$

with $s(x, c) := f(x) + \frac{1}{c}b(x)$.

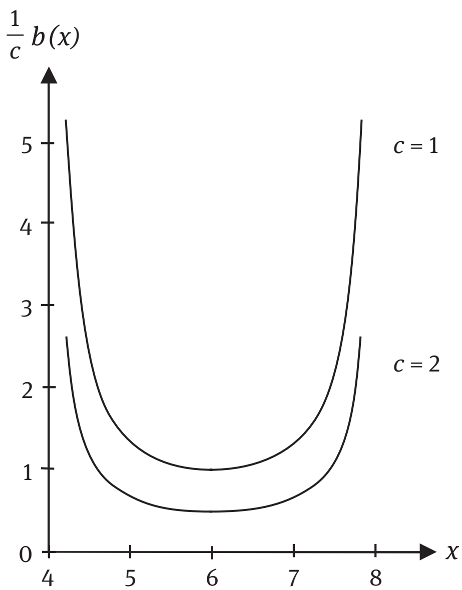


Fig. 10.7. Barrier terms.

Barrier method

(0) Choose a scalar $\varepsilon > 0$ and a number $c_0 > 0$. Set $k = 0$.

(1) Determine a solution x^k of problem

$$\min_{x \in int(M)} s(x, c_k)\,.$$

(2) If $\frac{1}{c_k} b(x^k) \leq \varepsilon$, stop ($x^k$ provides an approximation for a global minimum point of (10.1)). Otherwise go to step (3).

(3) Set $c_{k+1} = 10c_k$, $k = k + 1$ and go to step (1).

Example 10.16. For the problem

$$\min \frac{1}{100} e^x \qquad (10.14)$$
$$x \geq 2$$

we get

$$g(x) = 2 - x, \quad b(x) = \frac{1}{x-2}, \quad s(x, c) = \frac{1}{100} e^x + \frac{1}{c(x-2)}\,.$$

Figure 10.8 illustrates that $s(x, c)$ has a unique minimum point $x(c)$ such that $\lim_{c \to \infty} x(c) = 2$. Clearly, $x^* = 2$ is the minimum point of (10.14).

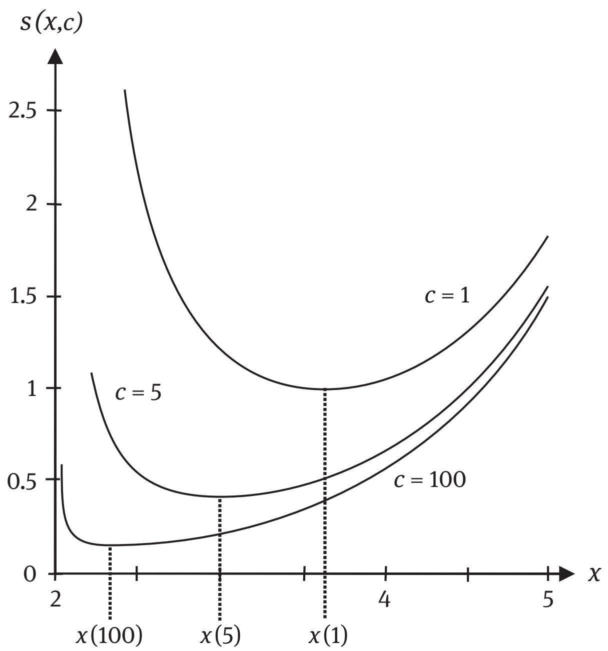


Fig. 10.8. Barrier function $s(x, c)$ of Example 10.16.

The table below shows some values $x(c)$ which can be calculated by the methods of Chapter 6.

Table 10.1. Solution of Example 10.16.

c	$x(c)$
1	3.6291
5	2.9986
100	2.3144
1000	2.1101
10 000	2.0361
100 000	2.0116

Example 10.17. Consider the problem with two variables

$$\min(x_1-3)^4+(2x_1-3x_2)^2$$
$$x_1^2-2x_2 \le 0\,.$$

We get $g(x) = x_1^2 - 2x_2$, $b(x) = \frac{1}{2x_2 - x_1^2}$,

$$s(x,c) = (x_1-3)^4+(2x_1-3x_2)^2+\frac{1}{c(2x_2-x_1^2)}\,.$$

Minimizing the function (x, c) by the methods of Chapter 7, we obtain the following results.

Table 10.2. Solution of Example 10.17.

c	$x_1(c)$	$x_2(c)$
0.1	1.549353	1.801442
1	1.672955	1.631389
10	1.731693	1.580159
100	1.753966	1.564703
1000	1.761498	1.559926
10 000	1.763935	1.558429
100 000	1.764711	1.557955
1 000 000	1.764957	1.557806

The sequence of points generated by the barrier method converges to the minimum point $x^* = (1.765071, 1.557737)^T$ which belongs to the boundary of the feasible region M (see Figure 10.9).

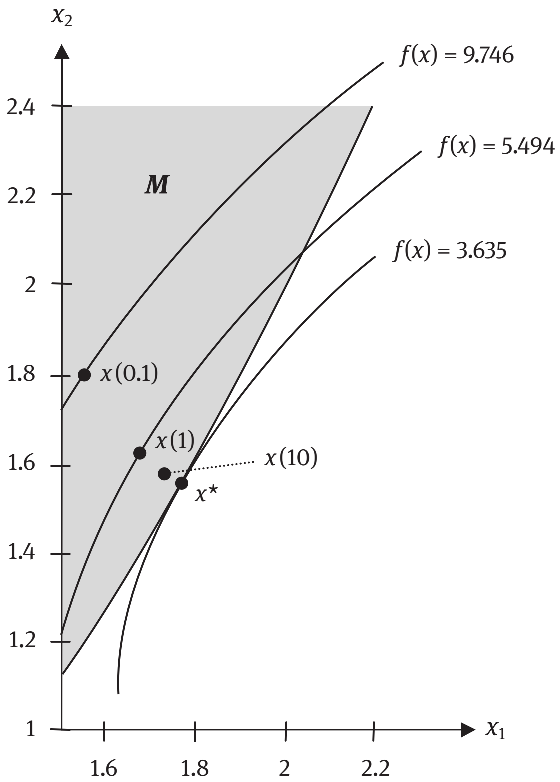


Fig. 10.9. Barrier method.

Exercise 10.18. *Solve the problem of Exercise 10.4 with the barrier method. Illustrate the solution process graphically.*

Exercise 10.19. *Apply the barrier method to the problem of Exercise 10.8 a.*

Since the barrier method constructs a sequence of interior points for (10.1), it is also called the *internal penalty method*. Under fairly weak assumptions (continuity of the functions, robustness of M), the method converges to the optimal solution of (10.1). Despite the fact that the problem $\min_{x \in int(M)} s(x, c)$ is formally restricted, from the computational point of view it can be considered unrestricted. Starting with an interior point of M, this problem can be solved by means of a search algorithm for unrestricted minimization. Because the value of the function $s(x, c) = f(x) + \frac{1}{c} b(x)$ grows indefinitely as x approaches the boundary of M, such a method (if carefully implemented), remains "automatically" inside the set M.

In practice the barrier method encounters diverse computational problems. For some problems it might be difficult to find an interior point, necessary to start a search algorithm. Rounding errors can cause serious problems in minimizing the function

$s(x, c)$, specifically when the search algorithm approaches the boundary of M. In this case the algorithm may easily "disappear" from the feasible region, when the step size is too large.

10.3 Sequential quadratic programming

Consider again the problem

$$\begin{aligned} &\min f(x) \\ g_i(x) &\leq 0 \quad \text{for } i = 1, \dots, m \\ h_j(x) &= 0 \quad \text{for } j = 1, \dots, l \end{aligned} \tag{10.15}$$

with twice continuously differentiable functions f, g_i, h_j. The so-called *sequential quadratic programming (SQP)* generates a sequence of not necessarily feasible points $x^0, x^1, x^2, \dots$ with $x^{k+1} = x^k + \lambda_k d^k$, that usually converges to an optimal point of (10.15). The name of the method due to Wilson, Han and Powell is explained by the fact that each search direction d^k is obtained by solving a quadratic problem.

For a vector $x^k \in \mathbb{R}^n$ and a symmetric positive definite matrix $B_k \in \mathbb{R}^{n \times n}$ we consider the following quadratic problem with vector variable $d \in \mathbb{R}^n$:

$$\left.\begin{aligned} \min F(d) &:= d^T \operatorname{grad} f(x^k) + \tfrac{1}{2} d^T B_k d \\ G_i(d) &:= d^T \operatorname{grad} g_i(x^k) + g_i(x^k) \leq 0 \quad \text{for } i \leq m\,, \\ H_j(d) &:= d^T \operatorname{grad} h_j(x^k) + h_j(x^k) = 0 \quad \text{for } j \leq l\,. \end{aligned}\right\} QP(x^k, B_k)$$

Since the objective function $F(d)$ is strictly convex, the problem has a unique solution. The functions $G_i(d)$ and $H_j(d)$ are the first Taylor approximations of g_i and h_j, respectively.

The following statement provides a stopping criterion for the algorithm below.

Theorem 10.20. *Regardless of the choice of the matrix B_k, it holds: If $d^k = 0$ is a solution of $QP(x^k, B_k)$ with vectors of Lagrange multipliers u^k and v^k, then (x^k, u^k, v^k) satisfies the KKT conditions of problem (10.15).*

Proof. The KKT conditions of $QP(x^k, B_k)$ are (note that d is the vector of variables):

$$\begin{aligned} \operatorname{grad} f(x^k) + B_k d + \sum_{i=1}^{m} u_i \operatorname{grad} g_i(x^k) + \sum_{j=1}^{l} v_j \operatorname{grad} h_j(x^k) &= 0 \\ u_i(d^T \operatorname{grad} g_i(x^k) + g_i(x^k)) &= 0 \quad \text{for } i \leq m \\ d^T \operatorname{grad} g_i(x^k) + g_i(x^k) &\leq 0 \quad \text{for } i \leq m \\ d^T \operatorname{grad} h_j(x^k) + h_j(x^k) &= 0 \quad \text{for } j \leq l \\ u_i &\geq 0 \quad \text{for } i \leq m\,. \end{aligned}$$

For $d = 0$ the system results in the KKT conditions of (10.15). □

There exist several variants of SQP. Below we present a basic version, where

$$\phi(x,u,v) = f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{j=1}^{l} v_j h_j(x)$$

is the Lagrange function of problem (10.15).

SQP algorithm:

(1) Choose $x^0 \in \mathbb{R}^n$ and $B_0 \in \mathbb{R}^{n\times n}$ (B_0 must be symmetric, positive definite and approximate well the matrix $H_x\phi(x^0,u^0,v^0)$, where u^0 and v^0 must be estimated). Set $k=0$.

(2) Determine a solution d^k of $QP(x^k,B_k)$ and the Lagrange multipliers u^k, v^k. If $|d^k| = 0 (\leq \varepsilon)$, stop. (Then x^k is optimal solution of (10.15) (see Theorem 10.20).) Otherwise go to step (3).

(3) Determine a solution λ_k of the problem $\min_{0\leq\lambda\leq 1} q(x^k + \lambda d^k, r)$, where $q(x,r)$ is the exact penalty function (10.10) and set $x^{k+1} = x^k + \lambda_k d^k$.

(4) Set (see (7.29))

$$\begin{aligned} y^k &= \operatorname{grad}_x \Phi(x^{k+1}, u^k, v^k) - \operatorname{grad}_x \Phi(x^k, u^k, v^k)\,, \\ p^k &= \lambda_k d^k\,, \\ B_{k+1} &= B_k + \frac{y^k y^{kT}}{y^{kT}p^k} - \frac{(B_k p^k)(B_k p^k)^T}{p^{kT}B_k p^k}\,. \end{aligned}$$

Set $k = k+1$ and go to step (2).

Before presenting a simple numerical example, we outline some basic ideas to motivate the algorithm: In Section 7.3 we studied Newton's method for unrestricted problems that determines a solution of the system $\operatorname{grad} f(x) = 0$. In the classical form, i.e. for a step size $\lambda_k = 1$, the procedure converges only if the starting point x^0 is sufficiently close to the optimal point. For an optimal step size we have global convergence, i.e. the method converges for each starting point (if the Hessian matrix is positive definite). The need to reduce computational work resulted in the development of quasi-Newton methods.

Similarly one can try to solve restricted problems by applying Newton's method to the system of *equations* that arises in the KKT conditions. Appropriate arrangements must ensure that the points x^k also satisfy the inequalities, at least for large values of k. It can be shown that the previous algorithm corresponds to this type of Newton's method, if $\lambda_k = 1$ and B_k is the Hessian matrix of the Lagrange function. As in the unrestricted optimization, the constant step size 1 can be substituted for the optimal step size. This is done with the aid of an exact penalty function. Instead of the function used in step (3), other exact penalty functions may be used.

To reduce the computational work, matrix B_k is constructed with the aid of an adequate recurrence formula, e.g. formula (7.29) of the BFGS method (see step (4)).

In contrast to unconstraint optimization, the matrix B_{k+1} is not necessarily positive definite, if B_k has this property. In practice, "correction formulas" are incorporated which ensure that each B_k is positive definite. The practical utility of enhanced SQP methods has been demonstrated by several numerical tests. Indeed, modern software for several purposes is based on SQP techniques (see, e.g. the site of the "NEOS Server" (Section 11.3)).

Example 10.21. We apply the SQP algorithm to the problem

$$\begin{aligned} &\min x_1^2 + (x_2+1)^2 \\ &x_1^2 - x_2 \le 0 \end{aligned}$$

with optimal point $x^* = (0,0)^T$. The Lagrange function is

$$\Phi(x_1, x_2, u) = x_1^2 + (x_2+1)^2 + u(x_1^2 - x_2)\,.$$

Thus

$$\begin{aligned} \operatorname{grad}_x \Phi(x_1,x_2,u) &= \begin{pmatrix} 2x_1 + 2x_1 u \\ 2(x_2+1) - u \end{pmatrix}, \\ H_x\Phi(x_1,x_2,u) &= \begin{pmatrix} 2+2u & 0 \\ 0 & 2 \end{pmatrix}. \end{aligned} \tag{10.16}$$

Because of simplicity we always choose $\lambda_k = 1$ in step (3).

Step (1):
We set $x^0 := \binom{-2}{1}$. Since the constraint is not active at x^0, the Lagrange multiplier u in (10.16) is chosen as $u = 0$. We obtain the initial matrix

$$B_0 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

and set $k := 0$.

Step (2):
Since $\operatorname{grad} f(x^0) = \binom{-4}{4}$, $\operatorname{grad} g(x^0) = \binom{-4}{-1}$ and $g(x^0) = 3$, we get the quadratic problem

$$\left.\begin{aligned} \min \quad & -4d_1 + 4d_2 + d_1^2 + d_2^2 \\ & -4d_1 - d_2 + 3 \le 0 \end{aligned}\right\} \quad QP(x^0, B_0)\,,$$

with the KKT conditions

$$\begin{aligned} 2d_1 - 4 - 4u &= 0 \\ 2d_2 + 4 - u &= 0 \\ u(-4d_1 - d_2 + 3) &= 0 \\ -4d_1 - d_2 + 3 &\le 0 \\ u &\ge 0\,. \end{aligned}$$

It is easily verified that there is no solution for $u > 0$. We obtain $u = 0$, $d_1 = 2$, $d_2 = -2$, thus $d^0 = \binom{2}{-2}$, $u^0 = 0$.

Step (3):

$$x^1 = x^0 + d^0 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

Step (4):

$$\begin{aligned}
y^0 &= \operatorname{grad}_x \Phi(x^1, u^0) - \operatorname{grad}_x \Phi(x^0, u^0) \\
&= \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} -4 \\ 4 \end{pmatrix} = \begin{pmatrix} 4 \\ -4 \end{pmatrix}, \\
p^0 &= \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \\
B_1 &= \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} + \frac{\binom{4}{-4}(4, -4)}{(4, -4)\binom{2}{-2}} - \frac{\binom{4}{-4}(4, -4)}{(2, -2)\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}\binom{2}{-2}} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \\
k &:= 1\,.
\end{aligned}$$

Step (2):
Since $\operatorname{grad} f(x^1) = \binom{0}{0}$, $\operatorname{grad} g(x^1) = \binom{0}{-1}$ and $g(x^1) = 1$, we obtain the quadratic problem

$$\left.\begin{aligned} \min d_1^2 \quad &+ d_2^2 \\ -d_2 + \quad &1 \le 0 \end{aligned}\right\} \quad QP(x^1, B_1)\,,$$

with the KKT conditions

$$\begin{aligned}
2d_1 &= 0 \\
2d_2 - u &= 0 \\
u(1 - d_2) &= 0 \\
1 - d_2 &\le 0 \\
u &\ge 0\,.
\end{aligned}$$

The unique solution of this system is $d_1 = 0$, $d_2 = 1$, $u = 2$, therefore $d^1 = \binom{0}{1}$, $u^1 = 2$.

Step (3):

$$x^2 = x^1 + d^1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Step (4):

$$
\begin{aligned}
y^1 &= \operatorname{grad}_x \Phi(x^2, u^1) - \operatorname{grad}_x(x^1, u^1)\\
&= \begin{pmatrix}0\\0\end{pmatrix} - \begin{pmatrix}0\\-2\end{pmatrix} = \begin{pmatrix}0\\2\end{pmatrix},
\end{aligned}
$$

$$
\begin{aligned}
p^1 &= \begin{pmatrix}0\\1\end{pmatrix},\\
B_2 &= \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix} + \frac{\binom{0}{2}(0,2)}{(0,2)\binom{0}{1}} - \frac{\binom{0}{2}(0,2)}{(0,1)\begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}\binom{0}{1}} = \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix},\\
k &:= 2\,.
\end{aligned}
$$

Step (2):
We obtain the quadratic problem

$$
\left.\begin{array}{l}\min 2d_2 + d_1^2 + d_2^2\\ -d_2 \le 0\end{array}\right\} \quad QP(x^2, B_2)
$$

with the solution $d^2 = \binom{0}{0}$. The algorithm ends, providing the optimal solution $x^2 = \binom{0}{0}$.

Exercise 10.22. *Apply the above algorithm to the problem*

$$
\begin{gathered}
\min -x_1 + x_2\\
x_1^2 - x_2 + 1 \le 0\,.
\end{gathered}
$$

Choose

$$
x^0 = \begin{pmatrix}0\\2\end{pmatrix}, \quad B_0 = \begin{pmatrix}2 & 0\\0 & 0\end{pmatrix}
$$

and set $\lambda_k = 1$ *in step (3). Show that the optimal point* $x^* = \binom{1/2}{5/4}$ *is determined after two iterations.*

11 Nondifferentiable and global optimization

Nowadays an introduction to the area of nonlinear optimization cannot be considered complete, if it does not even mention the two large areas of nondifferentiable and global optimization, studied in Sections 11.1 and 11.2, respectively. In recent decades, these branches have developed with immense rapidity (see the references in Section 11.3). We restrict ourselves in this chapter to introduce some specific cases and basic ideas to solve problems in these areas.

11.1 Nondifferentiable optimization

Except for some procedures in Sections 6.2 and 7.5, the solution techniques discussed earlier require that all functions of the optimization problem are at least once continuously differentiable. However, in many applications in economics, mechanics, engineering, etc. functions arise which do not have this characteristic. We present some examples in the following. Solution procedures for nondifferential problems are studied in Sections 11.1.2 and 11.1.3.

11.1.1 Examples for nondifferentiable problems

Piecewise linear objective function:

Example 11.1 (Composition of offers). A company that wants to acquire the amount a of a certain product P, requests the offers of n producers $P_1, \dots, P_n$. The producer P_j can provide the maximum amount m_j and his price for the quantity x_j of P is denoted by $f_j(x_j)\,(j = 1, \dots, n)$. If the company wants to minimize the total cost, the following separable problem (see Section 1.2.1) arises:

$$\begin{aligned} \min \sum_{j=1}^{n} f_j(x_j) \\ \sum_{j=1}^{n} x_j \le a \\ 0 \le x_j \le m_j \quad (j = 1, \dots, n)\,. \end{aligned}$$

The partial cost $f_j(x_j)$ is typically an increasing, concave and piecewise linear function. Suppose, e.g. that there is a cost of \$ 300 for the preparation of the production machines, the first thousand items of P are sold by the unit price of \$ 1 and for higher quantities, the producer P_j gives a discount of \$ 0.4 per piece. So we get the function $f_j(x_j)$ in Figure 11.1 which is not differentiable at the points $x_j = 0$ and $x_j = 1000$ (note that $f_j(0) = 0$).

As we have seen in Example 1.6, piecewise linear objective functions can also arise in transportation problems.

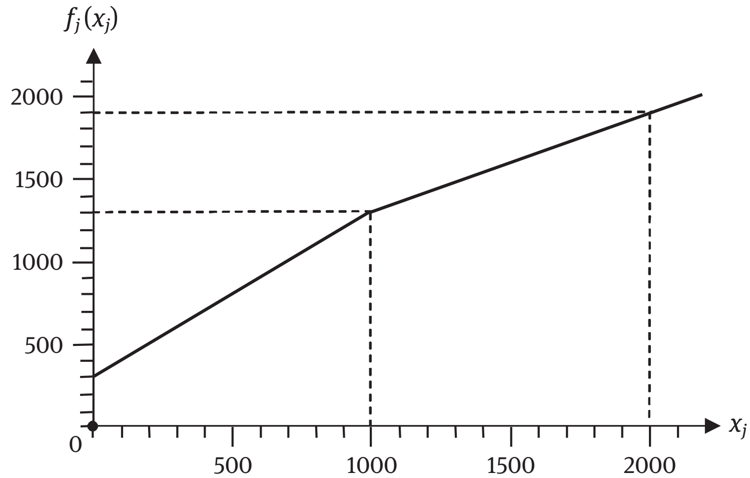


Fig. 11.1. Partial cost function.

Example 11.2. *Minimax problem:*
In many applications, the function to be minimized can be considered as the maximum of functions $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$, i.e. the problem has the form

$$\min_{x \in \mathbb{R}^n} \quad \max_{i=1,\ldots,m} \quad f_i(x) . \tag{11.1}$$

Even if all functions f_i are differentiable, $f(x) := \max_{i=1,\ldots,m} f_i(x)$ is usually not differentiable at "intersection points" of two functions f_i and f_j (see Figure 11.2).

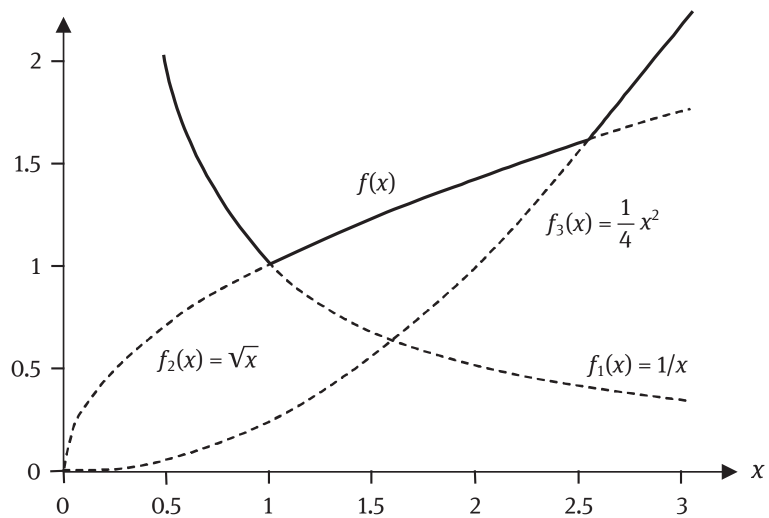


Fig. 11.2. Nondifferentiable function of a minimax problem.

A specific case of (11.1) is the *discrete linear Chebychev approximation*. Given a "complicated" function $h : \mathbb{R} \to \mathbb{R}$, n "simple" functions $g_1, \ldots, g_n$, and m points $t_1, \ldots, t_m \in \mathbb{R} (n < m)$. The task is to approximate the function h by a linear combination of the functions g_j such that the maximum distance between h and its approximation at the points t_j is minimized. We get the problem

$$\min_{x\in\mathbb{R}^n} \max_{i=1,\ldots,m} \left| h(t_i) - \sum_{j=1}^{n} x_j g_j(t_i) \right| \quad \text{for } i = 1, \ldots, m\,. \tag{11.2}$$

With the notation $f_i(x) := h(t_i) - \sum_{j=1}^{n} x_j g_j(t_i)$ for $i = 1, \ldots, m$, we can rewrite (11.2) as

$$\min_{x\in\mathbb{R}^n} \max \{f_1(x), -f_1(x), \ldots, f_m(x), -f_m(x)\}\ . \tag{11.3}$$

Example 11.3. Most methods to solve a restricted problem

$$\begin{aligned} &\min f(x) \\ &g_i(x) \leq 0 \quad \text{for } i = 1 \ldots m \end{aligned} \tag{11.4}$$

need a feasible point to start an iterative procedure (see, e.g. the introductions of Chapters 8 and 9). A feasible point x^* of (11.4) can be obtained by solving the problem

$$\min_{x\in\mathbb{R}^n} \max \{g_1(x), \ldots, g_m(x), 0\} \tag{11.5}$$

which is another specific case of (11.1).

Obviously the objective function $h(x) := \max\{g_1(x), \ldots, g_m(x), 0\}$ is zero, if and only if x is feasible for (11.4); $h(x)$ is illustrated in Figure 11.3 for $n = 2$, $m = 3$,

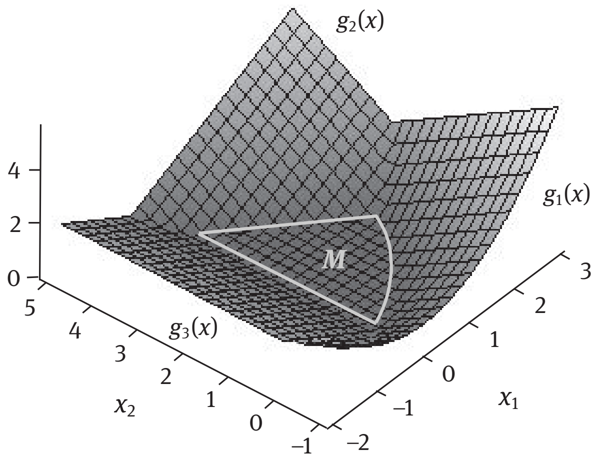


Fig. 11.3. Nondifferentiable bidimensional objective function.

$g_1(x) = (1/2)x_1^2 - x_2$, $g_2(x) = x_1 + x_2 - 4$, $g_3(x) = -x_1$. The function $h(x)$ is not differentiable at various "intersection points", i.e. at points where two of the four functions $g_1(x), g_2(x), g_3(x)$ and $g_4(x) := 0$ coincide. In particular, $h(x)$ is not differentiable at boundary points of the feasible region M of (11.4).

Decomposition:
In principle, a linear problem can be solved in two steps by writing it in the form

$$\begin{aligned} &\min c^T x + d^T y \\ &Ax + By \le b\,, \end{aligned} \tag{11.6}$$

where A, B, c, d, b are matrices and vectors of appropriate dimensions and x and y denote the respective vector variables. In the first step, a solution of the subproblem

$$f(x) := \min_y \{d^T y \,|By \le b - Ax\} \tag{11.7}$$

is determined for an arbitrary vector x, which is feasible for (11.6). In the second step, the "remaining problem"

$$\min_x \{c^T x + f(x)\}\,. \tag{11.8}$$

is solved. A solution of (11.6) is now given by (x^*, y^*), where x^* is a solution of (11.8) and y^* is a solution of (11.7) for $x = x^*$. The optimal value of (11.6) is $c^T x^* + f(x^*)$. The objective function $c^T x + f(x)$ of (11.8) is usually not everywhere differentiable.

Example 11.4. Consider the problem

$$\min \frac{1}{4}x + y \tag{11.9}$$

$$-2x - y \le -2 \tag{1}$$

$$-\frac{1}{2}x - y \le \quad 1 \tag{2}$$

$$\frac{1}{2}x - y \le \quad 5 \tag{3}$$

$$\frac{1}{2}x + y \le \quad 5 \tag{4}$$

$$-x \qquad \le \quad 0 \tag{5}$$

with optimal solution $(4, -3)^T$ (see Figure 11.4). It holds $c = \frac{1}{4}$, $d = 1$,

$$A = \begin{pmatrix} -2 \\ -1/2 \\ 1/2 \\ 1/2 \\ -1 \end{pmatrix}, \quad B = \begin{pmatrix} -1 \\ -1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \quad b = \begin{pmatrix} -2 \\ 1 \\ 5 \\ 5 \\ 0 \end{pmatrix},$$

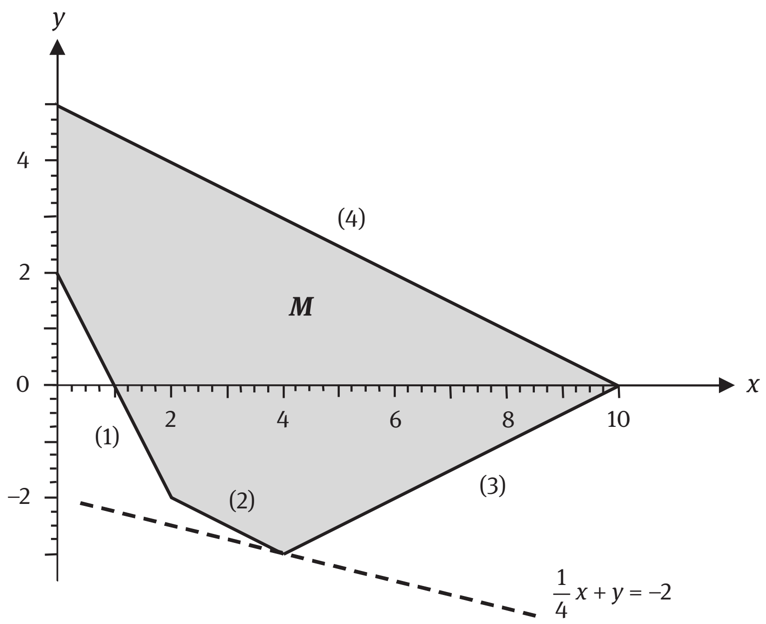


Fig. 11.4. Graphical solution of problem (11.9).

$$
\begin{aligned}
f(x) &:= \min_y \{d^T y \,|\, By \le b - Ax\} \\
&= \min_y \{y \,|\, y \ge 2 - 2x, y \ge -1 - (1/2)x, y \ge -5 + (1/2)x, y \le 5 - (1/2)x\} \\
&= \begin{cases} 2 - 2x & \text{for } 0 \le x \le 2 \\ -1 - (1/2)x & \text{for } 2 < x \le 4 \\ -5 + (1/2)x & \text{for } 4 < x \le 10\,. \end{cases}
\end{aligned}
$$

Note that the graph of $f(x)$ corresponds to the lower boundary of the set M in Figure 11.4. The objective function of (11.8) is therefore

$$
c^T x + f(x) = \frac{1}{4}x + f(x) = \begin{cases} 2 - (7/4)x & \text{for } 0 \le x \le 2 \\ -1 - (1/4)x & \text{for } 2 < x \le 4 \\ -5 + (3/4)x & \text{for } 4 < x \le 10 \end{cases}
$$

(see Figure 11.5). The function is not differentiable at the point $x = 2$ and at the minimum point $x^* = 4$. Thus the optimal solution of (11.9) is $(x^*, y^*)^T = (4, -3)^T$, which corresponds to the graphical solution in Figure 11.4.

In practice, the above decomposition is only useful for large scale linear problems, where the matrices A and B in (11.6) have a specific structure that facilitates the solution of the subproblem (11.7).

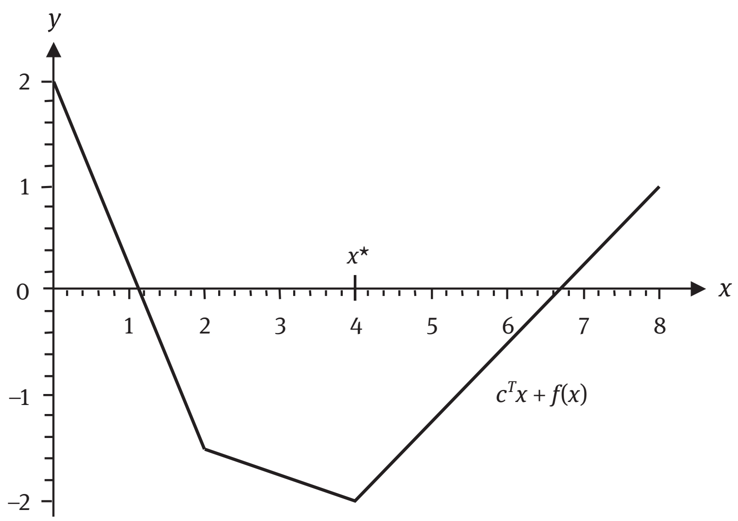


Fig. 11.5. Objective function of problem (11.8).

Besides the above examples, a nondifferential objective function may arise also in other applications: The exact penalty function (10.10) is generally not differentiable at all points. For example, for $r \neq 0$ the function (10.12) is not differentiable at $x = 1$. The objective function of the dual problem is often not differentiable (see Definition 4.25). The characteristics of this function are discussed, e.g. in Luenberger (2003, Section 6.3). Finally, the total transportation cost

$$f(x_1, x_2) = s \sum_{i=1}^{n} q_i \sqrt{(x_1 - a_i)^2 + (x_2 - b_i)^2}$$

to be minimized in the location problem of Example 1.18, is not differentiable at the points (a_i, b_i).

11.1.2 Basic ideas of resolution

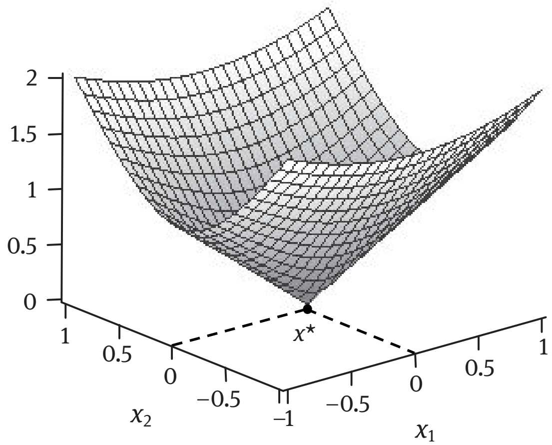
In the above examples and in most applications the objective function is differentiable at "almost all" points. This fact suggests the idea that, e.g. the gradient method or a quasi-Newton method could be applied to minimize such functions, expecting that the algorithm "usually" generates differentiable points. Furthermore, if an iteration point x^k is not differentiable, the iterative method could be restarted with a "differentiable point" close to x^k. However, the following examples show that such a procedure does not always work.

Example 11.5. The function

$$f(x_1, x_2) = \sqrt{x_1^2 + 3x_2^2} \tag{11.10}$$

is convex and everywhere differentiable except at the minimum point $x^* = (0, 0)^T$. For an arbitrary starting point the gradient method converges to x^* (see Figure 11.6).

(a) The function (11.10)



(b) Iteration points for $x^0 = (1, 0.5)^T$

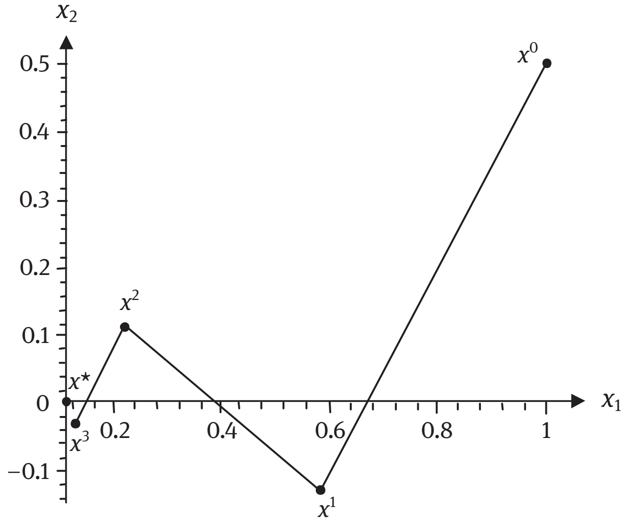


Fig. 11.6. Minimization of function (11.10) with the gradient method.

However, the problem is to establish an adequate stopping criterion. Since,

$$\operatorname{grad} f(x_1, x_2) = \frac{1}{\sqrt{x_1^2 + 3x_2^2}} \begin{pmatrix} x_1 \\ 3x_2 \end{pmatrix},$$

we get $|\operatorname{grad} f(x_1, x_2)| = \frac{1}{\sqrt{x_1^2+3x_2^2}} \sqrt{x_1^2 + 9x_2^2} \geq 1$ for any point $(x_1, x_2)^T \neq (0, 0)^T$, i.e. none of the points constructed by the algorithm satisfies the stopping criterion $|\operatorname{grad} f(x_1, x_2)| \leq \varepsilon$ (see the algorithm of Section 7.2).

Example 11.6. The function

$$f(x_1, x_2) = \begin{cases} \sqrt{52x_1^2 + 117x_2^2} & \text{for } x_1 \geq |x_2| \\ 4x_1 + 9|x_2| & \text{for } 0 < x_1 < |x_2| \\ 4x_1 + 9|x_2| - x_1^5 & \text{for } x_1 \leq 0 \end{cases} \tag{11.11}$$

is convex and differentiable at all points except the ray defined by $x_1 \leq 0$, $x_2 = 0$ (see Figure 11.7). The global minimum point is $x^* = (-0.8^{1/4}, 0)^T \approx (-0.946, 0)^T$. Starting the gradient method with a point of the region

$$R := \{(x_1, x_2)^T | x_1 > |x_2| > (4/9)^2 x_1\},$$

the sequence of iteration points stays in R and converges to $\overline{x} := (0, 0)^T$. This point is not even a local minimum point of f.

The last two examples show that the gradient method may fail to minimize a nondifferentiable function, even if the function is differentiable at all iteration points.

A simple approach to minimize a nondifferentiable function f is based on the approximation of f by a differentiable function g. For example,

$$f(x) = |x|$$

can be approximated by

$$g(x) = \sqrt{x^2 + \varepsilon}$$

with a "small" value of $\varepsilon > 0$. Consequently, the sum

$$f(x) = f_0(x) + \sum_{i=1}^{m} |f_i(x)| \tag{11.12}$$

with functions $f_i : \mathbb{R}^n \to \mathbb{R} (i = 0, 1, \ldots, m)$ can be approximated by the function

$$g(x) = f_0(x) + \sum_{i=1}^{m} \sqrt{[f_i(x)]^2 + \varepsilon}, \tag{11.13}$$

which is differentiable if all functions f_i are differentiable.

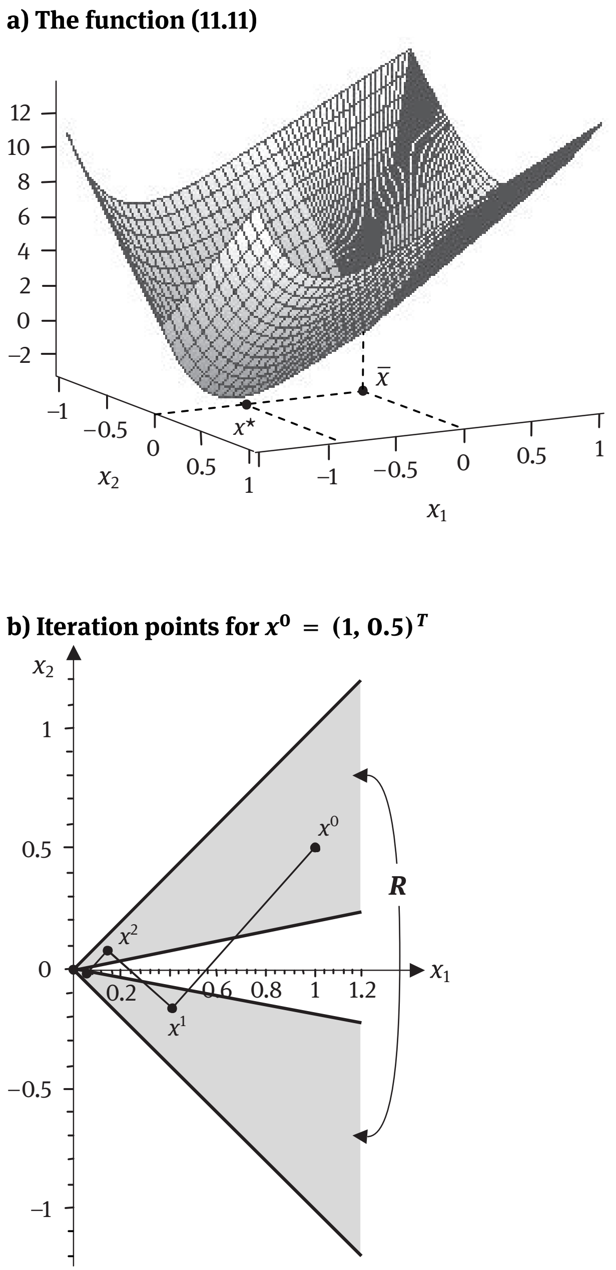


Fig. 11.7. Minimization of function (11.11) with the gradient method.

Many nondifferentiable functions of practical interest can be written as (11.12). For example, any continuous and piecewise linear function $f : \mathbb{R}^n \to \mathbb{R}$ can be expressed as

$$f(x) = a + bx + \sum_{i=1}^{m} \alpha_i |x - x_1| \\ (x_1 < x_2 < \cdots < x_m; a, b, \alpha_i \in \mathbb{R} \quad \text{for } i = 1, \ldots, m)\,, \tag{11.14}$$

which is a special case of (11.12) with $f_0(x) = a + bx$ and $f_i(x) = \alpha_i(x - x_i)$ for $i = 1, \ldots, m$. The differentiable approximation of (11.14) is then

$$g(x) = a + bx + \sum_{i=1}^{m} \sqrt{[\alpha_i(x - x_i)]^2 + \varepsilon}\,. \tag{11.15}$$

Example 11.7. The function

$$f(x) = \begin{cases} 7 - 3x & \text{for } x \le 2 \\ 5/3 - x/3 & \text{for } 2 < x \le 5 \\ -5 + x & \text{for } 5 < x \le 7 \\ -26 + 4x & \text{for } x > 7 \end{cases} \tag{11.16}$$

is continuous, convex and piecewise linear (see Figure 11.8). At first we write this function in the form (11.14): Since the nondifferentiable points are 2, 5 and 7, we obtain

$$f(x) = a + bx + \alpha_1|x - 2| + \alpha_2|x - 5| + \alpha_3|x - 7|\,. \tag{11.17}$$

The parameters $a, b, \alpha_1, \alpha_2, \alpha_3$ can be determined, comparing (11.16) and (11.17). For $x < 2$ we get

$$a + bx + \alpha_1(2 - x) + \alpha_2(5 - x) + \alpha_3(7 - x) = 7 - 3x \quad \Rightarrow$$

$$\begin{aligned} a + 2\alpha_1 \quad &+ 5\alpha_2 \quad + 7\alpha_3 \quad = 7 && \text{(i)} \\ b - \alpha_1 \quad &- \alpha_2 \quad - \alpha_3 \quad = -3\,. && \text{(ii)} \end{aligned}$$

Analyzing the regions $2 < x \le 5, 5 < x \le 7, x > 7$, we get similarly

$$\begin{aligned} a - 2\alpha_1 \quad &+ 5\alpha_2 \quad + 7\alpha_3 \quad = 5/3 && \text{(iii)} \\ b + \alpha_1 \quad &- \alpha_2 \quad - \alpha_3 \quad = -1/3 && \text{(iv)} \\ a - 2\alpha_1 \quad &- 5\alpha_2 \quad + 7\alpha_3 \quad = -5 && \text{(v)} \\ b + \alpha_1 \quad &+ \alpha_2 \quad - \alpha_3 \quad = 1 && \text{(vi)} \\ a - 2\alpha_1 \quad &- 5\alpha_2 \quad - 7\alpha_3 \quad = -26 && \text{(vii)} \\ b + \alpha_1 \quad &+ \alpha_2 \quad + \alpha_3 \quad = 4\,. && \text{(viii)} \end{aligned}$$

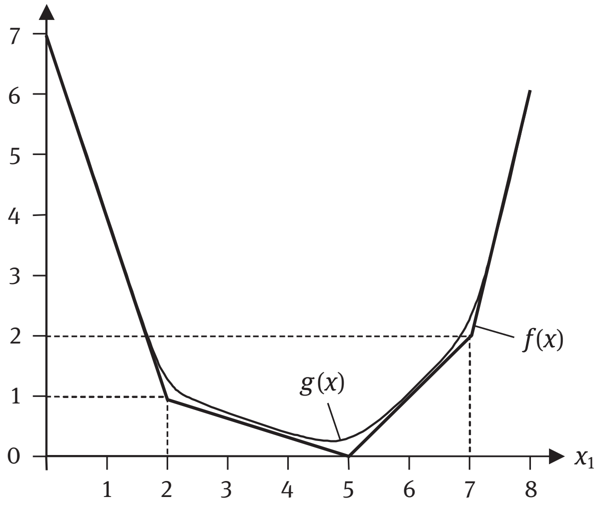


Fig. 11.8. Piecewise linear function f and differential approximation g for $\varepsilon = 0.1$.

The system (i)–(viii) can be easily solved. Comparing (ii) and (iv), (iv) and (vi), (vi) and (viii), we obtain $\alpha_1 = 4/3$, $\alpha_2 = 2/3$, $\alpha_3 = 3/2$, respectively. After this, the equations (i) and (ii) yield $a = -19/2$ and $b = 1/2$. Thus the function (11.16) can be written as

$$f(x) = \frac{-19}{2} + \frac{x}{2} + \frac{4}{3}|x-2| + \frac{2}{3}|x-5| + \frac{3}{2}|x-7|\,, \tag{11.18}$$

and according to (11.15) the differentiable approximation is

$$\begin{aligned} g(x) = {} & \frac{-19}{2} + \frac{x}{2} + \sqrt{[(4/3)(x-2)]^2 + \varepsilon} \\ & + \sqrt{[(2/3)(x-5)]^2 + \varepsilon} + \sqrt{[(3/2)(x-7)]^2 + \varepsilon} \end{aligned} \tag{11.19}$$

(see Figure 11.8).

Table 11.1 shows the minimum point x_ε of (11.19) for some values of ε:

Table 11.1. Minimization by differentiable approximation.

ε	x_ε
0.1	4.72
0.01	4.91
0.001	4.97

For $\varepsilon \to 0$ the points x_ε converge to the minimum point $x^* = 5$ of function (11.16).

In practice, the approach suggested by the above example causes serious problems regarding the stability of the procedure (see Chapter 5): A minimum point x^* of (11.12) usually satisfies $f_k(x^*) = 0$ for a $k \in \{1, \ldots, m\}$ (see Figure 11.8). For a small ε and a point x close to x^*, the kth term in the sum of (11.13) is then the square root of the small number $z = [f_k(x)]^2 + \varepsilon$. Since for $z \to 0$ the slope of the function $\sqrt{z}$ converges to ∞, tiny errors in z may cause large errors in $\sqrt{z}$.

We finally note that the minimax problem (11.1) can be transformed in the equivalent restricted form

$$\min_{f_i(x) \leq x_{n+1}} x_{n+1} \quad \text{for } i = 1, \ldots, m\,, \tag{11.20}$$

introducing another variable x_{n+1}. So, the nondifferentiability is eliminated, if all functions f_i are differentiable. However, this transformation does not “simplify” the problem. The feasible set of (11.20), which corresponds to the epigraph (see Definition 3.30) of the objective function $\max_{i=1\ldots m} f_i(x)$ of (11.1), can be quite complex. For example, the epigraph of the simple function f in Figure 11.2 is not convex.

Various algorithms for nondifferentiable problems try to adapt the basic ideas of the differential case (see, e.g. Wolfe (1975, p. 146)). Some notable methods are the bundle method, cutting plane and ellipsoidal methods. In the following we only illustrate some basic ideas of the former procedure.

11.1.3 The concept of bundle methods

Below we study the minimization of a function f, which is convex over the space $\mathbb{R}^n$, though a great part of the following theory remains valid under weaker conditions. The following considerations are based on Zowe (1985, Section 6). The convexity ensures that the unilateral derivative $D^+f(x, d)$ (see Definition 3.40) is defined at each point x and for each direction d. The first idea is to minimize f iteratively, by modifying the gradient method for a nondifferentiable function f. In this case, the maximum descent direction d^k is obtained in a natural manner by solving the problem

$$\min_{|d|=1} D^+f(x^k, d)\,, \tag{11.21}$$

since for $|d| = 1$ the derivative $D^+f(x^k, d)$ is a measure for the growth/decrease of f in the direction of d. Problem (11.21) is equivalent to

$$\min_{|d|\leq 1} D^+f(x^k, d)\,, \tag{11.22}$$

insofar as the optimal solutions are identical. But the latter is computationally easier, because the feasible set is convex. We now transform (11.22) by using (see Theorem 3.43 (iii))

$$D^+f(x, d) = \max_{a \in \partial f(x)} a^T d\,, \tag{11.23}$$

where $\partial f(x)$ denotes the subdifferential of f at x (see Definition 3.34). Combining (11.22) and (11.23), we obtain the problem

$$\min_{|d|\leq 1} \max_{a \in \partial f(x)} a^T d\,, \tag{11.24}$$

which can be rewritten with the aid of von Neumann's minimax theorem as

$$\max_{a \in \partial f(x)} \min_{|d|\leq 1} a^T d\,. \tag{11.25}$$

Since $-a/|a|$ is the solution of the "interior" problem, we obtain

$$\max_{a \in \partial f(x)} a^T \frac{-a}{|a|} = \max_{a \in \partial f(x)} -|a|\,, \tag{11.26}$$

which finally is equivalent to the minimum-norm problem

$$\min_{a \in \partial f(x)} |a|\,. \tag{11.27}$$

Using (11.27) to determine the search direction, we obtain the following basic procedure:
(1) Choose $x^0 \in \mathbb{R}^n$ and set $k = 0$.
(2) Calculate $d^k = -a^k/|a^k|$, where a^k is the solution of the problem $\min_{a\in\partial f(x^k)} |a|$.
(3) If $0 \in \partial f(x^k)$, stop. Otherwise go to step (4).
(4) Determine a solution λ_k of $\min_{\lambda\in\mathbb{R}_+} f(x^k + \lambda d^k)$.
(5) Set $x^{k+1} = x^k + \lambda_k d^k$, $k = k + 1$ and go to step (2).

The stopping criterion of step (3) is based on the fact that x^k is a global minimum point of f, if and only if $0 \in \partial f(x^k)$ (see Exercise 3.62). At differentiable points the iteration is identical to a step of the gradient method.

Example 11.8. We minimize the function

$$f(x_1, x_2) = |x_1| + |x_2|\,,$$

starting with $x^0 = (1, 0)^T$. A subgradient $a = (a_1, a_2)^T$ of f at the nondifferentiable point x^0 satisfies (Definition 3.34):

$$|x_1| + |x_2| \geq 1 + (a_1, a_2)\begin{pmatrix} x_1 - 1 \\ x_2 \end{pmatrix}.$$

With the methods of Section 3.4 it can be shown that the vector a is a subgradient, if and only if $a_1 = 1$ and $|a_2| \leq 1$, i.e. the subdifferential of f at x^0 is

$$\partial f(x^0) = \left\{ \begin{pmatrix} 1 \\ \alpha \end{pmatrix} \middle|\ |\alpha| \leq 1 \right\}.$$

The step (2) of the algorithm yields $a^0 = (1, 0)^T$ and $d^0 = -(1, 0)^T$. Searching in the direction d^0 gives the optimal solution $x^* = (0, 0)^T$ in the first step.

Example 11.9. We apply the algorithm to the function (11.11), starting with the point $x^0 = (0,0)^T$. Because of simplicity we determine the maximum descent direction directly by means of (11.21). Since

$$D^+f(x^0,d) = \lim_{t\to 0+} \frac{f(td)}{t}\,,$$

$$f(td) = \begin{cases} t\sqrt{52d_1^2 + 117d_2^2} & \text{for } d_1 \geq |d_2|\,,\\ t(4d_1 + 9|d_2|) & \text{for } 0 < d_1 < |d_2|\,,\\ 4td_1 + 9t|d_2| - (td_1)^5 & \text{for } d_1 \leq 0\,, \end{cases}$$

we get

$$D^+f(x^0,d) = \begin{cases} \sqrt{52d_1^2 + 117d_2^2} & \text{for } d_1 \geq |d_2|\,,\\ 4d_1 + 9|d_2| & \text{for } d_1 < |d_2|\,, \end{cases}$$

and the solution of (11.21) is $d^0 = (-1,0)^T$ with optimal value -4. Figure 11.7a illustrates that d^0 is in fact the maximum descent direction. The optimal solution is encountered in the first iteration.

The above algorithm does not always converge to the optimal solution. For example, if it is applied to the function (11.11) with a starting point in the region R of Example 11.6, the sequence x^k converges to the nonoptimal solution $\overline{x} = (0,0)^T$ (see Figure 11.7). Because (11.11) is differentiable for $x_1 > 0$, the above algorithm generates the same iteration points as the gradient method in Example 11.6. However, Example 11.9 suggests the following modification of the basic algorithm.

For an adequate neighborhood U_k of the point x^k we substitute the subdifferential $\partial f(x^k)$ in step (2) by the union of subdifferentials $D(U_k) := \bigcup_{y\in U_k} \partial f(y)$. If a point x^k is sufficiently close to $\bar{x}$, then $\bar{x} \in U_k$ and $\partial f(\bar{x}) \subset D(U_k)$, i.e. the ideal direction $(-1,0)^T$, corresponding to the point $\bar{x}$ (see Example 11.9) becomes a candidate for the search direction. By choosing this search direction, the modified algorithm can "escape from the attraction point" $\bar{x}$.

In enhancements of the bundle method used in commercial software, the search direction is determined in a sophisticated manner. Here we will confine ourselves to some basic ideas. Consider the collection of subgradients $g_1, g_2, g_3, \ldots$, determined in previous iterations such that the convex polytope $P :=\text{conv}(\{g_1, g_2, g_3, \ldots\})$ is a good approximation of $D(U_k)$. The subgradients of P represent the so-called *bundle*. The determination of the search direction by means of P is a quadratic minimization problem which can be easily solved. If the obtained direction is "good" (the line search will reveal this), the algorithm progresses in this direction. Otherwise, a better approximation of $D(U_k)$ is constructed by including another subgradient in the construction of the polytope P. For more details see Zowe (1985) and the literature cited there.

11.2 Global optimization

Next we study some nonlinear problems that have local minima beyond the global. Such problems occur frequently in applications.

Example 11.10. Consider the problem

$$\begin{aligned} \min -(x_1-1)^2 - (x_2-2)^2 & \\ 2x_1 + x_2 &\le 15 \\ (x_1-5)^2 + (x_2+5)^2 &\le 100 \\ x_1, x_2 &\ge 0 \end{aligned} \tag{11.28}$$

illustrated in Figure 11.9. The objective value is the negative squared distance to the point $(1,2)^T$. Thus the extreme point $(7.5,0)^T$ of M is the unique global minimum point of (11.28). The two extreme points $(0,0)^T$ and $(5,5)^T$ represent local minima. (Note that the extreme point $(0,\sqrt{75}-5)^T$ of M is *not* a local minimum point.)

A method that determines the search direction based on "local information" (derivatives at the iteration points) can only discover a local minimum, if it is applied to a problem with various minima, see the final remarks of Chapter 2.

The branch of mathematical optimization that develops algorithms to determine a global optimum in presence of other local optima, is called *global optimization*. We now introduce some special problems of this type which often arise in practice. In the Sections 11.2.2 and 11.2.3 some solution methods will be discussed.

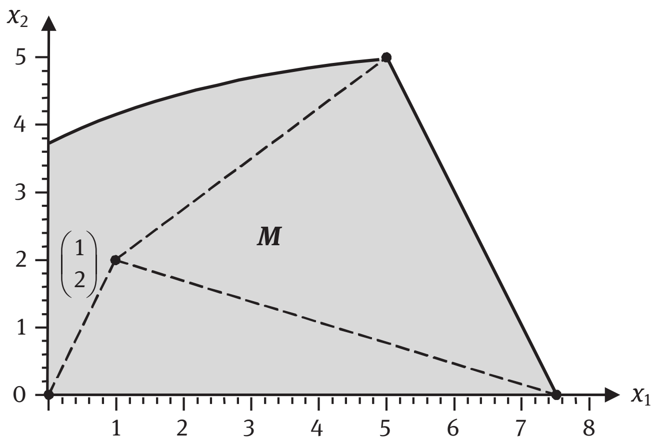


Fig. 11.9. Local and global minima of problem (11.28).

11.2.1 Specific cases of global optimization

Concave minimization

Given a problem of the form

$$\min_{x \in M} f(x)\,, \tag{11.29}$$

where f is a concave function and $M \subset \mathbb{R}^n$ a convex closed set.

The problem of the above example is of this form. Model (11.29) is used in several applications, e.g. in economics, when a concave cost function is to be minimized (Example 11.1), in diamond polishing, where the loss of material must be minimized, or in the production of computer chips. Moreover, some models of operations research, e.g. the binary linear programming problem can be transformed in a concave minimization problem (11.29) (see Horst and Tuy (1996, Section I. 2.3)).

Minimization of d.c. functions

A function of the form $f(x) = f_1(x) - f_2(x)$ with convex functions $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ is called a *d.c. function*, where the initials stand for "difference of convex functions". The minimization of d.c. functions is a very general problem, containing convex and concave minimization as special cases. It is interesting to observe, that any continuous function over a convex, closed bounded set, can be approximated arbitrarily well by a d.c. function (see Horst and Tuy (1996, Section I. 3)). To illustrate this, consider the function $f : [0, 2\pi] \to \mathbb{R}$ with $f(x) = \sin(x)$, which can be written as the difference of the convex functions (Figure 11.10)

$$f_1(x) = \begin{cases} \pi & \text{for } 0 \le x \le \pi \\ x + \sin(x) & \text{for } \pi < x \le 2\pi\,, \end{cases}$$

$$f_2(x) = \begin{cases} \pi - \sin(x) & \text{for } 0 \le x \le \pi \\ x & \text{for } \pi < x \le 2\pi\,. \end{cases}$$

Inverted convex constraints

A constraint $g_i(x) \le 0$ with a concave function $g_i : \mathbb{R}^n \to \mathbb{R}$ is called *inverted convex constraint*. The corresponding set $\{x \in \mathbb{R}^n | g_i(x) \le 0\}$ is the complement of a convex set. In many applications, convex and inverted convex constraints occur together, while the objective function is usually convex or concave.

Lipschitz optimization problems

Many algorithms have been developed for a problem of the type

$$\min_{x \in I} f(x)\,, \tag{11.30}$$

where f is Lipschitz continuous (see Section 7.6) and $I \subset \mathbb{R}^n$ is a so-called n-dimensional interval: For $a := (a_1, \dots, a_n)^T, b := (b_1, \dots, b_n)^T \in \mathbb{R}^n$ with $a_i \le b_i$ for

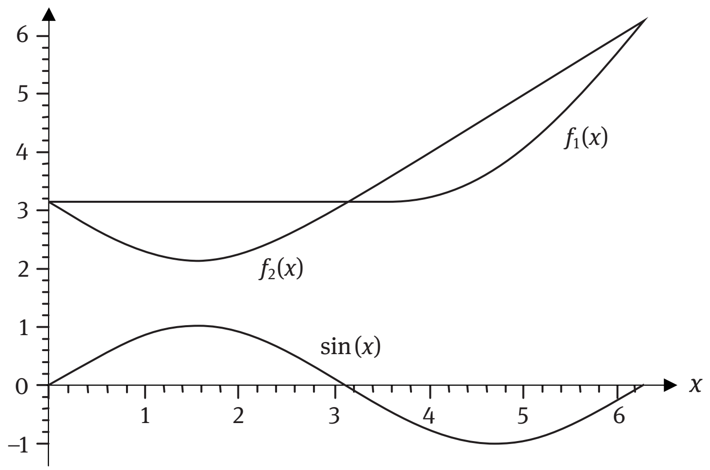


Fig. 11.10. D.c. function.

$i = 1, \dots, n$, the n-dimensional interval $I(a, b)$ is defined by

$$I(a, b) = \{(x_1, \dots, x_n)^T \in \mathbb{R}^n | a_i \le x_i \le b_i \text{ for } i = 1, \dots, n\} . \tag{11.31}$$

Other specific global optimization problems, which cannot be detailed here, are: biconvex optimization, complementarity problems (Section 9.2), fractional programming, minimax and maximin problems (Section 11.1), multiplicative programming and optimization with separable functions (Section 1.2.1). Further details on these subjects can be found in Horst and Tuy (1996) and the site of János D. Pintér: http://www.pinterconsulting.com/b.html.

11.2.2 Exact methods

The procedures for global optimization can be classified into exact approaches yielding an optimal solution and heuristics, which usually can provide only a "good" solution. We now study some exact methods, wherein only the first will be deepened.

Branch-and-bound method

This is a very general method for solving optimization problems, which is abbreviated as "B&B method". The term *branch* refers to the fact that subdivisions of the feasible set are constructed successively, while a *bound* denotes a lower or upper limit for the objective value of such a subdivision (see the summary in Figure 11.11).

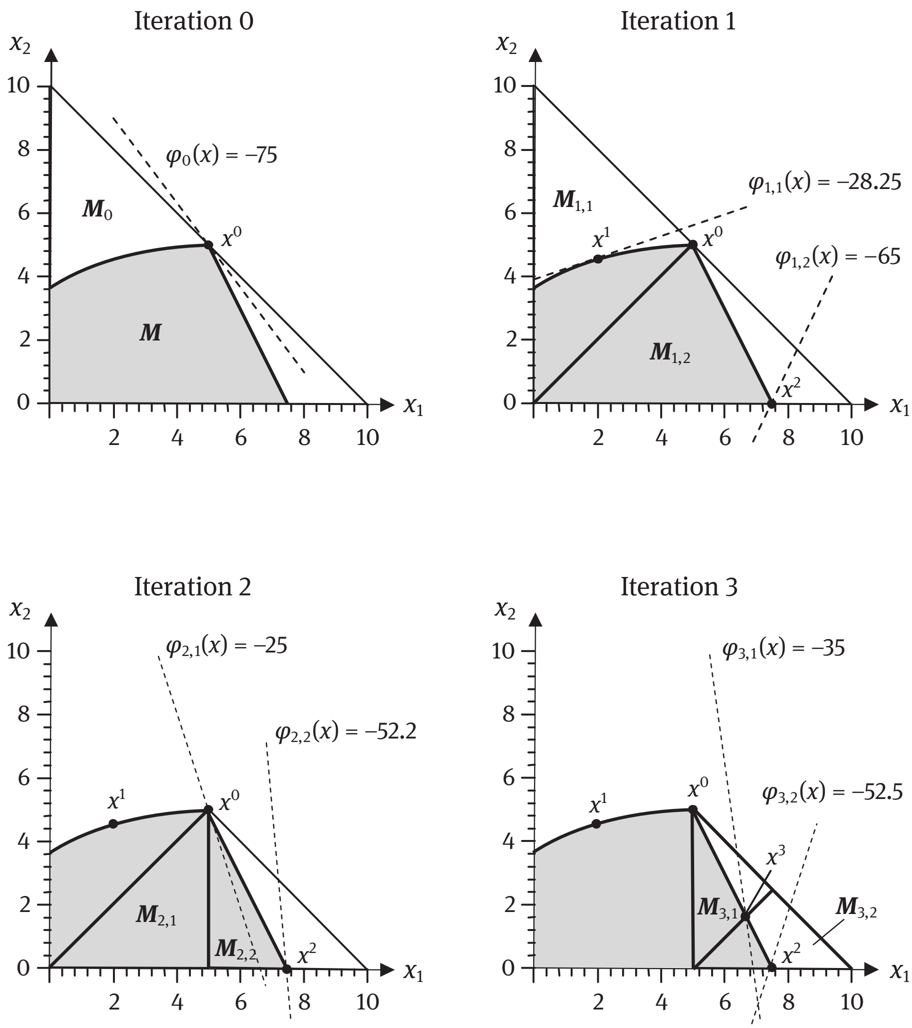


Fig. 11.11. Branch-and-bound method.

We present a rudimentary version of the B&B method to solve the global optimization problem $\min_{x\in M} f(x)$ (see Horst and Tuy (1996, Section IV.1)):

(1) Construct a relaxed feasible set $M_0 \supset M$ and subdivide M_0 into a finite number of subsets M_i, $i \in J$.

(2) For each M_i determine a lower and an upper limit $\alpha(M_i)$ and $\beta(M_i)$ for the minimum value of f over $M \cap M_i$, i.e. $\alpha(M_i) \le \inf_{i\in J}(M \cap M_i) \le \beta(M_i)$. Identify points in which f assumes the upper limit, i.e. determine feasible points $x^i \in M \cap M_i$ with $f(x^i) = \beta(M_i)$ for all $i \in J$. The values $\alpha := \min_{i\in J} \alpha(M_i)$ and $\beta := \min_{i\in J} \beta(M_i)$ are global limits for the minimum value of f, i.e. $\alpha \le \min f(M) \le \beta$.

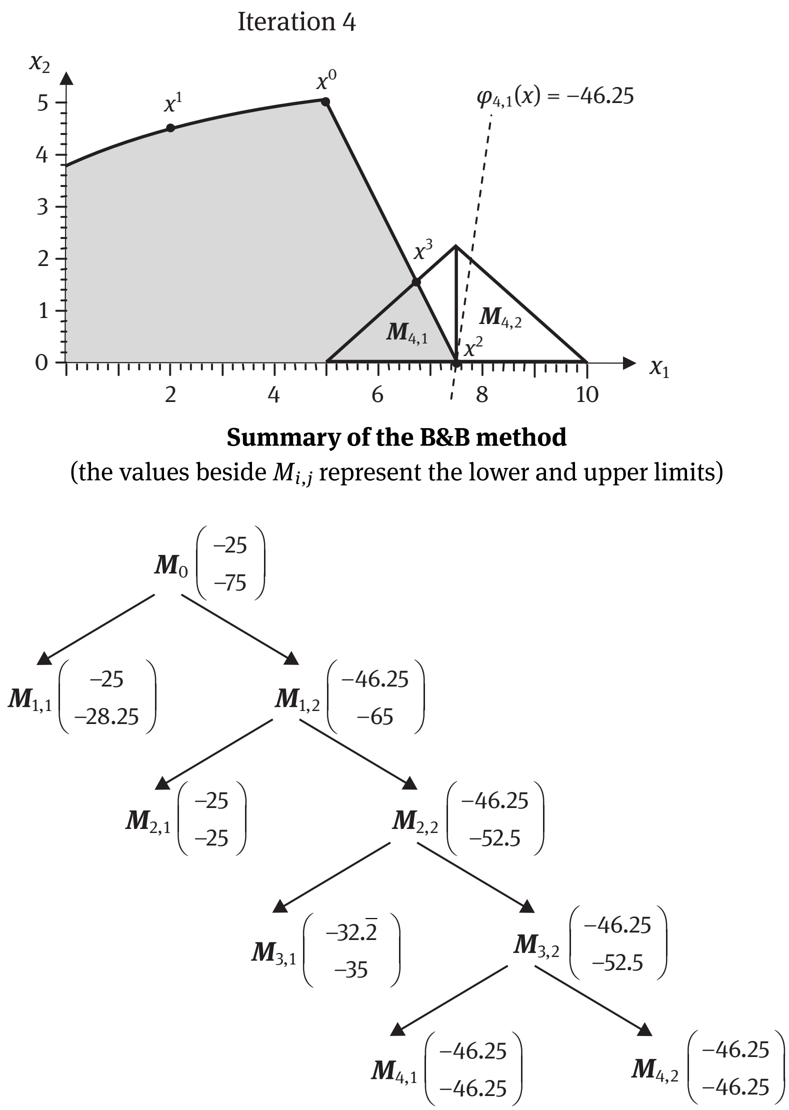


Fig. 11.11. Branch-and-bound method (continuation).

(3) If $\alpha = \beta$ (or $\beta - \alpha \leq \varepsilon$ for a predetermined $\varepsilon > 0$), stop. One of the points x^i determined in step (2) is optimal.

(4) Otherwise, eliminate the subsets M_i which cannot contain an optimal point ($\alpha_i > \beta$) and select thoroughly some of the remaining subsets ($\alpha_i < \beta$) to be further subdivided. Go to step (2).

Example 11.11. We illustrate the procedure, solving the problem (11.28).

Iteration 0:
As a relaxed feasible set we choose the triangle (simplex)

$$M_0 := \text{conv}\{(0,0)^T, (10,0)^T, (0,10)^T\} \quad \text{(see Figure 11.11).}$$

The other sets M_i are obtained by subdividing a given triangle along the bisector of the hypotenuse. To determine a lower limit we construct the affine linear function $\varphi_0(x_1, x_2) = a_1x_1 + a_2x_2 + a_3$ which coincides with f at the vertices of M_0. By solving the corresponding system of linear equations

$$\begin{aligned} a_3 &= -5 \\ 10a_1 + a_3 &= -85 \\ 10a_2 + a_3 &= -65\,, \end{aligned}$$

we get $\varphi_0(x_1, x_2) = -8x_1 - 6x_2 - 5$. Since f is concave, the function φ_0 underestimates f, i.e. $\varphi_0(x) \le f(x)$ for all $x \in M_0$. Now a lower limit α_0 can be determined by solving the convex problem

$$\min_{x \in M \cap M_0 = M} -8x_1 - 6x_2 - 5\,.$$

The optimal solution is $x^0 := (5,5)^T$ and the optimal value is $\alpha_0 = -75$. An upper limit is given by $\beta_0 := f(x^0) = -25$.

Iteration 1:
We subdivide M_0 into the triangles

$$M_{1,1} = \text{conv}\{(0,0)^T, (5,5)^T, (0,10)^T\}$$

and

$$M_{1,2} = \text{conv}\{(0,0)^T, (5,5)^T, (10,0)^T\}\,.$$

As in the previous iteration we determine affine linear functions $\varphi_{1,1}$ and $\varphi_{1,2}$, coincident with f at the vertices of $M_{1,1}$ e $M_{1,2}$, respectively. We obtain

$$\varphi_{1,1}(x_1, x_2) = 2x_1 - 6x_2 - 5 \quad \text{and} \quad \varphi_{1,2}(x_1, x_2) = -8x_1 + 4x_2 - 5\,.$$

Minimizing these functions over $M \cap M_{1,1}$ and $M \cap M_{1,2}$, respectively, we obtain the optimal points

$$x^1 := (5 - \sqrt{10}, \sqrt{90} - 5)^T \approx (1.84, 4.49)^T \quad \text{and} \quad x^2 := (7.5, 0)^T\,,$$

resulting in the lower limits

$$\alpha_{1,1} = 35 - 20\sqrt{10} \approx -28.25 \quad \text{and} \quad \alpha_{1,2} = -65\,.$$

The feasible points x^0, x^1 and x^2 constructed so far, provide upper limits for the minimum value of f over $M_{1,1}$ e $M_{1,2}$.

For $\beta_{1,1} := \min\{f(x^0), f(x^1)\} = -25$ and $\beta_{1,2} := \min\{f(x^0), f(x^2)\} = -46.25$ it holds evidently

$$\min f(M_{1,1}) \le \beta_{1,1} \quad \text{and} \quad \min f(M_{1,2}) \le \beta_{1,2}.$$

(Note that $M_{1,1}$ contains the points x^0 and x^1, and $M_{1,2}$ contains x^0 and x^2.)

Iteration 2:
The set $M_{1,1}$ can be eliminated: f has values $\ge \alpha_{1,1} \approx -28.25$ over this set, while $f(x^2) = -46.25$, i.e. $M_{1,1}$ cannot contain the solution of (11.28). We subdivide $M_{1,2}$ into the triangles

$$M_{2,1} = \text{conv}\{(0,0)^T, (5,0)^T, (5,5)^T\} \quad \text{and} \quad M_{2,2} = \text{conv}\{(5,0)^T, (5,5)^T, (10,0)^T\},$$

yielding the affine linear functions

$$\varphi_{2,1}(x_1, x_2) = -3x_1 - x_2 - 5 \quad \text{and} \quad \varphi_{2,2}(x_1, x_2) = -13x_1 - x_2 + 45.$$

Minimizing theses functions over $M \cap M_{2,1}$ and $M \cap M_{2,2}$, respectively, we obtain the solutions x^0 and x^2 with objective values $\alpha_{2,1} = -25$ and $\alpha_{2,2} = -52.5$. As upper limits we find

$$\beta_{2,1} := f(x^0) = -25 \quad \text{and} \quad \beta_{2,2} := \min\{f(x^0), f(x^2)\} = -46.25.$$

Iteration 3:
We can eliminate the set $M_{2,1}$ and subdivide $M_{2,2}$ into the triangles

$$M_{3,1} = \text{conv}\{(5,0)^T, (5,5)^T, (7.5,2.5)^T\}$$

and

$$M_{3,2} = \text{conv}\{(5,0)^T, (7.5,2.5)^T, (10,0)^T\},$$

which results in $\varphi_{3,1}(x_1, x_2) = -8x_1 - x_2 + 20$ and $\varphi_{3,2}(x_1, x_2) = -13x_1 + 4x_2 + 45$.

Minimizing these functions over $M \cap M_{3,1}$ and $M \cap M_{3,2}$, respectively, we obtain the optimal solutions $x^3 := (20/3, 5/3)^T$ and $x^2 := (15/2, 0)^T$ with objective values $\alpha_{3,1} = -35$ and $\alpha_{3,2} = -52.5$. The upper limits are

$$\beta_{3,1} := \min\{f(x^0), f(x^3)\} = -32.\bar{2} \quad \text{and} \quad \beta_{3,2} := \min\{f(x^2), f(x^3)\} = -46.25.$$

Iteration 4:
We eliminate $M_{3,1}$ and subdivide $M_{3,2}$ into

$$M_{4,1} = \text{conv}\{(5,0)^T, (7.5,0)^T, (7.5,2.5)^T\}$$

and

$$M_{4,2} = \operatorname{conv}\{(7.5, 0)^T, (7.5, 2.5)^T, (10, 0)^T\}.$$

We obtain

$$\varphi_{4,1}(x_1, x_2) = -10.5x_1 + 1.5x_2 + 32.5$$

and the minimum point over $M \cap M_{4,1}$ is x^2 with objective value $\alpha_{4,1} = -46.25$. The function $\varphi_{4,2}$ need not be calculated. Since $M \cap M_{4,2}$ consists of the unique point x^2, we obtain

$$\alpha_{4,2} = \beta_{4,2} = f(x^2) = -46.25.$$

Finally,

$$\beta_{4,1} = f(x^2) = -46.25.$$

So we have $\alpha = \beta$, and the algorithm terminates with the optimal solution $x^2 = (7.5, 0)^T$.

We still illustrate briefly, how the B&B method can be adapted to the Lipschitz optimization problem (11.30), taking advantage of the specific features of this problem. Since the feasible set is a multidimensional interval I, it is reasonable to construct sets M_i of this form. Starting with $M_0 := I$, the intervals will be subdivided successively into smaller ones according to diverse strategies. For at least one point of each interval, a value of f must be calculated that serves as an upper limit for the minimum value over this interval. Lower limits can be calculated by means of the relation (11.33) developed below:

If $M_{k,i}$ is the ith interval of the kth iteration and $x^{k,i}$ one of its points, then the Lipschitz continuity implies

$$\begin{aligned} &f(x^{k,i}) - f(x) \le L|x^{k,i} - x| \Rightarrow \\ &f(x) \ge f(x^{k,i}) - L\,|x^{k,i} - x|. \end{aligned} \tag{11.32}$$

If an upper limit A for $\sup_{x\in M}\{\,|\operatorname{grad} f(x)|\}$ is known (see (7.14)), we obtain

$$f(x) \ge f(x^{k,i}) - Ad(x^{k,i}), \quad \text{for all } x \in M_{k,i}, \tag{11.33}$$

where $d(x^{k,i})$ is a value satisfying $d(x^{k,i}) \ge |x^{k,i} - x|$ for all $x \in M_{k,i}$. Such a value can be easily determined. In particular, one can choose $d(x^{k,i}) := |b^{k,i} - a^{k,i}|$, if $M_{k,i}$ is the interval $I(a^{k,i}, b^{k,i})$.

Example 11.12. Consider the problem

$$\begin{aligned} &\min (x_1 - 2)^2 + (x_2 - 2)^2 \\ &\quad -2 \le x_1 \le 2 \\ &\quad -2 \le x_2 \le 2. \end{aligned} \tag{11.34}$$

Since $|\operatorname{grad} f(x)| = 2\sqrt{(x_1-2)^2+(x_2-2)^2}$, the smallest Lipschitz constant is the maximum value of this expression, i.e. $L = 2\sqrt{4^2+4^2} = 8\sqrt{2} \approx 11.31$. We choose the $M_{k,i}$ as squares and denote by $x^{i,k}$ the center of $M_{k,i}$. The value $d(x^{k,i})$ in (11.33) can be chosen as

$$d(x^{k,i}) = \frac{1}{2}\sqrt{c_{k,i}^2 + c_{k,i}^2} = c_{k,i}/\sqrt{2},$$

where $c_{k,i}$ is the length of a side of the square $M_{k,i}$. Now it follows from (11.33) that

$$f(x) \geq f(x^{k,i}) - 8c_{k,i}\,. \tag{11.35}$$

Iteration 0:
The center of $M_0 := M$ is $x^0 = (0,0)^T$. Relation (11.35) results in the lower limit $\alpha_0 = f(x^0) - 8 \cdot 4 = -24$, and as an upper limit we obtain $\beta_0 = f(x^0) = 8$.

Iteration 1:
We subdivide M_0 into the squares $M_{1,1}$, $M_{1,2}$, $M_{1,3}$ and $M_{1,4}$ (see Figure 11.12) with the centers

$$x^{1,1} = (-1,1)^T\,, \quad x^{1,2} = (1,1)^T\,, \quad x^{1,3} = (-1,-1)^T \quad \text{and} \quad x^{1,4} = (1,-1)^T\,.$$

For the minimum value of the objective function we obtain the lower limits $\alpha_{1,i} = f(x^{1,i}) - 8 \cdot 2$ and the upper limits $\beta_{1,i} = f(x^{1,i})$ for $i = 1,\dots,4$, i.e. $\alpha_{1,1} = -6$, $\alpha_{1,2} = -14$, $\alpha_{1,3} = 2$, $\alpha_{1,4} = -6$, $\beta_{1,1} = 10$, $\beta_{1,2} = 2$, $\beta_{1,3} = 18$, $\beta_{1,4} = 10$.

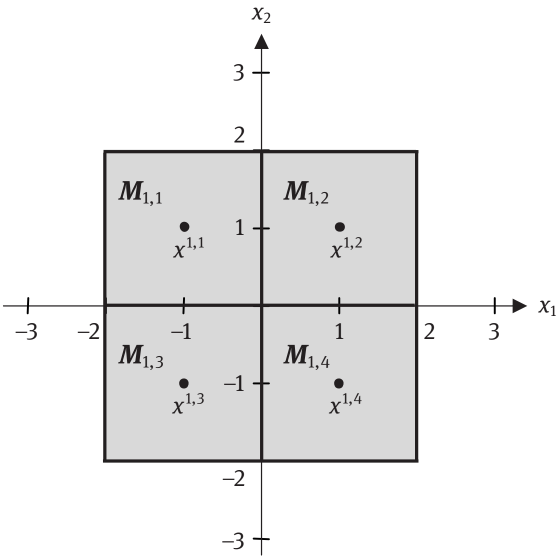


Fig. 11.12. Subdivisions of the feasible region.

In the next iteration the set $M_{1,3}$ can be eliminated, since the objective value over this set is ≥ 2, and the minimum value of f over $M_{1,2}$ is ≤ 2. Then the remaining sets $M_{1,1}, M_{1,2}$ and $M_{1,4}$ must be subdivided in smaller squares. The continuation of the example is left to the reader.

We finally list some other exact methods often used in practice.

"Naive" approaches

According to the name, these approaches are based on a simple idea, e.g. a search over a uniform grid, space covering sampling and pure random search. These methods converge under weak assumptions in the probabilistic sense, but are considered inadequate for large scale problems (see, e.g. Horst and Pardalos (1995) and Pintér (1996)).

Complete enumeration

These are search methods that determine all possible solutions, applicable to convex minimization and bilinear programming, among others. Naturally, complete enumeration may be impossible if the problem is too complex. For further details see, e.g. Horst and Tuy (1996, Section I.2.4).

Successive relaxation methods

The original problem is replaced by a sequence of "relaxed problems", which can be solved more easily. These problems are "refined" successively by means of cutting plane or more general techniques, such that the optimal solutions converge to the optimal solution of the initial problem. Successive relaxation methods can be applied to concave minimization and problems with d.c. functions, among others (see Horst and Tuy (1996, Section VI.3.1)).

11.2.3 Heuristic methods

The fact that classical optimization methods are frequently overburdened by the complexity of practical problems led to the development of several heuristic methods. We now turn to a universal approach which is not only applicable in global optimization, but also in several other areas of optimization, operations research, economics, artificial intelligence and social systems, etc.

Genetic algorithms

The basic idea is to apply biological principles of the natural evolution to an artificial system, for example an optimization model. The fundamentals of the theory are credited to Holland (1975). Recent literature and websites with diverse information are indicated in Section 11.3.

A genetic algorithm is an iterative procedure that simulates the development of a population of constant size. The individuals of the population are encoded by a string of symbols, called the *genome*, where the alphabet of symbols is usually binary. In the case of optimization the individuals correspond to "solution candidates" within a given *search space*. Essentially, a genetic algorithm operates as follows: the elements of an initial population are chosen at random from the search space. In each iteration, which is also called *evolutionary step* or *generation*, the individuals of the current population are encoded and evaluated according to a certain quality criterion, called *fitness*. In the case of optimization, the fitness is represented by the objective value of the solution candidate. To create the next generation, adequate individuals are chosen for the "reproduction". Usually the chance of being selected is proportional to the fitness. Thus, the most suitable individuals are more likely to be reproduced, while those with low fitness tend to disappear. In order to generate new individuals, corresponding to new points in the search space, genetically inspired operators are applied, the most important of which are called *crossover* and *mutation*. A crossover between two selected individuals (called parents) is realized with probability P_{cross} by exchanging parts of genomes. This operator aims to "move" the evolutionary process to "promising regions" of the search space. The mutation operator serves to prevent convergence to local optima (*premature convergence*), by randomly generating new points of the search space. This is performed by exchanging elements of the genome with a low probability P_{mut}. Genetic algorithms are iterative stochastic processes which do not necessarily converge to an optimal solution. As a stopping criterion one can choose a maximum number of iterations, previously specified, or the achievement of an acceptable level of fitness. The considerations are summarized in the following basic version of a genetic algorithm:

(1) Set $g = 0$ (generation counter) and determine the population $P(g)$.
(2) Evaluate the population $P(g)$, i.e. calculate the fitness of the individuals (objective values of the solutions).
(3) If the stopping criterion is satisfied, stop. Otherwise go to step (4).
(4) Determine $P(g+1)$ from $P(g)$ with the aid of crossover and mutation, set $g := g+1$ and go to step (2).

Example 11.13. We illustrate an evolutionary step on the basis of the problem (see Figure 11.13)

$$\max_{0\le x\le 31} x\sin(x/2) + 25\,. \tag{11.36}$$

From the above description it is clear that the search space M of a genetic algorithm must be finite. There is only a finite number of genomes with a given length.

For simplicity, we consider only integer values of the variable x, i.e. we choose $M = \{0, 1, \dots, 31\}$, resulting in a *"discrete approximation"* of the problem (11.36). We encode the values of x by a sequence a_4, a_3, a_2, a_1, a_0 of zeros and ones, such that

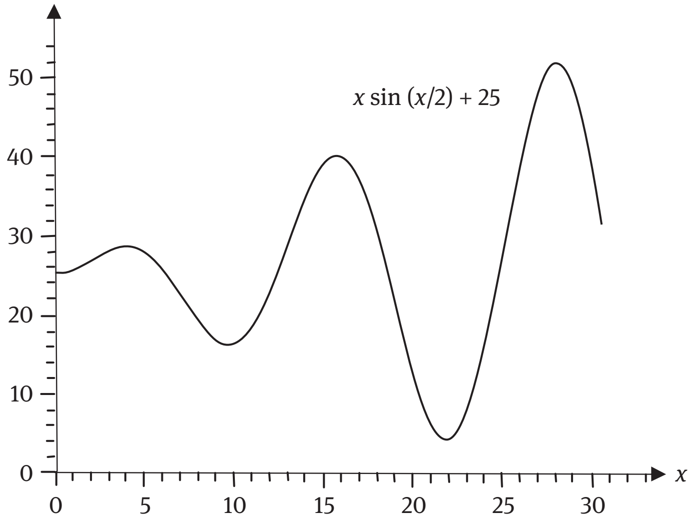


Fig. 11.13. Optimization problem (11.36).

$x = a_4 \cdot 2^4 + a_3 \cdot 2^3 + a_2 \cdot 2^2 + a_1 \cdot 2^1 + a_0 \cdot 2^0$. By using this binary code, the number 23 corresponds, for example, to the sequence (genome) 10111, because $23 = 1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$. Suppose that the initial population $P(0)$ consists of the four individuals in the following table, selected from the space M at random:

Table 11.2. Population $P(0)$.

x	Genome	Fitness $f(x)$	Relative fitness $f(x) / \sum f(x)$
5	00101	27.99	0.222
12	01100	21.65	0.172
19	10011	23.57	0.187
28	11100	52.74	0.419
Total		125.95	1

We select the individuals for reproduction with a probability proportional to the relative fitness, i.e. we select 5, 12, 19 and 28 with probabilities 22.2%, 17.2%, 18.7% and 41.9%, respectively. Suppose that the two pairs of "parents" are $\{19, 28\}$ and $\{5, 28\}$, i.e. the less fit individual 12 was not selected, and the fittest, 28, was chosen twice. Assume that a crossover between a pair of parents, which is realized with probability P_{cross}, results in exactly two "children". Suppose now that a crossover occurs between 19 and 28, causing a change in the second element of the genome (where the position of change was also determined randomly). This results in the genomes 11011 and 10100, corresponding to the children 27 and 20. Assume that no crossover occurs between 5 and 28, i.e. the two children are identical copies of the parents. Each of the four children 5, 20, 27 and 28 may be subject to an individual mutation with probability P_{mut}. Assume that 27 and 28 do not mutate, while 5 and 20 mutate at positions 5 and 2, respectively. This results in the genomes 00100 and 11100, corresponding to the individuals 4 and 28.

Table 11.3. Population $P(1)$.

x	Genome	Fitness $f(x)$	Relative fitness $f(x)/\sum f(x)$
4	00100	28.64	0.158
27	11011	46.72	0.258
28	11100	52.74	0.292
28	11100	52.74	0.292
Total		180.84	1

Note that this evolutionary step has increased the average fitness of the population from 31.49 to 45.21. Indeed, the individual 28 maximizes the fitness among all elements of the search space $M = \{0, 1, \ldots, 31\}$ (see Figure 11.13).

We emphasize here that the average fitness need not increase in each iteration since random mechanisms are underlying the reproduction. However, if the parameters P_{cross}, P_{mut}, etc., are well adjusted and if the method is applied to adequate problems (see below), a "stochastic convergence" to a population with good characteristics can be expected.

In the last years, genetic and evolutionary algorithms have been the subject of much applied research. The latter are similar to genetic algorithms, but dispense the coding by genomes (see Section 11.3). Indeed, such methods are flexible, robust, easily adaptable, do not need derivatives and can be efficiently combined with local search methods. However, the initial enthusiasm of the researchers was not justified by all practical applications. Genetic and evolutionary algorithms also have disadvantages and are not the most suitable procedures for all kinds of problems. For example, in solving linear problems they are much slower than the simplex method or the generalized reduced gradient method. In several practical tests, the computational time of "modern" procedures was at least a 100 times greater than that of the mentioned alternatives. Another problem is the lack of a concept of "optimal solution", i.e. there is no criterion to decide if a given solution is optimal or not. Therefore it is not certain when the ideal time is to finish the procedure. In practice, the algorithm has to be stopped by the user or when a certain predetermined criterion is satisfied.

In the following we outline some other heuristics of general character which can be applied to global optimization, among others.

Simulated annealing

It is essentially a modification of a search method in the maximum descent (ascent) direction. Starting with a random point of the search space, a random movement is performed. If the new iteration point is better, the movement will be accepted. Otherwise it is only accepted with a probability $p(t)$, where t is the time. For "small" values of t, $p(t)$ is close to one, but tends to zero as t increases. This is an analogy to the cooling of a solid. Initially, almost all movements are accepted, but when the "temperature" decreases, the probability of accepting moves that worsen the objective value is reduced. Such "negative moves" are occasionally necessary to leave a local optimum, but too many of them disturb the achievement of a global optimum. It can be proved that the method converges "almost always", but the result need not be a global optimum. As in the case of genetic algorithms it is difficult to decide if a given solution is optimal. It may be even difficult to identify a "good" solution. In practice it is common to use one of the following stopping criteria:

- A prefixed number of iterations is reached.
- The "temperature" falls below a certain minimum.
- The objective value reaches a given acceptable level or does not change any more significantly.

For further details see, e.g. Aarts and Lenstra (1997), Van Laarhoven and Aarts (1987).

Tabu search

Similar to the previous method it is a search method that allows the iterative generation of points that worsen the objective value. The essential idea is now to "forbid" (at least for some subsequent steps) search moves that lead to points already visited. The philosophy is to accept temporarily worse solutions to prevent the algorithms returning to "beaten paths". It is expected that new regions of the search space can be explored in this way, aiming to "escape" from local optima and ultimately determine a global optimum. Traditionally, tabu search has been applied to discrete problems, but can be also applied to continuous problems, using a discrete approximation (see Example 11.13 and Glover and Laguna (1997)).

"Globalized extensions" of local search methods

The basic idea is to start with an attempted search over uniform grids (see Section 11.2.2) or a random search. This first phase is to determine "promising" regions of the feasible set that deserve a more careful investigation. Then, in a second phase, a local search is performed which can be an exact method or a heuristic. Such hybrid procedures are used successfully in practice. The first phase may be replaced by a random selection of feasible points, which are adequate to start a local search. Such a selection is not trivial when the feasible region has a complex structure (see Aarts and Lenstra (1997), Glover and Laguna (1997), Van Laarhoven and Aarts (1987)).

11.3 References and software for Part II

For further reading on optimization methods we suggest Bazaraa and Shetty (1979) and Luenberger (2003). Several procedures of nonlinear programming are also studied in Friedlander (1994), Fritzsche (1978), Mahey (1987) and Mateus and Luna (1986). Comparisons of methods, test problems and some computer programs can be found in Himmelblau (1972), Schittkowski (1980, 1985) and Hock and Schittkowski (1981). Practical and numerical problems are discussed in Fletcher (1987) and Gill, Murray and Wright (1981). Among the more recent books may be cited Bazaraa, Sherali and Shetty (2006), Bertsekas (2004), Dahlquist and Björck (2003) and Gould and Leyffer (2003).

Stopping criteria for iterative procedures are introduced in Bazaraa and Shetty (1979, Section 7.2), the concept of convergence speed in Bazaraa and Shetty (1979, Section 7.4), Luenberger (2003, Section 6.7) and Fritzsche (1978, Section 4.5). Global convergence theorems are to be found in Dahlquist and Björck (2003, Section 6.3).

Various search methods for unrestricted problems, including the Golden Section algorithm, are studied in Wilde (1964), and Wilde and Beightler (1967), interpolation methods in Davidon (1959), Fletcher and Powell (1963) and in the site: http://www.korf.co.uk/spline.pdf. The conjugate direction method is credited to Fletcher and Reeves (1964). The method of Davidon, Fletcher and Powell was originally proposed by Davidon (1959) and improved by Fletcher and Powell (1963). Modern methods for unrestricted optimization (like inexact line search and trust region methods) are studied, e.g. in Conn, Gould and Toint (2000) and Gould and Leyffer (2003).

The methods of Sections 8.1.1 and 8.1.2 have been developed by Rosen (1960, 1961) and Zoutendijk (1960). As references for Sections 8.1.3 and 8.2 can be cited Fletcher (1987), Gill, Murray and Wright (1981) and Friedlander (1994).

An active set method for the quadratic problem is described in Luenberger (2003, Section 14.1). Lemke's method to solve the linear complementarity problem was published in Lemke (1968).

The use of penalty functions to solve restricted problems is usually credited to Courant. Camp (1955) and Pietrgykowski (1962) discussed this approach to solve nonlinear problems. But significant progress in the solution of practical problems is due to the classical paper of Fiacco and McCormick. The barrier function approach was originated by Carroll (1961), and Fletcher (1973) introduced the concept of the exact penalty function. A discussion of sequential quadratic programming can be found, e.g. in Fletcher (1987) and Nocedal and Wright (1999). Examples for nondifferential programming and solution methods like, e.g. the bundle method are studied in Clarke (1983), Lemarechal (1989), Wolfe (1975) and Zowe (1985).

Various specific cases of global optimization and diverse solution procedures are described in Horst and Tuy (1996) and in the article "The Relevance of Global Optimization" (http://pinterconsulting.com/l_s_d.html) of Janos D. Pintér which also provides several references about exact methods and heuristics.

Websites like “Nonlinear Programming Frequently Asked Questions” and the site http://www.mat.univie.ac.at/~neum/glopt.htlm of A. Neumaier contain diverse details about global optimization. Comprehensive information about all aspects of global optimization (theory, solution methods, software, test examples, applications and references) can be found in the online paper "Model Development, Solution, and Analysis in Global Optimization"of J. D. Pintér.

Introductory texts and summaries about genetic algorithms are to be found in the sites http://www.cs.bgu.ac.il/-sipper/ga_main.html of Moshe Sipper and http://aiinfinance.com of Franco Busetti. An overview on the application of evolutionary algorithms in constrained optimization is provided in Oliveira and Lorena (2002). The site http://www.solver.com of D. Fylstra treats among others the comparison between evolutionary algorithms and classical optimization. Information about software for global optimization can be found in http://plato.la.asu.edu/gom.html.

Diverse literature for nonlinear programming and information about software is presented in the site http://www.dpi.ufv.br/~heleno/sites.htm of the Federal University of Viçosa/Brazil. Certainly, one of the most useful websites for optimization is that of the NEOS Server: http://www.neos-Server.org/neos. This server provides state-of-the-art software free of charge to solve optimization problems, which have been sent to this site. However, only problems formulated in an appropriate modeling language are accepted.

We finally list some providers of nonlinear programming software. Except LANCELOT, all products are commercial, but most sellers offer a free trial version.

AIMMS (Advanced Interactive Multidimensional Modeling Software)
Seller: Paragon Decision Technology B. V., Netherlands
homepage: http://www.aimms.com
AIMMS is an algebraic modeling language that allows direct access to the fastest existing servers. The program permits that the user creates interactive models quickly and easily. The most notable advantage is that the user can concentrate on the specification of his application instead of worrying about the implementation. Via AIMMS the user has access to solvers like CPLEX, OSL, XA, CONOPT and MINOS.

GAMS (General Algebraic Modeling System)
Seller: GAMS Development Corporation, USA
homepage: http://www.gams.com
It is a sophisticated modeling system for operations research, applicable to large scale problems. The package allows the construction of complex models which can be quickly adapted to new situations. The language of GAMS is similar to common programming languages and enables a rapid familiarization of a user with programming experience. The creators value the easy handling of the program. For example, a whole set of related constraints can be introduced by means of a unique command. Location and type of an error are specified before running the program.

LINDO
Seller: LINDO Systems, Inc., USA
homepage: http://www.lindo.com
The program is easy to learn and solves preferentially linear problems but also some variants like quadratic problems. It provides online help for all commands. The user can learn by constructing his model. For better technical control the optimization procedure may be interrupted to provide the best current solution.

LINGO
Seller: LINDO Systems, Inc., USA
homepage: http://www.lindo.com
This product of LINDO Systems solves general nonlinear problems. The programming language allows a formulation as in mathematical texts. As in the case of GAMS, a series of similar constraints can be introduced by a unique command. The package provides various mathematical and probabilistic functions.

LOQO
Seller: Robert Vanderbei, Princeton University, USA
homepage: http://www.princeton.edu/~rvdb/
The program applies a primal dual interior point method. For convex problems a global optimum is attained. In the general case a local optimum close to the starting point is determined. The software solves large scale problems with thousands of variables and constraints.

MINOS
Seller: Stanford Business Software, Inc., USA
homepage: http://www.sbsi-sol-optimize.com
This is a software package for linear and nonlinear large scale problems. The program works more efficiently, when the constraints are linear. In this case, a reduced gradient method with quasi-Newton approximations to the Hessian matrix is used.

NPSOL
Seller: Stanford Business Software, USA
homepage: http://www.sbsi-sol-optimize.com
This software solves restricted nonlinear problems and is particularly efficient if the evaluation of the functions and gradients is complicated. In this case, NPSOL is superior to the program MINOS produced by the same company. NPSOL guarantees convergence for any starting point and is applicable to problems with up to some hundreds of variables and constraints.

OSL
Seller: OSL Development IBM Corporation, USA
homepage: http://oslsoftware.com
It is a software of high quality, where robust state-of-the-art algorithms have been implemented. Among other problems of operations research, quadratic problems can be solved. The software employs both solvers based on the simplex method as well as others, using an interior point method. The size of solvable problems is only limited by the available computer memory. OSL offers various possibilities to analyze the results like checking feasibility, sensitivity analysis and parametric analysis, etc.

LANCELOT
Contact: Nick Gould, Rutherford Appleton Laboratory, England
homepage: http://www.cse.scitech.ac.uk/nag/lancelot/howto.shtml
This noncommercial software solves large scale nonlinear problems. Versions of single and double precision are available.

Appendix: Solutions of exercises

Chapter 1

Exercise 1.2

We obtain

$$\begin{aligned}
\min\ -5x_1 + 2x_2^2 x_3 \\
-4x_1^3 + x_2 + 12 &\le 0 \\
x_1 x_3 + x_2^2 - 5 &\le 0 \\
-x_1 x_3 - x_2^2 + 5 &\le 0 \\
x_2 &\le 0\,.
\end{aligned}$$

Exercise 1.8

By setting

$$y_1 = \frac{x_1 + x_2}{2}\,, \quad y_2 = \frac{x_1 - x_2}{2}\,, \quad y_3 = \frac{x_3 + x_4}{2}\,, \quad y_4 = \frac{x_3 - x_4}{2}\,,$$

we get

$$\begin{aligned}
\min 2(y_1 + y_2)(y_1 - y_2) - (y_3 + y_4)(y_3 - y_4) \\
2y_1 \qquad + 2y_3 - 2y_4 &\le 30 \\
y_1 + y_2 \qquad + 2y_4 &\ge 20\,,
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
\min 2y_1^2 - 2y_2^2 - y_3^2 + y_4^2 \\
2y_1 \qquad + 2y_3 - 2y_4 &\le 30 \\
-y_1 - y_2 \qquad - 2y_4 &\le -20\,.
\end{aligned}$$

Exercise 1.11

The quantities of cake to be produced are given by (see (1.19)):

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} -3 & 0 & 1 \\ 0 & -4 & 0 \\ 2 & 0 & -5 \end{pmatrix} \begin{pmatrix} 44.68 \\ 28.24 \\ 41.32 \end{pmatrix} + \begin{pmatrix} 100 \\ 120 \\ 160 \end{pmatrix} = \begin{pmatrix} 7.28 \\ 7.04 \\ 42.76 \end{pmatrix}$$

and the constraints (1.14) are satisfied.

Exercise 1.12

(a)

$$Q = \frac{P + P^T}{2} = \frac{1}{2}\left(\begin{pmatrix} 1 & 4 \\ 2 & 5 \end{pmatrix} + \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}\right) = \frac{1}{2}\begin{pmatrix} 2 & 6 \\ 6 & 10 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 3 & 5 \end{pmatrix}.$$

It holds

$$(x_1, x_2)\begin{pmatrix} 1 & 4 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 + 6x_1x_2 + 5x_2^2 = (x_1, x_2)\begin{pmatrix} 1 & 3 \\ 3 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

(b) Obviously, $q = (3, 4)^T$. The matrix

$$Q = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

satisfies

$$\frac{1}{2}(x_1, x_2)\begin{pmatrix} a & b \\ b & c \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{2}ax_1^2 + bx_1x_2 + \frac{1}{2}cx_2^2 = -2x_1^2 + 5x_1x_2 + 8x_2^2.$$

This implies $a = -4$, $b = 5$, $c = 16$, i.e.

$$Q = \begin{pmatrix} -4 & 5 \\ 5 & 16 \end{pmatrix}.$$

Exercise 1.13

(i) (a) no (b), (c), (d) yes

(ii) (a) no

(b) no, but can be transformed in a separable problem

(c), (d) yes

(iii) (a), (b) yes (c), (d) no

(iv) (a), (b), (c) no (d) yes

(v) (a), (c) no (b), (d) yes

Exercise 1.16

With

$$f(x_1) := x_1^2 + 8000/x_1,$$

we get

$$f'(x_1) = 2x_1 - 8000/x_1^2 = 0 \Leftrightarrow x_1 = 4000^{1/3}.$$

Hence, $x_1^* = 4000^{1/3}\text{cm} = 15.9\,\text{cm}$, and $x_2^* = 2000/(x_1^*)^2 = 7.9\,\text{cm}$. The minimum amount of metal foil required is $(x_1^*)^2 + 4x_1^*x_2^* = 3(x_1^*)^2 = 756\,\text{cm}^2$.

Exercise 1.24

(i) The local minima are $x_1 = -3\pi/4$ and $x_2 = 5\pi/4$. Both are strict, but only x_1 is a global minimum point.

(ii) Each point of the interval $[0, 1]$ is a local and global minimum point. None of these points represents a strict minimum.

(iii) The unique local minimum point is $x^* = 0$. It is a strict and global minimum point.

(iv) Each point $x = (x_1, x_2)^T$ with $x_1 = 1 + 1/\sqrt{3}, x_2 \in \mathbb{R}$ is a local minimum point. None of these points is a strict or global minimum point.

Chapter 2

Exercise 2.9

(a) It holds

$$g^1 := \operatorname{grad} g_1(x^*) = \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \quad g^2 := \operatorname{grad} g_2(x^*) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and

$$\begin{aligned} Z(x^*) = L(x^*) &= \left\{ d \in \mathbb{R}^2 \setminus \{0\} \,\middle|\, d^T \begin{pmatrix} 1 \\ 4 \end{pmatrix} \le 0\,, \quad d^T \begin{pmatrix} 1 \\ 1 \end{pmatrix} \le 0 \right\} \\ &= \left\{ \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \in \mathbb{R}^2 \setminus \{0\} \,\middle|\, d_1 + 4d_2 \le 0, d_1 + d_2 \le 0 \right\}. \end{aligned}$$

(Theorem 2.6)

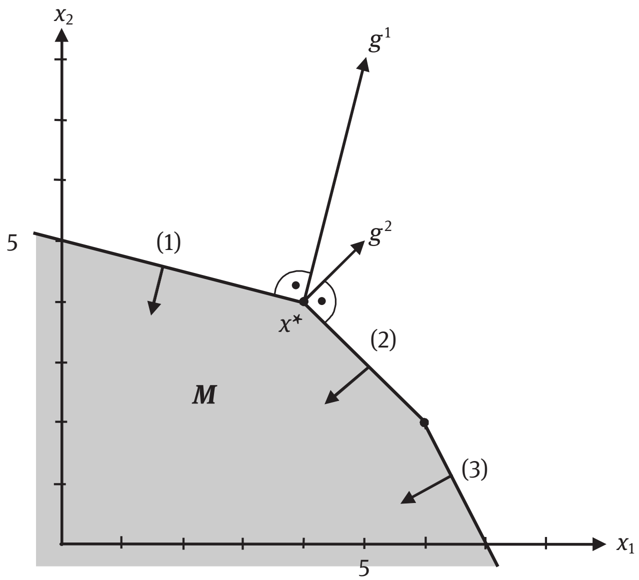


Fig. 2.7. Feasible set and gradients.

(b) It holds

$$\operatorname{grad} g_1(x) = \begin{pmatrix} 3(x_1-3)^2 \\ 1 \end{pmatrix}, \qquad g_1 := \operatorname{grad} g_1(x^*) = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$
$$\operatorname{grad} g_2(x) = \begin{pmatrix} 3(x_1-3)^2 \\ -1 \end{pmatrix}, \qquad g_2 := \operatorname{grad} g_2(x^*) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

Since the tangent lines to the curves (1) and (2) at the point x^* are horizontal, the cone of feasible directions is

$$Z(x^*) = \{d \in \mathbb{R}^2 | d_1 < 0, d_2 = 0\}\,.$$

From Theorem 2.5 it follows that

$$\begin{aligned} L_0(x^*) &= \left\{ d \in \mathbb{R}^2 \setminus \{0\} \,\middle|\, d^T \begin{pmatrix} 0 \\ 1 \end{pmatrix} < 0\,, d^T \begin{pmatrix} 0 \\ -1 \end{pmatrix} < 0 \right\} = \emptyset \subset Z(x^*) \subset \\ L(x^*) &= \left\{ d \in \mathbb{R}^2 \setminus \{0\} \,\middle|\, d^T \begin{pmatrix} 0 \\ 1 \end{pmatrix} \le 0, d^T \begin{pmatrix} 0 \\ -1 \end{pmatrix} \le 0 \right\} \\ &= \left\{ \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \in \mathbb{R}^2 \,\middle|\, d_1 \ne 0, d_2 = 0 \right\}\,. \end{aligned}$$

Therefore, we have the case $L_0(x^*) \ne Z(x^*) \ne L(x^*)$ (see Example 2.8).

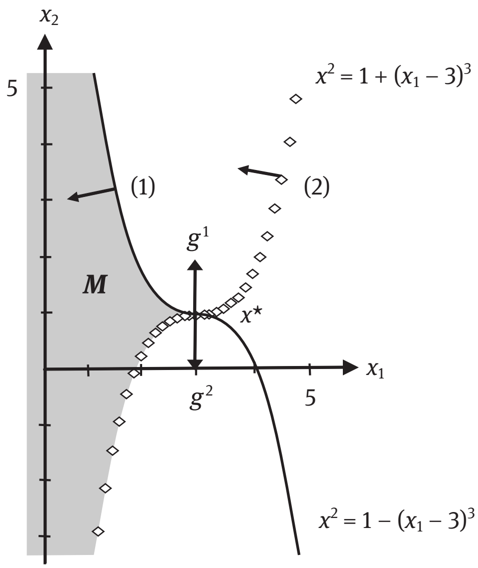


Fig. 2.8. Feasible region and gradients.

(c) Since M consists of the unique point $(0,0)^T$, we get $Z(x^*) = \emptyset$.
It holds $\operatorname{grad} g(x^*) = \binom{0}{0}$ and Theorem 2.5 implies

$$L_0(x^*) = \left\{ d \in \mathbb{R}^2 \setminus \{0\} \,\middle|\, d^T \begin{pmatrix} 0 \\ 0 \end{pmatrix} < 0 \right\} = \emptyset \subset Z(x^*) \subset L(x^*)$$
$$= \left\{ d \in \mathbb{R}^2 \setminus \{0\} \,\middle|\, d^T \begin{pmatrix} 0 \\ 0 \end{pmatrix} \leq 0 \right\} = \mathbb{R}^2 \setminus \{0\}\,.$$

This is the case $L_0(x^*) = Z(x^*) \neq L(x^*)$.

Exercise 2.14

(a) The condition $\operatorname{grad} f(x_1, x_2) = \binom{2x_1}{6x_2} = 0$ implies $x_1 = x_2 = 0$.

(b) $\operatorname{grad} f(x_1, x_2) = \binom{x_2 e^{x_1}}{e^{x_1}} = 0$ cannot be satisfied, since $e^{x_1} > 0$ for all x_1.

(c) $\operatorname{grad} f(x_1, x_2) = \binom{-\sin x_1}{2x_2} = 0$ is satisfied, if and only if $x_2 = 0$ and $x_1 = z\pi$ for an integer z.

Exercise 2.15

(a) It holds $\operatorname{grad} f(x_1, x_2) = \binom{2(1-x_1)}{2(1-x_2)}$.

Case 1: $x_1^* = x_2^* = 0$
We get $Z(x^*) = \{d \in \mathbb{R}^2 \setminus \{0\} \mid d_1, d_2 \geq 0\}$ (see Figure 2.9).
The relation

$$d^T \operatorname{grad} f(x^*) = (d_1, d_2) \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2d_1 + 2d_2 \geq 0$$

is satisfied for all $d \in Z(x^*)$. Thus $x^* = (0,0)^T$ satisfies the necessary condition.

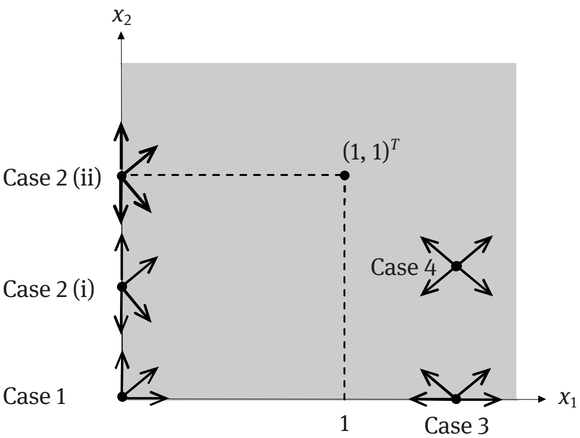


Fig. 2.9. Feasible directions at x^* in diverse cases.

Case 2: $x_1^* = 0, x_2^* > 0$
It holds $Z(x^*) = \{d \in \mathbb{R}^2 \setminus \{0\} | d_1 \geq 0\}$
(i) $x_2^* \neq 1$: Since $1 - x_2^* \neq 0$, the relation

$$d^T \operatorname{grad} f(x^*) = (d_1, d_2) \begin{pmatrix} 2 \\ 2(1 - x_2^*) \end{pmatrix} = 2d_1 + 2d_2(1 - x_2^*) \geq 0$$

is not satisfied for all $d \in Z(x^*)$.
(ii) $x_2^* = 1$: It holds

$$d^T \operatorname{grad} f(x^*) = (d_1, d_2) \begin{pmatrix} 2 \\ 0 \end{pmatrix} = 2d_1 \geq 0$$

for all $d \in Z(x^*)$, i.e. $x^* = (0, 1)^T$ satisfies the necessary condition.

Case 3: $x_1^* > 0, x_2^* = 0$
Similar to the previous case, only the point $x^* = (1, 0)^T$ satisfies the necessary condition.

Case 4: $x_1^*, x_2^* > 0$
It holds $Z(x^*) = \mathbb{R}^2 \setminus \{0\}$
(i) $x_1^* \neq 1$ or $x_2^* \neq 1$: The relation $d^T \operatorname{grad} f(x^*) = 2d_1(1 - x_1^*) + 2d_2(1 - x_2^*) \geq 0$ is not satisfied for all $d \in Z(x^*)$.
(ii) $x_1^* = x_2^* = 1$: Now it holds $d^T \operatorname{grad} f(x^*) = 0$, i.e. $x^* = (1, 1)^T$ satisfies the necessary condition.

Summarizing the above results, we can state that the points $(0,0)^T$, $(0,1)^T$, $(1,0)^T$, $(1,1)^T$ (and only these points) satisfy the necessary condition of Theorem 2.10.

(b) The problem of the exercise is equivalent to maximizing the squared distance between the points $(x_1, x_2)^T$ and $(1, 1)^T$ for $x_1, x_2 \geq 0$. Therefore, only the point $(0, 0)^T$ is a local optimum, and there exists no global optimum.

Exercise 2.20

(a) It holds

$$\operatorname{grad} f(x_1, x_2) = \begin{pmatrix} 3x_1^2 x_2^3 \\ 3x_1^3 x_2^2 \end{pmatrix}, \quad Hf(x) = \begin{pmatrix} 6x_1 x_2^3 & 9x_1^2 x_2^2 \\ 9x_1^2 x_2^2 & 6x_1^3 x_2 \end{pmatrix},$$

$$\operatorname{grad} f(x^*) = 0, \quad Hf(x^*) = 0.$$

Therefore, the condition of Theorem 2.19 is satisfied.

(b) For each ε-neighborhood

$$U_\varepsilon(x^*) = \{(x_1, x_2)^T \in \mathbb{R}^2 | x_1^2 + x_2^2 \leq \varepsilon\} \quad (\varepsilon > 0)$$

of x^* exists a point $\overline{x} \in U_\varepsilon(x^*)$ such that $f(\overline{x}) < f(x^*)$. For example, for $\overline{x} = (-\varepsilon/2, \varepsilon/2)$ it holds

$$f(\overline{x}) = (-\varepsilon/2)^3 (\varepsilon/2)^3 = -\varepsilon^6/64 < 0 = f(x^*).$$

Exercise 2.22

(a) We obtain

$$\operatorname{grad} f(x_1, x_2) = \begin{pmatrix} 6(x_1 - 2) \\ 10(x_2 - 3) \end{pmatrix}, \quad Hf(x_1, x_2) = \begin{pmatrix} 6 & 0 \\ 0 & 10 \end{pmatrix},$$

and the condition $\operatorname{grad} f(x_1, x_2) = 0$ implies $x_1 = 2$, $x_2 = 3$. The point $x^* := \binom{2}{3}$ satisfies the two conditions of Theorem 2.19.

Since $d^T Hf(x^*)d = 6d_1^2 + 10d_2^2 > 0$, for all $d \in \mathbb{R}^2\backslash\{0\}$(see Theorem 2.21), x^* is a local minimum point.

(b) It holds

$$\operatorname{grad} f(x_1, x_2) = \begin{pmatrix} -\sin x_1 \\ 2x_2 \end{pmatrix}, \quad Hf(x_1, x_2) = \begin{pmatrix} -\cos x_1 & 0 \\ 0 & 2 \end{pmatrix}.$$

A point $(x_1, x_2)^T$ satisfies condition (i) of Theorem 2.19, if and only if $x_1 = z\pi$ for an integer z and $x_2 = 0$. Such a point also satisfies condition (ii) of this theorem, if and only if

$$d^T Hf(x^*)d = (d_1, d_2) \begin{pmatrix} -\cos z\pi & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = -d_1^2 \cos z\pi + 2d_2^2 \geq 0$$

for all $d \in \mathbb{R}^2$.

Hence, the points $(z\pi, 0)^T$ with an odd integer z (and only these points) satisfy the two conditions of Theorem 2.19. They are local minimum points, since they satisfy the sufficient conditions of Theorem 2.21.

Exercise 2.23

It holds

$$f(x_1, x_2) = -(x_1 - 1)^2 - (x_2 - 1)^2, \quad \operatorname{grad} f(x_1, x_2) = \begin{pmatrix} 2(1 - x_1) \\ 2(1 - x_2) \end{pmatrix},$$
$$Hf(x_1, x_2) = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}.$$

As we have seen in Example 2.15 we need to check only the points

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Point $x^* = \binom{0}{0}$

It holds

$$\operatorname{grad} f(x^*) = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad Z(x^*) = \{d \in \mathbb{R}^2\backslash\{0\} \mid d_1 \geq 0\,, d_2 \geq 0\}\,.$$

If $d \in Z(x^*)$ and $d^T \operatorname{grad} f(x^*) = 2d_1 + 2d_2 = 0$, we have $d_1 = d_2 = 0$.

Therefore, $d^T Hf(x^*)d = 0$, i.e. condition (ii) of Theorem 2.16 is satisfied.

Point $x^* = \binom{0}{1}$
It holds

$$\operatorname{grad} f(x^*) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad Z(x^*) = \{d \in \mathbb{R}^2 \backslash \{0\} \,|\, d_1 \geq 0\}\,.$$

Setting, e.g. $d = \binom{0}{1}$, we get $d \in Z(x^*)$ and $d^T \operatorname{grad} f(x^*) = 0$, but

$$d^T Hf(x^*) d = (0, 1) \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -2 < 0\,.$$

Therefore, condition (ii) is not satisfied.

Point $x^* = \binom{1}{0}$
As in the previous case one can show that condition (ii) is not satisfied.

Point $x^* = \binom{1}{1}$
We get

$$\operatorname{grad} f(x^*) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad Z(x^*) = \mathbb{R}^2 \backslash \{0\}\,.$$

For example, for $d = \binom{1}{0}$ it holds $d \in Z(x^*)$ and $d^T \operatorname{grad} f(x^*) = 0$, but

$$d^T Hf(x^*) d = (1, 0) \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = -2 < 0\,.$$

Therefore, condition (ii) is not satisfied.

Exercise 2.30

(a) Since $3 > 0$ and $\left|\begin{smallmatrix} 3 & 2 \\ 2 & 4 \end{smallmatrix}\right| = 8 > 0$, the matrix is positive definite.

(b) We get

$$|Q - \lambda I| = \begin{vmatrix} 9-\lambda & 6 \\ 6 & 4-\lambda \end{vmatrix} = (9-\lambda)(4-\lambda) - 36 = \lambda(\lambda - 13)\,,$$

i.e. the eigenvalues are $\lambda_1 = 0$ and $\lambda_2 = 13$, and the matrix is positive semidefinite.

(c) We get

$$|Q - \lambda I| = \begin{vmatrix} 2-\lambda & -4 \\ -4 & 3-\lambda \end{vmatrix} = (2-\lambda)(3-\lambda) - 16 = \lambda^2 - 5\lambda - 10\,.$$

The eigenvalues are $\lambda_1 = \frac{5+\sqrt{65}}{2} > 0$ and $\lambda_2 = \frac{5-\sqrt{65}}{2} < 0$, therefore, the matrix is indefinite.

(d) It holds

$$15 > 0\,, \quad \begin{vmatrix} 15 & 3 \\ 3 & 1 \end{vmatrix} = 6 > 0\,, \quad \begin{vmatrix} 15 & 3 & 2 \\ 3 & 1 & 0 \\ 2 & 0 & 1 \end{vmatrix} = 2 > 0\,.$$

Thus the matrix is positive definite.

(e) Since $q_{2,2} = -3 < 0$, the matrix is not positive semidefinite.

Exercise 2.31

We obtain the following principal minors:

$$k=1,\quad i_1=1:\quad q_{1,1}=2$$

$$k=1,\quad i_1=2:\quad q_{2,2}=8$$

$$k=1,\quad i_1=3:\quad q_{3,3}=4$$

$$k=1,\quad i_1=4:\quad q_{4,4}=15$$

$$k=2,\quad i_1=1,\quad i_2=2:\quad \begin{vmatrix} q_{1,1} & q_{1,2} \\ q_{2,1} & q_{2,2} \end{vmatrix} = \begin{vmatrix} 2 & 3 \\ 3 & 8 \end{vmatrix} = 2\cdot 8-3\cdot 3=7$$

$$k=2,\quad i_1=1,\quad i_2=3:\quad \begin{vmatrix} q_{1,1} & q_{1,3} \\ q_{3,1} & q_{3,3} \end{vmatrix} = \begin{vmatrix} 2 & -1 \\ -1 & 4 \end{vmatrix} = 7$$

$$k=2,\quad i_1=1,\quad i_2=4:\quad \begin{vmatrix} q_{1,1} & q_{1,4} \\ q_{4,1} & q_{4,4} \end{vmatrix} = \begin{vmatrix} 2 & 5 \\ 5 & 15 \end{vmatrix} = 5$$

$$k=2,\quad i_1=2,\quad i_2=3:\quad \begin{vmatrix} q_{2,2} & q_{2,3} \\ q_{3,2} & q_{3,3} \end{vmatrix} = \begin{vmatrix} 8 & 2 \\ 2 & 4 \end{vmatrix} = 28$$

$$k=2,\quad i_1=2,\quad i_2=4:\quad \begin{vmatrix} q_{2,2} & q_{2,4} \\ q_{4,2} & q_{4,4} \end{vmatrix} = \begin{vmatrix} 8 & 1 \\ 1 & 15 \end{vmatrix} = 119$$

$$k=2,\quad i_1=3,\quad i_2=4:\quad \begin{vmatrix} q_{3,3} & q_{3,4} \\ q_{4,3} & q_{4,4} \end{vmatrix} = \begin{vmatrix} 4 & 6 \\ 6 & 15 \end{vmatrix} = 24$$

$$k=3,\quad i_1=1,\quad i_2=2,\quad i_3=3:\quad \begin{vmatrix} q_{1,1} & q_{1,2} & q_{1,3} \\ q_{2,1} & q_{2,2} & q_{2,3} \\ q_{3,1} & q_{3,2} & q_{3,3} \end{vmatrix} = \begin{vmatrix} 2 & 3 & -1 \\ 3 & 8 & 2 \\ -1 & 2 & 4 \end{vmatrix} = 0$$

$$k=3,\quad i_1=1,\quad i_2=2,\quad i_3=4:\quad \begin{vmatrix} q_{1,1} & q_{1,2} & q_{1,4} \\ q_{2,1} & q_{2,2} & q_{2,4} \\ q_{4,1} & q_{4,2} & q_{4,4} \end{vmatrix} = \begin{vmatrix} 2 & 3 & 5 \\ 3 & 8 & 1 \\ 5 & 1 & 15 \end{vmatrix} = -67$$

$$k=3,\quad i_1=1,\quad i_2=3,\quad i_3=4:\quad \begin{vmatrix} q_{1,1} & q_{1,3} & q_{1,4} \\ q_{3,1} & q_{3,3} & q_{3,4} \\ q_{4,1} & q_{4,3} & q_{4,4} \end{vmatrix} = \begin{vmatrix} 2 & -1 & 5 \\ -1 & 4 & 6 \\ 5 & 6 & 15 \end{vmatrix} = -127$$

$$k=3,\quad i_1=2,\quad i_2=3,\quad i_3=4:\quad \begin{vmatrix} q_{2,2} & q_{2,3} & q_{2,4} \\ q_{3,2} & q_{3,3} & q_{3,4} \\ q_{4,2} & q_{4,3} & q_{4,4} \end{vmatrix} = \begin{vmatrix} 8 & 2 & 1 \\ 2 & 4 & 6 \\ 1 & 6 & 15 \end{vmatrix} = 152$$

$$k=4,\quad i_1=1,\quad i_2=2,\quad i_3=3,\quad i_4=4:\quad \det(Q) = -1575.$$

The successive principal minors are 2, 7, 0, and -1575. Since there are negative principal minors, the matrix Q is not positive semidefinite.

Exercise 2.32
We get

$$\operatorname{grad} f(x_1, x_2) = \begin{pmatrix} 3x_1 + 4x_2 \\ 4x_1 + 10x_2 \end{pmatrix}, \quad Hf(x_1, x_2) = \begin{pmatrix} 3 & 4 \\ 4 & 10 \end{pmatrix}.$$

The condition $\operatorname{grad} f(x_1, x_2) = 0$ implies $x_1 = x_2 = 0$. Since $Hf(0,0)$ is positive definite, the point $(0,0)^T$ is the unique local (and global) minimum point.

Chapter 3

Exercise 3.3
Given convex sets $K, L \subset \mathbb{R}^n$ and points $x, y \in K \cap L$. Since K and L are convex, it holds $[x, y] \subset K$ and $[x, y] \subset L$, thus $[x, y] \subset K \cap L$, i.e. $K \cap L$ is convex.

Exercise 3.7
(i) From Definition 3.1 it follows that the statement

$$\sum_{i=1}^{m} \lambda_i x^i \in M \quad \text{for } \lambda_1, \dots, \lambda_m \geq 0\,, \quad \lambda_1 + \cdots + \lambda_m = 1\,, \quad x^1, \dots, x^m \in M \quad (*)$$

is valid for $m = 2$. Now we assume that $(*)$ is valid for $m = m_0$ with $m_0 \geq 2$. We show that under this assumption $(*)$ is also valid for $m = m_0 + 1$:
For $\lambda_1, \dots, \lambda_{m_0} \geq 0$, $\lambda_{m_0+1} > 0$, $\lambda_1 + \cdots + \lambda_{m_0+1} = 1$, we obtain

$$\sum_{i=1}^{m_0+1} \lambda_i x^i = (1 - \lambda_{m_0+1}) \sum_{i=1}^{m_0} \frac{\lambda_i}{1 - \lambda_{m_0+1}} x^i + \lambda_{m_0+1} x^{m_0+1}\,.$$

From the assumption follows that the sum on the right side belongs to M (note that $\sum_{i=1}^{m_0} \frac{\lambda_i}{1-\lambda_{m_0+1}} = 1$), therefore Definition 3.1 implies that $\sum_{i=1}^{m_0+1} \lambda_i x^i \in M$. The case $\lambda_{m_0+1} = 0$ is trivial.

(ii) Given two points $z^1, z^2 \in \operatorname{conv}(K)$. Then it holds that

$$z^1 = \sum_{i=1}^{m} \lambda_i x^i, \quad z^2 = \sum_{j=1}^{p} \mu_j y^j$$

with $\lambda_i \geq 0$ for $i = 1, \dots, m$, $\sum_{i=1}^{m} \lambda_i = 1$, $\mu_j \geq 0$ for $j = 1, \dots, p$, $\sum_{j=1}^{p} \mu_j = 1$ and $x^1, \dots, x^m, y^1, \dots, y^p \in K$. For λ with $0 \leq \lambda \leq 1$ we obtain:

$$\lambda z^1 + (1 - \lambda) z^2 = \sum_{i=1}^{m} \lambda\, \lambda_i x^i + \sum_{j=1}^{p} (1 - \lambda)\, \mu_j y^i\,.$$

Since $\lambda\, \lambda_i \geq 0$ for $i = 1, \dots, m$, $(1 - \lambda)\, \mu_j \geq 0$ for $j = 1, \dots, p$ and

$$\lambda\, \lambda_1 + \cdots + \lambda\, \lambda_m + (1 - \lambda)\, \mu_1 + \cdots + (1 - \lambda)\, \mu_p = \lambda + (1 - \lambda) = 1\,,$$

we get $\lambda z^1 + (1 - \lambda) z^2 \in \operatorname{conv}(K)$.

Exercise 3.13

(i) Given $x, y \in K_1 + K_2$. We obtain

$$x = x^1 + x^2, \quad y = y^1 + y^2$$

with $x^1, y^1 \in K_1$ and $x^2, y^2 \in K_2$. For λ with $0 \le \lambda \le 1$ we get:

$$\begin{aligned} \lambda x + (1-\lambda)y &= \lambda(x^1 + x^2) + (1-\lambda)(y^1 + y^2) \\ &= \lambda x^1 + (1-\lambda)y^1 + \lambda x^2 + (1-\lambda)y^2 . \end{aligned} \tag{$*$}$$

The convexity of K_1 and K_2 implies that

$$\lambda x^1 + (1-\lambda)y^1 \in K_1 \quad \text{and} \quad \lambda x^2 + (1-\lambda)y^2 \in K_2.$$

Hence, the expression $(*)$ is an element of $K_1 + K_2$, i.e. $K_1 + K_2$ is convex (see Figure 3.18).

(ii)

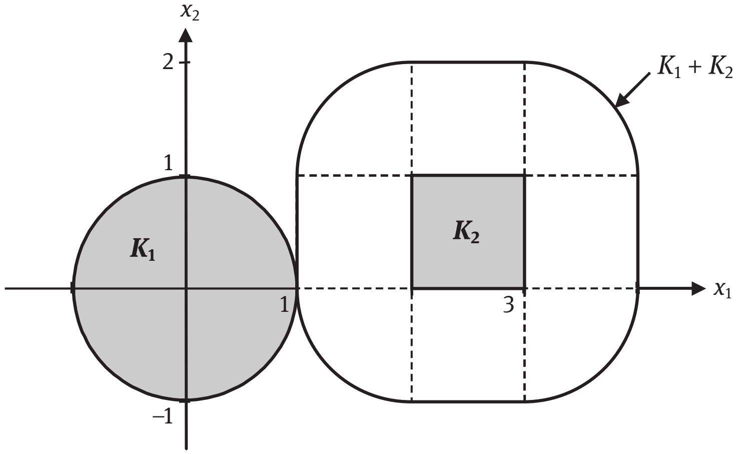


Fig. 3.18. Algebraic sum.

Exercise 3.15

(i) We have to prove the relation

$$f((1-\lambda)x + \lambda y) < (1-\lambda)f(x) + \lambda f(y) \quad \text{for } x \ne y, 0 < \lambda < 1. \tag{1}$$

This inequality is equivalent to:

$$\begin{aligned} [(1-\lambda)x + \lambda y]^2 &< (1-\lambda)x^2 + \lambda y^2 \Leftrightarrow \\ (1-\lambda)^2x^2 + \lambda^2y^2 + 2\lambda(1-\lambda)xy &< (1-\lambda)x^2 + \lambda y^2 \Leftrightarrow \\ [(1-\lambda)^2 - (1-\lambda)]x^2 + (\lambda^2 - \lambda)y^2 + 2\lambda(1-\lambda)xy &< 0 \quad \Leftrightarrow \\ -\lambda(1-\lambda)x^2 - \lambda(1-\lambda)y^2 + 2\lambda(1-\lambda)xy &< 0 . \end{aligned} \tag{2}$$

Dividing (2) by $-\lambda(1-\lambda)$ results in

$$\begin{aligned} x^2+y^2-2xy > 0 &\Leftrightarrow \\ (x-y)^2 > 0 &. \end{aligned}$$

Since $x \neq y$, the last relation is true.

(ii) It holds

$$\begin{aligned} g((1-\lambda)x+\lambda y) &= [f((1-\lambda)x+\lambda y)]^2 \le [(1-\lambda)f(x)+\lambda f(y)]^2 \\ &\le (1-\lambda)[f(x)]^2+\lambda[f(y)]^2 = (1-\lambda)g(x)+\lambda g(y) \\ &\text{for } x, y \in M, 0 \le \lambda \le 1 . \end{aligned}$$

(iii) It is sufficient to prove the statement for $a = b = 0$, i.e. for the function

$$f(x_1, x_2) = \sqrt{x_1^2+x_2^2} = \left| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right| . \tag{3}$$

In the general case, the graph of the function in (f) is only a dislocation of the graph of (3). The triangle inequality implies for $x^1, x^2 \in \mathbb{R}^2$, $0 \le \lambda \le 1$:

$$\begin{aligned} f((1-\lambda)x^1+\lambda x^2) &= |(1-\lambda)x^1+\lambda x^2| \le |(1-\lambda)x^1| + |\lambda x^2| \\ &= (1-\lambda)f(x^1)+\lambda f(x^2) . \end{aligned}$$

For example in the case $x^2 = 2x^1$, the two sides of the inequality are equal for all λ, i.e. the function in (f) is convex but not strictly convex.

Exercise 3.18

(a) The convexity of the functions f_i implies:

$$\begin{aligned} f((1-\lambda)x+\lambda y) &= \sum_{i=1}^m f_i((1-\lambda)x+\lambda y) \\ &\le \sum_{i=1}^m [(1-\lambda)f_i(x)+\lambda f_i(y)] \\ &= (1-\lambda)\sum_{i=1}^m f_i(x) + \lambda \sum_{i=1}^m f_i(y) \\ &= (1-\lambda)f(x)+\lambda f(y) . \end{aligned}$$

If at least one of the functions f_i is strictly convex, the inequality is strict.

(b) From the convexity of f it follows that:

$$\begin{aligned} \alpha f((1-\lambda)x+\lambda y) &\le \alpha[(1-\lambda)f(x)+\lambda f(y)] \\ &= (1-\lambda)\,\alpha f(x)+\lambda\,\alpha f(y) , \end{aligned}$$

i.e. the function αf is convex. In the case of strict convexity we can substitute "$<$" by "$\le$", i.e. αf is then strictly convex.

(c) Let $f_1, \dots f_m$ be concave functions over the convex set $M \in \mathbb{R}^n$. Then the sum $f(x) = \sum_{i=1}^m f_i(x)$ is concave over M. The function f is strictly concave, if at least one of the functions f_i is strictly concave.
Let f be (strictly) concave over the convex set $M \subset \mathbb{R}^n$. For $\alpha > 0$ the function αf is (strictly) concave over M.

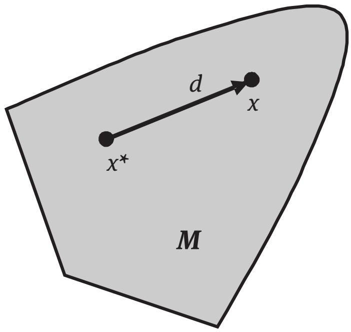


Fig. 3.19. Feasible direction.

Exercise 3.23

Since M is convex, we get for $0 < \lambda \le 1$:

$$x^* + \lambda d = x^* + \lambda(x - x^*) = (1 - \lambda)x^* + \lambda x \in M\,.$$

Therefore, d is a feasible direction at x^* (see Fig. 3.19).

Exercise 3.24

(i) For $x > 1$ it holds that $\ln x > 0$ and $\ln x$ is a concave function. Corollary 3.20 implies that $1/\ln x$ is convex. (This function is even strictly convex). Since $3e^x$ and $1/x$ are also strictly convex, the function f is strictly convex.
(ii) In spite of the discontinuity, f is strictly concave (see Figure 3.20).
(iii) The function is convex but not strictly convex (see Fig. 3.21).

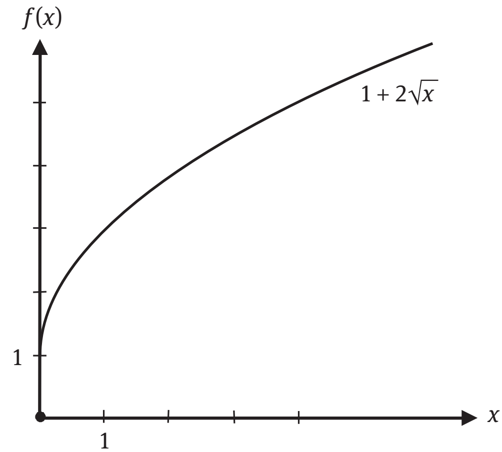


Fig. 3.20. Discontinuous concave function.

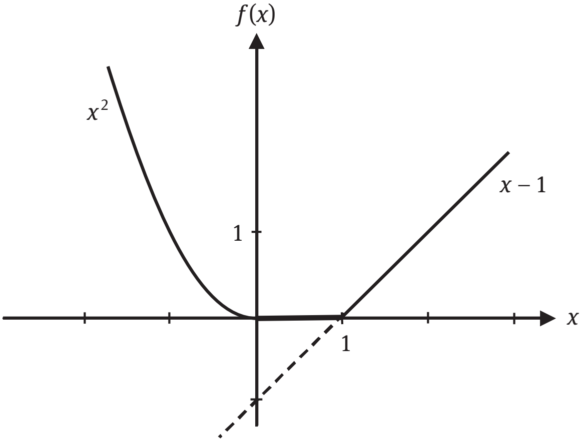


Fig. 3.21. Convex function.

Exercise 3.28

(a) It holds

$$\operatorname{grad} f(x_1, x_2) = \begin{pmatrix} 4x_1 - x_2 \\ -x_1 + 10x_2 \end{pmatrix},$$
$$Hf(x_1, x_2) = \begin{pmatrix} 4 & -1 \\ -1 & 10 \end{pmatrix}.$$

Since $4 > 0, \det \left(\begin{smallmatrix} 4 & -1 \\ -1 & 10 \end{smallmatrix}\right) = 39 > 0, Hf(x_1, x_2)$ is positive definite at all points (Theorem 2.29 (ii)). Thus, f is strictly convex.

(b) It holds

$$\operatorname{grad} f(x_1, x_2) = \begin{pmatrix} 2x_1x_2^2 \\ 2x_1^2x_2 \end{pmatrix},$$
$$Hf(x_1, x_2) = \begin{pmatrix} 2x_2^2 & 4x_1x_2 \\ 4x_1x_2 & 2x_1^2 \end{pmatrix}.$$

For example, $Hf(1, -1) = \left(\begin{smallmatrix} 2 & -4 \\ -4 & 2 \end{smallmatrix}\right)$ is not positive semidefinite (see Theorem 2.29 (iii)). Hence, f is not convex.

Exercise 3.29

(a) A continuously differentiable function f over the convex set $M \subset \mathbb{R}^n$ is *concave*, if and only if

$$f(x) \leq f(x^*) + (x - x^*)^T \operatorname{grad} f(x^*)$$

for all $x, x^* \in M$.

(b) (i) A twice continuously differentiable function f over the convex set $M \subset \mathbb{R}^n$ is *concave*, if and only if $Hf(x)$ is *negative semidefinite* for all $x, x^* \in M$.

(ii) If $Hf(x)$ is *negative definite* for all $x \in M, f$ is strictly *concave* over M.

Exercise 3.36

(a) Case $x^* = -1$: The relation (3.9) means

$$f(x) \geq 1 + a(x+1) \quad \text{for } x \in \mathbb{R},$$

which is equivalent to

$$\begin{aligned} -x &\geq 1 + a(x+1) \quad \text{for } x < 0, \\ x^2 &\geq 1 + a(x+1) \quad \text{for } x \geq 0. \end{aligned}$$

Setting, for example, $x = -2$ and $x = -0.5$, we obtain $a \geq -1$ and $a \leq -1$. Hence, $a = -1$ is the unique solution of this system.

(b) Case $x^* = 1$: Relation (3.9) means

$$f(x) \geq 1 + a(x-1) \quad \text{for } x \in \mathbb{R},$$

which is equivalent to

$$\begin{aligned} -x &\geq 1 + a(x-1) \quad \text{for } x < 0 \\ x^2 &\geq 1 + a(x-1) \quad \text{for } x \geq 0. \end{aligned} \tag{$*$}$$

The second inequality implies in particular

$$\begin{aligned} &x^2 - 1 \geq a(x-1) \quad \text{for } x > 1 \quad \Leftrightarrow \\ &x + 1 \geq a \quad \text{for } x > 1. \end{aligned}$$

Hence, $a \leq 2$. Similarly we get

$$\begin{aligned} &x^2 - 1 \geq a(x-1) \quad \text{for } 0 \leq x < 1 \quad \Leftrightarrow \\ &x + 1 \leq a \quad \text{for } 0 \leq x < 1. \end{aligned}$$

Therefore, $a \geq 2$. It follows that $a = 2$ is the unique solution of $(*)$.

(c) Case $x^* = 0$: Relation (3.9) means

$$f(x) \geq ax \qquad\qquad \text{for } x \in \mathbb{R},$$

which is equivalent to

$$\begin{aligned} &-x \geq ax \quad \text{for } x < 0 \\ &x^2 \geq ax \quad \text{for } x \geq 0 \quad \Leftrightarrow \\ &-1 \leq a \\ &x \geq a \quad \text{for } x > 0 \Leftrightarrow \\ &-1 \leq a \leq 0. \end{aligned}$$

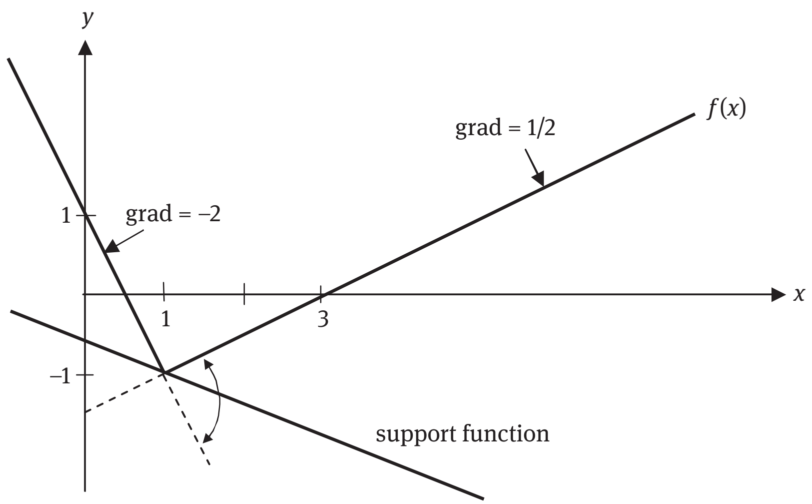


Fig. 3.22. Geometrical illustration of the subdifferential.

Exercise 3.38

The number $a \in \mathbb{R}$ is a subgradient of f at 1, if and only if a is the gradient of a support function of f at 1, i.e. if and only if $-2 \leq a \leq 1/2$ (see Figure 3.22).

Exercise 3.39

The relation (3.9) means

$$\begin{aligned}
&2|x_1| + 3|x_2| \geq a_1 x_1 + a_2 x_2 \quad \text{for } x_1, x_2 \in \mathbb{R} \quad \Leftrightarrow \\
&2|x_1| \geq a_1 x_1 \quad \text{and} \quad 3|x_2| \geq a_2 x_2 \quad \text{for } x_1, x_2 \in \mathbb{R} \quad \Leftrightarrow \\
&|x_1| \geq \left|\frac{a_1}{2} x_1\right| \quad \text{and} \quad |x_2| \geq \left|\frac{a_2}{3} x_2\right| \quad \text{for } x_1, x_2 \in \mathbb{R} \quad \Leftrightarrow \\
&|a_1| \geq 3 \quad \text{and} \quad |a_2| \leq 3\,.
\end{aligned}$$

Thus, $a = (a_1, a_2)^T$ is a subgradient of f at the point 0, if and only if $|a_1| \leq 2$ and $|a_2| \leq 3$.

Exercise 3.42

(i)
$$\begin{aligned}
f_{(y)}(t) &= f(x^* + ty) = f(2 + t, 3 + 2t) \\
&= 2(2 + t) - 3(2 + t)(3 + 2t) = -(2 + t)(7 + 6t)\,.
\end{aligned}$$

It holds

$$\begin{aligned}
f'_{(y)}(t) &= -(7 + 6t) - (2 + t)6\,, \\
Df(x^*, y) &= f'_{(y)}(0) = -7 - 12 = -19.
\end{aligned}$$

(ii) We get

$$f_{(y)}(t) = f(x^* + ty) = f(t, t) = \sqrt{t^2 + t^2} = \sqrt{2}|t|\,,$$
$$D^+f(x^*, y) = \lim_{t\to 0+} \frac{f_{(y)}(t) - f_{(y)}(0)}{t} = \lim_{t\to 0+} \frac{\sqrt{2}|t|}{t} = \sqrt{2}\,.$$

Similarly we obtain

$$D^-f(x^*, y) = \lim_{t\to 0-} \frac{\sqrt{2}|t|}{t} = -\sqrt{2} \neq D^+f(x^*, y)\,,$$

thus, the directional derivative $Df(x^*, y)$ does not exist.

(iii) We obtain

$$f_{(y)}(t) = f(x^* + ty) = f(ty_1, ty_2) = |ty_1| + |ty_2|\,,$$
$$D^+f(x^*, y) = \lim_{t\to 0+} \frac{|ty_1| + |ty_2|}{t} = \lim_{t\to 0+} \frac{|t||y_1| + |t||y_2|}{t} = |y_1| + |y_2|\,.$$

Exercise 3.45

We get

$$f_{(y)}(t) = f(x^* + ty) = |ty_1| + (ty_2)^2\,,$$
$$D^+f(x^*, y) = \lim_{t\to 0+} \frac{|t||y_1| + (ty_2)^2}{t} = \lim_{t\to 0+} (|y_1| + ty_2^2) = |y_1|\,.$$

The vector $a = (a_1, a_2)^T$ is a subgradient of f at x^*, if and only if

$$D^+f(x^*, y) = |y_1| \geq a_1y_1 + a_2y_2 \quad \text{for all } y_1, y_2 \in \mathbb{R}\,. \tag{$*$}$$

For $y_1 = 0$ we obtain

$$0 \geq a_2y_2 \quad \text{for all } y_2\,,$$

i.e. $a_2 = 0$. The relation $(*)$ reduces to

$$|y_1| \geq a_1y_1 \quad \text{for all } y_1 \in \mathbb{R}\,,$$

implying $-1 \leq a_1 \leq 1$. Hence, a vector $a = (a_1, a_2)^T$ is a subgradient of f at 0, if and only if $-1 \leq a_1 \leq 1$ and $a_2 = 0$.

Exercise 3.47

The unilateral derivatives are

$$D^+f(x^*, 1) = \lim_{t\to 0+} \frac{f(t) - f(0)}{t} = \lim_{t\to 0+} \frac{2t}{t} = 2\,,$$
$$D^-f(x^*, 1) = \lim_{t\to 0-} \frac{f(t) - f(0)}{t} = \lim_{t\to 0-} \frac{-t^3}{t} = 0\,.$$

The number a is a subgradient of f at 0, if and only if $0 \leq a \leq 2$, i.e. $\partial f(0) = [0, 2]$ (see Corollary 3.46).

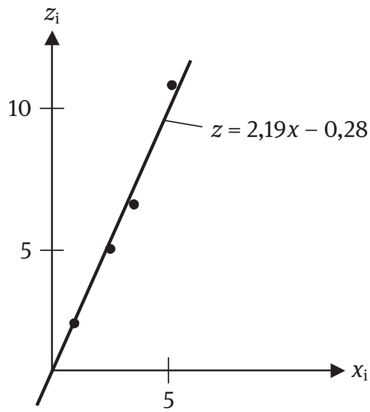


Fig. 3.23. Dispersion diagram.

Exercise 3.53

(a) Since the functions $f_i(\beta_1, \beta_2)$ of Example 3.52 are *not* strictly convex, the strict convexity of f does not follow directly from Theorem 3.16. But without loss of generality it can be assumed $x_1 \neq x_2$. In this case, the Hessian matrix of $f_1 + f_2$ is

$$Hf_1(\beta_1, \beta_2) + Hf_2(\beta_1, \beta_2) = \begin{pmatrix} 2(x_1^2 + x_2^2) & 2(x_1 + x_2) \\ 2(x_1 + x_2) & 4 \end{pmatrix},$$

having the determinant

$$8(x_1^2 + x_2^2) - 4(x_1 + x_2)^2 = 4(x_1 - x_2)^2 > 0\,.$$

Hence, $f_1 + f_2$ is strictly convex and Theorem 3.16 implies that $f = f_1 + f_2 + \sum_{i=3}^{m} f_i$ is strictly convex.

(b) We get

$$\sum_{i=1}^{4} x_i z_i = 94.25\,, \quad \overline{x} = 3\,, \quad \overline{z} = 6.3\,, \quad \sum_{i=1}^{4} x_i^2 = 44.5\,,$$

and (3.17), (3.18) imply

$$\beta_1 = 18.65/8.5 \approx 2.19 \quad \beta_2 \approx -0.28$$

(see Figure 3.23).

Exercise 3.56

The system (3.20) is of the form

$$\begin{aligned} x_1 - x_2 &= 1\,, \\ -x_1 + 5x_2 &= 1\,, \end{aligned}$$

and has the unique solution $x_1^* = 3/2$, $x_2^* = 1/2$.

Hence,

$$G(x_1^*) = \frac{3}{2}A - \frac{1}{2}B = \begin{pmatrix} 1/2 \\ 0 \\ 2 \end{pmatrix}, \quad H(x_2^*) = \frac{1}{2}C + \frac{1}{2}D = \begin{pmatrix} 1/2 \\ 1 \\ 2 \end{pmatrix},$$

and the distance between the straight lines is $|G(x_1^*) - H(x_2^*)| = 1$.

Exercise 3.57

It holds

$$\operatorname{grad} f(x_1, x_2) = \begin{pmatrix} 6x_1 + 2x_2 - 10 \\ 2x_1 + 10x_2 - 22 \end{pmatrix},$$
$$Hf(x_1, x_2) = \begin{pmatrix} 6 & 2 \\ 2 & 10 \end{pmatrix}.$$

Since $Hf(x_1, x_2)$ is positive definite for all x, f is strictly convex. The unique solution of the system $\operatorname{grad} f(x_1, x_2) = 0$ is $(x_1, x_2)^T = (1, 2)^T$. This is the unique global minimum point of f.

Exercise 3.60

(a) The function f is concave and $x^1, \dots, x^4$ are the extreme points of M (see Figure 3.24). It holds $f(x^1) = f(-1, 0) = 1 - 1 - 0 = 0$. In the same way we calculate $f(x^2) = -27, f(x^3) = -27$ and $f(x^4) = -42$. Therefore, x^4 is a global minimum point.

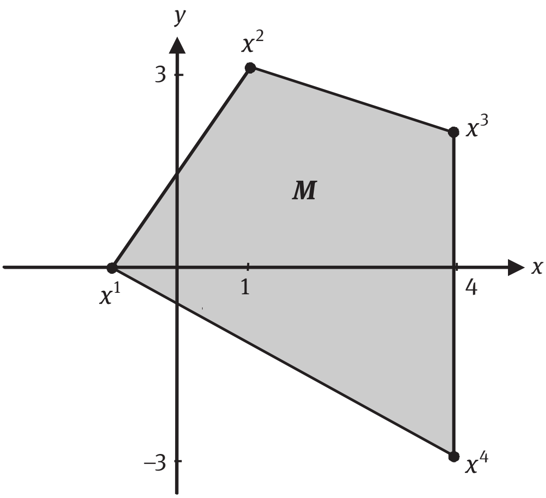


Fig. 3.24. Feasible region.

(b) A global minimum point $\binom{x_1}{x_2}$ can be found among the extreme points of M, i.e. satisfies

$$x_1^2 + x_2^2 = 1\,, \quad -1 \le x_1 \le 1\,. \tag{1}$$

By substituting x_2^2 in the objective function for $1 - x_1^2$, we obtain

$$1 - 2x_1^2 - 5(1 - x_1^2) = 3x_1^2 - 4\,. \tag{2}$$

Minimizing (2) results in $x_1 = 0$, and (1) implies $x_2^2 = 1$. Therefore, the problem of the exercise has the two global minimum points $(0, 1)^T$ and $(0, -1)^T$.

Exercise 3.61

The function $f(x) = 1/x$ is convex over the convex set $M = \{x \in \mathbb{R}^n | x > 0\}$, but no minimum point exists.

Exercise 3.62

The vector $a \in \mathbb{R}^n$ is a subgradient of the function f at x^*, if and only if

$$f(x) \ge f(x^*) + a^T(x - x^*) \qquad \text{for all } x \in M$$

(see Definition 3.34). Therefore, $0 \in \mathbb{R}^n$ is a subgradient of f at x^*, if and only if

$$f(x) \ge f(x^*) \quad \text{for all } x \in M\,.$$

Hence, the statements (i) and (ii) are equivalent. From Theorem 3.43 (iii) it follows that (ii) and (iii) are equivalent.

Exercise 3.63

Assume that x^1 and x^2 are two global minimum points, i.e.

$$f(x^1) = f(x^2) \le f(x) \quad \text{for all } x \in M\,. \tag{$*$}$$

The strict convexity of f implies

$$f((1 - \lambda)x^1 + \lambda x^2) < (1 - \lambda)f(x^1) + \lambda f(x^2) = f(x^1)$$

for all λ with $0 < \lambda < 1$, contradicting the statement $(*)$.

Exercise 3.64

It is advisable to review first the Exercise 3.7 (i). From Definition 3.14 it follows that

$$\begin{gathered} f\left(\sum_{i=1}^{m} \lambda_i x^i\right) \ge \sum_{i=1}^{m} \lambda_i f(x^i) \\ \text{for } \lambda_1, \dots, \lambda_m \ge 0\,, \quad \lambda_1 + \cdots + \lambda_m = 1\,, \quad x^1, \dots, x^m \in M \end{gathered} \tag{$*$}$$

is valid for $m = 2$. Assume now that $(*)$ is valid for $m = m_0$ with $m_0 \ge 2$. We prove that in this case $(*)$ is also valid for $m = m_0 + 1$:

For $\lambda_1, \dots, \lambda_{m_0} \geq 0, \lambda_{m_0+1} > 0,\ \lambda_1 + \cdots + \lambda_{m_0+1} = 1$ it holds that

$$\begin{aligned} f\Big(\sum_{i=1}^{m_0+1} \lambda_i x^i\Big) &= f\Big((1-\lambda_{m_0+1}) \sum_{i=1}^{m_0} \frac{\lambda_i}{1-\lambda_{m_0+1}} x^i + \lambda_{m_0+1} x^{m_0+1}\Big) \\ &\geq (1-\lambda_{m_0+1}) f\Big(\sum_{i=1}^{m_0} \frac{\lambda_i}{1-\lambda_{m_0+1}} x^i\Big) + \lambda_{m_0+1} f(x^{m_0+1}) \\ &\geq (1-\lambda_{m_0+1}) \sum_{i=1}^{m_0} \frac{\lambda_i}{1-\lambda_{m_0+1}} f(x^i) + \lambda_{m_0+1} f(x^{m_0+1}) = \sum_{i=1}^{m_0+1} \lambda_i f(x^i)\,. \end{aligned}$$

The inequalities follow from Definition 3.14 and the assumption, respectively.

Chapter 4

Exercise 4.4

For example, the point $x^* = 0$ of the set

$$M = \{(x_1, x_2)^T \in \mathbb{R}^2 | -x_1^2 + x_2 \leq 0\,, -x_1^2 - x_2 \leq 0\}$$

is regular, but the condition of Theorem 4.2 (b) is *not* satisfied: the vectors

$$g^1 := \operatorname{grad} g_1(x^*) = (0,1)^T \quad \text{and} \quad g^2 := \operatorname{grad} g^2(x^*) = (0,-1)^T$$

are linearly dependent (see Figure 4.15).

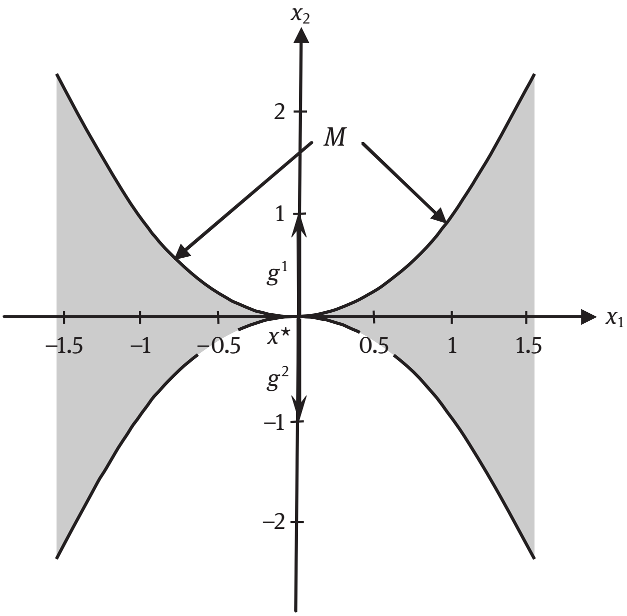


Fig. 4.15. Regular point $x^* = 0$ with linearly dependent gradients.

Sets of Exercise 2.9

(a) Since the constraints are linear, all points of M are regular.

(b) As in Example 4.3 it can be shown that $x^* = (3,1)^T$ is the only irregular point.

(c) The unique point $x^* = 0$ of M is irregular, since $Z(x^*) = \overline{Z(x^*)} = \emptyset \neq \overline{L(x^*)} = \mathbb{R}^2$. (Note that $g(x_1, x_2) = x_1^2 + x_2^2$ is convex, but the Slater condition is not satisfied.)

Exercise 4.9

The KKT conditions are

$$\begin{aligned} -1 - u_1 + u_2 &= 0 \\ -1 + u_1 + 2x_2 u_2 &= 0 \\ u_1(-x_1 + x_2 - 1) &= 0 \\ u_2(x_1 + x_2^2 - 1) &= 0 \end{aligned} \tag{1}$$

$$\begin{aligned} -x_1 + x_2 - 1 &\leq 0 \\ x_1 + x_2^2 - 1 &\leq 0 \\ u_1, u_2 &\geq 0\,. \end{aligned} \tag{2}$$

There is no solution for $u_2 = 0$, since the first equation of (1) would imply $u_1 = -1 < 0$, contradicting the condition $u_1 \geq 0$ in (2).

Case: $u_1 = 0, u_2 > 0$

The system (1), (2) reduces to:

$$\begin{aligned} -1 + u_2 &= 0 \\ -1 + 2x_2 u_2 &= 0 \\ x_1 + x_2^2 - 1 &= 0 \end{aligned} \tag{3}$$

$$\begin{aligned} -x_1 + x_2 - 1 &\leq 0 \\ u_2 &> 0\,. \end{aligned} \tag{4}$$

The unique solution of (3) is $(x_1, x_2, u_2) = (3/4, 1/2, 1)$, satisfying the inequalities (4).

Case: $u_1 > 0, u_2 > 0$

Now system (1), (2) reduces to:

$$\begin{aligned} -1 - u_1 + u_2 &= 0 \\ -1 + u_1 + 2x_2 u_2 &= 0 \\ -x_1 + x_2 - 1 &= 0 \\ x_1 + x_2^2 - 1 &= 0 \end{aligned} \tag{5}$$

$$u_1, u_2 > 0\,. \tag{6}$$

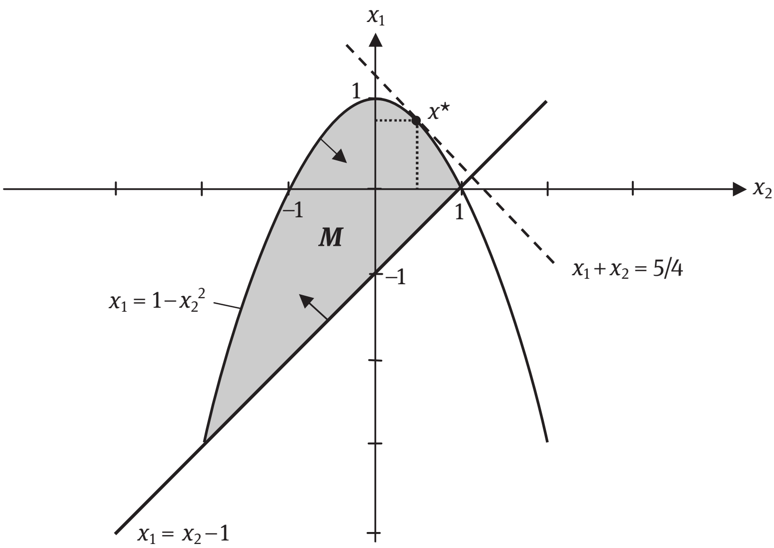


Fig. 4.16. Geometric resolution of Exercise 4.9.

The last two equations of (5) imply

$$x_1 = -3 \quad \text{and} \quad x_2 = -2 \quad \text{or} \quad x_1 = 0 \quad \text{and} \quad x_2 = 1\,.$$

In the first case, (5) implies $u_1 = -5/3, u_2 = -2/3$, in the second case, $u_1 = -1/3$, $u_2 = 2/3$. Both cases contradict (6).

Therefore, the unique solution of the KKT conditions (1), (2) is $(x_1, x_2, u_1, u_2) = (3/4, 1/2, 0, 1)$. Figure 4.16 shows that $x^* = (3/4, 1/2)^T$ is a global minimum point of the considered optimization problem.

Exercise 4.10

The KKT conditions are

$$\begin{aligned} 1 - 3x_1^2 u_1 - 3x_1^2 u_2 &= 0 \\ u_1 - u_2 &= 0 \\ u_1(x_2 - x_1^3) &= 0 \\ u_2(-x_2 - x_1^3) &= 0 \\ x_2 - x_1^3 &\le 0 \\ -x_2 - x_1^3 &\le 0 \\ u_1, u_2 &\ge 0. \end{aligned}$$

Because of the second equation, a solution must satisfy $u_1 = u_2$. For $u_1 = u_2 = 0$, the first equation is not satisfied. In the remaining case $u := u_1 = u_2 > 0$, the system reduces to

$$
\begin{aligned}
1 - 6x_1^2 u &= 0 && (1)\\
x_2 - x_1^3 &= 0 && (2)\\
-x_2 - x_1^3 &= 0 && (3)\\
u &> 0\,. && (4)
\end{aligned}
$$

The equations (2), (3) imply $x_1 = x_2 = 0$, which contradicts (1). Hence the KKT conditions have no solution.

The example does not contradict Theorem 4.6, since $x^* = 0 \in \mathbb{R}^2$ is not a regular point (see Example 4.3). An irregular point can be a minimum point without satisfying the KKT conditions.

Exercise 4.13

Without loss of generality let $A(x^*) = \{1, \dots, p\}$,

$$
\begin{aligned}
g^i &:= \operatorname{grad} g_i(x^*) \in \mathbb{R}^n \quad \text{for } i = 1, \dots, p\,,\\
h^j &:= \operatorname{grad} h_j(x^*) \in \mathbb{R}^n \quad \text{for } j = 1, \dots, k\,.
\end{aligned}
$$

We have to prove the following: if $g^1, \dots, g^p, h^1, \dots, h^k$ are linearly independent, then there exists a vector $z \in \mathbb{R}^n$, such that

$$
\begin{aligned}
z^T g^i &< 0 \quad \text{for } i = 1, \dots, p\,, && (*)\\
z^T h^j &= 0 \quad \text{for } j = 1, \dots, k\,.
\end{aligned}
$$

Let $g^1, \dots, g^p, h^1, \dots, h^k$ be linearly independent vectors of $\mathbb{R}^n$, then $p + k \le n$.

(a) Let $p + k = n$. Consider the nonsingular matrix $A \in \mathbb{R}^{n \times n}$, the columns of which are the vectors $g^1, \dots, g^p, h^1, \dots, h^k$. We choose $z \in \mathbb{R}^n$ such that $z^T A = (-1, \dots, 1, 0, \dots, 0)$ (where the components -1 and 0 appear p and k times, respectively).
Since A is nonsingular, z is uniquely determined such that

$$
\begin{aligned}
z^T g^i &= -1 && \text{for } i = 1, \dots, p\,,\\
z^T h^j &= 0 && \text{for } j = 1, \dots, k\,,
\end{aligned}
$$

i.e. z satisfies $(*)$.

(b) Let $p + k < n$, $r := n - p - k$. The Steinitz theorem ensures the existence of vectors $v^1, \dots, v^r \in \mathbb{R}^n$ such that

$$
g^1, \dots, g^p\,, \quad h^1, \dots, h^k\,, \quad v^1, \dots, v^r \qquad (**)
$$

are linearly independent. We consider the matrix A, the columns of which are the vectors $(**)$ and choose z such that

$$
z^T A = (-1, \dots, -1, 0, \dots, 0, c_1, \dots, c_r)
$$

(see above; $c_1, \dots, c_r$ are any real numbers). Since A is nonsingular, z is uniquely determined such that

$$\begin{aligned} z^T g^i &= -1 \quad \text{for } i = 1, \dots, p\,, \\ z^T h^j &= 0 \quad \text{for } j = 1, \dots, k\,, \\ z^T v^t &= c_t \quad \text{for } t = 1, \dots, r\,, \end{aligned}$$

i.e. z satisfies $(*)$.

Exercise 4.15
The KKT conditions are

$$\begin{aligned} -2 + u + 2x_1 v &= 0 \\ -1 + 2u - 2v &= 0 \\ u(x_1 + 2x_2 - 6) &= 0 \\ x_1 + 2x_2 - 6 &\le 0 \\ x_1^2 - 2x_2 &= 0 \\ u &\ge 0\,. \end{aligned} \tag{$*$}$$

Case: $u = 0$
System $(*)$ reduces to

$$\begin{aligned} -2 + 2x_1 v &= 0 \\ -1 - 2v &= 0 \\ x_1 + 2x_2 - 6 &\le 0 \\ x_1^2 - 2x_2 &= 0. \end{aligned}$$

The unique solution of the subsystem of equations is $(x_1, x_2, v) = (-2, 2, -1/2)$, satisfying the inequality.

Case: $u > 0$
Now $(*)$ reduces to

$$\begin{aligned} -2 + u + 2x_1 v &= 0 \\ -1 + 2u - 2v &= 0 \\ x_1 + 2x_2 - 6 &= 0 \\ x_1^2 - 2x_2 &= 0 \\ u &> 0\,. \end{aligned}$$

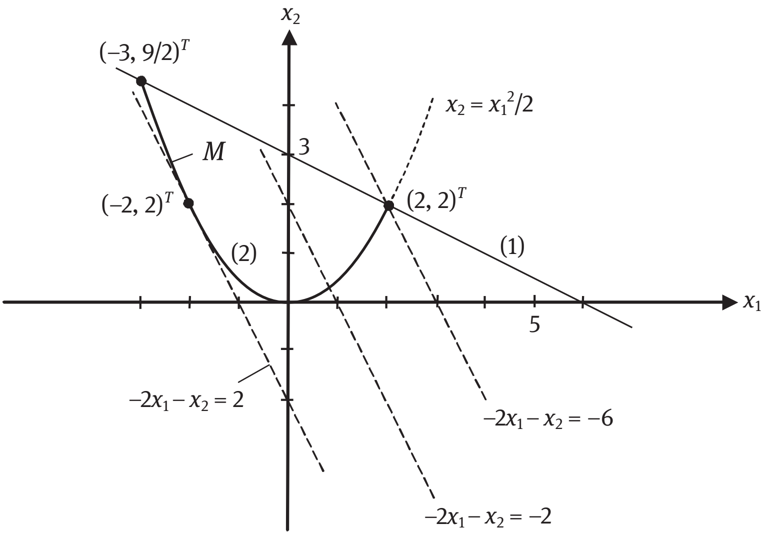


Fig. 4.17. Geometric solution of Exercise 4.15.

The subsystem of equations has the two solutions

$$(x_1, x_2, u, v) = (-3, 9/2, 1/5, -3/10)\,,$$
$$(x_1, x_2, u, v) = (2, 2, 4/5, 3/10)\,.$$

Both of them satisfy the inequality $u > 0$. Therefore, the conditions KKT $(*)$ have the three solutions:

$$(x_1, x_2, u, v) = (-2, 2, 0, -1/2)\,,$$
$$(x_1, x_2, u, v) = (-3, 9/2, 1/5, -3/10)\,,$$
$$(x_1, x_2, u, v) = (2, 2, 4/5, 3/10)\,.$$

The points $(2, 2)^T$ and $(-3, 9/2)^T$ represent local minima (see Figure 4.17), the former is also a global minimum point. (The point $(-2, 2)^T$ is a global maximum point.)

Exercise 4.16

We get

$$f(x) = \frac{1}{2}x^TQx + c^Tx\,,$$
$$\operatorname{grad} f(x) = Qx + c\,.$$

Setting

$$A = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix},$$

the constraints can be written as

$$h_i(x) = a_{1i}x_1 + \cdots + a_{ni}x_n - b_i = 0 \quad \text{for } i = 1, \dots, m\,.$$

Thus,

$$\operatorname{grad} h_i(x) = (a_{1i}, \dots, a_{ni})^T \quad \text{for } i = 1, \dots, m\,.$$

We obtain the KKT conditions (see Theorem 4.12):

$$Qx^* + c + v_1^* \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix} + \cdots + v_m^* \begin{pmatrix} a_{1m} \\ \vdots \\ a_{nm} \end{pmatrix} = 0$$

$$a_{1i}x_1^* + \cdots + a_{ni}x_n^* = b_i \quad \text{for } i = 1, \dots, m\,,$$

which can be written as

$$\begin{aligned} Qx^* + c + Av^* &= 0 \\ A^T x^* &= b\,. \end{aligned}$$

Exercise 4.17

(a) We obtain

$$Q = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}, \quad c = \begin{pmatrix} -3 \\ 5 \end{pmatrix}, \quad A = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad b = 5\,.$$

Thus

$$Q^{-1} = \frac{1}{5}\begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}, \quad A^T Q^{-1} A = 2\,, \quad A^T Q^{-1} c = -3\,,$$

and the global minimum point is

$$x^* = \frac{1}{5}\begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}\left[\begin{pmatrix} 2 \\ 1 \end{pmatrix}\frac{1}{2}(5-3) + \begin{pmatrix} 3 \\ -5 \end{pmatrix}\right] = \frac{1}{5}\begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 5 \\ -4 \end{pmatrix} = \begin{pmatrix} 19/5 \\ -13/5 \end{pmatrix}.$$

(b) The constraint results in $x_2 = 5 - 2x_1$. By substituting x_2 in the objective function we get

$$\begin{aligned} f(x_1) &= x_1^2 + x_1(5 - 2x_1) + \frac{3}{2}(5 - 2x_1)^2 - 3x_1 + 25 - 10x_1 \\ &= 5x_1^2 - 38x_1 + \frac{125}{2}\,, \\ f'(x_1) &= 10x_1 - 38x_1 = 0 \Rightarrow x_1 = 19/5 \Rightarrow x_2 = 5 - 2x_1 = -13/5\,. \end{aligned}$$

Exercise 4.18

(a) We have to minimize the function

$$f(x_1, x_2) = x_1x_2 + \frac{\pi}{4}x_1^2 + \frac{2x_2 + \pi x_1}{x_1x_2 + (\pi/4)x_1^2}V$$

subject to the constraints $g_1(x_1, x_2) = x_1 - x_2 \le 0$ and $g_2(x_1, x_2) = -x_1 \le 0$. The function f can be written as

$$\begin{aligned} f(x_1, x_2) &= x_1 x_2 + \frac{\pi}{4}x_1^2 + \frac{x_2 + \frac{\pi}{4}x_1 + \frac{\pi}{4}x_1}{x_2 + \frac{\pi}{4}x_1}\frac{2V}{x_1} \\ &= x_1 x_2 + \frac{\pi}{4}x_1^2 + (1 + \frac{\frac{\pi}{4}x_1}{x_2 + \frac{\pi}{4}x_1})\frac{2V}{x_1} \\ &= x_1 x_2 + \frac{\pi}{4}x_1^2 + \frac{2V}{x_1} + \frac{\pi V}{2x_2 + \frac{\pi}{2}x_1} . \end{aligned}$$

The partial derivatives are:

$$\begin{aligned} \frac{\partial}{\partial x_1} f(x_1, x_2) &= x_2 + \frac{\pi}{2}x_1 - \frac{2V}{x_1^2} - \frac{\pi V}{(2x_2 + \frac{\pi}{2}x_1)^2}\frac{\pi}{2} , \\ \frac{\partial}{\partial x_2} f(x_1, x_2) &= x_1 + \frac{\pi V}{(2x_2 + \frac{\pi}{2}x_1)^2} 2 . \end{aligned}$$

Since $\operatorname{grad} g_1(x_1, x_2) = (1, -1)^T$ and $\operatorname{grad} g_2(x_1, x_2) = (-1, 0)^T$, we obtain the KKT conditions (see Theorem 4.6):

$$\begin{aligned} x_2 + \frac{\pi}{2}x_1 - \frac{2V}{x_1^2} - \frac{\pi V}{(2x_2 + \frac{\pi}{2}x_1)^2}\frac{\pi}{2} + u_1 - u_2 &= 0 \\ x_1 - \frac{\pi V}{(2x_2 + \frac{\pi}{2}x_1)^2} 2 - u_1 &= 0 \\ u_1(x_1 - x_2) &= 0 \\ u_2 x_1 &= 0 \\ x_1 - x_2 &\le 0 \\ x_1 &\ge 0 \\ u_1, u_2 &\ge 0 . \end{aligned}$$

For $u_2 > 0$ the fourth constraint implies $x_1 = 0$, and then the objective function is not defined. Thus we get $u_2 = 0$ and the system reduces to

$$x_2 + \frac{\pi}{2}x_1 - \frac{2V}{x_1^2} - \frac{\pi V}{(2x_2 + \frac{\pi}{2}x_1)^2}\frac{\pi}{2} + u_1 = 0 \tag{1}$$

$$x_1 - \frac{\pi V}{(2x_2 + \frac{\pi}{2}x_1)^2} 2 - u_1 = 0 \tag{2}$$

$$u_1(x_1 - x_2) = 0 \tag{3}$$

$$x_1 - x_2 \le 0 \tag{4}$$

$$x_1 \ge 0 \tag{5}$$

$$u_1 \ge 0 . \tag{6}$$

We show at first that there is no solution for $u_1 = 0$. In this case (2) implies

$$\frac{\pi}{4}x_1 = \frac{\pi V}{(2x_2 + \frac{\pi}{2}x_1)^2}\frac{\pi}{2} . \tag{7}$$

By substituting the last summand in (1) for the left side of (7), we obtain

$$\begin{aligned} x_2 + \frac{\pi}{2}x_1 - \frac{2V}{x_1^2} - \frac{\pi}{4}x_1 = 0 \quad &\Leftrightarrow \\ x_2 + \frac{\pi}{4}x_1 = \frac{2V}{x_1^2} \quad &\Leftrightarrow \\ 2x_2 + \frac{\pi}{2}x_1 = \frac{4V}{x_1^2}\,. & \end{aligned} \tag{8}$$

Replacing the denominator of (2) by the right side of (8) implies

$$\begin{aligned} x_1 &= 2\pi V \frac{x_1^4}{16V^2} \quad \Leftrightarrow \\ x_1 &= \left(\frac{8V}{\pi}\right)^{1/3}. \end{aligned} \tag{9}$$

With (8) and (9) we now obtain:

$$2x_2 = \frac{4V}{x_1^2} - \frac{\pi}{2}x_1 = 4V\left(\frac{\pi}{8V}\right)^{2/3} - \frac{\pi}{2}\left(\frac{8V}{\pi}\right)^{1/3}.$$

Multiplying this equation by $(\frac{\pi}{8V})^{1/3}$ yields

$$2x_2\left(\frac{\pi}{8V}\right)^{1/3} = \frac{\pi}{2} - \frac{\pi}{2} = 0\,,$$

thus

$$x_2 = 0. \tag{10}$$

Since (9) and (10) contradict the constraint (4), the KKT conditions have no solution for $u_1 = 0$.
Assume now that $u_1 > 0$ in the conditions (1)–(6). Then (3) implies that $x_1 = x_2$. By summing the equations (1) and (2), we get

$$\begin{aligned} x_1 + \frac{\pi}{2}x_1 - \frac{2V}{x_1^2} - \frac{2V\pi^2}{(4+\pi)^2x_1^2} + x_1 - \frac{8V\pi}{(4+\pi)^2x_1^2} = 0 &\Leftrightarrow \\ \left(2+\frac{\pi}{2}\right)x_1 = \frac{2V}{x_1^2} + \frac{2V\pi^2}{(4+\pi)^2x_1^2} + \frac{8V\pi}{(4+\pi)^2x_1^2} &\Leftrightarrow \\ \frac{1}{2}(4+\pi)x_1 = \frac{2V}{x_1^2} + \frac{2V\pi}{(4+\pi)^2x_1^2}(4+\pi) &\Leftrightarrow \\ (4+\pi)x_1 = \frac{4V}{x_1^2} + \frac{4V\pi}{(4+\pi)x_1^2} &\Leftrightarrow \\ (4+\pi)^2x_1^3 = 4V(4+\pi) + 4V\pi &\Leftrightarrow \\ x_1^3 = \frac{16V + 8V\pi}{(4+\pi)^2} &\Leftrightarrow \\ x_1 = \left[\frac{16+8\pi}{(4+\pi)^2}\right]^{1/3} V^{1/3} \approx 0.931V^{1/3}\,. & \end{aligned}$$

With the aid of (2) we can calculate the Lagrange multiplier u_1:

$$u_1 \approx 0.362 V^{1/3} > 0\,.$$

Summarizing the above calculations, we observe that $x^* = (x_1^*, x_2^*)^T$ with

$$x_1^* = x_2^* = \left[\frac{16+8\pi}{(4+\pi)^2}\right]^{1/3} V^{1/3} \approx 0.931 V^{1/3}$$

is the unique "candidate" for a minimum point. This point is in fact the solution of the last problem of Example 1.17. (We will not analyze sufficient conditions).

(b) The optimal measures of the box are x_1^*, x_2^* (see above) and

$$x_3^* = \frac{V}{x_1^* x_2^* + (\pi/4)(x_1^*)^2} = \frac{V}{[1+(\pi/4)](x_1^*)^2} \approx 0.646 V^{1/3}\,.$$

The required amount of metal foil is

$$x_1^* x_2^* + \frac{\pi}{4}(x_1^*)^2 + (2x_2^* + \pi x_1^*)x_3^* \approx 4.641 V^{2/3}\,.$$

(c) Setting $x := x_1 = x_2$, the problem can be written as

$$\min f(x) := \left(1+\frac{\pi}{4}\right)x^2 + \frac{2+\pi}{(1+\frac{\pi}{4})x} V$$
$$x \geq 0\,.$$

The function f is convex and the derivative is

$$f'(x) = 2\left(1+\frac{\pi}{4}\right)x - \frac{(2+\pi)V}{(1+\frac{\pi}{4})x^2}\,.$$

The minimum point of f is given by

$$f'(x) = 0 \Leftrightarrow 2\left(1+\frac{\pi}{4}\right)x = \frac{(2+\pi)V}{(1+\frac{\pi}{4})x^2} \Leftrightarrow x^3 = \frac{1+\frac{\pi}{2}}{(1+\frac{\pi}{4})^2} V \Leftrightarrow$$
$$x^3 = \frac{16+8\pi}{(4+\pi)^2} V \Leftrightarrow x = \left(\frac{16+8\pi}{(4+\pi)^2}\right)^{1/3} V^{1/3}\,.$$

These are the same measures as in the original problem. This is to be expected, since the solution of part (a) must also satisfy $x_1 = x_2$.

Exercise 4.32

For the right side of (4.47) we obtain by using $(AB)^T = B^T A^T$:

$$\begin{aligned}
&\frac{1}{2}(c+Au)^T Q^{-1}(c+Au) - c^T Q^{-1} c - c^T Q^{-1} Au - u^T A^T Q^{-1} c - u^T A^T Q^{-1} Au - u^T b \\
&\quad = \frac{1}{2} c^T Q^{-1} c + \frac{1}{2} c^T Q^{-1} Au + \frac{1}{2} u^T A^T Q^{-1} c + \frac{1}{2} u^T A^T Q^{-1} Au - c^T Q^{-1} c - c^T Q^{-1} Au \\
&\quad\quad - u^T A^T Q^{-1} c - u^T A^T Q^{-1} Au - u^T b\,.
\end{aligned}$$

By joining the absolute, linear and quadratic terms of u, we get

$$-\frac{1}{2} c^T Q^{-1} c - u^T A^T Q^{-1} c - u^T b - \frac{1}{2} u^T A^T Q^{-1} Au = -\frac{1}{2} c^T Q^{-1} c - u^T h - \frac{1}{2} u^T P u\,.$$

Exercise 4.33

(a) Eliminating an additive constant, the primal problem can be written as

$$\min f(x) := \frac{1}{2}(x_1, x_2)\begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix} + (-4, 0)\begin{pmatrix}x_1\\x_2\end{pmatrix}$$

$$(2,1)\begin{pmatrix}x_1\\x_2\end{pmatrix} \le 2\,.$$

We get

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

$$P = A^T Q^{-1} A = (2,1)\begin{pmatrix}1/2 & 0\\0 & 1/2\end{pmatrix}\begin{pmatrix}2\\1\end{pmatrix} = 5/2\,,$$

$$h = AQ^{-1}c + b = (2,1)\begin{pmatrix}1/2 & 0\\0 & 1/2\end{pmatrix}\begin{pmatrix}-4\\0\end{pmatrix} + 2 = -2\,.$$

The dual problem

$$\max_{u\ge 0} -\frac{5}{4}u^2 + 2u$$

has the optimal solution $u^* = 4/5$. Using (4.46), we obtain the optimal solution of (P):

$$x^* = -Q^{-1}(c + Au^*) = \begin{pmatrix}-1/2 & 0\\0 & -1/2\end{pmatrix}\left(\begin{pmatrix}-4\\0\end{pmatrix} + \begin{pmatrix}2\\1\end{pmatrix}\cdot\frac{4}{5}\right) = \begin{pmatrix}6/5\\-2/5\end{pmatrix}.$$

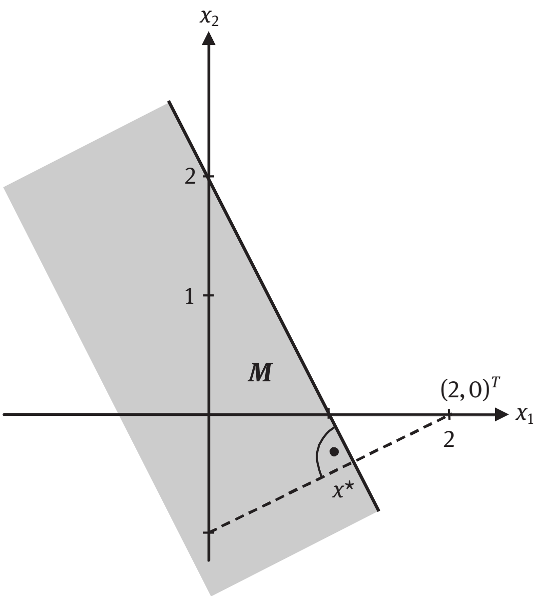


Fig. 4.18. Geometric solution of Exercise 4.33.

(b) The objective value $f(x_1, x_2) = (x_1 - 2)^2 + x_2^2$ is the squared distance between the points $(x_1, x_2)^T$ and $(2, 0)^T$ (see Figure 4.18). The point $x^* = (6/5, -2/5)^T$ minimizes the function f over the set $M = \{(x_1, x_2)^T | 2x_1 + x_2 - 2 \le 0\}$.

Exercise 4.34

(a) It holds

$$\begin{aligned} f(x_1, x_2) &= -3x_1 - x_2\,, \\ g(x_1, x_2) &= x_1^2 + 2x_2^2 - 2\,, \\ \Phi(x_1, x_2, u) &= -3x_1 - x_2 + u(x_1^2 + 2x_2^2 - 2)\,. \end{aligned}$$

For a given $u \ge 0$, the function Φ is strictly convex and the global minimum point x^* satisfies

$$\operatorname{grad}_x \Phi(x_1^*, x_2^*, u) = \begin{pmatrix} -3 + 2ux_1^* \\ -1 + 4ux_2^* \end{pmatrix} = 0\,,$$

i.e.

$$x^* = (x_1^*, x_2^*)^T = \left(\frac{3}{2u}, \frac{1}{4u}\right)^T. \tag{$*$}$$

We get

$$\inf_{x \in \mathbb{R}^2} \Phi(x, u) = -3x_1^* - x_2^* + u({x_1^*}^2 + 2{x_2^*}^2 - 2) = -\frac{19}{8u} - 2u\,,$$

and the dual problem is (see Definition 4.25):

$$\max_{u \ge 0} d(u) = -\frac{19}{8u} - 2u\,.$$

(b) It holds

$$d'(u) = \frac{19}{8u^2} - 2\,.$$

Setting $d'(u)$ equal to zero yields the optimal solution $u^* = \sqrt{19}/4 = 1.09$.

(c) The optimal solution x^* of (P) is obtained by solving

$$\min_{x \in \mathbb{R}^2} \Phi(x, u^*)\,,$$

(see the remark after Theorem 4.30), i.e.

$$x^* = \left(\frac{3}{2u^*}, \frac{1}{4u^*}\right)^T = \left(\frac{6}{\sqrt{19}}, \frac{1}{\sqrt{19}}\right)^T.$$

(d) We have to determine the point $x^* = (x_1^*, x_2^*)^T$ of the ellipse $x_1^2 + 2x_2^2 = 2$ with $x_1^* > 0$, in which the gradient of the tangent line is -3 (see Figure 4.19). For positive values of x_2, this ellipse is given by

$$x_2 = g(x_1) = \sqrt{1 - x_1^2/2}\,.$$

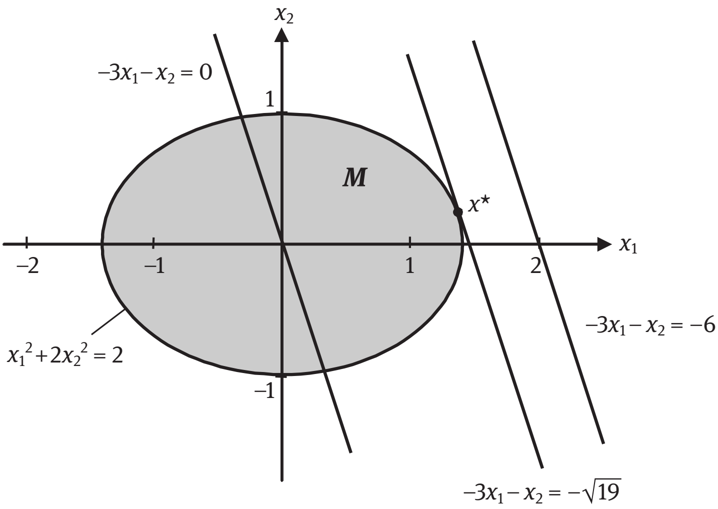


Fig. 4.19. Geometric solution of Exercise 4.34.

The condition

$$g'(x_1) = \frac{-x_1}{2\sqrt{1 - x^2/2}} = -3$$

implies $x_1^* = \frac{6}{\sqrt{19}}$, therefore $x_2^* = g(x_1^*) = \frac{1}{\sqrt{19}}$ which corresponds to the previous result. The optimal value is $-3x_1^* - x_2^* = -\sqrt{19}$.

Exercise 4.36

Substituting x_2 by $2 - x_1$ yields the one-dimensional problem

$$\max_{0 \le x_1 \le 2} 3(1 - e^{-2x_1}) + \frac{10 - 5x_1}{5 - x_1}\,.$$

The solution is given by

$$f'(x_1) = 6e^{-2x_1} - \frac{15}{(5 - x_1)^2} = 0\,.$$

It can be easily verified that the solution $x_1 \approx 0.9424$ of Example 4.35 satisfies this equation.

Exercise 4.41

(P) is the specific case of the primal problem of Example 4.40 with

$$Q = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad c = \begin{pmatrix} 0 \\ -6 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

The Wolfe dual is therefore

$$\begin{aligned} \max -x_1^2 - x_2^2 - 2u_1 - u_2 \\ 2x_1 + u_1 \geq 0 \\ 2x_2 - 6 + u_2 \geq 0 \\ u_1, u_2 \geq 0\,, \end{aligned}$$

equivalent to

$$\begin{aligned} \max -x_1^2 - x_2^2 - 2u_1 - u_2 \\ x_1 \geq 0 \\ x_2 \geq 3 \\ u_1, u_2 \geq 0\,. \end{aligned} \tag{DW}$$

(P) has the optimal solution $(x_1^*, x_2^*) = (0, 1)$ and the optimal value $z_{ot} = -5$, while (DW) has the optimal solution $(x_1^*, x_2^*, u_1^*, u_2^*) = (0, 3, 0, 0)$ and the optimal value $w_{ot} = -9$. Therefore, it holds $\Phi(\overline{x}, \overline{u}) \leq w_{ot} = -9 < z_{ot} = -5 \leq f(\hat{x})$ for any feasible points $\hat{x}$ and $(\overline{x}, \overline{u})$ of (P) and (DW), respectively. Thus, the weak duality theorem is satisfied and there exists a duality gap.

Exercise 4.42

The previous exercise shows a case in which the respective optimum values z_{ot} and w_{ot} are different. Therefore, one cannot formulate a statement analogous to Theorem 4.30.

Exercise 4.43

It holds $\Phi(x, u) = \frac{1}{2}x^T Qx + c^T x + u^T(A^T x - b)$, thus the Wolfe dual is

$$\begin{aligned} \max \frac{1}{2}x^T Qx + c^T x + u^T(A^T x - b) \\ Qx + c + Au = 0 \\ u \geq 0\,. \end{aligned}$$

By substituting c in the objective function (compare with Example 4.40) we obtain

$$\begin{aligned} \max -\frac{1}{2}x^T Qx - u^T b \\ Qx + c + Au = 0 \\ u \geq 0 \end{aligned}$$

or

$$\begin{aligned} \min \frac{1}{2}x^TQx + u^Tb \\ Qx + c + Au = 0 \\ u \geq 0\,. \end{aligned}$$

For $Q = 0$ results the standard form of a linear problem.

Exercise 4.48

The line $f(x) = z_{ot} = -9/4$ is the tangent line to the curve $g_2(x) = 0$ at the point $x^* = (1/2, 7/4)$. Therefore, this point is optimal. From the first two lines of (4.62) we obtain $u_1 = 1$ and $u_2 = 0$. Only the first constraint is active, thus

$$C(x^*) = \left\{ d \in \mathbb{R}^2 \,\middle|\, (d_1, d_2)\begin{pmatrix}1\\1\end{pmatrix} = 0 \right\} = \left\{ \alpha \begin{pmatrix}-1\\1\end{pmatrix} \,\middle|\, \alpha \in \mathbb{R} \right\}$$

and

$$d^T H_x \Phi(x^*, u^*) d = 2u_1^* d_1^2 = 2d_1^2 > 0 \text{ is satisfied for all } d \in C(x^*) \setminus \{0\}\,.$$

Exercise 4.49

The first-order necessary conditions are

$$c_1 + 2u\frac{x_1}{a_1^2} = 0 \tag{1}$$

$$c_2 + 2u\frac{x_2}{a_2^2} = 0 \tag{2}$$

$$c_3 + 2u\frac{x_3}{a_3^2} = 0 \tag{3}$$

$$u\left(\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \frac{x_3^2}{a_3^2} - 1\right) = 0 \tag{4}$$

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \frac{x_3^2}{a_3^2} - 1 \leq 0 \tag{5}$$

$$u \geq 0\,. \tag{6}$$

From the first three conditions it is clear that there is no solution for $u = 0$. For $u > 0$ it follows that

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \frac{x_3^2}{a_3^2} - 1 = 0 \tag{7}$$

$$x_i = -\frac{c_i a_i^2}{2u} \quad \text{for } i = 1, \dots, 3\,. \tag{8}$$

Substituting the x_i in (7) for the corresponding expressions in (8), we obtain

$$\frac{c_1^2a_1^2+c_2^2a_2^2+c_3^2a_3^2}{4u^2}=1 \Rightarrow$$
$$2u=\sqrt{c_1^2a_1^2+c_2^2a_2^2+c_3^2a_3^2}, \tag{9}$$

and combining (8) and (9) yields the solution candidate $x^*=(x_1^*,x_2^*,x_3^*)^T$ with

$$x_i^*=-\frac{c_1a_i^2}{\sqrt{c_1^2a_1^2+c_2^2a_2^2+c_3^2a_3^2}} \quad \text{for } i=1,\dots,3\,.$$

Geometrically we are minimizing a linear function over an ellipsoid M, therefore, x^* is a point of the surface of M. The cone of tangent directions is

$$C(x^*)=\left\{d\in\mathbb{R}^3\,\middle|\,\sum_{i=1}^{3}\frac{x_i^*d_i}{a_i^2}=0\right\},$$

which is the plane perpendicular to the vector $(\frac{x_1^*}{a_1^2},\frac{x_2^*}{a_2^2},\frac{x_3^*}{a_3^2})^T$. By shifting this plane to the point x^*, we obtain a supporting hyperplane of the ellipsoid M at x^*. Since the Hessian matrix

$$H_x\Phi(x^*,u^*)=\begin{pmatrix}\frac{2u}{a_1^2}&0&0\\0&\frac{2u}{a_2^2}&0\\0&0&\frac{2u}{a_3^2}\end{pmatrix}$$

is positive definite (i.e. the problem is convex), the second-order sufficient conditions are trivially satisfied.

Exercise 4.52

For the points $x^1=(2,2)^T$ and $x^2=(-3,9/2)^T$ we obtain

$$C(x^1)=\left\{d\in\mathbb{R}^2\,\middle|\,(d_1,d_2)\begin{pmatrix}1\\2\end{pmatrix}=0\,,(d_1,d_2)\begin{pmatrix}4\\-2\end{pmatrix}=0\right\}=\{0\}\,,$$
$$C(x^2)=\left\{d\in\mathbb{R}^2\,\middle|\,(d_1,d_2)\begin{pmatrix}1\\2\end{pmatrix}=0\,,(d_1,d_2)\begin{pmatrix}-6\\-2\end{pmatrix}=0\right\}=\{0\}\,,$$

i.e. the sufficient conditions are trivially satisfied. For $x^3=(-2,2)^T$ we get

$$C(x^3)=\left\{d\in\mathbb{R}^2\,\middle|\,(d_1,d_2)\begin{pmatrix}-4\\-2\end{pmatrix}=0\right\}=\left\{\alpha\begin{pmatrix}1\\-2\end{pmatrix}\,\middle|\,\alpha\in\mathbb{R}\right\}.$$

Since

$$H_x\Phi(x^*,u^*,v^*)=\begin{pmatrix}2v^*&0\\0&0\end{pmatrix}=\begin{pmatrix}-1&0\\0&0\end{pmatrix},$$

the second-order necessary condition is not satisfied.

Exercise 4.53
It holds

$$\begin{aligned}\operatorname{grad} f(x) &= (2(x_1-2), 2x_2, 2x_3)^T\,,\\ \operatorname{grad} g_1(x) &= (2x_1, 2x_2, 2x_3)^T\,,\\ \operatorname{grad} g_2(x) &= (0,0,-1)^T\,,\\ \operatorname{grad} h(x) &= (1,1,1)^T\,,\end{aligned}$$

and

$$\operatorname{grad}_x \Phi(x^*, u^*, v^*) = 0$$

implies

$$u_1 = 1\,, \quad u_2 = 0\,, \quad v = 0\,.$$

Hence,

$$\begin{aligned} C(x^*) &= \left\{ d \in \mathbb{R}^3 \,\middle|\, d^T \begin{pmatrix}2\\0\\0\end{pmatrix} = 0,\, d^T \begin{pmatrix}0\\0\\-1\end{pmatrix} \le 0,\, d^T \begin{pmatrix}1\\1\\1\end{pmatrix} = 0 \right\} = \left\{ \alpha \begin{pmatrix}0\\-1\\1\end{pmatrix} \,\middle|\, \alpha \ge 0 \right\}\,,\\ C^-(x^*) &= \left\{ d \in \mathbb{R}^3 \,\middle|\, d^T \begin{pmatrix}2\\0\\0\end{pmatrix} = 0,\, d^T \begin{pmatrix}0\\0\\-1\end{pmatrix} = 0,\, d^T \begin{pmatrix}1\\1\\1\end{pmatrix} = 0 \right\} = \{0\}\,,\\ C^+(x^*) &= \left\{ d \in \mathbb{R}^3 \,\middle|\, d^T \begin{pmatrix}2\\0\\0\end{pmatrix} = 0,\, d^T \begin{pmatrix}1\\1\\1\end{pmatrix} = 0 \right\} = \left\{ \alpha \begin{pmatrix}0\\-1\\1\end{pmatrix} \,\middle|\, \alpha \in \mathbb{R} \right\}\,.\end{aligned}$$

Chapter 5

Exercise 5.2
We obtain

Table 5.1. Error sequence.

k	r_k
0	0.9
1	0.648
2	0.3359232
3	0.090275517
4	$6.519735181 \cdot 10^{-3}$

Therefore, $q = 2$ is the smallest integer such that $\frac{r_{q+2}}{r_2} \le \frac{1}{10}$.

Exercise 5.4

In four-digit arithmetic we get

Table 5.2. Linear and quadratic convergence.

k	Linear	Quadratic
1	0.9000	0.9000
2	0.8100	0.7290
3	0.7290	0.4783
4	0.6561	0.2059
5	0.5905	0.0382

Exercise 5.5

(a) The relation $\frac{r_{k+1}}{(r_k)^2} \le \gamma$ implies

$$\frac{r_{k+1}}{r_k} \le \gamma r_k \,.$$

So it holds

$$\lim_{k\to\infty} \frac{r_{k+1}}{r_k} = \gamma \lim_{k\to\infty} r_k = 0\,,$$

since $\{r_k\}$ converges to 0.

(b) Clearly,

$$\lim_{k\to\infty} \frac{r_{k+1}}{r_k} = 0$$

implies $\frac{r_{k+1}}{r_k} \le \beta$ for each $\beta > 0$ and sufficiently large k.

Chapter 6

Exercise 6.4

The Golden Section algorithm results in the following iterations:

Table 6.2. Golden Section.

k	a_k	b_k	v_k	w_k	Case
0	0	3	1.146	1.854	B
1	1.146	3	1.854	2.292	A
2	1.146	2.292	1.584	1.854	B
3	1.584	2.292	1.854	2.022	B

Exercise 6.5

We obtain

$$\begin{aligned} L_{k+1} &= w_k - a_k &&= \beta L_k \quad \text{in case A and} \\ L_{k+1} &= b_k - v_k = b_k - a_k - \alpha L_k = (1-\alpha)L_k &&= \beta L_k \quad \text{in case B}. \end{aligned}$$

Thus,

$$\frac{L_{k+1}}{L_k} = \beta \quad \text{for } k = 0, 1, 2, \ldots$$

Exercise 6.6

We get $|x^* - \overline{x}| \leq \frac{L_k}{2}$ and

$$\frac{L_k}{2} \leq 10^{-4} \Leftrightarrow \frac{L_0}{2}\beta^k \leq 10^{-4}.$$

Since $L_o = 3$, the last relation is equivalent to

$$\beta^k \leq 1/15000 \Leftrightarrow k \geq -\ln 15000/\ln \beta = 19.98.$$

Hence, $k = 20$. The error sequence is 0.5, 0.073, 0.281, 0.062, ...

Exercise 6.7

It holds $f'(x) = 2x - 4$.

Table 6.3. Bisection algorithm.

k	a_k	b_k	m_k	$f'(m_k)$	$f'(m_k) < 0$?
0	1	4	2.5	1	No
1	1	2.5	1.75	−0.5	Yes
2	1.75	2.5	2.125	0.25	No
3	1.75	2.125	1.9375	−0.125	Yes
4	1.9375	2.125	2.03125	0.0625	No

Now it holds $\frac{L_{k+1}}{L_k} = 0.5$ for the lengths of the intervals. Similar to Exercise 6.6 we get

$$\frac{L_k}{2} \leq 10^{-4} \Leftrightarrow \frac{3}{2}0.5^k \leq 10^{-4} \Leftrightarrow k \geq -\ln 15000/\ln 0.5 = 13.87.$$

Therefore, $k = 14$.

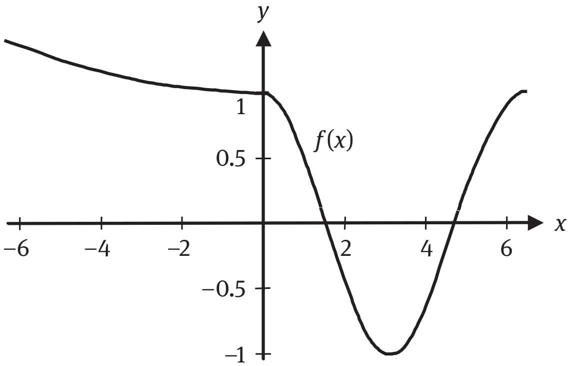


Fig. 6.9. Premature convergence of the bisection method.

Exercise 6.8

Consider the function $f : [-2\pi, 2\pi] \to \mathbb{R}$ with

$$f(x) = \begin{cases} 1 + x^2/100 & \text{for } -2\pi \le x < 0 \\ \cos x & \text{for } 0 \le x \le 2\pi\,. \end{cases}$$

Obviously, f is unimodal and differentiable and has the unique minimum point $\overline{x} = \pi$. Moreover, $f'(0) = 0$, i.e. the condition $f'(x) \neq 0$ for $x \neq \overline{x}$ is not satisfied (see Figure 6.9). Applying the bisection algorithm, we obtain $m_0 = 0$. Since $f'(m_0) = 0$, the algorithm ends with the *false* information that $\overline{x} = 0$.

Exercise 6.14

Since

$$f(x) = x^2 + \sqrt{x} - 15\,,$$
$$f'(x) = 2x + \frac{1}{2\sqrt{x}}\,,$$

we obtain:

$$x_{k+1} = x_k - \frac{{x_k}^2 + \sqrt{x_k} - 15}{2x_k + \frac{1}{2\sqrt{x_k}}}\,.$$

Therefore, $x_0 = 4, x_1 = 3.636, x_2 = 3.619, x_3 = 3.619$.

Exercise 6.15

(a) Since

$$g(x) = \frac{1}{x^2} + x\,,$$
$$g'(x) = -\frac{2}{x^3} + 1\,,$$
$$g''(x) = \frac{6}{x^4}\,,$$

we get:

$$x_{k+1} = \frac{4}{3}x_k - \frac{1}{6}x_k^4 .$$

Therefore, $x_0 = 1$, $x_1 = 1.167$, $x_2 = 1.247$, $x_3 = 1.260$, $x_4 = 1.260$.

(b) It can be easily verified that the conditions of Corollary 6.12 are satisfied for $a = 0.1$ and $b = 1.9$, i.e. the procedure converges to the minimum if $0.1 \le x_0 \le 1.9$. For $x_0 = 3$, the procedure converges to $-\infty$. We obtain $x_1 = -9.5$, $x_2 = -1370.2\dots$

Exercise 6.16

We have to determine a root of the function

$$h(t) = P(\text{erro II}) - P(\text{erro I}) = \Phi\left(\frac{t-180}{1.2}\right) + \Phi\left(\frac{t-176}{1.5}\right) - 1 .$$

The derivative is

$$\begin{aligned} h'(t) &= \frac{1}{1.2}\varphi\left(\frac{t-180}{1.2}\right) + \frac{1}{1.5}\varphi\left(\frac{t-176}{1.5}\right) \\ &= \frac{1}{1.2\sqrt{2\pi}}\exp\left(-\frac{(t-180)^2}{2.88}\right) + \frac{1}{1.5\sqrt{2\pi}}\exp\left(-\frac{(t-176)^2}{5.4}\right) \quad \text{and} \\ t_{k+1} &= t_k - \frac{h(t_k)}{h'(t_k)} . \end{aligned}$$

For $t_0 = 178$ we get $h(t_0) = -0.0443$, $h'(t_0) = 0.1922$, $t_1 = 178.2$. This value is already "close" to the root. The next iteration yields

$$h(t_1) = -0.0040 , \quad h'(t_1) = 0.1987 , \quad t_2 = 178.2 .$$

Exercise 6.18

In the first iteration we get

$$\begin{aligned} p(x_0) &= f(x_0) , \\ p'(x_0) &= f'(x_0) , \\ p(x_1) &= f(x_1) , \end{aligned}$$

equivalent to

$$\begin{aligned} 0.125a + 0.5b + c &= 2.070 , \\ 0.5a + b \quad &= -0.8196 , \\ 0.245a + 0.7b + c &= 1.980 . \end{aligned}$$

The solution is given by $a = 3.696$, $b = -2.668$, $c = 2.942$.

Hence, $x_2 = -b/a = 0.7219$. The second iteration yields

$$\begin{aligned} p(x_1) &= f(x_1) , \\ p'(x_1) &= f'(x_1) , \\ p(x_2) &= f(x_2) , \end{aligned}$$

equivalent to

$$\begin{aligned} 0.245a + 0.7b + c &= 1.980\,, \\ 0.7a + b &= -0.08976\,, \\ 0.2606a + 0.7219b + c &= 1.979\,, \end{aligned}$$

with the solution $a = 3.577$, $b = -2.594$, $c = 2.919$. Hence, $x_3 = -b/a = 0.7252$.

Exercise 6.19

The conditions (6.11) imply

$$\begin{aligned} p'(x_k) &= ax_k + b = f'(x_k)\,, \\ p''(x_k) &= a = f''(x_k)\,, \end{aligned}$$

i.e.

$$\begin{aligned} a &= f''(x_k)\,, \\ b &= f'(x_k) - x_k f''(x_k)\,. \end{aligned}$$

Hence, the polynomial has the minimum point

$$x_{k+1} = -b/a = x_k - \frac{f'(x_k)}{f''(x_k)}\,.$$

This is the recurrence formula of Newton's method.

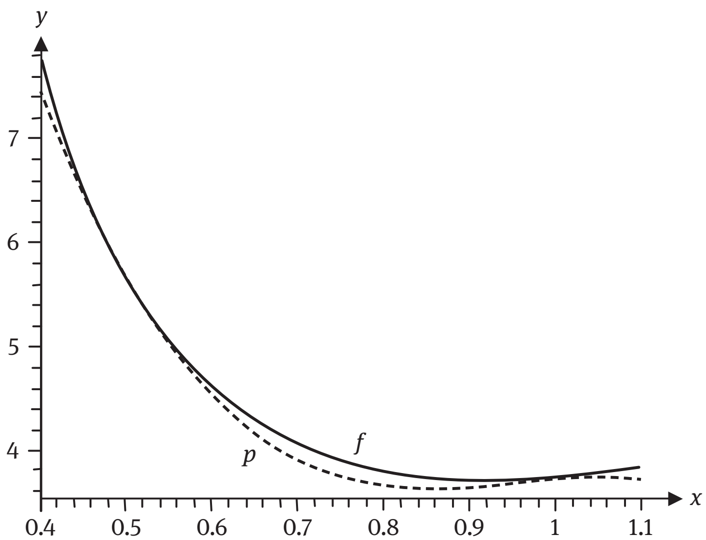


Fig. 6.10. Hermite interpolation polynomial.

Exercise 6.21

The system (6.13) results in

$$\begin{aligned}
\frac{0.5^3}{3}a + \frac{0.5^2}{2}b \quad +0.5c + d \quad &= 4 \quad + e^{0.5}\,,\\
0.5^2 a + 0.5b \quad +c \quad &= -16 \quad + e^{0.5}\,,\\
\frac{1}{3}a + \frac{1}{2}b \quad +c + d \quad &= 1 \quad + e\,,\\
a + b \quad +c \quad &= -2 \quad + e
\end{aligned}$$

and has the solution $a = -70.93$, $b = 136.5$, $c = -64.89$, $d = 23.98$. The Hermite interpolation polynomial (see Figure 6.10) is

$$p(x) = -23.64x^3 + 68.25x^2 - 64.89x + 23.98\,.$$

Chapter 7

Exercise 7.7

We obtain the eigenvalues $\lambda_1 = 5$, $\lambda_2 = 2$ and the corresponding eigenvectors $y^1 = (\sqrt{1/3}, \sqrt{2/3})^T$, $y^2 = (-\sqrt{2/3}, \sqrt{1/3})^T$, aligned with the principal axes of the ellipses (see Figure 7.12). For $z = 1$, the radii are $r_1 = \sqrt{1/5} \approx 0.447$ and $r_2 = \sqrt{1/2} \approx 0.707$.

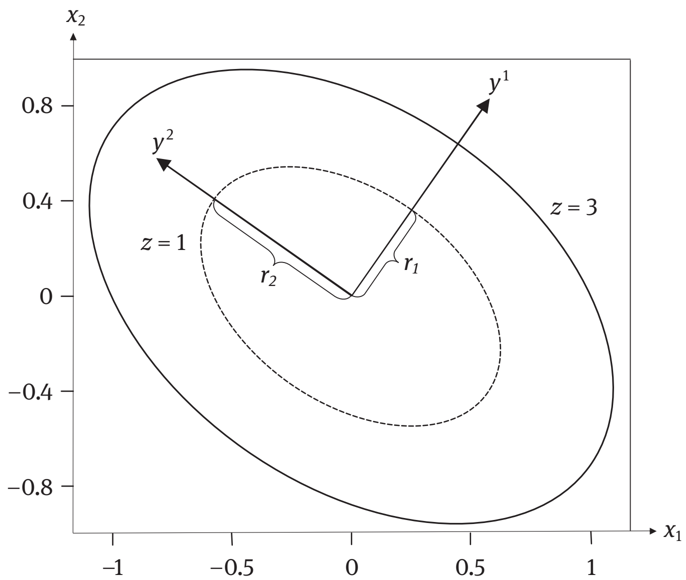


Fig. 7.12. Level curves of Exercise 7.7.

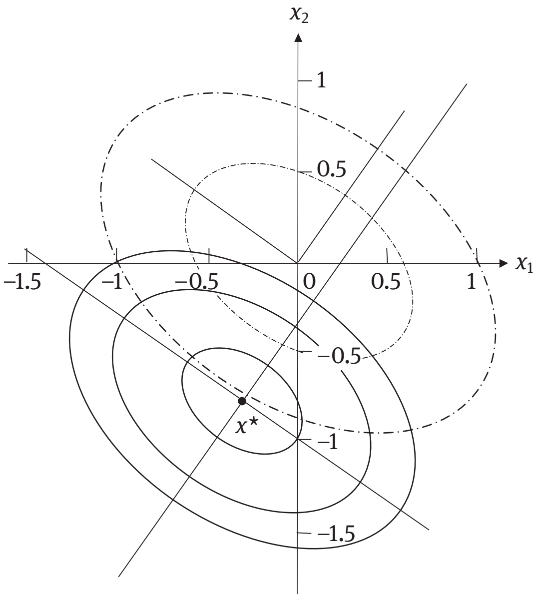


Fig. 7.13. Displaced ellipses.

Exercise 7.8

The center x^* of the displaced ellipses is the minimum point of the objective function (after addition of the linear term), i.e.

$$x^* = -(2Q)^{-1}c = -\begin{pmatrix} 6 & 2\sqrt{2} \\ 2\sqrt{2} & 8 \end{pmatrix}^{-1} \begin{pmatrix} 4 \\ 7 \end{pmatrix} = \begin{pmatrix} -\frac{4}{5} + \frac{7}{20}\sqrt{2} \\ \frac{21}{20} + \frac{\sqrt{2}}{5} \end{pmatrix} \approx \begin{pmatrix} -0.3050 \\ -0.7672 \end{pmatrix}.$$

Figure 7.13 shows these ellipses (level curves) for $z = -1, -2, -3$.

Exercise 7.9

Figure 7.14 illustrates the level curves of the function $f(x) = x^T Q x$ for the positive semidefinite matrix

$$Q = \begin{pmatrix} 9 & 6 \\ 6 & 4 \end{pmatrix}$$

and the indefinite matrix

$$Q = \begin{pmatrix} 2 & -4 \\ -4 & 3 \end{pmatrix}$$

(see Exercise 2.30). The curves are straight lines and hyperbolas, respectively.

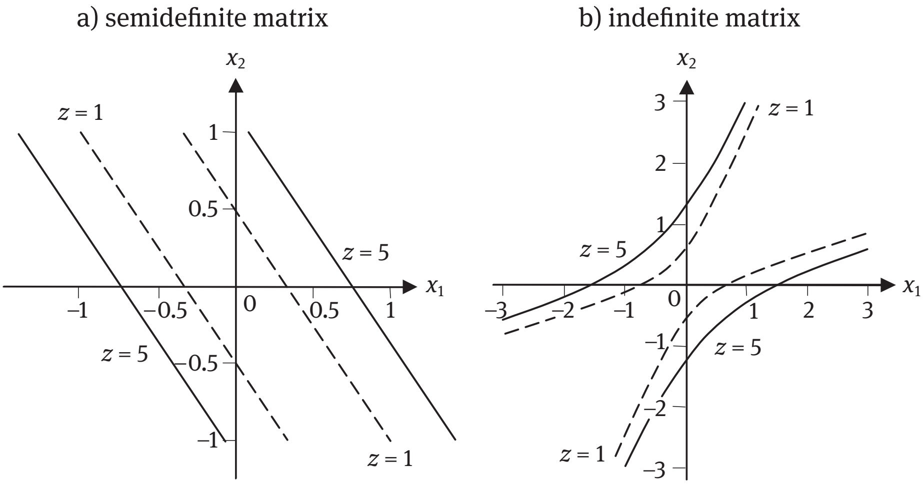


Fig. 7.14. Level curves of quadratic functions.

Exercise 7.12
The derivative of the one-dimensional function $g(\lambda) = f(x^k + \lambda d^k)$ is obtained by using the chain rule:

$$g'(\lambda) = [\operatorname{grad} f(x^k + \lambda d^k)]^T \operatorname{grad} f(x^k)\,.$$

This expression is zero for $\lambda = \lambda_k$, thus

$$g'(\lambda_k) = [\operatorname{grad} f(x^{k+1})]^T \operatorname{grad} f(x^k) = [d^{k+1}]^T d^k = 0\,.$$

Exercise 7.15
We obtain

$$d^0 = \begin{pmatrix} -4 \\ -4 \end{pmatrix}, \quad \lambda_0 = 0.397\,, \quad x^1 = \begin{pmatrix} 0.412 \\ -0.588 \end{pmatrix}.$$

Exercise 7.16
The Hessian matrix

$$Q = \begin{pmatrix} 4 & 0 \\ 0 & 6 \end{pmatrix}$$

has the eigenvalues $\mu = 4$ and $\nu = 6$. Theorem 7.13 states a linear convergence with rate not exceeding

$$\left(\frac{6-4}{6+4}\right)^2 = \frac{4}{100}\,.$$

Table 7.8. Gradient method.

k	x^k	$f(x^k)$	d^k	λ_k
0	$\binom{3}{2}$	30	$\binom{-12}{-12}$	1/5
1	$\frac{1}{5}\binom{3}{-2}$	1.2	$\binom{-2.4}{2.4}$	1/5
2	$\frac{1}{25}\binom{3}{2}$	0.048	$\binom{-0.48}{-0.48}$	1/5
3	$\frac{1}{125}\binom{3}{-2}$	0.00192		

We obtain $r_k = f(x^k) - f(x^*) = f(x^k)$, and with the aid of Table 7.8 we verify that

$$\frac{r_{k+1}}{r_k} = \frac{4}{100} \quad \text{for } k = 0, 1, 2\,.$$

Exercise 7.17
By setting

$$y_1 := \sqrt{2}(x_1 - 1)\,, \quad y_2 := 2(x_2 - 2)\,,$$

the objective function is transformed in

$$f(y_1, y_2) = \frac{1}{2}(y_1^2 + y_2^2) = \frac{1}{2}(y_1, y_2)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\,.$$

The eigenvalues of

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

are $\mu = 1$ and $\nu = 1$, i.e. the value of (7.18) is zero. Theorem 7.13 implies that $\frac{r_1}{r_0} \le 0$, i.e. $r_1 = 0$.

Since the error r_1 is zero, the problem is solved in one step. Alternatively we verify this, minimizing the transformed function by the gradient method. For $x^0 \in \mathbb{R}^n$ we get:

$$\begin{aligned}
d^0 &= -\operatorname{grad} f(x^0) = -x^0\,,\\
\lambda_0 &= \frac{x^{0T}x^0}{x^{0T}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}x^0} = 1 \quad \text{(see (7.16))}\,,\\
x^1 &= x^0 + \lambda_0 d^0 = x^0 - x^0 = 0\,.
\end{aligned}$$

Exercise 7.19

Table 7.9. Newton's method.

k	x^k	$\operatorname{grad} f(x^k)$	$Hf(x^k)$	$(Hf(x^k))^{-1}$
0	$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 4 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 2 & 0 \\ 0 & 12 \end{pmatrix}$	$\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$
1	$\begin{pmatrix} 0 \\ 2/3 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 32/27 \end{pmatrix}$	$\begin{pmatrix} 2 & 0 \\ 0 & 16/3 \end{pmatrix}$	$\begin{pmatrix} 1/2 & 0 \\ 0 & 3/16 \end{pmatrix}$
2	$\begin{pmatrix} 0 \\ 4/9 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 256/729 \end{pmatrix}$	$\begin{pmatrix} 2 & 0 \\ 0 & 64/27 \end{pmatrix}$	$\begin{pmatrix} 1/2 & 0 \\ 0 & 27/64 \end{pmatrix}$
3	$\begin{pmatrix} 0 \\ 8/27 \end{pmatrix}$			

Exercise 7.20

We obtain

$$\begin{aligned} f(x_1, x_2) &= (x_1 - 1)^2 + 2(x_2 - 2)^2, \\ \operatorname{grad} f(x_1, x_2) &= \begin{pmatrix} 2(x_1 - 1) \\ 4(x_2 - 2) \end{pmatrix}, \\ Hf(x_1, x_2) &= \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}. \end{aligned}$$

Thus,

$$\begin{aligned} x^1 &= x^0 - (Hf(x^0))^{-1} \operatorname{grad} f(x^0) \\ &= \begin{pmatrix} x_1^0 \\ x_2^0 \end{pmatrix} - \begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} 2(x_1^0 - 1) \\ 4(x_2^0 - 2) \end{pmatrix} \\ &= \begin{pmatrix} x_1^0 \\ x_2^0 \end{pmatrix} - \begin{pmatrix} x_1^0 - 1 \\ x_2^0 - 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}. \end{aligned}$$

Exercise 7.22

(a) We get

$$\begin{aligned} f(x_1, x_2) &= (x_1 - 2)^4 + (x_2 - 5)^4, \\ \operatorname{grad} f(x_1, x_2) &= \begin{pmatrix} 4(x_1 - 2)^3 \\ 4(x_2 - 5)^3 \end{pmatrix}, \\ Hf(x_1, x_2) &= \begin{pmatrix} 12(x_1 - 2)^2 & 0 \\ 0 & 12(x_2 - 5)^2 \end{pmatrix}. \end{aligned}$$

Thus,

$$\begin{aligned} x^{k+1} &= x^k - (Hf(x^k))^{-1} \operatorname{grad} f(x^k) \\ &= \begin{pmatrix} x_1^k \\ x_2^k \end{pmatrix} - \begin{pmatrix} \frac{1}{12(x_1^k-2)^2} & 0 \\ 0 & \frac{1}{12(x_2^k-5)^2} \end{pmatrix} \begin{pmatrix} 4(x_1^k-2)^3 \\ 4(x_2^k-5)^3 \end{pmatrix} \\ &= \begin{pmatrix} x_1^k \\ x_2^k \end{pmatrix} - \begin{pmatrix} \frac{1}{3}(x_1^k-2) \\ \frac{1}{3}(x_2^k-5) \end{pmatrix} = \frac{1}{3}\left(2x^k + \begin{pmatrix} 2 \\ 5 \end{pmatrix}\right). \end{aligned}$$

(b) Obviously the formula

$$x^k = \left(1 - \left(\frac{2}{3}\right)^k\right) \begin{pmatrix} 2 \\ 5 \end{pmatrix}$$

is correct for $k = 0$. By induction we obtain

$$\begin{aligned} x^{k+1} &= \frac{1}{3}\left[2x^k + \begin{pmatrix} 2 \\ 5 \end{pmatrix}\right] = \frac{1}{3}\left[2\left(1 - \left(\tfrac{2}{3}\right)^k\right)\begin{pmatrix} 2 \\ 5 \end{pmatrix} + \begin{pmatrix} 2 \\ 5 \end{pmatrix}\right] \\ &= \frac{1}{3}\left[2\left(1 - \left(\tfrac{2}{3}\right)^k\right) + 1\right]\begin{pmatrix} 2 \\ 5 \end{pmatrix} = \left[1 - \left(\tfrac{2}{3}\right)^{k+1}\right]\begin{pmatrix} 2 \\ 5 \end{pmatrix}. \end{aligned}$$

(c)

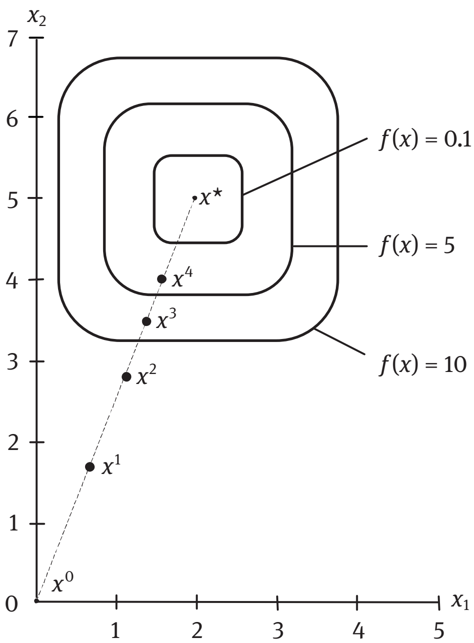


Fig. 7.15. Minimization process of Newton's method.

Exercise 7.23

Figure 7.15 shows that the algorithm starts the search in the direction of the minimum point $x^* = (2,5)^T$. Thus, the exact line search determines this point in the first iteration:

$$x^0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \operatorname{grad} f(x^0) = -\begin{pmatrix} 32 \\ 500 \end{pmatrix}, \quad Hf(x^0) = \begin{pmatrix} 48 & 0 \\ 0 & 300 \end{pmatrix},$$

$$d^0 = \begin{pmatrix} 48 & 0 \\ 0 & 300 \end{pmatrix}^{-1} \begin{pmatrix} 32 \\ 500 \end{pmatrix} = \begin{pmatrix} \frac{1}{48} & 0 \\ 0 & \frac{1}{300} \end{pmatrix} \begin{pmatrix} 32 \\ 500 \end{pmatrix} = \begin{pmatrix} 2/3 \\ 5/3 \end{pmatrix}.$$

The minimum point of the one-dimensional function

$$f(x^0 + d^0) = (\frac{2}{3}\lambda - 2)^4 + (\frac{5}{3}\lambda - 5)^4$$

is $\lambda_0 = 3$. Hence,

$$x^1 = x^0 + \lambda_0 d^0 = \begin{pmatrix} 2 \\ 5 \end{pmatrix}.$$

Exercise 7.27

It holds

$$\begin{aligned} \frac{d}{d\lambda} f(x^k + \lambda d^k) &= [\operatorname{grad} f(x^k + \lambda d^k)]^T d^k = [Q(x^k + \lambda d^k) + q]^T d^k \\ &= (g^k + \lambda Q d^k)^T d^k = g^{kT} d^k + \lambda d^{kT} Q d^k . \end{aligned}$$

The last expression is zero for

$$\lambda = \frac{-g^{kT} d^k}{d^{kT} Q d^k}.$$

Exercise 7.31

Table 7.10. Conjugate gradient method.

k	x^k	g^k	d^k	λ_k	β_k
0	$\binom{6}{3}$	$\binom{-6}{4}$	$\binom{6}{-4}$	13/3	4/9
1	$\binom{32}{-43/3}$	$\binom{8/3}{4}$	$\binom{0}{-52/9}$	3/26	
2	$\binom{32}{-15}$				

The point x^2 is optimal (see Theorem 7.26). The directions are Q-conjugate:

$$d^{0T} Q d^1 = (6, -4) \begin{pmatrix} 3 & 4 \\ 4 & 6 \end{pmatrix} \begin{pmatrix} 0 \\ -52/9 \end{pmatrix} = 0 .$$

Exercise 7.33

The vectors y^1, p^1 and the matrix M_2 have been calculated after the regular termination of the algorithm.

Since

$$M_2 Q = \begin{pmatrix} 3 & -2 \\ -2 & 3/2 \end{pmatrix} \begin{pmatrix} 3 & 4 \\ 4 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

the condition $M_2 = Q^{-1}$ is satisfied.

Table 7.11. DFP method.

k	x^k	g^k	y^k	p^k	M_k	d^k	λ_k
0	$\binom{6}{3}$	$\binom{-6}{4}$	$\binom{26/3}{0}$	$\binom{26}{-52/3}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\binom{6}{-4}$	13/3
1	$\binom{32}{-43/3}$	$\binom{8/3}{4}$	$\binom{-8/3}{-4}$	$\binom{0}{-2/3}$	$\begin{pmatrix} 3 & -2 \\ -2 & 7/3 \end{pmatrix}$	$\binom{0}{-4}$	1/6
2	$\binom{32}{-15}$	$\binom{0}{0}$			$\begin{pmatrix} 3 & -2 \\ -2 & 3/2 \end{pmatrix}$		

Exercise 7.34

The first two iterations for the two methods are summarized in the following tables.

Table 7.12. Conjugate gradient method.

Cycle	k	x^k	g^k	d^k	λ_k	β_k
1	0	$\binom{3}{5}$	$\binom{108}{500}$	$\binom{-108}{-500}$	0.01204	0.001543
	1	$\binom{1.700}{-1.020}$	$\binom{19.64}{-4.234}$	$\binom{-19.81}{3.471}$	0.1044	
2	0	$\binom{-0.3679}{-0.6575}$	$\binom{-0.1992}{-1.137}$	$\binom{0.1992}{1.137}$	0.6915	0.001839
	1	$\binom{-0.2302}{0.1288}$	$\binom{-0.04877}{0.008545}$	$\binom{0.04913}{-0.006454}$	5.640	
3	0	$\binom{0.04697}{0.09239}$				

Table 7.13. DFP method.

Cycle	k	x^k	M_k	d^k	λ_k
1	0	$\binom{3}{5}$	I	$\binom{-108}{-500}$	0.01204
	1	$\binom{1.700}{-1.020}$	$\begin{pmatrix} 0.9707 & -0.1657 \\ -0.1657 & 0.04129 \end{pmatrix}$	$\binom{-19.78}{3.466}$	0.1045
2	0	$\binom{-0.3679}{-0.6575}$	I	$\binom{0.1992}{1.137}$	0.6915
	1	$\binom{-0.2302}{0.1288}$	$\begin{pmatrix} 1.004 & -0.01156 \\ -0.01156 & 0.04904 \end{pmatrix}$	$\binom{0.04904}{-0.006442}$	5.640
3	0	$\binom{0.04697}{0.09239}$			

Exercise 7.35
Since the vectors x^k, g^k, y^k and p^k are identical with that of Exercise 7.33, we list only C_k, D_k, M_k, d^k and λ_k:

Table 7.14. Variant of the DFP method.

k	C_k	D_k	M_k	d^k	λ_k
0	$\begin{pmatrix} 4 & -8/3 \\ -8/3 & 16/9 \end{pmatrix}$	$\begin{pmatrix} 2 & -2/3 \\ -2/3 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 6 \\ -4 \end{pmatrix}$	13/3
1	$\begin{pmatrix} 0 & 0 \\ 0 & 29/18 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 26/9 \end{pmatrix}$	$\begin{pmatrix} 3 & -2 \\ -2 & 25/9 \end{pmatrix}$	$\begin{pmatrix} 0 \\ -52/9 \end{pmatrix}$	3/26
2			$\begin{pmatrix} 3 & -2 \\ -2 & 3/2 \end{pmatrix}$		

The vectors y^1, p^1 and the matrix C_1, D_1 and M_2 have been calculated after the regular termination of the algorithm.

Exercise 7.37
We obtain

$$x^0 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \quad f(x^0 + \lambda e^1) = |4 + \lambda| + 2(4 + \lambda)^2 + 1, \quad \lambda_0 = -4,$$

$$x^1 = x^0 + \lambda_0 e^1 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \quad f(x^1 + \lambda e^2) = |1 + \lambda|, \quad \lambda_1 = -1,$$

$$x^2 = x^1 + \lambda_1 e^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The point x^2 is obviously optimal.

Exercise 7.39
The sequence of points x^k converges to the optimal point $x^* = (2, 6)^T$ (see Table 7.15).

Exercise 7.43
(i) It holds $\sup_{x \in M} \{|\operatorname{grad} f(x)|\} = \sup_{x \in M} \{|2x|\}$. Therefore, the function f is not Lipschitz continuous for $M = \mathbb{R}$ and for $M = [-1, 3]$, the smallest Lipschitz constant is $L = \sup_{x \in [-1,3]} \{|2x|\} = 6$.
(ii) Now it holds $\sup_{x>0} \{|\operatorname{grad} f(x)|\} = \sup_{x>0} \{\frac{1}{x^2}\} = \infty$. Therefore, f is not Lipschitz continuous.
(iii) It holds $|\operatorname{grad} f(x)| = \left|\begin{pmatrix} 4x_1 \\ 3 \end{pmatrix}\right| = \sqrt{16x_1^2 + 9}$. Hence, f is not Lipschitz continuous for $M = \mathbb{R}^2$, and in the other case the smallest Lipschitz constant is $\sqrt{16 \cdot 5^2 + 9} = \sqrt{409}$.

Table 7.15. Cyclic minimization.

k	x^k	$f(x^k)$	λ_k
0	$\binom{3}{5}$	17	-1.3
1	$\binom{1.7}{5}$	0.1	0.1
2	$\binom{1.7}{5.1}$	0.09	0.03
3	$\binom{1.73}{5.1}$	0.081	0.09
4	$\binom{1.73}{5.19}$	0.0729	0.027
5	$\binom{1.757}{5.19}$	0.06561	0.081
6	$\binom{1.757}{5.271}$	0.059049	

Exercise 7.46

(a) For $x^k = \binom{x_1^k}{x_2^k}$ it holds $\operatorname{grad} f(x^k) = \binom{x_1^k}{9x_2^k}$, $d^k = -\binom{x_1^k}{9x_2^k}$.

Iteration 1:

$$x^0 = \begin{pmatrix} 9 \\ 1 \end{pmatrix}, \quad \operatorname{grad} f(x^0) = \begin{pmatrix} 9 \\ 9 \end{pmatrix}, \quad d^0 = -\begin{pmatrix} 9 \\ 9 \end{pmatrix}.$$

Armijo condition:

$$f\left(\begin{pmatrix} 9 \\ 1 \end{pmatrix} - \lambda^{(i)} \begin{pmatrix} 9 \\ 9 \end{pmatrix}\right) < 45 - \frac{1}{4}\lambda^{(i)} 162\,.$$

The first index i for which this condition is satisfied, is $i = 2$. Thus, $\lambda_0 = 1/4$.

$$x^1 = x^0 + \lambda_0 d^0 = \begin{pmatrix} 9 \\ 1 \end{pmatrix} - \frac{1}{4}\begin{pmatrix} 9 \\ 1 \end{pmatrix} = \begin{pmatrix} 6.75 \\ -1.25 \end{pmatrix}.$$

Iteration 2:

$$\operatorname{grad} f(x^1) = \begin{pmatrix} 6.75 \\ -11.25 \end{pmatrix}, d^1 = \begin{pmatrix} -6.75 \\ 11.25 \end{pmatrix}.$$

Armijo condition:

$$f\left(\begin{pmatrix} 6.75 \\ -1.25 \end{pmatrix} + \lambda^{(i)} \begin{pmatrix} -6.75 \\ 11.25 \end{pmatrix}\right) < \frac{477}{16} - \lambda^{(i)} \frac{1377}{32}\,.$$

The first index i for which this condition is satisfied, is $i = 3$. Thus, $\lambda_1 = 1/8$.

$$x^2 = x^1 + \lambda_1 d^1 = \begin{pmatrix} 6.75 \\ -1.25 \end{pmatrix} + \frac{1}{8}\begin{pmatrix} -6.75 \\ 11.25 \end{pmatrix} = \begin{pmatrix} 5.90625 \\ 0.15625 \end{pmatrix}.$$

(b)

$$Hf(x) = \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}, \quad d^k = -\begin{pmatrix} 1 & 0 \\ 0 & 1/9 \end{pmatrix}\begin{pmatrix} x_1^k \\ 9x_2^k \end{pmatrix} = -\begin{pmatrix} x_1^k \\ x_2^k \end{pmatrix} = -x^k\,.$$

Iteration 1:

$$x^0 = \begin{pmatrix} 9 \\ 1 \end{pmatrix}, \quad \operatorname{grad} f(x^0) = \begin{pmatrix} 9 \\ 9 \end{pmatrix}, \quad d^0 = -\begin{pmatrix} 9 \\ 9 \end{pmatrix}.$$

The Armijo condition

$$f\left(\begin{pmatrix} 9 \\ 1 \end{pmatrix} - \lambda^{(i)} \begin{pmatrix} 9 \\ 9 \end{pmatrix}\right) < 45 - \frac{90}{4}\lambda^{(i)}$$

is satisfied for $i = 0$. Thus, $\lambda_0 = 1$, $x^1 = x^0 + \lambda_0 d^0 = \binom{9}{1} - \binom{9}{1} = 0$. This point is already optimal.

Exercise 7.52

It holds

$$\operatorname{grad} f(x) = \begin{pmatrix} 2x_1^3 + 6x_1 \\ 8x_2 \end{pmatrix}, \quad Hf(x) = \begin{pmatrix} 6x_1^2 + 6 & 0 \\ 0 & 8 \end{pmatrix}$$

(positive definite for all x).

$k = 0$:

$$c^0 = \operatorname{grad} f(x^0) = \begin{pmatrix} 28 \\ 8 \end{pmatrix}, \quad Q_0 = Hf(x^0) = \begin{pmatrix} 30 & 0 \\ 0 & 8 \end{pmatrix}.$$

It holds $|Q_0^{-1}c^0| = \sqrt{(14/15)^2 + 1} > \Delta_0 = 1$. Since Q_0 is a diagonal matrix, we get $\lambda_1 = 30$, $\lambda_2 = 8$, $y^1 = \binom{1}{0}$, $y^2 = \binom{0}{1}$, $\alpha_1 = 28$, $\alpha_2 = 8$, and the equation

$$\Delta_0^2 = 1 = \frac{\alpha_1^2}{(\lambda_1 + \lambda^*)^2} + \frac{\alpha_2^2}{(\lambda_2 + \lambda^*)^2} = \frac{784}{(30 + \lambda^*)^2} + \frac{64}{(8 + \lambda^*)^2}$$

has the solution $\lambda^* \approx 5.201$. Thus,

$$s^0 = -(Q_0 + \lambda^* I)^{-1} c^0 = -\begin{pmatrix} \frac{1}{30+\lambda^*} & 0 \\ 0 & \frac{1}{8+\lambda^*} \end{pmatrix} \begin{pmatrix} 28 \\ 8 \end{pmatrix} \approx -\begin{pmatrix} 0.7954 \\ 0.6060 \end{pmatrix}.$$

The evaluation of the decrease results in

$$\rho_0 = \frac{f(x^0) - f(x^0 + s^0)}{f(x^0) - q_0(s^0)} \approx \frac{24 - 6.027}{24 - 7.840} \approx 1.112 \geq \beta,$$

thus,

$$x^1 = x^0 + s^0 \approx \begin{pmatrix} 1.205 \\ 0.3940 \end{pmatrix}$$

and the new convergence radius is $\Delta_1 = q\Delta_0 = 2$.

$k = 1$:
It holds

$$c^1 = \operatorname{grad} f(x^1) \approx \begin{pmatrix} 10.73 \\ 3.152 \end{pmatrix}, \quad Q_1 = Hf(x^1) \approx \begin{pmatrix} 14.71 & 0 \\ 0 & 8 \end{pmatrix},$$

and the point $-Q_1^{-1}c^1 \approx -\binom{0.7249}{0.3940}$ is in the interior of the trust region, i.e. $s^1 \approx -\binom{0.7294}{0.3940}$. Since

$$\rho_1 = \frac{f(x^1) - f(x^1 + s^1)}{f(x^1) - q_1(s^1)} \approx \frac{6.031 - 0.7042}{6.031 - 1.497} \approx 1.175 \geq \beta,$$

we set

$$x^2 = x^1 + s^1 \approx \begin{pmatrix} 0.4756 \\ 0 \end{pmatrix}, \quad \Delta_2 = q\Delta_1 = 4.$$

$k = 2$:
We get

$$c^2 = \operatorname{grad} f(x^2) = \begin{pmatrix} 3.069 \\ 0 \end{pmatrix}, \ Q_2 = Hf(x^2) \approx \begin{pmatrix} 7.357 & 0 \\ 0 & 8 \end{pmatrix}$$

and $-Q_2^{-1}c^2 \approx \binom{-0.4172}{0}$ is in the interior of the trust region, i.e. $s^2 \approx \binom{-0.4172}{0}$.
We obtain

$$\rho_2 \approx \frac{0.7042 - 0.01023}{0.7042 - 0.06408} \approx 1.084 \geq \beta,$$

therefore,

$$x^3 = x^2 + s^2 \approx \begin{pmatrix} 0.0584 \\ 0 \end{pmatrix}.$$

The algorithm converges to the optimal solution $x^* = 0$.

Exercise 7.53

We observe that α^* is the solution of the one-dimensional problem

$$\min_{\alpha \leq \Delta/|c|} g(\alpha) := \frac{1}{2}a\alpha^2 - b\alpha \qquad (*)$$

where $a := c^TQc > 0$ (since Q is positive definite) and $b := c^Tc \geq 0$. The function $g(\alpha)$ is a convex parabola, decreasing at $\alpha = 0$. Let $\overline{\alpha}$ be the minimum point of the problem $\min_{\alpha>0} g(\alpha)$, i.e. $\overline{\alpha} := b/a = \frac{c^Tc}{c^TQc}$. If $\overline{\alpha} \leq \Delta/|c|$, this point is the solution of $(*)$, i.e. $\alpha^* = \overline{\alpha}$. Otherwise, the function $g(\alpha)$ is decreasing over the interval $[0, \Delta/|c|]$, i.e. $\alpha^* = \Delta/|c|$. Hence, the Cauchy point is

$$s_C = -\alpha^* c \quad \text{with} \quad \alpha^* = \min\left\{\frac{c^Tc}{c^TQ_c}, \frac{\Delta}{|c|}\right\}.$$

For the first subproblem of Example 7.50 we obtain

$$Q = \begin{pmatrix} 14 & 1 \\ 1 & 2 \end{pmatrix}, \quad c = \begin{pmatrix} 7 \\ 5 \end{pmatrix}, \quad \Delta = 1.$$

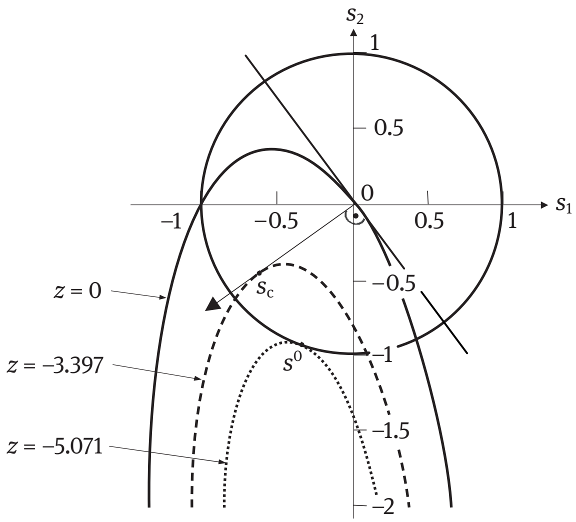


Fig. 7.16. Level curves of the first quadratic model of Example 7.50.

Therefore,

$$\frac{c^T c}{c^T Qc} = \frac{37}{403} \approx 0.09181, \frac{\Delta}{|c|} = \frac{1}{\sqrt{74}} \approx 0.1162, s_C = -\frac{37}{403}\begin{pmatrix}7\\5\end{pmatrix} \approx -\begin{pmatrix}0.6427\\0.4591\end{pmatrix}.$$

The objective values are $q(s_C) \approx -3.397$ and $q(s^0) \approx -5.071$. The points s_C and s^0 are illustrated in Figure 7.16, where the arrow through s_C indicates the maximum descent direction.

Exercise 7.54

(a) We get

$$Q = \begin{pmatrix}2 & 0\\0 & -2\end{pmatrix}$$

(indefinite), $c = 0$, and $\Delta = 1$. A global minimum point s^* must satisfy the following conditions:

$$(Q + \lambda^* I)s^* = -c \Leftrightarrow \begin{pmatrix}2+\lambda^* & 0\\0 & -2+\lambda^*\end{pmatrix}\begin{pmatrix}s_1^*\\s_2^*\end{pmatrix} = 0\,, \tag{1}$$

$$\begin{pmatrix}2+\lambda^* & 0\\0 & -2+\lambda^*\end{pmatrix} \text{ positive semidefinite}\,, \tag{2}$$

$$\lambda^* \geq 0\,, \tag{3}$$

$$\lambda^*(|s^*| - 1) = 0\,. \tag{4}$$

From conditions (2) and (4), it follows that

$$\lambda^* \geq 2, \tag{5}$$

$$|s^*| = 1. \tag{6}$$

Now it follows from (1), (5) and (6) that $s_1^* = 0$ and $s_2^* = \pm 1$. So we obtain the two global minimum points $s^1 := (0, 1)^T$ and $s^2 := (0, -1)^T$.

(b) We get

$$Q = \begin{pmatrix} 4 & -6 \\ -6 & 2 \end{pmatrix}$$

(indefinite), $c = \binom{2}{-1}$, and $\Delta = 1$. A global minimum point s^* must satisfy:

$$\begin{pmatrix} 4+\lambda^* & -6 \\ -6 & 2+\lambda^* \end{pmatrix} \begin{pmatrix} s_1^* \\ s_2^* \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}, \tag{1}$$

$$\begin{pmatrix} 4+\lambda^* & -6 \\ -6 & 2+\lambda^* \end{pmatrix} \text{ positive semidefinite}, \tag{2}$$

$$\lambda^* \geq 0, \tag{3}$$

$$\lambda^*(|s^*| - 1) = 0. \tag{4}$$

From Theorem 2.29 (ii) it follows that condition (2) is equivalent to

$$4+\lambda^* \geq 0, 2+\lambda^* \geq 0, \quad \text{and} \quad (4+\lambda^*)(2+\lambda^*)-36 \geq 0 \Leftrightarrow \lambda^* \geq \sqrt{37}-3 \approx 3.08.$$

But for $\lambda^* = \sqrt{37} - 3$, the system (1) has no solution, therefore,

$$\lambda^* > \sqrt{37} - 3, \tag{5}$$

i.e. the matrix in (1) is positive definite and hence invertible. Now (4) implies

$$|s^*| = 1, \tag{6}$$

and (1) implies

$$s^* = \begin{pmatrix} 4+\lambda^* & -6 \\ -6 & 2+\lambda^* \end{pmatrix}^{-1} \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \frac{1}{\lambda^{*2} + 6\lambda^* - 28} \begin{pmatrix} 2+\lambda^* & 6 \\ 6 & 4+\lambda^* \end{pmatrix} \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$
$$= \frac{1}{\lambda^{*2} + 6\lambda^* - 28} \begin{pmatrix} 2-\lambda^* \\ \lambda^* - 8 \end{pmatrix}. \tag{7}$$

Finally, it follows from (6) and (7)

$$\frac{\sqrt{4(\lambda^* - 1)^2 + (\lambda^* - 8)^2}}{\lambda^{*2} + 6\lambda^* - 28} = 1 \Rightarrow \lambda^* \approx 3.620,$$

i.e. the global minimum point is $s^* \approx \binom{-0.7673}{-0.6412}$ with objective value $f(s^*) \approx -2.257$ (see Figure 7.17).

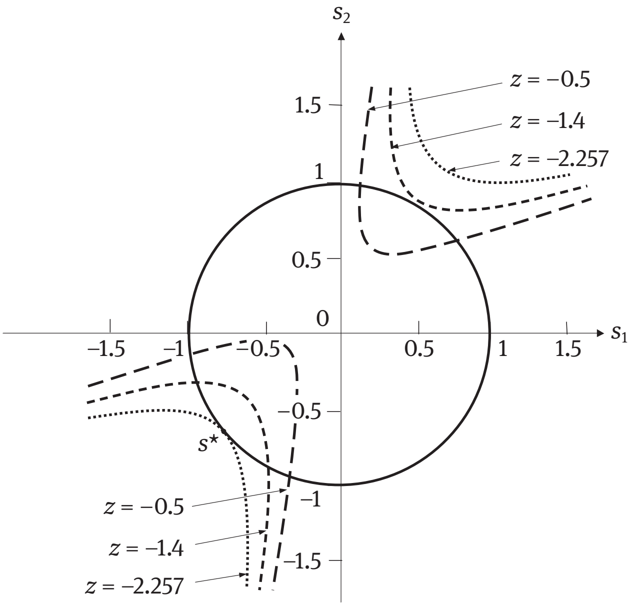


Fig. 7.17. Level curves of the function of Exercise 7.54 (b).

Chapter 8

Exercise 8.3

(a) Iteration 1:

$$\begin{aligned} k &= 0, \quad x^0 = (1/2, 1/4, 0)^T, \\ d^0 &= (-1, -1/2, 0)^T \text{ (see Example 8.1)}, \\ \overline{\lambda}_0 &= 1/2, \quad \lambda_0 = 1/2, \\ x^1 &= x^0 + \lambda_0 d^0 = (0, 0, 0)^T. \end{aligned}$$

Iteration 2:

$$k = 1, \quad x^1 = (0, 0, 0)^T,$$

$$A_1 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad P_1 = 0, \quad d^1 = 0, \quad \alpha^1 = -A_1 \operatorname{grad} f(x^1) = (0, 0, -2)^T,$$

$$A_1 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}, \quad P_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad d^1 = -P_1 \operatorname{grad} f(x^1) = (0, 0, 2)^T,$$

$$\begin{aligned} \overline{\lambda_1} &= 1/2, \quad \lambda_1 = 1/2, \\ x^2 &= x^1 + \lambda_1 d^1 = (0, 0, 1)^T. \end{aligned}$$

Iteration 3:

$$k = 2\,, x^2 = (0,0,1)^T\,,$$

$$A_2 = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}\,, \quad A_2 A_2^T = \begin{pmatrix} -3 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}\,,$$

$$P_2 = I - \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} = 0\,, \quad d^2 = 0\,.$$

Since grad $f(x^2) = 0$, it holds $\alpha^2 = 0$, and the algorithm terminates with the optimal solution x^2.

(b) Iteration 1: $\quad k = 0\,, \quad x^0 = (0, 1/2, 0)^T\,, \quad d^0 = (0, -1, 0)^T\,,$

$$\overline{\lambda_0} = 1/2\,, \quad \lambda_0 = 1/2\,,$$

$$x^1 = x^0 + \lambda_0 d^0 = (0,0,0)^T\,.$$

The following iterations are the same as in part (a).

Exercise 8.4

Each vector $v \in \mathbb{R}^n$ can be represented as

$$v = Z^*\beta^* + A^{*T}\alpha^*\,, \tag{*}$$

where $\alpha^* \in \mathbb{R}^q$ and $\beta^* \in \mathbb{R}^{n-q}$ are uniquely determined. Since $A^*Z^* = 0$, multiplying this equation by Z^{*T} yields

$$Z^{*T}v = Z^{*T}Z^*\beta^* + Z^{*T}A^{*T}\alpha^* = Z^{*T}Z^*\beta^*\,.$$

Hence,

$$\beta^* = (Z^{*T}Z^*)^{-1}Z^{*T}v\,.$$

The projection of the vector v into the space $Nu(A)$ is therefore:

$$Z^*\beta^* = Z^*(Z^{*T}Z^*)^{-1}Z^{*T}v\,,$$

and the projection matrix can be expressed as:

$$P^* = Z^*(Z^{*T}Z^*)^{-1}Z^{*T}\,.$$

Exercise 8.5

Iteration 1: $k = 0, x^0 = (1,2,1)^T$,

$$A_0 = (2,1,1)\,, \quad A_0A_0^T = 6\,, \quad (A_0A_0^T)^{-1} = 1/6\,,$$

$$P_0 = I - \frac{1}{6}\begin{pmatrix}2\\1\\1\end{pmatrix}(2,1,1)$$

$$= \begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} - \frac{1}{6}\begin{pmatrix}4&2&2\\2&1&1\\2&1&1\end{pmatrix} = \begin{pmatrix}1/3&-1/3&-1/3\\-1/3&5/6&-1/6\\-1/3&-1/6&5/6\end{pmatrix},$$

$$d^0 = -\begin{pmatrix}1/3&-1/3&-1/3\\-1/3&5/6&-1/6\\-1/3&-1/6&5/6\end{pmatrix}\begin{pmatrix}0\\2\\2\end{pmatrix} = \begin{pmatrix}4/3\\-4/3\\-4/3\end{pmatrix},$$

$$\overline{\lambda}_0 = \min\left\{\frac{-(0,-1,0)\begin{pmatrix}1\\2\\1\end{pmatrix}}{(0,-1,0)\begin{pmatrix}4/3\\-4/3\\-4/3\end{pmatrix}}, \frac{-(0,0,-1)\begin{pmatrix}1\\2\\1\end{pmatrix}}{(0,0,-1)\begin{pmatrix}4/3\\-4/3\\-4/3\end{pmatrix}}\right\} = \frac{3}{4},$$

$$\min_{0\le\lambda\le 3/4} f\left(x^0 + \lambda d^0\right) := \left(\frac{4}{3}\lambda\right)^2 + \left(1 - \frac{4}{3}\lambda\right)^2 + \left(1 - \frac{4}{3}\lambda\right)^2.$$

We obtain $\lambda_0 = 1/2$,

$$x^1 = x^0 + \lambda_0 d^0 = \begin{pmatrix}1\\2\\1\end{pmatrix} + \frac{1}{2}\begin{pmatrix}4/3\\-4/3\\-4/3\end{pmatrix} = \begin{pmatrix}5/3\\4/3\\1/3\end{pmatrix}.$$

Iteration 2: $k = 1\,, \quad x^1 = (5/3, 4/3, 1/3)^T$.

As above:

$$A_1 = (2,1,1)\,, \quad P_1 = \begin{pmatrix}1/3&-1/3&-1/3\\-1/3&5/6&-1/6\\-1/3&-1/6&5/6\end{pmatrix},$$

$$d^1 = -P_1\begin{pmatrix}4/3\\2/3\\2/3\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix},$$

$$\alpha^1 = -\left[(2,1,1)\begin{pmatrix}2\\1\\1\end{pmatrix}\right]^{-1}(2,1,1)\begin{pmatrix}4/3\\2/3\\2/3\end{pmatrix} = -2/3 < 0\,,$$

$$d^1 = -I\,\operatorname{grad} f(x^1) = -\begin{pmatrix}4/3\\2/3\\2/3\end{pmatrix},$$

$$\overline{\lambda}_1 = \min\left\{ \frac{-(-1,0,0)\begin{pmatrix}5/3\\4/3\\1/3\end{pmatrix}}{(-1,0,0)\begin{pmatrix}-4/3\\-2/3\\-2/3\end{pmatrix}}, \frac{-(0,-1,0)\begin{pmatrix}5/3\\4/3\\1/3\end{pmatrix}}{(0,-1,0)\begin{pmatrix}-4/3\\-2/3\\-2/3\end{pmatrix}}, \frac{-(0,0,-1)\begin{pmatrix}5/3\\4/3\\1/3\end{pmatrix}}{(0,0,-1)\begin{pmatrix}-4/3\\-2/3\\-2/3\end{pmatrix}} \right\} = \frac{1}{2},$$

$$\min_{0\le\lambda\le 1/2} f\left(x^1 + \lambda d^1\right) := \left(\frac{2}{3} - \frac{4}{3}\lambda\right)^2 + 2\left(\frac{1}{3} - \frac{2}{3}\lambda\right)^2 .$$

Clearly, the minimum point is $\lambda_1 = 1/2$. We set

$$x^2 = x^1 + \lambda_1 d^1 = \begin{pmatrix}5/3\\4/3\\1/3\end{pmatrix} - \frac{1}{2}\begin{pmatrix}4/3\\2/3\\2/3\end{pmatrix} = \begin{pmatrix}1\\1\\0\end{pmatrix}.$$

which is the solution of the exercise.

Exercise 8.7

$\overline{d}$ is solution of the problem

$$\begin{aligned} &\min 6d_1 + 4d_2 \\ &d_1 + d_2 \le 0 \\ &d^T d \le 1 \end{aligned}$$

and can be determined graphically. We obtain $\overline{d} = -\frac{1}{\sqrt{13}}\binom{3}{2}$ (see Figure 8.5 and Exercise 4.34).

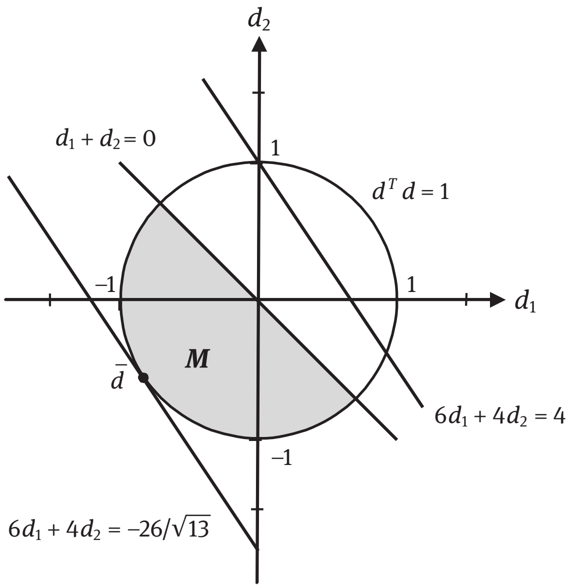


Fig. 8.5. Geometric solution of Exercise 8.7.

Exercise 8.8

By applying the Cauchy–Schwarz inequality, we get

$$-g^{*T}d \le |g^{*T}d| \le |g^*||d| \le |g^*| \quad \text{for all } d \text{ with } d^Td \le 1\,.$$

Hence,

$$g^{*T}d^* = -\frac{1}{|g^*|}g^{*T}g^* = -|g^*| \le g^{*T}d \quad \text{for all } d \text{ with } d^Td \le 1\,,$$

i.e. d^* is the solution of problem (8.14) with (8.16) instead of (8.15). This statement generalizes the observation of Exercise 8.7.

Exercise 8.9

The feasible region M of the problem is illustrated in Figure 8.6.

It holds

$$\operatorname{grad} f(x^*) = \begin{pmatrix} -2 \\ e^2 - 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 6.39 \end{pmatrix}.$$

Since both constraints are active at x^*, the optimal direction d^* is solution of

$$\begin{aligned} \min -2d_1 + (e^2 - 1)d_2 \\ -d_1 + d_2 \le 0 \\ -d_1 - d_2 \le 0 \\ d_1 \le 1 \\ -1 \le d_2\,. \end{aligned}$$

By solving this linear problem we obtain $d^* = (1, -1)^T$ (see Figure 8.6).

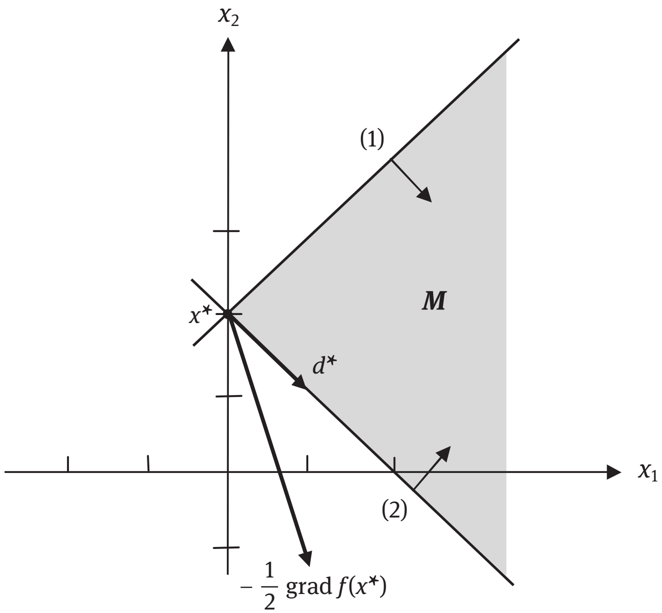


Fig. 8.6. Feasible region.

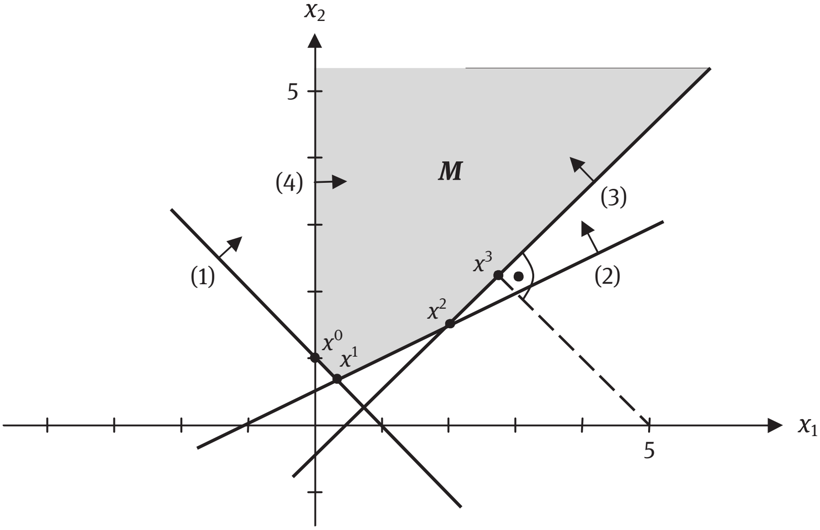


Fig. 8.7. Zoutendijk's method.

Exercise 8.10

We obtain

$$x^0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad d^0 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad \lambda_0 = \frac{1}{3},$$
$$x^1 = \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}, \qquad d^1 = \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}, \qquad \lambda_1 = \frac{5}{3},$$
$$x^2 = \begin{pmatrix} 2 \\ 3/2 \end{pmatrix}, \qquad d^2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \lambda_2 = \frac{3}{4},$$
$$x^3 = \begin{pmatrix} 11/4 \\ 9/4 \end{pmatrix}, \qquad d^3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The algorithm terminates with the optimal point x^3 (see Figure 8.7).

Exercise 8.12

Since $g^0 = \operatorname{grad} f(x^0) = (20, 12)^T$, d^0 is the solution of the linear problem

$$\begin{aligned} \min 20d_1 + 12d_2\,, \\ -d_1 + 5d_2 = 0\,, \\ -1 \le d_1 \le 1\,, \\ -1 \le d_2 \le 1\,, \end{aligned}$$

i.e. $d^0 = (-1, -1/5)^T$. We obtain $\overline{\lambda_0} = 10$, which can be verified graphically in Figure 8.4.

Since

$$f(x^0 + \lambda d^0) = (10 - \lambda)^2 + 3(2 - \lambda/5)^2\,,$$

the problem

$$\min_{0 \le \lambda \le \bar{\lambda}_0} f(x^0 + \lambda d^0)$$

has the solution $\lambda_0 = 10$. Therefore,

$$x^1 = x^0 + \lambda_0 d^0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Exercise 8.13
Since the solution in step (3) is always $d^k = 0$, we obtain the same iteration points as in Exercise 8.10.

Exercise 8.14
Consider the problem

$$\min x_1^2 + x_2^2$$

$$-x_1 + x_2 \le 0 \qquad (1)$$

$$x_1 + x_2 \le 10 \qquad (2)$$

$$x_2 \ge 0 \qquad (3)$$

with $x^0 = (9, 1)^T$ (see Figure 8.8).

If (8.16) is chosen to delimit $|d|$, Zoutendijk's method solves the problem in one step, i.e. $x^1 = (0, 0)^T$. The algorithm of Section 8.1.3 calculates $x^1 = (5, 5)^T$ and $x^2 = (0, 0)^T$.

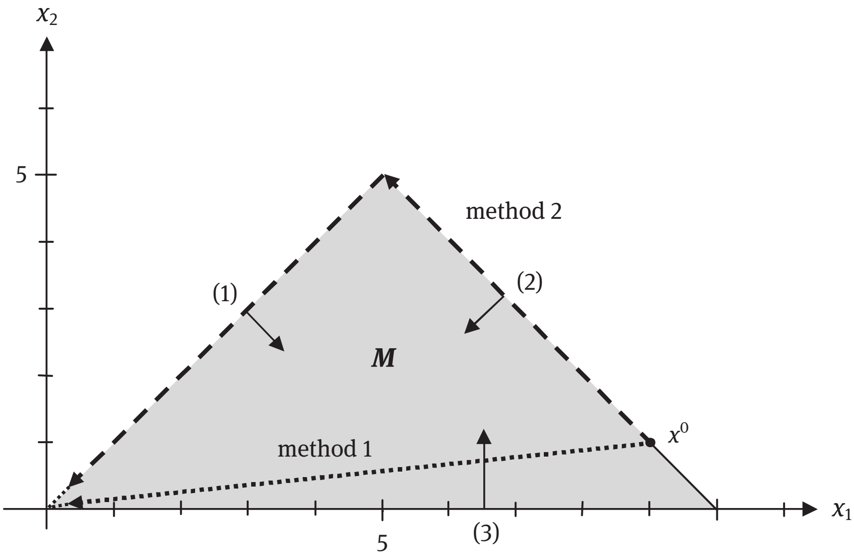


Fig. 8.8. Graphical illustration of both methods.

Exercise 8.16
It holds

$$f(x) = \frac{1}{2}x^T Q x + q^T x + c$$

with

$$Q = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad q = \begin{pmatrix} -2 \\ -2 \\ 2 \end{pmatrix}, \quad c = 3,$$

$$\operatorname{grad} f(x) = 2x + q\,.$$

Iteration 0:

$$g^0 = 2x^0 + q = (-2, -2, 8)^T\,.$$

$$\begin{aligned} \min\ & -2d_1 - 2d_2 + 8d_3 \\ & d_1 + d_2 + d_3 = 0 \\ & d_1 = d_2 = 0 \\ & -1 \le d_i \le 1 \quad (i = 1, 2, 3) \end{aligned}$$

has the solution $d^0 = (0, 0, 0)^T$.

$$\begin{aligned} \min\ & -2d_1 - 2d_2 + 8d_3 \\ & d_1 + d_2 + d_3 \le 0 \\ & d_1, d_2 \ge 0 \\ & -1 \le d_i \le 1 \quad (i = 1, 2, 3) \end{aligned}$$

has the solution $d^0 = (1, 0, -1)^T$.

$$\tilde{\lambda}_0 = \frac{-g^{0T} d^0}{d^{0T} Q d^0} = \frac{5}{2}, \quad \overline{\lambda}_0 = 3, \quad \lambda_0 = \frac{5}{2}, \quad x^1 = x^0 + \lambda_0 d^0 = \left(\frac{5}{2}, 0, \frac{1}{2}\right)^T.$$

Iteration 1:

$$g^1 = (3, -2, 3)^T,$$

$$\begin{aligned} \min\ & 3d_1 - 2d_2 + 3d_3 \\ & d_1 + d_2 + d_3 = 0 \\ & d_2 = 0 \\ & -1 \le d_i \le 1 \quad (i = 1, 2, 3) \end{aligned}$$

has the solution $d^1 = (0, 0, 0)^T$.

$$\begin{aligned}\min 3d_1 - 2d_2 + 3d_3&\\ d_1 + d_2 + d_3 &\le 0\\ d_2 &\ge 0\\ -1 \le d_i &\le 1 \quad (i = 1, 2, 3)\end{aligned}$$

has the solution $d^1 = (-1, 1, -1)^T$.

$$\tilde{\lambda}_1 = \frac{4}{3}, \quad \bar{\lambda}_1 = \frac{1}{2}, \quad \lambda_1 = \frac{1}{2},$$
$$x^2 = x^1 + \lambda_1 d^1 = \left(2, \frac{1}{2}, 0\right)^T.$$

Iteration 2:

$$g^2 = (2, -1, 2)^T.$$
$$\begin{aligned}\min 2d_1 - d_2 + 2d_3&\\ d_3 &= 0\\ -1 \le d_i &\le 1 \quad (i = 1, 2, 3)\end{aligned}$$

has the solution $d^2 = (-1, 1, 0)^T$.

$$\tilde{\lambda}_2 = \frac{3}{4}, \quad \bar{\lambda}_2 = 12, \quad \lambda_2 = \frac{3}{4},$$
$$x^3 = x^2 + \lambda_2 d^2 = \left(\frac{5}{4}, \frac{5}{4}, 0\right)^T.$$

Iteration 3:

$$g^3 = \left(\frac{1}{2}, \frac{1}{2}, 2\right)^T.$$

We add the equation

$$g^3 - g^2 = \left(-\frac{3}{2}, \frac{3}{2}, 0\right)^T,$$
$$\begin{aligned}\min \frac{1}{2}d_1 + \frac{1}{2}d_2 + 2d_3&\\ d_3 &= 0\\ -\frac{3}{2}d_1 + \frac{3}{2}d_2 &= 0\\ -1 \le d_i &\le 1 \quad (i = 1, 2, 3)\end{aligned}$$

has the solution $d^3 = (-1, -1, 0)^T$.

$$\tilde{\lambda}_3 = \frac{1}{4}, \quad \bar{\lambda}_3 = \frac{5}{4}, \quad \lambda_3 = \frac{1}{4}, \quad x^4 = x^3 + \lambda_3 d^3 = (1, 1, 0)^T.$$

Iteration 4:

$$g^4 = (0, 0, 2)^T.$$

We add a second equation

$$g^4 - g^3 = \left(-\frac{1}{2}, -\frac{1}{2}, 0\right)^T,$$

$$\begin{aligned}
&\min 2d_3 \\
&d_3 = 0 \\
-\frac{3}{2}d_1 + \frac{3}{2}d_2 &= 0 \\
-\frac{1}{2}d_1 - \frac{1}{2}d_2 &= 0 \\
-1 \le d_i &\le 1 \quad (i = 1, 2, 3)
\end{aligned}$$

has the solution $d^4 = (0, 0, 0)^T$ (even after the elimination of the additional equations).

$$\begin{aligned}
&\min 2d_3 \\
&d_3 \ge 0 \\
-1 \le d_i &\le 1 \quad (i = 1, 2, 3)
\end{aligned}$$

also has the solution $d^4 = (0, 0, 0)^T$. The algorithm terminates with the optimal solution x^4.

Exercise 8.18

(a) The pivoting of the constraints results in Table 8.2.

Table 8.2. Pivoting of Example 8.18.

x_1	x_2	x_3	b
1	–2	0	–1
3	–1	5	12
1	–2	0	–1
0	5	5	15
1	0	2	5
0	1	1	3

Therefore,

$$x_1 = 5 - 2x_3\,,$$
$$x_2 = 3 - x_3\,,$$

and the function h in (8.22) becomes

$$h(x_3) = (5 - 2x_3)^2 + 2(3 - x_3)^2 + 3(3 - x_3)x_3\,.$$

We get

$$h'(x_3) = 2(2x_3 - 5)2 + 4(x_3 - 3) + 9 - 6x_3 = 6x_3 - 23$$

and $x_3 = \frac{23}{6}$ is the unique minimum point of h.
Therefore, $x_1 = 5 - 2\frac{23}{6} = -\frac{8}{3}$, $x_2 = 3 - \frac{23}{6} = -\frac{5}{6}$ and the solution of the exercise is

$$(x_1, x_2, x_3) = \left(-\frac{8}{3}, -\frac{5}{6}, \frac{23}{6}\right)\,.$$

(b) It holds $x_1 = 6 - x_2$ and the function h is

$$h(x_2) = (6 - x_2)^2 + \frac{1}{4}x_2^4\,.$$

Thus,

$$h'(x_2) = 2x_2 - 12 + x_2^3$$

and $h'(x_2) = 0 \Leftrightarrow x_2 = 2$. Hence h is convex, $x_2 = 2$ is the unique minimum point of h and the solution of the original problem is $(x_1, x_2) = (4, 2)$.

Exercise 8.19

The considered plane consists of all points $x \in \mathbb{R}^3$, satisfying

$$6x_1 + 3x_2 + 2x_3 = 6\,.$$

Obviously, this plane contains the three points given in the exercise. Searching for the point of the plane which is closest to the origin, yields the problem

$$\min \sqrt{x_1^2 + x_2^2 + x_3^2}$$
$$6x_1 + 3x_2 + 2x_3 = 6\,,$$

equivalent to

$$\min x_1^2 + x_2^2 + x_3^2$$
$$6x_1 + 3x_2 + 2x_3 = 6\,.$$

We can express x_1 by $x_1 = 1 - \frac{1}{2}x_2 - \frac{1}{3}x_3$ and obtain

$$\begin{aligned} h(x_2, x_3) &= \left(1 - \frac{1}{2}x_2 - \frac{1}{3}x_3\right)^2 + x_2^2 + x_3^2\,, \\ \operatorname{grad} h(x_2, x_3) &= \begin{pmatrix} 2(1 - \frac{x_2}{2} - \frac{x_3}{3})(-\frac{1}{2}) + 2x_2 \\ 2(1 - \frac{x_2}{2} - \frac{x_3}{3})(-\frac{1}{3}) + 2x_3 \end{pmatrix}, \\ Hh(x_2, x_3) &= \begin{pmatrix} 5/2 & 1/3 \\ 1/3 & 20/9 \end{pmatrix}. \end{aligned}$$

Since $Hh(x_2, x_3)$ is positive definite, h is convex and the unique minimum point satisfies

$$\operatorname{grad} h(x_2, x_3) = 0 \Leftrightarrow \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 18/49 \\ 12/49 \end{pmatrix}.$$

We now obtain

$$x_1 = 1 - \frac{1}{2}x_2 - \frac{1}{3}x_3 = 36/49$$

and the minimum point for each of the two equivalent problems is

$$(x_1, x_2, x_3) = \frac{1}{49}(36, 18, 12)\,.$$

Therefore, the requested distance is

$$\sqrt{\left(\frac{36}{49}\right)^2 + \left(\frac{18}{49}\right)^2 + \left(\frac{12}{49}\right)^2} = \frac{6}{7}\,.$$

Exercise 8.20

(a) We apply Theorem 4.12. With

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix},$$

$$h_j(x_1, \ldots, x_n) := (a_{j1}, \ldots, a_{jn}) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} - b_j \quad \text{for } j = 1, \ldots, m$$

the problem (8.21) can be written as

$$\begin{gathered} \min f(x) \\ h_j(x) = 0 \quad \text{for } j = 1, \ldots, m\,. \end{gathered}$$

Since

$$\operatorname{grad} h_j(x) = (a_{j1}, \ldots, a_{jn})^T\,,$$

the KKT conditions are

$$\operatorname{grad} f(x) + v_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{1n} \end{pmatrix} + \cdots + v_m \begin{pmatrix} a_{m1} \\ \vdots \\ a_{mn} \end{pmatrix} = 0$$
$$h_j(x) = 0 \quad \text{for } j = 1, \ldots, m,$$

which is equivalent to

$$\begin{aligned} \operatorname{grad} f(x) + A^T v &= 0 \\ Ax &= b\,. \end{aligned}$$

(b) For the problem of Exercise 8.18 (a), the KKT conditions are:

$$\begin{array}{rrrrrl} 2x_1 & & & +v_1 & +3v_2 & = 0 \\ & 4x_2 & +3x_3 & -2v_1 & -v_2 & = 0 \\ & 3x_2 & & & +5v_2 & = 0 \\ x_1 & -2x_2 & & & & = -1 \\ 3x_1 & -2x_2 & +5x_3 & & & = 12\,. \end{array}$$

The last two equations yield

$$\begin{aligned} x_1 &= 5 - 2x_3\,, \\ x_2 &= 3 - x_3\,, \end{aligned}$$

and using this to substitute x_1 and x_2 in the first three equations, we obtain the system

$$\begin{array}{rrrl} -4x_3 & +v_1 & +3v_2 & = -10 \\ -x_3 & -2v_1 & -v_2 & = -12 \\ -3x_3 & & +5v_2 & = -9 \end{array}$$

with the solution $x_3 = \frac{23}{6}$, $v_1 = \frac{23}{6}$, $v_2 = \frac{1}{2}$. Therefore, the KKT conditions have the solution

$$(x_1, x_2, x_3, v_1, v_2) = \left(-\frac{8}{3}, -\frac{5}{6}, \frac{23}{6}, \frac{23}{6}, \frac{1}{2}\right).$$

Exercise 8.22

(a) It holds

$$\operatorname{Nu}(A) = \left\{ y \in \mathbb{R}^3 \,\middle|\, \begin{pmatrix} 1 & -2 & 0 \\ 3 & -1 & 5 \end{pmatrix} y = 0 \right\}.$$

The dimension is $n - m = 3 - 2 = 1$. For example, the vector $(2, 1, -1)^T$ represents a basis of $\operatorname{Nu}(A)$ and we can choose $Z = (2, 1, -1)^T$. Since

$$\operatorname{grad} f(x) = \begin{pmatrix} 2x_1 \\ 4x_2 + 3x_3 \\ 3x_2 \end{pmatrix},$$

the first-order necessary condition

$$\begin{aligned} Z^T \operatorname{grad} f(x) &= 0 \\ Ax &= b \end{aligned}$$

becomes

$$\begin{array}{rrrl} 4x_1 & +x_2 & +3x_3 & = 0 \\ -x_1 & -2x_2 & & = -1 \\ 3x_1 & -x_2 & +5x_3 & = 12\,. \end{array}$$

This system has the unique solution $x^* = (-8/3, -5/6, 23/6)^T$.
We get

$$Z^T Hf(x^*)Z = (2, 1, -1)\begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 3 \\ 0 & 3 & 0 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} = 6 > 0\,,$$

i.e. the sufficient optimality conditions are satisfied at x^*.

(b) We get

$$\mathrm{Nu}(A) = \left\{ y \in \mathbb{R}^2 \,\middle|\, (1, 1)y = 0 \right\}$$

having dimension $n - m = 1$. The vector $(1, -1)^T$ is a basis of $\mathrm{Nu}(A)$ and we can choose $Z = (1, -1)^T$. The necessary condition

$$\begin{aligned} Z^T \operatorname{grad} f(x) &= 0 \\ Ax &= b \end{aligned}$$

becomes

$$\begin{aligned} 2x_1 - x_2^3 &= 0 \\ x_1 + x_2 &= 6 \end{aligned}$$

and has the unique solution $x^* = (4, 2)^T$. Since

$$Z^T Hf(x^*)Z = (1, -1)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = 2 > 0\,,$$

the sufficient optimality conditions are satisfied at x^*.

Exercise 8.24

Iteration 1: $k = 0$,

$$Z^T \operatorname{grad} f(x^0) = \frac{1}{\sqrt{33}}\begin{pmatrix} 4 & -2 & 3 & 2 \\ -\sqrt{11} & -\sqrt{11} & 0 & \sqrt{11} \end{pmatrix}\begin{pmatrix} 8 \\ 2 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{33}}\begin{pmatrix} 28 \\ -10\sqrt{11} \end{pmatrix} \neq 0\,,$$

$$d^0 = -ZZ^T \operatorname{grad} f(x^0) = \frac{1}{33}\begin{pmatrix} -4 & \sqrt{11} \\ 2 & \sqrt{11} \\ -3 & 0 \\ -2 & -\sqrt{11} \end{pmatrix}\begin{pmatrix} 28 \\ -10\sqrt{11} \end{pmatrix} = \frac{1}{22}\begin{pmatrix} -74 \\ -18 \\ -28 \\ 18 \end{pmatrix},$$

$$\lambda_0 = 0.2\,, \quad x^1 = \frac{1}{55}(36, 37, -28, 18)^T\,.$$

Iteration 2: $k = 1$,

$$d^1 = \frac{1}{605}(-1038, -86, -476, 86)^T,$$
$$\lambda_1 = 0.2, \quad x^2 \approx (0.3114, 0.6443, -0.6664, 0.3557)^T.$$

Iteration 3: $k = 2$,

$$d^2 \approx (-0.4576, 0.0288, -0.2432, -0.0288)^T,$$
$$\lambda_2 = 0.2, \quad x^3 \approx (0.2199, 0.6501, -0.7151, 0.3499)^T.$$

Chapter 9

Exercise 9.2

It holds

$$Q = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}, \quad Q^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix}, \quad c = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad g^k = Qx^k + c.$$

Iteration 1: $$k = 0, \quad x^0 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \quad g^0 = \begin{pmatrix} 6 \\ 16 \end{pmatrix}, \quad A_0 = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix},$$

$$A_0 Q^{-1} A_0^T = \begin{pmatrix} 3/4 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}, \quad (A_0 Q^{-1} A_0^T)^{-1} = \begin{pmatrix} 4 & 4 \\ 4 & 6 \end{pmatrix},$$

$$\begin{aligned} d^0 &= -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} 6 \\ 16 \end{pmatrix} \\ &\quad + \begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 4 & 4 \\ 4 & 6 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} 6 \\ 16 \end{pmatrix} \\ &= -\begin{pmatrix} 3 \\ 4 \end{pmatrix} + \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \\ v^0 &= -\begin{pmatrix} 4 & 4 \\ 4 & 6 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} 6 \\ 16 \end{pmatrix} = \begin{pmatrix} -16 \\ -22 \end{pmatrix}. \end{aligned}$$

Eliminating the second row of A_0, we get

$$A_0 = (-1, 1), \quad A_0 Q^{-1} A_0^T = 3/4, \quad (A_0 Q^{-1} A_0^T)^{-1} = 4/3,$$

$$\begin{aligned} d^0 &= -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} 6 \\ 16 \end{pmatrix} + \begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \frac{4}{3}(-1, 1) \begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} 6 \\ 16 \end{pmatrix} \\ &= -\begin{pmatrix} 3 \\ 4 \end{pmatrix} + \begin{pmatrix} -2/3 \\ 1/3 \end{pmatrix} = \begin{pmatrix} -11/3 \\ -11/3 \end{pmatrix}, \\ \lambda_0 &= \frac{6}{11}, \quad x^1 = x^0 + \lambda_0 d^0 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}. \end{aligned}$$

Iteration 2:

$$k=1,\quad x^0=\begin{pmatrix}0\\2\end{pmatrix},\quad g^1=\begin{pmatrix}2\\8\end{pmatrix},\quad A_1=\begin{pmatrix}-1&1\\-1&-1\end{pmatrix},$$

$$A_1Q^{-1}A_1^T=\begin{pmatrix}3/4&1/4\\1/4&3/4\end{pmatrix},\quad (A_0Q^{-1}A_0^T)^{-1}=\begin{pmatrix}3/2&-1/2\\-1/2&3/2\end{pmatrix},$$

$$\begin{aligned}d^1&=-\begin{pmatrix}1/2&0\\0&1/4\end{pmatrix}\begin{pmatrix}2\\8\end{pmatrix}\\&\quad+\begin{pmatrix}1/2&0\\0&1/4\end{pmatrix}\begin{pmatrix}-1&-1\\1&-1\end{pmatrix}\begin{pmatrix}3/2&-1/2\\-1/2&3/2\end{pmatrix}\begin{pmatrix}-1&1\\-1&-1\end{pmatrix}\begin{pmatrix}1/2&0\\0&1/4\end{pmatrix}\begin{pmatrix}2\\8\end{pmatrix}\\&=-\begin{pmatrix}1\\2\end{pmatrix}+\begin{pmatrix}1\\2\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix},\end{aligned}$$

$$v^1=-\begin{pmatrix}3/2&-1/2\\-1/2&3/2\end{pmatrix}=\begin{pmatrix}-1&1\\-1&-1\end{pmatrix}\begin{pmatrix}1/2&0\\0&1/4\end{pmatrix}\begin{pmatrix}2\\8\end{pmatrix}=\begin{pmatrix}-3\\5\end{pmatrix}.$$

Eliminating the first row of A_1 yields

$$A_1=(-1,-1), A_1Q^{-1}A_1^T=3/4, (A_1Q^{-1}A_1^T)^{-1}=4/3,$$

$$\begin{aligned}d^1&=-\begin{pmatrix}1/2&0\\0&1/4\end{pmatrix}\begin{pmatrix}2\\8\end{pmatrix}+\begin{pmatrix}1/2&0\\0&1/4\end{pmatrix}\begin{pmatrix}-1\\-1\end{pmatrix}\frac{4}{3}(-1,-1)\begin{pmatrix}1/2&0\\0&1/4\end{pmatrix}\begin{pmatrix}2\\8\end{pmatrix}\\&=-\begin{pmatrix}1\\2\end{pmatrix}+\begin{pmatrix}2\\1\end{pmatrix}=\begin{pmatrix}1\\-1\end{pmatrix},\end{aligned}$$

$$\lambda_1=1,\quad x^2=x^1+\lambda_1d^1=\begin{pmatrix}1\\1\end{pmatrix}.$$

Iteration 3:

$$k=2,\quad x^2=\begin{pmatrix}1\\1\end{pmatrix},\quad g^2=\begin{pmatrix}4\\4\end{pmatrix},\quad A_2=(-1,-1),\quad (A_2Q^{-1}A_2^T)^{-1}=\frac{4}{3},$$

$$\begin{aligned}d^2&=-\begin{pmatrix}1/2&0\\0&1/4\end{pmatrix}\begin{pmatrix}4\\4\end{pmatrix}+\begin{pmatrix}1/2&0\\0&1/4\end{pmatrix}\begin{pmatrix}-1\\-1\end{pmatrix}\frac{4}{3}(-1,-1)\begin{pmatrix}1/2&0\\0&1/4\end{pmatrix}\begin{pmatrix}4\\4\end{pmatrix}\\&=-\begin{pmatrix}2\\1\end{pmatrix}+\begin{pmatrix}2\\1\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}.\end{aligned}$$

$$v^2=-\frac{4}{3}(-1,-1)\begin{pmatrix}1/2&0\\0&1/4\end{pmatrix}\begin{pmatrix}4\\4\end{pmatrix}=4\geq 0.$$

The algorithm terminates with the optimal solution x^2.

Exercise 9.3

Let

$$A=\begin{pmatrix}a_{11}&\cdots&a_{1n}\\\vdots&&\vdots\\a_{m1}&\cdots&a_{mn}\end{pmatrix},\quad b=\begin{pmatrix}b_1\\\vdots\\b_m\end{pmatrix}.$$

We define the functions $g_i(x)$ by

$$\begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \\ g_{m+1}(x) \\ \vdots \\ g_{m+n}(x) \end{pmatrix} = \begin{pmatrix} Ax - b \\ -x \end{pmatrix}.$$

The problem (9.6) can be written as

$$\begin{aligned} &\min f(x) \\ g_i(x) \le 0 \quad &\text{for } i = 1, \dots, n + m\,. \end{aligned}$$

Since

$$\begin{aligned} \operatorname{grad} f(x) &= Qx + c\,, \\ \operatorname{grad} g_i(x) &= \begin{pmatrix} a_{i1} \\ \vdots \\ a_{in} \end{pmatrix} \quad \text{for } i = 1, \dots, m \\ \operatorname{grad} g_{m+i}(x) &= -e^i \quad \text{for } i = 1, \dots, n \end{aligned}$$

(e^i is the ith unit vector of $\in \mathbb{R}^n$), we obtain the KKT conditions (see Theorem 4.6):

$$Qx + c + u_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{1n} \end{pmatrix} + \cdots + u_m \begin{pmatrix} a_{m1} \\ \vdots \\ a_{mn} \end{pmatrix} - v_1 e^1 - \cdots - v_n e^n = 0 \tag{1}$$

$$u_i(b_i - (a_{i1}, \dots, a_{in})x) = 0 \quad \text{for } i = 1, \dots, m \tag{2}$$

$$v_i x_i = 0 \quad \text{for } i = 1, \dots, n \tag{3}$$

$$Ax \le b \tag{4}$$

$$x \ge 0 \tag{5}$$

$$u, v \ge 0\,. \tag{6}$$

From conditions (4)–(6) it follows that $u_i, v_i, x_i \ge 0$ and $b_i - (a_{i1}, \dots, a_{in})x \ge 0$, i.e. the conditions (2) and (3) are equivalent to

$$u^T(Ax - b) = \sum_{i=1}^{m} u_i(b_i - (a_{i1}, \dots, a_{in})x) = 0$$

and

$$v^T x = \sum_{i=1}^{n} v_i x_i = 0\,.$$

Therefore, the above system (1)–(6) is equivalent to (9.9).

Exercise 9.6

We get

$$Q = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 1 \end{pmatrix} \text{ (positive definite)}, \quad c = \begin{pmatrix} -1 \\ 5 \\ -2 \end{pmatrix},$$

$$A = (2, 3, -1), \quad b = 5,$$

$$M = \begin{pmatrix} Q & A^T \\ -A & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 0 & 2 \\ 1 & 3 & 1 & 3 \\ 0 & 1 & 1 & -1 \\ -2 & -3 & 1 & 0 \end{pmatrix}, \quad r = \begin{pmatrix} c \\ b \end{pmatrix} = \begin{pmatrix} -1 \\ 5 \\ -2 \\ 5 \end{pmatrix}.$$

The pivoting is illustrated in Table 9.4, yielding the optimal basic solution $(z_0, z_1, z_2, z_3, z_4, w_1, w_2, w_3, w_4) = (0, 1/2, 0, 2, 0, 0, 15/2, 0, 6)$.

The solutions of the KKT conditions are:

$$x = (z_1, z_2, z_3)^T = (1/2, 0, 2)^T, \quad u = z_4 = 0,$$
$$v = (w_1, w_2, w_3)^T = (0, 15/2, 0)^T, \quad y = w_4 = 6.$$

The problem of the exercise has the solution $x = (1/2, 0, 2)^T$.

Exercise 9.7

We obtain the problem

$$\begin{aligned} \min x_1^2 + (x_2-4)^2 + (x_1-3)^2 + (x_2-4)^2 + (x_1-3)^2 + (x_2+1)^2 \\ x_1 + x_2 \le 3 \\ x_1, x_2 \ge 0. \end{aligned}$$

The additive constant in the objective function can be ignored. Therefore,

$$Q = \begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix}, \quad c = \begin{pmatrix} -12 \\ -14 \end{pmatrix}, \quad A = (1, 1), \quad b = 3,$$

$$M = \begin{pmatrix} 6 & 0 & 1 \\ 0 & 6 & 1 \\ -1 & -1 & 0 \end{pmatrix}, \quad r = \begin{pmatrix} -12 \\ -14 \\ 3 \end{pmatrix}.$$

The pivot steps are listed in Table 9.5 The optimal coordinates of the refinery D are $x_1 = z_1 = 4/3$ and $x_2 = z_2 = 5/3$.

Table 9.4. Lemke's method.

BV	z_0	z_1	z_2	z_3	z_4	w_1	w_2	w_3	w_4	r
w_1	–1	–2	–1	0	–2	1	0	0	0	–1
w_2	–1	–1	–3	–1	–3	0	1	0	0	5
w_3	–1	0	–1	–1	1	0	0	1	0	–2
w_4	–1	2	3	–1	0	0	0	0	1	5
w_1	0	–2	0	1	–3	1	0	–1	0	1
w_2	0	–1	–2	0	–4	0	1	–1	0	7
z_0	1	0	1	1	–1	0	0	–1	0	2
w_4	0	2	4	0	–1	0	0	–1	1	7
z_3	0	–2	0	1	–3	1	0	–1	0	1
w_2	0	–1	–2	0	–4	0	1	–1	0	7
z_0	1	2	1	0	2	–1	0	0	0	1
w_4	0	2	4	1	–1	0	0	–1	1	7
z_3	1	0	–1	1	–1	0	0	–1	0	2
w_2	$1/2$	0	$-3/2$	0	–3	$-1/2$	1	–1	0	$15/2$
z_1	$1/2$	1	$-1/2$	0	1	$-1/2$	0	0	0	$1/2$
w_4	–1	0	3	1	–3	1	0	–1	1	6

Table 9.5. Lemke's method.

BV	z_0	z_1	z_2	z_3	w_1	w_2	w_3	r
w_1	-1	-6	0	-1	1	0	0	-12
w_2	-1	0	-6	-1	0	1	0	-14
w_3	-1	1	1	0	0	0	1	3
w_1	0	-6	6	0	1	-1	0	2
z_0	1	0	6	1	0	-1	0	14
w_3	0	1	7	1	0	-1	1	17
z_2	0	-1	1	0	$1/6$	$-1/6$	0	$1/3$
z_0	1	6	0	1	-1	0	0	12
w_3	0	8	0	1	$-7/6$	$1/6$	1	$44/3$
z_2	0	0	1	$1/8$	$1/48$	$-7/48$	$1/8$	$13/6$
z_0	1	0	0	$1/4$	$-1/8$	$-1/8$	$-3/4$	1
z_1	0	1	0	$1/8$	$-7/48$	$1/48$	$1/8$	$11/6$
z_2	$-1/2$	0	1	0	$1/12$	$-1/12$	$1/2$	$5/3$
z_0	4	0	0	1	$-1/2$	$-1/2$	-3	4
z_1	$-1/2$	1	0	0	$-1/12$	$1/12$	$1/2$	$4/3$

Chapter 10

Exercise 10.7

(a) It holds

$$x_k = \frac{2r_k+1}{2r_k+4} = \frac{2\cdot 10^k+1}{2\cdot 10^k+4}\,,$$
$$p(x_k) = (g^+(x_k))^2 = (1-\frac{2r_k+1}{2r_k+4})^2 = \frac{9}{(2r_k+4)^2} = \frac{9}{(2\cdot 10^k+4)^2}\,.$$

Thus,

$$r_k p(x_k) \le 10^{-9} \Leftrightarrow \frac{9\cdot 10^k}{(2\cdot 10^k+4)^2} \le 10^{-9} \Leftrightarrow k \ge 10\,,$$

i.e. the algorithm terminates in the iteration $k = 10$.

(b) It holds

$$x^k = \frac{1}{r_k+1}\begin{pmatrix} r_k-1 \\ 2r_k-2 \end{pmatrix} = \frac{1}{10^k+1}\begin{pmatrix} 10^k-1 \\ 2\cdot 10^k-2 \end{pmatrix},$$
$$p(x^k) = (g_1^+(x^k))^2 + (g_2^+(x^k))^2 = (1-\frac{r_k-1}{r_k+1})^2 + (2-\frac{2r_k-2}{r_k+1})^2$$
$$= \frac{20}{(r_k+1)^2} = \frac{20}{(10^k+1)^2}\,.$$

Therefore,

$$r_k p(x_k) \le 10^{-9} \Leftrightarrow \frac{20\cdot 10^k}{(10^k+1)^2} \le 10^{-9} \Leftrightarrow k \ge 11\,,$$

i.e. the algorithm terminates in the iteration $k = 11$.

Exercise 10.8

(a) We get

$$g(x_1,x_2) = x_1 - x_2 + 1\,,$$
$$g^+(x_1,x_2) = \begin{cases} x_1 - x_2 + 1 & \text{for } x_2 \le x_1 + 1\,, \\ 0 & \text{for } x_2 > x_1 + 1\,, \end{cases}$$
$$q(x_1,x_2,r) = f(x_1,x_2) + r[g^+(x_1,x_2)]^2$$
$$= \begin{cases} (x_1-1)^2 + x_2^2 + r(x_1-x_2+1)^2 & \text{for } x_2 \le x_1 + 1\,, \\ (x_1-1)^2 + x_2^2 & \text{for } x_2 > x_1 + 1\,. \end{cases}$$

Since the function q is continuously differentiable and convex, a global minimum point satisfies the system $\operatorname{grad}_x q(x_1, x_2, r) = 0$. It holds

$$\frac{\partial}{\partial x_1} q(x_1, x_2, r) = \begin{cases} 2(x_1 - 1) + 2r(x_1 - x_2 + 1) & \text{for } x_2 \le x_1 + 1\,, \\ 2(x_1 - 1) & \text{for } x_2 > x_1 + 1\,, \end{cases}$$

$$\frac{\partial}{\partial x_2} q(x_1, x_2, r) = \begin{cases} 2x_2 - 2r(x_1 - x_2 + 1) & \text{for } x_2 \le x_1 + 1\,, \\ 2x_2 & \text{for } x_2 > x_1 + 1\,. \end{cases}$$

For $x_2 > x_1 + 1$ the system $\operatorname{grad}_x q(x_1, x_2, r) = 0$ has no solution, since $x_2 = 0$ and $x_1 = 1$ contradict $x_2 > x_1 + 1$. Therefore, $\operatorname{grad}_x q(x_1, x_2, r) = 0$ implies

$$2(x_1 - 1) + 2r(x_1 - x_2 + 1) = 0\,,$$
$$2x_2 - 2r(x_1 - x_2 + 1) = 0\,.$$

The unique solution of this system is the point

$$x(r) = \frac{1}{1 + 2r} \begin{pmatrix} 1 \\ 2r \end{pmatrix}.$$

The problem of the exercise has the solution

$$\lim_{r \to \infty} x(r) = (0, 1)^T\,.$$

(b) We have the constraint

$$h(x_1, x_2) := x_2 - x_1^2 = 0\,.$$

Hence,

$$q(x_1, x_2, r) = f(x_1, x_2) + r[h(x_1, x_2)]^2 = x_1 + x_2 + r(x_2 - x_1^2)^2\,.$$

The first-order necessary conditions for a minimum point of q are

$$\frac{\partial}{\partial x_1} q(x_1, x_2, r) = 1 - 2r(x_2 - x_1^2)2x_1 = 0\,, \tag{1}$$

$$\frac{\partial}{\partial x_2} q(x_1, x_2, r) = 1 + 2r(x_2 - x_1^2) = 0\,. \tag{2}$$

The condition (2) is equivalent to

$$2r(x_2 - x_1^2) = -1\,, \tag{3}$$

or

$$x_2 = x_1^2 - \frac{1}{2r}\,. \tag{4}$$

By substituting $2r(x_2 - x_1^2)$ in (1) for -1 (see (3)), we obtain

$$1 + 2x_1 = 0$$

or

$$x_1 = \frac{1}{2}. \tag{5}$$

Finally, (4) and (5) result in

$$x_2 = \frac{1}{4} - \frac{1}{2r}. \tag{6}$$

Therefore, the unique "candidate" for a minimum point of q is

$$x^* = \left(-\frac{1}{2}, \frac{1}{4} - \frac{1}{2r}\right)^T.$$

The Hessian matrix of q with respect to x is

$$H_x q(x, r) = \begin{pmatrix} 12rx_1^2 - 4rx^2 & -4rx_1 \\ -4rx_1 & 2r \end{pmatrix},$$

hence,

$$H_x q(x^*, r) = \begin{pmatrix} 2r + 2 & 2r \\ 2r & 2r \end{pmatrix}.$$

Since this matrix is positive definite for $r > 0$, x^* minimizes the function q. The problem of the exercise has the solution

$$\lim_{r\to\infty} \begin{pmatrix} -1/2 \\ 1/4 - 1/(2r) \end{pmatrix} = \begin{pmatrix} -1/2 \\ 1/4 \end{pmatrix}.$$

Exercise 10.9

(a) The KKT conditions of problem (10.7) are

$$\begin{aligned} f'(x) - u &= 0 \\ ux &= 0 \\ x &\geq 0 \\ u &\geq 0. \end{aligned}$$

The solution of (10.7) is $x^* = 1 + 1/\sqrt{3}$ (see Figure 10.3). By substituting x^* in the above system, we get $u = 0$, i.e. $u^* = 0$ is the Lagrange multiplier corresponding to the solution x^*. We have to show that

$$u^{(k)} := \beta r_k [g^+(x_k)]^{\beta-1} = 2r_k g^+(x_k)$$

converges to $u^* = 0$. We have seen in Example 10.3 that the minimum point of the function $q(x, r_k)$ is $x_k = 1 + 1/\sqrt{3}$ for a sufficiently large k. Therefore, $g^+(x_k) = 0$ for such a k and $\lim_{k\to\infty} u^{(k)} = 0$.

(b) The KKT conditions of (10.9) are:

$$\begin{aligned} 2(x_1+1) - u_1 &= 0 \\ 2(x_2+2) - u_2 &= 0 \\ u_1(1-x_1) &= 0 \\ u_2(2-x_2) &= 0 \\ x_1 &\geq 1 \\ x_2 &\geq 2 \\ u_1, u_2 &\geq 0\,. \end{aligned}$$

The optimal point of (10.9) is $x^* = (1,2)^T$. By substituting x^* in the above system we obtain $u_1^* = 4$ and $u_2^* = 8$. It holds

$$\begin{aligned} x^k &= \frac{1}{r_k+1}\begin{pmatrix} r_k - 1 \\ 2r_k - 2 \end{pmatrix}, \\ g_1^+(x^k) &= 1 - \frac{r_k-1}{r_k+1} = \frac{2}{r_k+1}, \\ g_2^+(x^k) &= 2 - \frac{2r_k-2}{r_k+1} = \frac{4}{r_k+1}\,. \end{aligned}$$

Therefore,

$$\begin{aligned} u_1^{(k)} &:= \beta r_k [g_1^+(x^k)]^{\beta-1} = \frac{4r_k}{r_k+1}, \\ u_2^{(k)} &:= \beta r_k [g_2^+(x^k)]^{\beta-1} = \frac{8r_k}{r_k+1}\,. \end{aligned}$$

Obviously, $u_1^{(k)}$ and $u_2^{(k)}$ converge to $u_1^* = 4$ and $u_2^* = 8$, respectively, if $k \to \infty$.

Exercise 10.10

We show first that $g^+(x)$ is convex, if $g(x)$ is convex. Therefore we have to prove the following (see Definition 3.14):

$$g^+((1-\lambda)x^1 + \lambda x^2) \leq (1-\lambda)g^+(x^1) + \lambda g^+(x^2) \tag{1}$$

for $x^1, x^2 \in \mathbb{R}^n$ and $0 < \lambda < 1$.

If $g((1-\lambda)x^1 + \lambda x^2) < 0$, the left side of (1) is zero and (1) is satisfied.

If $g((1-\lambda)x^1 + \lambda x^2) \geq 0$, the convexity of g implies

$$\begin{aligned} g^+((1-\lambda)x^1 + x^2) = g((1-\lambda)x^1 + \lambda x^2) &\leq (1-\lambda)g(x^1) + \lambda g(x^2) \\ &\leq (1-\lambda)g^+(x^1) + \lambda g^+(x^2)\,. \end{aligned}$$

Therefore, if all functions f, g_i and h_j are convex, the functions g_i^+ are also convex. Moreover, the functions $[g_i^+(x)]^\beta$ and $[h_j(x)]^\beta$ are convex for $\beta = 2$ and $\beta = 4$ (see Exercise 3.15). Finally, the Theorems 3.16 and 3.17 imply that $q(x, r)$ is convex.

Exercise 10.13

(a) The KKT conditions are

$$\begin{aligned}2(x_1+2)-u_1&=0\\2(x_2+4)-u_2&=0\\u_1(1-x_1)&=0\\u_2(2-x_2)&=0\\x_1&\geq 1\\x_2&\geq 2\\u_1,u_2&\geq 0\,.\end{aligned}$$

Setting $x_1=1, x_2=2$ we obtain the Lagrange multipliers $u_1^*=4, u_2^*=8$.

(b) It holds (compare with Example 10.4):

$$\begin{aligned}q(x,r)&=f(x)+r(g_1^+(x)+g_2^+(x))\\&=\begin{cases}(x_1+1)^2+(x_2+2)^2+r(1-x_1)+r(2-x_2) & \text{for } x_1\leq 1, x_2\leq 2\\(x_1+1)^2+(x_2+2)^2+r(1-x_1) & \text{for } x_1\leq 1, x_2>2\\(x_1+1)^2+(x_2+2)^2+r(2-x_2) & \text{for } x_1>1, x_2\leq 2\\(x_1+1)^2+(x_2+2)^2 & \text{for } x_1>1, x_2>2\,.\end{cases}\end{aligned} \tag{1}$$

(c) Since $q(x):=q(x,r)$ is not differentiable at the point x^*, we cannot apply Corollary 3.50 to prove that x^* is the minimum point. Below we show that for $r\geq 8$, the unilateral derivative $D^+q(x^*,y)$ is nonnegative for all $y\in\mathbb{R}^2$. As it was shown in Exercise 3.62, this implies that x^* is a global minimum point of $q(x)$.
It holds

$$D^+q(x^*,y)=\lim_{t\to 0+}\frac{q(x^*+ty)-q(x^*)}{t}\,.$$

For $t>0$ we obtain

$$\begin{aligned}q(x^*+ty)&=q(1+ty_1,2+ty_2)\\&=\begin{cases}(2+ty_1)^2+(4+ty_2)^2-rty_1-rty_2 & \text{for } y_1\leq 0, y_2\leq 0\\(2+ty_1)^2+(4+ty_2)^2-rty_1 & \text{for } y_1\leq 0, y_2>0\\(2+ty_1)^2+(4+ty_2)^2-rty_2 & \text{for } y_1>0, y_2\leq 0\\(2+ty_1)^2+(4+ty_2)^2 & \text{for } y_1>0, y_2>0\end{cases}\end{aligned} \tag{2}$$

$$q(x^*)=(1+1)^2+(2+2)^2=20\,.$$

So we obtain for the case $y_1\leq 0, y_2\leq 0$:

$$\begin{aligned}D^+q(x^*,y)&=\lim_{t\to 0+}\frac{4+4ty_1+t^2y_1^2+16+8ty_2+t^2y_2^2-rt(y_1+y_2)-20}{t}\\&=\lim_{t\to 0+}[4y_1+ty_1^2+8y_2+ty_2^2-r(y_1+y_2)]\\&=(4-r)y_1+(8-r)y_2\geq 0\quad(\text{since } r\geq 8)\,.\end{aligned}$$

Similarly we obtain $D^+q(x^*, y) \geq 0$ in the other cases of (2).
Alternatively, the part (c) of the exercise can be solved as follows: the function $f(x_1) := (x_1+1)^2 + r(1-x_1)$ is decreasing for $x_1 \leq 1$, since $f'(x_1) = 2x_1 + 2 - r < 0$ (note that $r > 8$), therefore,

$$(x_1+1)^2 + r(1-x_1) \geq f(1) = 4 \quad \text{for } x_1 \leq 1\,. \tag{3}$$

Similarly, the function $g(x_2) := (x_2+2)^2 + r(2-x_2)$ is decreasing for $x_2 \leq 2$, since $g'(x_2) = 2x_2 + 4 - r < 0$, therefore

$$(x_2+2)^2 + r(2-x_2) \geq g(2) = 16 \quad \text{for } x_2 \leq 2\,. \tag{4}$$

Moreover, we have the relations

$$(x_1+1)^2 \geq 4 \quad \text{for } x_1 > 1\,, \tag{5}$$

$$(x_2+2)^2 \geq 16 \quad \text{for } x_2 > 2\,. \tag{6}$$

The inequalities (3)–(6) imply that $q(x) \geq q(x^*) = 20$ for all $x \in \mathbb{R}^2$ (see (1)).

Exercise 10.18
We obtain

$$\begin{aligned} b(x) &= \frac{1}{x_1-1} + \frac{1}{x_2-2}\,, \\ s(x,c) &= (x_1+1)^2 + (x_2+2)^2 + \frac{1}{c}\left(\frac{1}{x_1-1} + \frac{1}{x_2-2}\right). \end{aligned}$$

Since $s(x,c)$ is continuously differentiable and convex for $c > 0, x_1 > 1, x_2 > 2$, the unique global minimum point $x(c)$ satisfies

$$\operatorname{grad}_x s(x,c) = \begin{pmatrix} 2(x_1+1) - \frac{1}{c(x_1-1)^2} \\ 2(x_2+2) - \frac{1}{c(x_2-2)^2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Therefore, the components of $x(c) = \binom{x_1(c)}{x_2(c)}$ satisfy the equations

$$\begin{aligned} (x_1(c)+1)(x_1(c)-1)^2 &= \frac{1}{2c}\,, \\ (x_2(c)+2)(x_2(c)-2)^2 &= \frac{1}{2c}\,, \end{aligned}$$

respectively. The equations can be solved, e.g. with Newton's method. Some points $x(c)$ are calculated in Table 10.3.

The sequence of points $x(c)$ converges to the optimal point $x^* = (1,2)^T$ if $c \to \infty$ (see Figure 10.10).

Table 10.3. Barrier method.

c	$x_1(c)$	$x_2(c)$
1	1.4516	2.3394
2	1.3277	2.2427
5	1.2126	2.1551
10	1.1524	2.1103
20	1.1089	2.0783
100	1.0494	2.0352
1000	1.0157	2.0112
10 000	1.0050	2.0035
100 000	1.0016	2.0011

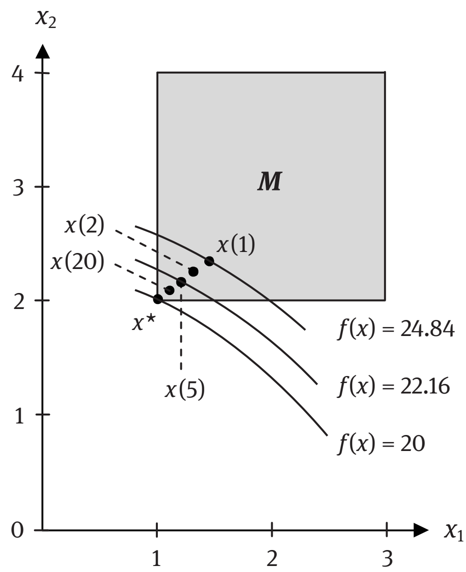


Fig. 10.10. Solution process.

Exercise 10.19
It holds

$$
\begin{aligned}
g(x) &= x_1 - x_2 + 1\,,\\
b(x) &= \frac{1}{x_2 - x_1 - 1}\,,\\
s(x, c) &= (x_1 - 1)^2 + x_2^2 + \frac{1}{c(x_2 - x_1 - 1)}\,.
\end{aligned}
$$

The unique global minimum point of $s(x, c)$ satisfies

$$
\operatorname{grad}_x s(x, c) = \begin{pmatrix} 2(x_1 - 1) + \frac{1}{c(x_2 - x_1 - 1)^2} \\ 2x_2 - \frac{1}{c(x_2 - x_1 - 1)^2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{1}
$$

Adding of the two equations yields

$$
\begin{aligned}
2(x_1 - 1) + 2x_2 = 0 \quad &\Rightarrow \\
x_2 = 1 - x_1\,. &
\end{aligned} \tag{2}
$$

By substituting x_2 in the second equation of (1) by $1 - x_1$, we obtain

$$
\begin{aligned}
2(1 - x_1) - \frac{1}{c \cdot 4x_1^2} = 0 \quad &\Rightarrow \\
x_1^2 - x_1^3 = \frac{1}{8c}\,. &
\end{aligned} \tag{3}
$$

With (2) and (3) we can calculate the minimum point $x(c) = \binom{x_1(c)}{x_2(c)}$:

Table 10.4. Barrier method.

c	$x_1(c)$	$x_2(c)$
1	−0.309017	1.309017
10	−0.106297	1.106297
100	−0.034756	1.034756
1000	−0.011119	1.011119
10 000	−0.003529	1.003529
100 000	−0.001117	1.001117
1 000 000	−0.000353	1.000353

The sequence of points $x(c)$ converges to the optimal point $x^* = (0, 1)^T$ if $c \to \infty$.

Exercise 10.22

The Lagrange function is

$$\Phi(x_1, x_2, u) = -x_1 + x_2 + u(x_1^2 - x_2 + 1)\,.$$

Thus

$$\begin{aligned} \operatorname{grad}_x \Phi(x_1, x_2, u) &= \begin{pmatrix} -1 + 2x_1 u \\ 1 - u \end{pmatrix}, \\ H_x \Phi(x_1, x_2, u) &= \begin{pmatrix} 2u & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Step (1)

$$x^0 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \quad B_0 = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \quad k := 0\,.$$

Step (2)

Since $\operatorname{grad} f(x^0) = \binom{-1}{1}$, $\operatorname{grad} g(x^0) = \binom{0}{-1}$ and $g(x^0) = -1$, we obtain the quadratic problem

$$\left.\begin{aligned} &\min -d_1 + d_2 + d_1^2 \\ &-d_2 - 1 \le 0 \end{aligned}\right\} \quad QP(x^0, B_0)\,.$$

The KKT conditions

$$\begin{aligned} 2d_1 - 1 &= 0 \\ 1 - u &= 0 \\ u(-d_2 - 1) &= 0 \\ -d_2 - 1 &\le 0 \\ u &\ge 0 \end{aligned}$$

have the unique solution $d_1 = 1/2$, $d_2 = -1$, $u = 1$. Therefore, $d^0 = \binom{1/2}{-1}$, $u^0 = 1$.

Step (3)

$$x^1 = x^0 + d^0 = \begin{pmatrix} 1/2 \\ 1 \end{pmatrix}.$$

Step (4)

$$\begin{aligned} y^0 &= \text{grad}_x\ \Phi(x^1, u^0) - \text{grad}_x\ \Phi(x^0, u^0) \\ &= \begin{pmatrix} -1+1 \\ 0 \end{pmatrix} - \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ p^0 &= \begin{pmatrix} 1/2 \\ -1 \end{pmatrix}, \\ B_1 &= \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} + \frac{\binom{1}{0}(1,0)}{(1,0)\binom{1/2}{-1}} - \frac{\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}\binom{1/2}{-1}(\frac{1}{2}, -1)\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}}{(\frac{1}{2}, -1)\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}\binom{1/2}{-1}} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \\ k &:= 1. \end{aligned}$$

Step (2)
We obtain

$$\left.\begin{aligned} &\min -d_1 + d_2 + d_1^2 \\ &d_1 - d_2 + 1/4 \leq 0 \end{aligned}\right\} \quad QP(x^1, B_1).$$

The KKT conditions

$$\begin{aligned} 2d_1 - 1 + u &= 0 \\ 1 - u &= 0 \\ u(d_1 - d_2 + 1/4) &= 0 \\ d_1 - d_2 + 1/4 &\leq 0 \\ u &\geq 0 \end{aligned}$$

have the solution $d_1 = 0, d_2 = 1/4, u = 1$. Therefore, $d_1 = \binom{0}{1/4}, u_1 = 1$.

Step (3)

$$x^2 = x^1 + d^1 = \begin{pmatrix} 1/2 \\ 5/4 \end{pmatrix} = x^*.$$

References

[1] Aarts, E.; Lenstra, J. K. (Ed.): *Local Search in Combinatorial Optimization*. Chichester: John Wiley and Sons, 1997.

[2] Abou-Taleb, N.; Megahed, I.; Moussa, A.; Zaky, A.: A New Approach to the Solution of Economic Dispatch Problems, *Winter Power Meeting*, New York, 1974.

[3] Avriel, M. *Nonlinear Programming*: Analysis and Methods. Englewood Cliffs, N. J.: Prentice Hall, 2003.

[4] Bazaraa, M. S.; Sherali, H. D.; Shetty, C. M.: *Nonlinear Programming: Theory and Applications.* New York: John Wiley and Sons, 2006.

[5] Bazaraa, M. S.; Shetty, C. M. *Nonlinear Programming - Theory and Algorithms*. New York: John Wiley and Sons, 1979.

[6] Bertsekas, D. P.: *Nonlinear Programming*. 2. ed. Massachusetts: Athena Scientific, 2004.

[7] Bracken, J.; McCormick, G. P.: *Selected Applications of Nonlinear Programming*. New York: John Wiley and Sons, 1968.

[8] Camp, G. D.: Inequality-Constrained Stationary-Value Problems, *Operations Research*, 3, pp. 548–550, 1955.

[9] Canon, M. D.; Cullum, C.; Polak, E.: *Theory of Optimal Control and Mathematical Programming*. New York: McGraw-Hill, 1970.

[10] Canon, M. D.; Eaton, J. H.: A New Algorithm for a Class of Quadratic Programming Problems, with Application to Control, *Siam J Control*, 4, pp. 34–44, 1966.

[11] Carroll, C. W.: The Created Response Surface Technique for Optimizing Nonlinear Restrained Systems, *Operations Research*, 9, pp. 169–184, 1961.

[12] Clarke, F. H.: *Optimization and Nonsmooth Analysis*. New York: John Wiley and Sons, 1983.

[13] Conn, A. R.; Gould, N. I. M.; Toint, Ph. L.: *Trust-region methods*. Siam/Mps *Series on Optimization*, Siam, Philadelphia, 2000.

[14] Da Silva, I. N.; Nepomuceno, L.; Bastos, T. M.: Designing a Modified Hopfield Network to Solve an Economic Dispatch Problem with Nonlinear Cost Function, *Ieee Trans. on Power Systems*, 2002.

[15] Dahlquist, G.; Björck, A.: *Numerical Methods*, Dover Publications, 2003.

[16] Davidon, W. C.: Variable Metric Method for Minimization, *Aec Research Development Report*, Anl-5990, 1959.

[17] Fletcher, R.: *Practical Methods of Optimization*. New York: John Wiley and Sons, 1987.

[18] Fletcher, R.: An Exact Penalty Function for Nonlinear Programming with Inequalities. In: *Mathematical Programming*, 5, pp. 129–150, 1973.

[19] Fletcher, R.; Powell, M. A.: Rapidly Convergent Descent Method for Minimization, *Computer Journal*, 6, pp. 163–168, 1963.

[20] Fletcher, R.; Reeves, C.: Function Minimization by Conjugate Gradients, *Computer Journal*, 7, pp. 149–154, 1964.

[21] Fox, R. L.: *Optimization Methods for Engineering Design*, Addison-Wesley, Reading, Mass., 1971.

[22] Friedlander, A.: *Elementos de Programação Não Linear*. Campinas: Editora da Universidade Estadual de Campinas, 1994.

[23] Fritzsche, H.: *Programação Não Linear: análise e métodos*. São Paulo: Editora da Universidade de São Paulo, 1978.

[24] Geoffrion, A. M.: Duality in Nonlinear Programming: A Simplified Applications-Oriented Development, *Siam Review*, 13, pp. 1–37, 1971.

[25] Gill, P.; Murray, W.; Wright, M. H.: *Practical Optimization*. New York: Academic Press, 1981.

[26] Glover, F.; Laguna, M.: *Tabu Search*. Boston; London: Kluwer Academic Publishers, 1995.

[27] Goffin, I. L.; Rousseau, I. M. (Ed.): *Mathematical Programming Study*, 20: Applications, 1982.

[28] Goldbarg, M. C.; Luna, H. P.: *Otimização Combinatória e Programação Linear*: Modelos e Algoritmos. Rio de Janeiro: Campus, 2000.

[29] Gould, N. I. M.; Leyffer, S.: An introduction to algorithms for nonlinear optimization. In: Blowey, J. F.; Craig, A. W.; Shardlow, T.: *Frontiers in Numerical Analysis*, 109–197, Springer Verlag, Berlin, 2003. http://www.numerical.rl.ac.uk/nimg/msc/lectures/paper.pdf.

[30] Haimes, Y. Y.: *Hierarchical Analyses of Water Resources Systems*: Modeling and Optimization of Large-Scale Systems. New York: McGraw-Hill, 1977.

[31] Held, M.; Karp, R. M.: The Travelling Salesman Problem and Minimum Spanning Trees, *Operations Research*, 18, pp. 1138–1162, 1970.

[32] Held, M.; Wolfe P.; Crowder, H.: Validation of Subgradient Optimization. In: *Mathematical Programming*, 6, pp. 62–88, 1974.

[33] Himmelblau, D. M.: *Applied Nonlinear Programming*. New York: McGraw-Hill, 1972.

[34] Hock, W.; Schittkowski, K.: Test Examples for Nonlinear Programming Codes. *Lecture Notes in Economics and Mathematical Systems*, 187, Springer, New York, 1981.

[35] Holland, J. H.: *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, Ann Arbor, Michigan, 1975.

[36] Horst, R.; Pardalos, P. M. (Ed.): *Handbook of Global Optimization*. Boston; London: Kluwer Academic Publishers, 1995.

[37] Horst, R.; Tuy, H. (Ed.): *Global Optimization*: Deterministic Approaches. Berlin; New York: Springer, 1996.

[38] Johnson, R. C.: *Mechanical Design Synthesis with Optimization Examples*. New York: Van Nostrand Reinhold, 1971.

[39] Karush, W.: *Minima of Functions of Several Variables with Inequalities as Side Conditions*. M S Thesis, Department of Mathematics, University of Chicago, 1939.

[40] Kuhn, H. W.; Tucker, A. W.: Nonlinear Programming. In: Neyman, J. (Ed.): *Proceedings 2nd Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, Calif.: University of California Press, 1951.

[41] Kuhn, H. W.: Nonlinear Programming. A Historical View. In: Cottle, R. W.; Lemke, C. E. (Eds.): *Nonlinear Programming*, 1976. pp. 1–26.

[42] Larsson, T.; Patriksson, M.; Strömberg, A. B.: Conditional Subgradient Optimization - Theory and Applications. *European Journal of Operational Research*, 88, pp. 382–403, 1996.

[43] Lemarechal, C.: Nondifferentiable Optimization. Nemhauser, G. L.; Rinnooy Kan, A. H. G.; Todd, M. J. (Eds.): *Handbooks in Operations Research and Management Science*, v. 1. Elsevier Science, Amsterdam, pp. 529–572, 1989.

[44] Lemke, C. E.: On Complementary Pivot Theory. In: Dantzig, G. B.; Veinott, A. F. (Eds.): *Mathematics of the Decision Sciences*. Providence, American Mathematical Society, 1968.

[45] Luenberger, D. G.: *Linear and Nonlinear Programming*, Reading, Massachusetts: Addison-Wesley Publishing Company, 2003.

[46] Mahey, P.: *Programação Não Linear*: Introdução à teoria e aos métodos. Rio de Janeiro: Campus, 1987.

[47] Mateus, G. R.; Luna, H. P. L.: *Programação Não Linear*. Belo Horizonte: V Escola de Computação, 1986.

[48] McCormick, G. P.: Optimality Criteria in Nonlinear Programming. In: Cottle, R. W.; Lemke, C. E. (Eds.): *Nonlinear Programming*. American Mathematical Society, 1976. pp. 27–38.

[49] Michalewicz, Z.; Vignaux, G. A.; Hobbs, M.: A Nonstandard Genetic Algorithm for the Nonlinear Transportation Problem. *Orsa Journal on Computing* 3(4), pp. 307–316, 1991.

[50] Minkowski, H.: *Gesammelte Abhandlungen*. Berlin: Teubner, 1911.

[51] Nocedal, J.; Wright, S. T.: *Numerical Optimization*. Berlin; New York: Springer, 1999. (Springer Series in Operations Research).

[52] Oliveira, A. C. M.; Lorena, L. A. N.: *Algoritmos evolutivos para problemas de otimização numérica com restrições*. São José dos Campos: Instituto Nacional de Pesquisas Especiais, 2002: http://www.lac.inpe.br/~lorena/alexandre/lworcap-Alexandre-lorena.pdf.

[53] Pietrgykowsky, T.: Application of the Steepest Descent Method to Concave Programming, *Proceedings of International Federation of Information Processing Societies Congress*, (Munich), North-Holland, Amsterdam, pp. 185–189, 1962.

[54] Pintér, J. D.: *Global Optimization in Action*. Boston; London: Kluwer Academic Publishers, 1996.

[55] Porto, R. L. L. (Org.). *Técnicas quantitativas para o gerenciamento de recursos hídricos*. Rio Grande do Sul: Editora da Universidade Federal do Rio Grande do Sul, 2002.

[56] Rockafellar, R. T.: *Convex Analysis*. Princeton, NJ: Princeton University Press, 1996.

[57] Rosen, J. B.: The Gradient Projection Method for Nonlinear Programming, Part I, Linear Constraints, *Siam J Applied Mathematics* 8, pp. 181–217, 1960.

[58] Rosen, J. B.: The Gradient Projection Method for Nonlinear Programming, Part Ii, Nonlinear Constraints *Siam J Applied Mathematics* 9, pp. 514 - 553, 1961.

[59] Sasson, A. M.; Merrill, H. M.: Some Applications of Optimization Techniques to Power Systems Problems, *Proceedings of the Ieee*, 62, pp. 959–972, 1974.

[60] Schittkowski, K.: Nonlinear Programming Codes, *Lecture Notes in Economics and Mathematical Systems* 183. Springer, New York, 1980.

[61] Schittkowski, K. (Ed.): *Computational Mathematical Programming*, Nato-Asi-Series. Springer, New York, 1985.

[62] Still, G.: *Semi-infinite Programming*: An Introduction: http://www.home.math.utwente.nl/~stillgj/sip.html.

[63] Topkins, D. M.; Veinott, A. F.: On the Convergence of Some Feasible Direction Algorithms for Nonlinear Programming, *Siam J Control*, 5, pp. 268–279, 1967.

[64] Van Laarhoven, P. J. M.; Aarts, E. H. L.: *Simulated Annealing*: Theory and Applications. Boston; London: Kluwer Academic Publishers, 1987.

[65] Wilde, D. J.: *Optimum Seeking Methods*. Englewood Cliffs, NJ: Prentice Hall, 1964.

[66] Wilde, D. J.; Beightler, C. S.: *Foundations of Optimization*. Englewood Cliffs, NJ: Prentice Hall, 1967.

[67] Wolfe, P.: On the Convergence of Gradient Methods Under Constraints, *Ibm J Research and Development*, 16, pp. 407–411, 1972.

[68] Wolfe, P.: A method of conjugate subgradients for minimizing nondifferentiable functions, *Mathematical Programming Study* 3, 1975, pp. 145–173.

[69] Wolfe, P.: A Method of Conjugate Subgradients for Minimizing Nondifferentiable Functions. In: Balinski, M. L.; Wolfe, P. (Eds.): *Nondifferentiable Optimization*, 1976.

[70] Yu, W.; Haimes, Y. Y.: Multi-level Optimization Conjunctive Use of Ground Water and Surface Water, *Water Resources Research*, 10, pp. 625–636, 1974.

[71] Zoutendijk, G. *Methods of Feasible Directions*. Princeton, NJ: Elsevier, Amsterdam and D Van Nostrand, 1960.

[72] Zowe, J.: Nondifferentiable Optimization. In: Schittkowsky, K. (Ed.): *Computational Mathematical Programming*. Nato Asi Series, Vol.15, Springer, Berlin, New York, 1985, pp. 323–356.

Index